Madrid

International Transport Forum

OECD

18th International Symposium on Transport Economics and Policy

The future for interurban passenger transport

Bringing citizens closer together

International Transport Forum

OECD

18th International Symposium on Transport Economics and Policy

16-18 November 2009
MADRID

The future for interurban passenger transport

Bringing citizens closer together

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 30 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

ISBN 978-92-821-0265-7 (print)
ISBN 978-92-821-0268-8 (PDF)
DOI 10.1787/9789282102688-en

Also available in French: *Les perspectives du transport interurbain de personnes: Rapprocher les citoyens*

Corrigenda to OECD publications may be found on line at: *www.oecd.org/publishing/corrigenda*.

INTERNATIONAL TRANSPORT FORUM

The International Transport Forum is an inter-governmental body within the OECD family. The Forum is a global platform for transport policy makers and stakeholders. Its objective is to serve political leaders and a larger public in developing a better understanding of the role of transport in economic growth and the role of transport policy in addressing the social and environmental dimensions of sustainable development. The Forum organises a Conference for Ministers and leading figures from civil society each May in Leipzig, Germany.

The International Transport Forum was created under a Declaration issued by the Council of Ministers of the ECMT (European Conference of Ministers of Transport) at its Ministerial Session in May 2006 under the legal authority of the Protocol of the ECMT, signed in Brussels on 17 October 1953, and legal instruments of the OECD. The Forum's Secretariat is located in Paris.

The members of the Forum are: Albania, Armenia, Australia, Austria, Azerbaijan, Belarus, Belgium, Bosnia-Herzegovina, Bulgaria, Canada, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, FYROM, Georgia, Germany, Greece, Hungary, Iceland, India, Ireland, Italy, Japan, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Mexico, Moldova, Montenegro, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, the United Kingdom and the United States.

The OECD and the International Transport Forum established a Joint Transport Research Centre in 2004. The Centre conducts co-operative research programmes addressing all modes of transport to support policy making in Member countries and contribute to the Ministerial sessions of the International Transport Forum.

Further information about the International Transport Forum is available on Internet at the following address:
www.internationaltransportforum.org

TABLE OF CONTENTS

SUMMARY OF DISCUSSIONS

SUMMARY CONTENTS

1. INTRODUCTION

The Symposium brought together leading transport researchers from around the world to explore a range of issues under the general theme of "the future for interurban passenger transport". A first set of papers investigates what drives demand for interurban passenger transport and infers how it may evolve in the future. The remaining papers investigate transport policy issues that emerge as key challenges from the long-run view on demand: when to invest in high-speed rail, how to regulate to ensure efficient operation, how to assign infrastructure to different types of users (e.g. cars and trucks), what role for information provision, and how to manage environmental impacts. Closing remarks summarized insights from the discussions from an academic and policy-making perspective.

In her opening remarks, Mrs **Concepción Gutierrez del Castillo**, Spanish Secretary of State for Transport, emphasized the importance of sustainability and equity as goals for transport policy, while maintaining its contributions to economic growth. Technological and organisational innovations are required to improve the sector's efficiency. Investments in high-speed trains, single sky agreements, and renewable forms of energy supply are all part of the solution. Many problems require an international approach.

Mr. **Jack Short**, Secretary-General of the International Transport Forum, suggested that, although progress has been made, there remains considerable scope for improvement in the contribution of transportation to economic welfare. Research has proven its value in improving policy in many instances, and continues to be important. In order to increase their impact, researchers need to focus more on implementation issues as this is the key challenge for policymakers in bringing new ideas into practice.

Mr. Short provided a quick overview of Symposium themes. It addresses fundamental questions concerning the shape of future passenger transport and whether current infrastructure and governance policies are appropriate. Big agglomerations are increasingly the motor of economic development. Growth will be stimulated by further agglomeration of economic activity in large cities, and high-quality transport between and inside metropolitan areas facilitates such agglomeration, so contributes to further growth. Deregulation, where it has occurred, has brought economic benefits. There is scope for further liberalisation in many transport markets. Investing in transport is not just a response to growing demand, but can be a force for driving growth if it is well targeted and makes good use of scarce financial resources. For this, improved appraisal is essential, with Cost-Benefit Analysis and Environmental Assessment used strategically to find good solutions across a comprehensive range of potential responses to capacity problems.

Mr. **Richard Thivierge**, Chair of the Joint Transport Research Committee underlined that the Symposium papers address key challenges for future transport policy: when to invest in high-speed rail, how to regulate to ensure efficient operation, how to assign infrastructure to different types of users (e.g. cars and trucks), what role for information provision, and how to manage environmental impacts.

2. THE SPATIAL DISTRIBUTION OF ECONOMIC ACTIVITY AND TRANSPORT: INSIGHTS FROM THE NEW ECONOMIC GEOGRAPHY

In his keynote speech, **Jacques Thisse** developed a framework to understand the long-run development of demand through insights on the location decisions of firms and workers. For firms, a key trade-off in deciding where to locate is between returns to scale in production and transport costs, the latter being understood broadly as trade costs. Concentrating production in cities allows exploiting scale economies, and is facilitated by declining transport costs. Low transport costs, between and inside cities, contribute to an uneven spatial distribution of production and of income. As economies become richer, taste factors have an increasing impact on location choices. For example, workers' dislike for relocating to cities may induce them not to move, with long commutes or lower growth as a consequence.

Thisse's analysis is quite different from the "fixed location" view that is common in transport economics. It increases awareness that decisions on what transport networks to develop – usually public decisions – have a direct and long-lasting impact on where economic activity will take place and how efficient it will be. This raises some questions for transport project appraisal: are the effects on location choice sufficiently reflected in assessments of infrastructure projects, and how does the framework inform our views on where to focus our efforts (e.g. urban *vs.* interurban infrastructure)?

Thisse's framework establishes a more direct link between transport and economic development than is present in much of transport economics, but at the same time it considers transport in a narrow sense as it emphasizes transport for trade and for commuting. Yves Crozet's presentation, discussed in Section 3, makes the point that in passenger transport other trip purposes matter as well. Furthermore, in analysing passenger transport, time spent in transport is a key factor next to monetary outlays. The latter are affected by subsidies, so that any change in funding policies may affect location choices and cities' growth potential.

3. WHAT DRIVES DEMAND FOR INTERURBAN TRANSPORT?

Yves Crozet pointed out that leisure transport and business travel, and more generally discretionary travel, represent an increasing share of trips. Past trends also reveal that with higher incomes came farther, faster, more frequent, and shorter duration trips. Recently there are signs of saturation of demand in some modes – notably car travel ("the golden age of cars may be over") –n some countries. There is no such saturation in overall mobility as there has been a switch to faster modes including high-speed rail and air transport. Associated with this modal shift is a move towards interurban trips, in a network of increasingly complementary cities. The variety of activities that can be accessed increases with faster transport, and with higher incomes the variety of activities consumed rises. With competing demands on the available amount of time, the opportunity cost of activities rises and their duration tends to decline, i.e. there is a trade off between duration and variety in the scheduling of leisure activities.

Will this pattern continue? Saturation could emerge in the sense that there is a limit to how many activities can be squeezed into a fixed time budget, or in the sense that people will come to dislike hectic lifestyles. But these factors are not very likely to curb demand soon. Instead, slower growth may follow from energy and environmental constraints. In the latter case, these constraints need to be imposed through policy. In many countries, there are measures to steer growing demand away from air travel towards high-speed rail, reflecting the view that this is the best compromise between growing demand and environmental requirements. The papers on high-speed rail (Section 4) question the wisdom of this approach, as they emphasize that high-speed rail makes sense in a limited set of circumstances only.

During the discussion the issue of time spent in intermodal connections was raised. Currently in Europe, time spent in accessing airports and sometimes new high-speed railway stations is longer than the core travelling time, by plane for example. This shows the potential for improvements in intermodal access and the importance of the issue of intermodality. Time-resources devoted to security checks at airports have also increased. To the extent such costs cannot be compressed, they will curb the growth in demand for fast transport modes.

David Gillen asks if demand for long distance air travel is likely to grow as it did before the 2008 shock. The answer is that several factors indicate that a more moderate growth path is likely due to less trade-oriented and slower growth for the world economy, higher energy prices, and environmental policy. Recovery is slow and we may be on the verge of a new macro economy, with profound impacts on the transport sector and international air travel in particular. For international air travel, GDP is not the main indicator (whereas it is for domestic air travel). Instead, changes in trade and foreign direct investment drive changes in air passenger kilometres. International air transport, by far the main component of air travel, is closely related to the growth of trade and the likely evolution of tourism (with trade-related traffic representing a declining share of volume but a large share of revenue).

In addition, air travel is stimulated by other factors than growth, notably deregulation and the concurrent changes in supply. These factors boost demand, but as deregulation permeates global markets its stimulating effects will wane over time. There also is a risk that protectionism will slow down movements towards open sky agreements. In sum, demand projections that are based on output mainly, and that implicitly assume growth will rebound to pre-crisis levels, likely overstate future growth. The ICAO, Airbus, and Boeing forecasts fall in this category. The economic swing has been of larger amplitude than previous bubble-bursts, and the fact that it affects a larger part of the world population means that long distance travel will be most affected.

Discussions focused on competition between high-speed train and air travel, stressing that competition potentially brings gains in efficiency. Competition stimulates modes to develop in market segments where they have a comparative advantage. High-speed rail outperforms conventional rail and the very large air market in a fairly narrow range of segments. Some of these segments rely on complementarities between air and rail, with fast trains providing convenient access to airports. The emergence of low cost airlines strengthens the number of destinations where competition exists and also reduces the number of short and medium distances where high-speed rail may be relevant. As will be emphasized below, where access charges for railway infrastructure are very high this deters competition.

4. ASSESSING HIGH-SPEED RAIL PROJECTS

Chris Nash pointed out that for new high-speed rail lines to be beneficial very high traffic volumes are required, of the order of nine million passengers per year on average (with variations depending on construction costs), a number not attained in all proposed projects. In markets with travel times of three hours or less between city centres, high-speed rail tends to capture at least 60% of the air plus rail markets.

Yield management means that prices exceed marginal costs. Whether it allows profitable operation, however, depends on access charges, which tend to be high (exceeding marginal cost, sometimes by a factor of 5) in a vertically separated environment. It is questionable from a social point of view if such high access charges make sense, given that they discourage use of very expensive infrastructure. If open access models of competition are accompanied by such charges, they may be outperformed by franchising models of competition.

High-speed rail is rarely worth it for higher speed alone but where a new line is required to accommodate growth the marginal cost of higher speed may be low enough to justify the high-speed option. The basic case for investment lies in added capacity, and the capacity of a high-speed line is vast. The benefits of released capacity in other rail travel and in airports (not so much in roads) need to be accounted for in assessments. Of course, such benefits occur only when there is congestion elsewhere, and alternative ways of expanding capacity need to be considered.

Environmental benefits are not a key argument in high-speed rail's favour. The energy intensity of high-speed rail is about twice that of conventional rail, an effect partly compensated by higher load factors. High-speed rail does not save energy, but may avoid CO2 emissions if power is produced with low emissions. The limited environmental bonus from high-speed rail is further diminished when emissions from the construction phase are included. For example, according to Mr. Crozet, the Dijon – Mulhouse line will need about 12 years of operation to compensate for emissions from construction. Numbers vary strongly across projects given the dependence of emissions on design choices (e.g. tunnels).

Network effects, i.e. volume changes in non-high-speed rail parts of the rail network, need to be accounted for and are potentially important. Such network effects tend to be substantially larger where high-speed rail shares a general purpose network, compared to the case of dedicated networks (as is dictated by technology in e.g. the case of maglev). Wider economic benefits, e.g. boosting agglomeration economies, are uncertain and vary greatly from project to project.

Katsuhiro Yamaguchi provides a stark example of the finding that the basic economic case for high-speed rail is one of very high levels of demand confronted with capacity constraints across modes. His analysis suggests that a maglev train connecting Tokyo, Nagoya and Osaka would be socially beneficial if the Japanese economy grew by 2-3% over the next 65 years. In that case, transport demand would grow so fast that even with the Maglev the volume of air transport would continue to grow. Irrespective of whether these assumptions are realistic, it deserves emphasis that the current maglev project has been proposed by the private high-speed rail company running trains on the potential maglev corridor. Its motive could be to move proactively to forestall competition from an alternate publicly funded proposal.

Ginés de Rus follows Nash in stating that not all proposed high-speed rail projects pass a cost-benefit test. Furthermore, he points out that public funds are getting scarcer and more money will be needed to repair and upgrade existing infrastructure, highlighting the need for careful project assessment.

In contrast to Nash, de Rus sees merit in the idea that prices should reflect all costs (not just marginal costs) in order to provide correct signals to investors (i.e., in this case, avoid overinvestment). Increased scarcity of public funds could mean more private sector involvement and heavier reliance on user charging to finance infrastructure. De Rus asks what this could mean for high-speed rail fares – and if fares increase, what that means for occupancy rates, which are key in making high-speed rail socially beneficial. Careful project assessment also requires considering a reasonable set of alternatives. For example, if high-speed rail generates benefits through relieving congestion elsewhere, should it be assumed that no improvements to charges for these other infrastructures are envisaged? In other words, should we go ahead with high-speed rail because airport or rail network access is priced inappropriately?

In the face of these remarks, it is difficult to explain the widespread enthusiasm for high-speed rail. De Rus points to co-financing arrangements for EU funds as one explanatory factor, with the potential of leveraging national funds with EU money diverting resources from projects that don't qualify for co-financing but show higher returns. This mechanism results in increased subsidies where investment costs are higher and revenues lower. Discussions ensued on what is the funding principle for high-speed rail, with stated objectives including European integration and cohesion, concerns not included in standard cost-benefit appraisal. Many experts, however, subscribe to the view that high-speed rail is not "beyond" cost-benefit appraisal.

While cost-benefit analysis is deemed to be indispensable, practice is not always satisfactory. In light of Thisse's remarks and given the size of a typical high-speed rail project, it is desirable to develop a systematic view on location effects. However, analytical and empirical constraints have prevented this from happening. Advances in this regard could have a considerable payoff. Experts pointed out that such advances don't necessarily mean increased complexity of models used, and expressed a preference for relying on simple models and scenarios in order to guarantee transparency and improve robustness.

5. GOVERNANCE: HOW MUCH (DE)REGULATION?

Botond Aba described how fiscal concerns in Hungary tend to be detrimental to the market position of public transport. Individual consumers tend to prefer cars over public transport and public investment in motorways caters to these preferences, leading to a strong modal shift towards cars. Car ownership and use creates an attractive base for generating public revenue. Public transport, while socially beneficial, cannot usually break even financially, meaning it is costly in terms of public funds. Aba contends that the budgetary implications of car and public transport travel drive transport policy, more than transport interests proper. A sustainability-oriented transport policy would require strong public involvement, with a focus on exploiting complementarities between public and private transport, rather than seeing them as competing modes.

Clifford Winston takes an almost diametrically opposed view, asking what the experience with deregulation in various parts of the US transport system tells us about the potential impacts of further deregulation and privatization. He argues that deregulation has delivered substantial benefits, and expects further improvement as the private sector continues adapting to the deregulated environment. Remaining inefficiencies due to poor public policy hamper the realization of the full benefits of deregulation. Where there is strong public involvement, e.g. in public transport and in infrastructure provision, performance declines, innovation is virtually absent, and funding tends to fall short.

Still according to Winston, the way forward is to continue reducing public involvement in the transport sector, through outright privatization of most functions. This will stimulate entry (boosting competition) as well as organizational and technological innovation, which are strongly stifled by regulation. The entry of Megabus in the US, which revived the coach market, can serve as a recent example. In general, any shortcomings of the market are thought to be small in comparison with government failure, so that deregulation or privatization is recommended even where cost structures may create problems (e.g. highways). Discussion filed to shed light on how private road monopolists would be deterred from rent seeking in the way they set charges for using roads. Adaptation to deregulation is slow and adaptation to privatization is slower. Frustration with the lack of quickly forthcoming benefits creates a threat of re-regulation (especially in times of crisis), implying a continuing distraction of entrepreneurial effort.

Long-distance coach services are an example of successful deregulation in Europe. **Didier van de Velde** shows that countries that adopted licensing approaches have witnessed the emergence of a profitable and competitive industry serving market segments not very well catered for by rail, air or car modes. Substitutability with rail is particularly weak, calling into question the rationale for policies in some countries to discourage coach services in order to protect rail, even if one would think such policies justifiable in principle. At the same time, competitive pressure from car and air as well as from potential entrants is strong enough to maintain competition even when the number of incumbents is small. Van de Velde was careful to point out that the (de-)regulatory model for coach services works well but is not necessarily transferable to other modes (notably rail), given major differences in technology, cost structures, and possibly the structure of demand.

De-regulation has progressed more slowly on Europe's railways. The team from the Universities of **Berlin and Dresden** assessed the merits of three models for market access in European long-distance passenger rail transport, characterised as "Tendered Concessions", the "Monopolistic Network Operator" and the "Open Market". Most empirical experience to date relates to the tendered concessions developed in Great Britain, with their strengths and weaknesses (see *Competitive Tendering of Rail Services*, ECMT/OECD 2007). Open access experience is still in its infancy but appears to be the preferred approach of the European Union for regulating international services, as apparent in Directive 2007/58/EC. This directive requires international services to be open for competition and permits cabotage, that is picking up domestic passengers on intermediate stops between terminals in different countries. Cabotage rights can be denied under EU rules, however, on routes operated by train companies under public service obligations with financial support from government. It is as yet unclear how compatible open access for international services will be with tendered concessions for domestic markets. This could be a problem particularly for networks in a country like the Netherlands where services are interwoven.

The paper includes a discussion of the 9 small scale attempts at entry in Germany, Europe's largest passenger market, over the last 15 years, none involving more than 2 train pairs. Two current cases are potentially more significant. Locomore Rail has announced plans to operate three daily trains from Hamburg to Cologne from August 2010 and has been successful in securing train paths from DB Netz. Keolis, backed by France's SNCF, plans services between Strasbourg, Frankfurt and between Hamburg, and Strasbourg, Frankfurt, Berlin, and Hamburg, comparable to DB InterCity services. Keolis has not yet received a confirmation of the train paths requested, with a decision to be made by the network subsidiary of DB by April 2010.

John Preston concurred that competition for long distance rail services remains relatively limited, noting that on-track competition, where it has occurred, seems to focus on niche markets which the incumbent operator has neglected. At the same time modelling work indicates that if track access charges are based on short run marginal cost, head-on competition may be feasible for densely trafficked routes but not necessarily socially desirable, with a tendency to result in too much service, at too high fares. By contrast, analysis of the niche open access entry in Britain providing direct services to new destinations, based on marginal cost based track access charges, does appear socially desirable. Capacity constraints on the main lines and at key terminals mean that such competition may be limited and there is the wider issue of whether these services are making the best use of limited capacity.

Off track competition in Great Britain has been able to attract sufficient numbers of bidders, has coincided with strong demand growth and can result in large premia being paid to the government. However, such competition is vulnerable to the winner's curse (i.e. in order to win bids have to present optimistic revenue forecasts that make them more likely to fail). The biggest revenue risk relates to GDP and risk sharing mechanisms that link premia / subsidies to GDP could perhaps avoid the worst problems experienced with franchises. Linking payments to GDP could also permit longer franchise periods, better suited to investment in new rolling stock.

Discussions on the papers concluded that the high fixed costs of providing passenger rail services, and especially high-speed services, condemns open access competition to a peripheral role. Open access entry is usually only possible where the entrant is required to pay charges for using infrastructure based on marginal, variable or avoidable costs. Seeking a significant contribution to fixed costs is likely to exclude entry. High-speed train services are usually charged high track access prices, covering a large part of fixed costs, making open access entry difficult in this market. Conversely if an open access operator paying only marginal costs took a large share of the market, network operations would be financially compromised. Infrastructure charges in Germany reflect these factors in basing prices on marginal costs for train operators that run only a small number of services a day on a route and charging much higher access prices for more frequent services. This structure of charges is partly a result of an regulatory decision that an early schedule of charges that spread fixed costs more evenly was anti-competitive.

It was acknowledged that all approaches to introducing competition into rail passenger markets pose challenging regulatory problems but competition for the market, through concessions, was viewed as more likely to succeed than competition in the market through open access train operations because it offers solutions for covering fixed costs. With either approach to introducing competition, the central importance of a credible and independent regulator was stressed. The need for a strong regulatory lead is particularly important when open access competition is expected to develop in circumstances where management of the infrastructure network is integrated with an existing train operator, for example through a holding company.

6. ASSIGNING INFRASTRUCTURE

Advanced transport systems consist of various modes, some of which use dedicated infrastructure. Increased product differentiation within rail transport has led to dedicated infrastructure for high-speed rail. By contrast, nearly all road infrastructure is general purpose and is shared by a very heterogeneous set of users. Could it make sense to assign parts of the road network to particular types of traffic? This issue is investigated in the papers by **Robert Poole** and **Robin Lindsey**, with a focus on car and truck traffic.

Poole observes that many High-Occupancy-Vehicle lanes still are underused, but argues that separate infrastructures can make sense when potential users differ strongly in their value of time. Car-only lanes can be justified in urban contexts where speeds are low, as this allows designing narrower lanes which in turn makes better use of existing rights-of-way and opens perspectives for using new rights-of-way (e.g. drainage channels, power line corridors). Truck-only lanes can be designed for heavy trailer combinations. Lindsey's formal analysis supports the possible case for separation, in the sense that an unregulated equilibrium on a general purpose facility tends to lead to integration, whereas the lowest-cost outcome could require separation because of crash risks or because of strongly differing values of time. Tolls can be used to match the unregulated and lowest-cost outcome. Lane access restrictions are less effective, however. For example, if cars are banned from one lane but trucks are not, then trucks can use both lanes and this raises costs.

7. HARNESSING INFORMATION TECHNOLOGY

Mr. Zimmermann explained that because the telematics market did not develop as expected a high tech initiative was taken in 2006 by the German authorities. The idea is to offer a complete range of information services both for private and public transport. Due to proprietary efforts, various interfaces and protocols had to be developed with algorithms for the transfer of data. Information has to be provided both before and while travelling. Floating data on secondary roads had to be put in place to guarantee that diversion on the secondary network does not lead to a loss of information. There has been some reluctance of public companies to provide data on incidents, but because of the interdependencies among service providers and the bad image associated with the lack of accurate data, the floating data system worked in the end. In this respect, providing information is a self reinforcing mechanism.

The discussion identified several unanswered questions, all of them important for any ITS evaluation: how to measure expected benefits of projects and of ITS in general; what elements might favour a Benefit/Cost ratio larger than 1; how to deal with instability when suggesting alternative routes may create more congestion on the diversion routes than it removes on the main route?

Mr. **Tapiador** and Mr. **Marti-Henneberg** tackled the problem intermodality in a specific context. As governments invest in high-speed rail, railway operators have to ensure access to this new type of services and link it to the railway system of the Nineteenth Century as a starting point. New railway stations also have to be built, often located on the outskirts of cities. In this context the private car ("Kiss and ride") is the preferred access mode, with taxis playing a very important role on the return journey. This shows that in dealing with accessibility and intermodality a wide range of modes has to be considered. Governments tend to focus on big investments whereas more simple and direct decisions can be quite effective to improve accessibility. At the same time, the authors argue that investments in information technologies may prove to be a very efficient way to strengthen intermodality at low costs.

The latter point provoked questions: to which extent are the costs of implementation of ITS really several orders of magnitude lower than in "hardware" (infrastructure/rolling stock)? Clear insight here is obviously important for deciding what to invest in.

8. SUSTAINABLE INTER-URBAN MOBILITY

As noted by De Rus and Nash, advocates of high-speed rail investments often place heavy emphasis on environmental benefits, especially when they divert significant shares of air travel. **Per Kågeson** tests this assertion by looking at the relative environmental (principally GHG) impacts of competing inter-urban modes, not at their present level of performance but at one more representative of their impact over the lifetime of high-speed rail infrastructure taking 2025 as the baseline. Many factors play a role in this assessment, including the amount of GHGs released during the construction of new infrastructure. Overall, however, it is the speed and resulting energy requirements for high-speed rail that dominate the final impact assessment. Kågeson notes that "it is odd that so much emphasis is placed on high-speed in the rail sector when so much focus has been on reducing speed for GHG savings on roads and in the air."

Does high-speed rail deliver on its claimed environmental benefits? High-speed rail can deliver GHG savings, especially when it replaces air travel, but after accounting for generated travel, high energy requirements and the carbon intensity of the marginal electricity used, these benefits are small and expensive. "Standard" passenger rail services may be "good enough" from both an environmental and economic perspective, especially where travel volumes are low and are not expected to grow significantly. These findings are robust across all but the most extreme assumptions so that in most cases it would be incorrect to attribute large-scale GHG benefits to high-speed rail.

Much of the debate regarding regulatory approaches to reducing GHG emissions from aviation has focused on the relative merits of a fuel levy versus a trading system but, as **Peter Morrell** points out, relative legal impediments to action on a global fuel levy and the EU decision to include aviation emissions within the European Trading System (ETS) has focused attention on the mechanics and economics of aviation GHG emissions trading. He points out that, as with other trading approaches, decisions regarding allocation regimes and distortionary impacts are important to understand when assessing overall performance -- not because they have an impact on overall emissions or costs but because they affect carriers differently and this can affect competitive conditions in the industry, which in turn affects emissions.

– Will carriers restructure their operations to avoid long inbound or outbound European flight segments in response to the new European rules? The answer is not straightforward since avoiding EU hubs may entail added fuel and time costs and may not fit with other commercial strategies (e.g. connecting with partner or code-share networks). In the examples Morrell cites, the cost penalty of the ETS charge is more-or-less matched by the fuel cost penalty of non-EU hubbing on the same point-to-point routes.

– Morrell asks how increased fares resulting from the added cost of permits might discourage travel and thus reduce aviation emissions. With 100% pass-through emissions could be 7.5% below what they otherwise would have been in 2020. However, it is not clear that operators would pass on 100% of the added costs. Carriers can use non-ETS routes, cargo and differentiated passenger markets to distribute the ETS burden so that not every fare increases by the costs of CO2-emissions caused by the flight. As pointed out in discussions, pass-through could also be lower at congested airports where its impact is likely to be a reduction in the landing slot rents accruing to incumbent airlines (OECD/ITF 2009), a view challenged by Morrell as failing to take account of the multi-dimensional outputs of airlines.

– Facing steeply rising abatement costs in aviation and a context where carbon prices will be largely set in the large power and electricity sectors, aviation is unlikely to reduce emissions in absolute terms and only slightly relative to transport volume. It would, however, pay for emission reductions in other ETS sectors by raising the cost of carbon permits. This is simply a reflection of differences in marginal abatement costs between sectors but, as pointed out in the discussion, it does raise the issue of the appropriateness of non-EU operators paying for EU emission reductions.

9. STRATEGIC ENVIRONMENTAL ASSESSMENT

There is considerable experience in applying strategic environmental assessment (SEA) to transport but, as **Rodrigo Jiliberto** notes, many of the procedures followed are ill adapted to the political decision making environment. A narrow legalistic approach is often used, treating SEA simply as a larger scale version of traditional Environmental Impact Appraisal (EIA). **Maria Partidario** observes that SEA was initially developed as a way to move environmental and social issues upstream in the planning and decision-making process and improve the context for subsequent project EIAs. But she argues that to be effective in changing outcomes, SEA has to cut its links with EIA and become an instrument that occupies a new space in strategic development processes, changing attitudes and establishing a direct role in the decision-making process.

She chose a case study of the selection of the site for a new airport for Lisbon to illustrate how SEA can change outcomes. Success in this case was in part conditioned because the government initiated a new SEA study as a means to achieve closure in an incremental planning process that had led to the selection of a number of unsuitable sites with the results contested by different interest groups. The SEA began by screening the entire region around Lisbon for suitable sites meeting criteria for accessibility, economic development potential and environmental sensitivity. The success of the process was attributed to a clear focus on the decision that needed to be made; not whether a new airport was required but where to locate it and how best to integrate it into the economic and environmental fabric of the region. The assessment was based on seven critical factors acknowledged by policy makers to be most relevant to the decision, and this enabled a much more structured approach to the studies that contributed to the SEA than is typical. Above all success was attributed to communicating clearly with politicians through the choice of indicators presented and the way in which summaries of the analysis undertaken was presented.

10. FINAL SESSION

Cristina Narbona Ruiz was the first speaker to intervene in the final session chaired by Francesc Robusté. She echoed David Gillen's view that the 2008 crisis is in many respects a rupture, and is accompanied by an environmental crisis. The failure of markets to properly regulate the global economy calls for a new political governance through transparency of information and accountability. A new paradigm is also needed because we are potentially facing some irreversible consequences of climate change. A green growth strategy is essential and it is at the same time a great challenge for politicians even if the economic costs of doing nothing would be higher than the costs of the measures to be implemented. In fact, the later we act, the more costly the measures to be taken will be. We have to gradually eliminate fossil fuel subsidies and move to carbon pricing. Part of the solution is also to move from an economy of ownership to one of service functionality and manage the demand for services. For example, in the transportation sector, public transport can no longer be seen as a second class choice.

Paolo Costa commented on the high-speed rail analyses discussed earlier, explaining that high-speed rail was part of the TEN-T programme to improve European integration through connecting the national networks and ensuring interoperability. A technical jump through new high-speed rail infrastructure was considered as the only way of strengthening public transport attractiveness while at the same time moving towards a decarbonized economy. Through the network effects and improved interoperability the long run positive return of these investments are undoubtedly positive for Paolo Costa, even if narrower economic assessments suggest negative social returns in some cases.

In response, **Chris Nash** agreed that profound changes in transport are required to meet sustainability. However, the contribution of high-speed rail in achieving European integration is very limited: the demand for such services comes from diversion of conventional trains and other modes, and is altogether not sufficient even with generated traffic to cover costs. Are mega projects such as high-speed rail the best way to achieve this European cohesion? Freight transport is also very important and Chris Nash questioned whether in the framework of TEN-T it would not have been wiser to concentrate on investment in freight transport even if HST frees capacity for some more conventional services. The high-speed rail system in Europe is characterized by high costs, a low level of interoperability, and technical complexity while at the same a consistent approach to questions such as adequate pricing for the use of infrastructure has still to be found. At this stage, insisting on cost recovery through high access charges is bound to produce socially suboptimal use of available infrastructure.

Francesc Robusté summed up the debate saying that sustainability is also a condition for economic growth and we cannot adopt a business as usual approach for future interurban transport. He added that on various points such as accessibility enhancement, cost benefit analysis, understanding future patterns of mobility, pricing and strategic decision-making the Symposium brought forward looking analysis that should help improve transport policy and transport services.

REFERENCES

ECMT (2007), *Competitive Tendering of Rail Services*, OECD Publishing, Paris.

ITF (2009), *Competitive Interaction between Airports, Airlines and High-speed Rail*, OECD Publishing, Paris.

OPENING SESSION – KEYNOTE SPEECH

How Transport Costs Shape the Spatial Pattern of Economic Activity

Jacques-François THISSE

CORE, Université Catholique de Louvain
Belgium
and
Ecole Nationale des Ponts et Chaussées
Paris

SUMMARY

1. INTRODUCTION

By its very nature, transport is linked to trade. Trade being one of the oldest human activities, the transport of commodities is, therefore, a fundamental ingredient of any society. People get involved in trade because they want to consume goods that are not produced within reach. The Silk Road provides evidence that shipping high-valued goods over long distances has been undertaken because of this very precise reason. But why is it that not all goods are produced everywhere? The reason is that regions are specialized in the production of certain products. The first explanation for *specialisation* that comes to mind is that nature supplies specific environments needed to produce particular goods. According to Diamond (1997), spatial differences in edible plants, with abundant nutrients, and wild animals, capable of being domesticated to help man in his agricultural and transport activities, explain why only a few regions have become independent centres of food production. Though relevant for explaining the emergence of civilization in a few areas, we must go further to understand why, in the wake of the Industrial Revolution, interregional and international trade has grown so rapidly.

Goods are not ubiquitous because regions are endowed with a *comparative advantage*. Specifically, this advantage stems from the ability of a region to supply a particular good at a lower opportunity cost than other regions, sometimes because its inhabitants have learned how to produce it by means of technologies unknown to others. Spatial heterogeneities among regions, such as the uneven distribution of immobile resources (natural harbours) and amenities (climate), as well as differences in the access to major transhipment points (e.g. the Great Lakes in Canada and the United States), may also be at the origin of a variety of comparative advantages. Each region thus specializes in the production of goods for which it has a comparative advantage and trades with regions specialized in the production of other goods. However, the existence of transport costs renders a whole range of goods for which neither region has a sufficiently important productivity advantage non tradable. In other words, the production cost advantage is not sufficient to overcome the disadvantage linked to the value of transport costs. As the magnitude of transport costs decreases, the range of tradable goods widens. Even though exogenous comparative advantages are important, it is my belief that they cannot by themselves explain the formation of big agglomerations and large trade flows across regions and countries. Furthermore, some of these heterogeneities (think of the supply of transport infrastructure) are not given by nature and should be treated as being endogenous.

Modern trade theory has underscored the fact that specialization may also be the outcome of activities displaying *increasing returns* (Helpman and Krugman, 1985). To understand how this works, it is important to recognize that increasing returns may arise for a variety of reasons. First of all, scale economies are said to be *internal* to firms when the productive efficiency of firms increases with the size of their output. One major reason for this is that firms are able to adopt more efficient technologies once their sizes have reached a minimum threshold. Firms may also increase their productivity through learning-by-doing economies that emerge over the production process itself. Less known, perhaps, is the concept of scale economies *external* to firms whose origin lies in the socio-economic structure of their close environment (Duranton and Puga, 2004). This includes a wide range of factors such as the access to specialized business-to-business services, the formation of a specialized labour force, the production of new ideas, based on the accumulation of human capital and face-to-face communications, and the availability of efficient and specialized infrastructure. Scale economies are the prime driver in the formation of cities where the division of labour and the specialization of tasks reach a level impossible to achieve with a dispersed population (Fujita and Thisse, 2002). It should then be clear that regions and cities get specialized in the production of specific goods because of the cost advantage generated by increasing returns, either internal or external to firms. Transport costs remain an impediment to trade, but market size matters here. Indeed, the existence of large local markets may overcome high transport costs through low average production costs.

Thus, we may safely conclude that the demand for the transport of commodities stems from the need to trade, which itself comes from the productive specialization of regions. All distance-related costs having dramatically decreased with technological advances in transportation and the development of the new communication technologies, it is easy to figure out why trade has grown at a fast pace. In addition, new and cheaper transport means impact on the location of firms and households. By changing the accessibility to input and output markets, lower transport and communication costs give them incentives to relocate. Therefore, it is legitimate to ask the question: *what is the impact of falling transport and communication costs on the location of economic activity?*

In order to say something relevant about the way a spatial economy is organized, it is necessary to assume that the production of goods involves increasing returns. If returns to scale are constant, allowing for the mobility of households and firms has a weird implication: all locations have the same relative prices and the same production structure. Indeed, in a world where activity can operate at arbitrarily small levels without efficiency losses, firms and households may reduce transport expenditures to zero by dispersing their activity across space. Every region then becomes an autarky, as it only needs to produce for its own domestic market. Hence, the standard economic paradigm combining constant returns and perfect competition is unable to account for the emergence and growth of big economic agglomerations and the existence of large shipments of goods.

Thus, the presence of increasing returns has a fundamental implication for the spatial structure of the economy: *not everything can be produced everywhere.* Therefore, it is no surprise that, in many real-world situations involving the location of large equipments, decision-makers face a trade-off between global efficiency and spatial equity (e.g. the proliferation of transport facilities is often the consequence of policies that put too much weight on spatial equity). Increasing returns have another major implication for the space-economy: lower transport costs may amplify or reduce the geographical advantage and disadvantage held by particular regions. Or, to put it differently, *a small exogenous comparative (dis)advantage can become a large endogenous comparative (dis)advantage*.

That said, what drives the location of firms and consumers is the existence of spatially dispersed markets. Accessibility is measured by all the costs generated by the various types of spatial frictions that economic agents face in the exchange process. Hence, it should be clear that *the way the space-economy is organized depends on the mutual interactions between mobility costs and scale economies*, the specification of which varies with the spatial scale (the world, the country or the city). In my opinion, the opportunity of developing interurban passenger transport must be evaluated within this framework because it strongly affects the type of mobility across cities that highly-skilled workers may choose.

The purpose of this paper is to discuss some of the main trade-offs at work at different spatial scales. Needless to say, within the format of this paper, I can cover only a few of the main ideas developed in economic geography and urban economics. The emphasis will be on the impact that falling transport costs have on microeconomic decisions on, and the resulting aggregate outcomes of, the location of firms and workers.

2. THE TRADE-OFF BETWEEN INCREASING RETURNS AND TRANSPORT COSTS

2.1 The optimal number and size of firms

The Industrial Revolution brought dramatically low transport costs as well as a huge increase in the size of production plants. The very first industrial plants had a very small optimal size. Indeed, as observed by Bairoch (1997): “In most manufacturing sectors, it was possible for a firm to have a competitive position with a very small size. The narrowness of the market, due to high transport costs, made it even easier to operate at a very low scale.” Things changed after the first half of the nineteenth century. The minimal size of a firm grew because of the use of increasingly diversified equipment, which then required many more workers. This growth in the size of firms was sustained by the expansion of markets areas, which in turn was possible because of the strong decline in transport costs. In brief, the interactions between these changes led to a gradual reduction in the number of firms, whose size increased. Take, for example, the case of Belgian steel enterprises: while their average workforce in 1845 was 26 people, it reached 446 people in 1930 (Bairoch, 1997). Hence, it is no surprise that the trade-off between increasing returns and transport costs is at the heart of location theory.

The trade-off between these two forces is easy to understand. First, as mentioned above, in the absence of increasing returns, one plant could be built in each consumption place so that there would be nothing to ship. Moreover, in the absence of transport costs, a single plant would be enough to satisfy the entire demand (except for the case where its marginal cost of production would increase). When transport costs increase with distance, this is formally equivalent to the case in which a fixed cost coexists with a growing marginal cost. Each plant supplies consumers located within a certain radius, the length of which depends on the relative level of the transport costs and the intensity of increasing returns, but those located beyond this radius are supplied by another unit.

The nature of this trade-off can be illustrated by considering the simple case of three spatially separated markets, **W**(est), **C**(entre) and **E**(ast), where the local demand for a given good is perfectly inelastic and normalized to **1**. Building one facility in a market requires **F** euro, while shipping one unit of the good between any two adjacent markets is equal to **T** euro. It is readily verified that the choice is between the following two options. First, building a facility in each market generates a total cost equal to **3F** since there is no shipping. Second, when a single facility is built, the optimal location is **C** and the corresponding cost **F + 2T**. The cost-minimizing solution, then, is to have a single facility if and only if

$$\mathbf{F + 2T < 3F \Leftrightarrow T < F.}$$

This inequality holds when **F** is high and **T** is low. Otherwise, it is optimal to have three facilities. This example is enough to understand that, on the one hand, high fixed costs favour the concentration of production in a small number of large units, as in modern developed economies; while, on the other hand, the situation in which high transport costs encourage the proliferation of small settlements across space characterizes preindustrial economies. Despite its simplicity, this example illustrates a very general principle: *strong scale economies in production* (large **F**), *low transport costs of* *commodities* (small **T**), *or both foster the agglomeration of economic activities in a small number of areas*.

By modifying slightly the example, it is possible to uncover another major principle of economic geography. Specifically, we assume that the common demand for the good is shifted upward from **1** to **D** units. The above inequality then becomes

$$\mathbf{F + 2DT < 3F \Leftrightarrow DT < F.}$$

Clearly, this ceases to hold when **D** is sufficiently large. Hence, when local markets are large (large **D**), it is optimal to supply each of them from a facility set up there. In other words, even when unit transport costs are low (small **T**), *the proximity to large markets matters for the location of firms*.

2.2 The optimal location of a firm

The simplest firm-location problem is the one in which the firm, which cannot be subdivided in smaller units because of increasing returns, buys one input in one market (**W**) and sells its output in another (**E**), with a link connecting the two markets. The optimal location of the firm, which minimizes the sum of transport costs, can be viewed as the equilibrium point of a system governed by two forces generated by the need for proximity to the product market and the factor market. The intensity of these two forces depends, on the one hand, on the quantities shipped ($\mathbf{w_1 > w_2}$) and, on the other, on the marginal cost of transport with respect to distance.

Assuming that input and output are shipped by means of the same transport mode, the value of the elasticity of the unit transport cost function **T** with respect to distance is an indicator of the degree of increasing returns in transportation. More precisely, a high value of this elasticity means that making the movement slightly longer increases its cost greatly. In this case, the value of transport costs is determined mainly by the distance covered when shipping goods. Such a situation describes quite well periods in which moving commodities was both dangerous and difficult, thus necessitating coaching inns for ground transport and coastal navigation for maritime transport. On the contrary, a low elasticity implies that the share of transport costs due to investments in infrastructure and equipment grows, so that distance matters less. Clearly, such a situation is characteristic of modern economies.

To start with, assume that the elasticity of the transport cost **T** is larger than **1**. In that case, the intensity of the pulling forces increases rapidly with distance, as illustrated in Figure 1a. Consequently, the system of forces is in equilibrium when the firm chooses the location where the marginal transport costs with respect to distance are equal: increasing the length of a trip is so costly that it is desirable for the firm to reduce the distance to the market with the higher marginal cost. This is why a place located in between the two markets is cost-minimizing. If the elasticity decreases to reach a value equal to **1**, the firm chooses to establish itself in the market with the highest weight (see Figure 1b where the bold line takes its lowest value at **W** since $\mathbf{w_1 > w_2}$). Because the intensity of the forces is now independent of the distances to the input and output markets, every intermediary location becomes suboptimal. This also holds when elasticity takes on values less than **1**, as the marginal cost of transport decreases with distance.

Figure 1a.

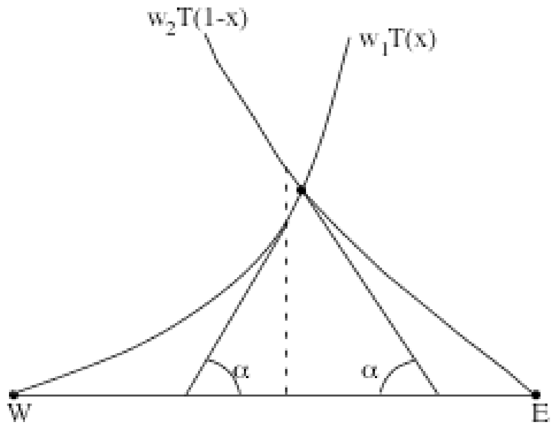


Figure 1.a: $w_1T'(x) = -w_2T'(1-x) = \alpha$

Figure 1b.

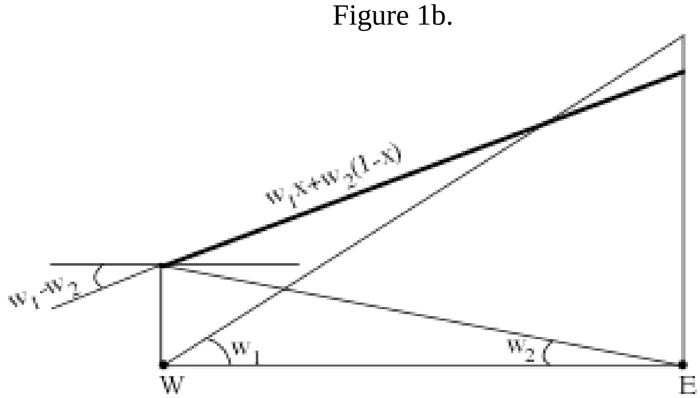


Figure 1.b: $w_1 > w_2$

The way in which distance has affected transport costs over time may then be described succinctly as follows. The long period during which all movements were very costly and risky was followed by another during which, thanks to technological and organizational advances, ships could cross longer distances in one go, thus reducing their number of stops. On land, it was necessary to wait for the advent of the railroad for appreciable progress to occur, but the results were the same. In both cases, long-distance journeys became less expensive and no longer demanded the presence of relays or rest areas. Such an evolution in technologies has favoured places of origin and destination at the expense of intermediate places. As this argument may be extended to the case of any transport network having several nodes and markets, we may confidently assert that *increasing returns in transport explain why places situated between large markets and transport nodes have lost many of their activities*. Stated in a different way, the construction of new and large transport infrastructures will be beneficial to the main centres it connects, but not the regions it crosses. But if the global morphology of the network is changed through new and bigger nodes (e.g. Singapore or Chicago), these infrastructures may affect the location of economic activity.

To sum up, scale economies in production and transport activities have combined to lead to the spatial concentration of human activities. In particular, the development of new transport technologies exhibiting a high degree of increasing returns strengthens the tendency toward more spatial polarization of high value-added activities.

3. THE MOBILITY OF FIRMS AND WORKERS

Countries and regions are affected not only by the growing mobility of commodities but also by that of production factors (e.g. capital and labour). What I want to stress here is that *lowering transport costs change firms' and workers' incentives to move*. It is, therefore, crucial to have a good understanding of how firms and workers react to these changes in order to assess the full impact of trade and transport policies. In this respect, it should be stressed that policy-makers often overlook the fact that their decisions impact on the location choices made by firms and households. These choices may lead to a new pattern of economic activity that vastly differs from the existing one. In particular, the economic geography approach to factor mobility highlights the fact that the mobility of factors need not reduce spatial inequality. It also stresses the fact that the mobility of firms and workers do not have the same impact on the global economy.

3.1 The home-market effect

Both economists and geographers agree that a large market tends to increase the profitability of the firms established in it. More generally, the idea is that locations that have good access to several markets offer firms a greater profit. Hence, it is reasonable to expect that the firms that set up in large regions enjoy higher profits than the ones installed in small ones. In brief, firms would seek the locations with the highest market potential where demand is high and transport costs low (Redding and Venables. 2004). The core region should, therefore, attract new firms, thereby heightening the inequalities between the core regions and the others. Nevertheless, as firms set up in the core regions, competition there is also heightened, thereby holding back the tendency to agglomeration. Consequently, the interregional distribution of firms is governed by two forces pulling in opposite direction: the agglomeration force is generated by firms' desire for *market access*, while the dispersion force is generated by firms' desire to avoid *market crowing*.

This question has been studied in a standard two-region, two-sector, and two-factor economy (Helpman and Krugman, 1985). The industrial sector produces differentiated goods under increasing returns and imperfect competition, using capital and labour, whereas the traditional sector produces one good under constant returns and perfect competition, using labour only. This setting combines the mobility of both commodities and capital, while consumers/workers continue to be immobile. Furthermore, the mobility of goods is imperfect because their shipments incur positive transport costs. It is therefore tempting to conclude that the region with the larger market will always attract firms for the reason that this location minimizes total transport costs to both markets. However, as said above, this argument ignores the fact that when more firms locate within the same region, local competition is intensified and profits are lower.

When one region is larger in terms of population and purchasing power, this push and pull system reaches equilibrium when this region hosts *a more than proportionate share of firms*, a result that has been coined the "home market effect" (HME). Because of its comparative advantage in terms of size, it seems natural that the larger region should attract more firms. What is less expected is that the share of firms exceeds the relative size of this region, thus implying that the initial advantage is magnified. This is because firms installed in the larger region have a better access to a bigger pool of consumers that allows them to produce at a lower average cost. Hence, contrary to general belief, capital does not necessarily flow from the regions where it is abundant to the regions where it is scare.

Moreover, the HME is amplified by decreases in transport costs: *more firms choose to set up in the larger region when transport costs decrease*. This somewhat paradoxical result can be understood as follows. On the one hand, lower transport costs makes exports to the smaller market easier, which allows firms to exploit more intensively their scale economies; on the other hand, lower transport costs also reduces the advantages associated with geographical isolation in the smaller market where there is less competition. These two effects push toward more agglomeration of the industrial sector, thus implying that, as transport costs go down, the smaller region gets de-industrialized to the benefit of the larger one. The HME is thus liable to have unexpected implications for transport policy, such as that implemented by the European Commission in its cohesion program. By making the transport of goods cheaper in *both* directions, the construction of a new infrastructure permits an increase in both imports to, and exports from, the smaller region. As seen above, a transport cost-reducing policy is likely to induce some firms to pull out of the smaller region, thus failing to reduce regional disparities. To some extent, this explains the disillusion regarding the effectiveness of policies that aim for a more balanced distribution of activities across the European Union (Midelfart-Knarvik and Overman, 2002).

It is well documented that on average firms and workers tend to be more productive in larger markets (Syverson, 2004). Once it is recognized that firms are heterogeneous in productivity, location choices act as a selection device. Specifically, decreasing transport costs lead to the gradual agglomeration of low-cost firms in the larger region because these firms are able to survive in a more competitive environment. In contrast, high-cost firms seek protection against competition from the low-cost firms by establishing themselves in the smaller region. This implies a higher productivity level in large markets than in small markets. However, as the global economy gets more and more integrated, the selection effect is turned upside down, the market access effect stressed above becoming the dominant force. Consequently, as transport costs decline, interregional productivity differences first increase and then decrease. Note also that the least efficient firms go out of business because global competition is too tough for them to survive in either region.

The HME cannot be readily extended to multi-regional set-ups because there is no obvious benchmark against which to measure the “more than proportionate” share of firms. But why should one bother about the existence of many regions instead of two? The new fundamental ingredient that a multi-regional setting brings about is that the accessibility to spatially dispersed markets varies across regions. In other words, the *relative position* of a region within the network of exchanges (which also involves cultural, linguistic and political proximity) matters. Any global (local) change in this network such as market integration (the construction of a major transportation link) is likely to trigger complex effects that vary in non-trivial ways with the properties of the graph representing the network (Thomas, 2002). When there are only two regions, the overall impact can be captured through the sole variation in transport costs. On the contrary, when there are many regions, a change that directly affects two regions generates general equilibrium effects that are unlikely to leave the remaining regions unaffected. In particular, a multi-regional setting should make it possible to study how lowering transport costs amplify or reduce the geographical advantage and disadvantage held by different regions.

Unfortunately, economic geography and urban economics do not have much to say regarding those questions, although the evidence shows that accessibility strongly affects the potential of regions and cities for development (Collier, 2007). To illustrate, Limão and Venables (2001) show that, in comparison with the median coastal country, the median landlocked country bears an additional transport cost of 55%, while its volume of trade at the same income level and distance decreases by 60%. Differences in accessibility have another facet which is often ignored: the level of human capital is higher in regions with a greater market access (Redding and Schott, 2003). With this in mind, it should be clear that accounting explicitly for a multi-regional economy with different transport costs is a critical issue (Behrens et al., 2010). Given the high analytical complexity of the problem, there is a need for computable and calibrated general spatial equilibrium models coping with several sectors and regions connected through a network having a specific design. In particular, what we have seen in section 2.2 shows that strategic choices on how to extend or reform transport networks is very likely to affect the location of firms in ways that should be carefully investigated through such models.

3.2 The emergence of a core-periphery structure

While firms bring with them the benefits of added production capability, the returns from physical capital need not be spent in the region where it is invested. By contrast, when human capital moves to a new region, workers bring with them both their production and consumption capabilities. As a result, their relocation simultaneously affects the size of labour and product markets in both the origin and the destination regions, expanding in the former and shrinking in the latter. Another major difference is that the mobility of capital is driven by differences in nominal returns, whereas workers move when there is a positive difference in real wages. Indeed, the gap in living costs matters to workers who consume in the region where they work, but not to capital-owners who consume their income in their region of residence, which need not be the region where their capital is invested. When some workers choose to migrate, their decisions change the relative attractiveness of both origin and destination regions. The resulting effects have the nature of externalities because workers do not account for them when making their decisions to move. Moreover, these externalities are pecuniary because prices fail to reflect the true social value of individual decisions when markets are imperfectly competitive.

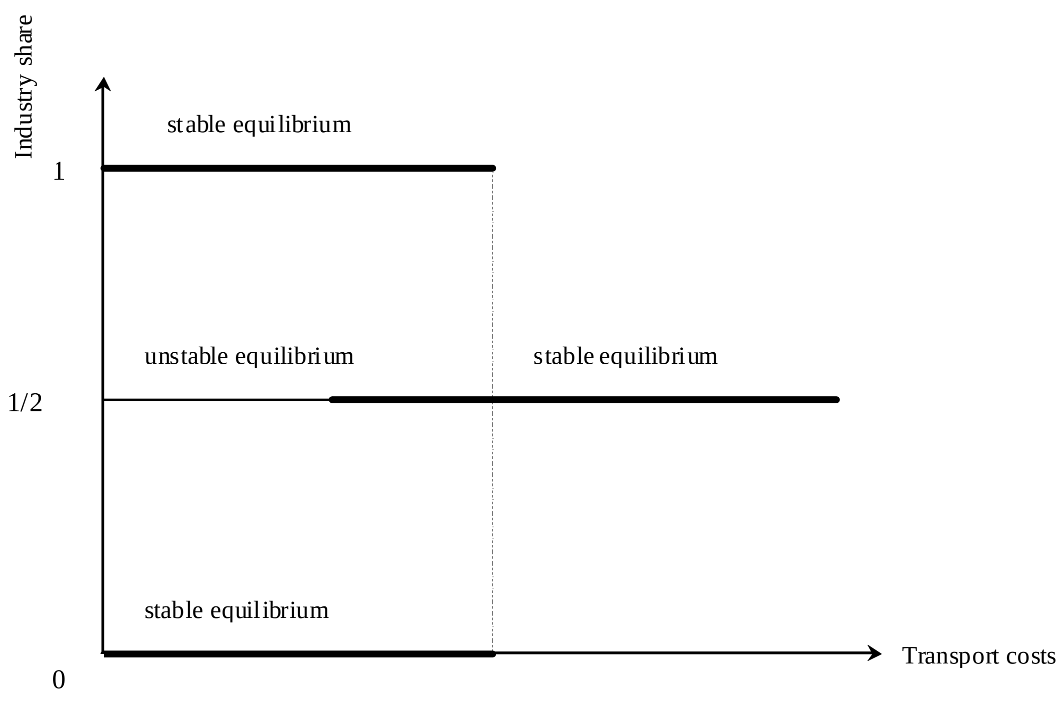
As in the foregoing, let us consider a two-region, two-sector, and two-factor economy. One production factor (unskilled labour) is spatially immobile and used as the input in the traditional sector; the second factor (skilled labour) is spatially mobile and used as the input in the industrial sector. In what has come to be known as the core-periphery model, two major effects are at work: one involves firms and the other workers. Assume that one region becomes slightly bigger than the other. First, a larger market size leads to a higher demand for the industrial goods. This generates a more than proportionate increase in the share of firms, which pushes nominal wages up. Second, the presence of more firms means a greater variety of local products as well as a lower local price index – a cost-of-living effect. Accordingly, real wages should rise, and this region should attract a new flow of workers. The combination of these two effects gives rise to a *cumulative causation* process that leads to the agglomeration of firms and skilled workers in a single region – the *core* of the economy, while the other region becomes the *periphery*.

Even though this process seems to generate inevitably a “snow ball” effect, it is not so clear that it will always develop according to that prediction. Indeed, the foregoing argument has ignored several key impacts of migration on the labour market. On the one hand, the increased supply of labour in the region of destination will tend to push wages down. On the other hand, the increase in local demand for industrial goods leads to a higher demand for labour. Thus, the final impact on nominal wages is hard to predict. Likewise, there is increased competition in the product market, which makes the region less attractive to firms. The combination of all those effects may lead to a “snowball meltdown”, which could result in the spatial dispersion of firms and workers.

Turning to the specific conditions for agglomeration or dispersion to arise, Krugman and others have shown that *the level of transport costs is the key-parameter* (Krugman, 1991; Fujita *et al*., 1999). On the one hand, if transport costs are sufficiently high, interregional shipments of goods are discouraged, which strengthens the dispersion force. The economy then displays a symmetric regional pattern of production in which firms focus mainly on local markets. Because the distribution of workers is the same within each region, spatial disparities vanish in that there are no interregional price and wage differentials. On the other hand, if transport costs are sufficiently low, then all firms will concentrate into the core, while the periphery will retain the traditional sector only. In this way, firms are able to exploit increasing returns by selling more goods in the region benefiting from the market expansion effects sparked by the migration of skilled workers without losing much business in the smaller market. Thus, the mobility of skilled labour is likely to exacerbate the HME discussed in section 3.1, the reason being that the size of local markets changes with labour migration. Figure 2 shows how sudden and big is the shift in the interregional distribution of the industrial sector.

Capital mobility and labour mobility are, therefore, not equivalent for the spatial organization of the economy. While spatial inequalities in section 3.1 reflect the exogenous distribution of capital-ownership, in the core-periphery setting they stem from the endogenous redistribution of human capital.

Figure 2



Despite its extreme nature, the above prediction provides a fairly neat description of the spatial unevenness of economic development observed in different periods and different continents. To illustrate, consider Bairoch’s (1997) estimates of the GDP per capita over the period 1800-1913 across European countries. This corresponds to a period of intense technological progress that preceded a long series of political disturbances.

Table 1: **Per capita GDP of European countries expressed in 1960 USD**

Countries	1800	1830	1850	1870	1890	1900	1913
Austria-Hungary	200	240	275	310	370	425	510
Belgium	200	240	335	450	55	650	815
Bulgaria	175	185	205	225	260	275	285
Denmark	205	225	280	365	525	655	885
Finland	180	190	230	300	370	430	525
France	205	275	345	450	525	610	670
Germany	200	240	305	425	540	645	790
Greece	190	195	220	255	300	310	335
Italy	220	240	260	300	315	345	455
Netherlands	270	320	385	470	570	610	740
Norway	185	225	285	340	430	475	615
Portugal	230	250	275	290	295	320	335
Romania	190	195	205	225	265	300	370
Russia	170	180	190	220	210	260	340
Serbia	185	200	215	235	260	270	300
Spain	210	250	295	315	325	365	400
Sweden	195	235	270	315	405	495	705
Switzerland	190	240	340	485	645	730	895
United Kingdom	240	355	470	650	815	915	1035
Mean	200	240	285	350	400	465	550
Coefficient of variation	0,12	0,18	0,23	0,31	0,38	0,39	0,42

Source: Bairoch (1997).

Even if the numbers given in Table 1 must be used cautiously, they reveal clear tendencies. First, in 1800, most countries, except the Netherlands and, to a lesser extent, the United Kingdom, had fairly similar incomes per capita. As the Industrial Revolution developed and spread across the continent, each country experienced growth: the average GDP increases from 200 dollars in 1800 to 550 dollars in 1913. However, this process affected countries in a very unequal way. This is shown by the rise of the coefficient of variation that rose from 0.12 to 0.42, which confirms the existence of strongly rising spatial inequalities. Second, countries with the highest growth rates are those located close to the United Kingdom, which became the centre of the global economy of the nineteenth century. This is readily verified by means of a regression of the logarithm of the GDP per capita on the logarithm of the distance to the UK, which shows that the impact of this variable is significantly negative. Moreover, the absolute value of this regression coefficient, which has the meaning of elasticity, rises from 0.090 in 1800 and reaches a peak equal to 0.426 in 1890 (and remains stable afterwards). Stated differently, before the Industrial Revolution, a decrease of 10% in the distance to the UK is accompanied by an increase of the GDP per capita equal to 0.9%. By World War I, this elasticity had reached 4.4%, thus showing how far spatial inequalities had evolved during the 19th century.

It is worth stressing that the emergence of the European core-periphery structure arose while transport costs were falling at a historically unprecedented pace. According to Bairoch (1997), on the whole, between 1800 and 1910, the reduction in the real average prices of transportation was on the order of 10 to 1. Therefore, while the European economy experienced a rapid growth, *this phenomenal decrease in transport costs was accompanied with an increasingly unbalanced geographical distribution of wealth.* At the interregional level, Pollard (1981) similarly observes that "*the industrial* *regions colonize their agricultural neighbours [and take] from them some of their most active and adaptable labour, and they encourage them to specialize in the supply of agricultural produces, sometimes at the expense of some pre-existing industry, running the risk thereby that this specialization would permanently divert the colonized areas from becoming industrial themselves.*"

Another important implication of the cumulative causation at work in the core-periphery model is the emergence of what can be called a *putty-clay geography*. Even though firms are a priori footloose, once the agglomeration process is set into motion, it keeps developing within the same region. Individual choices become more rigid because of the self-reinforcing nature of the agglomeration mechanism (the snowball effect mentioned above). In other words, the process of agglomeration sparks a *lock-in effect*. Hence, although firms and workers are (almost) freed from natural constraints, they are still connected through complex networks of interactions, which are probably more difficult to unearth than the old location factors related to the supply of natural resources.

4. THE BELL-SHAPED CURVE OF SPATIAL DEVELOPMENT

The core-periphery model overlooks many costs whose origin lies in the space-economy (e.g. the various congestion costs generated by the emergence of an agglomeration). It also leads to a very extreme prediction that might not be robust against the introduction of additional parameters. This is what I want to cover in this section through a few suggestive examples.

4.1 Vertical linkages

So far, agglomeration has been considered as the outcome of a cumulative causation process fed by the mobility of workers. However, agglomeration of economic activities also arises in contexts in which labour mobility is very low, as in most European countries. This underscores the need for alternative explanations of industrial agglomeration. One strong contender is the presence of input-output linkages between firms: the output of one firm can be an input for another, and *vice versa*. In such a case, the entry of a new firm in a region not only increases the intensity of competition between similar firms; it also increases the market of upstream firm–suppliers and decreases the costs of downstream firm–customers.

This is the starting point of Krugman and Venables (1995). Their idea is beautifully simple and suggestive: the agglomeration of the final sector in a particular region occurs because of the concentration of the intermediate industry in the same region, and conversely. Indeed, when firms belonging to the final sector are concentrated within a single region, the local demand for intermediate inputs is very high, thus making this region very attractive to firms producing intermediate goods. Conversely, because intermediate goods are made available at lower prices in the core region, firms producing final goods find that region very attractive. Thus, a cumulative process may still develop that leads to industrial agglomeration within the core region. In this alternative setting, new forces are at work. Indeed, if firms agglomerate in a region where the supply of labour is inelastic, then wages must surely rise. This in turn has two opposite effects. On the one hand, consumers' demand for the final product increases because they have a higher income. This is again a market expansion force, triggered now by higher incomes rather than larger populations. On the other hand, such wage increases also push toward the re-dispersion of firms. Indeed, when the wage gap between the core and the periphery becomes sufficiently large, some firms will find it profitable to relocate in the periphery, even though the local demand for their output is lower than in the core. The agglomeration is thus self-defeating, especially when transport costs are low because demand asymmetries have a weak impact on profits.

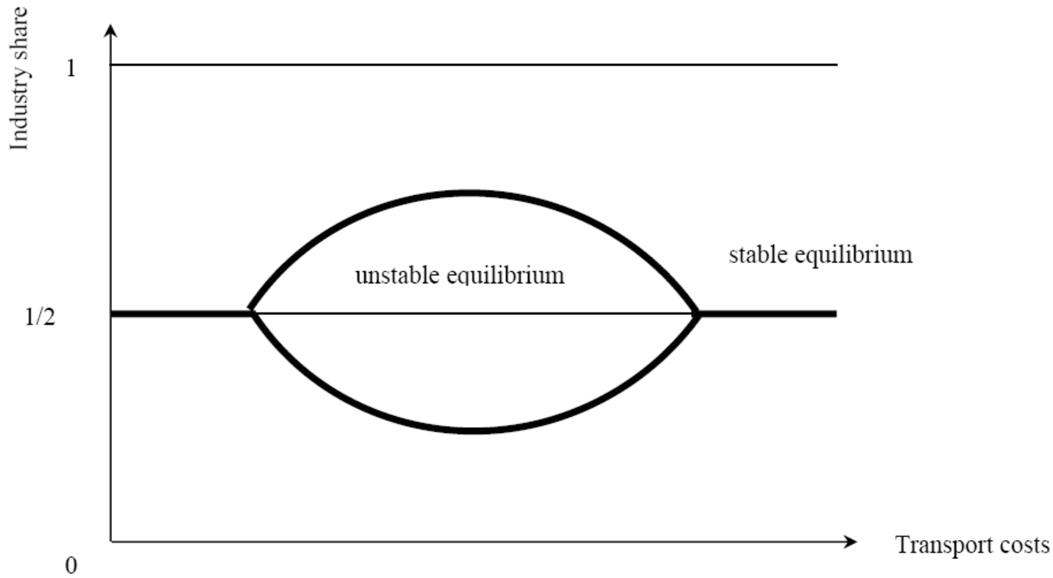
Thus, the set of equilibrium patterns obtained in the presence of vertical linkages is much richer than in the core-periphery model. In particular, if a deepening of economic integration triggers the concentration of industrial activities in one region, then beyond a certain threshold, an even deeper integration may lead to a reversal of this tendency. Some firms now relocate from the core to the periphery. In other words, the periphery experiences a process of reindustrialization. Simultaneously, the core might start losing firms, thus becoming de-industrialized. Therefore, *economic integration would yield a bell-shaped curve of spatial development*. By reducing the tension between the market outcome and the political concern for more spatial equity, the bell-shaped curve of spatial development lends support to a deeper integration of European economies.

4.2 Imperfect labour mobility

In the core-periphery model, workers are assumed to have the same preferences. It is highly implausible, however, that all individuals will react in the same way to a given real wage gap between regions. Some of them show a high degree of attachment to the region where they are born and will stay put even though they may guarantee to themselves higher living standards in another region. In the same spirit, lifetime considerations such as marriage, divorce and the like play an important role in the decision to migrate. Note also that regions are not similar and exhibit different natural and cultural features. Typically, individuals exhibit idiosyncratic tastes about such attributes, so that non-economic considerations matter to potentially mobile workers when they make their decision to move or not. In particular, as argued in hedonic theory of migration, once individual welfare levels get sufficiently high through the steady increase of income, workers tend to pay more attention to the non-market attributes of their environment.

Although individual migrations are difficult to model, it turns out to be possible to identify their aggregate impact on the spatial distribution of economic activities by using discrete choice theory. Recall that discrete choice models, which are widely used in transport analysis, aim at predicting the aggregate behaviour of individuals facing mutually exclusive opportunities such as modal choices. Using the logit model permits to assess the impact of heterogeneity in migration behaviour in that interregional migrations become sluggish (Tabuchi and Thisse, 2002). More precisely, as transport costs steadily decline, more and more skilled workers get agglomerated in one region for the reasons explained in the foregoing section, but the agglomeration process is now gradual and smooth. After having reached a peak in their spatial concentration, skilled workers gradually get re-dispersed. This is because the non-economic factors that drive the choice of a residential location become predominant and take over the economic forces stressed above, the intensity of which decreases with declining transport costs. As a result, the relationship between the degree of spatial concentration and the level of transport costs is bell-shaped (see Figure 3 for an illustration). Therefore, idiosyncratic factors in migration decisions act as a strong dispersion.

Figure 3



Hence, *within the EU polarization should arise on a relatively small scale*. For example, the analysis developed by Crozet (2004) suggests that Lombardy should attract firms within a radius ranging from 95 to 150 km from its centre. Consequently, this region is not expected to threaten any other major Italian region, since the largest city closest to Milan, i.e. Turin, is situated 141 km away, while Genoa and Rome are 164 and 576 km away, respectively.

The sticky mobility of European workers also has an implication that has been overlooked by policy-makers: the relative dispersion of the industrial sector caused by the heterogeneity of preferences is likely to generate efficiency losses at the macroeconomic level. These stem from larger trade flows and insufficient exploitation of scale economies. If so, the low mobility of European workers thus presents two opposite facets: on the one hand, it corresponds to workers' greater attachment to their region or country as embedded in their individual preferences; on the other hand, it gives rise to some losses with respect to productive efficiency, and these are liable to dampen European economic growth.

4.3 The spatial fragmentation of firms

A growing number of firms choose to break down their production process into various stages spread across different places. Specifically, the modern firm organizes and performs its activities in distinct locations, which altogether form a *supply chain* starting at the conception of the product and ending at its delivery. This spatial fragmentation of production aims at taking advantage of differences in technologies, factor endowments, or factor prices across places (Feenstra, 1998). The most commonly observed pattern is such that firms relocate their production activities in low-wage regions or countries, while keeping their strategic functions (e.g. management, R&D, marketing and finance) concentrated in a few affluent urban regions where the high-skilled workers they need are available.

In such a context, the development of new communication technologies is a major force that should be accounted for. It goes hand in hand with the growing role of transportation firms in the global logistics. With this in mind, two types of spatial costs must then be considered, namely communication costs and transport costs. Low transport costs allow firms producing overseas to sell their output on their home market at a low price. Equally important, but perhaps less recognized, is the fact that coordinating activities within a firm is more costly when headquarters and plants are physically separated because the transmission of information remains incomplete and imperfect (Leamer and Storper, 2001). However, lower communication costs make coordination easier and, therefore, facilitate the process of fragmentation. More precisely, in order to make low-wage areas more attractive for the set-up of their production, firms need both the development of new communication technologies and substantial decreases in transport costs.

Assume that each firm has two units, one headquarter and one plant. All headquarters are located in the same region and use skilled labour, whereas plants use headquarter-services together with unskilled labour. A firm is free to decentralize its production overseas by choosing distinct locations for its plant and headquarter. Two main scenarios are to be distinguished as they lead to very different patterns (Fujita and Thisse, 2006). When communication costs are high, all firms are national and established in the core region. Once communication costs steadily decrease, the industry moves toward a configuration in which some firms become multinational whereas others remain national. Eventually, when these costs have reached a sufficiently low level, the economy ends up with a de-industrialized core that retains only firms' strategic functions.

According to the value of communication costs, a fall in transport costs may lead to fairly contrasted patterns of production. When communication costs are high, reducing transport costs leads to a growing agglomeration of plants within the core, very much as in the core-periphery model. Hence, the core region attracts all activities. Things are totally different when communication costs are low. For high transport costs, most plants are still located within the core. However, once these costs fall below some threshold, the relocation process unfolds over a small range of transport cost values. This could explain why the process of de-industrialization of some developed regions seems, first, to be slow and, then, to proceed quickly, yielding a space-economy very different from the initial one. As suggested by the declining part of the bell-shaped curve, the welfare gap between the core and the periphery shrinks. Nevertheless, this catching-up process, which leads to a higher welfare level in the periphery, causes welfare losses in the core.

5. THE TRADE-OFF BETWEEN COMMUTING COSTS WITHIN THE CITY AND TRANSPORT COSTS BETWEEN CITIES

Tradable goods do not account for a very large fraction of the GDP of rich countries. On the contrary, many consumption goods and services are produced locally and not traded between regions. The forces pushing toward factor price equalization within every region thus lead to additional costs generated by the agglomeration of firms and workers within the same region. This in turn increases the cost of living in the larger region and may induce some workers to change place. A natural way to capture this phenomenon is to focus on the housing market where competition gets tougher as more people establish themselves in the same area, thus raising housing and land costs.

As mentioned above, a human settlement of a sizeable scale almost inevitably takes on the form of a city. Typically, a city possesses one main employment centre that gathers together firms, while workers are distributed all around it. Workers seek to reduce their *commuting costs* by choosing a living place in the vicinity of their working place. However, because of the scarcity of land, everybody cannot live close to the city centre. This in turn implies that workers must commute between the workplace and their living place. Competition for land among workers gives rise to a *land rent* that varies inversely with the distance to the city centre, thereby compensating workers living far from their workplace. In other words, there is a trade-off between commuting and housing costs: the former increasing with distance while the latter decrease (Fujita, 1989).

Land rent augmented by commuting costs defines what I call *urban costs*. In most developed countries, they stand for a large, and growing, share of households' budgets. In the United States, housing accounts on average for 20% of household budgets while 18% of total expenditures is spent on car purchases, gasoline, and other related expenses. The latter does not account for the cost of time spent in travelling, which keeps rising. We thus find it reasonable to claim that more than 30% of the income of US households is spent on urban costs. In France, between 1960 and 2000, housing and transportation expenses increased from 23% to 40% of household expenditures, which represents a growth of almost 75% despite an almost quadrupling of the real per capita income. Moreover, as predicted by urban economics, urban costs increase with city size. In the United States, urban costs are less than $15,000 per year in cities like Pittsburgh, Baltimore and Kansas City, but rise to nearly $20,000 per year in, e.g. San Francisco, Los Angeles and New York. Looking at French data reveals that, in 2000, urban costs represented more than 40% of individual incomes in Paris, but around 33% of individual incomes in medium-sized cities. Urban costs play a growing role in shaping the city, but we will see that they also have a strong impact on national urban systems and intercity trade flows.

5.1 The monocentric city

In the monocentric city, firms are agglomerated and form the central business district (CBD), inducing all households to commute between their working place and their residences. It is empirically well documented that firms seek proximity in order to enjoy the various types of benefits generated by the need for strategic information, such as knowledge spillovers, business communications and social interactions (Rosenthal and Strange, 2004). Knowledge, ideas and tacit information generate spillovers from one firm to another. Consequently, if economic agents possess different pieces of information, pooling them through informal communication channels can benefit everyone. Firms get agglomerated in a CBD when external economies are strong, commuting costs are low, or both. This is because firms are able to capitalize on the benefits generated by the various spatial externalities generated endogenously through non-market interactions among firms, without having to compensate workers for their high commuting costs. At the other extreme, firms and workers are mixed across locations, very much as in preindustrial cities endowed with poor urban transport systems. This configuration emerges as an equilibrium outcome when spatial externalities are weak, commuting costs are high, or both (Fujita and Thisse, 2002). In short, *high commuting costs fosters the dispersion of activities within the city*, *whereas low commuting costs leads to the specialization of land use between firms and households*. This is reminiscent of what we have seen in the core-periphery model in that lower mobility costs push toward more agglomeration.

But this is only one side of the coin. Let us return to the core-periphery setting discussed in section 3.2, and assume that a large share of the industrial sector is concentrated in a big city. If transport costs steadily decrease, the urban costs borne by workers within the core become too high to be compensated by a better access to the array of tradable goods. Therefore, dispersion arises once transport costs have reached a sufficiently low level by comparison with commuting costs. Lower urban costs in the periphery more than offset the additional transport costs to be paid for consuming the varieties produced in the core. Consequently, as the costs of shipping goods keep decreasing, the economy involves the following phases: dispersion, agglomeration, and re-dispersion. This is strikingly similar to the bell-shaped curve discussed in section 4. What triggers the re-dispersion of firms and workers is now the crowding of the land market. The relocation of the manufacturing sector away from large metropolitan areas toward medium-sized cities illustrates the impact that high commuting costs and low transport costs may have on firms' locations.

It should be clear that the re-dispersion phase depends on the strength of the spatial externalities among firms as well as on the efficiency of the urban transport means used by workers. The spectacular drop in commuting costs sparked by the near-universal use of cars has facilitated the agglomeration of activities within large cities, and then has delayed the interregional re-deployment of activities. So it is the relative evolution of interregional transport costs and intra-urban commuting costs that determines the structure of the space-economy. Stated differently, what matters for the global economy is not just the evolution of transport costs between regions; what goes on inside the different regions is also crucial.

5.2 The polycentric city

The foregoing argument suggests that workers and firms get re-dispersed because urban costs become very high in the core region. However, once it is recognized that big cities may become polycentric through the development of secondary business centres (SBDs), the average commuting costs and land rent borne by those working in a SBD are lower than those paid by the individuals working in the CBD. Simultaneously, because fewer workers commute to the CBD, the corresponding workers also bear lower urban costs. In sum, workers' welfare becomes higher when the city becomes polycentric. By the same token, firms are able to pay lower wages and land rents while retaining most of the benefits generated by urban agglomerations. For example, Timothy and Wheaton (2001) report substantial variations in wages according to intra-urban location (15% higher in central Boston than in outlying work zones, 18% between central Minneapolis and the fringe counties). Thus, we may expect the escalation of urban costs in large cities to prompt the redeployment of activities in a polycentric pattern.

For this to happen, however, firms located in SBDs must be able to maintain very good access to the inner city, which provides highly specialized business-to-business services (Porter, 1995), which in turn requires low communication costs. Indeed, SBDs have not eliminated the importance of the CBD. This is confirmed by Schwartz (1993) who observes that about half of the business services consumed by US firms located in suburbia are supplied in city centres. In the case of New York, Los Angeles, Chicago and San Francisco, this figure even grows to 65%. The same is true of France, as can be seen from the distribution of higher-order metropolitan functions (executives, engineers, and business service company management jobs, research, commerce, banking and insurance, art). These are more common in city centres than in their periphery. For example, for the Paris urban area, they make up 19.3% of employment within Paris itself, 15.7% in the suburbs, and 6.6% in the outside belt (Julien, 2002). These higher-order functions seek out central positions and major city centres retain specific features relative to SBDs. This implies that firms in SBDs incur an access cost to the main centre when they resort to these higher urban functions. Even if this cost is likely to have sharply fallen with the reduction in communication costs, allowance still has to be made for it.

By introducing communication costs, we account for the fact that agglomeration and dispersion across space may take two quite separate forms because they are now compounded by centralization or decentralization of activities within the same city. When commuting and communication costs are high, the space-economy is likely to be formed by several small cities. In contrast, when communication costs reach low values while commuting costs take intermediate values, large polycentric cities are likely to emerge. Therefore, by facilitating the formation of SBDs, the development of new information and communication technologies slows down the redispersion process. Stated differently, *employment decentralization within the metropolis allows the core regions to retain their primacy* (Cavailhès *et al.*, 2007). Such results shed light on the interplay between different types of spatial friction affecting the location of economic activities between and within urban agglomerations. Historical evidence shows that both trade and commuting costs have been decreasing since the beginning of the Industrial Revolution. Once again, what matters for the organization of the space-economy is the relative evolution of these two costs.

Nevertheless, the emergence of a handful of large polycentric cities dominating the European economic space is not inevitable. High-speed rail (HSR) provides fast and convenient travel between large and medium-sized cities by reducing the opportunity cost of being located in one city rather than another, especially when urban costs are high. If HSR is sufficiently cheap and fast, one can think of this transport mode as stimulating the emergence of several interregional urban systems within the EU. In this case, HSR would stabilize prevailing conurbation patterns within Europe by putting a brake on firms' and skilled workers' tendencies to agglomerate in big cities. This is in line with the European cohesion policy objectives.

All of this draws attention to two facts that policy-makers often neglect: on the one hand, local factors may change the global organization of the economy and, on the other, global forces may affect the local organization of production and employment. Stated in a different way, the local and the global interact to shape the entire economy. This relationship calls for a better coordination of transport policies at the city and interregional levels. In doing so, one should also account for the changes in new information and communication technologies as these ones influence the way firms conduct their business across space.

6. CONCLUDING REMARKS

(i) In 1885, Wilhelm Launhardt, a civil engineer who worked on the construction of transport infrastructures in Germany, noted that "the improvement of means of transport is dangerous for costly goods: these lose the most effective protection of all tariff protections, namely that provided by bad roads." And indeed, we have seen that a policy that systematically aims at improving the accessibility of a small region to the global economy runs the risk of being ineffective in promoting the development of this region. *The cumulative nature of the agglomeration process makes the resulting imbalanced pattern of economic activity particularly robust to various types of shocks*. In other words, affluent regions enjoy the existence of agglomeration rents that single-minded policies cannot easily dissipate. Consequently, the objective of the European Commission being to foster a more balanced distribution of economic activities across European regions, it should add more instruments to its policy portfolio.

(ii) However, we have also seen that the evolution of the space-economy depends on the interaction between several additional forces. The sluggish mobility of workers, the existence of non-tradable goods, the demand for intermediate goods, or the spatial fragmentation of firms, all suggest the existence of a bell-shaped curve linking regional disparities and spatial integration. Taking into account these new forces leads us to believe that a sufficiently extensive economic integration of the space-economy is likely to favour the development of several large urban regions, which could be spread over the entire territory of the EU. Eventually, spatial inequalities at the interregional level would be (partially) reduced through the redispersion of the industrial sector, very much as in the US where this sector is mainly located within medium- or low-population density areas (Glaeser and Kohlhase, 2004). By substituting long-distance commuting for the migration of skilled workers, high-speed rail may play a major role in this process. However, for the HSR to have a significant impact of the location of activities, it is crucial to connect cities that have a high potential of interaction. It would be naive to expect the HSR to become by itself the engine of regional development. On the contrary, such a transport policy must part of a broader and integrated portfolio of instruments. The European Commission and many national governments have spent enough money on building "cathédrales dans le desert."

(iii) During the last decade, the media have embraced the idea that we would be living in a world where the tyranny of distance, which weighed so heavily on human history, would be gone. The spectacular and steadily drop in transport costs since the mid-19th century, relayed by the retreat of protectionism and, more recently, by the near-disappearance of communication costs, is said to have freed economic agents from the need for proximity. In this way, technology and globalization would have joined together to make the traditional geography of activities obsolete, and transform yesterday's world with its peaks and troughs into a "flat world".

Recent empirical and theoretical work in economic geography shows a very different reality. While it is true that the importance of being close to natural resources has largely declined, thus giving firms and households more freedom, distance and location have not disappeared from economic life. For example, by showing that distance remains a major impediment to trade and interactions between spatially separated firms and consumers, the gravity model invalidates the idea that the tyranny of distance would be over (Head and Mayer, 2004). It is worth stressing, however, that market accessibility must be evaluated by *all* the costs generated by the various types of spatial frictions that firms and their customers face when trading goods. Such costs are called *trade costs*. Spulber (2007) refers to them as "the four Ts":

- *Transaction* costs that result from doing business at a distance due to differences in customs, business practices, as well as political and legal climates;

- *Tariff and non-tariff* costs, such as different anti-pollution standards, anti-dumping practices, and the massive regulations that still restrict trade and investment;

- *Transport* costs *per se*, because goods have to reach their consumption place, while many services remain non-tradable; and

- *Time* costs, as, despite Internet and video-conferences, there are still communication impediments across dispersed distribution and manufacturing facilities that slow down reactions to changes in market conditions, while the time needed to ship certain types of goods has a high value.

Transport policies cannot ignore this multi-facet of trade costs, nor their mutual interactions.

(iv) Despite more precise measurements of trade costs, economic geography still fails to provide an explicit description of the interactions between the transport and industrial sectors, or between carriers themselves. In particular, modelling explicitly the transport sector and the formation of freight rates through the strategic behaviour of carriers, as well as competition between transport modes, should attract more attention (Behrens et al., 2009). If trucking may reasonably be approximated by perfect competition in the wake of the Motor Carrier Act of 1980, which abolished most entry barriers and fare controls in the US, railroads are characterized by a small number of firms. Railroads are subject to high fixed costs, as they require heavy infrastructure, thereby creating natural oligopolies that behave strategically.

Moreover, integrating variables specific to the transport activity, such as density economies, market segmentation in the supply of transport services, logistic features, and scheduling considerations should also be addressed. All in all, it should be clear that a more realistic description of the transport sector would make economic geography and urban economics more appealing and relevant to transportation economists. This entire area is strongly under-analyzed and deserves much more attention in the future research agenda.

(v) Economic geography has chosen to focus on the historical trend of falling trade costs. Yet, one may wonder whether an increase in trade costs would bring the economy back to the initial situation. The answer is probably not. Even though the agglomeration process is not completely irreversible, the putty-clay nature of the space-economy and the existence of agglomeration rents imply a strong inertia in the location of economic activities. In this respect, it also worth stressing that economic geography models often exhibit hysteresis in which a lag occurs between the application and the removal of lowering trade costs and its subsequent effect on the location of agents.

(vi) How to design "optimal" transport policies remains the most difficult issue. Policy recommendations depend primarily on what decision-makers want to optimize: global efficiency, spatial equity, the ecological footprint, or a combination of all of them? Cities and industrial clusters are replete with different types of externalities, namely interactions that are not mediated by the market. Although the process of interaction goes both ways, individuals worry only about their role as "receivers" but neglect the fact that they are also "transmitters" to the others. As a result, the optimal distribution of firms is more concentrated than the equilibrium one (Fujita and Thisse, 2002). This may come as a surprise since the conventional wisdom is that market cities are too crowded in the vicinity of the centre. Note, however, that this conclusion does not take into account the various negative externalities generated by congestion and pollution. This makes the overall assessment of land-use patterns in cities especially hard. One clear recommendation emerges from theoretical and empirical studies: for the agglomeration economies to produce their effects, the intra-urban mobility is crucial. To avoid free-ridding and coordination failures, the optimal governance of cities should cover the whole area under consideration in order to permit the internalization of all costs and benefits (Cheshire and Magrini, 2009).

At the interregional level, the reasons for over- or under-agglomeration have more to do with linkages between firms and consumers-workers, through product and labour markets. Pecuniary externalities are critical because firms and workers do not account for the impact that their decisions to move have on the well-being of those who stay put as well as on those who live in the region of destination. Consequently, when migration flows are substantial, one may expect the interregional economy to be inefficiently organized. Preliminary analysis suggests that the mobility of firms and workers may yield a pattern of activities which is too concentrated. When some share of skilled workers finds it individually desirable to move to the larger region, the impact on the other skilled workers may be negative because the fiercer competition sparked on the local market is not outweighed by the better penetration of the smaller region. Hence, very much as in a huge prisoner's dilemma, the moving workers may end up being worse off after having moved than before moving. On the other hand, when the spatial economy is sufficiently integrated, the gains stemming from a better exploitation of scale economies become predominant, making the agglomeration of the industrial sector globally efficient. Note also that the over-agglomeration result does not account for the fact that technological progress brings about new types of innovative activities that benefit from being agglomerated, such as the R&D sector. This in turn may boost the growth rate of the global economy (Fujita and Thisse, 2002).

Last, we have seen that global forces are likely to affect the local organization of production and employment, whereas local factors may well change the global organization of the economy. This calls for the integration of the various types of spatial friction acting at different spatial scales. Such a task is probably out of reach for the time being, but it should guide us in setting the research agenda in transport analysis and in designing more effective policies.

BIBLIOGRAPHY

Bairoch, P. (1997), *Victoires et déboires, Histoire économique et sociale du monde du XVIe siècle à nos jours*, Paris: Editions Gallimard.

Behrens, K., C. Gaigné and J.-F. Thisse (2009), "Industry Location and Welfare When Transport Costs Are Endogenous", *Journal of Urban Economics*, 65, 195-208.

Behrens, K., A.R. Lamorgese, G.I.P. Ottaviano and T. Tabuchi (2010), "Beyond the Home Market Effect: Market Size and Specialization in a Multi-country World", *Journal of International Economics*, forthcoming.

Cavailhès, J., C. Gaigné, T. Tabuchi and J.-F. Thisse (2007), "Trade and the Structure of Cities", *Journal of Urban Economics*, 62, 383-404.

Cheshire, P. and S. Magrini (2009), "Urban growth drivers in a Europe of sticky people and implicit boundaries", *Journal of Economic Geography*, 9, 85-116.

Collier, P. (2007), *The Bottom Billion, 'Why the Poorest Countries Are Failing and What Can Be Done About It*', Oxford: Oxford University Press.

Crozet, M. (2004), "Do Migrants Follow Market Potentials? An Estimation of a New Economic Geography Model", *Journal of Economic Geography*, 4, 439-458.

Diamond, J. (1997), *Guns, Germs, and Steel: The Fate of Human Societies*, New York: W.W. Norton.

Duranton, G. and D. Puga (2004), "Micro-foundations of Urban Increasing Returns: Theory", in: J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: North Holland, 2063-2117.

Feenstra, R.C. (1998), "Integration of Trade and Disintegration of Production in the Global Economy", *Journal of Economic Perspectives*, 12 (4), 31-50.

Fujita, M. (1989), *Urban Economic Theory, Land Use and City Size*, Cambridge: Cambridge University Press.

Fujita, M., P. Krugman and A.J. Venables (1999), *The Spatial Economy. Cities, Regions and International Trade*, Cambridge, MA: The MIT Press.

Fujita, M. and J.-F. Thisse (2002), *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*, Cambridge: Cambridge University Press.

Fujita, M. and J.-F. Thisse (2006), "Globalization and the Evolution of the Supply Chain: Who Gains and Who Loses?" *International Economic Review*, 47, 811-836.

Glaeser, E.L. and J.E. Kohlhase (2004), "Cities, Regions and the Decline of Transport Costs", *Papers in Regional Science*, 83, 197-228.

Head, K. and T. Mayer (2004), "The Empirics of Agglomeration and Trade", in: J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Vol. IV, Amsterdam: North-Holland, 2609-2669.

Helpman, E. and P.R. Krugman (1985), *Market Structure and Foreign Trade*, Cambridge, MA: The MIT Press.

Julien, P. (2002), "Onze fonctions pour qualifier les grandes villes", *INSEE Première*, N° 840.

Krugman, P.R. (1991), "Increasing Returns and Economic Geography", *Journal of Political Economy*, 99, 483-499.

Krugman, P.R. and A.J. Venables (1995), "Globalization and the Inequality of Nations", *Quarterly Journal of Economics*, 110, 857-880.

Leamer, E.E. and M. Storper (2001), "The Economic Geography of the Internet Age", *Journal of International Business Studies*, 32, 641-655.

Limão, N. and A.J. Venables (2001), "Infrastructure, Geographical Disadvantage, Transport costs, and Trade", *World Bank Economic Review*, 15, 451-479.

Midelfart-Knarvik, K.H. and H.G. Overman (2002), "Delocation and European Integration: Is Structural Spending Justified?", *Economic Policy*, 35, 321-359.

Pollard, S. (1981), *Peaceful Conquest. The Industrialization of Europe 1760-1970*, Oxford: Oxford University Press.

Porter, M.E. (1995), "Competitive Advantage of the Inner City", *Harvard Business Review*, May-June, 55-71.

Redding, S. and P.K. Schott (2003), "Distance, Skill Deepening, and Development: Will Peripheral Countries Ever Get Rich?", *Journal of Development Economics*, 72, 515-541.

Redding, S. and A. Venables (2004), "Economic Geography and International Inequality", *Journal of International Economics*, 62, 53-82.

Rosenthal, S. and W. Strange (2004), "Evidence of the nature and sources of agglomeration economies", in: J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: North Holland, 2119-2171.

Schwartz, A. (1993), "Subservient Suburbia", *Journal of the American Planning Association*, 59, 288-305.

Spulber, D.F. (2007), *Global Competitive Strategy*, Cambridge: Cambridge University Press.

Syverson, C. (2004), “Market Structure and Productivity: A Concrete Example”, *Journal of Political Economy*, 112, 1181-1222.

Tabuchi, T. and J.-F. Thisse (2002), “Taste Heterogeneity, Labor Mobility and Economic Geography”, *Journal of Development Economics*, 69, 155-177.

Timothy, D. and W.C. Wheaton (2001), “Intra-urban Wage Variation, Employment Location and Commuting Times”, *Journal of Urban Economics*, 50, 338-366.

Thomas, I. (2002), *Transportation Networks and the Optimal Location of Human Activities: A Numerical Geography Approach*, Cheltenham, UK: Edward Elgar.

INTRODUCTORY REPORTS

Theme I:

Trends and Developments in Interurban Travel Demand

THE PROSPECTS FOR INTER-URBAN TRAVEL DEMAND

Yves Crozet

Transport Economics Laboratory
University of Lyon
Lyon
France

SUMMARY

INTRODUCTION

Mobility has increased enormously since the early days of the industrial era. Successive industrial revolutions have brought new, faster and relatively less expensive opportunities for both passengers and goods. If a contemporary of James Watt (1736-1819) or George Stephenson (1781-1848) were to return to Britain today, or to anywhere else in Europe, he would doubtless be astonished by the incredible mobility that is such an integral part of our activity schedules. His greatest surprise would not be at the number of our daily journeys (between three and four), or even the intensity – one might say the feverish pace – of our activity. Those features already existed in Europe's major capitals, and Paris traffic jams have been famous for centuries!

The great difference between our journeys and activity schedules and those of our forebears lies in the much longer distances we travel. By road, and even more so by rail and air, nowadays we can cover hundreds or even thousands of miles in a few hours. Inter-urban mobility is directly affected by these developments. Where international travel by coach and sailing ship used to take weeks, and intercontinental journeys sometimes even longer, we now count the time in hours. The transport revolution has played a major part in the economic history of the last two centuries (Niveau and Crozet, 2000), but it must be emphasized that the change has been gradual. Over two hundred years have passed between the stage-coach and the high-speed train, the clipper and the jet, during which technological progress and the higher speeds it enables have spread relatively slowly. Even with key technological revolutions like the railways, the automobile and the aeroplane, it took several decades for them to become available to the population at large.

From this slow percolation of technological progress into the way we live has arisen the idea that steadily increasing mobility is a structural given of modern society. Further, faster seems to have become the general rule, to such an extent that even space travel, so we are told, will become more widely available in the relatively near future. A few very wealthy people have already become the world's first space tourists.

It is the self-evident nature of this long-term trend towards increased mobility that we wish to examine in this report, since a number of factors could well undermine the relatively classic assumption that past trends will continue into the future.

The first factor that comes to mind concerns energy-related and environmental constraints. Can a world with seven billion inhabitants, and probably nine or ten billion to come, support a way of life currently available to only a minority of its people? Will we have enough energy? Fossil fuels are not inexhaustible. Moreover, and well before they start to give out, they make a major contribution to greenhouse gas emissions and are used extensively in all forms of transport.

Another issue, partly linked to the first, is that of the sustainability of economic growth. Higher mobility is directly linked to increased purchasing power and hence increased GDP. Aren't there limits to growth, as the Meadows report suggested thirty years ago?

A third question, that of lifestyles, though related to the other two, deserves particular consideration. It may be posed in an exaggerated form by supposing the first two problems to have been resolved. Even if we have plenty of cheap energy, without any major external effect, and steadily rising purchasing power, are we and our descendants certain to choose lifestyles in which mobility increases continuously? What will mobility actually look like in thirty to forty years' time?

To answer all these questions, and in so doing to paint a picture of inter-urban mobility in the relatively distant future, we shall start by looking back into the past. Understanding the trends of recent decades is essential to understanding how they could develop and change in the future and where the turning points or breaks might lie. In the first part, our glance in the mirror will be informed by a consideration of the macroeconomic dimensions of the coupling of economic growth and mobility (European Commission White Paper, 2001), not forgetting the microeconomic foundations that shed light on individual behaviour.

In the second part, we will look at factors that have so far appeared constant and at the saturation effects that could call them into question. The scenarios that emerge when the mitigation policies needed to address energy-related, environmental and economic constraints are added to these spontaneous saturation effects are not necessarily a carbon copy of past trends.

1. THE COUPLING OF ECONOMIC GROWTH AND MOBILITY: FROM THE MACROECONOMIC PROOF TO THE MICROECONOMIC FOUNDATIONS

Many retrospective studies show that the mobility of people (and goods) is closely correlated with economic growth, giving rise to the idea of coupling between mobility and standard of living. According to this idea, it is impossible to separate rising standards of living from increasing mobility, whether at macroeconomic level, that of nations, or microeconomic level, that of individual choices. By describing the basis for this coupling, we will highlight the key factors of transport demand, especially passenger demand for inter-urban mobility. We will look at the factors first from a macroeconomic standpoint, then from a microeconomic standpoint.

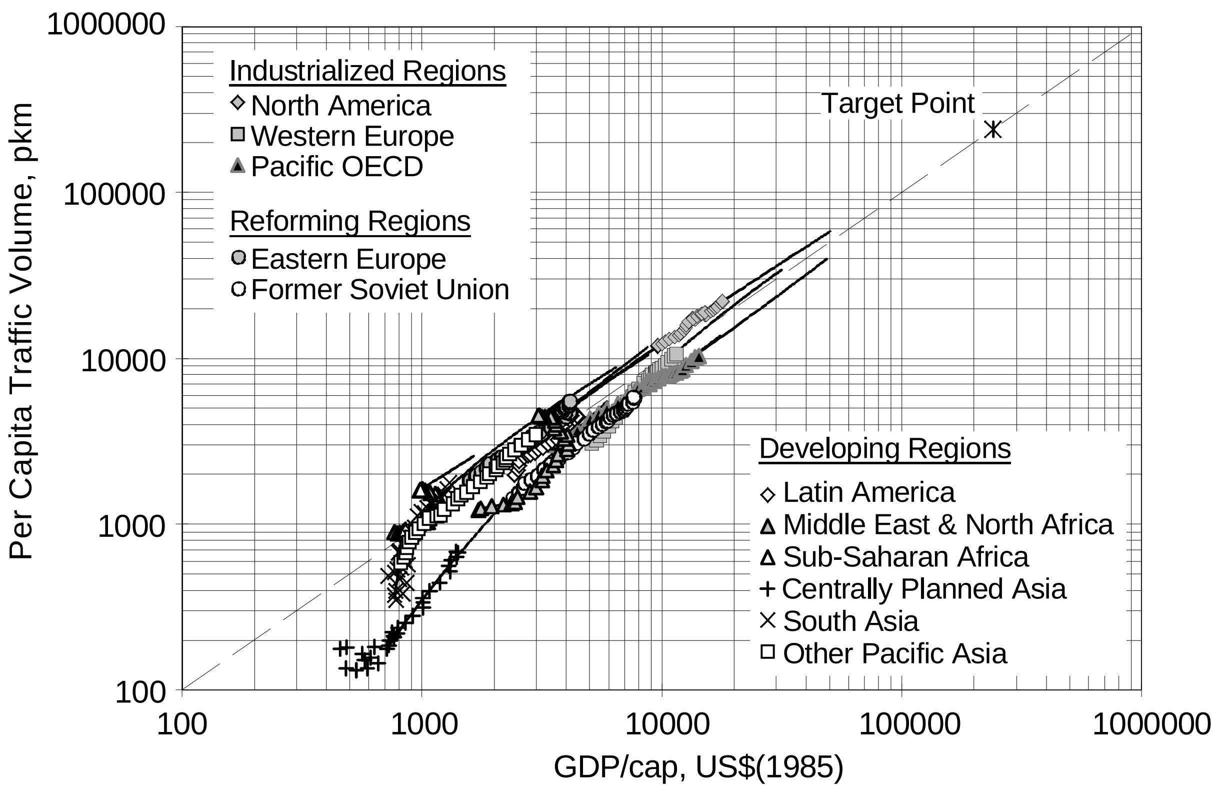
1.1. GDP per capita and transport demand: the "iron law" of coupling

When economists point out that this coupling has been a constant in recent economic history, whatever the country in question, they merely underline the part played by the key factors of economic growth and speed, i.e. the supply of transport and its technological capabilities in particular. We will begin by recalling the proof of coupling before showing that another factor must immediately be added to the key factor of economic growth, namely changes to the structure of transport supply.

1.1.1Coupling between economic growth and mobility: how things stand

After painstaking data collection, Schäfer and Victor (2000) formally established the direct link between economic growth and mobility in the chart below (Figure 1). Using GDP per inhabitant in constant 1985 dollars as a presentational device, they were able to construct a graph in which the first bisector gives a surprising equivalence between the level of GDP and total annual mobility per capita. As most countries are located close to the first bisector, or approach it over time (from 1960 to 1990), one could almost say "Tell me a country's GDP per capita and I will tell you the average distance travelled over a year: one kilometre per dollar of GDP per inhabitant"! As the chart is constructed on a logarithmic scale, we may directly deduce a distance/GDP elasticity of 1. In other words, a given percentage of growth in GDP per capita is matched by an identical percentage of growth in the distance travelled over a year.

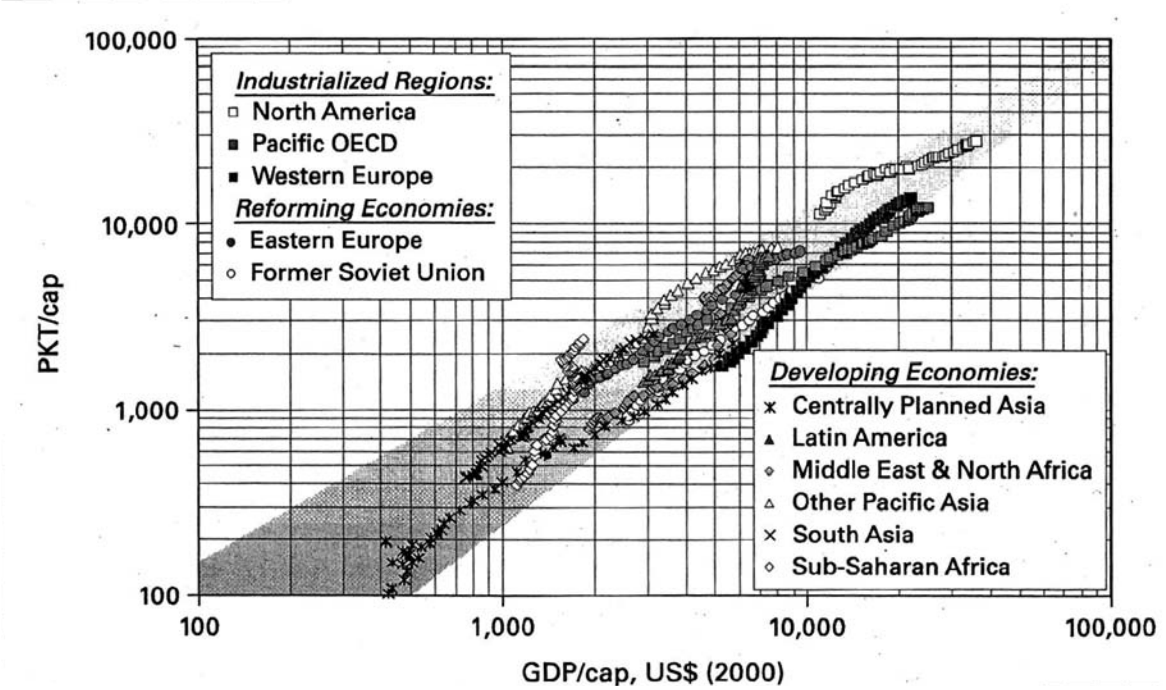
Figure 1. **Total mobility in passenger kilometres per year**
(Data 1960-90; Trends 1960-2050)



Source: Schäfer and Victor (2000); economic growth rates based on IPCC IS92a/e scenario.

The data were updated in a recent study (Schäfer *et al*., 2009), this time including data on personal mobility until 2005, as shown in Figure 2.

Figure 2. **Total mobility in passenger-kilometres per year**
(Data 1950-2005; Trends 2005-2050)



Source: Schäfer *et al.* (2009) *Transportation in a Climate-Constrained World*, MIT Press, p. 36.

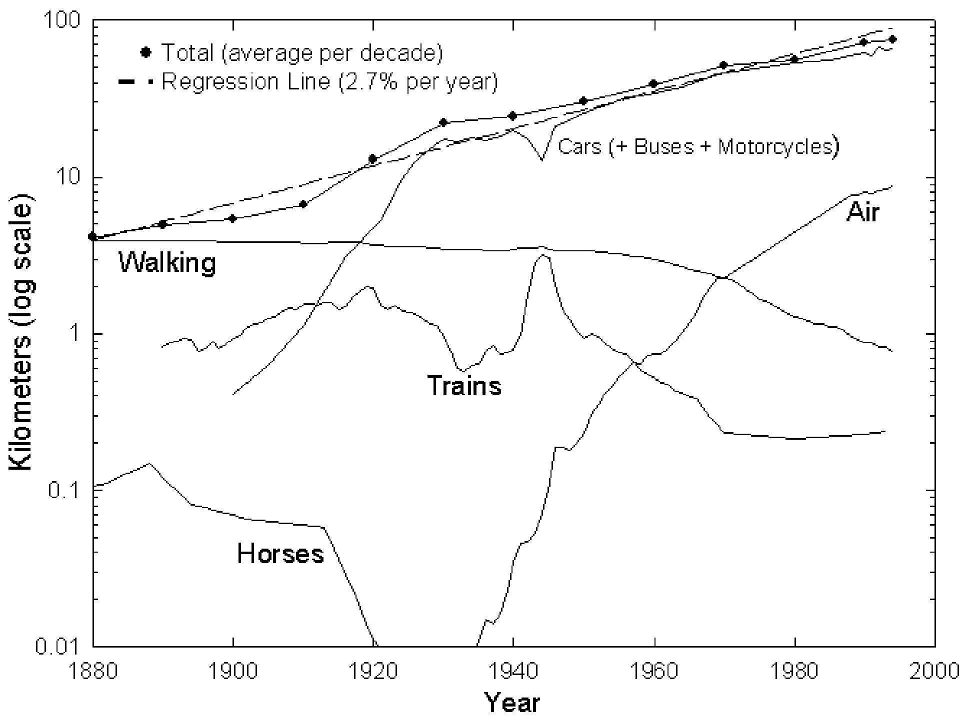
A comparison between Figures 1 and 2 shows, firstly, that coupling is both real and long-standing. In this version, however, taking into account a calculation of purchasing power parities based on constant 2000 dollars, the first bisector effect is eroded. It becomes more difficult to deduce the level of annual mobility per capita from the level of GDP per inhabitant. Taking a standard of living of USD 20 000 on the *x* axis, levels of mobility vary widely, from 10 000 kilometres a year for industrialised countries in the Asia-Pacific zone to 20 000 kilometres a year for North America. That makes it more difficult in Figure 2 to establish a target point like the one in Figure 1. Yet that is what the authors do in Chapter 2 of their book. After emphasising the differences between geographical zones and the fact that the level of GDP does not wholly explain the level of mobility, they nonetheless put forward the possibility of a "target point" that could correspond to a distance of 289 000 kilometres per person per year (180 000 miles a year, or 791 kilometres a day!) and a standard of living of USD 289 000 (constant 2000). This point at which the various countries would converge is no aberration from an economic standpoint. Among economic growth theorists, the idea that affluence is destined to spread on a global scale is frequently assumed (R. Solow). Of course, a level of GDP per inhabitant of nearly USD 300 000 (constant 2000) currently seems extravagant, especially when the world as a whole and the United States in particular is in the middle of a severe economic crisis. But it would be possible if economic growth ran at 3% a year for 75 years, which would multiply GDP per inhabitant eightfold; more or less what has happened in the United States over the last 75 years!

This would bring us back to the logic of alignment on the first bisector. However, the authors emphasize that their world is a hypothetical one that could exist only if the average door-to-door speed for air transport (including travel to the airport and to the final destination) rose from its present level of 270 kilometres per hour to 660 kph, with a transport time budget (TTB) of 1.2 hours a day. The question of speed and time transport budgets is therefore essential to an understanding of past trends and likely future changes.

1.1.2 The key role of speed and the transport system

According to the French economist, François Perroux, economic growth may be defined very simply: it is the growth of an indicator like GDP coupled with structural changes. But these structural changes are often neglected even though they play a key role in the process of per capita output growth. During industrialisation, overall productivity rises only because highly productive sectors account for a relatively greater share of total output. The same applies to mobility, as can be seen from the chart below illustrating the situation in the United States in the 20th century. We can see a steady rise in personal mobility (+2.7% a year), which tracks the rise in GDP per inhabitant over the same period. However, if the average daily distance travelled by an American has risen from 4 km in 1880 to nearly 80 km today (Schäfer, 2009) it is because fast modes have gradually replaced slow modes, allowing the average distance travelled by a person in a year to increase twentyfold.

Figure 3. **Distance travelled in km per person per day since 1800 in the United States**

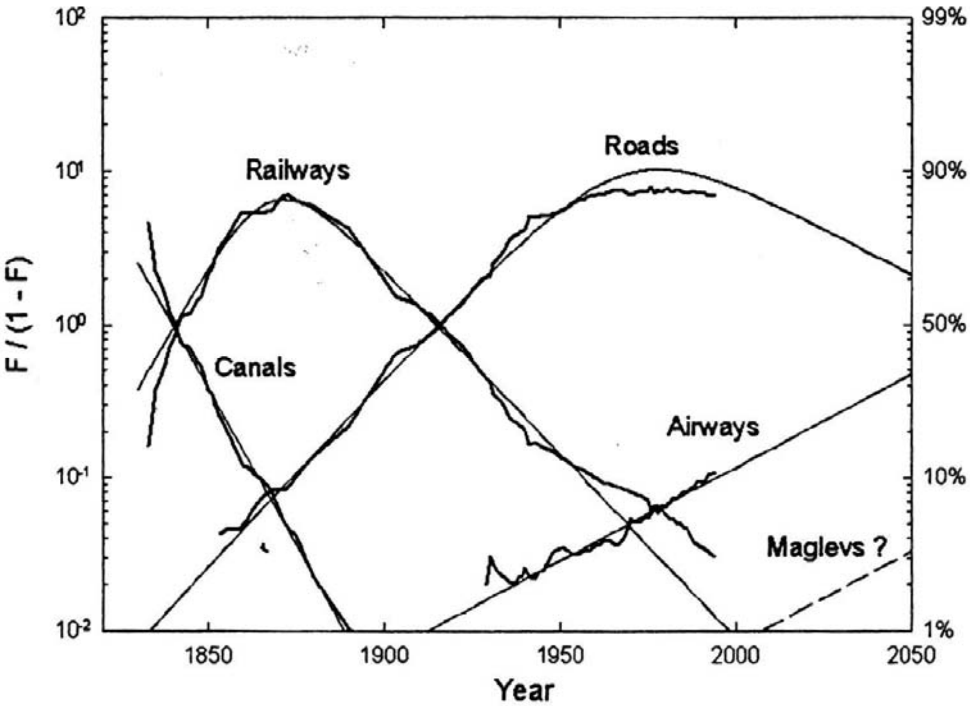


Source: Ausubel J.H., C. Marchetti, P.S. Meyer.

The fact that the coupling is constant therefore presupposes lasting structural changes. The average distance travelled by an American has steadily increased because the automobile has gradually replaced not just the train but also walking and horse-drawn carriages. The construction of a vast network of roads then highways has played a central role in this development. It is not enough for cars to be capable of going fast for journey speeds to rise: transport infrastructure also has to be suited to the capacities of the vehicles that use it.

From this standpoint of permanent structural change, the relative obsolescence that hit the railways in the early 20th century may now be affecting the automobile. In many developed countries, distances travelled by car are no longer increasing, not because total mobility has decreased but because some travel has shifted to faster modes like high-speed trains and aircraft. The growth in the relative share of air transport, perceptible in Figure 1, has been identified as a structural trend by Ausubel, who emphasizes the potential role of magnetic levitation trains[1]. For if it is necessary to continually develop the fastest modes, the history of transport could be depicted as a succession of technological waves. With each new wave, a new transport mode sees its market share increase at the expense of other, slower modes. Then, after reaching a certain level of development, it is itself superseded by another, faster mode.

Figure 4. **Total length of transport infrastructures in the US in market share**



Source : Grübler 1990 (an airline service is considered as a transport infrastructure).

Each new transport mode is faster than the previous one and hence increases the total volume of traffic. The mechanism derives from an implicit assumption that should really be made explicit: the relative constancy of time budgets devoted to mobility. In order for faster average travelling speeds to cause total traffic to rise, it must be assumed that at least some of the time savings are reinvested in additional distance. This hypothesis of the quasi-constancy of time budgets is familiar, in relation to daily mobility, as the Zahavi conjecture. Although it does not directly concern the interregional mobility that is our subject here, we can use the conjecture as an aid to comprehension. We may not yet be able to explain why, but the close link between economic growth and mobility is equivalent to an assumption that speed gains are reinvested in a trend increase in distance travelled (Crozet, 2005).

From the link between distance travelled and GDP, we can therefore move on to another link, namely the one between speed and GDP. If, like Schäfer, we start from the assumption that the total time budget devoted to transport does not decrease, or could even increase slightly, from 1 to 1.2 hours a day, economic growth should be accompanied by an increase in the average speed of travel. In the case of the target point mentioned earlier (289 000 kilometres a year for per capita GDP of USD 289 000), Schäfer *et al.* envisage a speed/GDP elasticity close to 1.

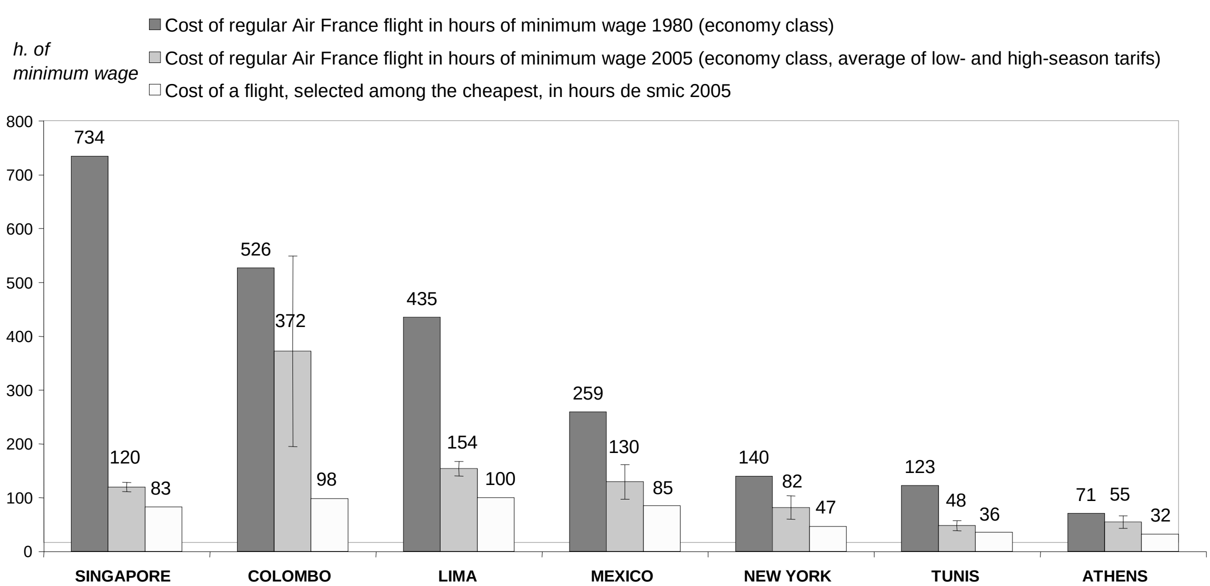
This brings us to the key macroeconomic relationship for understanding how the coupling became so entrenched in recent decades and how it could be called into question in the decades to come. How will the link between average travel speed and GDP evolve in the future? Will the speed/GDP elasticity gradually decline until a certain uncoupling is achieved or, as has been the case in recent decades, will it remain close to 1? In order to answer this question we need to introduce new factors that determine transport demand, including the cost or price of mobility, at the intersection between macro- and microeconomics.

1.1.3 Price and income effects: from the monetary cost to the generalised cost of transport

The target point mentioned by Schäfer and Victor corresponds to a total distance of over 700 kilometres per person per day. Although that is already the case for a handful of frequent fliers[2], is it realistic to suppose that such a lifestyle might become widespread? The question can be asked for the simple reason that transport has a cost not only for mobile individuals – a monetary cost and a time cost – but also for the community, which often has to subsidise infrastructure and in some cases current operations as well.

As regards the monetary cost, Schäfer *et al.* emphasize the trend decline in transport costs. The cost per kilometre of rail travel fell from 20 cents to 5 cents (at constant 2000 dollars) between 1882 and 2002. This fourfold reduction in the real cost should be taken in conjunction with the tenfold increase in per capita GDP over the same period. The experienced cost of mobility has fallen enormously. This combination of price effect and income effect has been a powerful stimulus to mobility. The same phenomenon can be seen in Figure 5 which shows, for France, the change in the price of an air ticket expressed in terms of the number of hours' work needed by a person paid the minimum wage.

Figure 5. **Price of air tickets from Paris to various destinations in hours of minimum wage equivalent (1980-2005)**



Source : Extract from thesis "Optimisation Spatio-Temporelle des Déplacements Touristiques", V. Bagard, LET 2005.

As we can see, the number of hours' work needed to buy a ticket for a typical flight has decreased considerably. The most spectacular fall is in an economy class flight to Singapore, which has dropped from 734 to 120 hours at the minimum wage in France. The reduction is lower for Colombo, a less popular destination for which high- and low-season price differences are still great – so much so, in some cases, as to wipe out the trend decline. It is also instructive to see from this chart that competitors to Air France exist, offering lower prices and leading to an almost tenfold reduction in the cost in terms of hours' work of a ticket to Singapore.

What we have here is a powerful factor behind the growth of air transport, especially as it is less avid for public subsidy than other modes. Most major airports are profitable. To a considerable extent, airport fees and en route charges cover public expenditure on air transport. The same cannot be said of rail transport, especially high-speed trains. The fact that trains require heavy ground infrastructure, which is not the case with aircraft, is a thorny problem for public finances and one to which we will return in the second part. If higher speeds require substantial investment in infrastructure, where is the money to come from? And to what extent can the cost be passed on to users? Should public transport subsidies, which are the rule in urban areas, be extended to inter-urban travel? As we can see, it is not possible to consider the distance/GDP or speed/GDP elasticity without also looking at the question of the cost, for both users and the public purse (Crozet, 2007).

Alongside the monetary cost, the second component of the generalised cost must also be taken into account, namely the cost of time spent in transport. Taking Schäfer's target point, which may serve here as an extreme illustration, travelling more than 700 km a day presupposes very high-speed transport modes. But 660 kph door-to-door may well be difficult to achieve. A significant increase in the time budget devoted to transport must therefore be envisaged. To lay the basis for a forward-looking consideration of inter-urban mobility, we cannot therefore satisfy ourselves with retrospective correlations between economic growth and mobility. We must look for factors that could call past trends into question, and in order to do that we need a better understanding of what motivates individual behaviour. Why does affluence cause us to increase our mobility, including perhaps our transport time budgets? And what mechanisms could undermine this trend?

1.2. When time becomes the "scarcest resource": the "iron law" of diminishing marginal utility

One of the main effects of increased purchasing power is to give us access to a growing number of goods and services. But constantly pushing back the limits of scarcity has not caused the problems of arbitrage that are at the very heart of economics to go away. Encapsulated for Milton Friedman in the famous "no free lunch" quip, the principles of economics do not cease to apply when abundance prevails. Quite the opposite in fact: the very fact that we have a host of goods and services before us will oblige us to make choices, and hence to abandon certain options in favour of others. What are the factors that guide transport demand where inter-urban mobility is concerned?

1.2.1 Intensification of consumption and growth of mobility

Mobility and mobility-related choices present economists with particular problems. The first is linked to the fact that transport is not as a rule sought for itself. Travel demand is derived, a form of joint consumption that is secondary to the linked activity. People do not generally travel for travel's sake but in order to do something else. However, calling travel secondary is probably too reductive for an understanding of transport demand. It would be more accurate to say that travel is subsidiary, insofar as it brings something more to the activity if only by making it possible. So there is something to be gained from studying the demand for travel in itself, taking account among other things of the costs it generates compared to the utility it procures. This can be regarded in two ways.

– From the traditional microeconomic standpoint of consumer choice, it is customary to draw a distinction between inferior, normal and superior goods. These categories help to describe the most commonly observed preferences. As E. Engel then H.H. Gossen showed over a century ago, when income increases consumption of inferior goods declines relative to the other categories. Symmetrically, the proportion of superior goods in household budgets will increase. This applies, for example, to spending on healthcare or education, which ultimately grows faster than income, in contrast with spending on food, which increases much more slowly. Spending on mobility traditionally lies between these two extremes and tends to fall into the "normal" category, where consumption rises more or less in line with income. That is precisely what Schäfer and Victor's chart tells us: reasoning in terms not of a proportion of income but of distance travelled, demand for mobility, a normal good, should increase at exactly the same pace as income.

– As we have already mentioned, however, this trend poses another problem of arbitrage if, like G. Becker or S. Linder, we extend the microeconomic reasoning to the scarce resource of time. If the average rise in speed means that distance travelled can increase in the same way as income without affecting the transport time budget, the arbitrage seems straightforward, in favour of the status quo represented by the constant transport time budget hypothesis. In other words, as time is a scarce resource whose value increases with income, the time component of the overall cost of transport also increases with income. This cost increase should militate against a rise in mobility unless it brings utility gains that exceed the cost increase.

We must therefore take a look at the utility gains resulting from increased mobility. To do so, let us see what S. Linder has to say on the subject. For him, the "leisured class" is not the one described by T. Veblen in the early 20th century. Like other people – even more so in fact –, the idle rich are confronted with the need to constantly choose between different options. The relative scarcity of time compared to the amount of available income is their chief concern. General affluence has extended this type of problem to a large proportion of the developed world's population, including the working population, to the point where time has become the "scarcest resource". As we recalled earlier, average income increased eight- to tenfold during the 20th century, and even more in many industrialised countries, while life expectancy has risen by only a third. As consumers, we therefore face de facto competition between the goods and services made accessible by higher incomes. Yet using many goods and services takes time. In order to solve this equation, we must achieve a trend increase in the quantity of goods and services used per hour available. That in turn means moving towards increasingly intensive lifestyles.

From this standpoint means of transport, especially fast modes, become a powerful way of intensifying consumption, not only because transport itself is a service but also because it gives access to a much wider range of goods and services. The expansion of tourism, especially to exotic destinations, is a perfect illustration. A few days' holiday by the Mediterranean or even much further afield, in the USA or the Maldives, for example, gives our activity schedules an intensity that bears no relation to what we can get from a visit to cousins in the next village. This leisure-related mobility is based on the same determinants as business mobility, the second key component of inter-urban mobility. Intensification processes are at work in both cases and mutually reinforce each other. The intensification of leisure activity (doing more in less time) becomes the pendant to the intensification of business activity in its classic form of higher productivity. The two movements combine to support economic growth, as if to serve as a reminder that the cause-and-effect relation of coupling goes not only from growth to mobility but also in the other direction.

Taking a look at some indicators of leisure activity, the figures speak for themselves.

- During the 1990s, the "leisure and culture" item in current expenditure rose by 16% in the UK, 13% in the USA, 2% in the Netherlands and 1% in France. Some activities very closely related to leisure, like theme parks, leisure centres and above all air travel, are expanding rapidly. The same applies to package tours and all modern forms of a tourism, which implies systematic recourse to market activities. The most significant outcome is the rise in the number of jobs directly or indirectly linked to leisure.
- For the vast majority of people, leisure time is not in contradiction with the consumer society. Although J. Dumazedier was right to point out that leisure was produced by the trend decline in working hours, his predictions about the "leisure civilisation" do not appear to have come to pass. Although working time has fallen on average on the scale of a lifetime, nevertheless we do not feel that we have more time. On the contrary, the abundance of available goods and services and the growing diversification of possible choices increase the pressure on our time budgets.
- The very notion of a time budget underlines the importance of the economic rationale in our behaviour. A philosopher like P. Sansot may sing the praises of slowness and encourage us not to let ourselves be devoured by the race against time characteristic of modern life, but his book has been only moderately successful. As Linder predicted, if we are dealing with a leisured class it is a harried one, flitting from one activity to the next thanks to mobility.
- What we can see here is the iron law of diminishing marginal utility, and its cutting edge becomes sharper as incomes rise. The greater our purchasing power, the faster the marginal utility of a given activity diminishes because other competing activities exist, made accessible by the higher income. Transport is a condition that allows access to these potential activities, especially if the speed increases and the relative price falls.

So it is not surprising that mobility should increase more or less in line with income, since it is merely the condition that allows the variety economy to develop (R. Gronau, 1975). We may also note that the same symmetrical movement animates both passengers and goods. If people do not travel to consume a particular good or service, the good or service comes to the consumer thanks to a mobility that is no less great than that of travellers – quite the opposite in fact!

1.2.2 Speed and the optimisation of activity schedules

Greater mobility is thus a logical sub-product of higher income. Higher speed is a coherent response to the quest for increasingly varied and intensive consumption. However, intensification in turn imposes particular constraints on activity schedules linked to the trend rise in the value of time. When income rises faster than the amount of time available, the value of time also increases, which means that the time budget we are willing to devote to each activity is potentially smaller. Let us take an example. If you spend four hours a day reading and then buy a television or a computer connected to the internet, the utility of the screen will be compared with that of reading. The time spent reading may well fall sharply, as we can see today among children and young people.

The key problem for individuals in today's world is therefore that of time management. Time is a scarce resource, so how should we allocate it to our various activities? One solution is of course to increase the total amount of time available, for example by cutting down on sleep or spending less time on what we regard as our least interesting activities. Lifestyle surveys tell us that the average amount of time we spend asleep has decreased by about an hour in less than a century. But, as Linder predicted, we have also greatly reduced the time we spend looking after our houses and the goods at our disposal. There are so many goods available to us that we are no longer able to devote a lot of time to each one[3].

Can this reasoning be applied to transport time? Since time is a scarce resource, couldn't we reduce our mobility in order to save time and increase the utility of our activities? That is the advice of the slowness devotee: give time more time, allow each activity time in which to flourish, don't flit continually from one activity to another. Even though it may sound sensible, we need to understand that singing the praises of slowness or duration, like the novelist Milan Kundera, calls into question the central assumption in microeconomics of diminishing marginal utility. That is not something to be taken lightly, since the opposite reasoning consists in supposing that the marginal utility of an activity increases, or at least does not diminish, with its duration. Is that realistic when the standard of living is rising? To answer that question it is crucial not to forget that transport demand is derived, a joint consumption associated with other activities. What is at stake is not primarily mobility *per se* but the growing diversification of activities.

For the time being, what we can see is not a reduction in transport time budgets but a reduction in the average duration of each of our activities. We do more things, spending less time on each. But the time devoted to transport does not diminish because maintaining it, together with higher speeds, is the precondition for the increase in the number of our activities. We will demonstrate the truth of this from the example of leisure, a powerful factor behind the rise in inter-urban mobility.

1.2.3 Rise in the value of time and fall in the average time spent on activities: a powerful factor of long-distance mobility

Farther, faster, more often, for shorter periods. Those, in a nutshell, are the trends that underlie our leisure behaviour, as specialists on the subject like J. Gershuny, F. Potier and J. Viard have shown. People take holidays more often but for shorter periods and travel further. How can we explain this paradox, this diversification of destinations coupled with a reduction in the length of stays?

The fact that the trend in our leisure behaviour is towards shorter stays, paradoxically with longer travel distances, is only one aspect of the development of the demand for variety (Gronau and Hamermesh, 2001). The distinguishing feature of modern lifestyles, and what makes them more attractive than previous forms, is the incredible variety of goods and services on offer. But faced with this variety, our choices result from the simple combination of a few key variables. The income level and the value of time combine with the speeds offered by different transport modes as shown in Figure 6[4]. Each axis corresponds to a key variable:

- the south axis represents the level of income;
- the west axis represents the value of time;
- the east axis represents the average distance travelled;
- the north axis represents the length of stay.

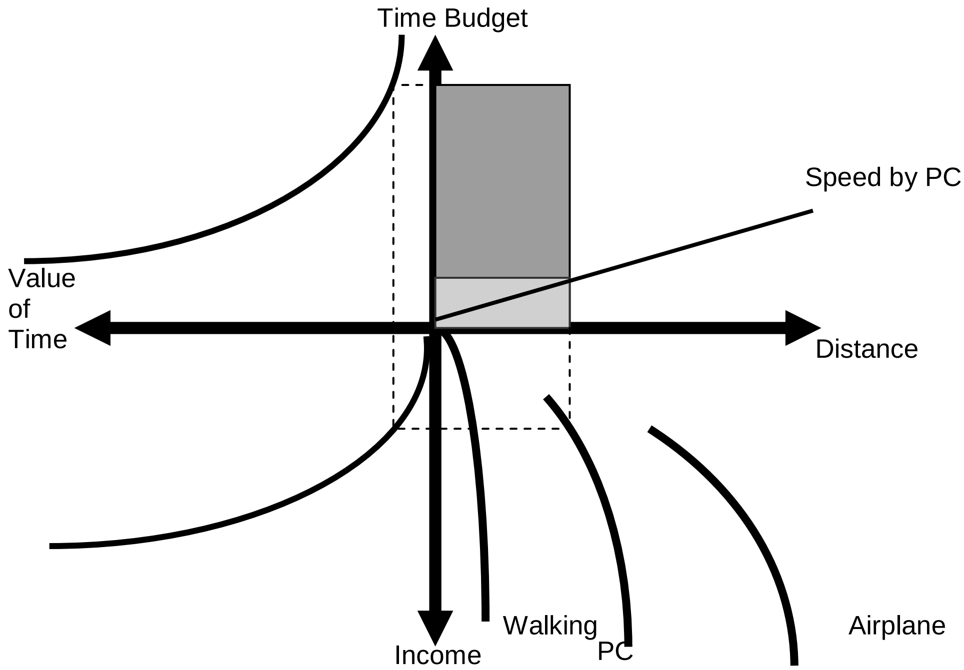
At the intersection of the axis pairs, each quadrant indicates the typical relations between the variables.

- The south-west quadrant assumes that the value of time increases exponentially with income. In other words, the richer we are, the scarcer and more valuable time becomes.
- The north-west quadrant follows on logically from the previous one. If income and the value of time both increase, the time budget we devote to each activity (in this case each leisure trip) will tend to decrease since the competition between the range of potential activities will cause the marginal utility of each activity taken separately to diminish more rapidly.
- The south-east quadrant shows the average speed offered by each transport mode, represented here by the average distance of possible journeys with a given mode. Walking offers few

possibilities at whatever income level. In contrast, rising income progressively gives access to increasingly expensive but increasingly rapid modes, such as road, high-speed rail and air travel.

– The north-east quadrant shows schematically the outcome of the interaction between the different variables, giving an average length of stay determined by the level of income, the value of time and speed (the distance of accessible journeys). All these are linked to a ratio which reveals that transport time represents a certain part of the total length of stay.

Figure 6. **Key variables for the length of holiday stays**

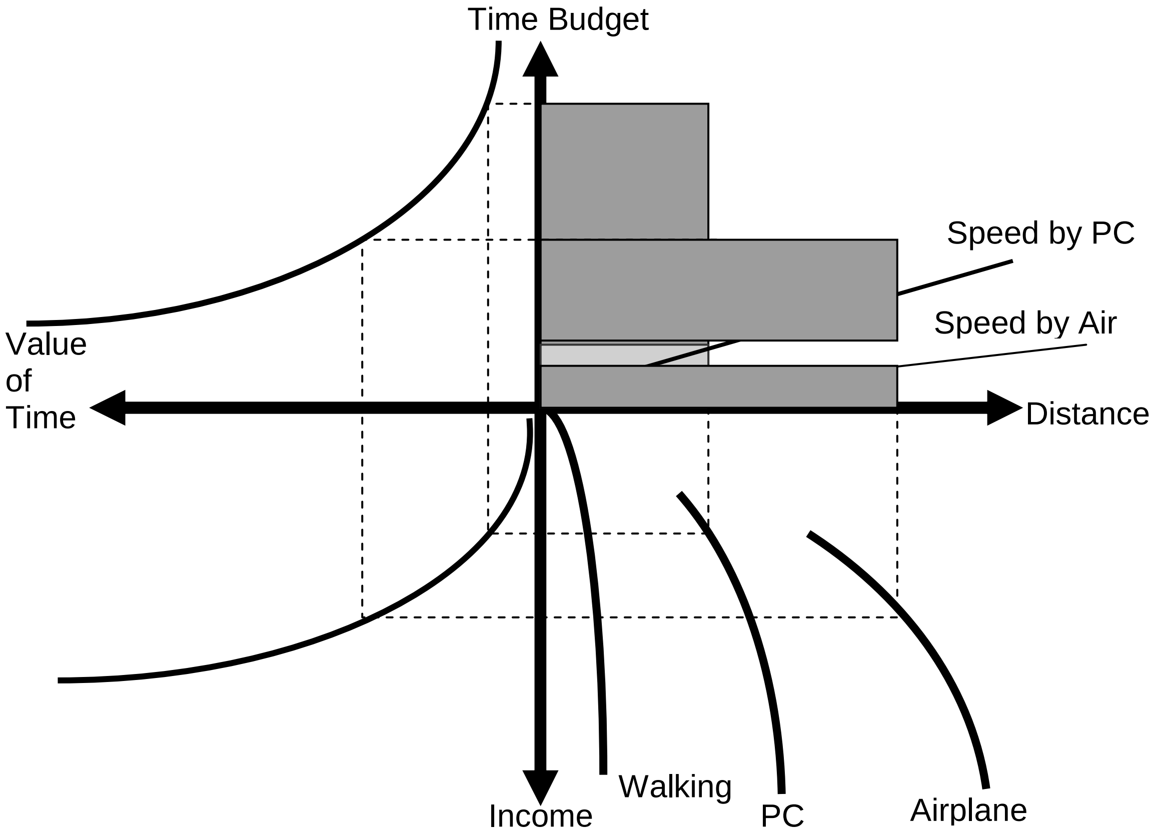


Source: After V. Bagard, 2005.

The stylised facts summarised in Figure 6 are typical of the way family holidays used to be in the 1960s or 70s: a car journey for a relatively long stay (two to three weeks) in the same place. The rise in incomes and in the value of time, combined with new, rapid transport modes, would gradually change this situation, as shown in Figure 7. Access to higher speed was first reflected in an increase in the average distance travelled. Holiday destinations became more and more exotic. But as the increase in speed went hand-in-hand with a rise in the value of time, and hence a reduction in the average length of stay, the result was not a fall but a rise in the ratio of journey time to total length of stay. At the risk of departing from the constancy assumption in this ratio (Mokhtarian, 2004), higher speeds result in the leisure sphere in an increase in transport time as a proportion of the total time spent on the activity. Given the increased utility drawn from the long-distance journey, a higher transport cost is accepted and the transport time budget is pushed up. It is one more reason why time scarcity becomes more acute with the increase in speed and income.

The businessmen and women and academics who read these lines are familiar with what is going on here. Thanks to the speed of air travel, they will often make a two- or three-day trip from one end of Europe to the other or from Europe to the United States for a conference, seminar or thesis committee meeting. The same rationale applies to business trips (which, let us remember, are included for statistical purposes in the general category of "tourism") as to family holidays: farther, faster, more often, for shorter periods. Will the trend continue in the years to come?

Figure 7. **Key variables for the length of holiday stays with access to air travel**



Source: After V. Bagard, 2005.

2. OUTLOOK FOR INTER-URBAN MOBILITY: SATURATION AND MITIGATION AT THE SERVICE OF DECOUPLING?

At a time when sustainable development stands at the top of the agenda, not only for governments but also for business and consumers, there is clearly something to be gained from asking whether mobility can keep on increasing indefinitely.

One simple answer to the question is sometimes given under the heading of degrowth, or zero growth. Proponents of this idea consider that coupling is not merely a correlation but a cause. Economic growth, they argue, lies behind mobility growth. For mobility to be more sustainable, all you have to do is stop growing (Georgescu-Roegen, 1979)! The reasoning behind such a view may seem seductive in its simplicity, though it verges on the simplistic: economic history teaches us that a relation between two variables is not necessarily linear over a long period. The real interest of the notion of sustainable development as described in the Brundtland Report lies in the fact that it goes beyond the simplistic idea that you have to stop growing in order to solve the problems. Sustainable development does not reject growth but seeks – and this is more difficult – to modulate its impacts, as is the case with the notion of mitigation now used extensively in research into environmental issues. In the transport sphere mitigation takes the form of decoupling, which boils down to studying the conditions under which the relationship between economic growth and personal mobility would no longer be linear. Let us therefore maintain the hypothesis of continuing economic growth, even if we are currently in the middle of a full-blown recession.

The fact that the current economic crisis has cut not only air travel but also high-speed rail and even motorway travel should not distract us from the need to take a long-term view. Even if the recession were to go on longer than hoped, and even if the recovery were to be slow, resulting in lower long-term trend growth, that does not mean that we should stop thinking about decoupling, if only because economic growth is continuing in many countries around the world, like China and India, and is accompanied by strong demand for mobility. Fast transport modes like high-speed rail and air travel are continuing to expand. Many countries are building new high-speed rail links. In the air transport sector, companies like Ryanair and EasyJet are carrying more and more passengers despite the crisis.

On the supply side, factors that encourage mobility growth will undeniably be present in the coming years. But it is worth recalling and comparing other factors that could impede the continuation of past trends and even lead to a certain uncoupling of economic growth and mobility.

- First, there is the environmental factor and the commitment to reduce greenhouse gas emissions. One outcome could be tighter restrictions on transport modes that consume the most fossil fuel, which emits large amounts of CO_2;
- Next comes public policy, which is very closely linked. Public policies, in the form of charging, taxation or regulation, can play an important role, especially by encouraging a shift towards transport modes that are not only cleaner but also use up less public space. Modal shift is often sought as a means of reducing the adverse effects of mobility. This would not be decoupling per se (i.e. economic growth with no mobility growth) but a relative decoupling resulting from a favourable structural effect. The replacement of existing technologies with new, cleaner technologies would allow for an increase in traffic while reducing the external

effects of transport, especially CO_2 emissions. The other question that arises, apart from that of the transport mode, is the cost of mobility. Higher energy prices together with less generous subsidies or new taxes, like a carbon tax, could encourage a certain degree of decoupling.

– Changes in individual behaviour will be decisive. Linked to public policy but also as a result of spontaneous changes in preferences, what can be expected from mobility demand? Can we look forward to a certain saturation of demand for inter-urban transport?

We will start in Section 1 by looking at individual behaviour and saturation before describing some scenarios for mobility in France to 2050. This will generate visions of the future (Section 2) in which saturation and mitigation are combined.

2.1. Decoupling and saturation: moving towards a change in individual preferences?

Taken literally, the phrase "farther, faster, more often, for shorter periods" poses logical problems. As we have seen, one trend effect of a rise in the number of activities is to reduce the amount of time spent on each one until it becomes very short. If it also leads to a trend increase in the ratio of transport time to activity time, it is easy to understand that the quest for utility cannot be a permanent quest for speed and more activities. Would it not be possible, then, to imagine a saturation effect which, by limiting the number of activities and hence of journeys, would encourage a minimum amount of time to be spent on each activity? Such an effect may already be at work in the industrialised world, especially in Europe, where automobile traffic has barely increased since the early 2000s. Is it the first sign of uncoupling linked to a saturation of demand for mobility?

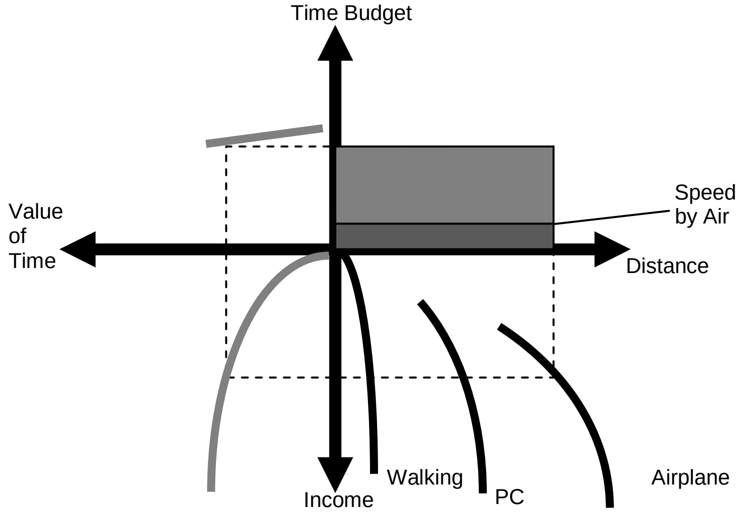
2.1.1 The limits to variety and to the fragmentation of activity schedules

With the effects of the economic crisis, a reduction in business travel has been observed since late 2008. Many firms have sought to cut travel expenses and to replace long-distance travel with communications and video-conferencing. Even before the recession started to bite, sociologists like S. Kesselring had observed a certain "disenchantment" among heavy business travellers. The growing amount of business travel and the associated cost in terms of fatigue is starting to become a specific human resource management problem in firms. In the academic world, we are starting to see thesis defences in which some committee members participate by videoconference. Likewise, with the economic crisis, travel agents and tour operators have noticed a fall-off in demand for travel to exotic destinations and symmetrically, especially in France, a preference for nearby tourist destinations.

This downturn in demand for long-distance transport, perceptible in the decline in air traffic, is for the time being consistent with the stylised situations shown in Figures 6 and 7. Lower income is logically reflected in a decrease in distance travelled and average journey speed, accompanied by a reduction in the value of time and a lengthening of stays. In this instance the trends are still driven by coupling, where economic growth and mobility move together in the same direction, whether up or down. The question is therefore whether the economic crisis is merely a parenthesis or whether it could herald a lasting shift in behaviour towards a certain frugality. Could we see in the future both a rise in income and a saturation of mobility? Figure 8 sketches an initial theoretical answer to that question. As we can see, the key issue is the value of time and its impact on the trend towards the fragmentation of stays.

If, as we can see here, the value of time grows not exponentially but rather logarithmically in relation to income, the relation between value of time and length of stay could take a different form, with the emergence of the equivalent of a minimum duration. The crux of the matter is whether such a hypothesis is realistic. What could prompt people living in developed countries to reduce mobility growth and the associated diversification of activities? The answer could well lie in the limits reached by the fragmentation of activity schedules and the related "zapping". An ageing population could be one factor that triggers such a trend reversal, though it should not be linked to the diminished physical capacities of older people. On the contrary, all the indicators point towards an increase in life expectancy without disability, and retired people are not those who least use cars, trains or aeroplanes for long-distance travel.

Figure 8. **Income, speed and the value of time: another relationship between the variables?**



Source: After V. Bagard, 2005.

What we need to envisage with ageing and affluence is rather a certain wisdom in the use of time, for example by questioning the tendency to reduce the average duration of each activity. Consumption could be intensified not by increasing the number of activities but by giving each one the amount of time it needs to flourish. As S. Linder has suggested, a wise attitude towards growing affluence does not only consist in constantly increasing the quantity of goods and services consumed per hour. For some activities, can we not also seek to preserve a minimum value for the ratio of time spent per quantity of goods or services consumed? The question is worth asking for long-distance travel, where transport time most eats into the length of stay. Among those who already have access to it, might we not see a trend saturation in this type of travel?

2.2. Is decoupling of GDP and passenger mobility already taking place?

Where car journeys are concerned, that question can be answered in the affirmative. If the most recent report from the European Environment Agency is to be believed (EEA Report No. 3, 2009), decoupling in relation to passenger mobility in Europe has already started.

Figure 9 shows that for passengers, unlike freight, GDP growth is generally significantly higher than the trend in overall traffic. The difference between the two confirms the decoupling hypothesis except in 2002, when coupling occurs. The new situation is mainly attributable to relative saturation.

Figure 9. **GDP and total passenger mobility in Europe**

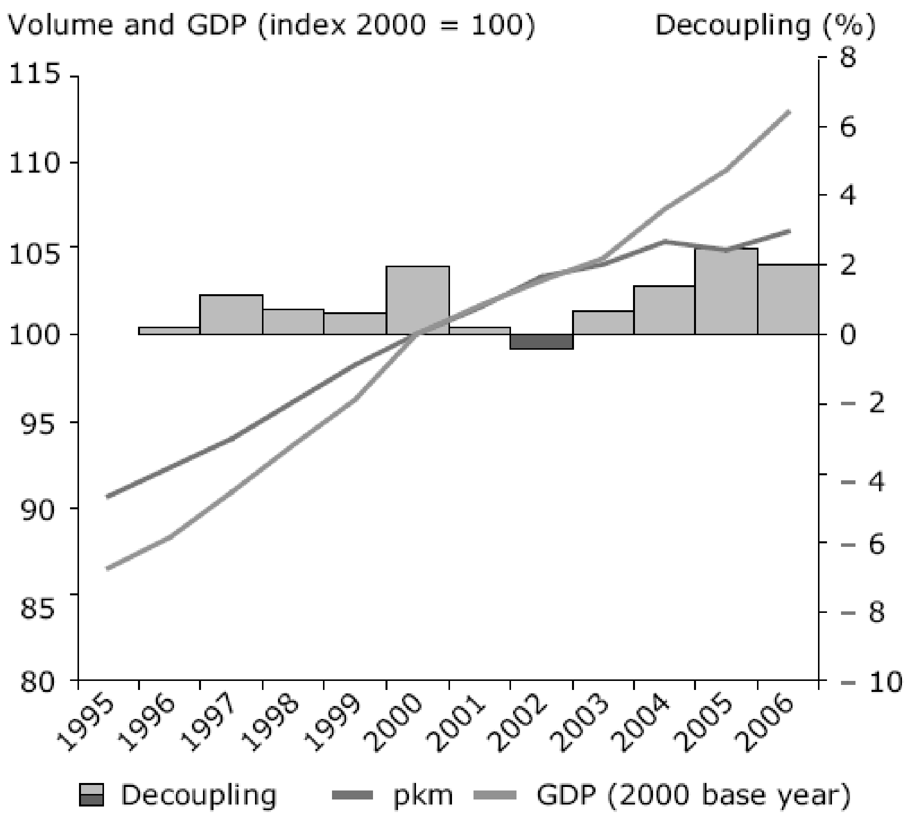


Table 1 shows passenger mobility in the major EU countries. In Germany, the UK, Italy and France, domestic passenger traffic has been more or less flat since the early 2000s.

Table 1. **Passenger traffic in the major EU countries (in billion passenger-kilometres)**

Year	**1995**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Germany	954.8	975.7	997.1	1001.9	996.6	1009.6	998.9	1014.1
France	737.3	812.2	840.1	848.9	853.1	855.3	848.1	848.7
Italy	745.7	867.2	860.0	854.8	854.6	865.2	840.2	845.5
UK	692.6	725.4	740.3	763.9	766.2	770.3	770.4	773.0

Source: European Environment Agency, 2009.

This relative levelling-off of mobility is all the more remarkable insofar as it occurred in a period of fairly significant economic growth. However, it also corresponds to a period of rising fuel prices that hit car drivers particularly hard. The phenomenon accelerated in 2008 when forecourt petrol prices soared in the space of a few months. The number of cars sold in Europe declined significantly over the same period. It was as though the automobile, which accounts for the vast majority of passenger kilometres, had reached a relative obsolescence marking the end of a golden age. Rising petrol prices, combined with constant congestion and speed limits, revealed a trend towards relative saturation. Journeys in urban areas were most affected, together with long-distance journeys facing competition from air and high-speed rail travel. So is this saturation of automobile use really the sign of decoupling or does it merely mark a transition towards fast modes like high-speed trains and aeroplanes?

2.1.3 The persistent growth of long-distance mobility

The European Environment Agency data in Table 1 must be set in context since they relate to domestic traffic in each country. The results are not the same if international traffic, especially air traffic, is included. Sufficient evidence can be obtained by comparing data on transport-related greenhouse gas emissions included in and excluded from the Kyoto Protocol.

- For the 27 countries of the European Union, the former rose from 779 to 992 million tonnes between 1990 and 2006, an increase of 27%. The spread around the average is considerable: -1% for Germany, +17% for France, but +100% for Portugal and +89% for Spain. Not all countries are at the same stage of economic development.
- Still for EU 27, emissions in the latter category rose from 176 to 305 million tonnes, an increase of 73%. Of this total, emissions from air transport alone rose from 66 to 131 million tonnes, with maritime transport accounting for the remainder.

Thus, all transport sector emissions for EU 27 between 1990 and 2006 rose from 955 to 1,297 million tonnes, over 36%. Of this amount, domestic and international air transport emissions rose from 83 to 157 million tonnes. They now represent 12% of total emissions, compared with 8.6% in 1990. This gives us two important signals.

- First, decoupling does not apply to demand for air transport – far from it, in fact. Until the recent economic crisis global air transport was rising faster than global GDP and, given the probably expansion of supply by airlines, the trend is most likely to continue in the years to come. The same is true of high-speed rail travel. Here again, traffic growth has been significantly higher than economic growth in recent years, to the point where many European countries (Spain, Italy, France and Portugal to name just four) are stepping up the construction of new high-speed rail links.

- Second, the very success of air transport will pose problems because of its growing contribution to greenhouse gas emissions. The problem is all the more crucial in that the mode is doubtless far short of reaching saturation. From the standpoint of significantly reducing greenhouse gas emissions, will it not be necessary to take restrictive measures, to go down the road of mitigation?

2.3. Decoupling and mitigation: towards a new set of collective preferences. Three scenario families for inter-urban mobility in France to 2050

The information presented in the following section is taken from projections drawn up for the French Ministry of Ecology and Sustainable Development (Château *et al.*, 2008). It is based on a TILT model (Transport Issues in the Long Term), the broad outlines of which are described in an annex. As always with projections, the model is not supposed to say what will happen: it is not predictive. Its interest lies in its capacity to link a large number of variables while seeking to maintain an overall coherence between them that takes account of various types of constraint which mobility will have to accommodate in the coming decades. More specifically, the approach uses the "backcasting" technique (Clement, 1995, Hickman & Banister, 2005). Bearing in mind the objective of reducing transport-related CO_2 emissions, an objective common to all industrial countries, what developments could take place in aspects such as mobility, modal split and public policy and how might they affect each other? As is customary in this type of work, we started by establishing a trend-based scenario, then developed two scenario families marking inversions of or breaks with previous trends.

2.3.1 Pegasus: trend scenario and key variables to 2050?

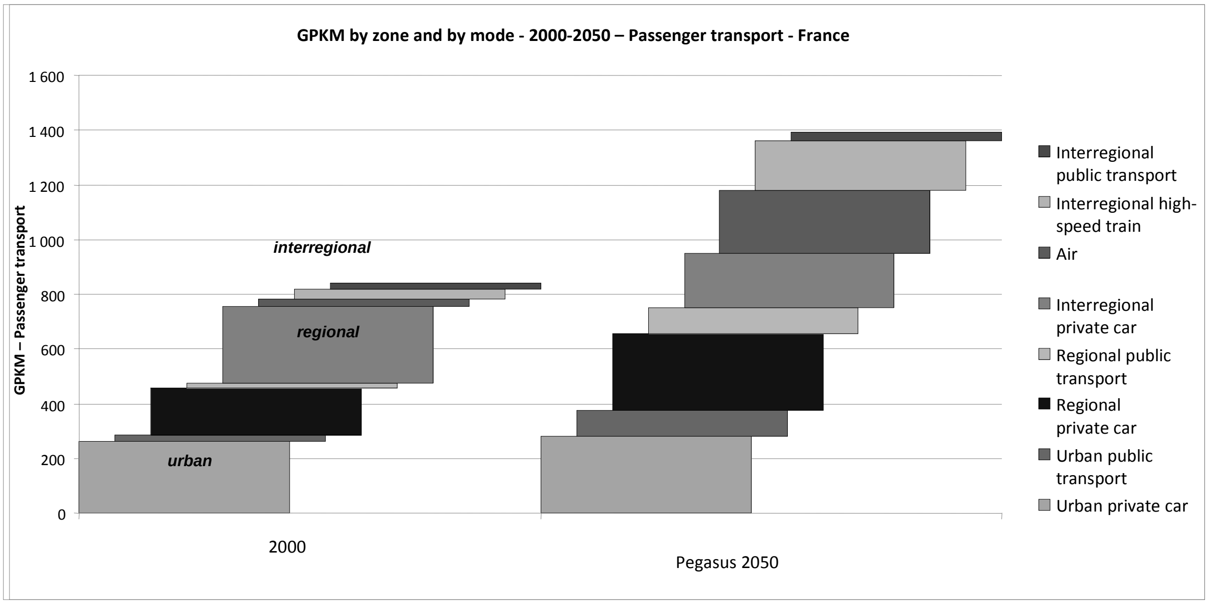
To underpin our projections, let us first assume that the current organisation of our economy and society will remain more or less the same. To encapsulate what is a simple extension of past trends, we named the scenario after a symbolic figure of Greek mythology: Pegasus, the winged horse that enabled Perseus to cover long distances quickly. Are we not already in such a situation, since the average French person nowadays covers over 14 000 km a year, or more than 40 km a day?

Let us start by looking at the results of the TILT model. The Pegasus scenario, which has an infinite number of variants, is summarized in Figure 10.

In relation to the baseline year (2000), the chart shows strong growth in regional and above all inter-urban passenger transport (over 40%). Urban traffic increases by "only" 25% and is marked by a sharp rise in the use of public transport. Growth in travel by high-speed train, bus, metro or tramway is much higher than growth in automobile travel. This corresponds to a shift in mobility choice towards collective modes, not primarily for environmental reasons but because they are the modes where improvements will be seen in the coming years, especially in terms of speed. For in this scenario family we have kept the idea that there is a non-zero elasticity between the average speed of travel and GDP. Rather than Schäfer's hypothesis of an elasticity close to 1, we have taken the actual speed/GDP elasticity in France over the period 1970-2000, namely 0.5, to deduce an arbitrary value of 0.33 for the period 2000-2050. In doing so, we have de facto incorporated a certain saturation of mobility. Because of the pursuit of speed gains we have not limited the growth in air transport, which is a highly effective way of increasing total distance travelled without increasing transport time budgets.

As Figure 10 shows, fast modes gradually replace slow modes. The modal choice shifts systematically towards faster modes (high-speed rail and air travel). As Figures 6 and 7 suggested, higher passenger mobility in terms of kilometres per capita per year is a direct consequence of higher average transport speeds. That is why the saturation rates of different transport modes vary in relation to the speed/GDP elasticity. In other words, relative saturation would occur for long-distance automobile travel. This has already been the case since the early 2000s in France, where the total volume of road and motorway traffic has remained more or less flat. Indicatively, in this scenario CO_2 emissions from passenger transport could be cut by two-thirds or a little more despite rising traffic (see Figure 12) thanks to advances in vehicle technology (automobiles and aircraft) and the emergence of second-generation biofuels. The substantial increase in TGV traffic plays a crucial role here. The scenario therefore concurs with the conclusions presented by Hickman and Banister in the VIBAT project. A forecasting exercise carried out for the UK to 2030, VIBAT indicates that half the targeted reduction in CO_2 emissions can be achieved through technological progress.

Figure 10. **Passenger mobility in France 2000-2050: Pegasus scenario**



However, reducing CO_2 emissions by a factor of three would not be sufficient to comply with Kyoto Protocol commitments and those that will doubtless be made at the Copenhagen climate change conference in late 2009. If global CO_2 emissions are to be halved by 2050, the countries that have been industrialised the longest will have to make a greater effort since they are chiefly responsible for past emissions. From that standpoint, let us take a closer look at scenarios that are more restrictive of personal mobility, especially inter-urban mobility. Changes of behaviour are needed in order to reduce CO_2 emissions by more than the amount made possible by technological progress alone. How are they to come about? To answer that question, we have made modifications to some key parameters in the model - modifications that are apparently benign but presuppose major changes in individual preferences.

The modifications introduced in the two new scenario families concern the following variables.

– First, we suppose that the speed/GDP elasticity becomes zero, **which represents a major break with previous trends**. It is reflected in a small increase in total distance travelled. In the first alternative scenario family, called Chronos, the increase in distance is mainly attributable to a 20% increase in transport time budgets. We have taken up one of the hypotheses put forward by Schäfer (2009), though without linking it to an increase in speed. It offers the possibility of continuing the increase in distance travelled, albeit at a slower pace and without any increase in the average speed. It is because the continuing embrace of mobility is time-consuming that this scenario family has been named Chronos.

– The second scenario family, baptised Hestia, makes the same assumption of a zero speed/GDP elasticity. But going further in the change of behaviour, it is not matched by an increase in transport time budgets. The reduction in average speeds will therefore severely limit the trend increase in distance travelled, indicating a return to proximity activities. This explains the name Hestia, the Greek goddess of hearth and home.

2.3.2 Chronos: lower road speeds but economic growth still coupled with mobility

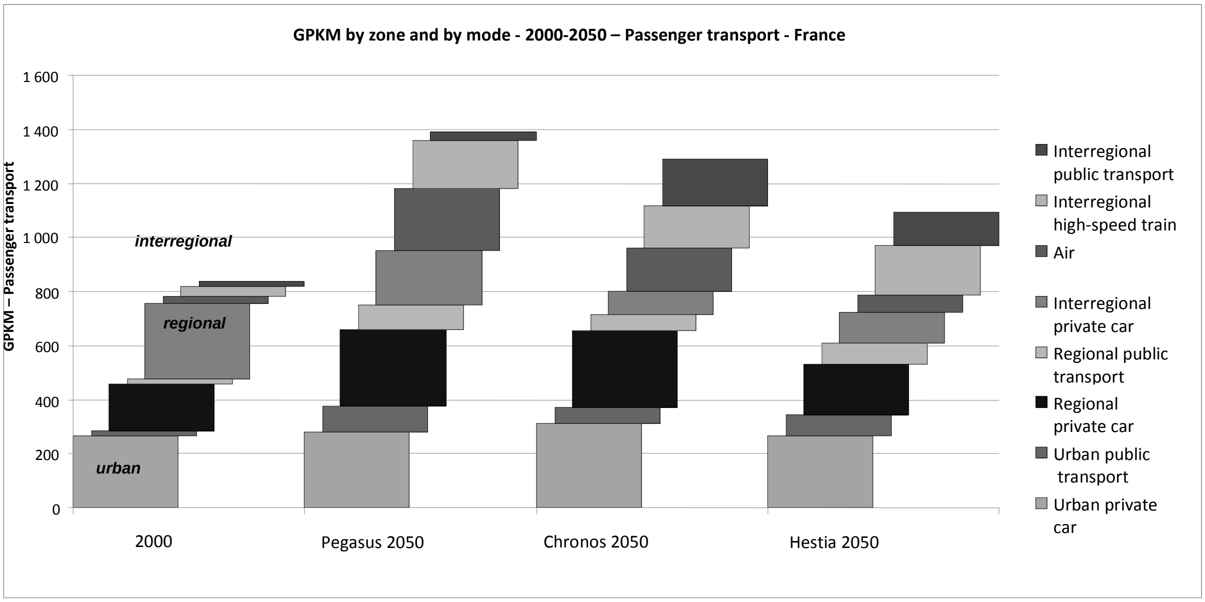
In Chronos, the underlying rationale for passenger travel is that a rise in the price of automobile use causes an increase in the use of public transport. The modal shift changes the household budget as the gains from the switch to a relatively less expensive mode are reinvested. Some of the gain will be reinvested in relocation (to get closer to public transport infrastructure) and some in fast long-distance transport services, especially air travel.

Thus, the system seeks to strike a balance by playing on the modal split in order to minimise cost. Chronos proposes an arbitrage between the need for speed (which increases because there is no saturation) and public limits on speed in the context of mitigation policies designed to encourage the use of cleaner transport modes and hence to improve the carbon footprint of transport as a whole. The public policy goal is therefore to achieve a large-scale modal shift, in favour of high-speed trains in particular, while keeping a more or less constant journey speed. In the French tradition of promoting high-speed trains, this is reflected in accelerated growth of rail travel while road speeds remain flat or even diminish. In this type of scenario, substantial investment is required in order to develop rail travel. Far-reaching changes to the organisation of the sector are also needed. So it comes as no surprise that in late 2007 the French president announced the construction of another 2 000 kilometres of high-speed railway lines.

The announcement was presented as an environmental response to the risks arising from an increase in air transport emissions. However, it is also a way of targeting speed gains on a particular mode, namely the high-speed train, and a particular type of travel, namely inter-urban journeys. The rise may be seen as offsetting the fact that the average speed of daily mobility journeys will fall, either because automobile mobility will be increasingly restricted or because the modal shift to local public transport will reduce the average journey speed. This scenario family therefore assumes the ongoing coupling of economic growth and mobility. As Figure 11 shows, total distances travelled increase almost as much as in the Pegasus trend-based scenario.

If economic growth and CO_2 emissions are decoupled (see Figure 12), it is mainly due to technological progress and a significant modal shift towards public transport. Nevertheless, the share attributed to air travel greatly changes the results. Although it is possible in the Chronos scenario family to approach Factor 4 for passengers, air transport must be severely restricted and replaced by high-speed rail. It is a rationale that we will find in an even more acute form in the Hestia scenario family.

Figure 11. **Passenger mobility 2000-2050: Pegasus, Chronos and Hestia scenarios**



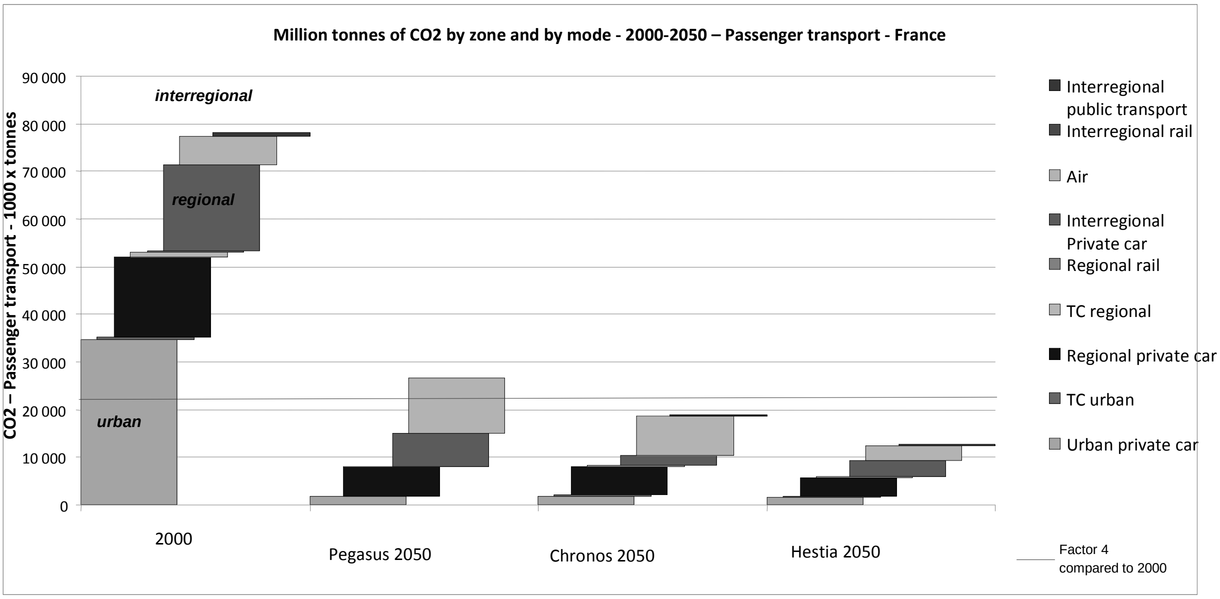
2.3.3 *Hestia: decoupling and mitigation. To what extent can air transport be restricted?*

A comparison of Figures 11 and 12 is instructive for more than one reason. We can see the key role played by restrictions on air transport in whether or not the objective of reducing CO_2 emissions by a factor of four is achieved. Air traffic increases sharply in the Pegasus scenario family and that has a knock-on effect on the sector's total emissions. In contrast, in the other two scenario families it is the drastic reduction in the relative share of air travel that makes it possible to achieve and even exceed the objective of a fourfold reduction in emissions, symbolised in Figure 12 by the horizontal line just above the 20 million tonnes of CO_2 mark.

The outlook in the Hestia scenario family is one of more restricted mobility. This is achieved not only through pricing and taxation but also through quantitative restrictions with, for example, the widespread introduction of tradable permits, reckoned to be more effective than a carbon tax. Facing what would amount to a complete break with the past, the system of individual preferences could have to change in favour of a reduction in distance travelled. Thus, an adaptation of the system through transport time (Chronos) would be replaced by a trend towards reduced distance (Hestia).

As we can see in Figure 11, the rationale is very similar to that of Chronos. The difference lies in the extent of the reduction in demand for transport by private car for regional and long-distance journeys. Once transport becomes too expensive, individuals express a preference for reduced distances because speed has become less accessible. If we look at Figures 6 and 8, this in fact brings us back to the logic of a reduction in purchasing power. The changing preference in favour of proximity does not come out of the blue but is the result of new constraints.

Figure 12. **Greenhouse gas emissions in 2050: Chronos and Hestia – Passengers**



Consequently, the increase in distance travelled is smaller in Hestia than in Chronos and Pegasus. In Hestia, proximity comes into play: the arbitrage concerns not only public policies that encourage the use of cleaner modes but also the geographical location of dwelling places and places for leisure activities and consumption. The main difference with the Chronos scenario therefore lies in the smaller rise in total distance travelled in relation to 2000. Passenger car traffic diminishes significantly but does not disappear altogether, in particular because air travel has been much more restricted than in the preceding scenario. But is such a decree of constraint possible? Backcasting shows us the path we ought to take, but as things stand at present there is little likelihood that we will do so, as the difficulty of reaching an international post-Kyoto consensus shows.

CONCLUSION

In 1825, when the British engineer, George Stephenson, put the first locomotive on rails (with a speed of 24 kph), the German philosopher, J.W. Goethe (1749-1832), expressed his concerns about the risks of the race for speed. Seeing it as diabolical, he coined the word "velociferic", suggesting that the quest for speed (velocity) had something in common with the devil (Lucifer). Has modern man assumed the guise of Mephistopheles? Nearly two centuries later, Milan Kundera picked up the same thread, quoting Goethe extensively in a novel (*Immortality*) in which he also insists on the death-dealing tendencies of speed, engaging in a regular critique of the road and the behaviour it induces in drivers.

How should we view these romantic strictures against the quest for speed after what we have just said about the past and future of mobility? At first sight, Goethe does not seem to have understood what was at stake. Higher speeds have profoundly changed standards of living and lifestyles, not always in a diabolical way. But Goethe and Kundera are probably right to suppose that there are limits to the quest for speed. There are physical and energy-related limits, as can be seen from the scrapping of supersonic commercial aircraft like Concorde. But there are also limits related to individual preference and the optimisation of activity schedules. That is why we are unlikely ever to attain the 791 kilometres per day envisaged by Schäfer in one of his hypotheses. However, that does not mean that personal mobility will level off in the years to come, especially where long-distance mobility is concerned. The accessibility gains offered by fast transport are such that demand for high-speed rail and air travel will remain strong. The extent of their relative growth will essentially depend on public policy, on the investment that public authorities are willing to finance or not, on the restrictions they might impose on the use of fossil fuels. Mitigation policies will have to be all the more proactive insofar as we are still a long way from reaching saturation point.

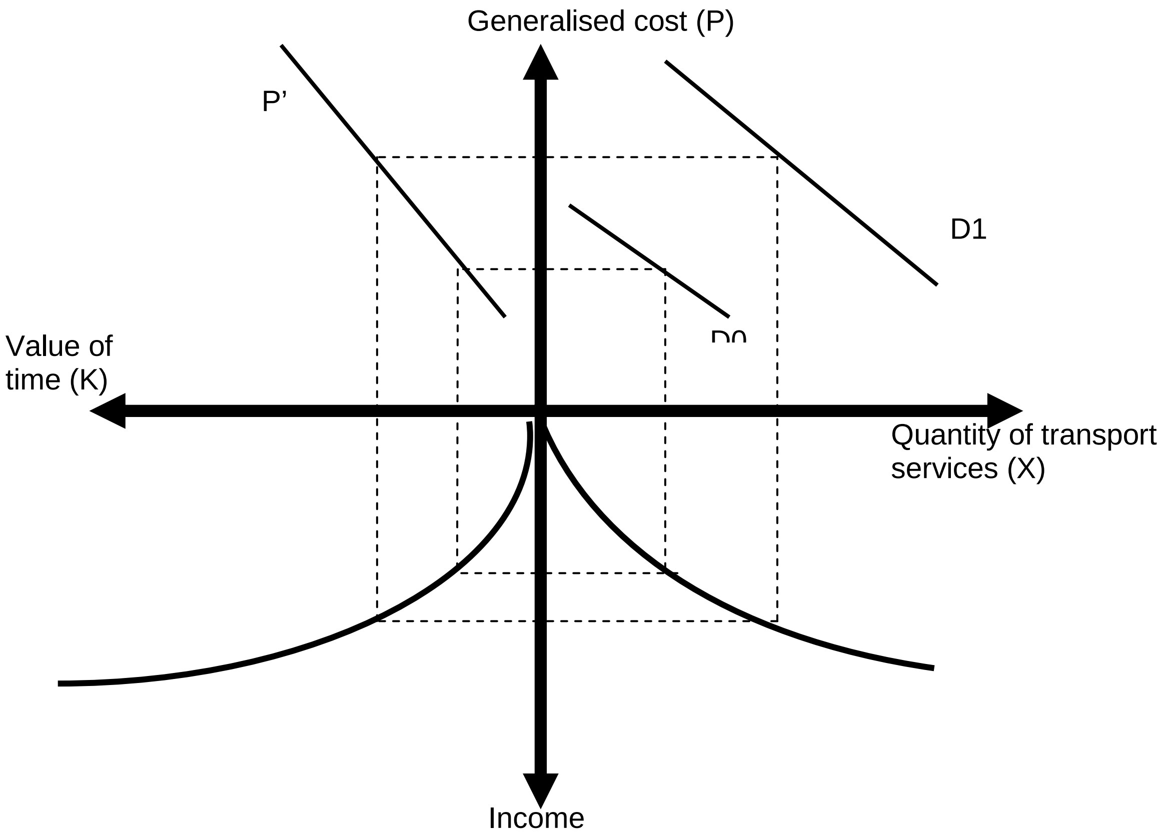
ANNEXES

Annex 1 : Spatio-temporal optimisation of recreational and business trips

Figures 6, 7 and 8 of the paper are derived from the thesis written by R. Gronau (1970) under the supervision of G. Becker, and from the thesis written by V. Bagard (2005) under the supervision of Y. Crozet.

R. Gronau's original model focused on long-distance transport demand by comparing bus and air transport when income increases. He examined the reasons for which we prefer a fast mode of transport and the logic on which they are based. Figures A and B summarize the stylised facts.

Figure A. **Stylised facts relating to demand for long-distance transport**
Initial situation where bus transport is the only option
(after Gronau, 1970)



The four key variables are income (Y), the value of time (K), the generalised cost of transport (P) and the quality of the transport services consumed (X). Between these four key variables in each quadrant lie the major stylised facts, whose logic will be easier to follow if we start with the income axis and then proceed in a clockwise direction around the diagram.

- The value of time increases more than proportionately where income K = f(Y) (bottom left-hand quadrant);
- The generalised cost P increases with the value of time for a given speed, in this case that of bus transport (top left-hand quadrant). The generalised cost (P’) also takes account of the cost of the ticket;
- Demand for transport is a decreasing function of the generalised cost (top right-hand quadrant);
- The quantity of transport consumed tends to rise with income because higher income levels provide access to new goods and services in new areas requiring greater mobility (bottom left-hand quadrant).

The main interest in Gronau’s reasoning lies in the emphasis it places on the two-fold impact of higher income. When individuals become wealthier, the increased value of time drives the generalised cost of transport upwards (top left-hand quadrant). However, higher income means access to a greater variety of consumer goods and services, which often require travel. Transport demand therefore rises from D0 to D1. The outcome is that if the bus is the only means of long-distance transport, the quantities consumed will rise but the increased cost in terms of time will act as a deterrent since the generalised cost rises rapidly if speeds remain low. This deterrent, which limits the quantity of transport consumed, is lessened if a significantly faster mode of transport, such as air transport, is available. In the latter case the quantity of transport services and, in particular, distances consumed can indeed rise sharply without increasing the amount of time spent travelling. A new balance is therefore struck, as shown in Figure B.

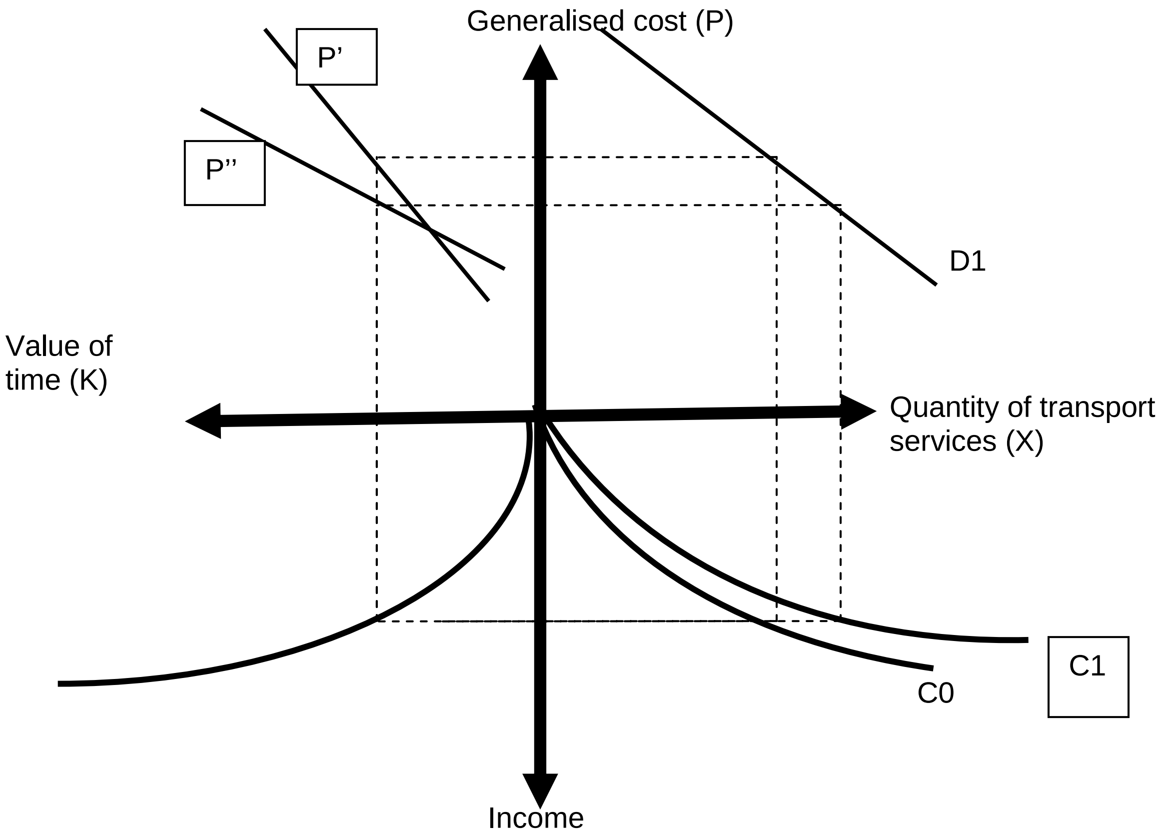
In this Figure, the new mode of transport, i.e. air transport, is responsible for two changes in typical relationships:

- In the top left-hand quadrant, the new line P” has a different gradient to line P’. This is due to the fact that the increased speed of air transport reduces the relative weight of time in the generalised cost. Since we have assumed that the cost of the ticket is not exorbitant, we obtain a relationship in which the generalised cost increases more slowly in relation to the value of time. To be more precise, when the value of time is low, the relative generalised cost of air transport is higher than that of the bus in that the only factor is the higher cost of the ticket. When the value of time increases, the generalised cost of air transport increases too, although at a slower rate given the shorter travel time.

- In the bottom right-hand quadrant, the impact of the lower generalised cost can be seen in the fact that for the same given income it is possible to consume a greater quantity of transport services. The relationship between the quantities X and income Y therefore changes from C0 to C1.

The new balance presented in Figure B takes account of these two changes. It can be seen that the outcome of this is a lower generalised cost of transport and an increased quantity of transport services for the same given income and therefore the same given value of time. Specialists will recognise an income effect, which has moved the position of the demand line, and a price effect related to the change in the structure of the generalised costs. The two effects combine to drive the quantities of transport services consumed upwards. Gronau finds an explanation here for the powerful development potential of air transport.

Figure B. **Stylised facts in demand for long-distance transport, from bus to air transport** (after Gronau, 1970)



On the basis of this diagram, V. Bagard's thesis sought to emphasize the consumption of time and space in relation to the consumption of recreational transport services. He therefore proposed different stylised facts given that the key variables had changed. While income and the value of time were retained (bottom and left-hand axes), the top and right-hand axes were changed:

- The top axis was used to represent the time budget allocated to the recreational activity, as well as its transport component. This total time budget is limited.
- The right-hand axis was used to represent the distances travelled every year.

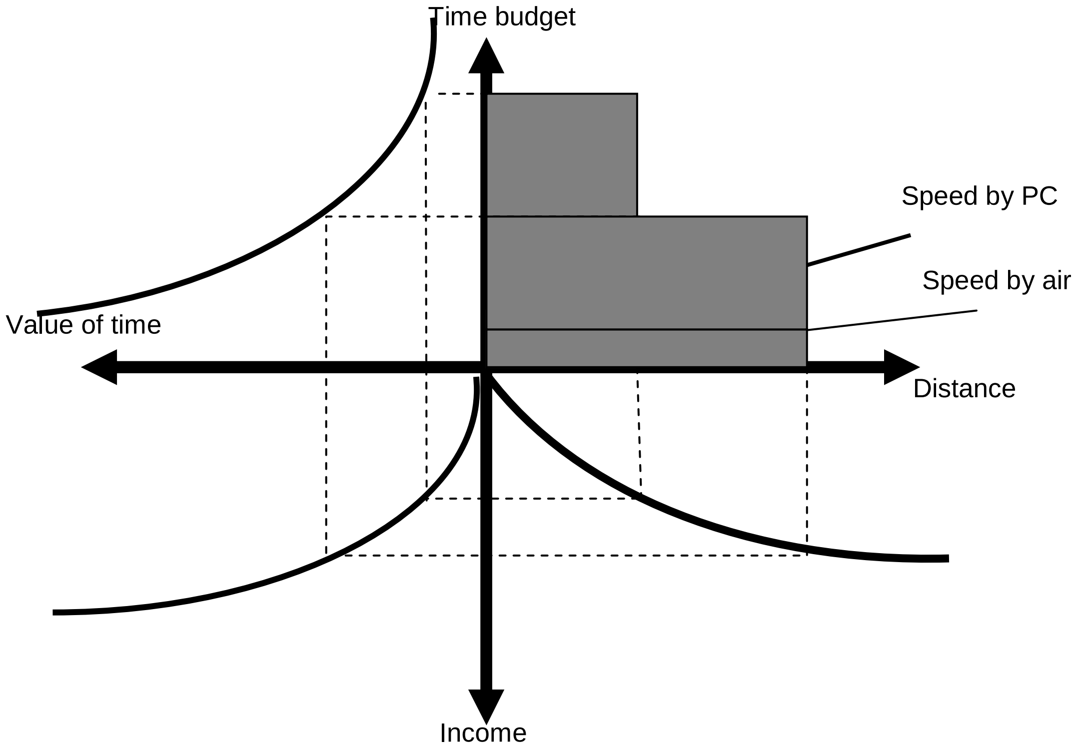
As Figure C shows, this produces the following relationships:

- Bottom left-hand quadrant: as with Gronau, the value of time rises commensurately with income;
- Top left-hand quadrant: the time budget allocated to a given activity decreases against income as a result of competition between activities;

– Top right-hand quadrant: the distance travelled depends on the average speed offered by the mode of transport (illustrated by gradient) and the value of the ratio between travel time and total recreational time;
– Bottom right-hand quadrant: distance increases with income because an increase in the latter provides access to increasingly faster modes of transport

.

Figure C. **Supply of speed and growth in distances for recreational travel**
(after V. Bagard 2005)



Figures 6 and 7 in the paper resume this line of approach but seek to stress the improbability of an exponential increase in distances in relation to income. Account does indeed have to be taken of the fact that speeds do not increase ad infinitum. For each trip, a given mode can only increase distance up to a certain level linked to the time budget available. Saturation mechanisms therefore do exist. This is what the bottom right-hand quadrant of Figure 7 shows in the paper. It can be seen that the increase in income is no longer accompanied by an exponential increase in distances for a given trip. The distance travelled increases in steps whenever a new and faster mode of transport emerges, which then itself levels off. This echoes the comment by A. Schäfer to the effect that the continued increase in distances travelled would require a sharp increase from 200 km/h to 600 km/h in door-to-door travel time for air transport, which would be highly unlikely! Saturation phenomena therefore do exist. Figure 8 considers another form of saturation which could combine with the previous form to slow growth in mobility. If the increase in the value of time were to level off too, like the increase in speeds, demand for trips over longer distances could indeed gradually become saturated. However, this threshold has not yet been reached, given that the share of the global population with access to fast modes of transport (high-speed train and air transport) still remains very low!

Annex 2:: The TILT model (Transport Issues in the Long Term)

The basis of the TILT approach lies on the proposition that a Speed/GDP elasticity implies different modal split possibilities. This is based on the growing importance of higher speeds as affluence and freight value grow (Schafer, A., Victor, D.G., 2000). Moreover, the modal split in transport is directly linked to the idea that modal speed, transport times, transport management and localizations determine modal shares. In this manner, transport modal saturation rhythms can be varied in the model - through public policies affecting localisations and the speed/GDP elasticity – which has proved to be fairly stable over time and very similar from one country to another (LET-ENERDATA, 2008).

Furthermore, in order to have a more precise view of the effects of public policies on each scenario, TILT has a microeconomic substructure that allows further analysis of demand determinants behind each scenario's modal split.

The TILT model has been designed to be a long-term equilibrium model by combining a macroeconomic and microeconomic structure in a backcasting approach that takes into account new motor technologies and facilitates sensitivity and impact assessments through five modules that work on three different geographical scales (urban, regional and interregional):

- A macroeconomic module based on a re-foundation of the energy-environment modelling structures in order to properly assess long-term modifications of demographics as well as social and cultural preferences in relation to transport needs.
- A microeconomic module based on a discrete choice and demand evolution that takes into account transport cost, infrastructure capacity and quality of service in order to asses changes in agents' transport choices.
- A vehicle fleet dynamic and technology evolution module that analyses technological impact based on market penetration probabilities and vehicles' survival rates for different motor technologies and different transport services (road, rail, sea, air, inland waterways).
- A public policy module that joins a sensitivity analysis (for policy categories) and multicriteria analysis (for specific public policies) in order to offer a detailed impact assessment of actions on CO_2 emissions.
- An impact assessment module based on an input-output equilibrium analysis that details impacts on employment and production by sector.

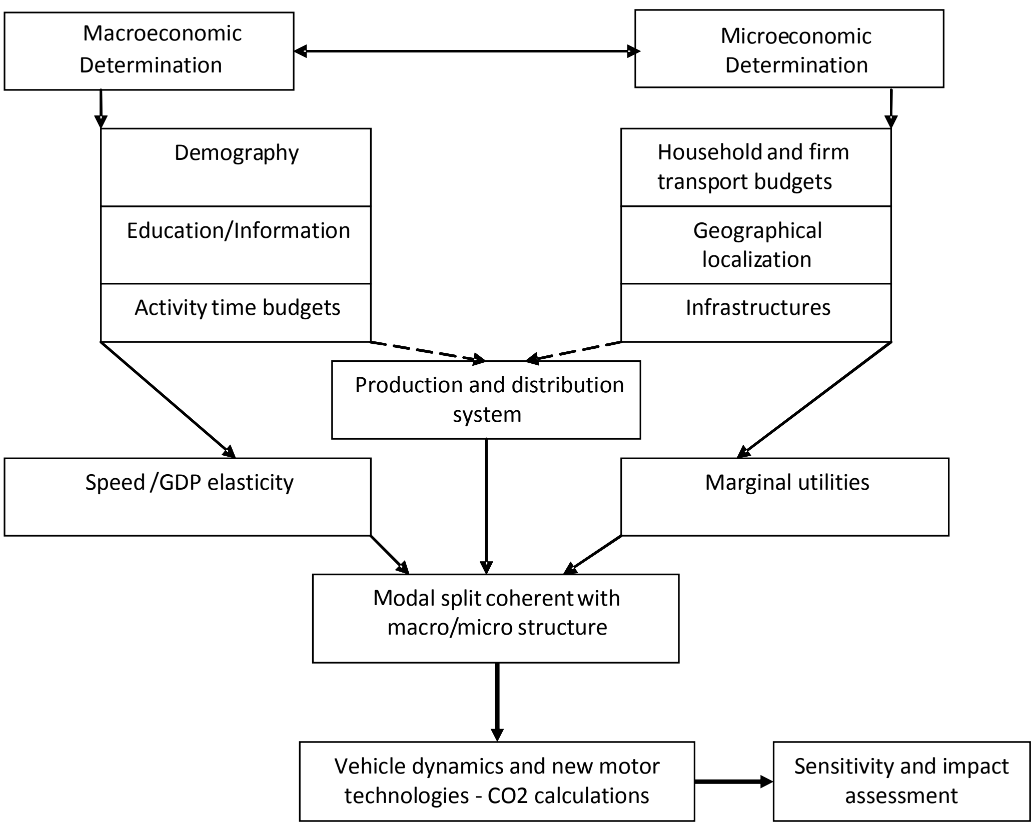
The TILT model structure enables the user to calculate energy consumption and pollutants emitted by transport activity (freight and passengers) on different geographical scales. The model has three important functions:

- Modelling passenger-kilometers and ton-kilometers coherent with a micro/macro equilibrium structure according to motor technology used for journeys and area of service.
- Modelling the vehicle park according to: age; motor technology; and year of production (for freight and passengers).
- Modelling and assessing public policy impacts on CO_2 emissions, infrastructure investment needs as well as overall impact on the economy.

By joining these three functions and the different TILT modules in a micro/macro equilibrium structure, it is possible to build scenarios that:

- Quantify the consequences of transport on the environment whilst detailing the systems' structure according to behavior and organizational changes, technology used, vehicle park dynamics, nature of a journey and vehicle age.
- Give a precise view of traffic by motor technology, gas consumption and emission levels for each type of transport according to service distances, type of vehicle and transport cost.
- Build policy pathways based that have different impacts in each scenario configuration and on the economy.

TILT Model Structure



These results coupled with the model's structure make TILT a powerful tool for building and exploring scenarios. The utility of the TILT model lays not only in its capacity to be flexible concerning political transport measures, changes in demography, behavioral differences as well as changes in transport structure and cost but also in its capacity to integrate new technologies' influence according to their year of entrance on the market and their ability to penetrate it. Furthermore, on the basis of its modelling structure, TILT is able to deliver a clear assessment of public policy sensitivity and infrastructure needs.

NOTES

1. On the potential of air transport and Maglev, see the papers presented at this Symposium respectively by D. Gillen and by K. Yamaguchi and K. Yamasaki.

2. The writer himself travels about 100 000 km a year, half of it by high-speed train and a quarter by air, representing nearly 275 km a day for an average transport time budget of about two hours a day.

3. This would explain the growing mess in young people's rooms and, increasingly, in the dwellings of young households.

4. This figure takes up and amends an analysis put forward by R. Gronau (1970) which took account of the generalised cost of transport (see Annex 2). As we want to emphasize the key issue of scarce time, we prefer to insist on the average length of stay and average travelling time.

BIBLIOGRAPHY

Ausubel, J.H., C. Marchetti and P.S. Meyer (1998), Toward green mobility: the evolution of transport, *European Review*, Vol. 6, No. 2, pp.137-156.

Bagard, V. (2005), Spatio-temporal Optimisation of Tourism Practices. Doctoral thesis in economics, supervised by Y. Crozet, 322 pages. www.univ-lyon2.fr

Banister, D., D. Stead, P. Steen, J. Akerman, K. Dreborg, P. Nijkamp and R. Schleicher-Tappeser (2000), *European transport policy and sustainable mobility*, Spon Press, 255 pp.

Banister, D., J. Pucher, M. Lee-Gosselin (2005), Making sustainable transport politically and publicly acceptable: Lessons from the EU, USA and Canada, in: Rietveld, P. and R. Stough (eds.), *Institutions and Sustainable Transport: Regulatory Reform in Advanced Economies*, Cheltenham, England: Edward Elgar Publishing, 2007, pp. 17-50. (http://www.itls.usyd.edu.au/)

Banister, D. and R. Hickman (2005), Towards a 60% Reduction in UK Transport Carbon Dioxide Emissions: A scenario building backcasting approach, VIBAT Study, www.ucl.ac.uk

Becker, G. (1965), Time and Household Production: a theory of the allocation of time, *Economic Journal*, 75, September, pp. 493-517.

Château, B., Y. Crozet, V. Bagard and H. Lopez-Ruiz (2008), Comment satisfaire les objectifs internationaux de la France en termes d’émissions de gaz à effet de serre et de pollution transfrontières ? Programme de recherche consacré à la construction de scénarios de mobilité durable. Rapport final. PREDIT, Paris, www.let.fr .

Clement, K. (1995), *Backcasting as a Tool in Competitive Analysis*, University of Waterloo, ISBM Report 24.

Commission of the European Communities (1992), *A Community strategy for “sustainable mobility”*, Green Paper, European Community Publications.

Commission of the European Communities (2001), *European transport policy for 2010: time to decide*, White Paper, European Community Publications, 136 pp.

Crozet, Y. (2005), Time and Passenger Transport, 127th ECMT Round Table, *Time and Transport*, OECD/ECMT, Paris, pp. 27-69.

Crozet, Y. (2007), Strategic Issues for the Future Funding and Operation of Urban Public Transport Systems, in: *Infrastructure to 2030*, Vol. 2, *Mapping Policy for Electricity, Water and Transport*, OECD, pp. 413-462.

Dumazedier, J. (1962), *Vers une civilisation du loisir ?* Paris, PUF. (In English, *Towards a Society of Leisure*? London, Macmillan, New York Free Press, 1967).

EEA, European Environment Agency (2009), Transport at a crossroads, TERM 2008: indicators tracking transport and environment in the European Union, EEA Report No. 3/2009, 52 pp.

European Conference of Ministers of Transport (ECMT) (1993), *Transport Growth in Question*, 12th International Symposium on Theory and Practice in Transport Economics, OECD Publications, 700 pp.

ECMT (1997), *Which Changes for Transport in the Next Century?*, 14th International Symposium on Theory and Practice in Transport Economics, OECD Publications, 509 pp.

ECMT (2000), *Key Issues for Transport Beyond 2000*, 15th International Symposium on Theory and Practice in Transport Economics, OECD/ECMT.

ECMT (2001), A baseline scenario for transport in Europe, OECD/ECMT, 54 pp.

Georgescu-Roegen, N. (1979), La décroissance : entropie, écologie, économie.

Gronau, R. (1970), *The Value of Time in Passenger Transportation: The Demand for Air Travel*, National Bureau of Economic Research, Occasional Paper N°109, Columbia University Press, New York and London, 74 pp.

Gronau, R. and D. Hamermesh (2001), *The Demand for Variety: a Household Production Perspective*, National Bureau of Economic Research, Working Paper No. 8509.

Gruebler, A. (1990), The Rise and Fall of Infrastructures: Dynamics of Evolution and Technological Change in Transport, Heidelberg: Physica.

Institute For Transport Studies, University of Leeds (2000), *Separating the Intensity of Transport from Economic Growth*, Report on the Workshop, University “La Sapienza”, Rome.

Kesselring, S. (2008), The mobile risk society, in: Canzler, Weert, Kaufmann, Vincent, Kesselring, Sven (eds.), *Tracing Mobilities*. Aldershot, Burlington: Ashgate, pp. 77-102.

Kesselring, S. and G. Vogl (2008), Networks, Scapes and Flows - Mobility Pioneers between First and Second Modernity, in: Canzler *et al.* (eds.): *Tracing Mobilities*, *op. cit*., pp. 163-180.

Linder, S. (1970), *The Harried Class of Leisure*, Columbia University Press, New York and London.

Mokhtarian, P.L. and C. Chen (2004), TTB or not TTB, That is the Question: A Review and Analysis of the Empirical Literature on Travel Time (and Money) Budgets, *Transportation Research A*, Vol. 38(9-10), pp. 643-675.

Niveau, M.A. and Y. Crozet (2000), *Histoire des faits économiques contemporains*, Paris, PUF, 847 pp.

Potier, F. (1998), *Les évolutions de la mobilité liée aux loisirs*, ECMT, Round Table No. 111, OECD, pp. 97-132.

Sansot, P. (2000), *Du bon usage de la lenteur*, Payot & Rivages.

Schäfer, A., J. Heywood, H. Jacoby and I. Waitz (2009), *Transportation in a Climate-Constrained World*, MIT Press, 329 pp.

Veblen, T. (1899), *Theory of the Leisure Class*, Penguin edition, 1994.

Viard, J. (2003), *Le sacre du temps libre*, Editions de l'Aube, 212 pp.

Viard, J. (2006), *Eloge de la mobilité*, Editions de l'Aube, 252 pp.

Zahavi, Y. and A. Talvitie (1980), Regularities in Travel Time and Money, *Transportation Research Record*, 750, pp. 13-19.

Zahavi, Y. (1979), *The "UMOT" Project.* Report for the US Department of Transportation and the Ministry of Transport of the Federal Republic of Germany.

INTERNATIONAL AIR PASSENGER TRANSPORT IN THE FUTURE

David GILLEN

Sauder School of Business
University of British Columbia
Vancouver BC
Canada

SUMMARY

*I am indebted to Stephen Perkins, David Starkie, Michael Tretheway and Jean Paul Rodrigue for comments and helpful discussion in developing this paper. I have also benefited greatly from the excellent research assistance of Adam Robertson and Chen Wei: I thank them.

1. INTRODUCTION

World stock markets fell further in mid-June 2009, when the World Bank and International Monetary Fund (IMF) both announced that the recovery from the current economic malaise would be longer rather than shorter. The World Bank stated that the world economy would contract 2.9%, compared with a previous forecast of a 1.7% decline. The Bank appears to be more pessimistic than the International Monetary Fund. The IMF is forecasting a global contraction of only 1.3% this year and growth of 2.4% in 2010. Furthermore, the World Bank cut its forecast for the US this year, calling for a 3% drop in the world's largest economy, after predicting a 2.4% contraction in March. Japan's Gross Domestic Product (GDP) is predicted to shrink by 6.8%, more than the previous prediction of a 5.3% decline. The Euro area's economy may shrink 4.5%, compared with the previous estimate of a 2.7% contraction. Global trade may drop by 9.7%, compared with a March forecast of a 6.1% decline.

In September 2009, Mr Bernanke told a Federal Reserve Board meeting that "the recession was technically over". He hastened to add that the recovery will be long and slow. This has been confirmed by IMF analysis that output per capita takes three years to recover after a banking crisis, and that seven years afterwards output is 10% lower than if the banking crisis had not occurred. Output is lower, trade is lower and trade and international air travel go hand-in-hand.

The forecasts and seemingly dire warnings of these leading financial and economic institutions that the world economies will take some time before starting on the road to recovery, is a triple blow to the world's international airlines. First, international aviation is driven in large part by GDP growth, and the nature and extent of the economic slowdown has led to substantial reductions in passenger traffic[1]. Secondly, airlines are by their nature cash-flow businesses and with fewer passengers now and in the future there is less cash, and this situation over a longer period threatens the survival of a number of carriers[2]. They have to be creative to survive: British Airways (BA) was asking employees to give some wage-free time, Air Canada simply asking for a USD 610 million bailout and most, if not all, carriers are significantly reducing capacity. Thirdly, international airlines have been shifting their business model as the low-cost carriers moved to capture a larger share of the domestic markets; legacy carriers started a few years ago to focus relatively more on long-haul, particularly high-yield traffic, both point-to-point and connecting, and this is the very traffic that is most affected by the current world economic crisis[3].

The objective of this paper is relatively straightforward, suggesting "what international air passenger travel will look like in five, ten or fifteen years, and why?" This requires answering two questions: what will be the principal determinants of the growth in international air travel and what impact will each of these drivers have on the growth rate? An imbedded question is: does history have anything to teach us or are there new forces at work? Canvassing the current aviation trade press finds two schools of thought. One takes the position that this a deep recession but a recession nonetheless and once world economies start recovering air traffic will go back to the typical growth of 4-5% annually. A second school is less sanguine, taking the position that it will not be business as usual when economies stop sinking and move to recovery. Any economic recovery is going to involve fundamental changes in institutions, rethinking polices regarding government participation in economies and changes in economic leadership in the world. There is also the hydra of protectionism, most prominent now in the US but certainly being practiced elsewhere, and the question of what will happen to foreign ownership restrictions that prior to 2009 were being seen as hurting rather than helping world airlines. All of this will change the momentum for international aviation.

The Organisation for Economic Co-operation and Development (OECD) in a recent paper (see OECD, 2009) has examined the economic downturn and the implications for the future development of GDP. This “development” refers to the magnitude and makeup of GDP. They distinguish three scenarios on how the economic crisis will affect global growth patterns. First, the crisis is an “accident” due to the breakdown in the financial system and once it is repaired it will be “business as usual”. Second, they refer to “retrenchment” describing a scenario of fundamentally changed global trade patterns; changes due to both an unsustainable system that was built on artificial financial foundations and due to policy responses. The “accident” and “retrenchment” scenarios are at each end of the “what will the world look like” spectrum. Somewhere in the middle lies an “adjustment” scenario which is characterised by a weaker outlook for global GDP growth, adjustments in global trade imbalances and weakened financial leverage. International air passenger travel would have different levels of growth and patterns of distribution; networks would change and with it carriers economic fortunes.

To understand where international air passenger travel may be heading in the medium to long term there are three sets of forces that should be investigated. First, what are the factors which have driven the growth in air travel in the past and what will those forces look like in the future? An examination of numerous air travel forecasting models indicates the key drivers as GDP and income growth. Closely linked to these factors are trade growth and foreign direct investment. There have been policy changes including the increasing liberalization of international aviation agreements, the changing business models of carriers, the expansion of alliances and the growth in long haul aircraft fleets. Given these were so important in the past will they be important in the future and what will they look like? If one believes in a model that an economic recovery will produce a set of world economies which will look much the same as what we saw in 2007-2008 then knowing the expected values and influences of old variables is what is important.

A second set of factors to consider arises from a possible change in world economies. What if the economies of 2010 and 2015 are not going to be the same as what we observed in 2007-2008? There may be new economic leaders, some or even many economies will undergo structural change and trade patterns of the past may be vastly different in the future. For example, there seems to be a consensus that the US economy will not see the levels of consumption it experienced in the post 2000 decade; savings will be higher in the US and spending may be rising in China. A new macroeconomic environment of particular importance will be the emerging role of the BRICs - Brazil, Russia, India and China - who, if they take over economic leadership, will alter international aviation networks considerably.

The third set of influences to be considered in assessing the future of international passenger air travel are those things - events, policies and economic and political environment which are new. What new forces will be at work in the future that will have an impact on international air travel? Certainly environmental issues will be a key factor, and a number of studies have investigated how emission trading schemes or carbon taxes would affect air travel particularly leisure travel. These studies have also investigated how such taxes or trading schemes may impact the structure of the networks and perhaps the industry itself. Other new forces will be technology such as improved engine fuel economy, biofuels, improved air traffic control (ATC) in the European Union (EU) and elsewhere such as free flight and integration under Eurocontrol, levying of airport and country specific taxes (e.g. United Kingdom and France), industry consolidation and the influence all of these would have on fares and service, and network reach and design.

The paper is organised in four main sections. Section two examines the travel demand forecasts of the past, what variables they relied on and what these variables are forecast to be going forward. In section three, we consider what a structural economic change might do to the future of air travel and section four examines how the "new" forces would impact air passenger travel. A summary and assessment for the future of air passenger travel is contained in section 5.

2. FORECASTING AIR PASSENGER TRAVEL DEMAND

A number of organisations, airframe manufacturers and agencies have provided forecasts of how they see aviation growing in the future. These forecasts by Airbus, Boeing and ICAO (International Civil Aviation Organisation) to name a few are summarised in Table 1; only values for international air passenger growth are included. All the values are fairly close with ICAO being seemingly more optimistic. These values are presumably reflecting some adjustments for the current economic crisis. Interestingly the Revenue Passenger Kilometres (RPK)/GDP growth ratio is approximately 1.6 for both Boeing and Airbus, which is what it has been over the past decade or so. This would seem to imply the airframe manufacturers are among those who take the view that, on balance, world economies will emerge from the recession in the same structural condition as before; a business-as-usual view or, as the OECD has named it, "the current crisis is an accident". In their work the OECD points out that even in the case in which globalisation continues there is substantial forgone economic growth; the effects of the slump are large with expected returns to previous growth rates not being realised for up to five years. If there is a shift from a globalisation regime, such as retrenchment, this implies a whole regime change with significant long-run consequences.

A particularly important insight from the OECD work is that even if a globalisation regime is retained, the pattern of globalisation must change since pre-crisis levels and patterns were not sustainable. Thus a stable, moderate and realigned globalisation regime may emerge but all of the forecasts reported in Table 1 do not reflect moderation or realignment.

Table 1. **Forecast growth in international air traffic 2008-2027 by various organisations**

Organization	EU-North America	Asia Pacific-Europe	Asia Pacific-North America
Boeing	4.7	5.7	5.6
Airbus	4.8	5.9	5.8
ICAO	4.5	5.8	6.0
Average 1990-2007	3.6	6.2	3.4

Sources: Boeing Current Market Outlook 2008-2027, Airbus, 2007 (2007-2026)
ICAO (2007) Outlook for Air Transport to 2025.

In Table 2, reproduced from Boeing's Current Market Outlook 2009, the expected growth in RPK between various regions is presented. It seems quite surprising that traffic growth between Latin

America and Asia Pacific and Africa will be so bullish. This reflects the expected growth in GDP in these regions; see Table 3. GDP growth has traditionally always been a significant driver in traffic growth and it appears there is a view that it will continue to do so – old drivers will be influential in the future. If one looks at the ratio of RPK to GDP across these sets of countries it varies from a low of 1.3 between Latin America and Africa to a high of 2.2 between Asia Pacific and Latin America; will these be the primary nodes of economic activity?

Table 2. **Growth in International Air Traffic Boeing 2009-2028**

	Africa	Latin America	Middle East	Europe	North America
Asia Pacific	9.2	9.1	6.3	5.5	4.9
North America	7.4	4.7	6.9	4.6	
Europe	5.4	4.3	5.5		
Middle East	6.1	-			
Latin America	5.5	-			

Source: Boeing Current Market Outlook 2009.

Table 3. **Assumed GDP Growth Rates for Boeing Air Traffic Forecasts**

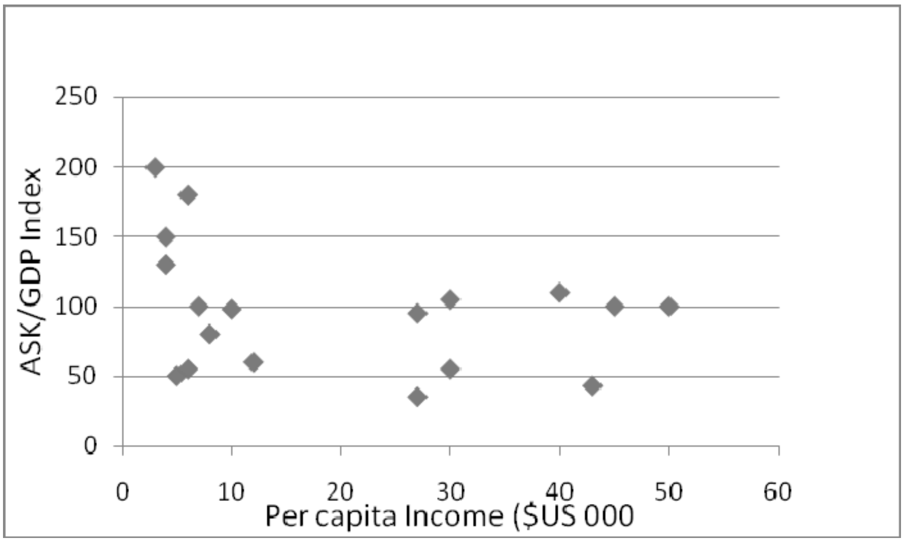
Region	**GDP growth**
Asia Pacific	4.4
North America	2.4
Europe	1.9
Middle East	3.8
Latin America	3.8
Russia and Central Asia	3.7
Africa	4.9
World	3.1

Source: Boeing Current Market Outlook, 2009.

Upon closer examination, it is clear the strength of the relationship between international passenger traffic growth and GDP *per se* has generally been overestimated due to a failure to account for changes in other strategic variables such as prices and network development and Open Skies air service agreements (see below). The measure of passenger growth with growth of GDP could be 1.5 or more[4]. However the reality is that while higher income countries generate more trips than lower income countries, air travel does not grow increasingly with wealth. Specifically, the air travel share of

GDP is independent of income. As Figure 1 shows, there is no clear relationship between the growth in passenger travel and the growth in income. This lends credibility to the elasticity of 1 value; air travel in general is not a luxury good, as people get richer they do travel more but they do not spend an increasing proportion of their income on air travel.

Figure 1. **Air travel share as a percentage of GDP**



Source: Swan (2009).

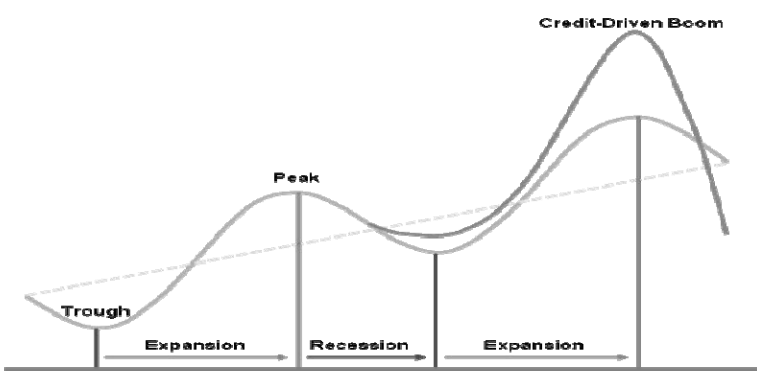
What are the other factors which have been important in the past? *First*, changes in trade regulations, trade liberalisation has led to what is termed globalisation. Firms take advantage of countries and regions comparative advantage, investing in other countries and increasing the amount of both merchandise trade and trade in services with the creation of international supply chains[5]. *Second*, changes in regulations, in this case international aviation air service agreements (bilaterals) with the result that fares come down, service expands and potentially there could be new firm entry. This improvement in service quality stimulates demand but the extent of the stimulus will depend upon the degree from which, and to which, markets liberalise. Piermartini and Rousova (2008) examined the impact of liberalising air transport services on air passenger flows in a sample of 184 countries. They find robust evidence of a positive and significant relationship between the volumes of traffic and the degree of liberalisation of the aviation market. An increase in the degree of liberalisation from the 25th percentile to the 75th percentile increases traffic volumes between countries linked by a direct air service by approximately *30%*. In particular, the removal of restrictions on the determination of prices and capacity and the possibility for airlines other than the flag carrier of the foreign country to operate a service are found to be the most traffic-enhancing provisions of air service agreements. The results are robust to the use of different measures of the degree of liberalisation as well as the use of different estimation techniques.

Gillen (2009) examined the case for Canada and estimated that the elasticity of *international* air passenger growth with respect to GDP was 0.45 (a 1% increase in GDP led to a 0.45 % increase in passengers), the elasticity with respect to 5th Freedoms was 0.15 (introducing 5th freedoms in a bilateral led to a .15% increase in numbers of passengers) and if an Open Skies agreement was inked, the elasticity was 0.66. Swan (2008) argues that the Open Skies effect happens only once (shifting the growth function) and has estimated that such events stimulate passenger growth over the long term by approximately 2% on average. However, it may be that such agreements can have direct (shift) and indirect effects as carriers adjust their networks and market structure changes. There can also be continuous effects if multiple trading blocks are liberalising sequentially[6]. However, there can be large differences depending on which markets are being considered. In the case of an Asia China Open Skies this would add 10% to passenger growth[7]. Korea would experience an estimated 6% boost from Open Skies, while Europe will see relatively small gains because of previous liberalisation; the changes are marginal (Swan, 2009).

In a recent study, Oum *et al*. (2009) make an important point that the liberalisation of air service agreements leads to expansion in markets but it also leads to more efficient continental and international networks which further stimulates traffic growth. The indirect efficiency effect would reinforce the direct effect of liberalisation on opening markets. The degree to which this would occur depends on the extent of liberalisation and the way it is done.

In the short to medium term, what changes would drive air traffic growth? Certainly the cycling of GDP around the long term trend is a key factor. This has been fairly regular in the past but over the last few decades the various asset and credit bubbles have increased the amplitude of the swings and the swings take longer to return to the trend. Figure 2 provides a stylised illustration of what appears to be happening currently. Traffic growth moves above and below the trend due to changes in the structure of economies as well as trade. Markets can change at different speeds.

Figure 2. **Trends in GDP growth and swings about the trend**

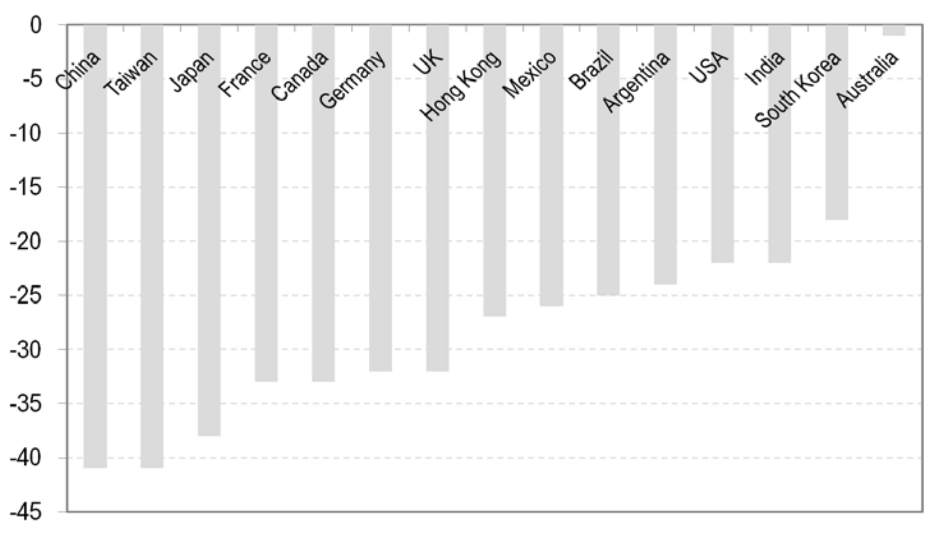


Source: Notteboom and Rodrigue (2009).

Notteboom and Rodrigue (2009) illustrate the sequence of three different market bubbles - high-tech, housing and trade. Each bubble accelerates the demand for international air travel and may increase the rate of growth. For example, high-technology industries and the finance sector, tend to be aviation intensive so a rapid growth in this sector leads to even more rapid growth in air travel than would be expected on average with growth in GDP. What is interesting about the three bubbles is each successive one encompassed a larger and larger population. The tech bubble involved relatively few people since only certain segments participated in this sector. It did certainly have a non-proportional impact on international air travel as assembly and manufacturing spread to Southeast Asia. The housing bubble, a consequence of Federal interest rates and financial policies in the US encompassed an entire nation and had consequences across many countries but principally in the US where it originated. The trade bubble was global and was driven in part, perhaps large part, by the housing bubble and the use of re-mortgages to increase consumption and purchase housing as well as a wide range of consumer goods in the US. Trade and the development of international supply chains drove an increase in international air travel.

The increasing amplitude in swings about the trend has resulted in higher costs for carriers. On the upswing, available capacity is expanded in increasing amounts and on the downswing this capacity drives fares lower and airline profits decline. The costs of adjustment increase. A second consequence is on consumer confidence which moves in short bursts generally lagging the GDP cycle but they move together. As the amplitude of the cycles about the trend increases it may be consumer confidence will take a longer time to re-establish itself and once it does a more conservative atmosphere may prevail[8]. There are the vagaries of war, flu viruses (SARS, Swine) and political disruption. These work through the cycle but again can be more troublesome as the cycle changes. For example, trade improves productivity, which has a positive impact on growth. If the bubbles reduce trade, the growth in GDP may slow more than proportionately due to loss of productivity.

Figure 3. **Year-over-year growth in total exports** (February 2009)



Source: Notteboom and Rodrigue (2009).

In the longer term, the growth in GDP and the growth in trade which exceeds GDP growth has driven international air passenger growth. The trend has been consistently upward and tied to growth in GDP but this growth is currently zero or negative in many cases. The growth rates of exports of many countries are also negative, as illustrated in Figure 3 for selected countries[9]. International air travel is following its traditional relationship with GDP and is also declining at double digits in some markets.

The five fundamental traditional drivers of long term international air passenger growth are GDP growth, political disruption, cost changes (e.g. fuel costs), service quality changes and trade growth. Political disruption would include terrorism, regime frictions such as with Iran and North Korea but also protectionism. While protectionism reduces trade growth (discussed below) it also appears in the form of reductions in foreign direct investment. Foreign ownership of "strategic assets" such as ports, energy and airlines are either up for review or simply prohibited. Such constraints increase capital costs and reduce trade in the long term. Political disruption and friction also increase costs in the form of security and regulation. These costs make shippers and service providers worse off and lessen trade and air travel. Cost changes particularly fuel costs is a long term threat. In the past growth in real fuel costs was zero or negative. In the future this will not be the case as the real cost of energy will go up and environmental taxes will become a permanent fixture. In the past cost reductions provided a 0.7% stimulus to passenger growth (Swan, 2009). It is unlikely this will continue and even advances in engine and fuel technology will not fully offset costs of raw materials inputs and taxes.

Quality changes occurred in the network over the last two to three decades. International networks reorganised with gateway hubs and airline alliances. This increased accessibility and stimulated traffic growth. A significant quality change was the growth in new markets; old markets did not simply get bigger but there were more routes opened and frequencies grew. Both of these outcomes stimulated traffic growth by one or more per cent. In the future the network will not be improving due to higher costs, hence bigger aircraft and less frequency; frequencies were a significant stimulus to traffic growth in the past. As trade growth slows frequencies decline, fewer routes are added (some abandonments may occur) and underserved cities continue to be underserved. All of this adds up to a negative net effect on past forecast traffic growth.

The slowing of trade growth over the longer term will also reduce the previous growth forecasts. As important will be the restructuring of trade as merchandise trade falls and trade in services grows somewhat. In the past trade growth was double that of GDP growth and added one to two% to forecast air traffic growth. In the short term with recession and trade reductions traffic growth will also be negative. In the longer term increased protectionism, a failure to reduce tariffs and increased costs from security and regulatory barriers will mean zero stimulus from the trend in the future.

The net impact of all of these factors could be traffic growth at 80% of what it was in the past; markets forecast to grow or actually growing at 4% will grow at 3.2%. This, as Swan (2009) contends, could occur with slowing trade growth, slower GDP growth, higher costs from fuel and taxes and a slowdown in route development, this in the business as usual model.

2.1. Empirical evidence on factors influencing international passenger traffic

In a number of papers there has been an attempt to assess the extent to which air travel is to be affected by current economic conditions. Oum *et al*. (2008), for example, estimate a model in which they include GDP growth, fuel prices and some dummy variables to reflect events such as SARS, 9/11 and Asian financial crises. They use aggregate data from 1980 to 2008 to examine how these factors listed affected total air travel – domestic plus international. They find the elasticity of air travel with respect to GDP is 1.58 but argue this value is inflated because it captures influences which were not included in the model such as increase services and new routes, the changes in air fares which would have been very important for domestic air traffic.

The model estimated in this paper uses data from 1996-2008 to look at international traffic only between eight regions; Africa, Asia, Europe, Middle East, Latin America, North America, South America and Southwest Pacific region. The dependent variable is revenue passenger kilometres. The explanatory variables include GDP growth, foreign direct investment into the region, total trade in merchandise and services, price of jet fuel, dummy variables to capture the influences of events such as SARS and 9/11 and a connectivity variable. The connectivity information was contributed by IATA who construct the index using information on flight frequency, seat per flight, number of destinations and a weighting factor which is designed to measure the importance of the airport. The "connectivity index is designed to measure how well a country, or region, is connected to the international air network. It is a measure of the number and economic importance of destinations served, the frequency of service to each destination and the number of onward connections available from each destination. Connectivity increases as the range of destinations and/or frequency of service increases. The index also reveals how connectivity changes over time. This index provides a measure of service improvements, route extensions and increased frequency. The results are reported in Table 4.

Table 4. **Panel fixed effects model**

Panel Fixed Effects Model		
8 cross section		
12 years; 1996-2007		
Variable (in Logs**)**	**Coefficient**	**T-Statistic**
Constant	-0.2849	-1.34
GDP	0.0652	2.10
Trade	0.8382	3.34
Connectivity	0.2201	2.37
Fuel Price	-0.2785	-3.34
Foreign Direct Investment	0.1306	2.28
Time	0.0884	1.80
9/11 Dummy	-0.1144	1.26
Adjusted R-sq	0.96	
Log Likelihood	168.96	

The results differ considerably from the model of Oum *et al*. (2008) but this model was estimated on only international air passengers whereas their model was estimated on total world air traffic. The model was composed of a panel data set with eight cross-sections (regions) and twelve years for each region. The variables are in logs so the coefficients can be interpreted as elasticities. Note the GDP elasticity is quite low, a mere 0.06, which is sensible in that the amount of international travel will be influenced, but only in a small way, by domestic growth. Also having trade, foreign investment and connectivity in the equation takes a good deal away from the magnitude of the coefficient. If one estimates essentially the same model as Oum *et al*. (2008), the estimated elasticity is only 0.31, considerably less than 1.58 of the Oum *et al*. model. What really matters for international travel is the amount of trade in merchandise and services; the elasticity is 0.83. Thus a drop in trade of 10% leads to a drop in international air travel of 8.3%. The next most important variable is connectivity, in which an increase in connectivity of 1% leads to a 0.2% increase in international air traffic.

Over the most recent three years in the data the connectivity index has risen on average by 8% across the world; thus boosting traffic growth by 1.6% on average. As connectivity declines through route abandonment, industry consolidation and capacity reduction, one can expect traffic to shrink accordingly.

The increase in jet fuel prices has a sizeable impact on international air traffic; the elasticity is -0.3, so a 10% increase in fuel prices leads to a 3% decrease in traffic. Estimates show that the elasticity of fuel prices with respect to increases in world oil prices is about 0.26 for auto fuel; because of differences in taxes, this elasticity in aviation would be higher, at 0.4 (see Gillen *et al.*, 2006).

Another important factor not previously considered is the magnitude of foreign direct investment (FDI) inbound; that is, foreign investment from outside the region. This is a rough measure of the degree of globalisation and as more investment takes place air traffic increases. The elasticity of international air traffic with respect to FDI is 0.13. A time-trend variable was inserted to pick up temporal trend effects and it shows a positive gradual increase in international air traffic.

What do these estimates indicate regarding future international air traffic growth? T Table 2 and 3 provide forecasts of interregional air traffic and growth in GDP, accordingly. This model indicates it is not GDP growth we should be looking at but rather trade in goods and services, changes in connectivity and changes in foreign direct investment. As well, fuel price increases and the application of fuel surcharges can have an impact. It is unlikely that fuel prices will reach the levels they did in summer 2008, but oil is trending upward over the longer term. The Energy Research Institute forecasts fairly steady prices for jet fuel in the next year. The IMF, however, forecasts a decline in GDP growth by 1.4% and an increase in 2010 of 2.5%. The IMF also forecasts FDI will fall by nearly 30% to 2010, and trade in goods and services will decline by 11% in 2009 and increase in 2010 by only 0.6%. These numbers suggest that international air traffic will fall in the near term and be weak in the foreseeable future.

3. INDUSTRIAL EVOLUTION OR REVOLUTION

What is unknown is what type of economies will emerge as the current economic crisis plays out; what will be the new macroeconomic and trade world? At present there appears to be both an industrial revolution and a carbon revolution. Together they could well reshape economies and trade into a set of multi-location global centres. The relative power of the US economy will decline with its old infrastructure and old factories and reversion to protectionism. The industrial revolution at the turn of the 18th century sprang from new technologies of transportation and communication and energy. The geography of trade and economic development was much influenced by coal and the geography of coal. This revolution took 100 years. If there will be another industrial revolution based on new technologies, environmental and energy efficiency will be central to competitiveness. Investments will need to be made in “soft infrastructure” of governance and reducing the friction of politics. How important will the comparative advantage be in driving trade? If economies in the BRIC countries create a set of multi-nodal economies where no one country really dominates, how will they trade, what do they trade and how does this drive air passenger travel?

There are two schools of thought on the evolution-revolution outcome. Some take the view that what we observe is a “blip” and those economies will return to normal. This might be regarded, as stated earlier, as the business-as-usual model. The OECD (2009) characterisation is that the current situation was an “accident” in financial markets and, once fixed, economies would return to their 2007 growth paths. The other school argues that a fundamental paradigm shift is taking place and what will emerge is a new macroeconomics and new trade flows. The extent of the change could vary from "retrenchment", in which trade flows and centres of production are radically altered, to a moderate "adjustment", which would see not a move away from globalisation but certainly a tempering of trade and economic growth. There are a number of factors that have come together to generate such an outcome. First, there is most likely an end of asset inflation and debt-derived consumption (at least temporarily). This will drive a re-equilibrium of trade flows, as well as standards of living to some degree. The "normal" of the last few years particularly in the US, which drove so much of what was taking place in globalisation and trade flows, was essentially a macroeconomic deception. Personal and government debt may, perhaps will, drive lower levels of consumption and discretionary mobility per capita. Second, energy prices are going to remain high and trend upwards; some analysts argue that oil may be at USD 100 by the end of 2010[10].

A third factor is the aging of the population, an issue that is often neglected. It could well be linked with two macroeconomic forces; an aging population is less mobile – an issue not considered by forecasting models – and, second, the retiring population is very likely to be much less wealthy than expected, as their two major assets, a house and a retirement plan, will be worth much less. For many, the expectations behind the quality of life in retirement are going to be readjusted substantially downward. In other cases, pension plans may go into default, waiting for government bailouts. This is most likely with defined benefit plans, and Europe is particularly vulnerable in this regard.

A startling statistic is that the US has 4.5% of the world's population and spent USD 10 trillion annually, while India and China have 40% of the world's population and spent USD 2 trillion annually. There is a USD 8 trillion gap and with the US faltering is it reasonable to believe the BRIC countries will make up the difference? The business-as-usual school must believe this to be the case.

In the US consumer spending rose from 67% of GDP in 1980 to 75% in 2007, while the household savings rate fell from 10% of income in 1980 to near zero in 2007. Household indebtedness went from 67% of income to 132%. These shifts in spending drove trade and resulted in the US having a current account deficit of nearly 6% of GDP by 2006. The financial crisis in 2008 led to the collapse of consumption, with more than USD 13 trillion in consumer wealth lost. However, the collapse has endured due to a shift to greater savings, up to 5% of income now. Some of this spending has been replaced by the fiscal stimulus in the US as well as elsewhere. But this offset is minor since it serves to stabilize, not replace, consumer spending and, secondly, much of the spending is national with requirements for domestically produced goods and services mandated. This rise in protectionism will exacerbate the lack of global growth whereas the US consumer had been its heart and soul for the past several years[11].

Many take the position that the new US model will be based more on export growth and less on consumption. This is in contrast to what fueled the boom previously and it is unlikely that growth in exports will compensate for the consumer sector. There are requirements that resources be shifted into production in tradable products and productivity to improve, particularly in export sectors. The externality of the US-led economic crisis on the rest of the world, notably Europe, will work against such export-led recovery. The resulting sluggish economy will see protectionism as a necessary condition to succeed. We see this increased protectionism in the US across many sectors and the financial and economic crisis has led to a shift left in the political spectrum, with a future of big government, parochialism and greater focus on domestic markets and less on developing trade.

The underlying causes of the economic recession and the current state of world economies leads some to a conclusion that the new macro economy is not going to look like the old macro economy (OECD, 2009). Centres of production will differ and trade patterns will change. Greater domestic production and consumption, particularly in the US, will lead to greater regional and domestic air travel with a relative decrease in international air travel. If airlines fail, consolidate and reduce service and capacity to survive, all of this will mean even less international air travel.

Growth in GDP and trade will continue to be important drivers of RPK from traditional factors. Industry behaviour in pricing, route development and network restructuring will have important but second-order effects. Unknown is what world economies will look like in the future and when economies will show positive growth? This is what the world looks like now (IMF, 2009); see Table 5. Looking at the table, what is notable is the value of 2009 figures in comparison to values in other economic downturns. The 2009 values are orders of magnitude larger for every indicator.

Table 5

Global Recessions: Selected Indicators of Economic Activity

(Percent change, unless otherwise indicated)

Variable	1975	1982	1991	Projected 2009	Average (1975, 1982, 1991)
Output					
Per capita output (PPP[1] weighted)	–0.13	–0.89	–0.18	–2.50	–0.40
Per capita output (market weighted)	–0.33	–1.08	–1.45	–3.68	–0.95
Other macroeconomic indicators					
Industrial production	–1.60	–4.33	–0.09	–6.23	–2.01
Total trade	–1.87	–0.69	4.01	–11.75	0.48
Capital flows[2]	0.56	–0.76	–2.07	–6.18	–0.76
Oil consumption	–0.90	–2.87	0.01	–1.50	–1.25
Unemployment[3]	1.19	1.61	0.72	2.56	1.18
Components of output					
Per capita consumption	0.41	–0.18	0.62	–1.11	0.28
Per capita investment	–2.04	–4.72	–0.15	–8.74	–2.30

Note: The 1991 recession lasted until 1993, using market weights; all other recessions lasted one year.

[1]PPP = purchasing power parity.

[2]Refers to change in the two-year rolling window average of the ratio of inflows plus outflows to GDP.

[3]Refers to percentage point change in the rate of unemployment.

Source: IMF 2009.

These figures underlie what we are seeing in terms of double-digit decreases in international air travel, except for the Middle East; it appears no major international airport or gateway has been spared, with even Dubai showing zero growth for the first period of 2009.

Below (see Figure 4) is an indication of what the IMF thinks may happen in various regions of the world and when such changes might occur (in the figures, years are on the horizontal axis and growth in GDP on the vertical axis). The key indices to watch are the ratios of government deficit to GDP, private savings to GDP (in March 2009, the US recorded the highest savings rate since 1946) and current account to GDP.

Both the benign and downside scenarios illustrated in Figure 4 are bad news for international aviation as GDP growth will be slow to recover in all regions and in both the US and Asia will turn down again in a few years. Industry restructuring is inevitable but the final outcome is highly dependent on regulations, domestic competition law enforcement and foreign ownership restrictions. Increased concentration may lead to higher fares and reduced route development, both of which will diminish traffic growth.

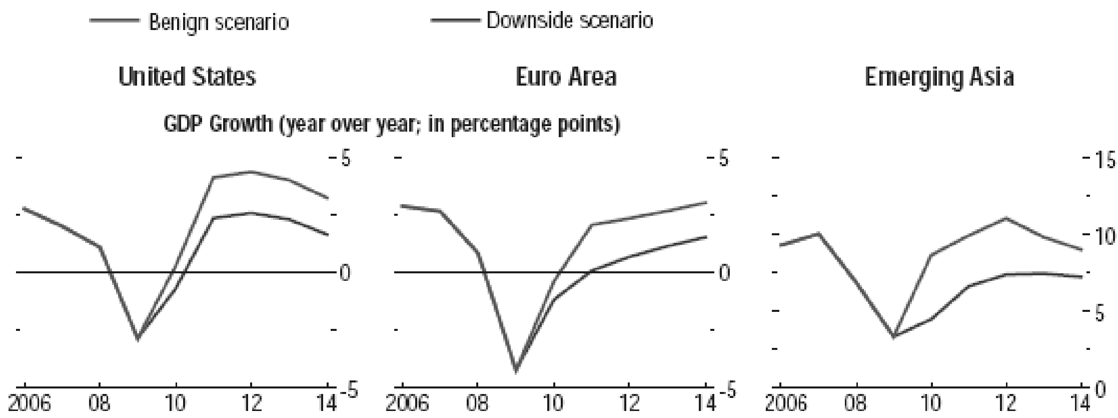
4. NEW FORCES INFLUENCING PASSENGER AIR TRAVEL

The most influential new factors which will affect air traffic growth will be environmental taxes, regulations and emissions trading schemes. As governments link their carbon strategy with their economic and energy strategy there will be direct impacts on the aviation sector, as well as indirect effects as economies and industry in general restructure, but also as the airline industry restructures. The introduction of carbon taxes or emissions trading will lead to changes in market structure which will affect fares, service and carrier profits. An issue of considerable debate is how much of the tax or cost of emissions permits will be passed through to consumers. If the emissions cost becomes a profits tax, this will result in some failures and potential consolidation. If it is fully passed through there will be some reduction in demand.

Gillen and Forsyth (2008) analyse outcomes under differing market structures assuming single price equilibrium and linear demands. Under competitive market conditions, the cost pass-through is 100%, with fares rising by the amount of the tax or permit cost allowance in the long run; in the short run, fares rise by less and airlines incur losses. Long-run equilibrium output is lower and fares are higher; in competitive markets traffic loss in the future may be from 0.7 to 1%.

On monopoly routes the pass-through is 50%, with profits falling and exit taking place from marginal routes. The impact on the long-term passenger forecast for these routes is minor. One would expect in the absence of government restrictions that such markets would evolve to be more competitive and therefore have a higher pass-through. In oligopoly, which would characterise the majority of international routes, if they were liberalised there would be incomplete pass-through, lower profits and less output. Growth is constrained. If the international routes have restrictive bilaterals, this is equivalent to the outcome with a slot-constrained airport. Fares are set in the market on the basis of bilateral restrictions: therefore any increase in costs due to allowances or carbon taxes will be a profit tax and fares will not change; any increases in costs are paid out of rents arising from bilateral restrictions. If rents are monopoly rents there is a 50% pass-through but if rents are scarcity rents there is no pass-through. In oligopoly in the long run there will be lower growth, lower growth than without the charge, firms will adjust to higher costs with exit from some routes. The route exit effect will reinforce the higher cost effect in reducing future air passenger growth, perhaps as much as 1%.

Figure 4



Another view with respect to cost pass through is provided by two studies, commissioned by the UK Department of Environment, Food and Rural Affairs in 2007, and 2008. These studies examined the impact of the EU Emissions Trading Scheme (ETS) on ticket prices and airline profits, respectively. What is notable in these studies is the claim there may be more than 100% pass-through under some circumstances. Specifically, the report claims cost pass-through could run between 80% and 150% and the key determinants are the level and elasticity of demand, the objective function of the airline (profit, sales or market share), the market structure and the types of rival (business model) participating in the market. In the majority of cases the pass-through is at or near 100%, a finding consistent with the literature. In cases where the pass-through exceeds 100% the demand elasticity is assumed to be constant and inelastic[12]. A greater than 100% pass-through is not possible on the average of fares, provided firms are profit maximising to begin with and, even in the case of differential pricing (yield management), no one price would be increased greater than the amount of the emissions charge with profit maximising firms[13]. The study also found, correctly in the author's view, that the method of allocation of the emission permits would have no effect on the magnitude of the pass-through.

The second key issue is what amount is passed through; how much ticket prices will rise depending on the cost of the permits or the level of the carbon tax. Scheelhaase and Grimme (2007) report that short-haul LCC fares would rise by 2.6% while short-haul legacy carrier fares would rise by 1.15% based on an assumed value for emission permits of €15, €20 and €30; the reality is there will be a range of fare increases which correspond to a range of permit prices. Their long-haul calculations of fare increases were airline specific; 3.3% for Lufthansa and 3.5% for the Emirates. Trucost, in a 2004 study, calculated the following for expected price increases: see Table 6. Oxera (2003) calculated that, on average, with CO_2 at €50/tonne, fares would go up 3.08% and passenger demand would fall by 3.02%.

Table 6. **Impact of cost increase from EU ETS on fares and demand**

Airline	Price Rise (€)	Fall in Demand (%)
Air France	2.43	2.43
Alitalia	6.33	7.60
Austrian	3.05	3.66
BA	0.45	0.45
Easyjet	3.45	5.17
Finnair	2.33	3.50
Iberia	3.35	4.01
KLM	0.83	1.00
Lufthansa	1.71	1.71
Ryanair	3.33	4.99
SAS	0.72	0.86
Swiss	1.82	2.73
Virgin Atlantic	2.98	4.46

Source: Developed based on Trucost (2004).

Albers *et al*. (2009) examine whether the EU ETS will result in a change in network configuration. In their analysis, they estimate that with a 100% cost pass-through, fares would increase by from 1% to 3.8% (long-haul flights), with the result that demand would fall by up to 3% but in most cases it was approximately 2% of countries' long-haul travel, primarily tourism. In their work they estimated that the 25 richest countries (by GDP per capita) account for 51% of world GDP, 15% of world population, 45% of world tourism GDP, 69% of international passenger volume and 70% of total passenger volume. The GDP impact of a 10% fuel tax would range from 0.03% for the US, 0.1% for Australia and 0.12% for South Africa.

An important sector which has a considerable impact on international passenger aviation traffic is the global investment and financial sector. The banking crisis has resulted in numerous bank and investment house failures. Profits collapse in a financial crisis as credit becomes more expensive, which means as firms have less to invest, the economy slows. Centres of activity migrate and with them the centres of finance; the exodus of personnel from the financial sector in London is a good example of the consequences of such shifts[14]. The Global Financial Centres (GFC) Index released in September 2009 indicated the top ten global financial centres had not changed from 2007 but they all had lost in ratings. A change in ratings illustrates some new dynamics in play. The top global financial centres have not changed since last year (London, New York, Hong Kong, Singapore, Zurich, Frankfurt, Geneva, Chicago, Tokyo and Sydney), but all except Singapore have lower ratings from the previous year[15]. Also new centres have emerged in China, the Middle East and Africa. Osaka has dropped 33 points in ranking while Bahrain and Johannesburg have gained 59 and 48 points respectively. The GFC index provides some support for the notion of shifts in paths of international passengers. What is needed is information on shifts in direct foreign investments as well.

5. SUMMARY

This paper had the objective of trying to understand "what international air passenger travel will look like in five, ten or fifteen years and what were the underlying drivers." This required answering two questions; to identify what will be the more important determinants of international passenger travel in the future and, secondly, to translate the impact of these factors into expected changes in future passenger growth. Identifying the drivers was relatively successful in determining which are most relevant; and how large each of the effects would be on traffic growth was less successful.

Three groups of factors were identified; the "old" variables which have been identified as driving air traffic growth, the new variables which may result from industrial revolution rather than evolution, and the "new" forces such as those resulting from the carbon strategy being adopted in the EU and which will be followed elsewhere[16].

Among the established key factors is, of course GDP growth. Some believe that, even with a retained globalisation regime, growth recovery is five years away. However, the return will not be "business as usual" for two important reasons; first, protectionism is growing and not just in merchandise trade. Restrictions on financial intermediation will prevent pre-crisis types of economic interactions from returning (OECD, 2009). Second, the crisis was a consequence of global imbalances, which have since been moderated. Global restructuring means those countries which were large exporters (China and German) will have to adjust. Exporting overcapacity will be absorbed by domestic demand, reduced output or changes in exchange rates.

The new forces of change both contribute to and deter traffic growth. Carbon taxes and cap and trade systems will reduce growth but not to a significant degree unless the number of permits is reduced or the carbon tax is increased. In the short to medium term neither is likely to occur. New air traffic control governance in conjunction with new hard and soft technologies, such as free flight and EU integration under Eurocontrol, will have a positive impact on growth without necessarily having an offset from emissions increases.

Boeing, in its Economic Outlook (2009), forecasts economic growth of 3.1%, a forecast growth in passengers of 4.1% and a growth in revenue passenger-km of 4.9%; this implies a ratio of 1.6 of RPK to GDP. This scenario is based on what appears to be a model of industrial "evolution" – the economic order will repeat itself in the recovery – and is predicated on lower fares, point-to-point service and higher frequency[17]. Boeing's forecast of these optimistic growth rates is based on a trend of increasing growth in RPK.

The trend that is observed in traffic growth has been driven by growth in GDP (more countries getting richer) and increasing competition and liberalisation, which reduces average fares and expands service in terms of route development and frequencies. The point of diminishing returns may have begun to set in for OECD countries which have liberalised aviation markets to a degree with growth tapering to a trend GDP growth. However, if we do have an industrial revolution taking place, how new economic and carbon/energy strategies will affect international air traffic growth is difficult to establish. It is not just GDP but the composition of GDP, it is not simply air service agreement liberalisation but the type of liberalisation and what the starting point is, it is a shift from trade in merchandise to trade in services and a shift from globalisation to regionalism and regional trade pacts. Globalisation is heavily based upon liquid capital markets. It is made to happen by consumers and traders. Traders depend on cheap reliable transportation for people and merchandise. The current crisis has revealed that globalisation means integration and integration can be fragile, as has become clear. Going forward, it may be that a more risk-averse world wishes for less integration. Protectionism may exacerbate such a shift.

Swan (2009) has pointed out that expenditures on air travel are, on average, 1% of GDP in developed and developing countries. This is for all air travel, not just international air travel. Oum *et al*. (2009) develop a set of forecasts for both intra- and interregional travel. Their model is based on measured impacts of GDP, liberalisation and exogenous events (e.g. wars) on air travel growth in the past. Interestingly, they forecast that interregional air travel growth will generally exceed intra-regional growth; the implication being that past influences will continue into the future and it is just a matter of when a recovery starts to take place[18].

Notteboom and Rodrigue (2009) have examined what is happening in liner shipping. They make the point that the current set of circumstances has no contemporary “frame of reference”; international aviation, like shipping, is facing a global and persistent decline and, as they say, this can lead to unintended consequences. In their view, liner shipping will undergo a paradigm shift rather than a contemporary recovery. International aviation has come through boom and bust cycles and has weathered the vagaries of war, pandemic and financial crisis, but international aviation, like shipping, will more likely than not undergo a paradigm shift as well.

While fuel prices and changes to air service agreements will have an impact on international aviation, the most important impact will come from industry and economic reorganisation. The shifts in trade flows and the potential for a reduced pace or even decline in globalisation and a shift to regionalisation will affect trade flows, and hence international aviation. The persistence of the current economic malaise - some have suggested a four to five year horizon before growth will recover - will lead to a number of firms failing[19]. Consolidation will take place with some capacity reductions either directly or through alliances, where the alliance will manage the capacity. This will lead to higher fares and less route expansion. Both will result in a reduction in international air travel. There is the prospect of LCC entry into international markets but this is dependent on liberalisation of air service agreements continuing and on an expenditure elasticity of one. In the past, the US was a major force for liberalising international air markets. It is unclear whether this will continue; the US economy is weak and there is less to be obtained from more liberalised markets, and the US is moving to economic protectionism which also lowers the return from liberalised air service agreements.

Trade is not the cause of the current economic crisis but it may be one of its casualties. Trade increased with globalisation, which created international supply chains – complex international networks for the manufacture of goods; goods cross borders many times from inception to final consumption. A decrease in demand is amplified across all borders because of these supply chains. This decrease may also lead to increased protectionism. It is this combination of factors which may make international aviation, as we know it, also a casualty.

Pre-crisis growth and trade patterns were inflated by global imbalances and therefore expectations of future trade growth should be moderated. Global economic activity in the future may well be less trade-intensive; moderate growth and moderate trade. This moderation may be a consequence of protectionism or exchange rate adjustments[20]. In either case, international passenger traffic is likely to decrease but, more importantly, there will be a shift in paths from pre-crisis periods. How this will play out is an open question. The old GDP elasticities of RPK demand were based on established patterns of trade and non-sustainable growth rates, so extrapolating from pre-crisis information is likely to be misleading. As economies begin to recover, the consensus is that recovery will be slow. This may lead to industry restructuring as marginal carriers cannot continue with the losses. This restructuring, may well lead to reductions in competition, so that gains made from liberalisation of air service agreements will be tempered and international air travel will be further impacted.

NOTES

1. This is not to suggest that domestic aviation activity is not responsive to GDP growth; it certainly is, but "visiting friends and relatives" (VFR) and leisure traffic also constitute a large part of domestic travel.

2. Airlines use cash from future customers to finance current production. Most businesses receive payment when the product or service is delivered.

3. It is also the "front of the plane" traffic which paid the premium yields and accounted for a sizeable proportion of overall revenue.

4. See Oum *et al.* (2009).

5. Trade in services tends to be relatively aviation-intensive.

6. This is the case with the EU approaching countries adjacent to EU member states and negotiating open skies agreements. Middle-Eastern and Mediterranean countries are the first candidates.

7. This figure would include short-haul international between China and Taiwan, Japan and Korea.

8. The speculation is that US consumers will save more and spend less while Chinese consumers will do the opposite.

9. Numbers are based on a calculation of annualized GDP growth for 1st quarter 2009, based on 4th quarter 2008 data.

10. Peak oil will assert itself; it remains to be seen if this will be gradual or sudden.

11. In Japan in the 1990s, demand was suppressed for a long period after the bubble.

12. There are two issues to consider: first, the analysis did not consider whether the route was slot-constrained and, second, is it reasonable to think that sensitivity to price would remain unchanged with a greater than 100% pass through. Finally, an assumed constant inelastic demand implies that a monopoly market would not be in equilibrium when the emissions charge is imposed.

13. In cases in which there is a predicted greater than 100% pass-through, firms cannot be profit-maximizing in the first place and it is not clear what objective function would generate this result.

14. British Airways has suffered considerably with the drop in premium traffic, much of it generated from London's financial district.

15. The report states that a change in rating of 1-10 points is considered insignificant, between 10-30 is a signal of changing competitiveness and >30 signifies major change.

16. These include carbon taxes or cap and trade.

17. The Boeing forecast was the only one that is current. Airbus' last available forecast is from 2007.

18. Their models were estimated on all traffic rather than separately for intra- and inter-regional traffic separately.

19. Although governments in many jurisdictions seem intent on protecting favoured or flag carriers.

20. Hummels (2009) also argues that rising energy costs, which increase the cost of transportation, environmental initiatives and changing channels of trade in merchandise will underlie the shift to moderation.

BIBLIOGRAPHY

Albers, Sascha, Jan-André Buhne and Heiko Peters (2009), Will the EU-ETS Instigate Airline Network Reconfigurations?, *Journal of Air Transport Management*, Vol. 16, pp. 1-6.

Access Economics Pty Limited (2007), Analysis of Aviation Specific Climate Change Policies on Developing Countries Dependent Long Haul Travel (Report to Department of Industry, Tourism and Resources).

Alamdari, Fariba E. and Damian Brewer (2008), Taxation Policy for Aircraft Emissions, *Transport Policy*, Vol. 1, Issue 3, June 1994, pp. 149-159.

Bruckner, Jan and Anming Zhang (2009), Airline Emissions Charges: Effects on Airfares, Service Quality and Aircraft Design, CESifo Working Paper Series No. 2547.

CE Delft (2007b), The Impact of the Use of Different Benchmarking Methodologies on the Initial Allocation of Emission Trading Scheme Permits to Airlines; Final Report to DfT Aviation Environmental Division and the Environment Agency, July, CE Delft and Manchester Metropolitan University.

Department for the Environment, Food and Rural Affairs (2008), A Study to Estimate the Impact of Emissions Trading on Profits in Aviation, Report to UK Department of Environment, Food and Rural Affairs, January.

Department for the Environment, Food and Rural Affairs (2007), A Study to Estimate Ticket Price Changes for Aviation in the EU ETS, Report to UK Department for the Environment, Food and Rural Affairs and UK Department of Transport, November.

Department for the Environment, Food and Rural Affairs (2006), Including Aviation into the EU ETS: Impact on EU allowance prices (report to Department for the Environment, Food and Rural Affairs and Department of Transport, UK), February.

Forsyth, Peter (2009), Climate Change Policies and Their Impacts on Airline Fares, Profitability and Emissions, paper presented to the 13th Hamburg Aviation Conference, Hamburg, February.

Forsyth, Peter and David Gillen (2007), Climate Change Policy and the Impact on Airfares, presentation to GARS Workshop on Aviation and Climate Change, Cologne, November.

Gillen, David (2009), Canadian International Aviation: Policy and Challenges, address to Calgary Chamber of Commerce and Van Horne Institute, May.

Gillen, David, W. Morrison and C. Stewart (2007), Air Travel Demand Elasticities: Concepts, Issues and Measurement in: Darin Lee (ed.), *Advances in Airline Economics: The Economics of Airline Institutions, Operations and Marketing*, Elsevier, The Netherlands.

Gillen, D., Cheng and E. Lin (2006), Forecasts of World Oil Prices in the Short, Medium and Long Term, Centre for Transportation Studies, Sauder School of Business, University of British Columbia, Research Report.

Global Financial Centres Index (2009), City of London, (www.cityoflondon.gov.uk).

Hummels, David (2009), Globalization and Freight Transport Costs in Maritime Shipping and Aviation. Paper prepared for the International Transport Forum, May 26-29, Leipzig, Germany.

Hummels, David (2007), Transportation Costs and International Trade in the Second Era of Globalization, *Journal of Economic Perspectives*, Vol. 21, No. 3 (Summer), pp. 131-154.

International Air Transport Association (IATA), *World Air Transport Statistics, 2003-2009.*

International Monetary Fund (2009), *Global Financial Stability Report Responding to the Financial Crisis and Measuring Systemic Risks*, World Economic and Financial Surveys, (April) Washington DC.

International Monetary Fund (2009), *World Economic Outlook: Crisis and Recovery*, World Economic and Financial Surveys, (April) Washington DC.

Lu, Cherie (2009), The implications of environmental costs on air passenger demand for different airline business models, *Journal of Air Transport Management*, Vol. 15, pp. 158-165.

Mayor, Karen and Richard Tol (2007), The Impact of the UK Aviation Tax on Carbon Dioxide Emissions and Visitor Numbers, *Transport Policy*, Vol. 14, pp. 507-514.

McKnight, Paul (2009), Airline Economics (mimeo, University of Western Ontario).

Morrell, Peter (2007), An Evaluation of Possible EU Air Transport Emissions Trading Scheme Allocation Methods, *Energy Policy*, Vol. 35, pp. 5562-5570.

OECD (2009) Transport Outlook 2009 Globalisation, Crisis and transport, Discussion Paper No. 2009-12, Joint Transport Research Centre of the OECD and the International Transport Forum, May.

Oum, Tae, Xiaowen Fu and Anming Zhang (2009), Air Transport Liberalization and Its Impact on Airline Competition and Air Passenger Traffic, Final Report, OECD International Transport Forum, Leipzig Germany May 2009).

Notteboom, Theo and Jean-Paul Rodrigue (2009), Economic Cycles and the Organizational and Geographical Attributes of Global Value Chains, presentation to Maritime Transport in Value Chains Workshop, Montreal June.

Piermartini, Roberta and Linda Rousova (2008), Liberalization of Air Transport Services and Passenger Traffic, World Trade Organization, Staff Working Paper ERSD-2008-06.

Schroder, Andreas (2008), Incorporating Aviation into the EU Emissions Trading Scheme, presentation to GARS Workshop, July.

Scheelhaase, Janina D. and Wolfgang G. Grimme (2007), Emissions Trading for International Aviation - An Estimation of The Economic Impact on Selected European Airlines, *Journal of Air Transport Management*, Vol. 13, pp. 253-263.

Swan, William (2009), Reforecasting under Stress (presentation to International Forum on Shipping, Ports and Airports, Hong Kong Polytechnic University, May.

Swan, William (2008), Forecasting Asia Air Travel with Open Skies (mimeo Seabury Group, Seattle, USA).

Theme II:

Adapting the Intermodal Network to the Passenger Market: Long-term Planning and Assessment

WHEN TO INVEST IN HIGH-SPEED RAIL LINKS AND NETWORKS ?

Chris NASH

University of Leeds
Leeds
United Kingdom

SUMMARY

1. INTRODUCTION

Definitions of high-speed rail (HSR) differ, but a common one is rail systems which are designed for a maximum speed in excess of 250 kph (UIC, 2008). These speeds invariably involve the construction of new track, although trains used on them can also use existing tracks at reduced speeds.

A number of countries have upgraded existing track for higher speed, with tilting technology on routes with a lot of curves. However such trains do not normally run at speeds above 200 km p h. Their rationale is to upgrade services at relatively low cost in countries which have sufficient capacity to cope with increased divergence of speeds on routes shared with all forms of traffic. Most of the countries which adopted this strategy initially, such as Britain and Sweden, are now considering building HSR.

The only form of totally new technology that has come close to being implemented is maglev. However, no country yet uses such a system for inter city transport. It was proposed to introduce such a system between Hamburg and Berlin, but this project has been abandoned; it is still under discussion for the Tokyo-Nagoya route in Japan. The technology is capable of very high speeds, but apart from cost considerations, it has the inflexibility that the trains are not able to make use of a section of new infrastructure and then to transfer to existing tracks to finish their journey. The latter mode of operation is a feature of all new high-speed rail systems worldwide, even where – as in Japan and Spain – the new lines are built to a different track gauge from the existing lines (Spain uses bogies capable of adjustment to the different gauge, whilst Japan has undertaken installation of limited sections of multi gauge track). Maglev technology has its greatest chance where there is sufficient traffic to justify both a new self contained route and the existing one, and the most likely corridor to satisfy that requirement in the near future is the Tokaido corridor in Japan.

Thus the only high-speed inter city projects to have been completed to date use conventional rail technology with purpose built new lines for some but not all of the route network . That is therefore the focus of this paper.

In the next section we consider the motivation behind the introduction of HSR around the word. We then examine evidence on its impact on mode split. Following this consider the approach to appraisal of HSR followed by some actual examples. We then discuss a model that has been constructed to identify the key parameters that determine its social viability. After this we consider network effects and track access pricing before reaching our conclusions.

2. MOTIVATION FOR THE INTRODUCTION OF HIGH-SPEED RAIL

The first country in the world to build a dedicated line for new high-speed trains (originally at 210 kmph, so not satisfying the above criterion) was Japan. The background to this was that the original Tokaido line was narrow gauge (3 feet 6 inches) and unsuitable for high speeds. It was also at capacity. It was the twin desire for a big increase in capacity in one of the most densely used corridors in the world, and for a major improvement in journey time to be competitive with air that led to the approval of the construction of a new high-speed line at standard gauge. The Tokaido Shinkansen started running between Tokyo and Osaka on 1st October 1964 and was an immediate success, carrying 23 million passengers in its first year and leading to demands for its extension countrywide (Matsuda, in Whitelegg *et al.*, 1993). Wider economic considerations such as regional development and equality led to the development of Shinkansen investment on progressively less busy and profitable routes. When Japanese railways were reorganised as a set of separate regional commercial organisations in 1987, the high-speed infrastructure was placed in a separate holding company (the Shinkansen holding company) and the new operating companies were charged for its use on the basis of ability to pay, thus permitting cross subsidy between profitable and unprofitable routes. (Ishikawaka and Imashiro, 1998). Whilst this decision was later reversed and the Shinkansen sold to the operating companies in order that it should appear on their balance sheets, the principle of basing the charge on ability to pay rather than historic construction cost was maintained.

The success of the Japanese high-speed system, particularly in gaining market share from air, was undoubtedly a major factor inspiring European railways to follow the same path. The next in line was France, where intensive economic and technical research led to the proposal to build a new high-speed line from Paris to Lyons. Again the background was a shortage of capacity on the route in question plus the growing threat of competition from air (Beltran, in Whitelegg *et al.,* 1993). In 1981 the TGV Sud-Est between Paris and Lyon opened with speeds up to 270km/h. The name Sud-Est was itself designed to emphasise the network effects of this line, which as well as serving the Paris-Lyons market carried trains for a large number of destinations beyond Lyons. From this beginning plans were developed for a network of lines with the justification being largely in transport cost-benefit analysis terms although hopes were also raised for wide regional economic impacts (Polino, in Whitelegg *et al.* 1993). The idea that high-speed trains should be open to everyone, at reasonable fares (democratisation of speed) was an important part of the philosophy and helped the popularity of TGV with the general public. Subsequent developments have seen extensions to Marseille and Nice, the TGV Atlantique Paris-Bordeaux, Paris-Lille-London/Brussels and most recently Paris-Strasbourg.

The background to the introduction of high-speed rail in Germany was somewhat similar; a perceived shortage of capacity in the face of growing demand, accentuated by particular bottlenecks on north-south routes which had become more important following partition. Again the growing threat of air and car competition also led to a perceived need for high speed to satisfy the marketing requirement of “twice as fast as car; half as fast as plane” (Aberle, in Whitelegg *et al.*, 1993). However, the geography of Germany did not lend itself to development of a single key route, but rather of new sections of track where particular bottlenecks occurred. These were designed for both freight and passenger traffic, although their use by freight has been small. Although construction started in 1973, it was held up by environmental protests. Not until 1985 was a new design of high-speed train (the ICE) introduced. Gradually these trains were extended to cover the principal inter city routes throughout Germany, with long stretches of running on conventional track upgraded for 200 kmph. Thus the marketing of the ICE is very different from that of the French TGV; a lot of shorter journeys are made on it, reservations are not compulsory and load factors averaging 50% as opposed to the French 70% are tolerated.

The geography in Spain is more like that of France, with long distances between the major cities and even less intermediate population. Given the relatively low quality of the inherited infrastructure, Spanish Railways were rapidly losing market share to air and car. High speed was seen as a way of enabling rail to compete, as well as promoting regional economic development (Gomez-Mendosa, in Whitelegg *et al.* 1993). Whilst construction of the first line, Madrid-Seville, was hastened to serve the International Exhibition in Seville in 1992, construction of a whole network of lines was encouraged by Keynesian policies of relieving large scale unemployment by a major public works programme. The aim is to link Seville-Madrid-Barcelona to the French TGV system, and for that reason the network is being built to standard gauge even though other main lines on the Iberian peninsula are broad gauge.

Italy took its first steps towards construction of dedicated high-speed lines early with the Rome-Florence Direttissima, work on which started in 1966 and the first section of which opened in 1976 (Giuntini, in Whitelegg *et al.* 1993) but it was not until 1985 that a team was set up explicitly to study high-speed rail, leading ultimately to plans for a network of lines.

The early development of high-speed rail in Europe was entirely at the national level, using domestically produced technology (France, Germany and Italy each produced their own high-speed rolling stock using national manufacturers). However, the advantages of linking lines into a European inter-operable network were realised, and the concept emerged of a 15 000 km network of high-speed routes emerged, linking all the major cities of Europe (CER, 1989). The 1993 Treaty of Maastricht called for a network of Trans-European lines, linking the existing high-speed lines. Of major strategic importance were the new line between Brussels and Cologne, the extension of TGV Sud-Est to the Spanish border, the planned Alpine crossing between Lyon and Turin and links between the French and German networks (TGV Est). Recognition that such lines would benefit not just the countries in which they were built but the European Union more generally led to their designation as part of the Trans European Network, and a large share of the limited European funds made available for transport infrastructure has been directed towards them. Peripheral countries have also received substantial funding for high-speed rail from regional and cohesion funds, designed to reduce economic and social inequality within Europe.

By 2006, high-speed trains in Europe were carrying 84 billion passenger-km per annum, of which more than half was in France (UIC, 2008a). In the meantime, high-speed rail has been extended to more countries in Asia, including Korea, Taiwan and China.

3. IMPACT ON MODE SPLIT

This section will briefly consider impacts on rail market share. Detailed results on market shares are available for the early impact on mode split of the Paris-Lyon and Madrid-Seville lines. TGV Sud-Est between Paris and Lyon was opened in two stages between 1981 and 1983. The train journey time was first reduced by around 30%, after the opening of the Northern section, and the implied journey time elasticity was around -1.6. However, the time elasticity was around -1.1 for a journey time reduction of around 25% on the opening of the Southern section of the route. The cause of this lower elasticity was because the transfer from air had been largely completed in the first phase when rail was fast enough to provide effective competition. The Spanish AVE service introduced in April 1992 reduced rail journey times between Madrid and Seville from around 6½ hours to 2½ hours, making what was a very unattractive service into one which competes effectively with air.

Table 1 indicates the market shares of plane, train and road before and after the introduction of high-speed rail on these two routes. The impact on rail market share is very large, particularly in Spain where the improvement in rail journey time was larger. Much more traffic is extracted from air than road. It should be noted that the figures will have been influenced by a significant amount of newly generated traffic. Wilken (2000) reports that surveys of AVE passengers indicated that 15% of the additional rail traffic was newly generated, whilst according to Bonnafous (1987) no less than 49% of the additional traffic on Paris Lyons in the first four years was generated traffic. In other words, whilst there was indeed a substantial transfer from air, the reduction in road mode share was largely caused by the generation of rail traffic, rather than direct transfer.

Table 1. **Before and after high-speed market shares**

	TGV Sud-Est		**AVE Madrid-Seville**	
	Before	**After**	**Before**	**After**
Plane	31%	7%	40%	13%
Train	40%	72%	16%	51%
Car and Bus	29%	21%	44%	36%

Source: COST318 (1996).

More up-to-date figures are quoted by SDG (2006) and Campos and Gagnepain (2007) for the air-rail mode split, showing that where rail journey times are reduced below four hours, rail share of the rail-air market increases rapidly with further journey time reductions, and rail tends to have a market share of at least 60% and sometimes effectively drives air out of the market when rail journey times are below three hours. Future trends are found to depend on a wide variety of factors including the introduction of environmental charges on air transport and trends in air and rail costs.

It should be stressed that this evidence is from countries where for most people a city centre rail station is more convenient than an airport: where development is low density with weak city centres and poor public transport this may not be the case.

Kroes (2000) points out that the available evidence concerning modal shift relates to traffic that is not transferring at the airport to another plane. There is very little evidence on the transfer market. However, the increasing integration of rail with air with high-speed rail stations at airports such as Paris, Brussels, Frankfurt and Amsterdam offers the prospect of much greater rail penetration into this market, especially if ticketing and baggage handling is better integrated.

4. APPRAISAL OF HSR

The process of appraisal requires comparison of a base case with a series of options. It is necessary to be clear what the base case is and to ensure that a realistic range of options is examined. A base case that literally assumes a 'do-nothing' situation may be very unfavourable, particularly in the face of growing traffic, and thus exaggerate the case for undertaking a particular option; on the other hand the base case should not be padded out with unnecessary investments, as that may have the same effect. In general the base case should be a 'do minimum' and other likely investments should be examined as alternative 'do something' options. These alternatives should be compared on an incremental basis to see whether the additional cost of moving to a more expensive option is justified, and the phasing and timing of options should also be examined. The fact that a particular option is better than the base case is thus not in itself evidence that it is desirable.

In the case of high-speed rail, the base case should therefore include such investment as is necessary to keep the existing service running, and consideration should be given to how to deal with any exogenous growth in traffic. This might mean investing in additional rolling stock or revising fares structures and levels. More major changes should be considered as alternative do something options. These might include upgrading existing infrastructure, purchase of a fleet of new tilting trains or indeed construction of additional road or airport capacity. There will also be options regarding high-speed rail – how far to extend the new line; to which alternative points to run the new trains, what service frequency and pricing policy to adopt. It is essential to examine sufficient alternatives to be confident that the best alternative has been identified. The range of potential options makes appraisal of high-speed rail a difficult task.

It is also necessary to consider the timing of investment. High-speed rail might turn out to have the highest net present value, but if the demand for HSR and the other benefits from it are forecast to grow then it might still be better to postpone the investment.

HSR involves construction of new lines, stations, etc. and purchase of new rolling stock, and additional train operating costs and externalities (mainly noise, air pollution and global warming effects). The principal benefits from, HSR are:

- time savings;
- additional capacity;
- reduced externalities from other modes;
- generated traffic;
- wider economic benefits.

Time savings are generally split into business, commuter and leisure. A relatively high proportion of HSR traffic is likely to be travelling on business, although questions have been raised on whether the full business value of time should be applied in this case on two grounds:

- many long distance business trips start and end outside normal working hours;
- when travelling by train it is possible to work on the way (Hensher, 1977).

However, research has shown that firms are willing to pay something like the full business value of time even in these circumstances, presumably because of the benefits they perceive in shortening long working days and having staff less tired (Marks, Fowkes and Nash, 1986).

Additional capacity is obviously only of value if demand is exceeding the capacity of the existing route. But in those circumstances additional capacity may be of value not just in allowing for growth between the cities served by the high-speed line, but also, by relieving existing lines of traffic, for other types of service such as regional passenger or freight. Of course, this raises the further option of building new capacity not for high-speed passenger but for regional passenger or freight traffic. If new capacity is to be built anyway, then it is the incremental benefit of high speed versus the incremental cost that has to be considered, a comparison which is likely to make high speed look much more attractive than if the entire cost of new lines has to be justified on the basis of higher speeds. There is also clear evidence (Gibson *et al.* 2002) that running rail infrastructure less close to capacity benefits reliability; it may also lead to less overcrowding on trains. Both of these features are highly valued by rail travellers and especially business travellers (Wardman, 2001).

Typically as illustrated in the previous section a substantial proportion, but not all, of the new traffic attracted to rail will be diverted from other modes – mainly car and air. To the extent that infrastructure charging on these modes does not cover the marginal social cost of the traffic concerned there will be benefits from such diversion. It is frequently argued that high-speed rail has substantial environmental advantages since it diverts traffic from road and, particularly, air, where greenhouse gas emissions are much greater. On the other hand, as noted above, a substantial proportion of the traffic is typically newly generated or diverted from conventional rail, where given lower speeds, one might expect energy consumption to be lower. Of course high-speed rail is invariably electrically powered, which gives the possibility of using a carbon free source of energy, whereas inter urban road and air transport are currently tied to oil. Electrically powered trains are also free from local air pollution, except for small particulate matter from braking, at the point of use, although the visual intrusion and noise from a new high-speed line are often the subject of controversy.

One of the few studies to break down emissions in detail by type of train, as well as type of air and car transport is C E Delft (2003). They produce the following results:

Table 2. **Energy consumption by mode 2010**

	Intercity train	High-speed train	Air (500km)	Diesel car on motorway
Seating capacity	434	377	99	5
Load factor	44%	49%	70%	0.36
Primary energy (MJ per seat km)	0.22	0.53	1.8	0.34
(MJ per passenger km)	0.5	1.08 (0.76*)	2.57	0.94

*At 70% load factor.
Source: CE Delft (2003).

In other words, high-speed rail has a substantial advantage over air transport, is similar to car and very much worse than conventional rail. Recent unpublished work for Network Rail suggests that on a heavily used new high-speed line from London to Manchester, energy embodied in the infrastructure might add some 15% to these figures; obviously for a less well used line the increase could be substantially more. However. whilst the load factor given for high-speed rail of 0.49 may be typical of Germany, where high-speed trains spend a lot of their time running at conventional speeds on traditional track, and seat reservations are not compulsory, both the French TGV and Eurostar, with long non stop runs, compulsory seat reservations and sophisticated yield management systems, claim load factors similar to the 70% shown for air. A load factor of 70% reinforces the advantage over air and brings HSR below car, but it is still 50% higher than conventional rail. Given the sort of combination of mode switching and generation found above, the savings and costs tend to cancel out and the introduction of high-speed rail cannot lead to a substantial energy saving; where there is little diversion from air, it will undoubtedly lead to an increase. So the claim of HSR to reduce greenhouse gases must rest on a non fossil fuel source of electricity generation, as is currently the case in some countries (e.g. France, with a high share of nuclear and Switzerland with a lot of hydro) but not others such as Britain.

Diverting traffic from road does not simply affect greenhouse gases, but also reduces road noise, accidents, local air pollution and congestion. The following table (Table 3) presents the unit values for these costs for a petroleum car, as estimated for a major European corridor in the European research project GRACE (GRACE, 2005 Deliverable 7). Whilst the off peak costs are quite similar between routes, the peak costs are much larger and more variable, being dominated by congestion costs which vary greatly from route to route.

Table 3. **Marginal social cost and prices for long distance car transport**

Milano-Chiasso
Interurban petrol GRACE car petrol EV

	Peak	**Off-Peak**	**Night**
Noise	0.007	0.011	0.035
Congestion	0.147	0.002	0.001
Accident	0.015	0.015	0.015
Air pollution	0.001	0.001	0.001
Climate change	0.005	0.005	0.005
W&T	0.016	0.016	0.016
	0.191	**0.050**	**0.073**

Chiasso-Basilea
Interurban petrol GRACE car petrol EV

	Peak	**Off-Peak**	**Night**
Noise	0.004	0.007	0.021
Congestion	0.194	0.003	0.001
Accident	0.008	0.008	0.008
Air pollution	0.001	0.001	0.001
Climate change	0.005	0.005	0.005
W&T	0.032	0.032	0.032
	0.244	**0.056**	**0.068**

Basel-Duisburg
Interurban petrol GRACE car petrol EV

	Peak	**Off-Peak**	**Night**
Noise	0.005	0.009	0.027
Congestion	0.123	0.002	0.001
Accident	0.008	0.008	0.008
Air pollution	0.001	0.001	0.001
Climate change	0.005	0.005	0.005
W&T	0.019	0.019	0.019
	0.161	**0.044**	**0.061**

Duisburg-Rotterdam
Interurban petrol GRACE car petrol EV

	Peak	**Off-Peak**	**Night**
Noise	0.009	0.014	0.043
Congestion	0.122	0.002	0.001
Accident	0.006	0.006	0.006
Air pollution	0.001	0.001	0.001
Climate change	0.005	0.005	0.005
W&T	0.020	0.020	0.020
	0.163	**0.048**	**0.076**

Table 4 shows what motorists pay for these routes (it is doubtful whether vehicle excise duty should be included here, as it is a fixed cost of car ownership and is unlikely to influence the decision to drive on a particular journey). It is found that in the peak there is a significant benefit of up to 10 eurocents per kilometre from removing cars from untolled roads, whilst in the off peak cars pay around their marginal social cost on untolled roads and more than that where a toll is payable. Of course, a higher shadow price of carbon would affect this comparison but as is seen greenhouse gas costs are not a large part of the total. In other words for road transport, the biggest issue concerns congestion. But it is unlikely that there will be a large net benefit from relief of road congestion unless the road is congested in the off peak as well as the peak.

Table 5 shows similar estimates for social costs of air transport, taken from the IMPACT study. In the case of air, the absence of fuel tax means that there is normally no charge for environmental externalities, although this is crudely allowed for in some countries (including Britain) by a departure tax. In the absence of a departure tax there is an uncovered cost of perhaps 1.5 eurocents per passenger km on a 500 km flight, or a total of 7.5 euros. In other words, diversion of 1 million passengers from air might give a benefit of 7.5 million euros. It will be seen in the next section that this is not a very great contribution to the costs of HSR.

Table 4. **Road transport prices**

Road transport corridor segment	**Km**	**Car passenger toll € km**	Fuel tax gasoline € km	**Vehicle excise duty per km car gasoline**	**Total price (€)**
A8-A-9 Milano-Chiasso (I)	50	0.055	0.064	0.013	0.132
E35 Chiasso-Basilea (CH)	279	0.093	0.053	0.010	0.156
A5-E35 Basel-Duisburg (D)	584	0.046	0.056	0.012	0.114
E35-A25 Duisburg-Rotterdam (NL)	204	-	0.058	0.020	0.078

Source: GRACE D7.

Table 5. **Externalities - air (eurocents 2 000 per passenger-km)**

	Air Pollution	**Climate Change**	
Flight Distance (km)	**Direct Emissions**	**Direct Emissions**	**Indirect Emissions**
<500 km	0.21	0.62	0.71
500 – 1000	0.12	0.46	0.53
1000 – 1500	0.08	0.35	0.40
1500 – 2000	0.06	0.33	0.38
>2000	0.03	0.35	0.40

Noise costs per landing or take-off (Schiphol)

	40 seater	**100 seater**	**200 seater**	**400 seater**
Fleet average	180	300	600	1200
State of Art	90	150	300	600

Source: IMPACT 2008 Handbook.

The other key issue for air is charging for slots at congested airports. The allocation of slots by grandfather rights, and charging structures based on average costs of running the airport (or less) means that charges may not reflect the opportunity cost of slots or the costs of expanding capacity. Where shortage of capacity is acute and the cost and difficulty of expanding capacity high, as at Heathrow, this may be a significant factor.

In other words, the biggest external benefits of HSR are likely to come where road or air are highly congested and expansion on those modes difficult and expensive, including in terms of environmental costs. Of course, HSR construction has its own external costs in terms of noise, land take and visual intrusion which must be set against these benefits. External costs for air are much higher for the shorter route due to their concentration on take-offs and landings.

Generated traffic leads directly to benefits to users, which are generally valued at half the benefit to existing users using a linear approximation to the demand curve. But there has been much debate as to whether these generated trips reflect wider economic benefits that are not captured in a traditional cost benefit analysis. Leisure trips may benefit the destination by bringing in tourist spending, commuter and business trips reflect expansion or relocation of jobs or homes or additional economic activity. The debate on these issues centres on whether these changes really are additional economic activity or whether it is simply relocated. In a perfectly competitive economy with no involuntary unemployment, theory tells us that there would be no net benefit. In practice, there are reasons why there may be additional benefits. For instance, if the investment relocated jobs to depressed areas, it may reduce involuntary unemployment. However, it is common for high-speed rail to favour central locations, and if the depressed areas are at the periphery, this is the opposite of what is desired.

It is generally accepted that reducing transport costs may lead to benefits or costs that are not reflected in a standard cost-benefit analysis, due to market imperfections such as uncompetitive labour markets or agglomeration externalities (Graham, 2005). SACTRA (1999) suggested that wider economic benefits of schemes would not generally exceed 10-20% of measured benefits, whilst a specific study of the TENS network suggested that it would not change regional GDP by more than 2% (Brocker, 2004). On the other hand there may be specific cases where effects are much larger. The impact of HSR on Lille (with its uniquely favourable location) is often cited, whilst a study of a proposed high-speed route in the Netherlands found wider economic benefits to add 40% to direct benefits. (Oosterhaven and Elhorst, 2003). Vickerman (2006) concludes that whilst high-speed rail may have major wider economic benefits, the impact varies greatly from case to case and is difficult to predict.

5. ACTUAL CASE STUDIES

There are relatively few published *ex post* cost-benefit analyses of specific high-speed rail projects, One of the few published studies, for Madrid-Seville, which opened with less than 3 million trips per annum and is still carrying only of the order of 5 million trips p.a., found the project not to be justified. (de Rus and Inglada, 1997). A summary of the appraisal is given in Table 6. In this case, it appears that the social benefits of the line do not even cover the costs of operation, so that having built it, it would have been better to have left it unused, initially at least! Neither shadow pricing labour to allow for relief of unemployment, nor a general increase in costs on all modes of transport change this unfavourable result significantly. It will be seen that no value is given for environmental benefits, although we have seen above that this is not likely to be large. There is also no benefit given for the capacity released on conventional rail or at airports; perhaps in the circumstances of Spain, this has little alternative use, although that is not always the case, as will be seen below.

Table 6. **Cost benefit analysis of the Madrid-Seville HSR**

	Social benefit of HST*	GDP growth rate (3%)	Project life (40 years)	Shadow prices for labour	Increase of 25% in generalized costs of car, train and bus
Costs					
Infrastructure	-237.761	-237.761	-237.761	-200.575	-237.761
Residual value	17.636	18.546	5.816	17.636	17.636
Trains	-58.128	-61.003	-61.700	-58.128	-58.128
Maintenance	-41.410	-41.410	-45.022	-41.410	-41.410
Operation	-135.265	-140.575	-155.516	-135.265	-135.265
Time savings					
deviated traffic:					
– Conventional train	37.665	39.950	44.582	37.665	55.119
– Car	4.617	4.898	5.469	4.617	9.779
– Bus	1.958	2.079	2.321	1.958	2.867
– Air transport	0	0	0	0	0
Generated traffic	86.718	92.080	102.951	86.718	92.703
Costs savings					
– Conventional train	18.505	19.629	21.906	18.505	18.505
– Air transport	19.020	20.157	22.460	19.020	19.020
– Bus	1.680	1.783	1.990	1.680	1.680
– Car operating costs	17.412	18.471	20.618	17.412	17.412
– Congestion	4.896	6.284	7.486	4.896	4.896
– Accidents	4.128	4.363	4.867	4.128	4.128
Net present value of HST	-258.329	-252.509	-259.533	-221.143	-228.819

* Project life (30 years), GDP growth (2.5%), social discount rate (6%)

Source: de Rus and Inglada (1997).

Table 7. ***Ex post*** **appraisal of French high-speed line construction**

		Sud Est	**Atlantique**	**Nord**	**Inter Connection**	**Rhone Alpes**	**Mediter-ranean**
Length (km)		419	291	346	104		259
Infrastructure cost	Ex ante	1 662*	2 118	2 666	1 204	1 037	4 334
(m euros 2003)	Ex post	1 676	2 630	3 334	1 397	1 261	4 272
	% change	+1	+24	+25	+16	+22	-1
Traffic (m pass)	Ex ante	14.7	30.3	38.7	25.3	19.3	21.7
	Ex post	15.8	26.7	19.2	16.6	18.6	19.2
	% change	+7.5	-12	-50	-34	-4	-11.5
Financial return (%)	Ex ante	15	12	12.9	10.8	10.4	8
	Ex post	15	7	2.9	6.5	n.a.	n.a
Social return (%)	Ex ante	28	23.6	20.3	18.5	15.4	12.2
	Ex post	30	12	5	13.8	n.a.	n.a.

Source: Conseil Général des Pont et Chaussées (2006), Annex 1.

As commented above, France is one of the countries with the most experience of HSR, and it is also a country which is systematic in conducting cost benefit analyses of all transport projects. More recently, an ex post evaluation of French HSR projects has been undertaken and compared with the *ex ante* appraisals (Table 7). It will be seen that all the lines considered were expected to have acceptable financial and social rates of return, and to carry at least 15 million passengers per annum. In practice, the out turn rates of return are generally lower, mainly because of higher infrastructure costs and lower traffic levels than forecast in some cases. However, the only line for which the social case turned out to be marginal was the TGV Nord, where the major shortfall in traffic was mainly due to extreme over estimation of Eurostar traffic through the Channel Tunnel.

6. KEY PARAMETERS INFLUENCING THE CASE FOR HSR

De Rus and Nombela(2007) and de Rus and Nash (2007) have explored the key parameters determining the social viability of high-speed rail, and in particular the breakeven volume of traffic under alternative scenarios. They built a simple model to compute capital costs, operating costs and value of time savings for a new self contained 500 km line at different traffic volumes. Typical costs were estimated using the database compiled by UIC (Table 8). A range of time savings from half an hour to one and a half hours was taken, and a range of average values of time from 15 to 30 euros per hour. Other key assumptions are the proportion of traffic that is generated, and the rate of traffic growth.

Table 8. **Estimated costs of a 500 km HSR line in Europe (2004)**

	Cost per unit (€ thousand)	Units	Total cost (€ million)
Capital costs:			
Infrastructure construction[(1)] (km)	12 000-40 000	500	6 000-20 000
Rolling stock (trains)	15 000	40	600.0
Running costs (p.a.):			
Infrastructure maintenance (km)	65	500	32.5
Rolling stock maintenance (trains)	900	40	36.0
Energy (trains)	892	40	35.7
Labour (employees)	36	550	19.8

Source: de Rus and Nash (2007).

Table 9 shows the breakeven volume in terms of millions of passengers per annum in the first year, assuming all travel the full length of the line, under a variety of assumptions about the other factors. Note that benefit growth may occur because of rising real values of time as incomes rise, as well as traffic growth. With exceptionally cheap construction, a low discount rate of 3%, very valuable time savings and high values both for the proportion of generated traffic and for benefit growth, it is possible to find a breakeven volume as low as 3 million trips per annum, but it is doubtful whether such a favourable combination of circumstances has ever been found. Construction costs of 30 million per km will carry this up to 7 million, and a reduction of the value of time savings to a more typical level to 4.5 million; lower benefit growth and levels of generated traffic will take the result to 4.3 million. An increase in the rate of discount to 5% would take the value to 4.4 million. In other words, it appears to be the construction cost that is the key determinant of the breakeven volume of traffic; all the other adjustments considered have a similar smaller impact. All of these adjustments together would raise the breakeven volume to 19.2 million trips per annum, and even worse scenarios can of course be identified. On the other hand a more modest increase of capital costs to £20 million, with a high value of time savings but a discount rate of 5%, 30% generated traffic and a 3% annual growth in benefits leads to a breakeven volume of 9 million. This represents a realistic breakeven volume for a completely new, self-contained high-speed line in favourable circumstances.

Table 9. **Breakeven demand volumes in the first year (million passengers) under varying assumptions**

Construction cost (£k per km)	**Rate of interest (%)**	**Value of time saved (euros)**	**% generated traffic (%)**	**Rate of benefit growth (%)**	**Breakeven volume (m. pass.)**
12	3	45	50	4	3
12	3	30	50	4	4.5
30	3	45	50	4	7.1
12	3	45	30	3	4.3
12	5	45	50	4	4.4
30	5	30	30	3	19.2
20	5	45	30	3	8.8

These representative breakeven volumes ignoring any net environmental benefits, but we have given reasons above to expect these to be small. What they also ignore is any network benefits in terms of reduced congestion on road and air, and also within the rail sector, and that issue will be considered further in the next section.

Construction costs vary enormously from case to case, as can be seen from Table 8, with Spain having the lowest costs and Britain the highest (Steer, Davis and Gleave, 2004). Some of these cost differences are inevitable, as a result for instance of land prices, although these do not usually account for more than around 5% of the costs of an HSR project. A very major contributor to costs is the amount of tunnelling involved, and generally the costs of entering large cities are high. The British high-speed link to the Channel Tunnel is the most expensive high-speed line ever built, largely because of the lengthy tunnelling at the approach to the London terminal to avoid environmental objections. If these costs can be avoided, for instance by using existing under or unutilised rail infrastructure, then the case can be considerably improved, even if this means a compromise regarding speeds (Whilst it may be thought unlikely that such infrastructure exists in the neighbourhood of large cities, this is not necessarily so; for instance British cities do often have such infrastructure as a result of rationalisation of rival lines built by competing companies in the early days of development of the rail system).

7. NETWORK EFFECTS

Laird, Mackie and Nellthorp (2005) demonstrate how network effects may take place within the transport sector, leading to costs and benefits beyond the project being considered, as a result of the presence of one or more of the properties of economies of scale, scope or density, congestibility and consumption externalities. How far do such benefits improve the case for high-speed rail?

We have already considered network effects on road and air infrastructure, but are there also network effects within the rail sector? Essentially the argument is that once one stretch of high-speed rail has been built, extending it further will add to traffic on the existing stretch, reducing unit costs and increasing unit revenues and benefits. At the same time, by relieving conventional lines of fast passenger trains, capacity may be released which enables other services, passenger or freight, to be improved, although their finance may be seriously weakened by taking away their most profitable traffic.

The point may be illustrated with a study for Britain which examined a whole range of alternative routes, from a short new line from London to Birmingham (under 200 km), with trains continuing to other destinations over conventional lines, to a route continuing via Leeds and Newcastle to Edinburgh and Glasgow (around 750 km).

Britain only first began considering HSR, except for the link to the Channel Tunnel, in 2002, with a study undertaken by Atkins for the Strategic Rail Authority (Atkins, 2003). The Atkins study took place in a context of rapid growth in both passenger and freight traffic in recent years, leading to forecasts of severe overcrowding on both long distance services and London commuter services, and a lack of capacity for further growth in freight. Thus a major objective of the scheme was to relieve existing routes, as well as providing faster more competitive services between the major cities. This rather general remit led to the need to generate and study a wide range of options. Altogether some fourteen options were studied in depth, the main issues being whether to have a single route north from London which might split further north to serve cities up the east and west sides of the country, or two have two separate routes, and how far north to go. The obvious starting point would be a new route from London to the heavily populated West Midlands (the initial section of route would carry no fewer than 12 trains per hour in each direction in 2016 for much of the day). The further north the line was extended, the less heavily used the new sections would be, but this effect might be offset by the fact that these extensions attract additional traffic on to the core part of the network. It is a feature of British geography that most of the main cities in Britain could be served by a single line or a short branch off it.

It was forecast that the new line if built to its extremities would attract nearly 50 million passenger trips per year in 2015, although most of these would only use part of the route. This high figure reflects the high population density of Great Britain and the large number of origin-destination pairs that the line would serve. Of these passengers around two thirds would be diverted from existing rail routes and the remainder split almost equally between diversion from other modes and newly generated trips. Most of the forecast diversion occurred from car – the forecast of diversion from air was surprisingly low given experience of the impact of HSR on air traffic elsewhere.

Results of the appraisal of two options are shown in Table 10. Option 1 is the line from London to the West Midlands. which is the obvious first phase of any high-speed rail programme in Great Britain, and is seen to be well justified in its own right. But option 8, the extension through to both Manchester on the West Coast and right through to Scotland via the East Coast are also shown to be justified. It is obviously important, however, to examine the issue of timing and phasing. The study showed that, if feasible, immediate construction of the whole line was the best option.

Table 10. **Appraisal of Options 1 and 8 (£bn PV)**

	Option 1	**Option 8**
Net revenue	4.9	20.6
Non financial benefits	22.7	64.6
Released capacity	2.0	4.8
Total benefits	**29.6**	**89.8**
Capital costs	8.6	27.7
Net operating costs	5.7	16.3
Total costs	**14.4**	**44.0**
NPV	15.3	45.7
B/a	2.07	2.04

Source: Atkins, 2003) Summary Report, addendum, Table 2.1, with errors corrected.

Although net revenue more or less covers operating costs for both options, the capital cost can only be justified by non financial benefits and released capacity. Some 75% of the non financial benefits take the form of time savings or reduced overcrowding with the remainder mainly taking the form of reduced road congestion and accidents. On balance it was thought that the non quantified environmental benefits were slight. It is an interesting question whether more of the user benefits could be captured as revenue by more sophisticated yield management techniques than the simple fare structure modelled. Such yield management methods are already in use on other high-speed services, including Eurostar services between London, Paris and Brussels.

Table 11. **Unit costs and revenues**

Option	**HSR train-km(2016)**	**Capital cost per train-km**	**Net revenue per train-km**
1	55 474	2.58	1.47
8	162 067	2.85	2.12

Table 11 compares unit costs and unit incremental revenues. The capital cost per train km of the larger option is somewhat higher than for option 1, for the obvious reason that the density of trains on the route diminishes once the junction with the branch to Birmingham is reached (train-km per route-km would fall from around 300 per day to nearer 200 when we move from option 1 to option 8). However, the incremental revenue per train-km also rises quite substantially, even though on average the additional route is less intensively used than the initial stretch. The reason for this is clearly that the longer route attracts more traffic raising both mean fares and load factors on the first section of the route. Thus the more extensive network covers a much greater share of its costs from incremental revenue than the more limited network.

Table 10 also shows the estimated value of the improvements in services and increased traffic that could be carried on existing lines as a result of the construction of the new HSR line. These improvements would mainly affect London commuter services and freight traffic, where, in the absence of new capacity, severe constraints on capacity would apply. Naturally, these benefits are assessed to cover a much greater share of the capital cost of option 1, which duplicates the heavily used West Coast Main Line into London, than further north.

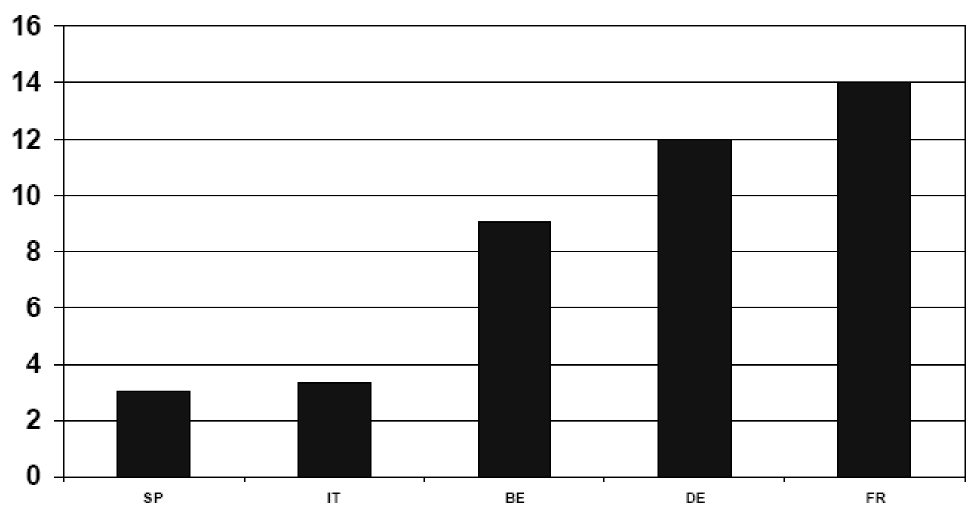
8. PRICING POLICY

To the extent that HSR is built with government funding, the opportunity cost of that funding should be taken into account by use of a shadow price of public funds, or by requiring a benefit-cost ratio well in access of one. Where private financing is involved, this will need to be serviced, and the most obvious source of income for this is via track access charges.

The method of financing high-speed rail can also be significant in determining the outcome. UIC (2008) find that the access charges levied on train operators vary substantially, but absorb between 25-45% of the revenue of high-speed rail operators. As such, they significantly affect the competitive position of rail as opposed to other modes.

Some typical track access charges for HSR are illustrated in Figure 1.

Figure 1. **Typical access charges for high-speed passenger trains € per train-km in 2008**



Source: ITF (2008), based on the approach of ECMT (2005).

In Britain, variable track access charges are based on estimated short run marginal wear and tear cost, and for a class 390 pendolino tilting train, running at 200 kmph on conventional track, the current charge is around 14p per vehicle mile. This amounts to roughly 1 euro per train km or 2 euros per 1000 gross tonne km. This figure is based on a cost allocation model resting on engineering judgement. The only econometric study of rail infrastructure costs which produces separate figures for high-speed passenger services is the Quinet and Gaudry work for France (Gaudry and Quinet, 2003). They find a value of around 2 euros per train km for high-speed and other inter city services, and of 3 euros per train km for other passenger trains. To this must be added a small amount of external cost; where track capacity is scarce, a more substantial scarcity charge may be justified. Nevertheless, it therefore appears that charges in Belgium, Germany and particularly France (as well as through the Channel Tunnel and to London) may substantially exceed marginal cost, even if environmental costs are charged for.

Of course, marginal social cost pricing in the rail sector is only optimal to the extent that it is adopted on competing modes as well. To the extent that air transport is not charged appropriately for scarce runway capacity and for environmental costs, there may be a case for charging rail below marginal cost on routes that are competitive with air.

The impact of high track access charges may be minimised by means of Ramsey-Boiteux pricing (Ramsey, 1927; Boiteux, 1956). Essentially this means pricing up more in those market segments which are least sensitive to price. This is permitted under the EU Directive on track access charges (2001/14), provided there is no discrimination between different operators competing for the same market segment.

Crozet (2007) calculates the value of the optimum mark up, assuming that the shadow price of public funds is 1.3 (Crozet, 2007). For the French high-speed network, the optimal mark up would range between 3.2 and not more than twice the marginal cost, for elasticities of 0.7 (Paris-Lyon) and 1.5 (Paris-Nice), respectively. That is, even allowing for the opportunity cost of government funds, infrastructure charges for high-speed lines should not be higher than 6.4 €/train-km taking 2 €/train-km as an upper limit to the marginal infrastructure cost per train km for high-speed rail and a price elasticity of 0.7 and if there is no environmental charge (which arguably should be the case given the general absence of environmental charges in air transport). As seen from Figure 1, the typical mark ups for access to high-speed lines in France greatly exceed these levels. The impact of high track access charges on the new route could be even more problematic if open access competition is permitted on the existing lines at much lower charges.

Adler, Pels and Nash (2008) modelled competition between rail and air on a number of Trans-European Network corridors where investment in high-speed rail is either underway or proposed, using a game theory model to compute Nash equilibria. They assumed competition between low cost and conventional airlines but no within mode competition on rail.

Where high-speed rail was introduced with a low track access charge of 2 euros per train km, they found high-speed rail to be socially worthwhile, even though a profit maximising monopoly rail operator would use much of the benefit to raise price rather than increase market share (although, as noted above, a sophisticated yield management system might be able to achieve both of these aims simultaneously). However, when access charges were raised to 10 euros per train km, services ceased to be profitable and would not operate without subsidy. In general, a high access charge will limit the frequency of service offered below the optimal level, and thus also limit the benefits.

On the other hand, a fixed charge as part of a two part tariff could make a major contribution towards the costs of building the network. However, such a charge is problematic if open access competition is to be introduced. What contribution should new entrants make to the fixed charge? The answer provided by the literature is that the new entrant should pay for the reduction in profitability of the existing operator (Baumol, 1983), but such a system is hard to administer. On the other hand, a franchising system – including a cap on the fares to be charged – can reconcile the desire to make a contribution to fixed costs with a wish to charge for track access at marginal cost; in this case the contribution could come from the willingness of the franchisee to pay for the franchise

9. CONCLUSIONS

Most successful applications of high-speed rail seem to arise when there is both a need for more rail capacity and a commercial need for higher speeds. It seems difficult to justify building a new line solely for purposes of increased speed unless traffic volumes are very large, but when a new line is to be built, the marginal cost of higher speed may be justified; conversely the benefits of higher speed may help to make the case for more capacity. It follows from the above that appraisal of HSR will need to include assessment of the released capacity benefits for freight, local and regional passenger services and the changes in service levels on the conventional lines. It also follows that the case for HSR is heavily dependent both on future economic growth and on the assumption that demand for long distance passenger and freight transport will continue to increase. If long run economic recession, or environmental constraints prevent this from occurring then far less new HSR will be justified than in a 'business as usual' scenario. Already the current recession will have at least delayed the case for some new lines, although increased government spending to reflate the economy may have the opposite effect.

High-speed rail is more successful at competing with air than car, and there is evidence for the widely quoted three-hour rail journey time threshold (although this evidence predates the increased security and congestion at airports which is believed to have increased this threshold). Where rail journey times can be brought close to or below three hours HSR can be expected to take a major share of origin-destination aviation markets.

Of the measured indirect benefits of HSR investment, congestion is the most significant. Relief of road congestion is, however, unlikely to be a major part of the case for high-speed rail except where chronic congestion is spread throughout the day along much of the route. Relief of airport capacity through transfer of domestic legs from air to rail is potentially more important where capacity is scarce and expansion is difficult, costly and has a serious environmental impact, as in the case of Heathrow.

Environmental benefits are unlikely to be a significant part of the case for high-speed rail when all relevant factors are considered, but nor are they a strong argument against it provided that high load factors can be achieved and the infrastructure itself can be accommodated without excessive environmental damage. A key factor here is the approach to cities, where the choice may be between use of conventional tracks at reduced speed or expensive tunnelling.

The issue of wider economic benefits remains one of the hardest to tackle; such benefits could be significant, but vary significantly from case to case, so an in-depth study of each case is required.

The breakeven volume of passengers to justify a new high-speed line is very variable, ranging from 3 million to 17 million in the first year of operation under possible assumptions examined, but typically even under favourable conditions at least 9 million passengers per annum will be needed. Whilst it appears that all the French high-speed lines comfortably exceeded this volume, it is clear that some proposals are being developed where traffic is very much less dense (the Madrid-Seville line, for instance, carried less than 3m passengers in its second year of operation and is still only at around the 5 million level). The most important variable in determining the breakeven volume is the construction cost, which varies enormously according to circumstances.

It is important to consider network effects. The benefits of a high-speed line may be maximised by locating it where it may carry traffic to a wide number of destinations using existing tracks beyond the end of the high-speed line, whilst extensions to an existing network lead to greater benefits than isolated new lines by attracting increased traffic to the network as a whole. Obviously this implies technical compatibility between HSR and existing rail as a prime requirement.

ACKNOWLEDGEMENTS

The author wishes to acknowledge helpful comments on an earlier draft by Peter Mackie, Stephen Perkins, Emile Quinet, Gines de Rus and Tom Worsley. Responsibility for the final version is, of course, solely the author's own.

BIBLIOGRAPHY

Adler, Pels and C. Nash (2008), High-speed rail and air competition: game engineering as a tool for cost-benefit analysis. Unpublished paper from the FUNDING project.

Atkins (2003), *High Speed Line Study*, London.

Baumol, W.J. (1983), "Some Subtle Issues in Railroad Deregulation", *International Journal of Transport Economics*, Vol. 10, pp. 341-355.

Boiteux, M. (1956), Sur la gestion des monopoles publics astreints à l'équilibre budgétaire, *Econometrica* 24, pp. 22-40, published in English as "On the management of public monopolies subject to budgetary constraints", *Journal of Economic Theory*, 3, pp. 219-240.

Bonnafous, A. (1987), The Regional Impact of the TGV, *Transportation*, Vol. 14, pp. 127-137.

Brocker, J. *et al.* (2004), IASON, Deliverable 6.

CE Delft (2003), To shift or not to shift, that's the question. The environmental performance of the principal modes of freight and passenger transport in the policymaking context, Delft.

Campos, J. and P. Gagnepain (2007), Measuring the intermodal effects of high-speed rail. Unpublished, University of Gran Canaria.

Commission of the European Communities (1990), The European High-Speed Rail Network. Report of the High Level Group, Brussels.

Community of European Railways (1989), Proposals for a European High-Speed Network, Paris.

Conseil Général des Ponts et Chaussées (2006), Les Bilans LOTI des LGV Nord Europe et Interconnexion Ile de France (accessed on www.rff.fr).

COST318 (1998), Interaction between High-Speed Rail and Air Passenger Transport: European Commission: Directorate General of Transport.

Crozet, Y. (2007), "Infrastructure charging within the French railways sector: a new challenge", presented at the 11th World Conference on Transportation Research, Berkeley, June 2007.

de Rus, G. and V. Inglada (1997), Cost-benefit analysis of the high-speed train in Spain. *The Annals of Regional Science*, 31, 175-188.

de Rus, G. and C.A. Nash (2009), In what circumstances is investment in HSR worthwhile? In: G de Rus (ed.), *Economic Analysis of High-Speed Rail in Europe*, Fundacion BBVA, Madrid.

de Rus, G. and G. Nombela (2007), Is investment in high-speed rail socially profitable? *Journal of Transport Economics and Policy*, 41(1) 3-23.

ECMT (2005), Charges for the Use of Infrastructure in ECMT Railways. Report and Recommendations. ECMT, Paris.

Gaudry, M. and E. Quinet (2003), *Rail track wear-and-tear costs by traffic class in France,* Université de Montréal, Publication AJD-66.

Gibson, S., G. Cooper and B. Ball (2002), "Capacity charges on the UK rail network", *Journal of Transport Economics and Policy,* 36, 2, 341-354.

GRACE (2005), Generalisation of Research on Accounts and Cost Estimation. European Commission project under the Transport RTD of the 7th Framework Programme, Institute for Transport Studies, University of Leeds.

Graham, D.J. (2005), Wider Economic Benefits of Transport Improvements: Link Between Agglomeration and Productivity, Imperial College, London.

Hensher, D.A. (1977), *Value of Business Travel Time*, Pergamon Press, Oxford.

Ishikawa and Imashiro (1998), *The Privatisation of Japanese National Railways*, Athlone, London.

INFRAS/IWW (2004), External Costs of Transport: Update study. Final Report, Zurich/Karlsruhe.

International Transport Forum (2008), *Charges for the Use of Rail Infrastructure*, OECD, Paris.

Kroes, E. (2000), *Air-Rail Substitution in the Netherlands*, Hague Consulting Group.

Laird, J.J., J. Nellthorp and P.J. Mackie (2005), Network effects and total economic impact in transport appraisal, *Transport Policy*, 12, pp. 537-544.

Marks, P., A.S. Fowkes and C.A. Nash (1986), Valuing Long Distance Business Travel Time Savings for Evaluation: A Methodological Review and Application. PTRC Summer Annual Meeting.

Oosterhaven, J. and J.P. Elhorst (2003), Indirect economic benefits of transport infrastructure investments, in: Dullaert *et al.* (eds.), *Across the border: building on a quarter century of transport research in the Benelux*, Antwerp, de Boeck.

Ramsey, F. (1927), "A contribution to the theory of taxation", *Economic Journal*, 37/1.

SACTRA (Standing Advisory Committee on Trunk Road Investment) (1999), *Transport and the Economy*, London.

SDG (2004), High-Speed Rail: International Comparisons, Final report. Commission for Integrated Transport.

SDG (2006) Air and Rail Competition and Complementarity. Final report. European Commission, DGTREN.

Union Internationale des Chemins de Fer (2008), *High-speed rail. Fast track to sustainable mobility*, UIC, Paris.

Union Internationale des Chemins de Fer (2008), Infrastructure Charges for High Performance Passenger Services in Europe, UIC, Paris.

Vickerman, R. (2009), Indirect and wider economic impacts of high-speed rail, in: G. de Rus (ed.), *Economic Analysis of High-Speed Rail in Europe*, Fundacion BBVA, Madrid.

Wardman, M. (2001), "A review of British evidence on Time and Service Quality", *Transportation Research E*, Vol. 37, No. 2, pp. 107-128.

Whitelegg, J. *et al.* (1993), *High-Speed Trains – Fast Tracks to the Future*, Leading Edge Publications in association with Stockholm School of Economics, Hawes.

Wilken, D. (2000), Areas and Limits of Competition between High-Speed Rail and Air. Paper presented at the Think-Up Project Workshop, Dresden.

THE HIGH-SPEED INTER-CITY TRANSPORT SYSTEM IN JAPAN: PAST, PRESENT AND THE FUTURE

Katsuhiro YAMAGUCHI
Kiyoshi YAMASAKI

Graduate School of Public Policy
The University of Tokyo
Japan

SUMMARY

EXECUTIVE SUMMARY

With the advent of Shinkansen in 1964, a unique inter-city transport network emerged, in which high-speed railway and air transport developed simultaneously in Japan, giving rise to modal choice between them based on price and speed.

Looking ahead, the next generation of high-speed transport, the Maglev, is on the horizon. In order to capture the full impacts of Maglev technology, simulation analysis with a dynamic spatial nested logit model was conducted. From this we identify a significant opportunity for the Maglev Super-Express between Tokyo, Nagoya and Osaka, but net benefits would exceed net costs only when approximately 2-3% annual economic growth is achieved over the next 65 years in Japan. If such an economic condition is realised, the total air transport market would also continue to grow, despite strong competition from the Shinkansen/Maglev system.

Another point of interest is Maglev's impact on reducing global warming. CO_2 emissions from Maglev are about one-third of those from air transport. The introduction of the Maglev Super-Express in inter-city transport, however, also attracts passengers from Shinkansen which has five times lower CO_2 emission intensity than air transport. Indeed, our simulation analysis shows that total CO_2 emissions from high-speed inter-city transport increase when the Maglev Super-Express is introduced. The increase in total CO_2 emissions from electricity users, including the Maglev Super-Express, could be mitigated through efforts by the energy conversion sector to reduce the CO_2 content of the electric power supply, for instance, by increasing the use of nuclear energy. Further research on assessing the possible impact of capacity constraint on the existing network, not considered in this paper, would facilitate deeper understanding of future high-speed intercity transport systems.

1. INTRODUCTION

The increasing value of time in modern society has brought high-speed railway and air transport to the forefront of today's inter-city transport. With the advent of Shinkansen in 1964, Japan has unveiled the significant potential of high-speed railways in inter-city travel. The ICE in 1991 and TGV in 1993 have opened a new era for Europe, and at the start of the 21st century South Korea, followed by China, has introduced their system. This year, the United States' President (USA) has announced his vision for high-speed railways.

Unlike in the USA, where air transport has long stood as the dominant inter-city transport mode, air transport in Japan developed side-by-side with Shinkansen. Liberalization and infrastructure development have helped Japan to establish an extensive network for the air transport market, filling the gap in market segments that Shinkansen cannot fill. The two different modes of transport, high-speed rail and air transport, have provided Japan with a modern inter-city transport system with the unique feature of extensive competition between them.

Looking ahead, we see a new technology for the next generation of high-speed transport, the Maglev. A business plan to introduce the Maglev system between Tokyo and Nagoya by 2025 has recently been released. We thus need to anticipate a new high-speed inter-city transport system with three different modes of travel.

This paper highlights the historical landmarks of how high-speed railway and air transport developed in Japan, and takes a look beyond the horizon of future inter-city transport. Various transport statistics are compiled and analysed in an attempt to underpin the characteristics of these transport modes. We also set up a dynamic spatial nested logit model to assess the nation-wide impact of the Maglev Super-Express.

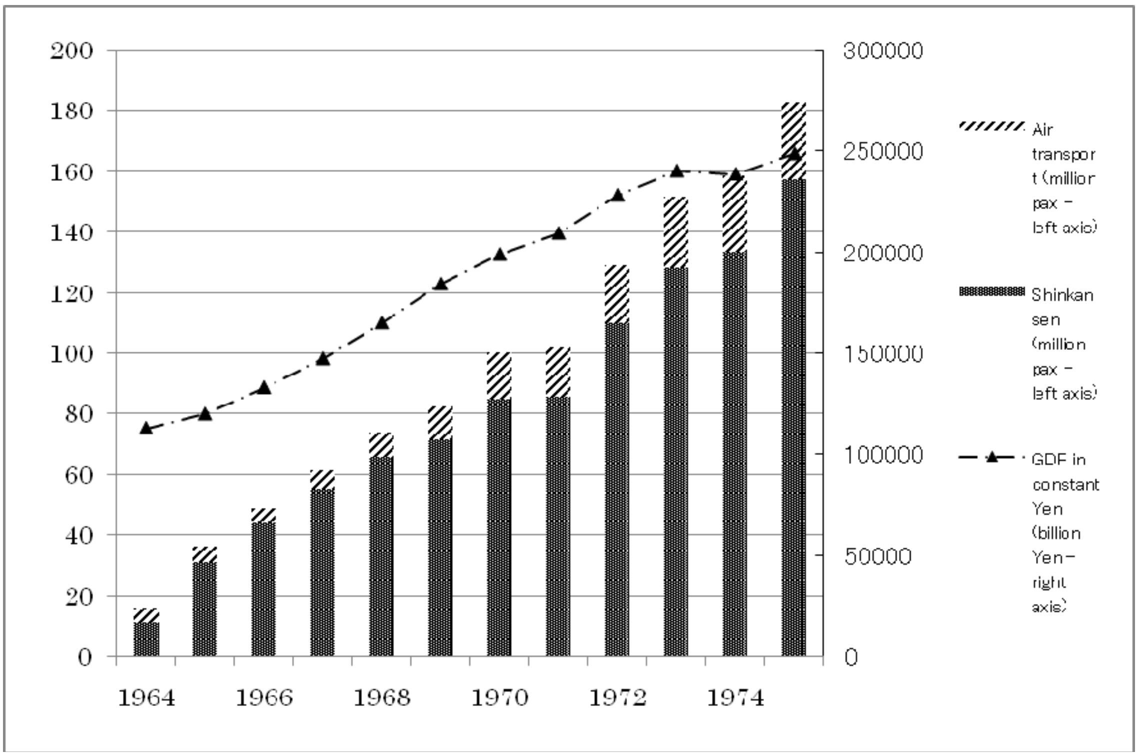
2. THE EVOLUTION OF HIGH-SPEED INTER-CITY TRANSPORT IN JAPAN

2.1. 1960-70

In October 1964, in the era when the maximum speed on the railway system was 120km/h, Shinkansen with a maximum speed of 210 km/h was considered as the super-express "dream come true". The previous seven-hour trip between Tokyo and Osaka, 550 km in length, was cut to four hours and ten minutes by the initial bullet train. At first, ten "Hikari" super-express trains that only stopped at Nagoya and Kyoto between Tokyo and Osaka, and ten "Kodama" express trains that stopped at other stations were operated. The first fleet consisted of twelve cars with a total of 987 seats. The capacity of passenger railway transport between Tokyo and Nagoya increased by 42% even though the rapid train service on the existing network was reduced by more than 30%. Within one year

Shinkansen was speeded up to shorten the trip between Tokyo and Osaka to three hours and ten minutes. Frequency was increased to 55 round-trips per day. The fare between Tokyo and Osaka by Hikari was 2 480 yen. In six months, Shinkansen's ridership reached 11 million. In particular, speed and price significantly attracted business trip-makers. Figure 1 shows that by 1970 the annual Shinkansen passenger ridership reached 85 million.

Figure 1. **Demand for air transport and Tokaido Shinkansen in passenger-kilometres (1964-1975)**



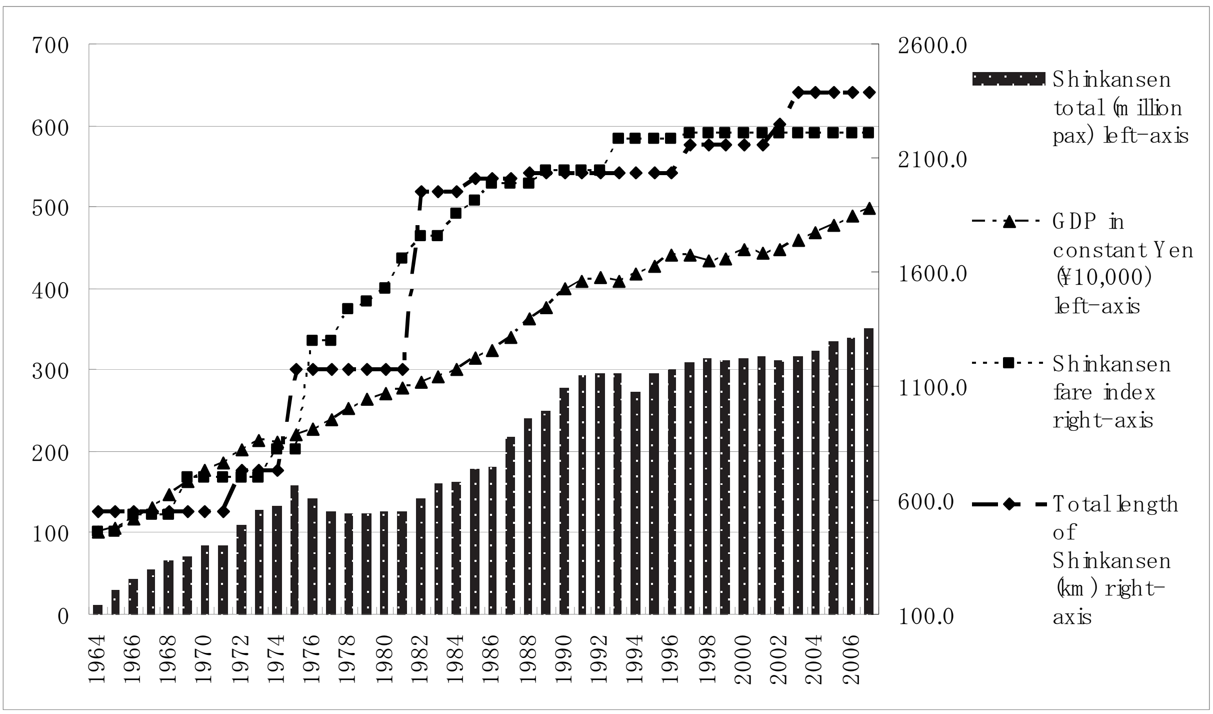
At the initial stage of air transport development, the national flag carrier, Japan Air Lines (JAL), operated on international routes and domestic trunk routes. Routes between Tokyo, Osaka, Sapporo, Fukuoka and Okinawa were designated as domestic trunk routes. Other airlines were assigned to operate on domestic local routes. An increase in demand and severe airline competition called for a new framework to secure fair competition and the orderly development of the market. A 1970 policy recommendation by the Transport Policy Council under the Ministry of Transport[1] and the Ministerial Order of 1972 outlined the subsequent regime for air transport in Japan. Under this so-called 45/47 regime[2], JAL was to serve on international and domestic trunk routes, All Nippon Airways (ANA) on domestic trunk and local routes and Toa Domestic Airlines (TDA)[3] on domestic local routes. This regime continued to be the framework for Japanese air carriers until the mid 1980s.

When Shinkansen started its operation in 1964, air transport was at the initial stage of introducing turbo-jet aircrafts. The first turbo-jet to fly in the domestic market was the Conveyer 880 on the Tokyo-Sapporo route in 1961. By 1964, Boeing 727 and DC8 joined the fleet of Japanese air carriers. The Tokyo-Osaka route, however, was still operated by turbo-prop aircrafts when Shinkansen started its operation. In those days, the average speed of domestic air transport was 333 km/h and it took an hour and forty-five minutes to fly from Tokyo to Osaka. During the first six months of Shinkansen's operation, 3.6 million passengers, equivalent to 14% of the Tokyo and Osaka air transport market, shifted to rail. Despite the dramatic success of Shinkansen, air transport marked rapid growth in the subsequent years. By 1970, the annual number of air transport passengers was above 15 million.

2.2. 1970-90

In 1972, Shinkansen was stretched to Okayama, 150 km west of Osaka, and then in 1975 to Hakata in North Kyushu, 644 km from Osaka. Now, Shinkansen was composed of 553 km of Tokaido Shinkansen and 644 km of Sanyo Shinkansen. Between 1965 and 1975, Shinkansen enjoyed 15% annual growth in passenger ridership and reached 157 million by 1975.

Figure 2. **Historical data regarding Shinkansen (1964-2007)**

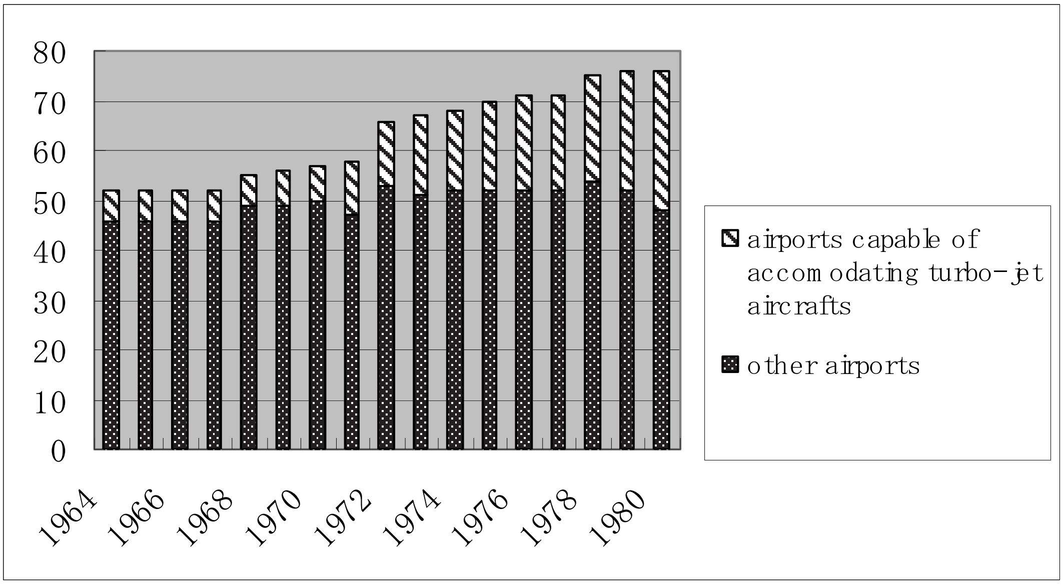


In the following years, however, Shinkansen demand started to decline. Apart from the economic downturn due to the exchange rate reform of 1971 and the oil crisis in 1973, Japan National Railways (JNR) was suffering from a huge financial deficit, accumulating year by year. Investment, maintenance and operation costs were basically self-managed by JNR. The rapid motorisation in urban and regional transport led JNR into severe financial distress. In particular, the expansion of the rail network in rural areas amplified the problem. JNR's accumulated losses skyrocketed from 83 billion yen in 1965 to 678 billion yen in 1975 and was still growing fast. The government and JNR took steps to alleviate their financial difficulties by increasing fares. A one-way Shinkansen ticket from Tokyo to Osaka, initially set at 2 480 yen, was hiked to 5 050 yen by 1974 and reached

10 800 yen by 1981; a four-fold increase in 17 years. JNR's price hike had over-ridden the CPI and the Tokyo-Osaka air fare, which rose by 2.7 times and 2.3 times, respectively, during the same period. Railway fares continued to be increased until JNR was privatised in 1987. By then, a Shinkansen ticket from Tokyo to Osaka cost 13 100 yen. The historical data depicted in Figure 2 illustrates the effect of the price hikes.

Demand for air transport had also stagnated during the late 1970s but not as severely as for Shinkansen. Turbo-jet aircraft, with faster speeds and greater capacity than turbo-prop aircraft, were introduced rapidly. As shown in Figure 3, the number of airports accommodating turbo-jet aircraft was increased from six in 1965 to 28 in 1980.

Figure 3. **Number of airports in runway categories (1964-1980)**



Class One international airports in Tokyo and Osaka were built and funded 100% by the government[4]. The central government was also task ed to own and operate Class Two airports in major cities, such as Sapporo and Fukuoka. Two-thirds of the funding was assured by central government and the rest covered by local government. Class Three airports in local cities were built and managed by local governments with half of the investment subsidized by central government. In 1967, the first of the Five-Year Airport Construction Plans was adopted. In 1970, central government established a Special Account for Airport Development, to invest and maintain the Class One and Two airports and subsidize the Class Three airports. The financial sources for the Special Account were twofold. One source was direct income from landing fees and 11/13 of the jet fuel tax levied on domestic air transport operation, sourced through the General Account of the Japanese Government. This accounts for 70%-80% of the total revenue. The rest is composed of generic funds from the General Account and provisions from the local government for Class Two airports. In the 1980s, government loans were injected into the Special Account for Airport Development to finance large investments in Haneda Airport. In 1966, the New Tokyo International Airport Agency (Narita) was established by the government. After twelve years of difficulty, Narita Airport was opened in 1978. International flights were basically shifted from Haneda to Narita, giving room to facilitate untapped demand in the domestic air transport market.

In the 1980s, Japan steadily recovered from the economic shocks. Rapid growth was experienced in both the international and domestic air transport markets. In 1985, the Transport Policy Council reviewed the 45/47 framework and recommended that the government should to turn to a pro-competitive policy. The operation of multiple numbers of airlines on routes was liberalized on high-density routes. The threshold demand level, allowing two airlines (double tracking) and three airlines (triple trucking) to operate, was set out by the Ministry of Transport. Thresholds of double/triple tracking were cut down in 1992 and in 1996 for the further promotion of competition. In 1997, the threshold itself was abolished so that any number of airlines could enter into any route regardless of the volume of that route. As a consequence, the ratio of available seats on routes with multiple numbers of airlines against total available seats in the domestic air transport market rose from 53% in 1985 to 80% in 1999. The new aviation policy, set out in 1985, also allowed airlines other than JAL to operate on international routes and JAL was privatised.

Domestic airfares were regulated to control airfares based on cost. When the airlines applied for an increase in airfares due to inflation or an upspring in the price of fuel, etc., the overall cost of airline operation was reviewed by the government. An airfare increase was only allowed up to the level justified by aggregate cost under efficient operation. Such an "aggregate cost formula" was common for public utilities.

2.3. 1990-Present

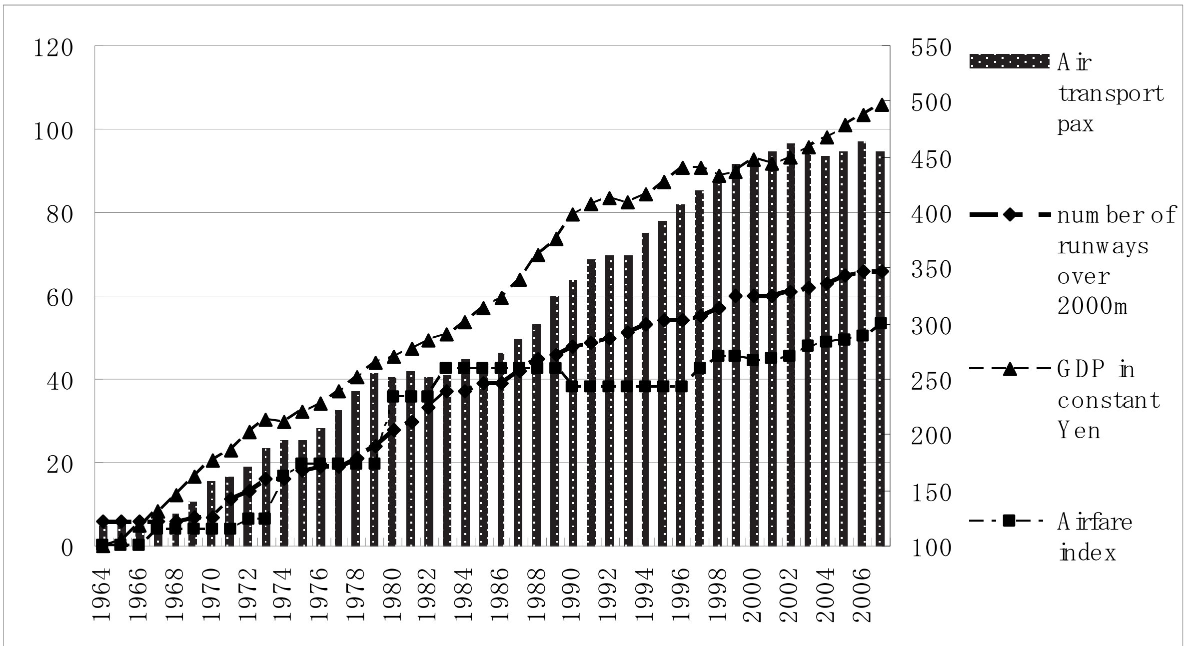
2.3.1 Liberalization in the air transport market

Due to the burst of the "economic bubble", the Japanese economy plunged into recession and prices became deflationary in the early 1990s. The opening of Kansai International Airport in 1994 would have been welcomed more if it were not for the great depression. The private sector was facing difficulties, with deteriorating demand and prices. Public utilities including transport services, however, tried to pass excessive costs to the consumer by raising prices. As from 1994, strong criticism over price hikes for public utilities pushed the regulatory reform of public utilities into a policy agenda. Amidst countervailing forces, airfare regulation was deregulated to introduce a "zone airfare scheme". This allowed airlines to obtain automatic approval within a specific zone. The new zone airfare system provided airlines with flexibility when setting air fares. Seasonal differences and flight-by-flight pricing were now possible. In 1996, the airlines' applications were approved under the new regulation. Under the new price regulation regime, incumbent airlines increased the normal fares for trunk routes while introducing various discount fares, such as advance booking discounts and frequent flyer programmes (FFPs). Despite the introduction of various discount fares, normal airfare hikes on trunk routes such as Tokyo-Fukuoka and Tokyo-Sapporo were confronted with strong criticism in the Fukuoka and Sapporo regions.

This opened a window of opportunity for entrepreneurs to set up new airlines. Airport capacity expansion of the highly congested Haneda Airport was under construction. In March 1997, a new runway was opened and 40 landing slots per day were added. These slots were allocated to airlines in two stages: July 1997 and April 1998. At that time, there were six projects launched to raise new airlines and the first two to be in the market were Skymark Airlines in September 1998 and Hokkaido International Airlines (AIR DO) in December 1998. They entered into Tokyo-Fukuoka and Tokyo-Sapporo routes respectively. Apart from subsidiaries of the major three air carriers, it was indeed the first new air carrier entry in 35 years. At the launch of their services, the two airlines set out much lower airfares compared to incumbent carriers. Skymark offered half the normal fare and AIR DO was 36% below the incumbents. This "everyday low fare" strategy became popular and their load factor rose as high as 80%. On the other hand, incumbent carriers suffered a sudden drop in passengers.

These routes were lucrative trunk routes with many business travellers. The incumbent carriers started to offer discount fares on flights just before and after the flights of new entrants. They also upgraded their frequent flier programmes. These counter measures were quite effective and by March 1999 the incumbent carriers regained their demand at the same level as that of the previous year. Enhanced competition facilitated annual passenger increase in Tokyo-Fukuoka route and Tokyo-Sapporo route, by 16.3% and 9.4%, respectively. From then on, a pro-competitive slot allocation policy at congested airports such as Haneda Airport became an important agenda for the Ministry of Transport. The new policy was introduced to review slot allocation in congested airports every five years. Figure 4 illustrates the historical trend in air transport. It could be observed that despite economic stagnation in the mid-1990s, air transport experienced moderate growth due to market stimulation from deregulation.

Figure 4. **Historical data regarding air transport (1964-2007)**



In Japan, deregulation in the transport sector has been implemented in steps. In December 1996, with a view to accelerate deregulation in every transport sector and to promote administrative reform, the Ministry of Transport decided to abolish supply/demand testing in the entire transport sector by the end of the century. Based on the report from the Transport Policy Council of April 1998, the air transport market was totally liberalized while measures for maintaining essential air services to remote islands and the rule for slot allocation in congested airports were reinforced. Having set out necessary measures for liberalization, the Civil Aeronautics Law was amended and put into effect in February 2000, so that supply/demand regulation policy was abolished and a licence for each route was no longer needed. The airfare regulation was also deregulated from approval regulation to prior notification. With regard to the congested airports, slot allocation was adopted, subject to review every five years based on pre-set allocation criteria.

According to Yamaguchi (2005), from 1980-98, the accumulated increase in consumer surplus from deregulation and public investment related to air transport amounted to 1.2 trillion and 3.5 trillion yen, respectively.

2.3.2 JNR reform and Shinkansen

The year that Shinkansen started its operations was the year that the JNR's severe financial problems became apparent. In 1964, JNR reported its first operating loss, which then grew year by year. By 1966, the capital reserve dwindled and net losses started to accumulate. In 1971, JNR reported an operating loss before depreciation. Fares were raised almost every year. Total government subsidies reached 6.6 trillion yen. Despite these measures, long-term debt reached 37.1 trillion yen, of which 15.5 trillion yen was JNR's accumulated loss. In 1987, the government put an end to JNR's financial crisis through privatisation. The JNR's reform package of 1987 was composed of the following:

a) Privatisation of JNR into six regional passenger railway transport corporations and one freight transport corporation;
b) Shinkansen would be held by a special-purpose government agency and leased to JR companies;
c) 11.6 trillion yen of the total 37.1 trillion yen long-term debt would be borne by major JR companies and the rest, 25.5 trillion yen, by a special-purpose government agency.

In 1993, JR East was floated on the stock market, followed by JR West and JR Central in 1996 and 1997, respectively. In 1991, Shinkansen assets, spun-off in the 1987 JNR reform package, were bought back by the three JR companies. The final solution to the 25.5 trillion yen long-term debt, borne by a special-purpose government agency, was achieved in 1998.

A law stipulating a nationwide plan for Shinkansen development was enforced in 1970. Under the plan, agreed in 1973, an extension of the network – northwards to Sapporo in Hokkaido and southwards to Kagoshima in Kyushu – and the development of the Hokuriku Shinkansen, connecting Tokyo and Osaka via Nagano and Toyama, were included in the development plan phase. These new routes were christened *Seibi-Shinkansen*.

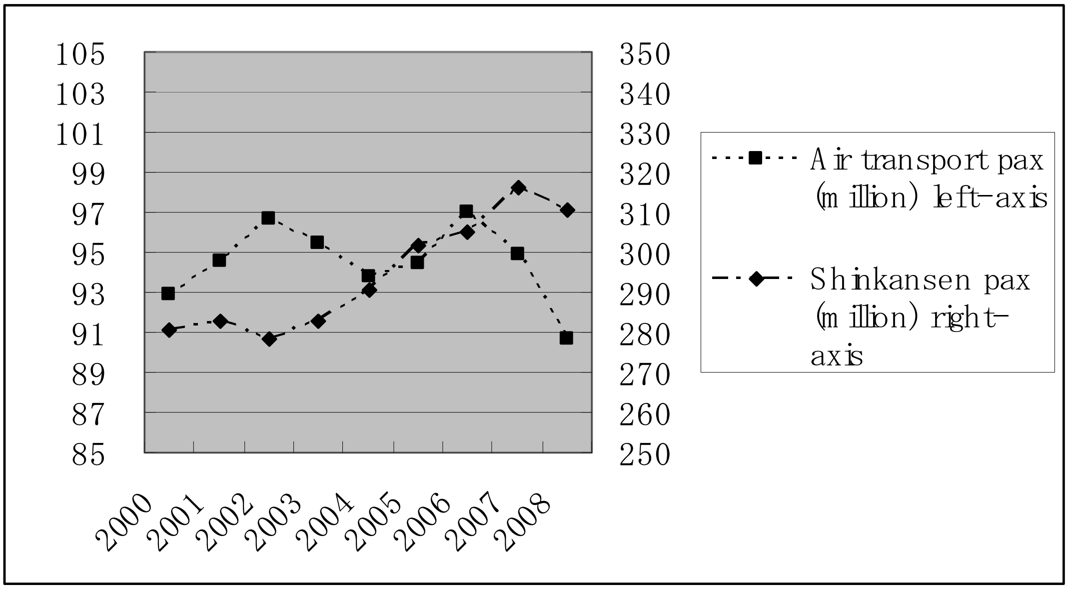
Over-investment was one of the major causes of financial turmoil for JNR. Thus, an important feature of the new Shinkansen funding scheme was to avoid a new financial crisis. A funding scheme, established in 1989 for the extension to Nagano – the first of the routes to be constructed as *Seibi-Shinkansen* – comprised 50% JR investment, 35% by central government and 15% by local government. The funding scheme was revised in 1996 so that JR would only bear investment costs up to a level where they would still benefit. The rest of the investment would be covered by the government: two-thirds by central government and one-third by local government.

2.4. Towards the future

2.4.1 Shinkansen and air transport

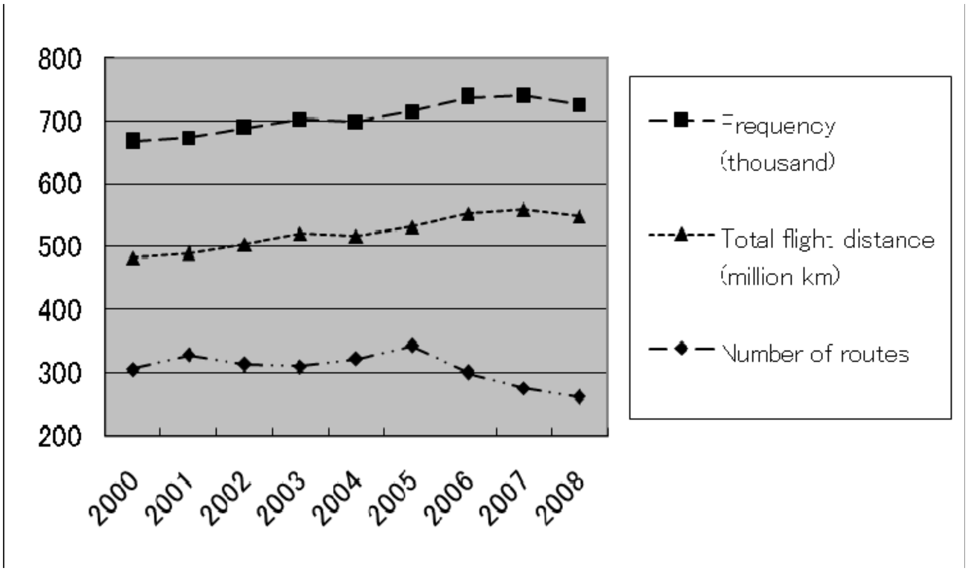
With the turn of the century, Shinkansen constantly increased its demand and, in recent years, a complementary relationship with air transport has continually been manifested. Figure 5 shows the recent annual number of Shinkansen passengers in comparison with those of air transport.

Figure 5. **Recent trend of passengers on air transport and Shinkansen (2000-2008)**



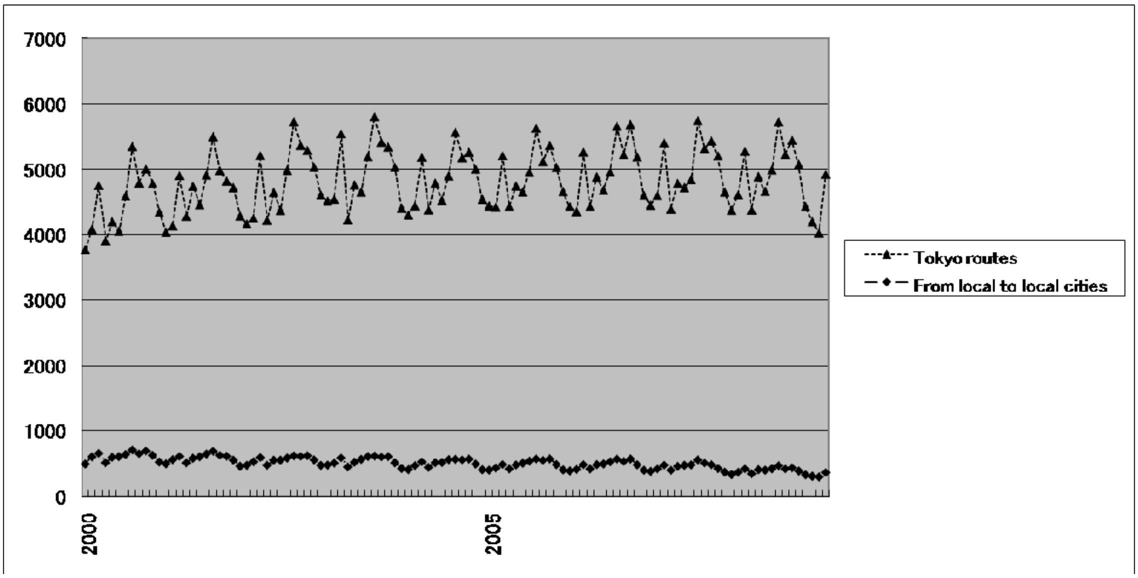
The extension of the existing Shinkansen under operation currently represents a total of 2 387 km. 1 173 km of the *Seibi-Shinkansen* network are unfinished, and due to constraints on government funds, it is estimated that it will take about ten years to be completed. Apart from the *Seibi-Shinkansen*, the Maglev Super-Express is planned to be built as part of the grand design of the national Shinkansen network, as stipulated under the National Shinkansen Law of 1970. The major difference between *Seibi-Shinkansen* and the Maglev Super-Express is that the latter is declared to be self-financed by JR Central.

Figure 6. **Recent trend of air transport (2000-2008)**



Since the turn of the century, except for 2005 when Chubu Centrair International Airport was opened, the total number of routes for domestic air transport has seen a gradual decline. On the other hand, as depicted in Figure 6, total frequency and total flight distances have increased. Routes to and from Tokyo (Haneda) are increasing in capacity and demand, while other routes, local-to-local routes in particular, are losing both. Route concentration has led the overall average frequency per route to increase by about 30% between 2000 and 2008. Figure 7 shows the trend in the number of monthly passengers on routes to and from Tokyo and local-to-local cities. While demand for Tokyo routes increased by 10%, local routes decreased by 35%.

Figure 7. **Monthly number of passengers in thousands on routes to and from Tokyo and between local cities (Jan. 2000- Mar. 2009)**



As for total domestic air transport demand, with the rise of fuel costs, the average fare (yield) per passenger-kilometre has increased from 15.0 yen/per km in 2002 to 17.6 yen/per km in 2008. As a result, the total number of passengers has declined from 96.7 million in 2002 to 90.7 million in 2008. The merger of JAL and JAS in 2002 also had an impact on the market in general. Figure 8 illustrates the recent trend.

Figure 8. **Recent trend of passengers and average price of air transport (2000-2008)**

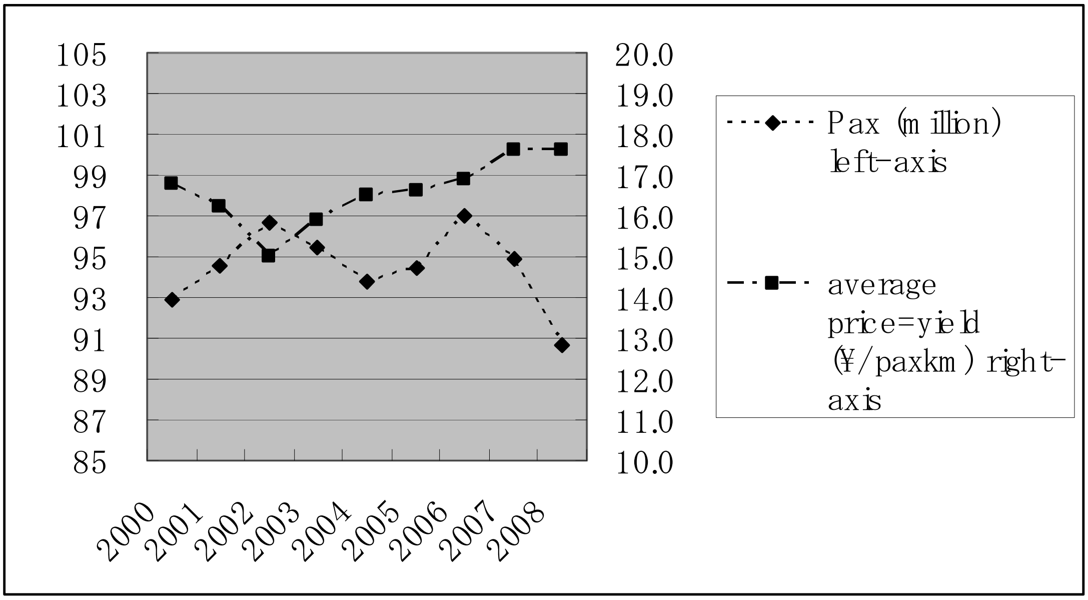


Figure 9 shows the profound effect of the world-wide economic downturn since September 2008 on air transport and Shinkansen. Both transport modes have experienced unprecedented decreases in demand in recent months. Speculators view February 2009 as the lowest point. There are hopes that the transport market, mirroring the general economic activity, will rebound in the foreseeable future.

Figure 9. **Percentage change of monthly passengers on air transport and Shinkansen (March 2007- March 2008)**

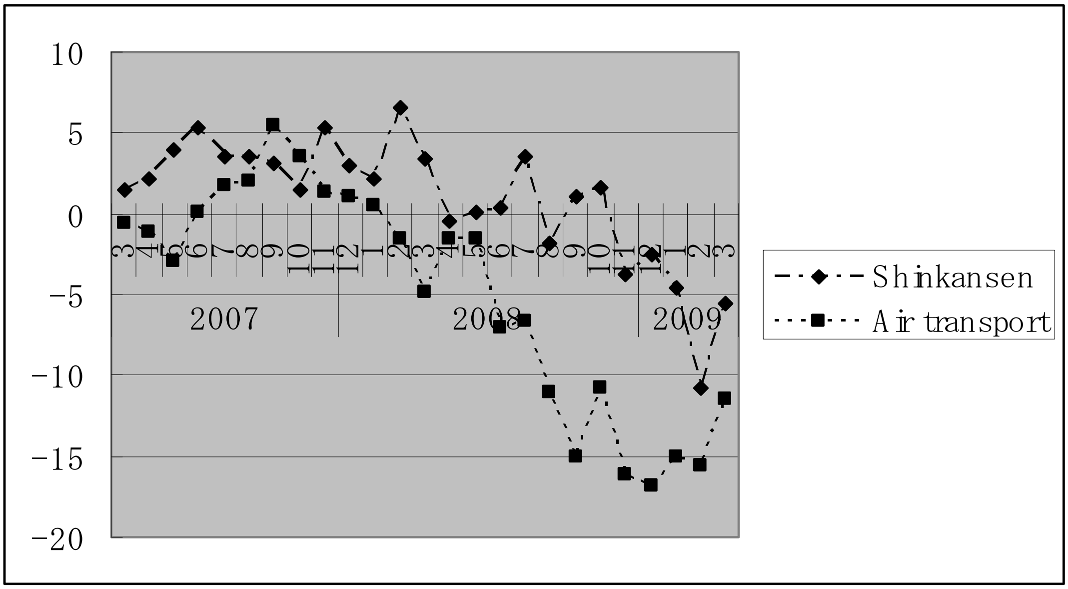
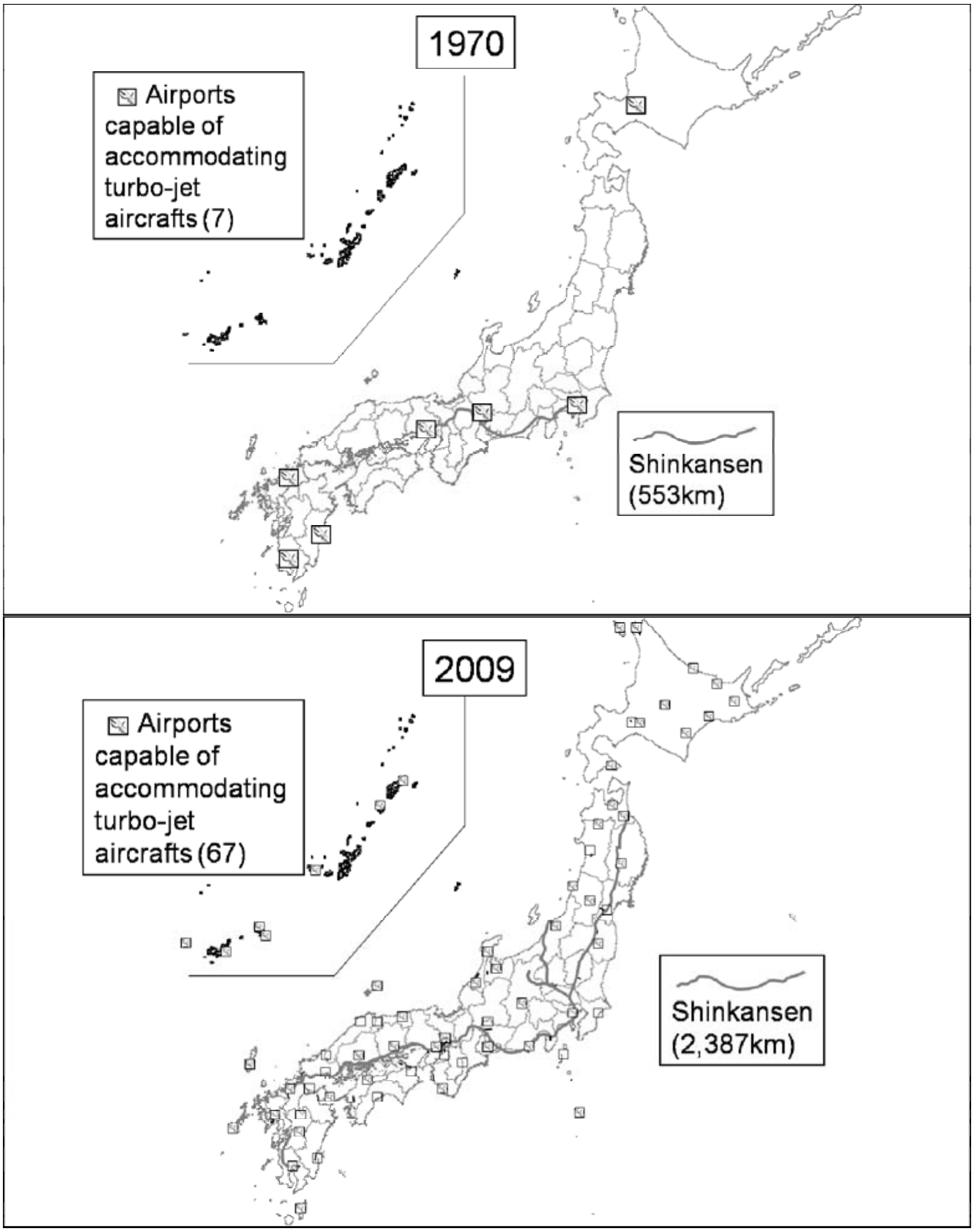


Figure 10 gives snap-shots of the Shinkansen network and airports in 1970 and 2009. It should be noted that regional airport development has basically come to an end. Now there is a need to facilitate the increase of capacity in the Tokyo metropolitan area. In 2010, landing slots in Tokyo Haneda Airport and Narita International Airport are to be increased substantially. In particular, the opening of the fourth runway at Haneda Airport is expected to have a profound impact on domestic and near-by East Asian inter-city air transport. In 2009, there were 806 domestic flights and 24 international charter flights operating daily at Haneda Airport. Domestic flights should be increased to 826 in October 2010 and then to 880 within six months thereafter.

Figure 10. **Shinkansen network and airports in 1970 and 2008**



Back in 1978, when Narita International Airport was opened, international scheduled flights were basically shifted away from Haneda Airport. With the 2010 expansion, Haneda Airport will accommodate 40 international scheduled flights daily to major near-by East Asian cities during the day and another 40 international flights between late evening and early morning. Furthermore, another 72 flights should eventually be added, the allocation of which is still to be determined.

2.4. The Maglev

The technology of the super-conductivity magnetic levitated super express, the so-called "Maglev", goes back to 1962. Ten years after the start of the research project in JNR, the first test operation was undertaken on a 220-metre strip test guideway at a research centre in Kunitachi, Tokyo. In 1974, construction of a 7-kilometre testing lane was initiated in Miyazaki, where test runs were conducted until the test bed was switched to Yamanashi in 1996. In the current 42.8 km stretch of test-course in Yamanashi, a maximum speed of 581 km/h was recorded in 2003 and in that year the government technology committee announced that the Maglev Super Express was now technologically feasible. By 2006, accumulated test runs had exceeded 500 000 km and in 2007, the test course was designated to be part of the commercial path of Chuo Shinkansen. That year, JR Central announced that they planned to open the Tokyo-Nagoya Maglev Super Express by 2025, and would be the sole investor in the 500 trillion yen project.

Chuo Shinkansen is listed as one of the routes to be developed under the National Shinkansen Development Law. The Maglev Super Express planned by JR Central is an integral part of the Chuo Shinkansen. Currently, there is debate over which specific route the Chuo Shinkansen should take. Local governments are requesting diversion of the route to local cities which would inevitably increase the construction cost of the overall Maglev infrastructure.

Table 1. **Comparison of Shinkansen, Maglev (plan) and air transport**

	Tokyo-Nagoya (366km*)			**Tokyo-Osaka (553km*)**		
	Time	**Fare**	**CO2/pax**	**Time**	**Fare**	**CO2/pax**
Shinkansen (Nozomi)	103min	10 780yen	5.2kg	156min	14 050 yen	7.9kg
Maglev (plan)	40min	(11 780 yen)	15.7kg	60min	(15 000 yen)	23..8kg
Air	-	-	-	68min	13 600 yen	68.8kg

*Distance in railway mileage.

Table 2. **CO_2 intensity**

Mode	**CO_2 -g/paxkm**
Shinkansen	14.2
Maglev	43.0
Air	124.5

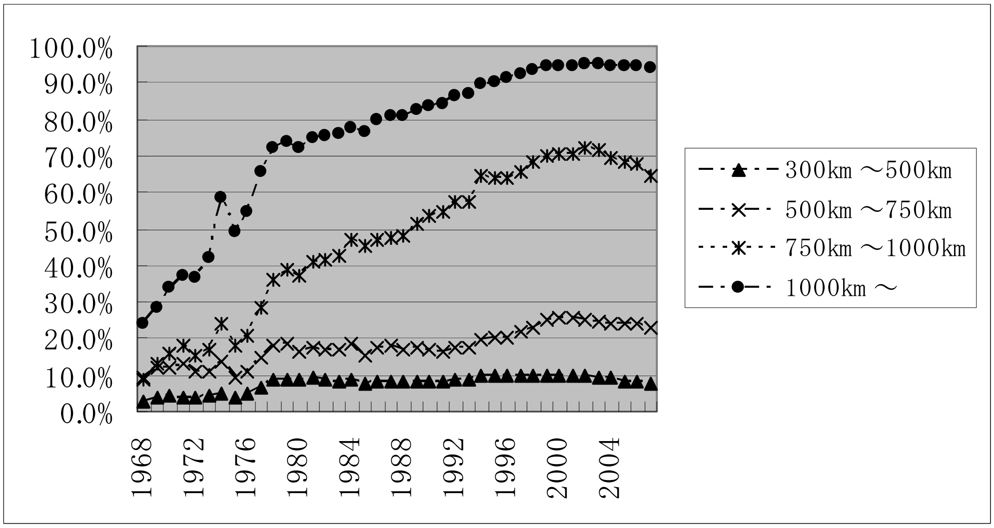
3. MARKET CHARACTERISTICS OF A HIGH-SPEED, INTER-CITY TRANSPORT SYSTEM

3.1. Average travel distance of Shinkansen and air transport

Originally, Shinkansen was utilised for long-distance travel, the majority of journeys exceeding 300 km. By 2007, however, more than half of Shinkansen ridership was for trips of less than 300 km. The average distance declined from 319 km in 1968 to 234 km in 2007. The breakdown of Shinkansen average travel distances into segments is as follows: Tokaido=308 km, Sanyo=251 km, Tohoku=168 km, Jouetsu=126 km, Hokuriku=82 km, Kyushu=103 km. Only Tokaido Shinkansen is maintaining an average ridership of more than 300 km.

On the other hand, the average trip length for domestic air transport has increased over time: 605 km in 1968 and 881 km in 2007. Average distances for Shinkansen and air transport have been diverging over the years. As a result, the modal share of air transport in long-distance travel has been increasing, as depicted in Figure 11.

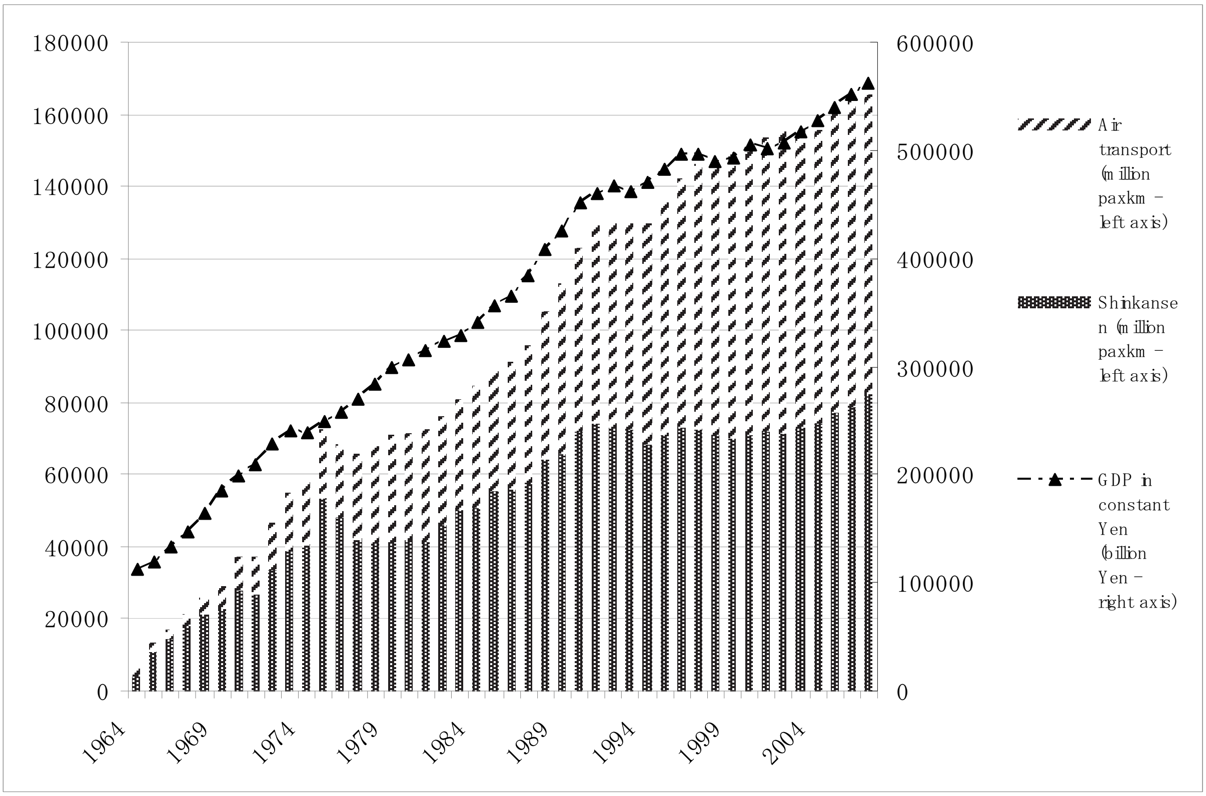
Figure 11. **Trend in share of air transport in distance groups**



3.2. Modal split between Shinkansen and air transport

From Figure 12, the aggregate demand growth of Shinkansen and air transport has basically paralleled that of GDP. When Shinkansen ridership growth stagnated between 1975 and 1985, air transport seems to have filled the gap. In order to clarify this modal choice relationship, the following logit model was estimated.

Figure 12. **Trend in GDP and passenger kilometres of Shinkansen and air transport**



3.2.1 Logit model

Here we conduct a logit model analysis using pooled historical data. Let U_k be the utility of choosing transport mode k composed of deterministic portion V_k and random variable δ so that,

$$U_k = V_k + \delta .$$

There are two transport modes, railway (R) and air transport (A). Let V_k be a function of price and defined as follows:

$$V_k = \alpha + \beta p_k$$

Where:

p_k represents fare of mode k, and
α, β are parameters.

The probability of choosing air transport or railway would be:

$$P_A = \frac{\exp(V_A)}{\exp(V_R) + \exp(V_A)} \text{ and } P_R = \frac{\exp(V_R)}{\exp(V_R) + \exp(V_A)}.$$

Let X be total demand of air transport and railway. Then,

$X_A = s_A X = P_A X$ and $X_R = s_R X = P_R X$. Thus,

$X_A / X_R = P_A X / P_R X = P_A / P_R = \exp(V_A) / \exp(V_R)$.

Taking the natural log of both sides, the formula to be estimated is as follows:

$$\ln[X_A / X_R] = \ln[P_A / P_R] = \alpha + \beta(p_A - p_R) + \varepsilon$$

Where ε is the error term.

3.2. *Description of data*

Ridership statistics are available for both Shinkansen and air transport. While route segment data is available for air transport, railway on-board segment data, including that of Shinkansen, however, is not available. It is not possible to discern how many passengers get onboard Shinkansen at Tokyo and get off at Osaka from railway statistics.

In order to identify inter-prefectural transport, a Regional Passenger Flow Survey has been conducted annually since 1960. Through this survey it is possible to know how many people travelled between and within the 47 prefectures. A breakdown into different modes of travel is provided. Therefore, it is possible to know how many people travelled between Tokyo Prefecture and Osaka Prefecture. When a multi-modal trip is made, each rider on an individual mode is counted as one. Also, the purpose of travel is unknown. However, even given these limitations, the survey does give valuable inter-prefectural data.

In order to complement the unknown factors, a Trunk Route Passenger Flow Survey has been conducted every five years since 1990. The latest survey was conducted in 2005. This detailed survey is conducted for a single day in autumn and compiled into 207 zones. The level of transport service between zones is compiled from publicly available timetables.

There are two datasets for X. Data-set A is composed of the number of annual passenger-kilometres performed by Shinkansen and air transport (1965-2007). Data-set B is composed of the total number of trips made over 300 km by railway and air transport (1968-2007). As for transport cost p, Shinkansen and the airfare between Tokyo and Osaka are chosen as representative price data (1964-2007). Prices are inflation-adjusted by the Consumer Price Index. These data are pooled and regressed by the ordinary least-square method.

3.2.3 *Result of the estimate*

The estimates of β for the two datasets are -1.2 and -1.7, respectively, and both statistically significant (Table 4). They are consistent with past studies.

Table 3. **Modal split parameter**

Parameters	Data set A		Data set B	
	Parameter	t-ratio	Parameter	t-ratio
Constant(α)	0.070	1.194	0.399	4.682*
Transport cost（β）	-1.242	-11.804*	-1.711	-9.535*
R^2	0.699		0.705	
Sample size	43		40	

*Significant at 1% level.

Average own-price elasticity $|\beta p_k(1-s_k)|$ and average cross-price elasticity $|\beta p_k s_k|$, calculated from estimated parameter and data sets A and B, are listed in Table 4. These figures are consistent with past studies.

Table 4. **Average price elasticity**

	Data set A	Data set B
Own price elasticity (average)	0.70	0.89
Cross price elasticity (average)	0.94	1.51

The transport demand share between air transport and Shinkansen, or travel over 300 km by air transport and railway, is significantly correlated with the relative price difference. In this model, however, spatial conditions and speed factors are ignored. In order to analyse the air-rail relationship in a more comprehensive manner, we need to develop a spatial model that breaks region into zones, as well as to take different trip purposes into account. Looking into the future, there is also a need to consider changes in population, economic growth and new technology for inter-city transport. In the following section, we develop a nationwide inter-city transport demand model to assess the impact of the Maglev Super-Express.

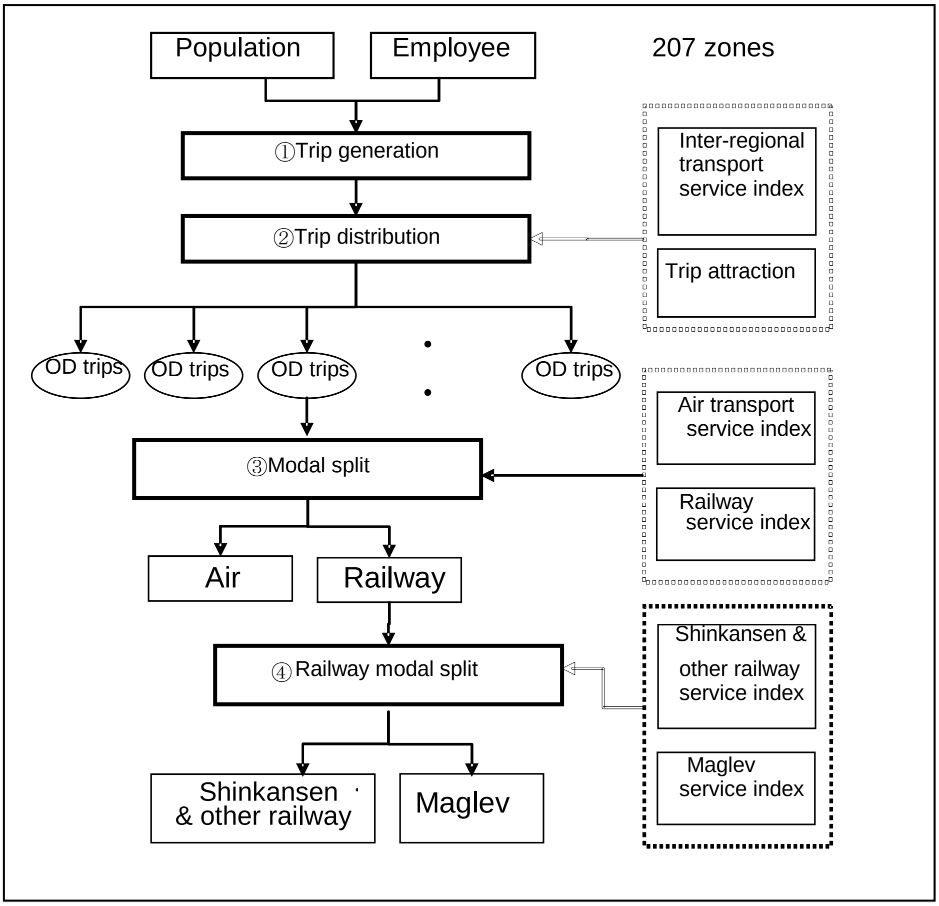
4. SIMULATION ANALYSIS OF A FUTURE HIGH-SPEED INTER-CITY TRANSPORT SYSTEM WITH MAGLEV

4.1. Model structure

The model is structured in four stages, as illustrated in Figure 13.

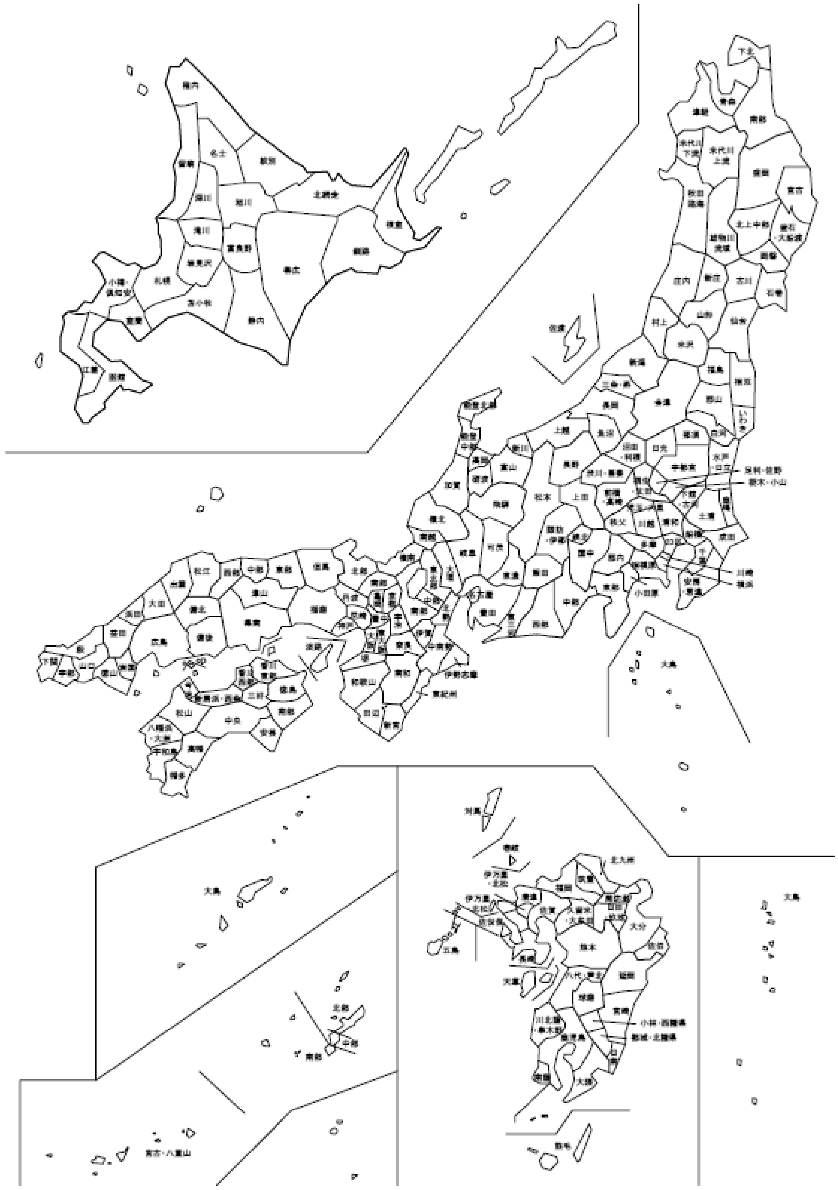
1. National trip generation model;
2. Zone-to-zone trip distribution model;
3. Air *vs.* rail modal split model;
4. Shinkansen and other railways *vs.* Maglev choice model.

Figure 13. **Model structure**



The spatial inter-city demand model is developed by breaking Japan into 207 zones, as depicted in Figure 14. The model is separated into three different trip purposes: business, tourism and private.

Figure 14: **Japan in 207 Zones**



4.2. Trip generation model

4.2.1 Model structure

Trip generation is modelled as a function of population and trips per capita. For business travel, the number of employees is used for population.

$$T_{im} = POP_{im} GA_{im} \qquad (1)$$

T_{im}: trip generation in zone i with trip purpose m

POP_i: population of zone i

GA_{im}: per capita number of trips from zone i trip purpose m

4.2.2 Trip generation model

Trip generation per capita is modelled as a function of level of service and price and income elasticities. The parameter is calibrated so that current trip generation per capita of that zone matches the model value.

$$GA_i = \beta_{0i} q_i^{\beta_1} (1+\mathrm{n}\eta) \qquad (2)$$

β_{0i}: parameter to be caliberated from current level of GA_i and accessibility index to other zones q_i

β_1: price elasticity

q_i: accessibility index derived form the trip distribution model

n: annual GDP growth rate

η: income elasticity

4.3. Trip distribution model

The objective of the trip distribution model is to allocate trips generated in a specific zone (zone i) to other destinations. We use a nested logit model to calculate the proportion of trips to destinations. From zone i, the probability of zone j being selected as a destination (P_{ij}) depends on the utility level of a trip between zone i and zone j (Vij) among the available destinations. The utility level of a trip between zones i and j depends on the service level of transport modes between the two zones (q_{ij}), and the attraction factor of the destination zone j (S_j). q_{ij} is derived from the log-sum of the transport mode selection model described below. The aggregation of trips destined to zone j is used as the attraction factor of zone *j*.

Parameter θ_1^D used in the log-sum factor is an estimated figure from the Annex.

$$q_i = \frac{1}{\theta_1^D} \ln\left[\sum_j \exp\left(V_{ij}\right)\right] \qquad (3)$$

$$P_{ij}^D = \frac{\exp\left(V_{ij}\right)}{\sum_j \exp\left(V_{ij}\right)} \qquad (4)$$

$$V_{ij} = \theta_1^D q_{ij} + \theta_2^D \ln S_j \qquad (5)$$

$$q_{ij} = \frac{1}{\theta_1^S} \ln\left[\exp\left(\theta_1^S q_{ij}^A + \theta_2^S\right) + \exp\left(\theta_1^S q_{ij}^R\right)\right] \qquad (6)$$

P_{ij}^D: probability of choosing zone j as destination

V_{ij}: utility of travelling between zones i and j

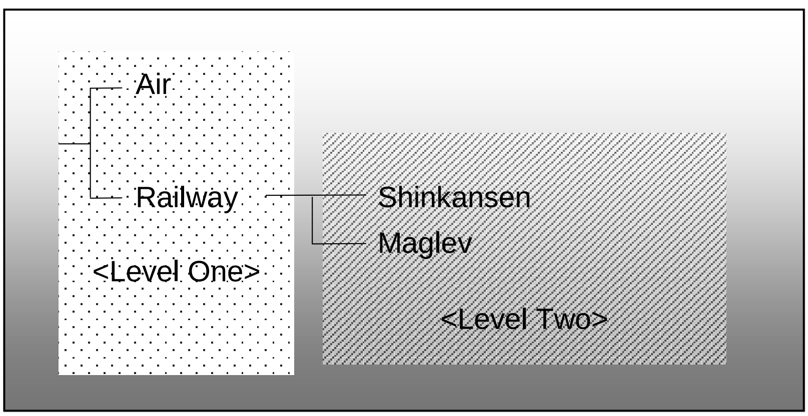
S_j: aggregate trip destination of zone j

q_{ij}: log sum value of travelling between zones i and j

4.4. Transport mode selection model

The transport mode selection model gives the modal split of the total trips between zones. We use a nested logit model. As depicted in Figure 15, the model is structured to provide two basic transport modes – “Air” and “Railway” – and a choice of “Shinkansen and other railway[5]” and “Maglev” for “Railway”.

Figure 15. **Transport mode selection model structure**



4.4.1 Level One

The probability of transport mode k being chosen for trips between zones ij is expressed in the form of an aggregate multi-nominal logit function. V_{ij} is the deterministic portion of the utility associated with mode k.

q_{ij}^{K} is the generalised price, composed of time factor and out-of-pocket costs. The value of time w is set exogeneously from past research (see Annex for details). In the case of the railway, the generalised price is the weighted average of Shinkansen and Maglev. θ_1, θ_2 are parameters to be estimated.

Following the utility function U_{ij}^{k} of travelling between zones i and j by transport mode k, composed of a deterministic portion $V_{ij}^{k} = \alpha + \beta p_{ij}^{k}$ and a random variable, assume that,

$$U_{ij}^{k} = \beta p_{ij}^{k} + \alpha + \varepsilon_{ij}^{k} \quad (7),$$

where $p_{ij}^{k} = M_{ij}^{k} + \theta T_{ij}^{k}$ is the generalised cost of travelling between zones i and j by transport mode k,

M_{ij}^{k} is the travel fare between zones i and j by transport mode k,

θT_{ij}^{k} is the product of θ, value of time, and T_{ij}^{k}, the time it takes to travel between zones i and j by transport mode k,

α is constant and β is a parameter, and

ε_{ij}^{k} is a random variable with Gumbel distribution.

Then, the probability of choosing mode travel by transport mode k between zones i and j could be expressed as follows:

$$P_{ij}^{k} = \frac{\exp\left(V_{ij}^{k}\right)}{\sum_{k=A,R} \exp\left(V_{ij}^{k}\right)} \quad (8)$$

Thus, when X_{ij} is the total travel demand between zones i and j, the demand function of transport mode k would be:

$$x_{ijk} = P_{ij}^{k} X_{ij} = \frac{\exp\left(V_{ij}^{k}\right)}{\sum_{k=A,R} \exp\left(V_{ij}^{k}\right)} X_{ij} \quad (9)$$

$$P_{ij}^{A} = \frac{\exp\left(V_{ij}^{Air}\right)}{\exp\left(V_{ij}^{Air}\right)+\exp\left(V_{ij}^{Rail}\right)} = \frac{\exp\left(\theta_1^S q_{ij}^A + \theta_2^S\right)}{\exp\left(\theta_1^S q_{ij}^A + \theta_2^S\right)+\exp\left(\theta_1^S q_{ij}^R\right)} \quad (10)$$

$$q_{ij}^R = P_{ij}^{Linear}\cdot q_{ij}^{Linear} + P_{ij}^{Shinkansen}\cdot q_{ij}^{Shinkansen}$$

$$q_{ij}^A = w\cdot t_{ij}^A + p_{ij}^A$$

P_{ij}^k: Probability of choosing transport mode k between zones i and j

(k = A; air transport, k=R; railway)

V_{ij}^k: Utility when choosing transport mode k between zones i and j

q_{ij}^R: Log-sum value of railway from the Level Two model

4.4.2 Level Two

$$P_{ij}^S = \frac{\exp\left(V_{ij}^S/\lambda\right)}{\exp\left(V_{ij}^S/\lambda\right)+\exp\left(V_{ij}^M/\lambda\right)} = \frac{\exp\left(\theta_1^S q_{ij}^S/\lambda\right)}{\exp\left(\theta_1^S q_{ij}^S/\lambda\right)+\exp\left(\theta_1^S q_{ij}^M/\lambda\right)} \quad (11)$$

$$q_{ij}^S = wt_{ij}^S + p_{ij}^S,\ q_{ij}^M = wt_{ij}^L + p_{ij}^M$$

P_{ij}^S, P_{ij}^M: Probability of choosing S (Shinkansen) or M (Maglev)

V_{ij}^K: Deterministic portion of utility when travelling by K ($K \in S, M$)

q_{ij}^R: Log sum of travelling by railway

λ: correlation factor between S and M

The nested logit model is used to reflect consumer preferences for Shinkansen and Maglev that are a closer substitute than air transport and railway in general. Thus, in the second stage of modal choice, λ is a parameter that gives the level of correlation between the two alternatives, Shinkansen and Maglev. The higher the λ the more the two choices are independent, and adding Maglev as an alternative is valued higher by tripmakers. Since we do not have observable data on the degree of independence between Shinkansen and Maglev, we shall use an exogenous value of 0.8 as λ.

4.5. Parameter estimation and exogenous values

Parameter estimation is conducted for the trip distribution model and modal split model. They are detailed in the Annex.

Price elasticity in the trip generator model is taken from past surveys. We use the following values. See Annex for a list of price elasticity values in past surveys.

Table 5. **Demand elasticity**

	Business	Sightseeing	Private
Price elasticity (β1)	0.7	1.5	1.5

Income elasticity in the trip generation model is also taken from past research. Income elasticity of 1.78 is used in the model based on Murakami *et al.* (2006). See Annex for a list of income elasticity values in past surveys.

4.6. Future setting of socio-economic factors and service characteristics of Maglev

4.6.1 Population and economic growth

Future estimates of population at city level are given by the National Institute of Population and Social Security Research. According to this estimate, the national population is expected to decrease from 127 million to 119 million; an approximately 6% decrease[6]. City level data aggregated to 207 zones indicate that while metropolitan areas such as Tokyo, Yokohama, Toyota (in the Nagoya region) and Amagasaki (in the Kansai region) increase their population, other areas suffer a decline.

As for economic growth, the current economic situation makes it difficult to specify robust economic prospects. Thus, we consider a number of scenarios with annual growth rates ranging from 0.5% to 3% in 0.5% intervals. The base year of the data set used in the model is 2005. The Maglev Super-Express inauguration year is set at 2025. A standard project duration of fifty years is used for the Maglev Super-Express so that the project is evaluated through the year 2075.

4.6.2 Service characteristics of Maglev

The following trip-time reduction and price increase between Tokyo-Nagoya and Tokyo-Osaka is used as a scenario for a future demand estimate.

Table 6. **Service characteristics of the Maglev Super-Express**

	Tokyo-Nagoya	Tokyo-Osaka
Time	40 minutes	60 minutes
Cost	1 000 yen increase	1 000 yen increase

Note: Twenty minutes are added at the transfer point when the Maglev Super-Express and other rail transport are used in a single journey.

4.6.3 OD zones that are affected by the introduction of Maglev

We need to assign OD zones that are affected by the introduction of Maglev. It is clear that OD pairs that are geographically irrelevant to the Tokyo-Nagoya-Osaka corridor need to be eliminated. Using NITAS, we identify OD pairs that currently take trips via Tokaido Shinkansen. Potential OD pairs that are currently not taking Tokaido Shinkansen but may choose Maglev once it is introduced are also included in the simulation.

4.6.4 Metropolitan zones

Three major metropolitan regions include the following prefectures. They comprise the metropolitan areas of Tokyo, Osaka and Nagoya, respectively.

Table 7. **Three metropolitan areas and prefectures**

	Tokyo Region	Hanshin Region	Chukyo Region
Prefecture	Tokyo-Kanagawa-Chiba-Saitama	Nara-Kyoto-Osaka-Hyogo	Aichi-Mie-Gifu

4.7. Result of the simulation

4.7.1 Impact of the Maglev Super-Express on modal split

Table 8 shows the estimated annual number of trips for the national total in 2025. Due to the decrease in population, benchmark figures without Maglev decrease by 2% compared to the 2005 population case. With the introduction of the Maglev Super-Express between Tokyo and Nagoya, the nation-wide modal split, for Shinkansen and Maglev combined, shifts from 75.6% to 76.1%. Table 9 depicts the estimated annual number of trips for the corridor between the Tokyo and Hanshin regions in 2025. There is a much larger impact in this corridor, the modal split for Shinkansen and Maglev combined changing from 78.6% to 81.4%. When the Maglev Super-Express connects Tokyo and Osaka via Nagoya, then 84.4% would be shared by Shinkansen and Maglev combined. Although introduction of the Maglev Super-Express does have a strong impact on air transport, more significant is the impact on Shinkansen. Indeed, more than half of Shinkansen trips will be taken away by Maglev in the corridor between the Tokyo and Hanshin regions.

Table 8. **Estimated annual number of trips (in millions) – national total in 2025**

	Air	Shinkansen	Maglev	Total
Without Maglev	84 (24.4%)	261 (75.6%)	- -	345
With Maglev Tokyo-Nagoya	83 (23.9%)	216 (62.6%)	46 (13.4%)	345
With Maglev Tokyo-Osaka	81 (23.4%)	200 (57.9%)	64 (18.6%)	346

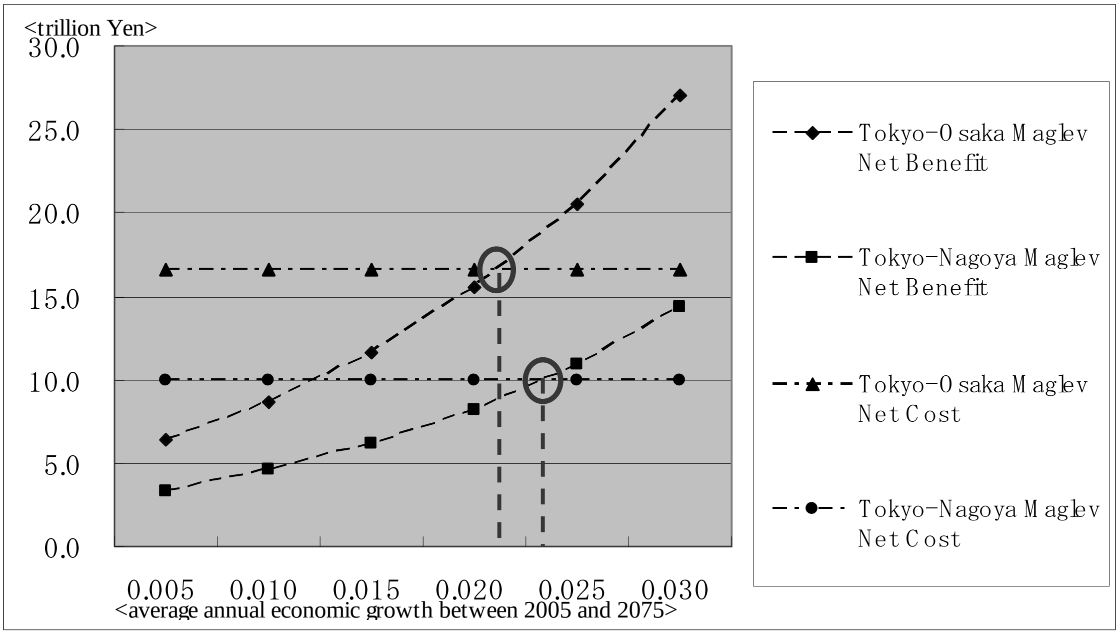
Table 9. **Estimated annual number of trips (in millions) – between Tokyo and Hanshin regions in 2025**

	Air	Shinkansen	Maglev	Total
Without Maglev	8 (21.4%)	31 (78.6%)	- -	39
With Maglev Tokyo=Nagoya	7 (18.6%)	13 (32.8%)	19 (48.7%)	40
With Maglev Tokyo=Osaka	6 (15.6%)	11 (26.4%)	24 (58.0%)	41

4.7.2 Benefits and costs of the Maglev Super-Express

The future benefits of introducing the Maglev Super-Express depend on the level of economic growth. We conducted a sensitivity analysis of net benefits with an annual growth rate ranging from 0.5% to 3% in 0.5% intervals. As for cost, we used data from a joint report by the Japan Railway Construction, Transport and Technology Agency (JRTT) and JR Central in July 2009, which revealed construction, maintenance and repair costs for the Tokyo-Nagoya Maglev Super-Express with a 50-year project duration[7]. It could be observed from Figure 16 that net benefit exceeds net cost when economic growth is above the 2.0% to 2.5% range. It should be noted that net benefit is calculated in comparison to the BAU case without any capacity constraint for Shinkansen or air transport. The net benefit will be greater if capacity constraint exists. With regard to annual economic growth, over 2% is a challenging target but not an inconceivable one. Future economic prospects, released by the Cabinet Office of Japan in January 2009, indicate a number of different GDP growth rate cases. Depending on the speed of recovery of the world economy, Japan is expected to grow at approximately 1.5% to 2% and above for the next decade. Demand growth from emerging economies such as China and India is promising. New opportunities in environmental business, nano-technology and robotics, among others, are expected to generate growth in the Japanese economy throughout the 21st century.

Figure 16. **Net benefit and cost of Maglev introduction (trillion yen)**



4.7.3 The impact of the Maglev Super-Express on CO_2 emissions

The environmentally friendly nature of Maglev technology should be noted. The CO_2 emission intensity of the Maglev Super-Express is one-third that of air transport. One of the expectations of introducing the Maglev Super-Express is its capability of mitigating CO_2 emissions from high-speed intercity transport. This, however, is not precisely the case. Because the Maglev Super-Express, with a CO_2 emission intensity five times higher than Shinkansen, would attract a considerable number of passengers, not only from air transport but also from Shinkansen, total CO_2 emissions from high-speed intercity transport would increase by 2.7% with the Maglev Super-Express between Tokyo-Nagoya and 4.9% between Tokyo-Osaka. If, however, the Shinkansen capacity constraint diverts considerable demand towards air transport, these estimates would need to be revised. We leave this question to future analysis. Also, there is a possibility that the increase in CO_2 from Shinkansen and Maglev could be mitigated by reducing the CO_2 content of the electric power supply. Due to the low utilisation of nuclear energy, the CO_2 content of electric power supplies in Japan is five times higher than that in France. There is potentially a large scope for substantial reductions in CO_2 emissions from this perspective.

5. CONCLUSION

In this paper we revisited the evolution of high-speed inter-city transport in Japan and conducted a simulation analysis of introducing the next-generation transport mode, the Maglev. In a unique market in which both high-speed railways, the Shinkansen and air transport developed simultaneously, modal choice based on price and speed has been manifested very clearly. So in assessing the impact of the Maglev Super-Express, planned to be introduced between Tokyo and Nagoya by 2025, we need to take into account the differences in price and speed characteristics of the existing and new transport modes.

From the simulation analysis, by a dynamic spatial nested logit model, we identify a significant opportunity for the Maglev Super-Express between Tokyo, Nagoya and Osaka. Accumulated social welfare and operational revenue, however, was found to exceed the net investment, maintenance and repair costs only when approximately 2%-3% annual economic growth is achieved for the next 65 years. If such economic conditions are realised, the total air transport market would also continue to grow, despite strong competition from the Shinkansen/Maglev system.

One other finding was Maglev's impact on CO_2 emissions. Maglev could not take advantage of its CO_2 emissions intensity being considerably lower than that of air transport. This is because Maglev attracts more passengers from Shinkansen, which has a five times lower CO_2 emissions intensity. An increase in total CO_2 emissions from electricity users, including the Maglev Super-Express, could be mitigated by the energy conversion sector's efforts to reduce the CO_2 content of electric power supplies through an increase in the utilisation ratio of nuclear energy, for instance.

More analysis is needed to unveil the full impact of high-speed inter-city transport improvements. In particular, we need to take capacity constraint into consideration. When economic growth triggers additional trips, capacity constraint in the existing Shinkansen network, for instance, may divert considerable demand to air transport. If this is the case, we need to alter the BAU case and reassess the net benefits and impact on CO_2 emissions. Furthermore, productivity gains, migration effects and national land-use efficiency are some of the themes that have not been covered by this paper. We look forward to further developments in such areas of research.

6. ACKNOWLEDGEMENTS

We express appreciation for research and support from the International Transport Policy Research Unit (ITPU), Graduate School of Public Policy, The University of Tokyo. We are also grateful to Kazuki Iwakami for his contributions to the data analyses and Tae Hoon Oum for his enthusiasm for intercity transport analysis.

NOTES

1. As of January 2001, the Ministry of Transport was integrated with the Ministry of Construction, etc., to form the Ministry of Land, Infrastructure, Transport and Tourism (MLIT).

2. 45/47 stands for 1970 and 1972 in Japan's Showa era.

3. In 1988, the name was changed to Japan Air Systems (JAS). In 2002 it was merged with JAL to form the current Japan Airlines Inc.

4. Apart from the two Class One airports, there are currently three others. New Tokyo International Airport, currently Narita International Airport, was constructed as a 100% government-owned agency, while Kansai International Airport, opened in 1994, and Chubu International Airport, opened in 2005, were PFIs.

5. Hereafter referred to as "Shinkansen".

6. Since there is no estimate for regional employees, we take the 2005 value as constant.

7. Tokyo-Osaka Maglev Super-Express costs were estimated by route length, since no official figures had been released as of July 2009. Both net benefit and net cost are present values at year 2025, depreciated by 4% per annum.

ANNEX

The estimation of parameters for trip distribution and the modal split model is conducted as follows.

1. Trip distribution model

1.1. Model to be estimated

The distribution model is in the following form. In order to derive the function to be estimated we give a benchmark destination J_i for every i. The relative probability of allocation of trips to destination j ($i \neq j$) *vis-à-vis* benchmark destination J_i, leaving out OD pairs without any trips, are pooled as samples.

$$\ln\left(\frac{P_{ij}^C}{P_{iJi}^C}\right) = V_{ij} - V_{iJi} = \theta_1^D q_{ij} + \theta_2^D \ln S_j - \theta_1^D q_{ij} - \theta_2^D \ln S_{Ji}$$

$$= \theta_1^D \left(q_{ij} - q_{iJi}\right) + \theta_2^D \left(\ln S_j - \ln S_{Ji}\right)$$

$$= \theta_1^D \left(q_{ij} - q_{iJi}\right) + \theta_2^D \left(\ln \frac{S_j}{S_{Ji}}\right)$$

Ji: a random benchmark destination from zone i ($j \neq Ji$)

S_j: total trip destination to zone j

q_{ij}: log sum of trip between zones i and j

The distribution model is estimated by the weighted least squares method.

$$Y = \left\{\ln\left(\frac{P_{ij}^C}{P_{iJi}^C}\right)\right\} = \theta_1^D \left(q_{ij} - q_{iJi}\right) + \theta_2^D \ln\left(\frac{S_j}{S_{Ji}}\right)$$

$$\sqrt{w_i} Y = \sqrt{w_i} \left\{\ln\left(\frac{P_{ij}^C}{P_{iJi}^C}\right)\right\} = \sqrt{w_i}\theta_1^D \left(q_{ij} - q_{iJi}\right) + \sqrt{w_i}\theta_2^D \ln\left(\frac{S_j}{S_{Ji}}\right)$$

$\sqrt{w_i}$: squared root of trip generation at zone i

1.2. Description of data

Table 10. **List of data**

	Item	Definition of data	Source of data
Zone data	Population	Population of the zone	National Population Census (2005, MHLW)
	Employment	Number of employees in the zone	National Population Census (2005, MHLW)
	Trip attraction factor	Aggregate number of destination trips to the zone	Inter-regional Travel Survey (2005, MLIT)
Inter-zone data	Number of O-D trips	O-D trip between zones by major transport modes and purpose of travel	Inter-regional Travel Survey (2005, MLIT)
	OD travel time	Time of travel between zones	NITAS：National Integrated Transport Analysis System (2008, MLIT)
	OD travel cost	Fares paid for travel between zones (including access and egress)	Survey of Air Passengers (2005, MLIT), JTB timetable (2005, JTB)

1.3. Result of the parameter estimation

The result of the parameter estimation is shown in Table 11. Parameters are statistically significant and R^2 is at an acceptable level. The parameter for generalised cost (θ_1^D) is negative, as we had expected.

Table 11. **Trip distribution parameter**

Trip distribution parameter	Business		Tourism		Private	
	Parameter	t-ratio	Parameter	t-ratio	Parameter	t-ratio
Generalized cost(θ_1^D)	-0.294	-97.688	-0.286	-59.157	-0.361	-89.392
Trip attraction(θ_2^D)	0.765	122.545	0.703	75.642	0.551	66.505
R^2	0.684		0.531		0.642	
Sample size	11 334		7 194		7 732	

2. Modal split model

2.1. Model to be estimated

The probability of selecting air transport *vs.* rail could be expressed in the following form.

$$\frac{P_{ij}^{\mathrm{A}}}{P_{ij}^{R}} = \frac{\dfrac{\exp(V_{ij}^{\mathrm{A}})}{\exp(V_{ij}^{\mathrm{A}}) + \exp(V_{ij}^{\mathrm{R}})}}{\dfrac{\exp(V_{ij}^{\mathrm{R}})}{\exp(V_{ij}^{\mathrm{A}}) + \exp(V_{ij}^{\mathrm{R}})}} = \frac{\exp(V_{ij}^{\mathrm{A}})}{\exp(V_{ij}^{\mathrm{R}})}$$

$$\ln\left(\frac{P_{ij}^{\mathrm{A}}}{P_{ij}^{\mathrm{R}}}\right) = \ln\left(\frac{P_{ij}^{\mathrm{A}}}{1-P_{ij}^{\mathrm{A}}}\right) = V_{ij}^{\mathrm{A}} - V_{ij}^{\mathrm{R}} = \theta_1^S (q_{ij}^{\mathrm{A}} - q_{ij}^{\mathrm{R}}) + \theta_2^S$$

A larger weight is placed for OD pairs with a high trip volume. We use the squared root of the total OD trips between zones ij (w_{ij}). θ_1^S should be negative since higher generalised costs reduce the incentive to choose that mode. Parameters θ_1^S, θ_2^S are estimated with the weighted least squares method.

$$\sqrt{w_{ij}}\ln\left(\frac{P_{ij}^{\mathrm{A}}}{1-P_{ij}^{\mathrm{A}}}\right) = \sqrt{w_{ij}}\theta_1^S (q_{ij}^{\mathrm{A}} - q_{ij}^{\mathrm{R}}) + \sqrt{w_{ij}}\theta_2^S$$

w_{ij}: total number of trips between zones i and j

$$q_{ij} = \frac{1}{\theta_1^S}\ln\left[\exp\left(\theta_1^S q_{ij}^{\mathrm{A}} + \theta_2^S\right) + \exp\left(\theta_1^S q_{ij}^{\mathrm{R}}\right)\right]$$

$$q_{ij}^{\mathrm{A}} = p_{ij}^{\mathrm{A}} + wt_{ij}^{\mathrm{A}}、 q_{ij}^{\mathrm{R}} = p_{ij}^{\mathrm{R}} + wt_{ij}^{\mathrm{R}}$$

q_{ij}^{A}: generalized cost of air transport, q_{ij}^{R}: generalized cost of railway

q_{ij}: expected generalized cost of travelling between zones i and j

w: value of time

2.2. Description of data

In addition to data used for estimating the trip distribution model, the following value of time factor from the existing literature is used to convert travel time into monetary value. This parameter is used by MLIT in its air transport demand model for airport planning in Japan and is estimated from disaggregate data on air transport passengers.

Table 12. **Value of time**

	Business	Sightseeing	Private
Time value in yen/hr	4 193	3 642	3 133
Time value in yen/min	69.88	60.70	52.22

2.3. Result of the parameter estimation

The result of the parameter estimation is listed in Table 13. Parameters are statistically significant. As we had expected, parameter θ_1^S is negative.

Table 13. **Modal split parameter**

Modal split parameter	Business		Tourism		Private	
	Parameter	t-ratio	Parameter	t-ratio	Parameter	t-ratio
Transport cost(θ1)	-1.433	-48.688	-0.846	-13.028	-1.113	-20.495
Constant(θ2)	-1.479	-27.462	-0.932	-11.259	-1.449	-24.511
R^2	0.699		0.303		0.487	
Sample size	1 670		588		955	

3. Price elasticity for trip generation model

Following is a list of major surveys of demand elasticity that were referenced.

Table 14. **Survey of demand elasticity**

		Leisure Travel	Business Travel
(i)	Air Passenger Travel (Cross-section)	1.52	1.15
	Intercity Rail Travel (Cross-section)	1.40	0.70
(ii)	Air Passenger Travel	1.10-2.70	0.40-1.60
	Intercity Rail Travel	1.40-1.60	0.60-0.70
(iii)	Air Passenger Travel (Short)	1.52	0.7

Sources:
(i) Oum, Waters and Yon (1992);
(ii) Oum, Waters and Yong (1990);
(iii) IATA and Inter VISTAS Consulting Inc. (2007).

4. Income elasticity for trip generation model

Following is a list of major surveys of income elasticity for the air transport market in Japan that were referenced.

Table 15. **Survey of income elasticity**

		Income elasticity
(i)	Ohashi *et al.* (2003)	1.50
(ii)	Yamaguchi (2005)	1.44
(iii)	Murakami *et al.* (2006)	1.78

BIBLIOGRAPHY

IATA and Inter VISTAS Consulting Inc. (2007), Estimating Air Travel Demand Elasticities, Final Report, International Air Transport Association (IATA).

Ohashi, Tadahiro, F. Takuma, K. Tsuchiya, K. Yamaguchi (2003), "Effects of Deregulation and Airport Development on Japanese Domestic Air-passenger Market", Applied Regional Science Conference (in Japanese).

Oum, Tae Hoon, W.G. Waters and Jong-Say Yong (1990), "A Survey of recent estimates of price elasticities of demand for transport", Working Paper No. 359, World Bank.

Oum, Tae Hoon, W.G.Waters and Jong-Say Yon (1992), "Concepts of Price Elasticities of Transport Demand and Recent Empirical Estimates, An Interpretive Survey", *Journal of Transport Economic and Policy*, Vol. 26, No. 2.

Yamaguchi, Katsuhiro (2005), "Policy Impact Analysis and Performance Management in Air Transport Policy", Japan Society of Transportation Economics (in Japanese).

Watanabe, Kenta, H. Nomiyama, T. Yamashita and K. Yamamoto (2009), "Estimate of externalities of Chuo Maglev Super-Express", Policy Research Paper GraSPP-P-09-001 and ITPU-P-09-001, Graduate School of Public Policy, The University of Tokyo (in Japanese).

INTERURBAN PASSENGER TRANSPORT: ECONOMIC ASSESSMENT OF MAJOR INFRASTRUCTURE PROJECTS

Ginés DE RUS

University of Las Palmas de Gran Canaria
University Carlos III de Madrid
Spain

SUMMARY

ACKNOWLEDGEMENTS

The author is indebted to Kurt Van Dender for useful comments and suggestions, to Jorge Valido for his work as research assistant and to Carlos Huesa and José Luis Pertierra for providing data and information on some intercity bus concessions affected by the introduction of high-speed rail. The use of the AENA database for air traffic is also gratefully acknowledged.

ABSTRACT

The future of interurban public transport will be significantly affected by public sector decisions concerning investment in infrastructure, particularly the construction of new high-speed rail lines in medium-distance corridors where cars, buses, airplanes and conventional trains are the competing modes of transport. The distribution of traffic between the alternative modes of transport depends on the generalized prices, which fundamentally consist of costs, time and government's pricing decisions. High-speed rail investment, financed by national governments and supranational institutions such as the European Union (EU), has drastically changed the previous equilibrium in the affected corridors. This paper discusses the economic rationale for allocating public money to the construction of high-speed rail infrastructure and how the present institutional design affects the selection of projects by national and regional governments, with deep long-term effects in these corridors and beyond.

Keywords: infrastructure, incentives, project evaluation, high-speed rail, intermodal competition.

1. INTRODUCTION

This paper addresses a crucial issue for the future of interurban passenger transport networks, i.e. the influence of public decisions on large infrastructure investments that will change the present equilibrium in intercity transport. It will focus mainly on the massive investment in high-speed rail (HSR) infrastructure that some national governments and supranational organisations, such as the European Commission, are helping to make through direct investment or by co-financing national projects under very favourable conditions.

The future of interurban transport is expected to be dominated by strict budget constraints and the introduction of efficiency-oriented policies affecting pricing and investment decisions, such as the application of polluter-pays and user-pays principles and the planning of infrastructure on a strict economic basis. The ultimate objective is to have an "integrated and sustainable transport system" that promotes economic growth and social cohesion (European Commission, 2009).

Investment in infrastructure requires significant public funds. The type of assets invested in transport infrastructure are essentially irreversible and subject to cost and demand uncertainty, so the optimal timing is a key economic issue, since the investment decision can be delayed in most cases (Dixit and Pindyck, 1994). These characteristics give a significant value to the option to invest, which is in the hands of governments that own the land or can expropriate it. In the case of intercity transport, most of the corridors are already in operation and investments in large projects, such as high-speed rail infrastructure, can be viewed as a change in the generalized cost of travelling (time and cost savings, reliability, comfort and safety, etc.) with respect to the situation prevailing *without* project (de Rus and Nash, 2007; de Rus, 2008).

Infrastructure and services do not follow the same long-term planning criteria. Private service operators, including car owners, decide how much and when to invest in new capacity, and this also includes technology. Private airlines decide which type of aircraft to buy depending on their demand expectations and business strategies. There is strong evidence that the competitive air transport industry works reasonably well (Morrison and Winston, 1995; 2005). This is also true of bus transport, at least under a concession regime (Nash, 1993; Mackie and Preston, 1996; Preston, 2004).

On the other hand, roads, airports, ports and railway tracks and stations ultimately belong to the public sector (with some exceptions), and although many crucial transport decisions are in the hands of private operators subject to market discipline, the public sector can heavily influence future modal split and the configuration of transport networks through investment, pricing and regulatory decisions affecting capacity.

This is the case with high-speed passenger trains operating largely within the public sector both in the areas of infrastructure and services. The construction of new lines in the European Union (EU), China's announcement that it intends to spend $162 billion to expand its railway system and the decision of the US government to include HSR passenger services as a centrepiece of national transport policy has given a new endorsement to this technology that may promote the expansion of railways in intercity transport.

From an economic perspective, the question is quite simple: is HSR socially worth it? And the obvious answer is: it depends. HSR is a rail technology that allows trains to travel faster than cars, buses and conventional trains, but more slowly than commercial aviation. Like any other technology, HSR is not inherently good or bad. Its social value resides in its ability to solve transport problems that are significant enough to justify its opportunity cost. Cost-benefit analysis can help answer this crucial question, but we do not need to go any further to maintain that the economic case for HSR investment depends on the prevailing conditions in the intercity corridor where the construction of the new line is planned, in particular the level of demand, the degree of congestion, value of time, expected time savings from diverted traffic, generated traffic and the net external effects.

The context in which the social appraisal of projects is carried out cannot be ignored in the economic analysis of major infrastructure projects. The institutional design is a key element for understanding public decision-making when different levels of governments are involved, as it is the case in the EU or generally when the national and regional governments of the same country do not necessarily share the same objectives, particularly with regard to where public investment should be made.

This paper addresses these long-term planning and assessment issues which affect the future of interurban transport. In Section 2, the long-term challenges in intercity transport are considered, by looking at the differences between the alternative modes of transport in medium-distance corridors where air, rail and road compete and where public investment decisions concerning infrastructure deeply affect market equilibrium. In Section 3 we discuss the conditions under which public investment in HSR infrastructure can be socially worthwhile. In Section 4, the incentives associated with national or supranational funding are considered, showing the relevance of the institutional design affecting the funding of large infrastructure projects. Finally, conclusions are drawn in Section 5.

2. LONG-TERM PLANNING FOR INTERURBAN PASSENGER TRANSPORT

Medium-distance intercity corridors (around 500 km) with road, air and rail transport in open competition have a modal split equilibrium that is very sensitive to small changes in the generalized prices of the alternative modes of transport. The differences between these modes of transport are quite obvious, but they have several things in common. On the supply side, they all need infrastructure to provide services combining vehicles, labour and energy under private or public ownership, and with infrastructure and operations vertically integrated or unbundled; and on the demand side, they all involve a transport service carrying passengers who have to pay different generalized prices in terms of money, time, quality and safety.

Air, maritime and road transport are vertically unbundled and different operators use a common infrastructure, sometimes with free access and sometimes with payment of an access fee (toll, price, tariff, etc.). Usually the operators are private and the infrastructure is public or privately operated under a concession contract. Road, air and maritime transport services are vertically separated from the infrastructure operator, and railways are unbundled in some cases and vertically integrated *de facto* in the case of high-speed trains operated by a single firm with the exclusive use of dedicated infrastructure. Buses and cars share the same roads, competing airlines share airports and high-speed rail is technically operated as a single business, even if, from an organizational standpoint, the maintenance and operation of the infrastructure are separated from service operations.

HSR has other advantages over airlines beyond vertical integration (with subsidized prices), reflecting some structural differences. Airports and airlines would still serve a large number of markets using the same airport capacity, and it is not clear that airport congestion management would be better with vertical integration. The HSR advantage in this case is that capacity is used to serve a very small number of markets (O-D pairs), and this makes it possible to reach very high levels of reliability.

These differences on the supply side have significant impacts on the demand side. The vertical integration of infrastructure and operation in the case of HSR is a significant advantage with respect to air transport in terms of the generalized costs of travel. HSR is more reliable than air transport, and access and waiting time much less cumbersome. Airport and airlines managers do not necessarily have the same objectives and, as a matter of fact, the generalized cost advantage of HSR lies outside the travel-time segment of the trip. In the case of roads, the differences are even clearer. Road infrastructure and operations are vertically separated. In contrast with the single operator of HSR, there are many users driving their own cars with free access (sometimes paying a toll) to a limited-capacity infrastructure. Road transport has the advantage of reducing access and waiting time to almost nothing and the cost disadvantage appears in the travel-time segment.

Investment in HSR changes the equilibrium in the interurban corridor through its impact on the generalized price of rail travel. Compared with conventional rail, HSR services barely affect access, egress and waiting time. The main impact is on travel time with a magnitude depending on the prevailing operating conditions of the conventional rail (one hour or more when conventional trains run at 100 km/h, but around half an hour when the operating speed is 160 km, over a distance of 450 km (Steer Davies Gleave, 2004). Road passengers travelling medium distances benefit from travel-time reductions but lose in terms of access, egress and waiting time. The comparison of the generalized costs of HSR and air shows a contrasting picture with respect to road. HSR is competitive over medium distances, but loses its competitive edge for long distances (Campos and Gagnepain, 2009).

Time savings are not the only consequence of HSR investment. The reduction in the generalized cost of travel generates new trips, and the diverting of traffic from other modes of transport may contribute to the reduction of congestion, accidents and environmental externalities. Unfortunately, the net impact on the alternative modes is not necessarily positive. The reduction of congestion is one effect on those who continue to use the previous mode of transport, but the reduction of operations in response to lower demand volumes affects negatively the adjustment to travel preferences of those users.

Before we discuss the benefits of HSR and the social value of channelling public funds to develop it, it is helpful to see the dominant trends concerning the future of interurban passenger transport. An "integrated and sustainable transport system" is the declared objective of most transport programs all over the world. It is far from evident what that objective means. It can include different actual transport policies with different degrees of public intervention, particularly with regard to investment decisions and pricing.

The development of a transport network is the result of the interplay of private and public decisions within a context of sometimes unpredictable changes in society and particularly in the economy. For long-term planning purposes, it is worth looking at the discussion of future trends in European transport by the Focus Groups for the European Commission in connection with the development of the White Paper on transport policy (European Commission, 2009) looking 40 years ahead. We are not interested here in some of the predictions, which are impossible to verify at present. Nevertheless, it is very informative to find out how they understand transport issues and what their public policy recommendation is insofar as this vision informs European transport policy[1].

The present context is one of tighter budget constraints, a situation that is going to worsen in the future given the present economic recession and growing public deficits. Increased ageing and the growing dependency rate, on one hand, and the need to devote more funds to repairing, upgrading and renewing existing infrastructure, on the other, will reduce the funds available for the transport sector and users will have to pay more than in the recent past, both for the internalization of externalities and cost recovery.

The following summarizes some of their positions on infrastructure and pricing policies:

- The importance of transport for economic development, the growth of transport demand and the need to maintain and upgrade existing capacity as well as constructing new capacity require direct charging for transport services. Both the user-pays and the polluter-pays principle will have to be translated into practical pricing decisions.

- Tighter budget constraints and the introduction of user charging will promote private participation. Private operators will assist in the construction and operation of transport infrastructure. The regulatory framework is crucial in order to provide the right incentives to get the best results from private participation.

- Planning of infrastructure plays a decisive role in ensuring coherent and uniform development at the European level. The construction of new infrastructure should be conditional upon the existence of real needs, as determined by the economic appraisal of projects.

– Infrastructure design should facilitate the use of environmentally-friendly energy resources and be integrated with land planning and transport solutions. Co-modality should be encouraged through a common and integrated ticketing system, common terminals and platforms, etc.

– Economic efficiency requires that prices reflect all costs. External costs have to be internalized and this has an impact in the short run by promoting an efficient use of existing infrastructure, and in the long run by providing long-term signals to investors that will gradually transform the transport system. Pricing is more effective in changing modal split than other policies.

– The efficient use of the network can be achieved through liberalization, which facilitates market entry and reduces administrative barriers. This would be especially helpful in the case of railways. Regulations to correct market failure should be designed to remove the considerable barriers to a level playing field in the transport sector (especially in the context of intermodal and international competition).

In a situation of intermodal competition with road, air and rail transport fighting for customers, it is useful to analyse how HSR investment responds to these long-term objectives.

3. HIGH-SPEED RAIL INVESTMENT AS AN IMPROVEMENT IN INTERURBAN TRANSPORT

In a given corridor, a HSR project has total infrastructure costs equal to *I* in the base year, and thanks to the supply of high-speed trains using this infrastructure, social benefits (net of annual maintenance and operating costs), denoted by *B*, are generated in the first year of operation. These net benefits grow annually at a rate of *g*. The infrastructure has a lifespan of *T* years and the discount rate is equal to *i*. Within this framework and assuming that *i* is greater that *g*, investment is socially worthwhile if the following condition is satisfied:

$$B\left[1-(1+g)^{T}(1+i)^{-T}\right](i-g)^{-1} > I \quad (1)$$

Two key values in expression (1) are the rates *g* and *i*. Expression (1) simplifies to (2) when the project lasts forever:

$$B(i-g)^{-1} > I \qquad (2)$$

Let us simplify and assume that condition (2) is satisfied or, alternatively, that the growth rate of net benefits is higher than the social discount rate $(g > i)$. In both cases the net present value is positive, though in the second case any positive value of *B* is compatible with a positive NPV. In practical terms, this last case translates into a very favourable case for HSR investment, as the net present value is positive even starting with a low demand volume.[2] In this case of exponential growth of net benefits, a positive net present value is not a sufficient condition to accept the project. The question "is HSR socially worthwhile?" cannot be answered without addressing the problem of optimal timing.

Even disregarding the additional benefits of relevant information which reveals when the investment is postponed, we have to address the question of optimal timing. Unless the benefits of the first year are greater than the opportunity cost of the investment, it is better to delay the investment decision even if the net present value is positive. Ignoring for simplicity the net benefit in year T+1, it is socially worthwhile to invest in HSR when:[3]

$$B > Ii \quad (3)$$

For a given social discount rate condition (3) is satisfied if the first year's net social benefits of introducing HSR in a corridor offset the opportunity cost of allocating I to this project rather than to other social needs. In expression (3), B accounts for the net social benefits in the first year, and this basically includes the time savings obtained by diverting traffic to the new mode, the benefits of generated traffic, the increase in quality and safety, the reduction in congestion, accidents and other negative externalities in alternative modes of transport, the release of additional capacity for other kinds of traffic (e.g. rail freight and long distance in airports) and the change in the operating and maintenance costs of moving the volume of passengers in the corridor because of the project (excluding the investment costs).

These net benefits in the first year depend heavily on the specific characteristics of the corridor and how the new line affects the generalized cost of travel. In order to offset the investment opportunity costs, a significant volume of demand is required in the corridor to offset the high cost of the investment. The cost of constructing one kilometre of HSR infrastructure ranges from 12 to 40 million euros, with an average of 18 euros, and these values do not include planning and land costs and main stations. The costs are quite sensitive to the terrain characteristics and the need to cross high-density urban areas (Campos and de Rus, 2009).

The benefits in the first year of operation are very sensitive to the ability of high-speed trains to divert traffic from highly congested modes of transport. The introduction of a HSR line in a 500 km corridor with an uncongested road and good air transport connexions is hard to justify unless several conditions are met: a high volume of demand shifting from the other modes of transport, a significant reduction in total trip time, the generation of new demand, the reduction of negative externalities and a high willingness to pay for these benefits.

The expected time saving (and its composition) obtained with a HSR project is very sensitive to the original transport mode in which passengers were travelling previously. A passenger shifting from road to HSR saves travel time but increases access, egress and waiting time. On the other hand, a passenger shifting from air transport to HSR increases his travel time and saves access, egress and waiting time.[4] The passenger shifts if the HSR generalised price is lower than in the original mode, and this can happen, even if the total trip time increases, when the HSR fare is low enough to offset the longer trip (de Rus, 2008).

The existence of network externalities is another alleged direct benefit of HSR (see Adler *et al.*, 2007). Undoubtedly, a dense HSR network offers more possibilities to rail travellers than a less developed one. Nevertheless, we are sceptical of the economic significance of this effect. We do not argue against the idea that networks are more valuable than disjointed links. The point is that when there are network effects it should be included in the benefits at a route level already discussed. Although rail passengers gain when the wider origin-destination menu is in a denser network, the utility of a specific traveller who is travelling from A to B does not increase with the number of passengers unless the frequency increases, and this effect (a sort of Mohring effect) is captured at a line level.

Time savings come from diverted traffic. Generated traffic increases total travel time but produces benefits insofar as the passengers are willing to pay the generalized cost of travel. Diverted traffic has other intermodal effects beyond the ones already described. These effects are the indirect effects of HSR on passengers who continue to use their original transport mode.

Indirect effects are the impact of HSR on secondary markets, whose products are complements to or substitutes for the primary market. For simplicity's sake, we are focusing on the alternative modes of transport affected by the introduction of HSR. Are users of the alternative modes better off with HSR? What about the producers? It is important to distinguish here between transfers and real resource changes. We have already seen the direct benefits that society gains from the introduction of HSR, but users who remain attached to their former modes of transport may be affected positively or negatively depending whether there are distortions on these modes of transport. The same is applicable to other economic agents.

The critical issue is whether price is higher or lower than marginal social cost in the alternative mode of transport. When the price is below the marginal cost in the original transport mode, the diversion of traffic to the new transport mode benefits society.[5] This could happen because suboptimal congestion, or pollution, is reduced. However, the opposite might occur, and the indirect effect could be negative when the price is above the marginal cost, for example, if the reduction of demand in the original transport mode forces the operators to reduce the level of service, thereby increasing the generalized cost of travel.

The key point is whether the original transport mode was optimally priced. Although it has been argued that the reduction of road and airport congestion is a positive effect of HSR, this is only the case if there is a lack of optimal pricing. When road and airport congestion charges internalise the external marginal costs, there are no indirect benefits from the change in modal split. This can be viewed from another perspective. The justification of HSR investment based on indirect intermodal effects should be first compared with a "do something" approach, consisting of the introduction of optimal pricing (user and polluter-pays principles).

It should also be mentioned that, given for example the impossibility of road pricing, a second-best case for HSR investment, based on indirect intermodal effects, requires significant effects of diverted traffic on the pre-existing traffic conditions in the corridor. This means the combination of significant distortion, high demand volume in the corridor and sufficiently high cross-elasticity of demand in the alternative mode with respect to the change in the generalised cost.

The assumption that the price is equal to the social marginal cost means that the loss of traffic by conventional modes of transport does not affect the utility of those who continue to use these modes of transport, nor the welfare of producers or workers in these modes. This would mean that operators are indifferent to a 50 per cent loss in patronage, or workers to losing their jobs, because in both cases they are receiving the exact amount of their opportunity costs. There are many reasons to abandon this assumption, one of which is the existence of unemployment, but we will concentrate here on how the reduction of demand in air and bus transport affects user's utility when the operators respond to lower demand by reducing the service level.

Figures 1 to 10 and Tables 1 to 4 (see Annex) show how the introduction of HSR in some corridors reduced demand for airlines and bus operators and how the airline industry responded by adjusting the supply to the external shock in demand. There is a remarkable difference between the effects of the reduction of service in both modes of transport. Bus operators cannot change their basic regulated timetables because they operate under a concession contract. Although they cut the level of service when demand diminishes, the reduction in supply does not affect frequencies since the suppressed services leave at the same time as approved in the basic regulated timetable. However, it can be argued that although users are barely affected by the short-term adjustment of bus operators, financial difficulties will emerge later in contract renegotiations or when concessions expire. This means that users and/or taxpayers (or workers) will have to pay for the adjustment in the medium-term.[6]

Airlines operate in open competition so the short-term adjustment to the external shock in demand produced by the introduction of HSR services is a reduction in the number of operations. This affects frequencies, firstly because the reduction in demand is substantially higher; secondly, because airlines are not subject to public service obligations and so the adjustment is legally feasible; and thirdly, because of the nature of flight operations (slots required for take-off and landing), frequencies are necessarily affected when services are cut. The reduction in the number of flights per hour increases total travel time when passengers arrive randomly, or decreases utility when they choose their flight in advance within a less attractive timetable.

Finally, it should be stressed that intermodal competition is based on the generalized price of travel. Modal choice may be affected by the competitive advantage of each mode of transport, but the comparative advantage can reflect two completely different facts in this case. It may, for example, reflect a technological advantage with respect to the trip length, but it may also be explained by the charging policy in use. The impact on market share in medium distance-corridors may be substantial depending on whether the government charges variable costs or aims for full cost recovery, or something in between, depending on the severity of budget constraints.[7]

The final equilibrium in medium-distance (or even in short-distance) corridors will not only be the result of the free interaction of supply and demand. Governments will have a strong influence on the final modal split because the construction of public infrastructure is critical in transport, and particularly in the case of HSR. Once the HSR infrastructure is built the short-run marginal cost is considerably lower than the average cost (see Campos and de Rus, 2009, Campos *et al*., 2009) and the crucial question is whether society is willing to pay the total costs (including capacity) of a new mode of transport in the light of the actual travelling conditions in a particular corridor and the alternatives available for improving the present situation.

4. FUNDING OF TRANSPORT INFRASTRUCTURE AND ITS EFFECTS ON PROJECT SELECTION

The construction of a high-speed rail network is an expensive task. It is an investment that has the following characteristics: it is large-scale, irreversible and costly. The decision to invest public funds in the construction of HSR lines is subject to cost, and especially, demand uncertainty. The irreversible nature of the decision and the profound impact on equilibrium in the corridor where the new project is to be built makes the economic appraisal of the project quite relevant. It is therefore judicious to examine how institutional design affects the final choice in the allocation of public money in interurban corridors.

National and supranational governments are supporting the implementation of this new rail transport technology with public funds. To understand the impact of this public support on the investment decision, it is useful to distinguish two levels in the process of funding major infrastructure projects. The first relates to the institutional design, in which supranational and national governments (or national and regional governments) agree on the projects to be financed. The second is related to the selection of contracts for the construction and operation of the infrastructure. This level includes the relationship between the national (or regional) government benefitting from the project and the operator(s) responsible for the construction and operation of the project.[8]

The co-financing system in the EU is the so-called "funding-gap" method consisting of a type of cost-plus financing mechanism in which the difference between the investment costs and the discounted revenues (net of operating costs) of the project are partially covered by the supranational organisation. The European Commission finances a percentage (the co-funding rate) of this financial gap. The incentive embedded in this mechanism is perverse, since the subsidy increases with total investment costs and decreases with net revenues. This financing mechanism penalizes the internalization of externalities and congestion, leads to excessive demand and biases the capacity size and the choice of technology.

Let us suppose that a country facing a problem of capacity in its transport network is considering mutually exclusive projects, including the construction of a new HSR line that can apply for financial support from a supranational agency. The country is governed by a politician, who must decide upon the main characteristics of the project (let us say HSR or upgraded conventional train), make a cost-benefit analysis and then present these elements to the supranational planner in order to obtain the funds for construction of the infrastructure.

The effects of the present system of co-financing in the EU, or any other system in which a national government pays for the infrastructure in the national budget and the regional government decides which type of project is to be financed, can be modeled as follows (de Rus and Socorro, 2009). Assume that there are only two periods. During the first period, the new rail infrastructure is constructed. During the second period, the citizens of the country use it. The real construction costs are paid by the national government. We know that actual costs do not necessarily coincide with the minimum investment cost. To minimize construction costs requires an effort on the part of the politician, which has a cost for him.

It is not uncommon for national governments to be better informed than the supranational agency about the transport problem and the set of alternatives available and therefore about the minimum investment cost required to solve the problem. For this reason, we assume that the supranational planner cannot observe (or verify) either the minimum investment cost, or the effort exerted by the politician in order to be efficient. Moreover, the national government has to decide on the price to be charged for the use of the new infrastructure and consequently the number of users. There are also operating and maintenance costs, which are privately known, and in many cases there are different technologies and/or capacity sizes with significant cost differences.[9]

Once we abandon the idea of a benevolent supranational planner with perfect information and assume that the utility function of the politician depends on his own private income (only obtained if the politician is governing the country), we can explain more fully some of the evidence concerning the national government's decisions on expensive infrastructure.[10] The higher the welfare of voters in the second period, the higher the probability of re-election. The welfare of voters in the second period is the sum of their consumer surplus and the value of social expenditures.

The fixed costs/total cost ratio in HSR projects can be 50 per cent or higher (Campos *et al.*, 2009), so these projects are always candidates for supranational funding. In a world of perfect information, the supranational agency would maximize social welfare by forcing the national government to exert the maximum level of effort, thereby minimizing project costs and introducing marginal social cost pricing. In the real world, efforts and marginal costs are not observable and the behaviour of the national government will respond to the incentives of the financing mechanism.

With the present funding gap mechanism (as with any other cost-plus financing system), it is costly to be efficient. Governments have no incentive to minimize investment costs or to introduce optimal pricing. There is a bias in favour of expensive, latest technology mega-projects and pricing will depart from user-pays or polluter-pays principles, since the higher the price for the use of the new national infrastructure, the lower the consumer surplus of voters will be, and the lower the probability of re-election. Consequently, the politician will choose maximum number of users and will not charge for the external costs.

The evidence supports these conclusions. It is remarkable that member countries have promoted the construction of some HSR lines when the demand was too low to pass a strict cost-benefit analysis as well as other transport infrastructure such as roads or ports. An *ex post* evaluation of a sample of projects co-financed by the Cohesion Fund in the period 1993-2002 concludes that national governments have been focusing primarily on timely commitment of the available funding, paying less attention to the technical content and economic priority of projects (ECORYS Transport, 2005). The evaluations generally fail to assess the quantitative contribution of the project to the declared objectives. Problem descriptions and analyses are sometimes lacking.

Moreover, it was generally impossible to determine whether projects were technically sound, and this deficiency led to problems such as improper designs; technical changes after the project was approved but before construction was started; late changes to design/tender dossiers; late beginning of implementation; cost overruns due to additional activities for the contractor, who was then in a good position to claim additional costs; longer implementation periods than foreseen; and too many requests for extension of the implementation period. The document concludes that "the evaluators have found only pragmatic criteria for the co-financing rate. In addition some basic dilemmas exist between general policy objectives and the rules applied for calculation of the co-financing rate. In particular the polluter-pays principle is only partially adopted since increasing user charges is discouraged by the present system of determining the co-financing rate" (ECORYS Transport, 2005).

These disappointing results are not completely unexpected. As we have already discussed, national governments are in general better informed than supranational planners about the costs and benefits of the infrastructure projects to be constructed in their own regions, and they do not necessarily share the same objectives. Governments may have incentives to manipulate project evaluation in order to obtain more funds from the supranational planner. In a context of asymmetric information and different objectives, the relationship between national governments and supranational planners cannot be modelled in a conventional cost-benefit analysis framework.

The existence of information asymmetries and conflicting interests requires a different approach in which incentives are explicitly accounted for. Florio (2006) proposes to move away from the current low-powered incentive EU co-financing mechanism, essentially a partial reimbursement of investment cost scheme, towards a more incentive-based system.

As argued in de Rus and Socorro (2009), a fixed-price financing mechanism may provide the necessary incentives to reduce costs and charge the socially optimal price. Moreover, with the funding-gap method, cost-benefit analysis is simply a bureaucratic requirement to enable national governments to obtain supranational funds. However, with the fixed-price financing mechanism, cost-benefit analysis is a very useful tool for governments to allocate the supranational funds in the most efficient way.

The fixed-price mechanism, in this context, is an *ex ante* fixed quantity of external funding unrelated to costs and revenue. The idea of the fixed-quantity financing mechanism is to make national governments responsible for insufficient revenues and cost inefficiencies, since they receive a fixed amount of funding and are the residual claimants for effort. The incentive to introduce optimal pricing is now high as the costs of inefficient pricing are also suffered by the politician.

It is worth stressing that by giving national governments an *ex ante* fixed amount of funds, the European Commission loses its influence on the selection of projects. This is not the position of the European Commission, which establishes infrastructure investment priorities for the member countries. An intermediate solution is to replace the funding-gap method with an alternative financing scheme based on *ex ante* fixed-quantity funding linked to generic objectives such as investing in “accessibility” or “minimizing the total social cost of transport” in selected corridors, a mechanism that should be dissociated in any case from costs and revenues and the selection of any specific technology. The risk of building socially unprofitable HSR lines would be dissociated from the co-financing mechanism, since the selection of the most expensive (and perhaps inappropriate) project will now have a completely different opportunity cost for national governments.

5. CONCLUSIONS

The future of interurban transport will be determined by the interaction of consumer preferences, technological developments and the availability of resources to meet mobility needs. Competition between firms and modes of transport, subject to the minimum regulation required both to internalize externalities and guarantee a basic level of accessibility, will shape transport networks in the years to come. However, public intervention is not confined to price regulation or equity issues in transport. Public infrastructure construction can exert a remarkable influence on the future form of interurban transport corridors.

The high-speed rail investment decisions taken and the subsequent infrastructure pricing policies set by the public sector have a profound impact on the allocation of resources in the transport sector and the rest of the economy. It seems obvious that high-speed rail infrastructure is an appropriate option for some corridors but a very expensive one in low-traffic areas where the alternative modes of transport can satisfy demand at much lower cost.

The challenge is to design an institutional framework that helps to find the best options for society, beyond the special interests of industry groups and politicians. To reinforce the use of cost-benefit analysis as a requirement for approving new infrastructure is clearly insufficient. Because of asymmetries of information and conflicting interests, there is a need for a new incentive mechanism that will help overcome the current situation in which the member country-supranational government relationship (or that of regional and national governments) creates a bias in favour of the most expensive and modern technology over more efficient and less expensive solutions, new construction over maintenance and upgrading, and free access over the introduction of efficient pricing based on the polluter-pays and user-pays principle.

NOTES

1. Transport policy priorities have changed over the past decades. In the 1960s, the emphasis was on network and capacity expansion; from the 1970s onward, efficiency was more important than new construction; from the 1980s onward, the negative externalities of transport emerged strongly; in the 1990s, the focus was on the potential of new technologies for network improvement (Vreeker and Nijkamp, 2005).

2. It is not unusual for the construction of HSR in low demand corridors to be defended on the basis of optimistic traffic projections.

3. With a NPV>0 and in the case of “accept-reject”.

4. An explanation of the time (and quality) advantage of HSR over air transport is contingent on differences in security procedures, and it should not be taken for granted that these differences will remain as they are.

5. We assume that the price is equal to, or greater than, the marginal cost in the new transport mode.

6. This argument can be extended to conventional rail services negatively affected by the introduction of HSR.

7. We shall not discuss here which type of pricing criteria should be followed. For a discussion of the justification of short-run *vs.* long-run marginal cost pricing in transport, see Rothengatter (2003) and Nash (2003). The effects on HSR prices when infrastructure investment costs are included in prices can be seen in de Rus (2008).

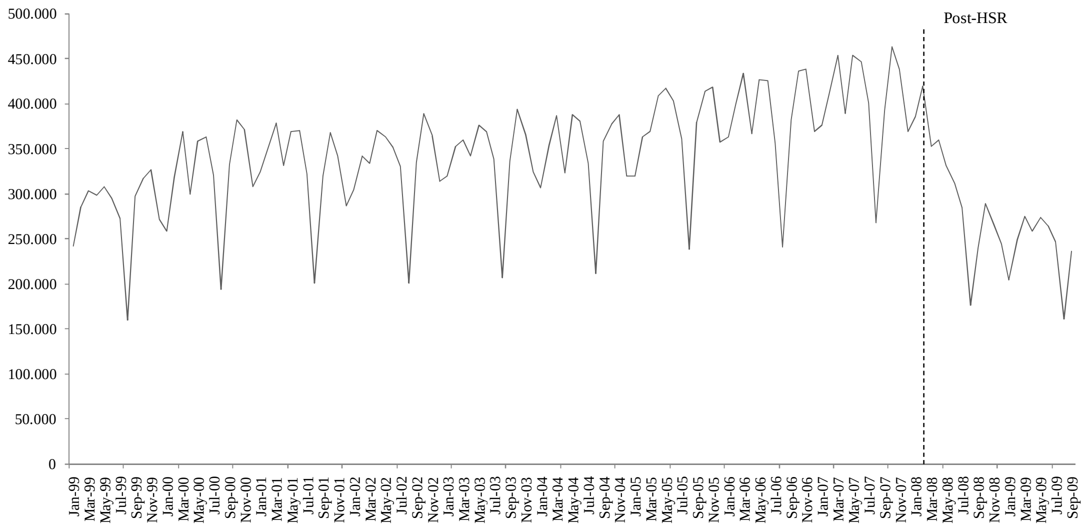
8. This second level has been widely analysed in the economic literature (Laffont and Tirole, 1993; Bajari and Tadelis, 2001; Guasch, 2004; Olsen and Osmundsen, 2005).

9. Cost overruns are common in large infrastructure projects and it has been shown that the deviation is not only explained by unforeseen events (Flyvbjerg *et al.,* 2003).

10. The implementation of the user-pays and the polluter-pays principles and the reduction of public expenditure have significant political costs (Sobel, 1998). Downs (1957), Niskanen (1971) and Becker (1983), have often assumed that legislators attempt to maximize electoral support. Even if re-election may not be the primary factor motivating their legislative behaviour, it is still true that legislators react in predictable ways to the electoral costs and benefits of their choices. Thus, legislators will favour actions that increase the probability of their being re-elected over decisions that lower it (Sobel, 1998; Robinson and Torvik, 2005).

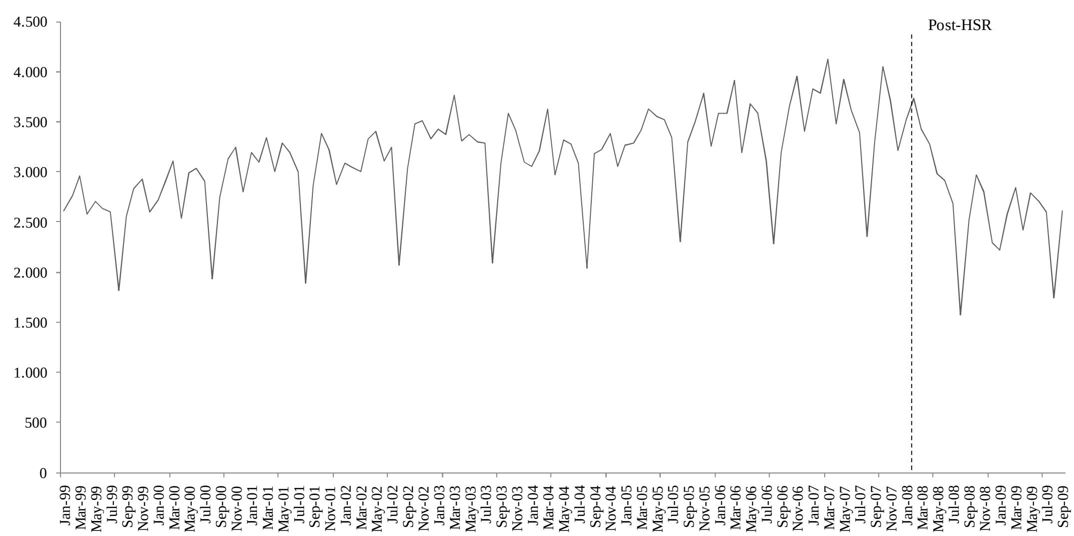
ANNEX

Figure 1. **Madrid-Barcelona air passenger-trips per month**



Source: built from data in www.aena.es

Figure 2. **Madrid-Barcelona commercial flights per month**



Source: built from data in www.aena.es

Table 1. **Madrid-Barcelona (passengers-trips)**

Variable	Coefficient	Std. Error	t-Statistic
T	1064	65**	16.26046
D1	-2914	9522	-0.306032
D2	36026	9520**	3.784228
D3	64210	9543**	6.728295
D4	38762	9540**	4.063172
D5	66535	9537**	6.976728
D6	57925	9534**	6.075548
D7	19253	9532*	2.019896
D8	-100908	9530**	-10.58831
D9	20683	9529*	2.170561
D10	68821	9742**	7.064316
D11	56634	9741**	5.813673
HSR	-150368	6914**	-21.74971
C	260965	7932**	32.90185

T: month; D1-D11: monthly dummies; HSR: dummy for High Speed Rail.

R-squared: 0.901392; Adjusted R-squared: 0.890245; Durbin-Watson stat: 1.032179;
*,** significant at the 5 or 1 per cent level.

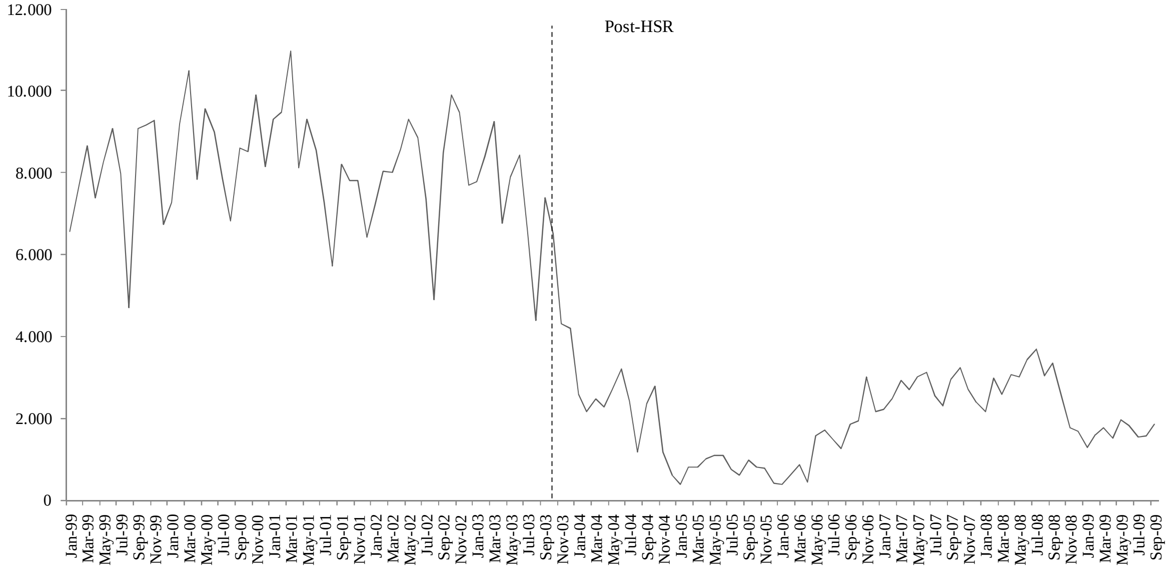
Table 2. **Madrid-Barcelona (commercial flights)**

Variable	Coefficient	Std. Error	t-Statistic
T	8	0.55**	15.06314
D1	177	80*	2.212122
D2	244	80**	3.042595
D3	526	80**	6.557876
D4	174	80*	2.173615
D5	375	80**	4.673040
D6	263	80**	3.287075
D7	106	80	1.321722
D8	-917	80**	-11.43899
D9	11	80	0.131378
D10	406	82**	4.954149
D11	412	82**	5.027874
HSR	-1037	58**	-17.84494
C	2550	67**	38.23676

T: month; D1-D11: monthly dummies; HSR: dummy for High Speed Rail.

R-squared: 0.880377; Adjusted R-squared: 0.866855; Durbin-Watson stat: 1.100381; *,** significant at the 5 or 1 per cent level.

Figure 3. **Madrid-Zaragoza air passenger-trips per month**



Source: built from data in www.aena.es .

Figure 4. **Madrid-Zaragoza commercial flights per month**

Source: built from data in www.aena.es

Table 3. **Madrid-Zaragoza (passenger-trips)**

Variable	Coefficient	Std. Error	t-Statistic
T	-1	5	-0.232599
D1	-87	475	-0.183369
D2	482	475	1.015930
D3	970	475*	2.043228
D4	139	475	0.293170
D5	872	475	1.837054
D6	928	475	1.954525
D7	129	475	0.270871
D8	-1049	475*	-2.206355
D9	644	476	1.354009
D10	1269	486*	2.612176
D11	972	486*	2.002266
HSR	-5973	381**	-15.67043
C	7706	388**	19.86720

T: month; D1-D11: monthly dummies; HSR: dummy for High Speed Rail.

R-squared: 0.899041; Adjusted R-squared 0.887629; Durbin-Watson stat: 0.843296; *,** significant at the 5 or 1 per cent level.

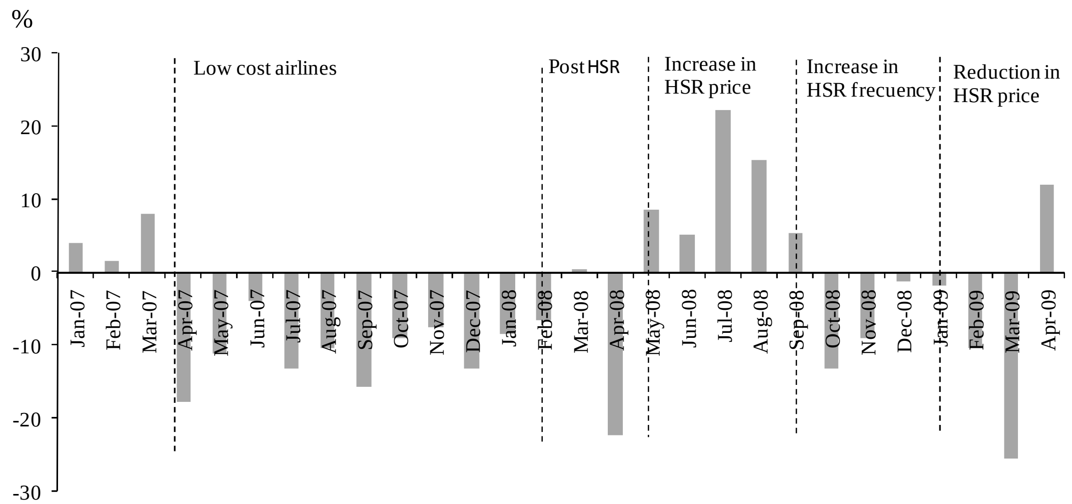
Table 4. **Madrid-Zaragoza (commercial flights)**

Variable	Coefficient	Std. Error	t-Statistic
T	0.098	0.16	0.598539
D1	9	15	0.597563
D2	5	15	0.341518
D3	16	15	1.054459
D4	8	15	0.518867
D5	17	15	1.142293
D6	14	15	0.885964
D7	9	15	0.605989
D8	-25	15	-1.638183
D9	11	15	0.687313
D10	32	16*	2.069965
D11	24	16	1.546399
HSR	-157	12**	-12.75281
C	228	12**	18.27449

T: month; D1-D11: monthly dummies; HSR: dummy for High Speed Rail.

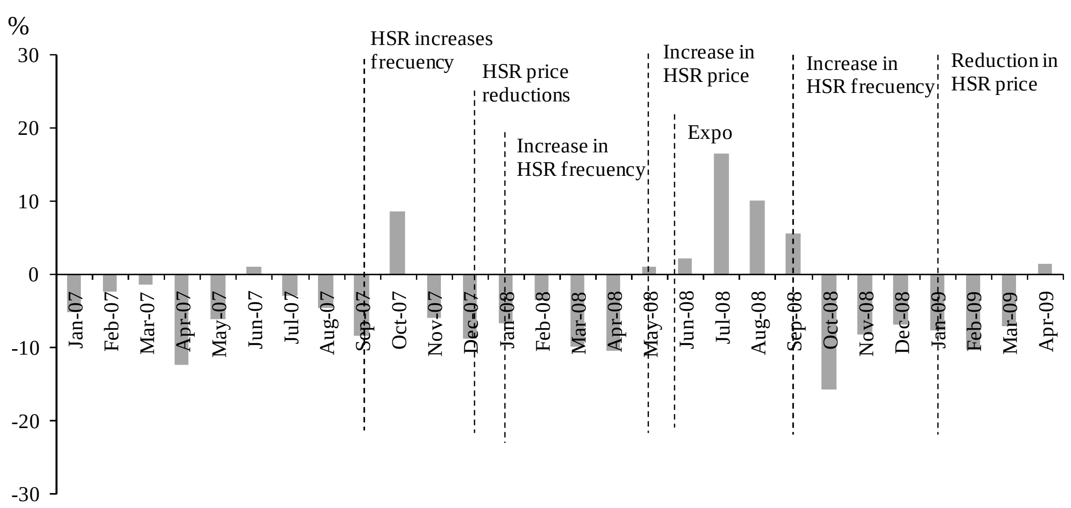
R-squared: 0.839659; Adjusted R-squared: 0.821534; Durbin-Watson stat: 0.626277; *,** significant at the 5 or 1 per cent level.

Figure 5. **Madrid-Barcelona (scheduled bus services)**
Changes in demand per month (base year: 2006)



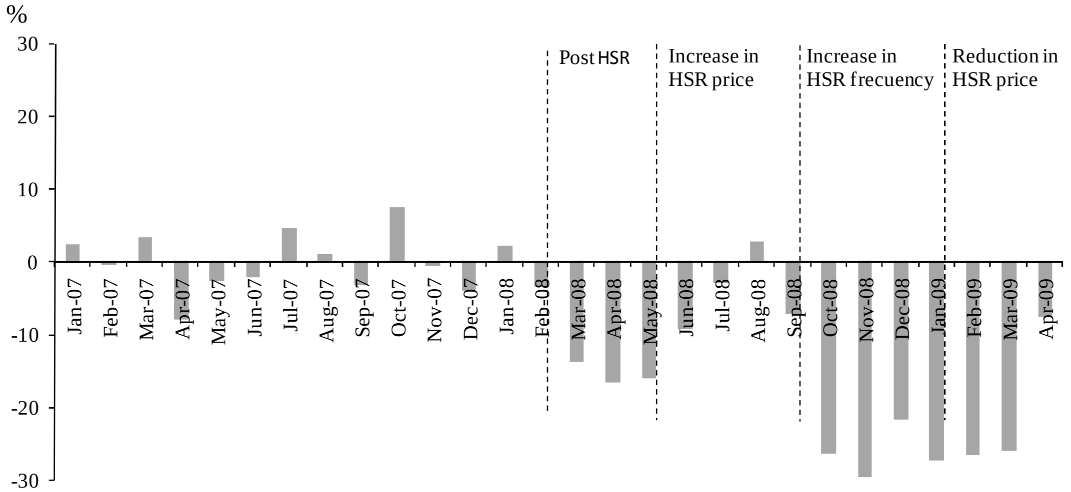
Source: built from data provided by FENEBUS.

Figure 6. **Madrid-Zaragoza (scheduled bus services)**
Changes in demand per month (base year: 2006)



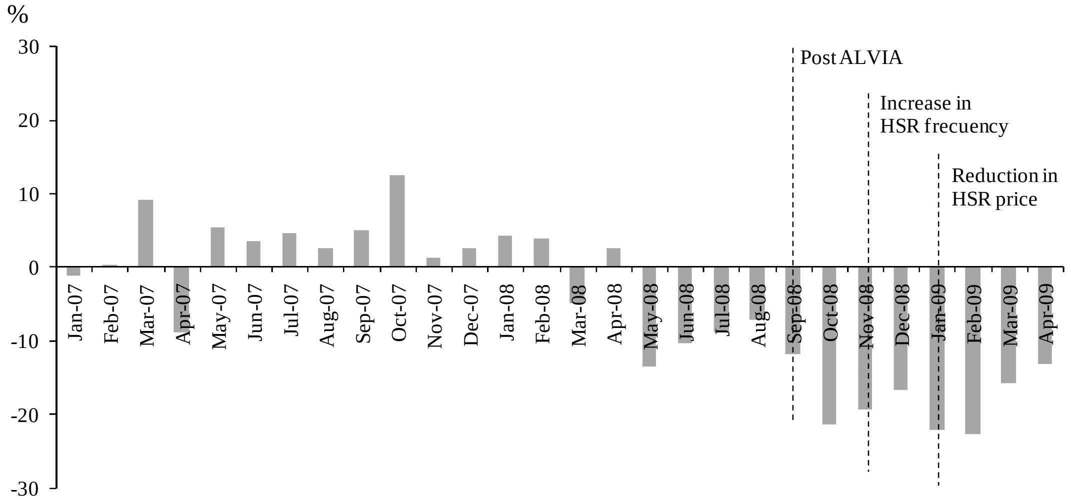
Source: built from data provided by FENEBUS.

Figure 7. **Zaragoza-Barcelona (scheduled bus services)**
Changes in demand per month (base year: 2006)



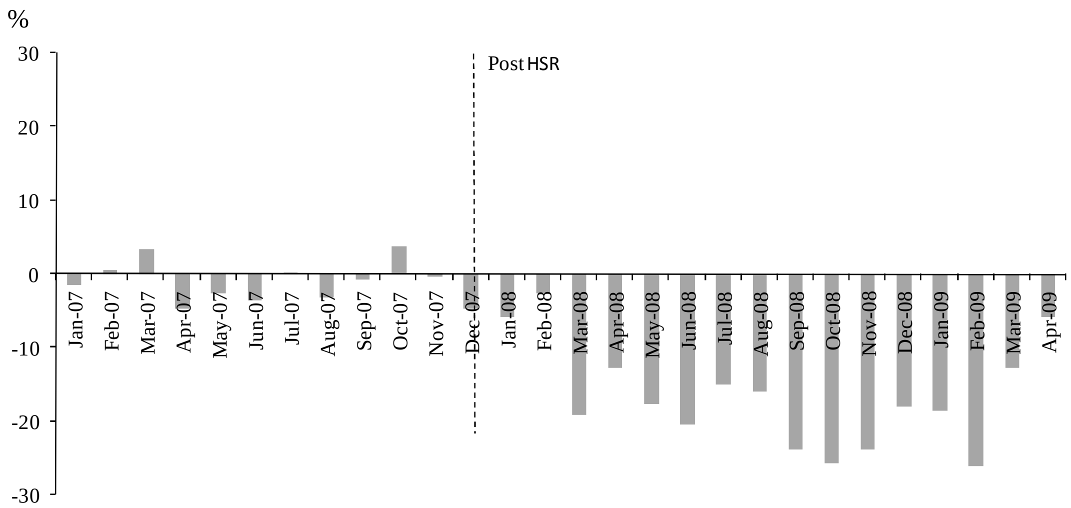
Source: built from data provided by FENEBUS.

Figure 8. **Madrid-León (scheduled bus services)**
Changes in demand per month (base year: 2006)



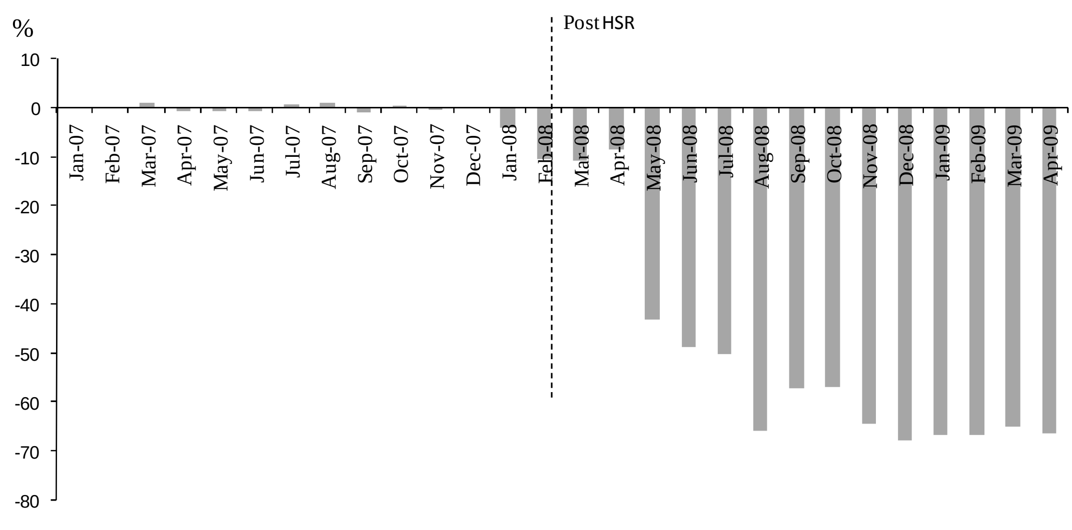
Source: built from data provided by FENEBUS.

Figure 9. **Madrid-Valladolid (scheduled bus services)**
Changes in demand per month (base year: 2006)



Source: built from data provided by FENEBUS.

Figure 10. **Lleida-Barcelona (scheduled bus services)**
Changes in demand per month (base year: 2006)



Source: built from data provided by FENEBUS.

BIBLIOGRAPHY

Adler, N., C. Nash and E. Pels (2007), “Infrastructure pricing: The case of airline and high-speed rail competition”, Paper presented at the 11th World Conference on Transport Research, Berkeley.

Bajari, P. and S. Tadelis (2001), “Incentives versus transaction costs: a theory of procurement contracts”, *RAND Journal of Economics*, 32 (3), pp. 387-407.

Becker, G.S. (1983), “A theory of competition among pressure groups for political influence”, The Quarterly Journal of Economics, 98 (3), pp. 371-400.

Campos, J. and P. Gagnepain (2009), “Measuring the intermodal effects of high-speed rail”, in: G. de Rus (ed.), *Economic Analysis of High-Speed Rail in Europe*, BBVA Foundation.

Campos, J. and G. de Rus (2009), “Some stylized facts about high-speed rail: A review of HSR experiences around the world”, *Transport Policy*, 16 (1), pp. 19-28.

Campos, J., G. de Rus and I. Barron (2009), “The cost of building and operating a new high-speed rail line”, in: G. de Rus (ed.), *Economic Analysis of High-Speed Rail* in Europe, BBVA Foundation.

de Rus, G. (2008), “The Economic effects of high-speed rail investment”, International Transport Forum, OECD, Paris. Discussion Paper No. 2008-16.

de Rus, G. and C.A. Nash (2007), “In what circumstances is investment in high-speed rail worthwhile?”, Working Paper 590, Institute for Transport Studies, University of Leeds.

de Rus, G. and P. Socorro (2009), “Infrastructure investment and incentives with supranational funding”. Paper presented at the VIIIth Milan European Economy Workshop, University of Milan, in the framework of the EIBURS project.

Dixit, A.K. and R.S. Pindyck (1994), *Investment under uncertainty*, Princeton University Press.

Downs, A. (1957), *An economic analysis of democracy*, New York: Harper& Row.

ECORYS Transport (2005), “Ex post evaluation of a sample of projects co-financed by the Cohesion Fund (1993-2002)”, Synthesis Report for the European Commission, DG Regional Policy.

European Commission (2009), “The future of transport”, Focus Groups’ Report, Brussels.

Florio, M. (2006), “Cost-benefit analysis and incentives in infrastructure planning and evaluation: a research agenda for the EU cohesion policy”, 5th Milan European Economy Workshop, 26-27 May 2006.

Flyvbjerg, B., N. Bruzelius and W. Rothengatter (2003), *Megaprojects and risk: An anatomy of ambition*, Cambridge University Press.

Guasch, J.L. (2004), *Granting and renegotiating infrastructure concessions: Doing it right*, Washington, D.C.: WBI Development Studies, The World Bank.

Laffont, J.J. and J. Tirole (1993), *A theory of incentives in procurement and regulation*, MIT Press, Cambridge, MA.

Mackie, P. and J. Preston (1996), *The local bus market: A case study of regulatory change*, Avebury.

Morrison, S.A. and C. Winston (1995), *The evolution of the airline industry*, The Brookings Institution, Washington, D.C.

Morrison, S.A. and C. Winston (2005), "What's wrong with the airline industry? Diagnosis and possible cures", Hearing before the Subcommittee on Aviation, Committee on Transportation and Infrastructure, United States House of Representatives.

Nash, C.A. (1993), "British bus deregulation", *The Economic Journal*, 103 (419), pp.1042-1049.

Nash, C.A. (2003), "Marginal cost and other pricing principles for user charging in transport: a comment", *Transport Policy*, 10 (2) pp. 345-348.

Niskanen, W.A. (1971), *Bureaucracy and representative government*, Chicago: AldineAtherton, Aldershot.

Olsen, T. and P. Osmundsen (2005), "Sharing of endogenous risk in construction", *Journal of Economic Behavior & Organization*, 58 (4), pp. 511-526.

Preston, J. (2004), *The deregulation and privatisation of public transport in Britain: twenty years on*, Transportation Research Foundation.

Robinson, J.A. and R. Torvik (2005), "White elephants", *Journal of Public Economics*, 89 (2-3), pp. 197-210.

Rothengatter, W. (2003), "How good is first best? Marginal cost and other pricing principles for user charging in transport", *Transport Policy*, 10 (2), pp. 121-130.

Sobel, R.S. (1998), "The political costs of tax increases and expenditure reductions: Evidence from state legislative turnover", *Public Choice*, 96 (1-2), pp. 61-80.

Steer Davies Gleave (2004), *High-speed rail: international comparisons*. Report prepared for UK Commission of Integrated Transport, London.

Vreeker, R. and P. Nijkamp (2005), "Multicriteria evaluation of transport policies", in: K.J. Button and D.A. Hensher (eds.), *Handbook of transport strategy, policy, and institutions*, Elsevier, Oxford.

Theme III:

Competition and Regulation of Interurban Travel: Towards New Regulatory Frameworks?

COMPETITION OR CO-OPERATION IN PUBLIC TRANSPORT

Botond ABA

Institute for Transport Sciences Ltd. (KTI)
Budapest
Hungary

SUMMARY

1. INTRODUCTION

Transport, and particularly public transport, is a regular subject for academic and policy analyses. Here the focus is generally not so much on the positive results of transport innovation (for example, innovations in bus transport and its dynamic development), but more on the problems caused by that innovation - for example, subways and high-speed trains - and the ensuing social strains.

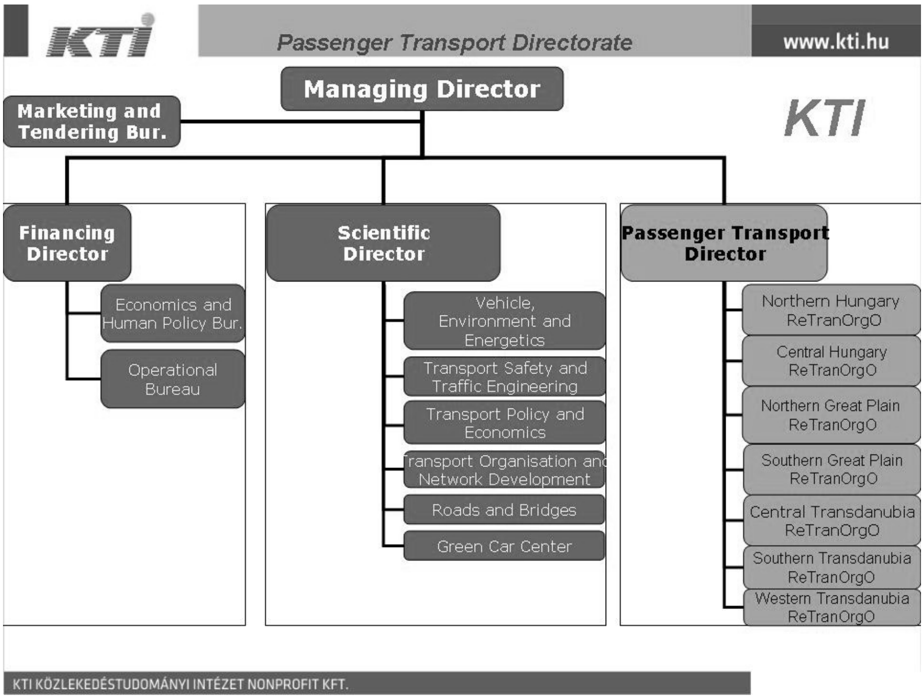
The recent economic crisis has particularly highlighted social contradictions within a globalised economy, with markets contracting and enterprises competing to hold their monopolistic status or even to survive. Focusing especially on the transport sector, competition is defined in terms of enterprises, old and new, attempting to influence and convince passengers to select and use their transport regularly. In this, individual means of transportation are often chosen by the public. This reality has historically benefited, and continues to benefit, the automobile industry in particular, thereby yielding it increasing power and influence over the whole transport industry and the economy at large. The author’s hypothesis in this paper, is that, in economic terms, the public transport sector is very different from other, more “conventional” categories of the market sector. While personal utility and real value do not determine the market as a whole (supplies and demand included), the phenomena of the pseudo-market does. Moreover, the sustainability of public transport does not depend on the individual defined in terms of individual passengers, but in terms of the community and its common wealth.

2. PRESENTATION OF THE INSTITUTE FOR TRANSPORT SCIENCES LTD (KTI)

KTI celebrated its 70th anniversary in 2008. In the last eight years, during which the strains on the market increased, it has focussed its attention on the public transport market. During this period, the Institute’s fields of interest have increased, with the traditional areas (automobile transport and road research) being complemented by railway transport, transport policy and economics research.

In 2007, the Passenger Transport Directorate was established, with seven regional offices spread over the country. Together, they evaluate interurban public transport performances and processes and control the fulfilment of public service contracts.

Figure 1: **Organisation of KTI**



The traditional focus of KTI has been on research and measurement. More specifically, in the examination of the passenger transport sector, our young researchers have been making continuous efforts to analyse and reveal more comprehensive modes of understanding.

Our colleagues, who have been dealing with the various topics of focus, are briefly introduced below:

– Gábor Albert and Árpád Tóth, who developed the “concurrency index”, which determines the competition between different transport modes and quantifies the potential of alternative routes;

– Balázs Ács, who deals with the quantification of the economic background and experiences of long-distance bus transport;

– Dr. Maria Heinczinger and her team, who have been studying “The impacts of flat rate and discounts introduced in PT on taking rail and on division of modality”.

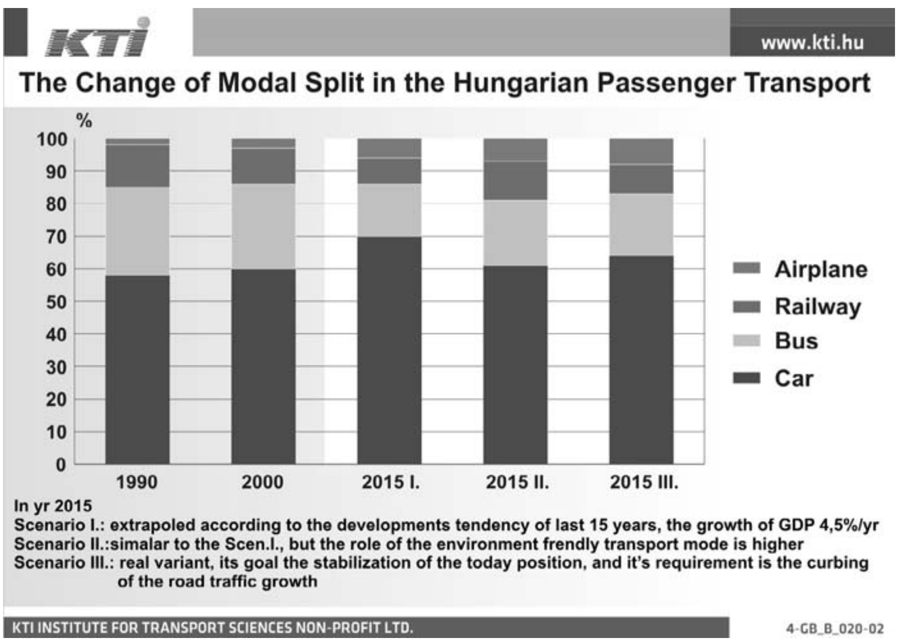
– In Hungary, it is not only the KTI experts who are analysing and researching these questions. One of the KTI’s associates is László Kormányos (Hungarian State Railways/MÁV). who, in his Ph.D., developed scientific models presenting service improvement and technical evolution by using the integrated periodic timetable.

3. OVERVIEW OF THE PASSENGER TRANSPORT SECTOR IN HUNGARY

In terms of the development of their transport sectors, the experiences of the new European Union's member countries do have some common attributes, but are nevertheless substantially different. The following basic statements appropriately define the passenger transport sector of Hungary and its evolution since 1990:

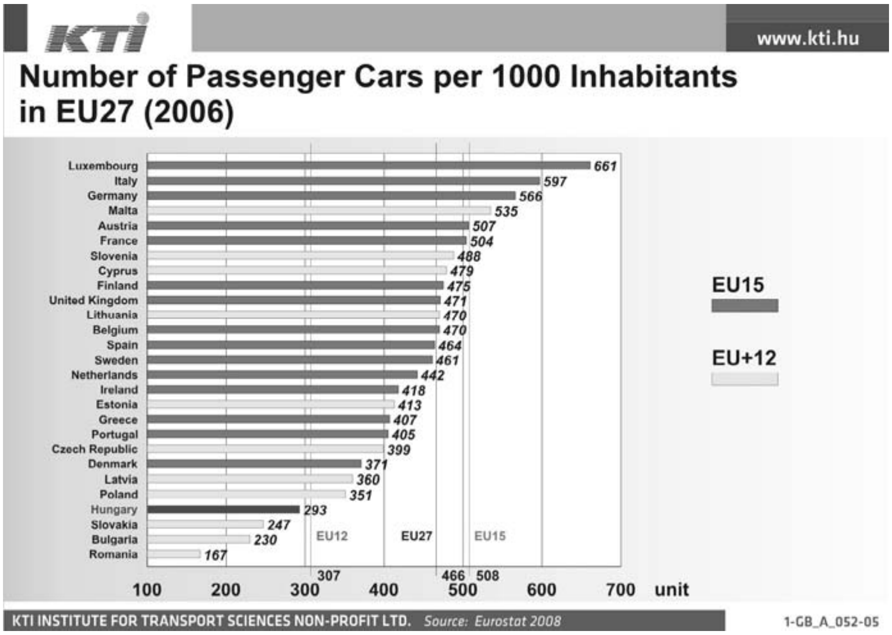
– Hungary, as with other eastern-European countries, initially had a very high modal split: the number of privately-owned cars was low and public transport was the dominant form of transportation.

Figure 2: **Modal split forecast**



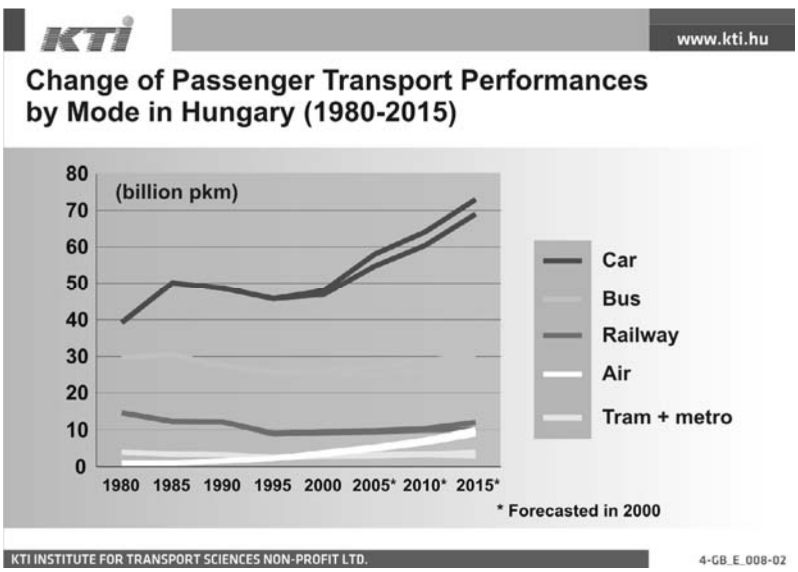
– Compared with the EU-27 as a whole, the number of cars used is still very low.

Figure 3: **Number of automobiles per 1 000 inhabitants in the EU27**



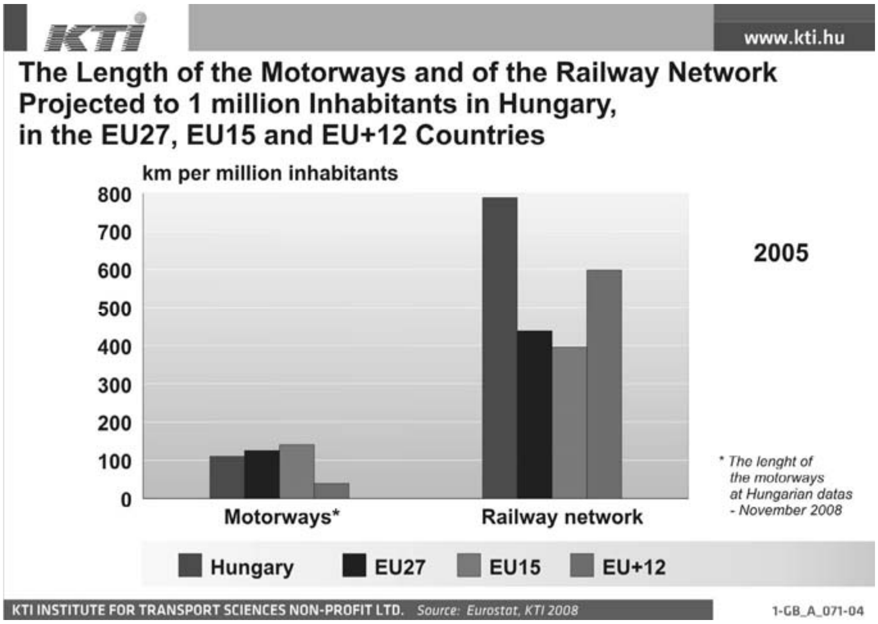
– No real uniform market initially existed in central and eastern European countries. In Hungary, the former Hungarian Planning Institute divided public services between the two state-owned groups of enterprises (the MÁV and the National Bus Transportation Companies/Volán).

Figure 4: **Forecast of passenger transport performance**



– **The railway sector had a significant infrastructural potential**. There was also a marked lack of motorways, and intercity bus transport had a very low infrastructure potential.

Figure 5: **Infrastructure potential**



– The length of the motorway network grew radically, thereby causing changes in distribution: a growing demand for motorway use had the effect of launching the intercity bus services, developed to reach high speeds on the newly-built motorways. Furthermore, the private sector also showed interest in this new development.

Figure 6: **Development of motorways**

KTI www.kti.hu

Length of Motorways at End of Year (km)

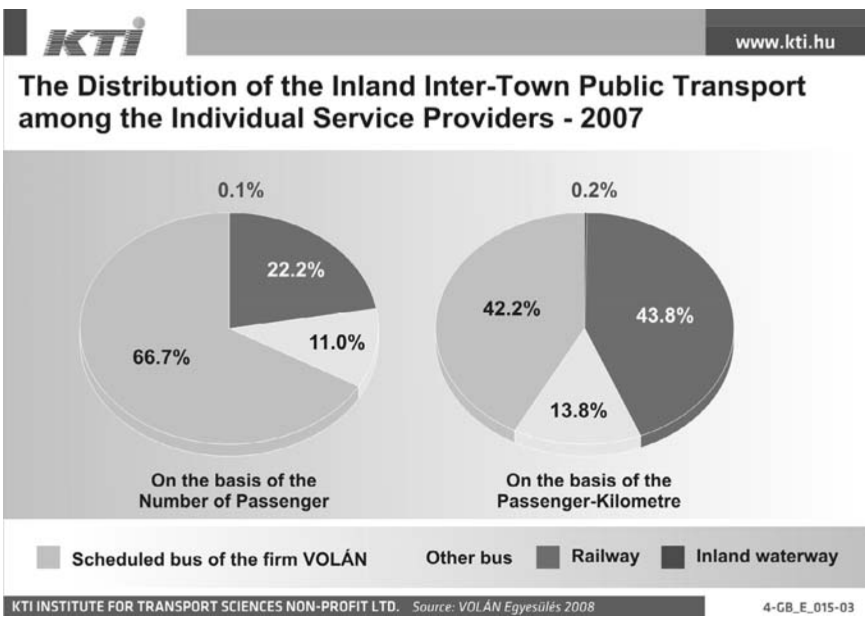
	1995	2000	2005	Growth between 1995-2005
EU27	47 969	54 790	61 565	+28,3%
EU15	47 467	51 563	57 565	+21,3%
EU+12	2 502	3 227	4 000	+59,9%
HU*	335	448	636	+89,9%

*1112 km (November 2008)

KTI INSTITUTE FOR TRANSPORT SCIENCES NON-PROFIT LTD. Source: Eurostat, KHEM 2008 1-GB_A_028-07

– However, the market did not open up to the private sector. Most long-distance public transport operators are still state-owned participants.

Figure 7: **Distribution of domestic interurban public transport**



– When Hungary acceded to the EU, the transport sector was characterised by a lack of domestic regulation and long-term public service contracts for most of the state-owned enterprises (which is still the case today).
– It is hard to determine when the market will open up. It is said that 2012 could represent a turning point.
– The lack of specific international benchmarking makes the situation problematic for the authorities concerned. (Real data could scarcely be determined because of market interests.)

Consequently, the following questions arise with regard to the Hungarian passenger transport sector:

3.1 Question 1: Is there any competition?

The answer is a rather surprising but definite **yes**.

Figure 4 clearly shows that the number of automobiles has increased immensely since 1990, generating a higher market share in the public transport sector.The modal split has failed since 1980, when it was 50-50%. If we follow the evolution of the past 15 years, by 2015 public transport's market share should have decreased by 30%, unless a drastic intervention takes place in market processes.

It is important to highlight that this drop from 50 to 30% is no more dramatic than the change in city transport, where the modal split fell from 82% to 60-55% between 1988 and 2008. It is useful to take into consideration the analogies of city transport, which show, more clearly than interurban traffic indicators, the characteristics of the enterprise-based economy and market and social distortions.

Before asking who the beneficiary of this competition is, it is worth determining its nature and how social traditions could be changed in this respect.

3.2 Question 2: What kind of competition are we talking about?

Although not a new statement, we often forget that in the passenger transport sector there are dual purchaser relations. The passenger as an end-user of the service feels the market effects – service quality and cost – directly. Depending on the social settings of different countries (for example, the Netherlands, Italy or Hungary), road users act differently in terms of:

– Journey time (see Figures 8 and 9);
– Access and (egress) walking time (which is almost zero in the case of individual transport);
– Accessibility or frequency of services;
– Waiting and transfer times;
– Own service quality through personal accessories (for example, Internet use);
– Journey cost.

As a KTI survey of passenger complaints concluded, the above statements' effects on feelings of insecurity sharply influence the market. It seems that today's passengers are more sensitive to indirect values, such as comfort, safety and stress effects, than to primary ones, such as travel time and charges. Although the measurement of these values is not based on scientific exactitude, they are clearly reflected by tendencies.

In practice, these questions of quality lie within the competency and responsibility of the relevant authorities, as soon as any intervention occurs in the market processes. Here, actions taken to regulate market access and compensation for losses and payments are considered as interventions, whether carried out by a state, a regional council or an optional, self-monitoring market.

In Hungary, strong competition developed between individual and public transportation (PT) systems. The economic crisis has had an impact on both sectors but in different ways. While in the PT sector public financing restrictions have forced the State, as the responsible authority, to radically reduce expenses, individual transport users have not had the same economic and cognitive perception of the economic crisis. For instance, people in Hungary have been "envious" of citizens in other countries, where the introduction of "scrapped car subvention" schemes has encouraged individual transportation.

Figure 8: **Accessibility of motorways**

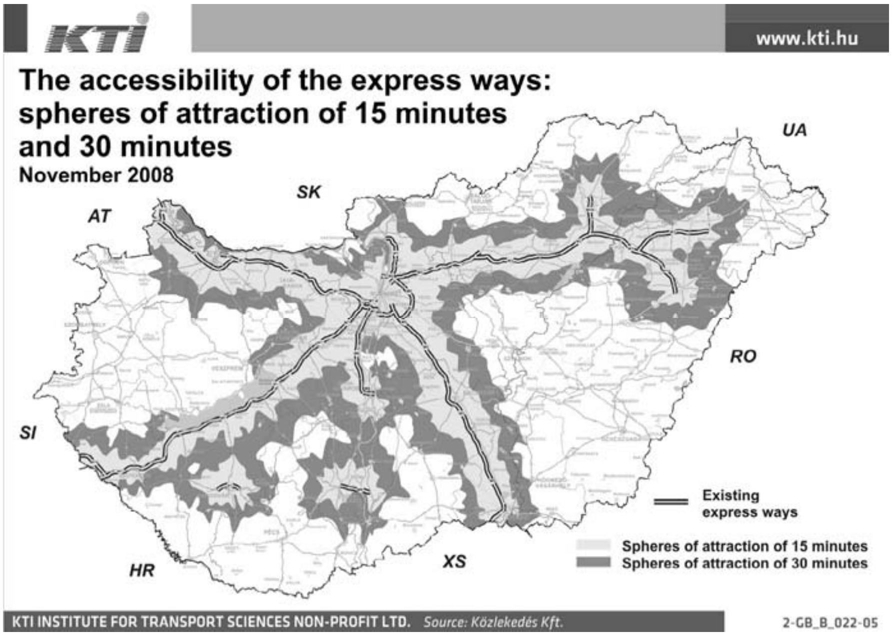


Figure 9: **Accessibility of chief county towns**

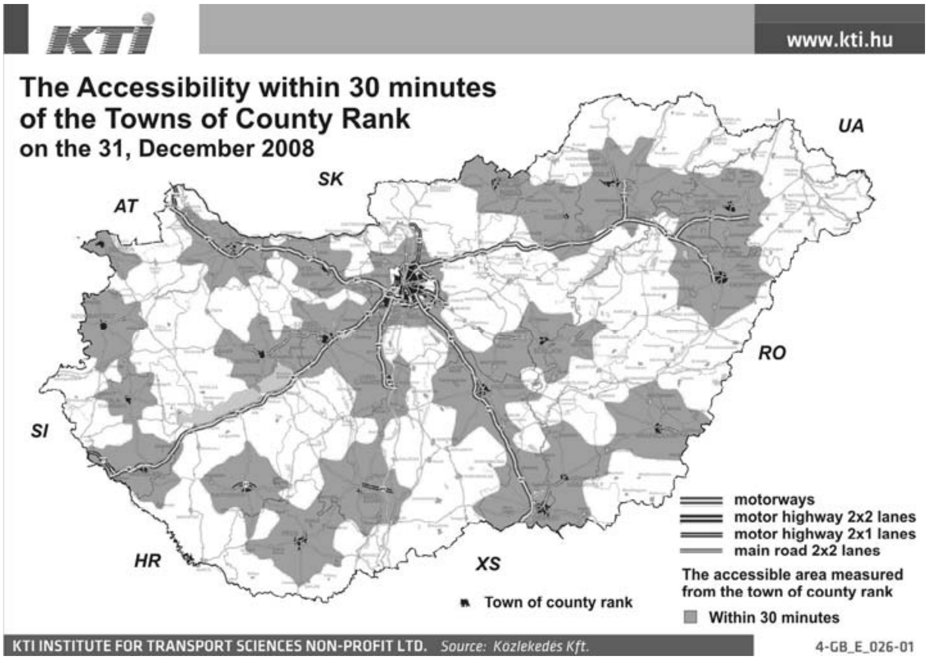


Figure 10: **Pacific Electric Railway car piles at junkyard on Terminal Island, California, 1956**

In 2008, Hungarian households spent a total of approximately EUR 1 billion on financing public transport operations, as provided by EU regulations. In 2009, this amount was blocked. In order to fight the global economic crisis, in 2010 the EU, in co-operation with international financial institutions, plans to reduce the normal level of investment by 10-15% over three years.

As a general rule, decreasing public financing generates strong competition in the transport sector. This competition is not put into practice simply as active lobbying. In the author's point of view, it is a pseudo-market phenomenon, in particular distorting market relations, including the redistribution of transport performance, while operators try to reduce loss-generating services. In

Hungary, for example, bus operators were invited to scrap 3 500 services in 2008 and 2009. As a result, the economic recession seems to have revealed a severe confrontation between lobbying groups.

At the same time, there has barely been any competition between bus operators. The aim before the crisis was to "gain" new passengers from the railway sector by providing a better service structure. However, today, bus operators are either trying to acquire passengers from each other or giving up their territories, thereby incurring losses. This is true not only for small, private enterprises but also for large, state-owned ones with long-term territorial contracts.

The following table summarizes the main advantages and disadvantages of an evolving competition:

Advantages	**Disadvantages**
It more likely reflects real demand in rush hours and on used lines.	Off-peak times remain with no service (or are overbid).
According to route demand, a quicker access time is assured.	There are blank territories and no side-trip routes (for example, farmsteads).
Chain efficiency improves: operators "give up" passengers where their service is not sufficiently effective.	Last-mile problems: who is responsible for a door-to-door service?
(To choose the right volume efficiency and to build an integrated service system in a company or in a group of companies).	Efficiency deterioration in high overhead cost monopolies.
The required effect of volume efficiency is increased. (Earlier, there were train services even with no passenger demand.)	(Instead of having 3-4 bus services, a train service would be indicated.)

Conclusions:

- A national transport culture influences passengers' market decisions and the rationalisation of transport operators' choice;
- Passengers very quickly accept reliable service conditions;
- The economic interests of transport operators are often in contradiction with passengers' interests;
- Public financing and its disproportionate distribution distorts the transport operators' efficiency.

Overall, the question is one of efficiency and how to prioritise it, between:

- The State/region/council's efforts focused on social welfare;
- Transport operators focused on profit;
- Passengers, vindicating their individual interests.

3.3 Question 3: "To be or not to be?" To compete or to co-operate?

To answer this question, it is necessary, and sufficient, to analyse these two market methods - competition and co-operation – in the context of the above priorities.

It is widely believed that in the context of economic competition, the winner is the consumer, namely, the passenger, as competition favours the individual passenger's interests.

This statement is evident in the real market environment. However, this advantage does not obviously manifest itself in the pseudo-market phenomena of public transport. Two factors could be further determined for clarification:

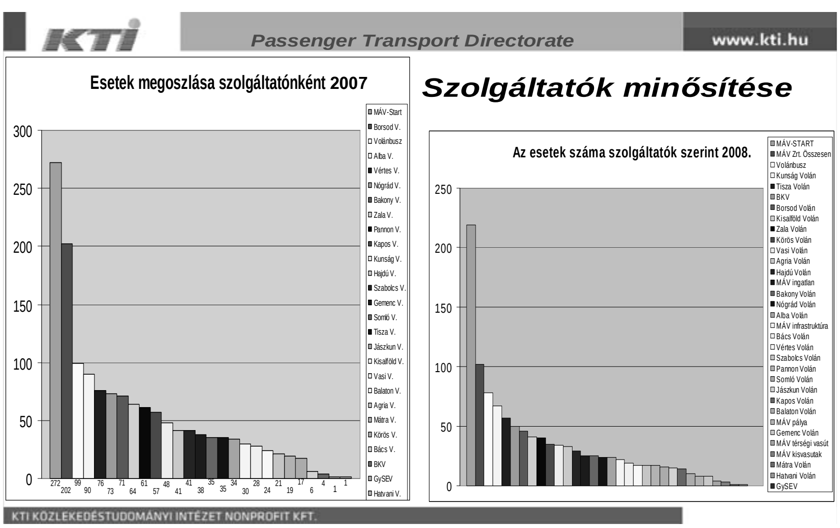
- Transport operators do not directly affect passengers' value (there are always exceptions), but they compete for market shares. After having signed public service contracts, the extent to which transport operators comply depends on how strongly the public service obligation (PSO) is controlled. It is evident that in the private sector (supposing the best of intentions) efforts towards profitability arise. Here, the profitability target and the improvement in modal shift are contradictory.
- Passengers do not evaluate public costs spent on transport, especially if resources are blocked by an economic crisis, contrary to individual transport, where only fuel expenses count. (Deformations of this kind of evaluation and economic rationalisation belong to more important and more harmful pseudo-market phenomena.)

In the presence of a responsible authority or municipality, only one aim counts: what is the relationship between public transport responsibility, social solidarity and market economics?

Accordingly, success is influenced by three factors:

- Constituent, namely passenger, satisfaction can be evaluated not only during public elections. The monitoring of the evolution of passenger complaints in terms of quantity and quality is a great method to evaluate the efficiency of responsible authorities.

Figure 11: **Last two years' evaluation of transport operators**

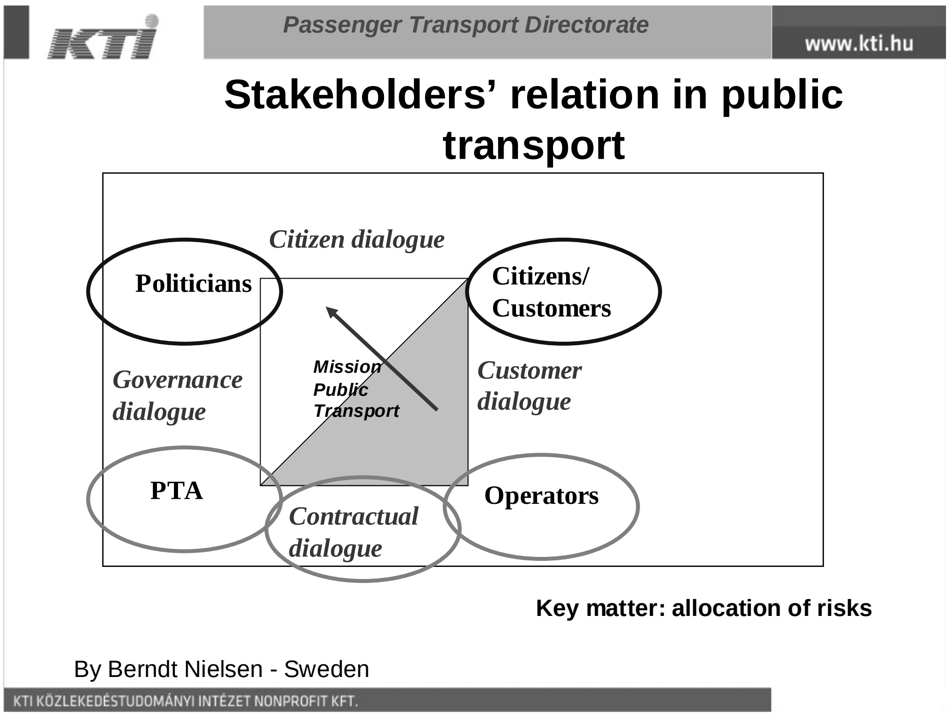


- The interests of responsible authorities are evident: to *decrease public expenses* and to improve "public efficiency".

– A *demand for market domination* has emerged, whereby procedures would be calculable, comprehensible and feasible. Assuring this is the obligation of the responsible authorities. Unfortunately, a demand for little regulation from the private sector results in states maintaining unpredictable market relations.

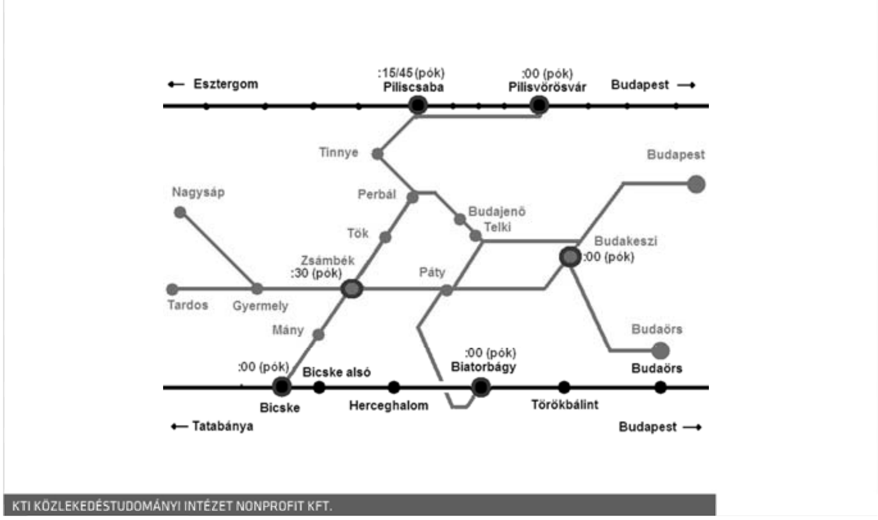
Transport operators' interests seem simple and direct: to make a great profit and to secure capital return. The Hungarian pseudo market has the special feature that every state-owned transport operator is interested in profit minimizing. At the same time, state-owned institutions have exclusive rights for risk allocation.

Figure 12: **Stakeholders' relationship by Berndt Nielsen**



In a state of co-operation, it is much easier to find common interests. It is evidently impossible to fulfill all passenger transport demand with one mode only. All actors, the responsible authority, transport operators and passengers, are interested in *choosing an optimal way to change modes*, with all of them equally benefiting from co-operation. However, if participant interests are substantially diverse then co-operation will quickly fall through.

Figure 13: **Intermodal system with periodic timetable (ITF) in Western-Buda area**



KTI has found different ways for choosing the optimal way of changing a mode of transportation. One of them is the integrated periodic timetable. Two train operators (the MÁV Hungarian State Railways Company and the ROE Train Company), two bus operators and one of the subcontractors tightly co-operate with each other on the network below.

Competition and co-operation are not opposite categories but joint elements for a rational economy. *Consequently, it is inappropriate to oppose co-operation and competition.* Instead:

- Real market competition has to be created (so as not to maintain an artificial pseudo market function);
- An optimal level of system co-operation is required (where average use time is minimal and the smallest public cost means, in financial terms, "the highest profit");
- Optimal market control tools need to be used in order to create common interests shared by all actors, including the creation of legal regulations as well as an intervention mechanism which is accountable, maintainable and, in risk terms, equally safe for everyone.

The question of what is the optimal solution is the subject of serious professional debates. *The aim here is to bring transport operators' interests towards maintainability rather than profit maximization*. The conditions which have to be included in the contract are as follows:

- In case of extra profit, the responsible authority is entitled to levy concessions;
- In case of accepted losses, the rate of public financing, its calculation and disbursement must be controlled;
- In terms of winning services, the conditions for cross-financing must be clear;
- There must be the possibility to use efficient protection against market influence and manipulation of market relations;
- Participants must (voluntarily and compulsorily) restrict monopoly acquisition.

It is also important that all tariffs be set on a par concerning both individual and public financing.

At the end of the last century, modal split was at 60%. This figure is taken from a representative survey elaborated by researchers of the city transport sector. These researchers were curious to analyse the population's opinion at the time, concerning the rate of individual and public financing with regard to transportation costs.

The result was, predictably, 60%:40%. One might expect that the 60% who travelled by public transport preferred public financing, while car-users would support the "pay as you go" principle. Yet, when evaluated, it turned out that the rate of car users' *vs*. public transport users' answers was the same: 60%:40%. This could be interpreted as solidarity being significantly high in society and public transport financing widely well-tolerated at the time. Only international benchmarking could answer whether such a high demand for public financing in Hungary was a specific historical heritage or a truly permanent social demand.

3.4 Question 4: How to regulate competition and enforce co-operation while conforming to the market?

Requirement 1: Analysing the situation of public service

From the data given, it is possible to conclude that there is a need for intervention.

One of the possible measurements is a concurrency index created by KTI. This index measures *how current public services could possibly be replaced by another sector also having a current service*. Its advantage is that the above verbal evaluation also gives an objective numerical evaluation and, by virtue of this, the possibility to create a ranking of different services.

As a result, parallel supply could optimally be distributed and/or reduced. The index shows the interrelation of two public transport operators, between the same two given points, from the passenger's, rather than the transport operator's, point of view.

It does not provide a ranking, but shows the possibility of replacing one of the operators with another.

The concurrency index refers to a number: the value is zero if you cannot reach point B from point A (namely, if there is no substitute operator at either end) or if the alternative mode is unacceptable. The value of 100 corresponds to both operators having the same proven characteristics and thus being equal. The value can be over 100 if the alternative operator is better than the one being replaced. The software examines those cases where:

- Bus only or train only are involved;
- The train ride is without transfers, or the bus has a maximum of one transfer;
- In the case of a transfer, the waiting time (conditionally total access time + waiting time) can be defined by one parameter, or when a longer waiting time is highly acceptable.

Requirement 2: Exact definition of the target position

One of the methods and measurements is a model worked out by László Kormányos, evaluating the mobility supply. The model prefers the supply market; namely, it is a measurement *oriented towards passengers.*

For this analysis, the system of relations, the network aspect and the integrated transport chain are defined, taking into consideration the following parameter systems:

– Average access time (hour);
– Average access frequency (access /hour);
– Complex timetable-structure index (ITF index) (%)
    - frequency index(P),
    - symmetry index(S)
    - transfer index(C).

The *vector method* defines the value vector of mobility supply which, by averaging and weighting, could be quantified in terms of relations and complex networks. It could thus be produced with the parameters of basic mobility supply.

$$\underline{v} = \begin{bmatrix} v_t \\ v_f \\ v_I \end{bmatrix} = \begin{bmatrix} \dfrac{\sum_{i=1}^{n} t_{n_i} \cdot n_i}{\sum_{i=1}^{n} n_i} \\ \dfrac{n}{t} \\ \dfrac{N_p}{N_ö} \cdot \dfrac{N_s}{N_{öa}} \cdot \dfrac{N_k + N_{cm}}{N_ö} \end{bmatrix}$$

The model evaluates the mobility supply for optional public transport relations (one or more relations, even in terms of the whole train network or the overall network of public transport modes). According to mobility demand, *alternative accessibility is at the core of the evaluation model.* In terms of the network aspect, the importance of the analysed mobility supply is defined in terms of the qualitative analysis of the transfers and access time (symmetry, etc.). *Comparative versions with the evaluation of mobility supply (timetable) and the evaluation of parameters and value changes could be determined* (Kormányos, 2009).

Requirement 3: Apart from the "soft" and neutral market tools of state regulation, competition capability and willingness for co-operation among the transport operators could be modified

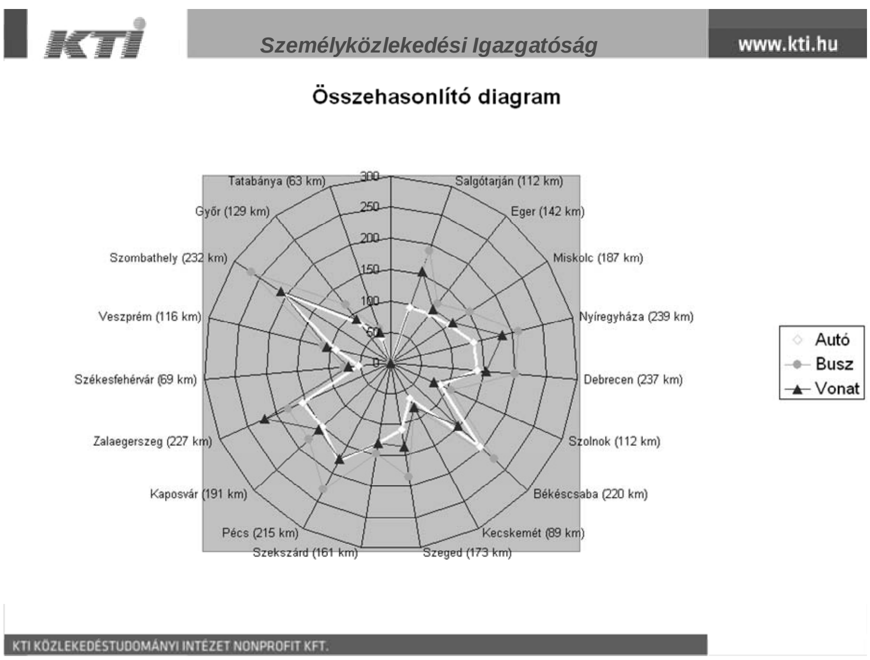
The analysis of market competition capability, and taking advantage of it in terms of infrastructure use, is very important for market accessibility. As shown below, cost efficiency in Hungary is determined by infrastructure fees.

Usage (route) cost of infrastructure per 1 000 seat-kilometres

Mode	**Cost**
Bus transport	~ EUR 3
Train transport	~ EUR 13

The intensive use of infrastructure results in an indirect market manipulation factor, the use of travel time. Disadvantages, such as the reduction of mass and the deterioration of volume efficiency, also have to be mentioned.

Figure 14: **Access time among competitive sectors (Private car, bus and train)**



The last means of market control mentioned in this study is through tariffs and their effect on competition capacity. It must be clear that one of the most remarkable ways of maintaining a balance between supply and demand and between the different transport modes is through tariff policy.

Figure 15: **Comparison of full price tickets (bus, rail, common)**

Figure 16: **Number of passengers before and after tariff reforms**

Transport operators are usually unwilling to use over-complicated tariff systems. However, Figures 15 and 16 show that the unification of the tariff system not only affects demand and competitiveness. The dramatic deterioration in bus transport depends more on the simplification of the discount system. Here, the more than 43 types of discount were simplified to the detriment of bus passengers. Nevertheless, the relative increase in rail tariffs was higher. But, the average travel distance explains the real effect: 19 kilometres in the case of buses and 56 km for trains.

4. CONCLUSION

To conclude, Figure 17 below summarizes rail and bus services in Hungary. It shows the remarkable differences existing between the rail and bus sectors. These seem to indicate that as long as the state contribution to the railway sector remains as high as is shown below, no real competition can be forecast.

Figure 17: **Summary of Hungarian bus and train services**

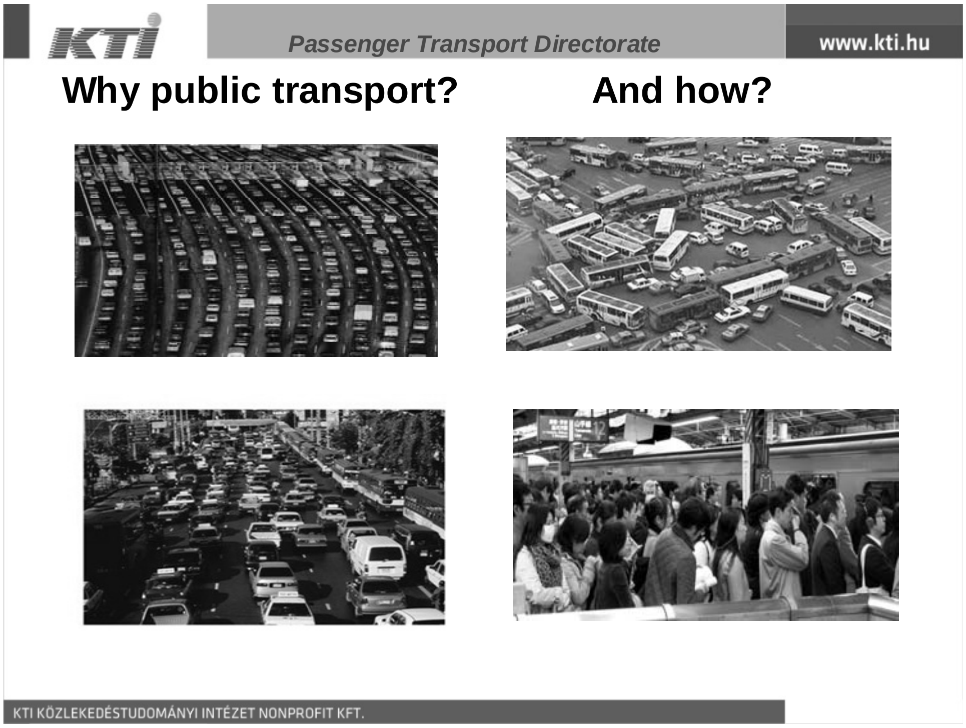
KTI Passenger Transport Directorate www.kti.hu

Rail and bus services in Hungary

Characteristics	Rail	Bus
Transport performances	Passenger: 150 million Passenger-km: 9 billion	Passenger: 480 million Passenger-km: 9 billion
Proportion in regional public transport	Passenger: 22 % Passenger-km: 5 %	Passenger: 77 % Passenger-km: 50 %
Travel distance (average)	Cca. 60 km	Cca. 18 km
State contribution (for 1 passenger on 1 km)	Cca. 15-25 HUF	Cca. 1 HUF
State subsidy since 2000	1111 Md HUF	44 Md HUF
Public service contract	1-1 year	By 2012
Service level changes	1+3 times per year	3-8 times per year

KTI KÖZLEKEDÉSTUDOMÁNYI INTÉZET NONPROFIT KFT.

(And finally)



(Is this really a thing of the past?)

LESSONS FROM THE US TRANSPORT DEREGULATION EXPERIENCE FOR PRIVATIZATION

Clifford WINSTON

Brookings Institution
Washington, DC

USA

SUMMARY

* This paper draws on material in my book, *Last Exit*: *Privatization and Deregulation of the U.S. Transportation System*.

1. INTRODUCTION

Travellers throughout the world are generally dissatisfied with their country's transportation system because of the significant highway congestion, air travel delays, unreliable public transit service, and so on, which they are forced to endure. Public officials have sought to address such problems by increasing government spending on transportation; but it has become quite clear that most, if not all, countries cannot spend their way out of their transportation problems.

The failure of the public sector to manage and operate transportation systems efficiently has spurred some countries to explore whether expanding the role of the private sector could improve the performance of their transportation modes and infrastructure. Examples include privatized railroads in various countries in Europe, privatized subways in Tokyo and Hong Kong, privatized airports in London and Sydney, and privatized highways in a few parts of the United States.

Of course, the limited privatization of transportation that has occurred around the world is not pure privatization because governments have maintained a presence by instituting some form of regulation such as price caps and limits on entry. Thus, considerable uncertainly remains about the economic effects of privatizing and deregulating part of or an entire transportation system and how policymakers should manage the transition to privatization to maximize its effectiveness.

The purpose of this paper is to suggest how the US experience with deregulating its intercity transportation system can identify important considerations for all countries that wish to pursue privatization. Transportation deregulation in the United States gave private railroad, trucking, bus, and airline companies the freedom to set prices, choose which markets to serve, and what level of service to provide. Because US firms were saddled with inefficiencies that developed over decades of regulation, their adjustment to deregulation has been difficult and time consuming. Nonetheless, deregulation has succeeded to a notable extent in the short run and could provide even greater benefits in the long run.

Privatization would give companies that were formerly in the public sector, such as public buses, railways, airports, and highways, the freedom to set prices, raise capital, and offer service in a competitive environment. Based on the deregulation experience, privatization could generate large benefits by enabling transportation providers to develop efficient practices, to be more responsive to consumers' preference, and to implement new technologies in a timely fashion. At the same time, privatized firms would have to overcome inefficiencies that are even greater than those that deregulated firms had to overcome because they were managed and operated by the public sector. Policymakers should be aware of this fundamental challenge and, if possible, take steps to ameliorate the difficulties that privatized firms would inevitably encounter.

2. TRANSPORTATION DEREGULATION IN THE UNITED STATES

Privatization and deregulation ar transformative policies where the government transfers (through a sale) the parts of the transportation system that it owns and operates to private firms and does not regulate those firms' prices, service, and expansion and contraction of their networks (entry and exit).[1]

With the exception of transferring the northeast freight rail system, Conrail, back to the private sector, the United States has not had recent experience with privatizing any part of its transportation system; but its recent experience with partially deregulating intercity transportation—railroads, trucking, airlines, and buses—has given us an opportunity to accurately assess the economic effects of that policy and to identify some important issues related to privatization.[2] As indicated by the term partial deregulation, policymakers did not deregulate every aspect, economic and otherwise, of carrier operations. For example, freight railroads are still subject to maximum rate regulations. In addition, policymakers did not reform public infrastructure policies to ensure that each mode's infrastructure would be in accord with carriers' adjustments to deregulation. For example, airports did not introduce congestion pricing even though airlines' accelerated development of hub-and-spoke route structures increased the demand for scarce runway capacity during peak travel periods throughout the day.

Two important considerations should guide interpretations of the evidence from deregulating the US intercity transportation system. First, because regulation and deregulation never occurred at the same time at the national level,[3] the most accurate way to measure the economic effects of deregulating a transportation industry is a counterfactual analysis that estimates the price, cost, and service changes that are solely attributable to deregulation and thus *would not have occurred* had the industry still been regulated. Second, as noted, the intercity transportation industries are still subject to some government regulations and some, if not all, firms that were subject to regulation have not fully shed their regulatory bequeathed operating practices and capital structure.

It is therefore useful to distinguish between the short-run and long-run effects of deregulation on the performance of an intercity transportation industry. In the short run, the industry has not been completely deregulated and may be subject to other government policies that compromise its performance under (partial) deregulation. In addition, firms that existed in the industry prior to deregulation have not fully adjusted their operations and investments to the deregulated environment. In the long run, the industry is fully deregulated and firms have optimized their operations and investments to this environment.

3. THE SHORT-RUN EFFECTS OF DEREGULATION

Beginning with the 1978 Airline Deregulation Act, prices, service, entry and exit in the intercity transportation industries were substantially deregulated. However, travellers are still experiencing the short-run effects of airline deregulation because carrier competition and operations have been constrained by the lack of available gates at some congested airports; inefficient airport pricing and investment have allowed travel delays to grow, especially at hub airports, which handle far more operations under deregulation than they did under regulation; various hearings on and potential regulatory interventions in airline service and competition have partly diverted managements' focus from improving carrier operations; and tensions between managers of legacy carriers and labor continue to exist because the "rent sharing" mentality that developed under regulation has persisted under deregulation.[4]

The nation is still experiencing the short-run effects of railroad deregulation because maximum rate guidelines have not resolved the captive shipper problem—that is, some shippers have access to only one railroad; the threat of some form of rate-regulation has, at times, diverted the attention of rail managers from improving carriers' operations; and railroads have not completed the task of optimizing their networks and realizing greater economies of density by abandoning and consolidating the extensive track network that was built under regulation and by building new lines to serve high-volume shippers. And the nation is still experiencing the short-run effects of trucking deregulation because inefficient highway pricing and investment has caused delivery times to become longer and less reliable, which makes it more difficult for truckers to provide high-quality service to facilitate shippers' just-in-time inventory policies.

Despite being adversely affected by the lingering effects of regulation and deficient infrastructure, the intercity transportation industries have significantly improved their efficiency under deregulation and benefited users by reducing prices and providing better service.[5] The key steps in the industries' process of adjustment have been the entry of new firms and the expanded entry by incumbent firms that has increased competition, and the freedom and incentive to improve operations and service quality to users. Deregulation also has its critics who point to financial crises, losses to labor, degradations in service, and the like as indicative of its failings.

Entry and price changes. Intercity transportation firms compete at the market or route level. It is often thought that the number of firms in a market is the most accurate indication of the level of competition; but deregulation showed that the identity of the firms may be as, if not more, important than the number of firms in determining the intensity of competition.

Competition increased in the deregulated airline industry because more (equivalent-sized) carriers competed on airline routes over given distances and because of the growth of new low-cost (low-fare) carriers such as Southwest Airlines. Morrison and Winston (2000) found that Southwest sharply reduced fares on routes that it serves, on routes that it could potentially serve (i.e. Southwest serves one or both of the airports on the route but not the route), and on routes where it supplies adjacent competition (i.e., Southwest serves origin and destination airports that are within say fifty miles of the origin and destination airports that make up a given route).

Competition increased in the deregulated LTL (less-than truckload) trucking industry because of the growth of low-cost (nonunion) regional carriers and because of increased competition from alternative small shipment carriers such as UPS and Federal Express. The TL (truckload) sector has always consisted of unregulated competitors in the form of private trucking. Still, competition in this sector intensified following deregulation because of the growth of national mega-carriers (also called advanced truckload carriers), such as Schneider National and Landstar, and because private carriers were given the opportunity to transport other firms' freight.

The railroad industry has not experienced entry of new carriers since deregulation. Nonetheless, railroads have had to contend with additional competition provided by advanced truckload carriers, and they have enhanced their own competitiveness by accelerating the development of intermodal (truck-rail) service. Moreover, competition among railroads has increased because a large fraction of deregulated rail traffic moves under contract rates, thereby enabling shippers in many instances to play one railroad off against another when they negotiate rates.

In the most intense case, two railroads compete directly for a shipper's traffic if their tracks traverse directly into the shipper's plant or if they have access to the shipper through reciprocal or terminal switching. As pointed out by Grimm and Winston (2000), shippers that are captive to one railroad may benefit from locational competition supplied by a nearby carrier. For example, a shipper may be served by Railroad A but could threaten to locate a new facility on or build a spur line to Railroad B as a bargaining chip to obtain a lower rate from Railroad A or to get Railroad B to commit to a reduced rate. Shippers could also stimulate railroad competition in some cases through product or geographic competition. For example, an industrial site served only by Railroad A in a given market may be able to use a substitute product shipped from a different origin by Railroad B, or the site could obtain the same product from an alternative origin served by Railroad B. Finally, small shippers that may not be able to get railroads to compete intensely for their traffic may improve their bargaining position by using third-party logistics firms, which achieve cost savings for shippers by leveraging the volumes of all their clients to obtain discounts from carriers.

Consumers benefited from lower prices generated by new sources of competition in the intercity transportation industries, including incumbent firms, new entrants, and alternative modes. And those gains were magnified because competition also caused firms to operate more efficiently and to pass on much of the cost savings to consumers in lower prices. Deregulated competition has been sufficiently intense to cause airline fares on low-traffic density (non-hub) routes to fall (Morrison and Winston (1997) and to cause rail fares to approach long-run marginal cost in duopoly markets for coal transportation [Winston, Dennis, and Maheshri (2008)].

Improvements in operations and service. Deregulation enabled intercity transportation carriers to simultaneously improve the efficiency of their operations and their service to travellers and shippers. Freed from entry and exit regulations, airlines have accelerated the development of hub-and-spoke route networks that feed travellers from all directions into a major airport (hub) from which they take connecting flights to their destinations. Carriers use hub-and-spoke route systems to increase load factors and reduce average costs and, by increasing the number of feasible flight alternatives, to offer travellers much greater service frequency. For example, an additional aircraft departure from a spoke airport to a hub airport can increase the number of flight alternatives on many connecting routes.

Railroads have improved the design of their networks to channel more traffic on a given route and have made greater use of double stack rail cars and intermodal operations to reduce costs and provide faster and more reliable service to shippers.[6] Trucking firms have also improved the efficiency of their networks, reduced costs, and provided faster and more reliable service to shippers.

Carriers have also made much greater efforts, sometimes with the aid of advances in information technology, to tailor their services to travellers' and shippers' varied preferences. Airlines have developed revenue (yield) management systems, which have helped carriers increase load factors by offering travellers a wide range of fares from discount fares with various travel restrictions to much higher fares with no travel restrictions. Airlines' computer reservation systems have helped to improve scheduling and flight reservations. Travellers are able to access those systems on airlines' websites to book their travel, thereby obtaining the lowest discount fares, to print their boarding passes and avoid the check-in line at the airport, and to receive real-time schedule information.

Railroads and trucking firms have negotiated thousands of price-service contracts with shippers that align their services with shippers' production and inventory policies and that make more efficient use of their own capacity. For example, shippers can sharply reduce their rates by including backhaul shipments in their contracts. Third-party logistics firms analyze shipper distribution patterns and logistics costs and use sophisticated software to determine the lowest-cost routes and the carriers with the lowest rates. Trucks and railroads also use computer information systems to route their cargo more efficiently and to track shipments.

It could be argued that carriers' adoption of advances in information technology would have occurred regardless of deregulation. But the benefits from those advances were realized because deregulated firms had the financial incentive and operating freedom to design new networks and to engage with customers to determine their preferences. Under regulation, they had little financial incentive or competitive pressure to do so, and regulators certainly were not able to design regulations to stimulate innovative activity.

Criticisms of deregulation. Intercity transportation deregulation has attracted its share of critics—although generally not from academia—who allege that the benefits from the policy have not been widely shared and that the deregulated transportation industries have been subject to service meltdowns and financial crises, which raise questions about their long-term viability. In fact, the benefits from deregulation have been broadly shared among consumers, while the problems that firms have experienced are either part of their long-run adjustment or not attributable to deregulation.

Price regulation benefitted certain travellers by, for example, keeping airline fares below marginal cost on short-haul routes and cross-subsidizing them with fares above marginal cost on long-haul routes, and benefitted certain shippers by preventing railroads from raising rates on bulk commodities. Thus, if economic deregulation improved pricing efficiency, it was not expected to benefit every traveller and shipper. Surprisingly, in the process of improving the cost efficiency of the intercity transportation system, the benefits to consumers from deregulation have been more broadly distributed than expected. And for the most part, consumers' losses can be explained by economic rather than anti-competitive forces.

About 80 per cent of airline passengers (accounting for 90 per cent of passenger miles) fly on routes with lower average real fares since deregulation. Roughly 90 per cent of the difference in the gains to travellers can be explained by the higher costs of serving travellers on low-density routes, where smaller planes have a higher cost per seat-mile and fly with lower load factors (Morrison and Winston (1999)). As noted, deregulation reduced railroad rates, on average, and some small shippers have been able to share in those benefits by using third-party logistics firms. All modes have improved their service quality in the deregulated environment except when their operations have been compromised by public infrastructure inadequacies (e.g., airline travel times have increased because of inefficient runway pricing and investment). Moreover, the benefits from deregulation have been achieved without compromising any mode's safety record (Savage (1999)).

Labor benefited from price and entry regulation because unions' wage demands were not tempered by market forces. However, consumers' gains from deregulation do not primarily consist of transfers from labor. Peoples (1998) concludes that deregulation of railroads, trucking, and airlines caused wages to fall in those industries and resulted in a USD 10.3 billion (1991 dollars) welfare loss to labor, which amounts to roughly 20 per cent of the gains to consumers.

A fundamental challenge facing the intercity transportation industries is to match their capacity with demand. The unpredictability of demand could be particularly problematic for an industry that must invest in capacity long before actual demand materializes. If demand is lower than expected, firms may have to significantly cut prices to fill the available capacity. If demand is higher than expected, firms with the greatest capacity are likely to gain market share. The airline industry has made capacity commitments roughly two years in advance because of the lead times needed to acquire aircraft. Railroads and trucking firms face much shorter lead times when they invest in capacity.

Since it was deregulated in 1978, the airline industry has suffered huge financial losses because of overcapacity that was attributable to the early 1980s and 1990s recessions and to the September 11, 2001 terrorist attacks. It has also suffered losses from the sharp increase in fuel prices in 2008 that substantially raised the cost of carrier capacity. Of course, macroeconomic contractions, terrorist attacks, and spikes in fuel prices are not attributable to deregulation. In fact, industry losses may have been greater if carriers did not have the flexibility to respond to those shocks by adjusting fares and capacity throughout their networks.

Railroads are able to contract with shippers to align their cars and equipment with shippers' demand and to reduce their vulnerability to financial problems caused by overcapacity. But railroad consolidations in the aftermath of deregulation, such as the Union Pacific and Southern Pacific merger and Norfolk Southern's and CSX's acquisition of Conrail, have resulted in service disruptions because the acquiring carrier did not effectively integrate the acquired carrier into its operations. Fortunately, rail operations have improved quickly after the service disruptions and shippers' rates were not elevated because network capacity was restored [Winston, Maheshri, and Dennis (2009)]. In the future, railroads that are involved in consolidations will hopefully take measures to avoid such disruptions.[7]

Finally, airlines have been sharply criticized for their lengthy delays, and in some cases for holding their passengers "hostage" on a tarmac for several hours. But, as noted, air travel delays reflect to a large extent inefficient pricing and investment policies, while extreme delays suggest that an airport is indifferent toward the quality of service that its users receive. In my view, a private commercial airport would seek to develop a reputation for safeguarding travellers and would find it in its interest to prevent airlines from forcing passengers to remain in their aircraft for an excessive period of time (e.g. more than an hour or so) before taking off. Public airports have little economic incentive to reduce travellers' delays and discomfort and are therefore bystanders while passengers are stuck on *their* infrastructure for hours.

4. THE LONG-RUN EFFECTS OF DEREGULATION

In the long run, the benefits to consumers from intercity transportation deregulation will increase as firms are no longer saddled by three short-run constraints: suboptimal public infrastructure, counterproductive residual regulations, and inefficient practices and investments developed during the regulatory environment. The transportation industries cannot address the first and second constraints on their own. Indeed, privatization could significantly ameliorate the first constraint. Unfortunately, even an optimistic assessment would conclude that it would take decades to do so; in other words, the full benefits of deregulation are many years away.

For their part, the intercity transportation industries continue to adjust to the deregulated environment and improve their operations and investments. Through its travails with exogenous economic and non-economic shocks, the airline industry has become more resilient and efficient. It is improving its ability to match capacity with demand under a variety of difficult circumstances. For example, during the past several years airlines have reduced overbooking and denied boarding to fewer passengers by charging higher fees to change flights. But despite some thirty years of deregulation, the industry has yet to be profitable during an economic downturn. In addition, its labor relations are still contentious and it is not well-positioned to compete as effectively as possible in a deregulated global airline market. When those problems are adequately addressed, the industry will, at long last, have shed the inefficiencies of regulation, fully adjusted its operations to the US deregulated environment, and enhanced consumer welfare even further.

The railroad industry has greatly improved its financial performance under deregulation, but it has not earned a normal rate of return on its invested capital on a consistent basis.[8] To achieve that goal, carriers are slowly modernizing their equipment and optimizing their plant size by pruning their networks of unprofitable markets and investing in potentially profitable ones.[9] Rail will therefore continue to make progress in improving its service times and reliability, reducing its costs, and benefiting shippers. The industry’s structure has also not fully adjusted to deregulation. It is possible that more rail mergers will be proposed until only two (highly efficient) Class I railroads remain in the industry. This end-to-end restructuring would create two transcontinental railroads, but still leave two large railroads in the East and two in the West, thereby having little effect on competition. Indeed, this may be the final equilibrium for the US rail freight industry.

The trucking industry has alleviated the serious shortage of long-distance drivers by increasing the use of intermodal operations and increasing compensation. For-hire truckers have significantly reduced their empty mileage under deregulation and they can make further progress by continuing to consolidate loads and by attracting more traffic from private trucking.[10]

5. IMPLICATIONS FOR PRIVATIZATION

By relaxing the federal government's control over airlines', railroads', and truckers' pricing, entry, and exit decisions, deregulation has tried to improve social welfare by accomplishing three goals for consumers and firms: first, to enable them to behave more efficiently within the technological "frontier;" second, to enable them to behave more efficiently as firms innovate and expand the frontier; and third, to enable them to respond more effectively to external shocks to reduce their costs.

Deregulation of the intercity transportation system has accomplished the first goal to a significant extent as firms have improved their basic operations and reduced prices, while heterogeneous consumers have selected price-service packages that are aligned with their varying preferences. Deregulation has made some progress in accomplishing the second goal as firms have successfully implemented advances in information technology to improve their operations. And firms and consumers—in particular, airlines and air travellers—have adjusted their behavior to reduce the cost of economic shocks that have occurred since deregulation began.

Because deregulation is a long term process, firms and consumers have not completely adjusted to it. First, regulation constrained and strongly influenced firms' operations and technology. Economists and other observers have underestimated the time that firms have required to optimize their pricing and service decisions to unregulated competition, to learn how to adjust those decisions to changes in the business cycle, and to shed inefficient operating practices, technology, and counterproductive frictions with labor and their competitors that may seek to gain a political advantage. Firms that have never been regulated occasionally make erroneous and costly business decisions; not surprisingly, deregulated firms have made their share of mistakes and have required considerable time to learn from those mistakes and how to respond to changes in their competitive and macroeconomic environment.

Second, it has been argued that regulation stymies innovation and technological advance [e.g. Gallamore (1999)] and that deregulation provides greater incentives and opportunities for firms to innovate. At the same time, the timing and location of technological advances is difficult to predict. Intercity transportation technology has improved under deregulation; but even after decades of deregulation, it is likely that further innovations that would not occur under regulation await the future.

Finally, the government must adjust its actions in light of deregulation. Counterproductive residual regulations, the threat of re-regulation, and inefficient infrastructure policies have undermined the performance of the deregulated intercity transportation industries.

Similar to deregulation, privatization has the potential to improve the performance of transportation services and infrastructure that are provided in the public sector by giving private firms the opportunity to develop efficient operations and to introduce technological innovations in a timely fashion. In the process, consumers could reap substantial gains.

But privatization differs from deregulation in at least two important respects. First, it would enable private firms to provide transportation services that were formerly provided by the public sector, but unlike deregulated firms most of the private firms would have little, if any, experience competing in those services. Second, unlike deregulated firms, private firms would inherit to a large extent the public sector's highly inefficient operations, investments, and technology.

Thus, transportation firms in a privatized environment are likely to face even greater challenges and more uncertainties in their adjustment to unregulated competition than private deregulated firms may face in their adjustment. Based on US carriers' experience with intercity transportation deregulation, privatized firms' adjustment process would most certainly be time consuming and far from error free.

Policymakers who have an interest in pursuing privatization should appreciate the magnitude of the adjustment process that firms in their country would have to endure to become efficient competitors. Accordingly, they should not maintain or implement policies that may compromise adjustments. And they, as well as the public, must be patient while firms try to overcome mistakes and setbacks that are bound to occur. At the same time, the potential long-run benefits from privatization will hopefully justify the intervening struggle.

NOTES

1. The government may retain some control over firms' exit through the application of bankruptcy and merger and acquisition laws.

2. Recent leases of US highway facilities to the private sector, which are subject to regulations, do not constitute privatization.

3. Regulation and deregulation have simultaneously occurred at the state level. Comparisons of prices and service across states with different regulatory policies have been used to predict and assess the effects of deregulation.

4. Carriers were able to earn excess profits because regulation elevated fares and prevented entry. Labor unions' wage and work rule demands reflected their desire to share in carriers' rents. Deregulation has made it much more difficult for carriers to earn excess profits, but labor and the legacy carriers still have an adversarial relationship that can be traced to their hard fought negotiations during regulation. Carriers that entered the airline industry after deregulation have had to contend much less with this history when they negotiate with labor.

5. Morrison and Winston (1999) summarize the empirical evidence on the economic effects of airline, railroad, and trucking deregulation. Borenstein and Rose (2007) and Winston (2006) provide recent surveys of the evidence for airlines and railroads, respectively. Much less empirical evidence is available for the economic effects of intercity bus transportation.

6. Bitzan and Keeler (2007) estimate that freight railroads have reduced annual costs by as much as USD 10 billion from increased traffic densities attributable to deregulation.

7. Winston, Maheshri, and Dennis (2009) indicate that future consolidations may arise because the remaining major carriers in the west, Burlington Northern and Union Pacific, may merge with a major carrier in the east, CSX or Norfolk Southern, to form two transcontinental railroads.

8. The railroad industry's profitability is a controversial issue. However, it does appear that the industry's returns on investment have been below its cost of capital (Grimm and Winston (2000)).

9. Daniel Machalaba, "New Era Dawns for Rail Building," *Wall Street Journal*, February 13, 2008 points out that for the first time in nearly a century, railroads are making large investments in their networks—adding sets of tracks, straightening curves that force engines to slow, and expanding tunnels for bigger trains.

10. There has been little analysis of the intercity bus industry's adjustment to deregulation. But as noted by Schwieterman (2007), the industry has started to assert itself some 25 years after being deregulated by expanding service in several national markets.

BIBLIOGRAPHY

Bitzan, John D. and Theodore E. Keeler (2007), "Economies of Density and Regulatory Change in the US Railroad Freight Industry," *Journal of Law and Economics*, Vol. 50, February, pp. 157-179.

Borenstein, Severin and Nancy L. Rose (2007), "How Airline Markets Work…Or Do They? Regulatory Reform in the Airline Industry," NBER working paper 13452, September.

Gallamore, Robert E. (1999), "Regulation and Innovation: Lessons from the American Railroad Industry," in: Jose A. Gomez-Ibanez, William B. Tye and Clifford Winston (eds.), *Essays in Transportation Economics and Policy: A Handbook in Honor of John R. Meyer*, Brookings Institution, Washington DC, pp. 493-529.

Grimm, Curtis and Clifford Winston (2000), "Competition in the Deregulated Railroad Industry: Sources, Effects, and Policy Issues," in: Sam Peltzman and Clifford Winston (eds.), *Deregulation of Network Industries: What's Next?*, Brookings Institution, Washington DC, pp. 41-71.

Morrison, Steven A. and Clifford Winston (2000), "The Remaining Role for Government Policy in the Deregulated Airline Industry," in: Sam Peltzman and Clifford Winston (eds.), *Deregulation of Network Industries*: *What's Next?*, Brookings Institution, Washington DC, pp. 1-40.

Morrison, Steven A. and Clifford Winston (1999), "Regulatory Reform of US Intercity Transportation," in: Jose A. Gomez-Ibanez, William B. Tye and Clifford Winston (eds.), *Essays in Transportation Economics and Policy: A Handbook in Honor of John R. Meyer*, Brookings Institution, Washington DC, pp. 469-492.

Morrison, Steven A. and Clifford Winston (1997), "The Fare Skies: Air Transportation and Middle America," *The Brookings Review*, Fall, pp. 42-45.

Peoples, James (1998), "Deregulation and the Labor Market," *Journal of Economic Perspectives*, Volume 12, Summer, pp. 111-130.

Savage, Ian (1999), "The Economics of Commercial Transportation Safety," in: Jose A. Gomez-Ibanez, William B. Tye and Clifford Winston (eds.), *Essays in Transportation Economics and Policy: A Handbook in Honor of John R. Meyer*, Brookings Institution, Washington DC, pp. 531-562.

Schwieterman, Joseph P. (2007), "The Return of the Intercity Bus: The Decline and Recovery of Scheduled Service to American Cities, 1960-2007," DePaul University, School of Public Service Policy Study, December.

Winston, Clifford (2006), "The United States: Private and Deregulated," in: Jose A. Gomez-Ibanez and Gines de Rus (eds.), *Competition in the Railway Industry*: *An International Comparative Analysis*, Edward Elgar, Cheltenham, United Kingdom, pp. 135-152.

Winston, Clifford, Scott Dennis and Vikram Maheshri (2008), "Duopoly Equilibrium Over Time in the Railroad Industry," Brookings Institution Working Paper, November.

Winston, Clifford, Vikram Maheshri and Scott Dennis (2009), "Long-Run Effects of Mergers: The Case of US Western Railroads," Brookings Institution Working Paper, May.

LONG-DISTANCE BUS SERVICES IN EUROPE: CONCESSIONS OR FREE MARKET?

Didier VAN DE VELDE

Delft University of Technology,
and
inno-V consultancy

Amsterdam
The Netherlands

SUMMARY

1. INTRODUCTION

Long-distance coach services are not the most glamorous part of Europe's long distance passenger transport system. High-speed rail or airlines attract much more political and media attention. Rail and air are much more visible and require much more (public) investment in highly visible infrastructures. Coaches on the contrary disappear in general traffic and do not require public investments, except perhaps in suitable coach stations at attractive places in urban centres. Yet, long-distance "express" coaches cater for a substantial part of the mobility of Europe's less-wealthy citizens, at least in those countries that have appropriately (de)regulated this branch of activity.

Few international studies have been published on this topic. The report from the 114th Round Table organised by the ECMT in 1999 (ECMT, 2001) was one such study, covering Britain, Poland, Sweden and the Eurolines organisation. National studies on the topic are scarce too, except perhaps in Britain, Sweden and Norway – three countries with a well-functioning deregulated coach market.

This paper makes a review of the current situation in the interurban passenger transport market by coach in Europe, describing for a number of selected countries the regulatory setting, the main market actors, the main developments have taken place in the last decade or two and a number of resulting challenges, especially in terms of regulation. The paper starts with a chapter on country cases. The next chapter summarizes the main facts and trends that appear out of this review. The last chapter draws a few conclusions.

2. COUNTRY CASES

This section presents an overview of the regulatory setting and main market actors in a selected number of European countries. For each country, recent evolutions and a number of main challenges are also presented. A few countries have been selected to provide, together, a good illustration of the diversity and similarities on the interurban passenger coach market in Europe, with a focus on the Western part of Europe. These countries are: Great-Britain, France, Germany, Spain, Italy, Poland, Norway and Sweden. The presentation of each country focuses on the national interurban coach operations, being for most countries the main part of the market. International coach services are another substantial part of the scheduled coaching business in Europe. We devote a separate section to *Eurolines* as a main part of the international passenger coach services takes place under the flag of this brand.

2.1. Scope and definitions

The passenger transport services reviewed in this paper are long-distance, scheduled passenger coach services. Only regular, scheduled passenger transport services are covered, meaning that the touristic coaching sector, or private hire and organised package tours, are not covered here. The services that will be described here are services open to everyone and operated according to a published timetable, i.e. similar to local public transport, trains and airplanes.

The words "long-distance coach services" need, perhaps, some further definition in view of the diversity that can be encountered across Europe. "Long-distance coach services", also called "express buses" or "interurban coaches", have in common that they cater for transport needs outside urban agglomerations, usually from city to city, often also serving towns not well served by rail on their way. Operations are generally done with coaches, not by buses, although the concepts "coach" and "bus" do not necessarily exist distinctly in the various European languages. The exact definition of long distance coach services varies also from country to country. This would not be very interesting if it were not for the fact that these definitions also determine the regulation under which services will fall. Distance is often a main criterion to fall under the regulatory regime applicable to long distance coach services, but these distances can be highly different: over 15 miles in Britain, or over 100 km in Sweden. Other countries often adopt an administrative distinction, where long-distance is defined as those services crossing the borders of the regional transport authorities, as in Italy, Norway or as in Sweden, where both definitions are combined.

It would be nice to be able to compare the size, modal shares and modal shift in interurban passenger travel in Europe. Unfortunately, statistics of national and international interurban passenger transport are difficult to compile. Numerous differences in definition exist from country to country, making international comparisons hazardous. Census data for travel surveys often do not include trips made by foreign nationals. Local and regional buses are often aggregated with coach statistics, making this data rather useless for the purpose of the analysis presented here. Differences in the definition of what "interurban" is, make international comparison of modal shares unreliable. The following table should therefore only be seen as a mere illustration of the limited size of mobility by bus and coach compared to the share of mobility by car. It is also striking to see the similarity of the modal shares for bus (and coach) and for train. However, "bus" includes here urban buses, regional buses and interurban coaches. While the share of interurban coaches could be 50% or more of the total of the category "bus" – this is probably the case in Spain with its extensive coach network and relatively limited rail services – this percentage can however be much lower in other countries.

Table 1. **Modal shares (in passenger-km)**

Country	Year	Bus	Car	Train
EU-15	1997	8.9	84.5	6.6
EU-15	2007	8.7	84.1	7.1
EU-25	1997	-	-	-
EU-25	2007	9.3	83.6	7.1
Germany	2007	6.4	85.8	7.8
Spain	2007	13.9	80.9	5.2
France	2007	5.5	84.9	9.6
Italy	2007	11.9	82.4	5.7
Norway	2007	7	88	4.9
Poland	2007	9.6	83.6	6.8
Sweden	2007	7.2	84.1	8.7
UK	2007	6.3	87.3	6.4

Source: Eurostat (2009).

United Kingdom

The British coach market was fully deregulated by the 1980 Transport Act. The only requirement (besides licensing requirements) to create services was that an authorisation was to be requested 28 days before starting the operations. This 28-day prior authorisation was scrapped under the 1985 Transport Act, hence no prior notice is now needed for operation of an express route.

Today, *National Express* is the main supplier of express coach services. *National Express* was privatised in the 1980s and was the former monopolist on the express coach market as part of the former *NBC (National Bus Company)*. Most services are operated by local contractors under the *National Express* brand rather than directly by *National Express* with its own staff.

The deregulation provided intense competition in the 1980s on some relations. A company called *British Coachways*, grouping six existing operators, attempted to establish a network to compete with *National Express*, but this failed as early as 1983 (Robbins, 2007). By the middle of the 1980s, most competitors of *National Express* stopped their services as innovations such as lower fares or higher comfort had been copied by *National E*xpress. Additionally, *National Express* initially had access to most of the coach stations and refused access to other operators. An important issue at the time was control on the *Victoria Coach Station* in central London. The coach station was subsequently transferred to the control of the London transport authority (now *Transport for London*).

The decline of competition in the 1980s after a strong competitive period resulted in a *de facto* monopoly by *National Express.* This disappeared in 2003 with the arrival of *Megabus.com* as a no frills, low-cost, coach brand provided by the *Stagecoach group* – another main player in local public transport in the UK. *Megabus* strives to differ from *National Express* by concentrating on low income target groups such as students, young people without driving license, or elderly people. Its network is less extensive, has lower frequencies and focuses on main relations between London and main cities. It uses pre-booking with yield management in its pricing strategy, copying the success observed in the low-cost airline business and to some extent by *National Express*. *Megabus* also tries to get closer to its target groups by remaining outside the coach stations used by *National Express* and by stopping closer to where its target groups are located, on the curbside or on university campuses. A higher propensity of the target groups to use the internet, and the strategy of *Megabus* to sell tickets via the internet meant that access to a coach station, as a central information and access point to the coach network, became less essential (Robbins, 2007), furthermore it also contributed to save costs.

Coach travel in Great Britain represents a substantial share of mobility, but a clear accounting of the market share of the coach sector is difficult as statistics tend to combine (local) bus services in the accounts.

National Express remains dominant in this market, despite the entry of *Megabus.* A study of the competition between *National Express* and *Megabus* on the relation London – Bournemouth, which also showed that car ownership and access to car usage is about 50% lower for *Megabus* users, calculated a market share on this relation of 79% for *National Express* and 21% for *Megabus* (Robbins, 2007).

One of the main challenges for new entrants on the British coach market was to find a niche that *National Express* had not yet occupied. Another challenge was the implementation of appropriate channels for ticket sales. The increasing usage of the internet was a chance for *Megabus.* It facilitated market access for the company as it did not have to rely on the access to existing travel agents, where *National Express* already had an advantage. Most of the ticket sales of *Megabus.com* are now done through the Internet (Robbins, 2007).

Sweden

Long-distance coach services are defined in Sweden as those running at least 100 km and crossing at least one county border. This market is now completely deregulated and non-subsidised. Deregulation took place in two steps, where the first step involved a reversal of the "burden of proof". From 1993 on the national railway carriers *SJ* had to prove that the opening of a coach line would damage seriously the railway business, or counties had to prove that it would seriously damage county bus routes (contracted and subsidised) rather than the entrant having to prove that it would damage neither the railway nor the regional bus services. Although this first step was neutral from an aggregate welfare point of view, it also led to gains for low income (low value of time) customers, while the railways lost a little revenue (SIKA, 1997; Jansson *et al.* 1997). The second deregulation step took place in 1999 with a full deregulation of the market with, however, a continued possibility for the country passenger transport authority to prevent coaches from picking up and setting down passengers in certain cases when these travel only within their area of authority.

Three main players dominate the Swedish market, providing 79% of the total supply in coach-km. A further 25 operators also provided long-distance coach services in Sweden in 2007. The main coach station in Stockholm catered for 25 operators serving 53 routes in 2002 (BR, 2002). *Swebus Express* is the main operator in Sweden. It is owned by *Concordia Bus*, a Swedish company active in all Nordic countries. Interestingly, *Concordia* originates in Norway, as a joint venture between a Norwegian regional operator in Oslo and *National Express* in 1997. *National Express* sold its share in 1999 after which *Concordia* acquired *Swebus* in 2000 from the British *Stagecoach. Stagecoach* itself had bought *Swebus* from the Swedish state railway *SJ* in 1996. *Concordia Bus* states, in its 2008 annual report, having a market share of 50% in the express coach market, and a market share of 5% in collective transport, with rail having 75%. *Svenska Buss* is a co-operation company owned by five regional Swedish bus and coach operators. *Säfflebuss and Bus4you* are now owned by the Norwegian group *Nettbuss*, itself part of *NSB* (the national Norwegian rail operator) and operate under the name of *GoByBus*. *Ybuss* is a Swedish privately owned operator co-operating with *Swebus Express.* Many other regional operators exist.

About 90% of the Swedish long-distance coach services are run on a commercial basis, the rest is run under contract from a transport authority. Coach services represented in 2005/6 about 5% of the number of long-distance trips and 6% of the passenger-km. Cars (66% in passenger-km in 2005/6) and train (about 15% in passenger-km in 2005/6) both have a larger market share in trips and in mobility, as do airplanes (11% in passenger-km in 2005/6) in terms of kilometres only, as few people use buses for very long-distances (above 600 km) (SIKA, 2008). The relative share of collective means of transport increased in the period 1993-1998, in line with the deregulation of both inland air traffic and long-distance coach services. While mobility increased by 9% during this period, that of collective transport grew by 13%. This evolution was reversed between 1998 and 2004 (growth of 9% for collective means of transport while total mobility grew by 13%), but the mobility by bus remained stable during that period, illustrating a shift from air to rail (Banverket, 2006). As in other countries, passengers are mainly students, elderly and low-income population groups.

Long-distance coach services are often included in regional fare integration schemes managed by the county passenger transport authorities. This means that local customers can use the long-distance buses as part of the total regional network and under the same fare conditions. This constitutes an interesting additional source of revenue for the coach operators, while it constitutes an interesting additional service for the customers of the county passenger transport authority.

The results of the deregulation started in 1993 and fulfilled in 1999 are perceived to be positive. Coach services are seen in Sweden as a welcome addition to the rest of the public transport system. Deregulation has become a part of the Swedish passenger transport system and it will gain importance in the near future. The national railway system is currently being deregulated, with open access being implemented in a stepwise approach from this summer (2009) until 2010, in line with the European-wide deregulation and liberalisation of the international rail passenger market. Proposals for a deregulation of the local and regional bus transport are currently being discussed. If these plans go ahead, this is certain to have substantial influence on the possibilities for a further development of the express coach network in Sweden.

Norway

The current express coach services have evolved from the old authorisation regime, and the pre-existing local public transport services. These local public transport services were and are regulated by the counties, and appeared – historically – on the basis of route authorisations initiated by operators. Many of these routes are subsidised by local authorities. Competitive tendering is also used since the end of the 90s for unprofitable area- or route-based contracts.

The current regulatory regime for long-distance coach services is in place since 2003 and represents an almost complete deregulation. It is the result of a gradual liberalisation that began in the 1990s after new initiatives for route co-operation by existing operators started at the end of the 1980s. The express services developed from existing transport companies initiating new and faster transport services crossing the boundaries of their traditional (county) areas to provide more attractive bus connections. The extension of services mostly occurred in partnerships between the transport companies involved in the areas served. Other routes resulted from the extension to Oslo of the former long-distance rail feeder routes.

In the first instance, these initiatives lead to some resistance from the side of the authorities, who feared the weakening of local public transport, and from the national Norwegian railway company *Norges Statsbaner (NSB)* which also wanted to avoid competition. However, the introduction of interurban express coach services through this market initiative lead to a high popularity amongst users (Leiren and Fearnley, 2008) and the fears for excessive competition between coach and rail appeared unfounded (Hjellnes COWI, 1999).

Whereas route authorisations used to be issued by the national government, this competence was decentralised to the counties. In practice, all requests for authorisations are granted as long as quality standards of operations are fulfilled. The counties can, though, impose some regulation to protect subsidised local public transport services. However, it seems that in many cases counties have adapted the local services to the existence of the express services and chose to "buy" specific additions to the express services to fulfill local needs (school transport, lengthening routes, etc.)

Today, most interurban coach services in Norway are organised via *NOR-WAY BUSSEKSPRESS*, which is a marketing organisation owned by 40 member companies running the different coach lines, some of which are run in co-operation with one another. The members are each responsible for the design of the services regarding timetables and fares. The main competitor to *NOR-WAY BUSSEKSPRESS* is *TIMEkspress*, a coach brand of *NSB* in Southern Norway. The services are run by *Nettbuss*, the coach operator of *NSB*, which also offers further coach brands like *Komfortbussen*, *Bus4you* and *Flybussen* (the airport express bus brand of *NOR-WAY BUSSEKSPRESS*). It should be noted that *Nettbuss* also runs further interurban services for *NOR-WAY BUSSEKSPRESS*. Other competitors are *Lavprisexpressen* and *Konkurrenten*.

The main part of the express coach services is run commercially and market access is *de facto* free. The current express coach network is seen more as a useful complement than as a competitor to the rest of the public transport services. Studies conducted in Norway showed that most passengers are new or attracted from using the car, rather than from train and airplane services (see Hjellnes COWI, 1999; Strand, 1991). Studies also showed that public transport usage has, on the whole, increased on the corridors with train/coach competition. In total, ridership has more than doubled between 2002 and 2007, while productivity has reached levels significantly higher than in neighbouring Sweden, that has also deregulated its market but with less possibility for co-operation between operators (Alexandersson *et al.,* 2009).

Policy documents, such as the National Transport Plan 2010-2019 reiterate that these services are welcome additions to the public transport system, as it allows servicing areas that would otherwise not benefit from public transport. Also, it is seen to contribute to a better environment and to fewer accidents by reducing car traffic.

Leiren and Fearnley (2008) identify two challenges that currently face the express coach market. The first is the subsidization issue. Express coaches do receive some public payments in a number of cases (providing fare rebates, some pupils transport or service to areas that would otherwise not be served). The current European regulation prescribes to submit subsidised services to competitive tendering. The simple application of this rule would threaten the nature of the industry, replacing it with a more centrally planned system. The regulation can however be limited to those services exceeding some limits regarding the height of subsidy or amount of kilometres produced. An additional element is that some authorities report that buying additional services from commercial long-distance operators proves to be cheaper than organising a separate local contract (competitively tendered) (Leiren and Fearnley, 2008). A danger, though, of having too much such influences on the long distance coach market is that this could lead to less attractive services, effectively taking the economic basis for those services away. Cleary, further political decisions are required here.

Another issue mentioned by Leiren and Fearnley (2008) is that the Norwegian competition authority (and then also European bodies) started to investigate some existing co-operation between operators. Yet, co-operation is often perceived to be beneficial. A study by the Norwegian Transport Economic Institute (Leiren *et al.* 2007) showed that the coach network, including the routes with co-operation within that network, lead to substantial welfare gains (NOK 1.5 billion per year). The National Transport Plan 2010-2019 now mentions an intention to exempt the co-operations that appeared before 2003 from this control. Here, too, further (political) choices will need to be made to decide how this issue should be regulated in the Norwegian context.

Poland

Public transport in Poland was, prior to 1990, organised in a similar fashion to that in other former communist countries. The State Road Transport (*PKS*) was the main carrier of passengers and goods by road. Before 1988, passenger transport operations in more than one region ("voivodship") required a permit from the Ministry of Transport. Permanent permits were only given to state carriers. Other road operators' access to the market was limited to single or periodic permits. It was the Act on Economic Activity (1988) that liberated many fields of activity, including road transport (Taylor and Ciechanski, 2008).

PKS was split into four state-firms in the early 1980s: one national PKS and three regional companies. The organisational structure of PKS counted numerous local branches receiving subsidies from the state budget. In 1990, the four firms were disbanded and all 233 branches became individual enterprises (Taylor and Ciechanski, 2008). There was little interest from foreign investors. Less than half of all firms were subsequently privatised, the most popular form involving a privatisation to the company's employees. The only main international concern interested was Veolia, which has taken control of 11 PKS companies as of mid-2006 (Taylor and Ciechanski, 2008). The limited interest in privatisation by foreign investors could be linked to the rapid decline in ridership. By 2005, public transport ridership was only one-third of the 1989 figure, due to the extensive development of individual motoring (Taylor and Ciechanski, 2008)

Little competition appeared at the national level. A new company started in 1994: *Polski Express*, as a subsidiary of Britain's *National Express Group* and targeting mainly connections not well served by rail. This company experienced serious economic difficulties later on (Taylor and Ciechanski, 2008).

In the late 1990s, real competition came from private "independent" operators having small numbers of buses, usually of lower standard, serving the most profitable routes. These activities led to a worsening of the economic situation of local PKS companies (more involved in local and regional transport). In some areas, local PKS companies went out of business. PKS remained, though, dominant, accounting for 92% of passengers and 95% of scheduled bus and coach services in Poland (Taylor and Ciechanski, 2008).

Also, *Polbus-PKS* was created in 1995 as a reaction to *Polski Express*. *Polbus PKS* was set up by 21 PKS companies and a couple of private companies as a marketing company, inspired by the example of *NOR-WAY Bussekspress.* It aimed at providing a modern coach network for domestic services, with a unified sales and information system throughout the services of its member companies across Poland. The company started providing long-distance services, especially where rail links were unattractive (Taylor and Ciechanski, 2008). *Pekaes Bus*, set up in 1996 as spin-off of *PKS*, also provided long-distance services. It was subsequently taken over by *Veolia Eurolines Polska*.

Komornicki (2001) reported on the substantial supply of semi-legal and illegal bus connections between Poland and neighbouring countries at the beginning of the 1990s. He reported that this problem (lack of quality certification, accidents, etc.) was considerably reduced from 80 to 20% of the market by 1998. It would be interesting to know whether the issue has completely vanished, now Poland has become a member of the EU, and whether the issue has reappeared further east.

Spain

Long-distance concessions are granted by the national government on an exclusive basis. The length of those concessions varies between 8 and 20 years. Regional inter-urban bus concessions are awarded by regional governments. In both cases, contracts are now mainly granted by means of competitive tendering, although direct contracting is/was possible in some circumstances, but mainly in urban transport. Until 1990, both long-distance and inter-urban services were under state control and concessions were awarded directly, without tendering. A reform was introduced with the decentralisation to the Autonomous communities (Regions) and a reform of the passenger transport legislation (in 1987 and 1990). As a result of this, the 113 existing concessions for long-distance services (all of them not subject to tendering) could be extended until at least 2007, most were extended until 2013 and some until 2018. New concessions for the provision of services on routes insufficiently served, or replacing illegal lines, have to be awarded through public tendering.

Numerous coach operators exist on the Spanish market. In 1988 *ENATCAR* was created as a public company, taking over all coach services of the national railway carrier *RENFE*. This operator was subsequently privatised to *ALSA*, who is the main supplier of long distance coach services – with nearly 10% of the market – and offering a wide range of differentiated services. The company is privately owned, member of *Eurolines*. It was sold to British *National Express Group* in 2005. *ALSA* is expanding its influence in the Spanish bus sector, integrating the second national transport operator *Continental Auto* in 2007.

Coaches have traditionally a strong position in Spain’s long-distance public transport market. Reliable statistics seem to be absent, but the coach market is believed to be four times larger than that of the train when measured in passenger-km (García-Pastor *et al*. 2003). The further development of the Spanish high-speed network may bring a change in this situation though, as did low cost airlines. The 2008 annual report of *National Express*, the owner of *ALSA*, mentions these competitive pressures and their response to the entry of low-cost airlines and the development of high speed rail by varying their frequency, adapting their prices and altering their network to provide complementary services. Furthermore, they also announced the launch of new services, with revised on-board catering and offering on-board WiFi, being the first transport mode in Spain to offer this facility.

Despite the usage of competitive tendering, Spain’s long-distance coach services are all profitable, in the sense that they do not receive public subsidy. García-Pastor *et al.* (2003) report that, according to a study for the Spanish Ministry of Development (Consultrans, 1999), the competitive tendering initiated in the 1990s did, however, have positive effects on service quality and ticket prices for these concessions. According to this study, extended concessions appeared to have 46% higher passenger-km fares than that in tendered concessions.

Italy

The Italian interurban coach market can be divided into national and regional services. National interurban coach services *(linee extraurbane statali)* operate commercially on routes of 200-1 200 km between the larger cities located in different regions. The legal regime applicable to those services has recently been modified, with a decree from November 2005 aimed at opening the market. Services do not receive any subsidy. Coach operators now have to apply for an authorisation at the ministry before starting new services. While the former regime did not allow competition, the new authorisation regime is supposed to make competition on the road possible. However, little competition seems to have taken place since, and successive changes in government seem to have delayed the deregulation.

Regional interurban coach services *(linee extraurbane regionali)* serve routes of 30-300 km between larger cities located within the same region. Most regional routes are still directly awarded concessions, subsidised and have regulated routes and fares. Some deregulation is also planned here as some regions have developed regional legislation that follows the national decree. However, this competition seems often restricted to those services that do not interfere with existing subsidised regional services operating under concession contracts.

Operators differ significantly in size. No national operator dominates the market at the moment. *Sitabus,* as a large operator, is owned by the national train operator *Trenitalia*. One of the largest wholly privately owned operators is *Arriva Italy*. The rest of the market is highly fragmented, with a large number of local operators, often owned by regions and municipalities, but privately owned operators exist too. Most operators are based in one region and offer, in addition to regional services, connections with Rome or other main Italian cities. Few companies offer nationwide services. *Arriva* entered the market by taking over 11 regional companies, many of which operate in the interurban market.

Some services are supplied in complement to the existing high-speed trains of *Trenitalia*. *Sita*, as a subsidiary of *Trenitalia*, is a main supplier of such services. Some services include high-quality seats and on-board internet facilities. Other services are directly competing with long-distance train services. Fares on those routes are comparable to the (highly subsidized) railway fares on those relations from the North to the South of the country. These fares are generally lower than domestic airfares.

Recently, *Ibus* was initiated, a co-operation between nine operators integrating marketing and ticketing activities, and also member of Eurolines.

France

There are essentially no long-distance express coach services in France. The regulation of public transport is allocated to the State for interregional passenger transport services, and these are the monopoly of the national railway company *SNCF*. Regional and local transport services are organised by the *départements* (to be compared with counties) and by (co-operation of) municipalities. Most of these services are submitted to competitive tendering. Express services exist at the level of the *departments,* when ordered by the respective transport authority, but no services are operated on a national scale on real long distances.

As a result of this, and although some competition exists between *Sniff*'s train services – in particular its *TV* high-speed train services – and the airline business, there is no such competition between rail and road. Market entry by market initiative by individual transport operators is, for the time being and since the enactment of the current transport legislation in 1982, not foreseen and prevents explicitly direct competition to *SNCF* services. As a result, such entry is *de facto* impossible. Gaining the agreement of *SNCF* seems illusory, as the company has always been opposed to the idea.

This may change in the near future, as the current political majority announced, in July 2009, its intention to introduce a number of amendments to the current legislation to allow international coach services some degree of sabotage on the French territory and, more importantly, to allow a full liberalisation and deregulation of the long-distance coach business at the national level. The idea would be to introduce a system of largely deregulated authorisations, which would effectively abolish the monopoly of *SNCF* by the end of 2009. Incidentally, this would also be the moment when the liberalisation of international railway services decided at the European level would come to force.

With this in mind, *SNCF* reportedly started changing its mind, perhaps seeing also some opportunities for its own bus and coach subsidiary (*Keolis*), that is currently expanding its activities not only in France but also in the rest of Europe. Additionally, *SNCF* may benefit from replacing some of its loss-making interregional services by more profitable coach services (Kramarz, 2009). The expectation is that, as in other countries, students, less wealthy customers, and people with a lower value of time would be the main beneficiaries of such liberalisation. This, though, would require the appearance of a national network of services, which is perhaps still far-fetched. Yet, here too, some combination with existing, but less profitable, regional services may lead to win-win situations if the regulation allows such combinations.

Germany

The basic regulatory principle of the German express coach market is that of free market initiative by transport operators. The market is, however, strongly regulated by the National law on public transport. That law restricts direct, on-the-route, competition between transport operators and provides some protection to incumbent operators. Supplying new, more or less parallel, services is only allowed when these represent a significant improvement over existing services.

Today, there is an extensive interurban coach network with West Berlin as hub. Most of those services are a relic of the division of Germany. West Berlin, as part of the Federal Republic of Germany, was located inside the territory of the German Democratic Republic. The bus services provided connections between West Berlin and other cities in the Federal Republic These connections are operated by *Berlin Linien Bus*, a joint venture of various coach operators partly owned by *DB*. Most connections are served once a day. Every journey must have Berlin as starting point or destination. Services starting in Berlin cannot be boarded at other stops, and buses to Berlin can only be left in Berlin (Maertens, 2008).

Other providers are *Touring* and *Public Express*. *Touring* – owned by Eurosur (a joint venture of the Spanish and Portuguese bus operators *Alsa*, *Linebus* and *Socitransa*) operates a night service from Hamburg via Kassel, Frankfurt and Darmstadt to Mannheim. Other services mostly go to other European countries. These represent most of the services performed by *Touring*. *Touring* runs these under the flag of *Eurolines.* Further national services are provided by *Public Express*, who offers coach services between Bremen, Oldenburg and Groningen in the Netherlands, and also between Bremen and Aurich. Another segment of regular coach services are airport express buses. Many regional airports are served by such coach services (Maertens, 2008).

The evaluation of the potential of interurban coach services shows that interurban coaches would provide travel possibilities for people with lower incomes (Maertens, 2008). The services of current suppliers like *Touring* and *Public Express* confirm this. It is obvious that *Public Express*, for instance, focuses on students and families. Students receive discounts of approximately 50%. Adults travelling with children are allowed one child to travel for free. *Touring* wants to attract travellers by offering low fares for those who book the journey well in advance (Touring, n.d.).

Suppliers as *Touring* and *Public Express* show that there are market parties who want to expand but are hindered by the current law. A problem with the existing market entry regulation is that the required level in quality of improvement remains unclear. The extension of the existing German interurban coach services is also confronted by resistance coming from the established passenger transport branch. This causes long court trials between operators of new services and incumbents. Consequently, the efforts of *Touring* to extend its interurban coach connections are hindered by claims from the German railway company *Deutsche Bahn* (DB). A recent example is the attempt by *Touring* to open a coach service competing with *DB* on the Frankfurt-Cologne route (Köhler, 2009). However, one can also observe that in 2009 a new company, *AutobahnExpress*, has managed to obtain a number of authorisations for routes linking Potsdam, Dresden, Leipzig, Halle, Kassel and Göttingen via motorways.

Two political parties (the liberal democratic FDP and the Green Party) tried to promote the idea of deregulation for interurban coach services in 2005 and 2006, but the Parliamentary Committee on Transport, Building and Urban Affairs rejected both requests (Maertens, 2008). However, deregulation of interurban coaches is now likely to go ahead as the recent coalition agreement of the Federal Government (CDU, CSU and FDP), published in the autumn of 2009, includes the formal intention to deregulate this market.

Eurolines

Eurolines is a joint venture of European coach operators which organises most of the international coach services inside Europe. The brand name *Eurolines* groups 35 independent coach companies, operating in 32 countries and providing together Europe's largest regular coach network. *Eurolines* developed common quality standards for all its members, and harmonized the sales and travel conditions. The network currently connects over 500 destinations, covering the whole of the continent and Morocco.

Eurolines was founded in 1985 as a competitor to *Europabus*, that had been created by several European rail companies in 1965 to prevent other coach operators from competing with their rail services (Eurolines, 2008). While the transport services of *Europabus* remained limited, *Eurolines* developed its market by providing services on international relations with significant demand, starting with the travel needs of migrant and guest workers coming from Spain and Portugal. Various initiatives in the various countries, such as *Budget Bus* in the Netherlands (and many more examples in other countries), were eventually bundled together under the common flag of *Eurolines* as marketing brand for regular international services.

The *Eurolines Organisation* is an International Non-Profit Organisation, according to the Belgian law. Membership is open to (groups of) companies operating international scheduled passenger services by coach. Decisions concerning *Eurolines Services*, as commercial daughter of the *Eurolines Organisation*, are made by a council of directors of all *Eurolines* member companies. Next to this council, an executive committee – consisting of nine directors of the member companies – guides the implementation of new product developments. A main challenge for *Eurolines* has been the differences in national legislation pertaining to the operation of passenger coach services as, for example, differences in fuel taxes and the rules for value added taxation created a lot of bureaucracy (Bochar, 2001).

It is interesting to note that *Veolia* has now acquired a significant position in *Eurolines*, as it owns the brand in Belgium, France, the Netherlands and Portugal and operates the brand, in partnership, in Scandinavian countries, Poland and Spain.

3. MAIN TRENDS AND CHALLENGES

Before drawing a few conclusions in the next chapter, this chapter will summarize the main facts and trends that appear out of the countries reviewed above. Where relevant, a few challenges will also be sketched.

3.1. Organisational forms in long-distance passenger transport

There are two main families of organisational forms for passenger transport services. The first one, that I call "market initiative" regimes, are those organisational forms where it is essentially transport operators that come up with ideas of markets to be served (van de Velde, 1999). Operators are free, in such regimes, to suggest new services and request permission to operate them. In its pure form, the request for permission is a mere formality. "Authorisations" to operate (sometimes called "licences" and sometimes, unfortunately, called "concessions") are then granted without further analysis by transport authorities of whether the market "needs" the additional service, whether another operator already provides similar services, whether fares are appropriate, etc. But this regime can also be combined with various forms of regulatory interventions, limiting the free access to the market by various requirements pertaining to the non-parallelism to existing services, to fare integrations, to the protection of railway rights, etc. The principle remains that of market initiative, but regulation can be so tight as to effectively prevent any entry.

The alternative for market initiative is "authority initiative" (van de Velde, 1999). Here, it is a transport authority that is charged with the creation of the transport services. The authority can then provide the services itself, with its own staff or company, or it can concede these services to an operator of its choice, which then usually takes place via competitive tendering. The essential difference with "market initiative" is that this regime prohibits any spontaneous initiative from market actors. It grants all rights of service creation to the authority. If the authority does not take any initiative, nothing happens and nothing can – legally – happen. The private sector can be involved, but this requires the authority first to realise that a transport service is needed, then to specify its characteristics (in a more or less detailed way) and finally to organise a competitive procedure (tendering) in order to award the service to an operator under contract for a specific period of time. Such contracts are then called "concessions" or sometimes, confusingly, "franchises". They are often exclusive, but this is not necessarily the case.

In the countries presented, and as can be seen in Table 2, deregulated market initiative is clearly dominant. France and Germany, two potentially main European markets, are still effectively closed but an opening of both markets is expected. Concrete steps for liberalisation and deregulation are being taken in France. This is quite striking, as France has organised the rest of its local and regional passenger transport system on the basis of a strict authority initiative and competitive tendering regime that leaves no space for the free market. But this move will require a change in the current legislation.

Germany, too, is likely to move to a deregulated regime as this is now included in the new coalition agreement for the Federal Government. Existing market actors are pressing in that direction: a few routes already exist as the existing legislation does not make it fundamentally impossible – although quite difficult – to enter the market at this moment. This means that it should be a relatively simple move to realise as soon as all actors agree. A main resulting change will then be the necessary abolition of the railway protection that is currently fiercely defended by *DB*. It could be that the opening of the international railway markets in 2010 will also facilitate that shift in position.

As a result, Spain appears to be the only country in the sample (and apparently also in the rest of Europe) that bases its regime not on the free market, but on a regime of concessions awarded by competitive tendering. It should be noted, though, that the current concessions are more the result of historical rights that were probably initiated by the market, and not the authority, at their origin. Some of the existing concessions have already been submitted to tendering, but the bulk is still to come.

At the level of the operators, the organisational form is characterised by a diversity of arrangements, many of which are hybrid. Operators provide services with their own vehicles and staff, but also often make use of sub-contractors (such as *National Ex*press in Great-Britain). This can be observed in most countries analysed. This, incidentally, allows smaller family-based operators to participate in larger network services. But smaller operators are also present individually on the market, as can be seen in, for example, Sweden and Italy.

Table 2. **Organisational forms in long-distance coach transport in Europe**

Authority initiative		**Market initiative**	
Public sector	**Private concessions**	**Regulated**	**Deregulated**
• France: long-distance monopoly for rail (SNCF), no long-distance coach services, competitively tendered coach services at the local/regional level	• Spain: exclusive concessions awarded increasingly by competitive tendering to private operators, both at the national and regional level	• Germany: market initiative in theory but almost no possibility of entry in practice to due protection of railway monopoly	• Great-Britain: open market, free competition, no railway protection • Sweden: open market with some regulatory rights for regional transport authorities, no railway protection • Norway: open markets with some regulatory rights for regional transport authorities, no railway protection • Italy: open market, but still in starting stage • Poland: open market but still many state-owned operators present.

Another dominant feature of this market is the “marketing co-operation”. Individual operators, conscious of the existence of demand-side network effects present in this industry, bundle their products under an attractive brand name – allowing them to realise a wider service coverage and higher product attractiveness together than would be the case as an isolated provider. *Eurolines* is the best example of this at the European level. *NOR-WAY Bussexpress* is another example at the scale of one country. Further examples can be found in Sweden. At the extreme, the marketing co-operation, which is a type of commercial franchise (contrary to the usage of the word "franchise" in the context of competitively tendered concessions), approximates the model of the main operator subcontracting most of its operations to local operators. Yet main differences exist, and lie in the balance of power and attribution of risks between the small contractors and the main contractor, or the assembly of operators in the case of co-operation.

3.2. Performances

As can be read in the cases presented, the liberalisation and deregulation of the coach sector is perceived to be a success in those countries that have implemented it. Some countries went for a "big bang" approach, as the UK in 1980. Others, very much in their tradition, went for a more gradual approach, as Sweden; or pragmatic approach, as Norway; or incomplete approach, as Italy. It remains to be seen how France and Germany will tackle to current reluctance to deregulate.

The competitive tendering alternative to the market, as used in Spain, also delivered good results according to the studies reported in the case study. Tendered concessions proved to have lower fares than extended (negotiated) concessions. This does not allow us to draw a conclusion on the relative advantage of tendering above a deregulated market, as one has to remember that the Spanish concessions are exclusive and therefore lack the competitive pressure present in countries such as Norway, Sweden, Poland or Britain. An international comparison of service levels, quality and fares would be needed to be able to judge this.

Contrary to most of the rest of public transport, long-distance coaches are operated on a commercial basis in the sense that subsidization is almost non-existent. The railway sector on the contrary, even protected from competition from the coach sector, mostly requires subsidization – if not directly in operations, then at least through part of the infrastructure expenses.

Another attractive aspect of the coach sector can be found in several publications in the fact that coaches produce little pollution per passenger-km, and reach a safety level in terms of accidents that is comparable to that of the train and airplanes – which is substantially lower than the car system (see, e.g., ECMT, 2001). The proponents of deregulation then combine this argument with the observation that in countries that have removed railway protection and deregulated the coach sector, coaches tend to capture more passengers from the car than from rail, to conclude that coach deregulation would be beneficial from an overall transport policy point of view.

3.3. Markets served

Providers of long-distance coach services focus rather clearly on specific target groups: students, elderly, people with no access to cars, and poorer people in general. Swedish and British studies have shown the advantage of deregulation for these groups, while showing at the same time the limited impact on the rail system in terms of passengers captured. Rail and coach seem to cater for people with different values of time in terms of long-distance travelling. Some studies even show that direct competition between both modes in one corridor tends to result in a growing market for both at the expense of the car.

In addition to this, coach tends also to serve quite successfully relationships that are not available by rail, in particular by providing direct links between airports and various areas.

Eurolines is obviously the main player for international services in Europe. Clear statistics do not exist, making the presentation of clear observations on this market difficult. Poland, and probably also the Czech republic and other former communist countries not reviewed here, appear to be main players in term of the European network of coach services. This has much to do with the current propensity of workers from Poland and other Central and Eastern European countries to seek job opportunities elsewhere in (mainly) Western Europe. Family visits, tourism and further exchanges are responsible for the growth of these markets since the fall of the Iron Curtain. As such, this development is very similar to market developments that could be observed several decades ago and that were responsible for the development of the coaching business at that time, especially in relation to the transport flows between Spain and Portugal and the rest of Europe.

Two main European countries still have only very limited coach services, except for the international *Eurolines* services: France and Germany. Changes are clearly overdue here, and political momentum is now building up for change in France and probably also in Germany.

Smaller European countries have, in view of their sheer size, limited or no long-distance coach services. This is the case in the Netherlands, Belgium of Switzerland. Denmark does have a few long-distance coach relations from Copenhagen to Jutland. It should be noted that the railway services are rather excellent in these countries for the distances considered. These countries are of course also well served by the *Eurolines* services.

3.4. Network effects, monopolies, barriers to entry and regulatory needs

Network effects need to be recognised in this industry. The marketing co-operations presented above appear out of the market process in a profitable, competitive and open market. This is an indication of their desirability. But it is also an indicator of their questionability from a regulatory point of view. The British, Norwegian and also the *Eurolines* case show the attractiveness of this concept for the passengers (information, ticketing, image, attractiveness, etc.) The sheer existence of co-operation between providers in a competitive market is prone to attract suspicions from competition authorities, as can be seen in the Norwegian case reported earlier or in the acquisition of the Scottish *Citylink* by *National Express* at the end of the 1990s (followed by a forced divesture) as reported by White (2008).

Yet, as already stated by the 1998 Round Table on Interurban coaches (ECMT, 2001), these co-operations and the resulting conglomerates of operators do not seem to lead to abuse of dominant positions. This is due to the strength of the intermodal competition with (mainly) the car, low-cost airlines and, to a lesser extent, rail. Furthermore, as exemplified by the *Megabus* entry in Britain, the market seems to remain sufficiently contestable in terms of intramodal competition. It is important to stress that this lack of concern can only be true inasmuch as entry barriers are appropriately removed. This relates to non-discriminatory access to coach stations, fair licensing requirements, and fair authorisation procedures. This also requires non-exclusive route authorisations, or a very clever authorisation-issuing authority (which is perhaps too much to ask for in many cases). Last but not least, it requires the enforcement of a fair access to existing marketing organisations where these have a dominant position on some markets. Indeed, competing marketing organisations could also exist, but their viability will very much depend upon the size of the market to be served.

The fair access to coach stations seemed to be more of a problem in the 1980s than it appears to be nowadays. None of the studies accounted for in this research mentioned coach station access as a main issue. Stations can be owned either by the public sector or by a main operator but need, and seem, to be accessible according to fair rates. Furthermore, the relevance of stations as places to find information and buy tickets has lost much of its relevance with the advances reached in internet sales in this sector.

While coaches are often operated through numerous local, small, family operators, one can also observe the continuous expansion of a few main European-wide operators. While the traditional model of small operators as sub-contractors of larger brand-holder or member of market association does not yet seem to be threatened, it will also be interesting to see whether this model will lose in importance and be gradually replaced by larger operators. The expansion of *Veolia*, as main French group, is currently very visible all across Europe. The British *National Express* is a second example, although less prominent. Earlier expansions of international groups, such as *Stagecoach*, have been witnessed in Sweden, but the events showed that these expansions could be very volatile. The future will tell, but a point for further study, in terms of regulatory preoccupations, is whether expanding large conglomerates pose a larger competitive threat to the coaching market rather than the co-operations *per se*.

The European Union adopted, in 2007, a new regulation on public service obligations in the passenger transport sector (Regulation 1370/2007) that is applicable to the end of 2009. In short: services granted exclusive rights or financial support must be submitted to competitive tendering. There are a number of (complex) exceptions to this, however (see van de Velde, 2008 for more details). Deregulated markets without exclusive rights are not directly affected by the Regulation. Furthermore, compensation for fare rebates can continue to exist, when accessible to all operators. This would mean that the long-distance coach sector is not affected by the measure as it operates without subsidy and without exclusivity. The main exception is Spain, but as competitive tendering has been chosen as the awarding mechanism in that country, it all seems compatible with the Regulation. However, as could be seen in, for instance, Sweden and Norway, long-distance coaches do not exist in isolation from other public transport services. As it happens, regional transport authorities in countries with low densities of population, and under whose responsibility local public transport falls, have many times discovered the mutual benefits that may exist between local buses and long-distance coaches. Integrating coach services with regular local services can realise service improvements (speed) for local customers, and can allow serving remote areas that would otherwise not be served if a long-distance bus did not stop on its way to the next remote main city. Combining both types of services often requires subsidization of the long-distance coach. A problem may appear here with the new Regulation when this amount, or the size of the contract, would exceed some threshold, forcing the authority to use competitive tendering, which would only be counterproductive in this case. There is no clear view at the moment on the extent of this problem, but it could constitute a (probably minor) challenge in the years to come if pragmatism cannot be used.

3.5. Towards further deregulation: challenges for the near future

The points of view and opinions in markets that are currently closed, such as France and Germany, have much in common with what could be heard in countries such as Sweden and Norway before their own deregulation: the railways needed to be protected against coach competition as the opening up of that market would result in losses of attractiveness for the rail system by substantial losses in passengers. The facts proved different in those countries; rail hardly suffered and coaches opened up new markets with people that could not afford the train anyway. Main opportunities are certainly present in these two countries as well. And France is, unexpectedly, likely to be the first of the two to open the market.

A new factor for the near future is the opening of the international passenger rail market for competition in 2010. The European Parliament forced this unexpected move. It is until now rather unclear how this will work in practice. International passenger train services will, from then on, be able to operate in an open-access regime. However, existing national public service concessions will benefit from some protection. It is as yet unclear how this will be interpreted in the various countries. Germany, as main promoter of this liberalisation, seems favourable to a simple opening. The position in France is likely to be much more restrictive. This deregulation of the international rail travel could have some effect on the international coaching business, although it is likely that here, too, just as for the deregulation of the national coach markets, coach and train will hunt for customers with different values of time.

4. CONCLUSIONS

The title of this paper is “Long-distance Coach Services in Europe: concessions or free market?” The review illustrates that the clear choice of most countries is for a free market, and neither for a system of competitively tendered concessions, nor for a regime of exclusive rights.

Those countries that have not yet opened up their markets are also more likely to move to a deregulated regime rather than a system of tendering concessions. In short: competitive tendering does not seem to be the most favourite choice in this market. Deregulation has shown that it can work and the markets seem to remain sufficiently competitive, both in an intermodal and intramodal sense.

BIBLIOGRAPHY

Alexandersson, G., N. Fearnley, S. Hultén and F. Longva (2009), "Impact of regulation on the performances of long-distance transport services: a comparison of the different approaches in Sweden and Norway", *11th International Conference on Competition and Ownership in Land Passenger Transport*, Delft, the Netherlands, 20-25 September.

Banverket (2006), "Järnvägens roll i transportförsörjningen", Banverket, Stockholm.

Bochar, D. (2001), "Eurolines or the pan-European coach network of regular lines services: an introduction", In: ECMT, Economic Research Centre Round Table 114, *Regular Interurban Coach Services in Europe*, pp. 7-44, OECD, Paris.

BR (2002), "Expressbussen efter avregleringen - Utveckling och erfarenheter", Svenska Bussbranschens Riksförbund, Stockholm.

Consultrans (1999), "Present and future situation of the concession system of road passenger transportation under the control of the General Administration of the State", Report for the Spanish Ministry of Development, Madrid.

ECMT (2001), Economic Research Centre Round Table 114, *Regular Interurban Coach Services in Europe*, OECD, Paris.

García-Pastor, A., C. Cristóbal-Pinto, J.-D. González and M. López-Lambas (2003), "The Spanish situation of road public transport competition", *European Transport Conference*, Strasbourg, 8-10 October**,** 12 pp.

Hjellnes COWI (1999), "Evaluering v konkurranseflater for ekspressbussruter. Endelig rapport for empiriske undersøkelser", COWI, Oslo.

Jansson, K., I. Vierth and J. McDaniel (1997), "Economic analysis of the deregulation of coach services in Sweden through model simulations of historic & hypothetical competitive situations", *European Transport Conference*, Association for European Transport.

Köhler, M. (2009), "Vier Jahre Rechtsstreit um einen Linienbus nach Köln", *Frankfurter Allgemeine Zeitung*, 11 March 2009.

Komornicki, T. (2001), "The development of international bus transport in central Europe: the case of Poland", in: ECMT, Economic Research Centre Round Table 114, *Regular Interurban Coach Services in Europe*, OECD, Paris.

Kramarz, F. (2009), "Pour des bus Greyhound à la française ", *Les Echos* - *Blogs*, 8 July.

Leiren, M.D. and N. Fearnley (2008), "Express coaches - the story behind a public transport success", *European Transport Conference*, Leeuwenhorst Conference Centre, The Netherlands, 6-8 October.

Leiren, M.D., H. Samstad, N. Fearnley and H. Minken (2007), "Ekspressbusruter - ett samensatt marked", 904/2007, Transportøkonomisk Institutt, Oslo.

Maertens, S. (2008), "Intercity-Busverkehr in Deutschland – Notwendigkeit und Perspektiven einer Liberalisierung", *Fachgespräch Potentiale des Fernlinienverkehrs in Deutschland*, Münster, 2 June 2008, Institut für Verkehrswissenschaft.

Robbins, D. (2007), "Competition in the UK express coach market 25 years after deregulation: the arrival of Megabus.com", *European Transport Conference*, Leeuwenhorst (The Netherlands), 17-19 October 2007, Association for European Transport.

SIKA (1997), "Effekter av avreglering av långväge busstrafik", 1997:6, SIKA, Stockholm.

SIKA (2008), "Långväga buss 2007", SIKA Statistik 2008:21, SIKA, Stockholm.

Strand, S. (1991), "Konkurransen mellom tog og ekspressbuss", 0078-1991, Transportøkonomisk Institutt, Oslo.

Taylor, Z. and A. Ciechanski (2008), "What Happened to the National Road Carrier in a Post-Communist Country? The Case of Poland's State Road Transport", *Transport Reviews*, 28**,** 619–640.

van de Velde, D.M. (1999), "Organisational forms and entrepreneurship in public transport (Part 1: Classifying organisational forms)", *Transport Policy*, 6**,** 147-157.

van de Velde, D.M. (2008), "A new regulation for European public transport", *Research in Transportation Economics*, 22**,** 78-84.

White, P.R. (2008), *Public transport: its planning, management and operations*, 5th edition, Routledge, Abingdon.

LONG-DISTANCE PASSENGER RAIL SERVICES IN EUROPE: MARKET ACCESS MODELS AND IMPLICATIONS FOR GERMANY

Thorsten BECKERS[a], Fabian HAUNERLAND[b], Christian von HIRSCHHAUSEN[a,b], and Matthias WALTER[b,*]

[a]Technische Universität Berlin
[b]Dresden University of Technology
Germany

SUMMARY

ABSTRACT

This paper focuses on classifying market access for long-distance passenger rail services in Europe into three main models and discusses the advantages and disadvantages of each of these models. The “Tendered Concessions” model aims to introduce competition for the market by which operators are selected in a tendering procedure. The “Monopolistic Network Operator” model aims to sustain network effects by granting a concession to one operator. The “Open Market” model enhances operators’ entrepreneurship by providing opportunities to plan services based on open access to the network. We present the strengths and opportunities, risks and threats without favouring any one model. Classifying the many design options and their different impacts will help to structure the ongoing policy discussion. The paper also gives an overview of the organisation of long-distance passenger railway markets in selected European countries, and discusses the development of Germany’s long-distance rail passenger services in particular.

Keywords: Long-distance passenger rail transport, market access, open access, competitive tendering.

1. INTRODUCTION

The liberalization of European rail transport markets has been on the agenda of politicians, academics, and industry for the last 20 years. Whereas infrastructure aspects and freight transport were the primary focus, the regulation of passenger transport has largely been a secondary concern. Directive 91/440/EEC explicitly addressed the vertical disintegration of national railway incumbents, requiring at least accounting separation. The First Railway Package, effective after 2003, concentrated on improving the effectiveness of recent legislation. The Second Railway Package, effective after 2004, included, among other issues, safety and interoperability. The latest regulation, the Third Railway Package, contains Directive 2007/58/EC which aims at opening the market for international passenger rail services after 1st January 2010. Of importance for domestic travel, the Directive includes the possibility of passenger carriage within countries along international routes. Exceptions refer to the protection of routes served with public service contracts. Open access for domestic services is not mandatory, but is used to augment competitively tendered services (see Griffiths, 2009, for the British example). The key question in designing a model for market access is: should access to long-distance passenger rail markets be open, or via concessions, or possibly a mix? In this paper, we address this under-researched question.

The remainder of the paper is organised as follows: In Section 2 we give the necessary background on rail passenger transport in Europe. We consider the value chain, differentiate long-distance services from other services, and point to the issue of non-profitable interregional services.

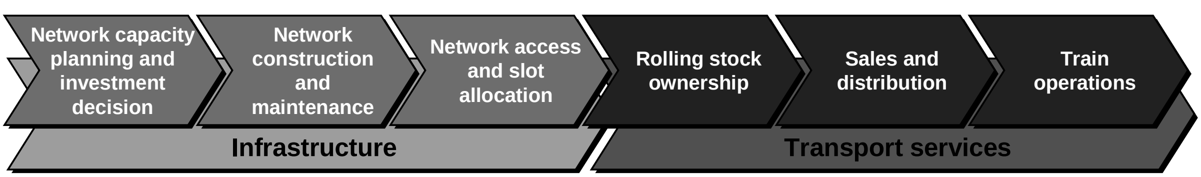
Furthermore we look at issues such as ownership of the infrastructure network and vertical separation, because of the interdependencies between the design of access to the infrastructure and the potential for competition for long-distance rail passenger services. Section 3 presents and analyses the different models of market access for long-distance rail passenger services using country examples. This includes a franchising model with a number of concessions as well as a single concession to a monopolistic network operator. An opposing model is the "Open Market" approach in which companies can introduce new services for any route for network slots they are awarded. Section 4 is dedicated to a more intensive case study of Germany, Europe's largest market for rail services. It includes an analysis of the current situation and an outlook on future developments. Section 5 concludes.

2. SECTOR BACKGROUND

2.1. The value chain and axes of competition

Analysing the different forms of market access for long-distance rail passenger services cannot be conducted without considering the other steps of the value chain in the railway sector, which are displayed in Figure 1. The infrastructure is a non-contestable natural monopoly consisting of network capacity planning and the investment decision, network construction and maintenance, and network access management and slot allocation. Much network construction and maintenance can be outsourced to a competitive market for construction, maintenance and renewal activities. The provision of transport services includes rolling stock ownership, ticket sales and distribution, and train operations.

Figure 1. **The value chain of the rail transport sector**



Source: Author's illustration.

A first point of contact between network activities and train operations is assignment of network access. In open access train operators must apply for network slots. Possible organisational forms for the responsible authority are integrating with the network operator or remaining a separate agency. Based on the allocation of network access, the next step is timetable preparation. It requires co-ordination with other rail services provided by the same operators or others.

Rolling stock is procured after the allocation of network access or a successful competitive tendering. The procurement process for new rolling stock is long-term, and markets for second-hand rolling stock hardly exist. Different technological requirements across countries and tracks make it difficult to resell stock. Rolling stock ownership represents a competitive market with significant entry barriers. To mitigate such effects, the rolling stock can be owned by independent companies founded only for this purpose, or by authorities.

Based on the allocation of network access and the availability of rolling stock, ticketing, sales, and marketing can be organised. Sales can be classified into off-board and on-board. In the case of longer travel times, on-board sales can be a reasonable option for passenger convenience and reduced transaction times. Further options are represented by a markup of on-board ticket prices, or the imposition of pre-booking for increased planning reliability. On the other hand, ticket vending machines throughout the countrywide network represent an essential facility which is generally too costly to duplicate. Moreover, a well-established and popular Internet platform for scheduling and sales is a competitive advantage. At issue, however, is how much these services should be centralised to provide fair, non-discriminatory access and to provide a united "face" to customers. The degree of state involvement in such a united face is another crucial point in the evaluation of the different models of market access.

The actual nucleus of the value chain is train operations. However, one should remember that around 50% of total costs are already predetermined by track and station access costs, energy costs, and marketing and sales costs (Monopolkommission, 2009, pp. 49 and 94; and Preston, 2008). Train operations can represent a state-granted monopoly or an open market. The quality and type of service are partly predetermined in competitively tendered services or can be freely chosen in any open access services. Quality and service of train operations are strongly related to the rolling stock and the tracks.

2.2. Differentiation between long-distance and regional services

The differentiation of long-distance and regional transport services is a crucial point in the setup of a model for market access for long-distance rail passenger services. Popular distinction criteria are represented by type of service, travel distance, and profitability.[1] Using type of service, all high-speed trains, intercity, eurocity and night trains are classified as long-distance, with the rest being urban, local, or regional.

Travel distance could classify all trips over a certain threshold, e.g. 50 km, as long-distance, depending on country characteristics. However, this would require complex data collection and the service classification could only be based on the majority of passengers.

A third possible distinction criterion is profitability. Urban, local, and regional services are usually characterised by some form of state provision, i.e. an enterprise in public ownership or public procurement. Through its nature as a public service obligation, the provider receives subsidies and can sell tickets at a price below cost recovery. To minimize these subsidies, countries such as Germany, Great Britain, the Netherlands, and Sweden use competitive tendering.[2] In principle, this procedure is also possible for long-distance services, and is another crucial point in determining how much the different forms of market organisation could allow improved integration of subsidy instruments.[3] Further differentiation of long-distance services is possible through different services classes, on-board service, stop frequency, and so on.

2.3. Handling of non-profitable interregional services

Since there are no generally accepted distinction criteria between long-distance and regional services, there is a gap into which unprofitable interregional services fall. The handling of such services is especially interesting when looking at different models of market access. In this respect it is important to consider public service contracts. Directive 2004/18/EC defines these as contracts between a service provider and a contracting authority. The term public service obligation is used for public service contracts that offer an auction for subsidies and award the winning company a monopoly to operate a specified route with subsidies for a specified period.

In countries where a concession for the entire long-distance network is given, presumably non-profitable sections are included. In Great Britain competitive tendering is applied to both profitable and not profitable services, resulting in a concession fee for the former and a subsidy for the latter. The Swedish national railway SJ decides whether or not a service is profitable. If the state-owned enterprise decides not to operate a service, competitive tendering is introduced. In Germany, where federal law obligates the state to provide regional transport services, there is no legal base for public assistance for long-distance services. Instead, regional authorities define parts of abandoned long-distance services as regional services which allows them to maintain service quality with public financial aid to the engaged operator. Italy uses public service contracts to ensure services that otherwise would not operate.

2.4. Infrastructure organisation

There are strong interdependencies between the design of market access for long-distance rail passenger services and the organisation of the infrastructure network.[4] The key characteristics for Austria, France, Germany, Great Britain, Italy, the Netherlands, and Sweden are presented in Table 1.

Directive 91/440/EEC mandates a separation between train operators and the infrastructure manager. Although by now all EU member countries' railway markets have undergone such separation, there are different degrees. In Great Britain, the Netherlands and Sweden, there is full independence between infrastructure functions and long-distance passenger train operations. In Austria, Germany, and Italy, a holding structure exists which comprises both infrastructure and train operating functions. Both options are explicitly allowed following the Directive. France has a formal separation but, by means of contracts, important segments of the infrastructure maintenance are still the responsibility of the SNCF.

The rail networks in all considered countries are owned by the state. After the negative experiences in Great Britain during privatisation in the 1990s, no other country has privatised its tracks or stations. Although the organisational structure of the infrastructure management varies widely, this does not affect the general acceptance of public responsibility for the railway infrastructure.

Table 1. **Characterisation of European long-distance passenger rail transport markets**

	Austria	France	Germany	Great Britain	Italy	Netherlands	Sweden
Degree of infrastructure - transport services separation	Holding integration	Partial separation	Holding integration	Full separation	Holding integration	Full separation	Full separation
Network ownership	Public	Public	Public	Public (after private intermezzo)	Public	Public	Public
Market dominance and operator ownership	100% ÖBB Personen-verkehr (state-owned), entries between Vienna and Salzburg announced by Westbahn (hourly service) and Fair Train (every 3 hours)	100% SNCF (state-owned)	99% DB – Fernverkehr (state-owned)	Oligopoly of private train operating companies (apart from temporary re-nationalisations) (Merkert, 2009)	100% Trenitalia (state-owned), considerable entries announced by NTV (Rome-Milan) and DB in co-operation with ÖBB (Munich-Verona)	NS (state-owned) and NS-/KLM joint-venture HSA (90% state-owned) together 100%	SJ (state-owned) dominating, some smaller railway undertakings present

Concessions	No	No	No	Competitive tendering for regionally delimited areas and long-distance connections, franchised under National Rail	No	Two concessions granted to NS until 2015, and to HSA until 2024	Concessions only on routes where SJ refuses to operate commercially
Open access	Yes	No	Yes	Yes, if not primarily abstractive	Yes	No	Only for night trains, change announced for 2009/2010
Degree of market opening	Access for all operators given	None	Access for all operators given	Access for all operators through competitive tendering	Access for international groupings given	None	Purely commercial day services reserved by law to SJ, change announced for 2009/2010
	Austria	**France**	**Germany**	**Great Britain**	**Italy**	**Netherlands**	**Sweden**

Source: Own illustration according to Alexandersson (2009), Alexandersson and Hultén (2009), Alexandersson *et al.* (2009), company websites, Holvald (2009), and Monopolkommission (2009, p. 56).

2.5. Organisation of long-distance services in selected European countries

The long-distance passenger rail market in most countries is still dominated by the state-owned incumbent which used to be (and in many countries still is) the only operator. The prominent exception is Great Britain, where there are several train operating companies (TOCs) active on long-distance routes. Most routes are tendered by the UK Department for Transport and operated as franchises. These companies operate under their own brands, but offer a common Internet platform *National Rail* through the Association of Train Operating Companies (ATOC). Open access operators are presented on the *National Rail* website as well. Specific tickets are valid on trains by different operators. Apart from temporary re-nationalisations (Merkert, 2009a), there is no publicly owned British TOC, but some TOCs are partially owned by foreign, state-owned enterprises, e.g., the German DB Regio owns part of Wrexham & Shropshire, and French Keolis owns shares in several British TOCs (Nash and Smith, 2007, and Alexandersson, 2009)**.**

There is more than one long-distance train operator in Sweden, Germany, and the Netherlands. In Sweden, SJ operates profitable services on its own account. On routes which SJ refuses to serve, the state organises competitive tendering to find the most economic operator. As a consequence there are several active train operators, although SJ offers the majority of services. Except for winning tendering and introducing night services, train operators other than SJ are so far not allowed to participate in the market, but a change to more open access was announced for 2009-2010 (Alexandersson and Hultén, 2009).

In Germany there is open access to the entire rail network, but Deutsche Bahn subsidiary DB Fernverkehr AG holds a 99% share in long-distance rail passenger transport. However, the incumbent is not the only active operator. After a few unsuccessful attempts, there are currently three enterprises which together comprise less than 1% of the long-distance market (Holzhey *et al.*, 2009, p. 99).

In the Netherlands long-distance rail services are split into two concessions operated by HSA and the national Dutch railway NS. In other European countries, only the state-owned incumbent serves the whole long-distance network. In some countries such as France the exclusivity is due to protective legislation which guarantees exclusive rights to the traditional railway undertaking.

Directive 91/440/EEC mandates an opening for “international groupings”. The term describes any association of two or more railway undertakings from different EU member countries for the purpose of providing international transport services. In countries with exclusivity rights of the incumbent, this means that only international services may be operated. The right to form international groupings is not limited to public train operators, and theoretically one international grouping could operate in all EU member countries, as long as each service is at least transnational.

Austria and Italy grant rail access to all operators. While no significant market entries have been observed, market entries in both countries have been announced by private operators for the near future.[5]

3. MODELS OF MARKET ACCESS

3.1. "Tendered Concessions" model

We define our first model for market access as the "Tendered Concessions" model, featuring Great Britain as the prominent, contiguous example. In this model the network of operations is structured into a reasonable number of sub-networks based on a demarcation along regions or traffic flows and tracks. There is competition for the market for each sub-network in terms of competitive tendering. The tendering authority typically has extensive design and decision responsibilities. It determines the level of supply as obligation in the tender documents. This includes routes, frequency of service, capacity, operation times, and (minimum) requirements of service quality. The aim of the competitive tender is to find the operator which best fulfills a list of criteria, with price often being the most important or single criterion. Quality aspects can represent an additional decision criterion with weights assigned to price and further aspects. In the case of profitable lines, the goal is to find the operator which pays the maximum concession fee to the state. Track access charges are regularly predetermined by the national track access charging scheme. One can generally differentiate between three kinds of contracts:

- Under a management contract, the operational as well as the revenue risks are borne by the authority.
- Under a gross-cost contract, the service operator bears the operational risk. This means a high degree of tariff regulation by the authority and relatively low incentives for the service provider to attract as many passengers as possible.
- Under a net-cost contract, the provider acts as the most entrepreneurial, and bears the risk on both the cost and revenue sides. The predefined contractual agreement between the authority and the operator has the advantage that performance control through a system of indicators can be established and pressure is applied through penalties on subsidies or payments.

However, as the quantity of services is mostly predetermined by state authorities, entrepreneurship is generally on a low level and the influence on costs and market development can be restricted. In the case of Britain, only the management changes when a franchise is awarded to a new operator (Smith *et al.*, 2009). While this introduces some employment certainty for existing personnel, it also establishes bargaining power and information advantages over the new management. The procedure additionally reduces the possibilities to influence costs and processes, and has direct effects on franchise bids. Hence, it is doubtful whether the sole exchange of top management is suitable for producing substantial changes in the service provision. This in turn is related to franchise duration, since increased durations lead to possibilities for better innovation for the new management. Obviously, the franchise duration must be aligned to the rolling stock life-cycle with replacement and refurbishment dates. The importance of franchise duration may also be dependent on rolling stock ownership. The British model introduced rolling stock companies (ROSCOs) as providers of the rolling stock. Such an additional level in the value creation removes a barrier to market entry, because the service provider is no longer concerned about the financing of rolling stock. In contrast, it is doubtful whether this supports rolling stock innovation (Yvrande-Billon and Ménard, 2005) and even more important, it introduces new transaction costs (Merkert, 2009b).

The "Tendered Concessions" model exhibits further transaction costs in the coordination of many players, particularly when there are network overlaps or connection points. Hence this model has high disadvantages in comparison with a single network operator when the network effects are high. On the other hand, a star-like (radial) network structure such as in Great Britain facilitates a franchise demarcation along regions or traffic flows and tracks. This is emphasized even more with a station structure, i.e. London, where different routes serve different stations without long-distance service interchanges. The "Tendered Concessions" model, possibly with common concessions for both long-distance and regional rail, improves the timetable coordination between different services. It can even remove any artificial distinction between long-distance and regional rail services. Under the model the allocation of subsidies to rail services operators or the concession payments to the authority can be carried out fairly and transparently. For example, government can decide to pay subsidies only to regional rail services which may be considered as public service obligations, but there is no general need to classify a service into long-distance or regional.

Furthermore, the "Tendered Concessions" model allows the state to compensate subsidies for non-profitable lines through the collection of concession fees on profitable lines. Hence, the state can balance expenses and revenues. We note, however, that there may be some additional costs for the authorities that manage competitive tendering. In this model, profitability and losses are bounded and state-regulated. This originates primarily in the design of auctions and reverse auctions where interested companies compete to win the franchise, and potentially articulate similar bids restricting the margin a priori. Moreover, cap and collar regimes such as in Great Britain limit the risks and chances (Kain, 2007, and Preston, 2008). The British Department for Transport will skim 80% of the revenues lying above the 106%-level of the train operating company's original forecast.[6] Hence, through a system of risk-sharing the rail operator's profit is limited. This conclusion only holds for a sufficiently high number of bidders that contribute to a market outcome with an efficient production of services, a bid that corresponds to average production costs plus an opportunity cost of capital that is normal in the market and a quality of services that is at least as good as before. Consequently, in the absence of corruption and collusion competitive tendering with a sufficiently high number of bidders will represent a competition substitute.[7] Unfortunately the substitute is frequently threatened by a decreasing number of bidders after an initial phase of tenderings (see e.g., Augustin and Walter, 2009, for an example from the German bus market).

A critical issue in the "Tendered Concessions" model is that the state and its authorities define the level of service concerning routes, frequency, quality, etc. This leaves room for political influence, e.g., routes or cities served because of localised political ambitions.

An extension of the "Tendered Concessions" model is possible through open access services proposed by private companies without any subsidy requests. Such services can provide direct connections without the need for transfers. Since these services also compete with franchised services, the regulation authority must decide whether they are not primarily abstractive from the franchisees' revenues (Griffiths, 2009). This procedure requires an intelligent institutional design to secure a fair and efficient decision process and furthermore introduces additional transaction costs. In Britain, open access services approved by the Office of Rail Regulation (ORR) have gained a share in passenger revenues of only 0.6% (ORR, 2009). A simulation of the different market entry strategies by Preston *et al.* (1999) shows that on-track competition in Great Britain usually reduces welfare resulting from increased consumer benefits, but greater profit reductions for operators. With the Directive 2007/58/EC coming into effect on January 1, 2010, open access will certainly play a more important role in continental Europe, but it is doubtful whether the directive will affect countries in island positions such as Great Britain. The only track connection to continental Europe is the Eurotunnel, and London is the first major stop so that cabotage does not play a large role in Great Britain.

3.2. Network concession for a monopolistic operator

A single network concession for a monopolistic operator under a performance-based contract represents a different type of concession. Such a concession appears to be relatively suitable when substantial network effects with many interconnection points and crosslines may be opposed to a split in concession areas. A very dense network can, for example, be found in the Netherlands with frequent timetable intervals similar to suburban transit systems. A monopolistic network concession can have the effect that it prohibits market entry of competitors and can be used to strengthen the position of a monopolistic public company. In the absence of competition, incentives for a public monopolistic network operator for efficient performance will be quite low. Hence, the challenge in this case is to design an institutional setting which facilitates efficiency-oriented governance of the monopolistic public company.

In the Netherlands, a 10-year concession contract was directly awarded to the Dutch railway undertaking NS in 2005 (Van de Velde *et al.*, 2009). The contract is part of a major railway reform in which infrastructure management is defined as a government responsibility, and passenger transport is targeted as a non-subsidized commercial activity. Non-profitable regional lines were separated from the NS network and tendered (Alexandersson, 2009). The concession contract is monitored by performance indicators, and NS must propose improvement values. The importance of such performance indicators has to be highlighted in case of the permanence of a monopolistic network. Difficulties can arise when trying to identify a reasonable benchmark company. This is necessary when improvements are not only compared to a firm’s own base level, but also to best practices. As monopolistic operator, an international comparison has to be made (see, e.g., Coelli and Perelman, 2000), but this is difficult because of different operating environments, different purchasing powers, and so on.

The early results of the Dutch concession contract are that investment activities intensified, and performance in terms of customer satisfaction and punctuality improved. From 2009 on, NS should pay for the concession. In fact, the concession for NS is complemented by a second concession for train operations on the high-speed railway link (HSL-South) between Amsterdam, Schiphol Airport, Den Haag, Rotterdam, and Brussels. This concession was granted in a competitive tendering to High Speed Alliance, a joint-venture of NS and AirFrance-KLM, for the period from 2009 to 2024 (NS, 2009, pp. 101, 120).

Open access has not been planned as an option in domestic services, although Directive 2007/58/EC naturally makes this possible for international services. This raises the issue of sustainability of the NS monopoly. Facing these seemingly contradictory approaches in national and European regulation, we note that the Netherlands have always been critical to opening up the national market because of the country’s assertion that a single company with exclusive rights is more capable of efficiently serving such a densely used network.

3.3. “Open Market” model

The “Open Market” model is based on the concept of competition in the market. All European countries that have introduced open access as the primary market entrance possibility, such as Germany, Italy, and Austria, still face the existence of a single monopolistic network operator, although some market entries have now been announced for the near future.

The “Open Market” model assumes that competition in the market, or at least the threat of competition, results in a creative product offering, technological innovation, and downward pressure on costs and prices. In theory, it emphasizes the entrepreneurship of operators, because they plan and determine routes, frequency, quality, and are assumed to operate as profit-takers. Hence, political influence in state-owned companies could be a critical aspect. Full profit orientation can lead to a cutback of service offerings in rural and remote areas, because it may emerge that they are unprofitable to operate. Nonetheless, if it is desirable to operate these services, the state faces a conflict between its open access orientation and the necessity to subsidize. A suitable solution to determine the minimum level of subsidies is competitive tendering, but this is not part of the “Open Market” model. However, non-profitable lines may be separated from the open market. This provokes strategic behavior: companies could shut low-profit or loss-generating interregional routes under the certainty that they will be publicly procured. They do not have to fear any negative network effects in terms of connections to profitable lines.

Given that local and regional services are likely to remain public service obligations, the “Open Market” model for long-distance services requires a reasonable differentiation between these two types of services. However, such a differentiation is not straightforward. Possible distinction criteria were given in Subsection 2.2, but all are either difficult to measure or give providers room for strategic behavior.

On the one hand, indirect subsidization is also possible via reduced track charges. On the other hand, the state could potentially gain track access charges that are cost recovering. In either case, it is not possible for the state to compensate subsidies through concession fees, in contrast to the “Tendered Concessions” model. For very efficient and well-positioned firms, it is possible to gain high profits, which may once again attract entrepreneurs and creative product offerings.

Reducing service offerings in rural and remote areas also represents opportunities for competitors. These lines offer a market niche with limited capital requirements in comparison to a major high-speed trunk route, and the risk of direct competition with the incumbent, or even predatory behavior, is low.

The discussion about cutbacks in rural and remote areas initiates a controversy in how far the “Open Market” model should be augmented with obligations for operators to serve regions, to provide special rates for low-income customers, to provide interconnections with other means of transport, etc. These obligations are all part of the larger question about how to accomplish welfare enhancements in the “Open Market” model. More generally, the profit orientation in this model can lead to an increase in ticket prices in comparison to the politically influenced ticket prices of state-owned European incumbents.

The “Open Market” model closely relates to another institutional aspect of European railway organisation: the separation of infrastructure and operations. There is a long-standing discussion on the advantages and disadvantages of vertical integration vs. unbundling. Empirical results have confirmed the presence of economies of scope between a network and train operations for a majority of European railways (Growitsch and Wetzel, 2009). However, it is doubtful if vertical integration is necessary to exploit these economies of scope. Hirschhausen *et al.* (2004) found that only a few critical transaction processes that demand a hierarchical organisation are existent.

In practice, supporters, particularly labour unions in Germany, have pointed out the benefits of internal labour markets in the case of vertical integration.[8] In contrast, there is a strong discrimination potential against competitors. This potential is especially relevant for the “Open Market” model in long-distance passenger rail transport with the importance of network effects, but is less relevant where network effects are insignificant and the transport is less sensible to the particular time slot for network access, e.g., for block train traffic in freight transport. It is also less relevant for tendered services, because the track allocation takes place on an upstream level. An independent network operator will try to maximize network utilisation and enforce the development of bottlenecks, while a vertically integrated railway company will try to maximize the firm's entire profits. If unbundling is politically not enforceable, then the minimum requirement for a functioning market is to implement an effective access regulation.

The discussion concerning non-discriminatory network access is related to two more aspects. First, transparency with respect to free capacities is necessary. This could easily be implemented with an Internet-based information system (Monopolkommission, 2009). Second, the "Open Market" model also requires careful consideration of the long-term planning reliability for network access. Once procured, rolling stock may be difficult to resell, and the deployment on other tracks may be impossible due to different technological requirements. Hence, as investment in rolling stock is specific and secondary markets are almost non-existent, it is important to have ensured slots on tracks for a sufficiently long period to recover the investment, e.g., for a minimum of 10 years.

Network access is not the only monopolistic bottleneck in the "Open Market" model. Alexandersson and Hultén (2009) emphasize the need for an independent booking and ticketing system. An independent authority may also be desirable for timetable planning. Finally, in comparison to other sectors, such centralised institutions and state intervention tend to limit the "free-ness" of this market.

Critical to all market access models, network effects can play a very important role in long-distance passenger rail transportation. In the "Open Market" model, additional offerings selected by cherry- picking can lead to service terminations of the incumbent because of revenue abstraction, to a reduction of network effects, or to increased network congestion. Following this, the beneficial former network effects such as interconnection possibilities, integrated vehicle scheduling, and cost advantages can be harmed or even destroyed.[9] Thus, the "Open Market" model can present disadvantages for consumers and can lead to inefficiencies from a welfare economic perspective.

On the other hand network effects may be so beneficial to the incumbent that on-track competition never develops. Another negative impact on potential competition results from scarcity of network capacities. In consequence, the network operator might have strong market power. This will be especially problematic for consumers as well as from a welfare economic perspective in the case of a monopolistic network operator which is privatised and aims at maximizing profits.

An additional possibility of the incumbent to foreclose competition is implementation of strategic behavior against potential newcomers, e.g., the incumbent can invest in rolling stock only for the purpose of deterrence. In general, strategic behavior of the different players can be expected and will cause net-costs from a welfare economic point of view.

We note that due to intermodal competition effects the controversy about the extent of intramodal competition in the railway market may be of little significance. The most important competitors are motorised individual transport (MIT), air transport, and express coaches. However, these means of transport sometimes address different target groups, and we note that they partially serve different sub-segments of travelling. MIT is attractive because of its flexibility but may be inadequate for long trips and a lesser alternative for business travellers and the socially deprived.

Friederiszick *et al.* (2009) find a high competition intensity between low-cost airlines and Deutsche Bahn (DB). Holding the view that there is very low potential for on-track competition for long-distance passenger rail services, Friederiszick *et al.* (2009) conclude that international railway alliances such as Railteam are no threat to competition. With respect to intermodal competition, the results of Friederiszick *et al.* (2009) are not generalisable, because air service plays no role in many routes that are less than 300 km, or that are point-to-point connections between cities with no substantial air connections. This in particular holds for a decentralised urban settlement structure such as Germany (in contrast to France with its star-like travel flows to Paris). Friederiszick *et al.* (2009), whose research has been financed by DB, have been criticised for their sample selection, e.g., Monopolkommission (2009, p. 78) particularly questions their short observation period (January 2006 until October 2007) with less emphasis on winter months.

Express coach services can be an alternative for young people, seniors, low-income earners, and others who are not as sensitive to travel times. It is questionable how much of a competitive threat express coach services represent to railways, or if they merely induce new traffic and entice passengers away from motorised individual transport (Walter *et al.*, 2009). One option is the provision of non-profitable interregional lines with economical express coach services.

4. CASE STUDY: GERMANY

Germany's railway market is the largest in Europe, and a prominent example of an "Open Market" model for access to the long-distance network. However, the market that has developed so far is characterised by niche competition rather than open market features.

A major reorganisation of German railways was conducted in 1994 with the *Bahnreform*. The first stage of this railway reform consisted of three basic principles. The first was to reorganise the formerly West German railway *Bundesbahn* and the East German railway *Reichsbahn* into a new, primarily state-owned, corporation. The second concerned the delegation of responsibility for regional railway services to the federal states. The third, and most important for studying market access models, was to introduce non-discriminatory market access for private companies. Germany was hence at the forefront for providing open access to the long-distance passenger rail transport market.

To date there has been no substantial on-track competition. Holzhey *et al.* (2009) count 9 attempts to enter the market in 15 years of liberalization, all of which are small-scale and consist of at most 2 pairs of trains per day. Five of these services ceased after operating for a very short time. The remaining services have in common the ability to serve routes that were previously operated by some kind of Deutsche Bahn train, in particular the so-called InterRegio lines. These were abolished beginning in 1999 because of profitability problems (Link, 2004). The underlying concept of the InterRegio (and also of its competitive successors) was to connect the many medium-sized towns and vacation areas with metropolitan areas. The services stopped frequently (thus were slower than InterCity or high-speed trains) and were also cheaper. The more utilised lines were reorganised into InterCity lines, while the rolling stock partly remained the same and prices were increased.

Another condition for the start of the few commercial services was the introduction of competitive tendering for regional rail services. The four services have used rolling stock from their regional operations and two, Harz-Berlin-Express (Veolia) and Vogtland-Express (Arriva), represent an extension of lines operated under a public service obligation. The InterConnex Leipzig-Berlin-

Rostock (Veolia) was equipped with long-distance rolling stock after the first four years of operation. It is also the only service directly competing with DB long-distance offerings. All of these services can be distinguished from DB offerings by their longer travel times and lower prices (Séguret, 2009).

An exception is the night train between Berlin and Malmö in Sweden. Unlike the other services mentioned which are provided by subsidiaries of international integrated private transport companies, Berlin Night Express is operated jointly by Georg Verkehrsorganisation and SJ.

Interestingly, the four current long-distance offerings by DB’s competitors are connections to Berlin through the eastern part of Germany. Two reasons for the existence of these routes may be the East German settlement structure which has only three larger agglomeration areas (Berlin, Dresden, and Leipzig) and the low percentage of business travellers which make them unattractive for DB. A third reason may be the price sensitivity in regions with lower per capita incomes.

The low level of competition intensity can be attributed to four factors. The first is DB’s vertically integrated structure with discrimination potential and information advantages, in particular through information exchange between long-distance operations and the network. The DB infrastructure subsidiaries directly control 35% of total costs for long-distance service operations, such as access charges, traction power, etc. (Holzhey *et al.*, 2009, p. 102). This cost issue is particularly relevant, since the sector is said to yield only low profit margins. However, this could also be related to the incumbents’ business models. Low-cost airlines, for example, have been able to earn high profits from a similar market situation in aviation.

The second factor is network access. Congestion is already a problem in Germany, and it has been attenuated due to the present financial crisis and resultant decline in freight transport. The focus of past network investments has been on new high-speed lines, e.g., Frankfurt-Cologne or Munich-Nuremberg-Erfurt-Berlin, whereas main junctions, e.g., in Frankfurt and Cologne, are congested, intersections exhibit obstacles (Vieregg, 2004), long-distance, freight and regional traffic are forced to share congested track sections, and many lines are speed-restricted because of poor track. Moreover, transparency concerning free capacity could be improved. Holzhey *et al.* (2009, p. 115) have proposed a visualised network capacity timetable that is open to all interested companies. The instrument of framework contracts could be improved through more flexibility, longer lead times, and the prioritisation against other awarding criteria (Monopolkommission, 2009, pp. 7, 61).

The third factor is the expansion strategy of local authorities that have begun to procure interregional services. Good examples are the so-called regional services on the Elsterwerda-Berlin-Stralsund route with a line length of over 400 km and the service between Munich and Nuremberg that serves the new high-speed line between Ingolstadt and Nuremberg with former long-distance rolling stock[10]. Although these services may constitute travel improvements, they also signal that there is no need for private initiatives for commercial lines, and they complicate the discovery of appropriate connections (Monopolkommission, 2009, p. 58).

The fourth factor is the impact of today’s financial crisis that has made it more difficult to finance rolling stock investments. However, two recent announcements of market entry may represent a new strategy. In October 2009, the private newcomer locomore rail announced plans to operate three daily trains from Hamburg to Cologne after August 2010, meaning that it has already successfully applied for track capacity. Comfort and travel time should be comparable to DB InterCity services, and tickets should be cheaper[11]. locomore is supported by the US investment firm Railroad Development Corporation. A potential strategy to reach competitive travel times and to save access costs may be to stop at alternative stations instead of running into bottlenecks and loops such as the main stations of Dortmund and Bremen.

A less advanced, but potentially more dangerous competitor for DB is Keolis, which is backed by SNCF, Axa Private Equity, and a Canadian pension fund. Also in October 2009, it announced services between Strasbourg, Frankfurt and Hamburg, and Strasbourg, Frankfurt, Berlin, and Hamburg comparable to DB InterCity services. Keolis has not yet received a confirmation for track access. This decision will be made by the network subsidiary of DB by April 2010, so that services could start at the earliest in December 2010.

As a starting station, Strasbourg offers Keolis the possibility to use existing French maintenance facilities and to span a real international network of train connections. However, possible market distortions follow from the (partial) state ownership of both Keolis and DB which compete with private operators.

These announcements both incorporate a new strategy for market entry compared to the previously introduced peripheral services of Veolia and others. Both potential entrants would serve trunk routes that are characterised by competitive average speeds without the imperative use of expensive high-speed vehicles.[12] Competition in the high-speed segment up to 300 km/h may also be limited by the close international cooperation and joint ventures in this segment, such as Railteam, Thalys, etc.

However, it is important to bear in mind that the market organisation and the regulatory setup are by no means finalised. The coalition agreement of the new German government further assumes a vertically integrated DB under a holding company, in order to facilitate a common job market.[13] The transport and logistics subsidiaries will be privatised as soon as capital markets recover. However, shifting the profits from the network to the holding will not be permitted, and the infrastructure will get a more independent management. Dual mandates with the same manager holding positions in both the holding and the network subsidiary will not be permitted. Further objectives of the railway policies mentioned in the coalition agreement are: a stronger regulator; harmonization of the rules on a European level; and the examination of a highly synchronised countrywide timetable with infrastructure investments in specific bottlenecks (Deutschland-Takt). The issues remaining are the extent to which a partially privatised monopolist can exercise market power to raise prices and to abandon services in rural areas, and how the potential on-track competition can serve to mitigate such effects.

The suggestions from DB competitors (Holzhey *et al.*, 2009, p. 113) aim to completely change the organisation of Germany's long-distance passenger rail market. One option may be the introduction of concessions for all routes and marketing of all services under a common brand. Another option is to focus on concessions for interregional lines to establish a second long-distance network alternative to the expensive high-speed segment. A third option is the systematic support of long-distance services by track access charges where peripheric routes are subsidized through higher charges on high-demand routes. A careful evaluation is necessary to determine the ability of these options to resolve critical long-distance passenger rail market concerns. It must however be clear that the introduction of concessions would renounce the "Open Market" model practiced in Germany so far.

The coalition agreement also includes liberalization of express coach services in Germany. Until now, these services have been heavily restricted to single connections, mainly to and from Berlin (Walter *et al.*, 2009). Express coach services could fill the gap left by abandoning trains on less-frequented routes with bus units that are smaller than trains. On the other hand, market entry is also likely to focus on trunk routes with great passenger potential and interest in low prices.

5. CONCLUSION

This paper has classified the models for market access in European long-distance passenger rail transport into the "Tendered Concessions" model, the "Monopolistic Network Operator" model and the "Open Market" model. Noting that each European country will pursue its own approach aligned to regional circumstances, nonetheless our classification can help to structure the ongoing discussion. We have presented the models' strengths, opportunities, risks, and threats without favoring any one model. There are very different design options which have very different impacts. Empirical experience with the "Tendered Concessions" model in Great Britain has progressed the most, while open access experience is still in its infancy.

Open access appears to be the preferred regulation for international services, as manifested through Directive 2007/58/EC. With this directive, cabotage is possible, but only when the routes served under public service obligations are not distorted. It remains unclear whether open access for international services may distort tendered concessions in domestic markets, hence, if these two contradictory regulations coincide. This may be a smaller problem for geographically or technologically isolated rail markets, e.g., Great Britain, but could be a larger problem for networks highly integrated in a central European country like the Netherlands.

NOTES

1. These distinction criteria are all used in the German market.

2. Regulation (EC) No. 1370/2007, coming into effect on 3 December 2009, strives to stimulate competition in passenger transport markets and specifies competitive tendering as the standard award procedure. However, rail services are excluded from this rule, and direct awards with tenures of up to 10 years are possible (15 years when competitive tendering is used).

3. Subsidies can accompany institutional problems, such as the need for funding through tax collection. We do not further consider such aspects.

4. We define long-distance services as any rail services that are not classified as urban, suburban or regional services in Directive 91/440/EEC.

5. See Monopolkommission, 2009, p. 56, and www.oepnvwettbewerb.de/index.php?option=com_content&task=view&id=1872&Itemid=27.

6. The cap and collar regimes are controversial because they can provoke strategic behaviour in the estimation of revenues and costs (Preston, 2008).

7. With a high number of bidders, an efficient market outcome is more or less guaranteed. With a low number of bidders, competitive pressure can be still high enough to lead to an efficient market outcome, but this is more uncertain.

8. This argument has enjoyed renewed attention in the current financial crisis, because the cargo subsidiary of Deutsche Bahn experienced a sharp recession, with a subsequent decline in the need for personnel.

9. Service terminations because of revenue abstraction do not necessarily lead to welfare decreases. However, this is likely to lead to decreased network effects which, in turn, imply welfare disadvantages.

10. In contrast to the ICE high-speed service on this line the additional train offers more stops with the accompanied increased travel time.

11. http://www.abendblatt.de/wirtschaft/article1253055/Hamburg-Berlin-SNCF-fordert-Bahn-heraus.html, retrieved November 2, 2009.

12. The use of high-speed vehicles represents the third major market entrance strategy.

13. www.fdp-bundespartei.de/files/363/091024-koalitionsvertrag-cducsu-fdp.pdf, p. 29 f.

BIBLIOGRAPHY

Alexandersson, G. (2009), Rail Privatisation and Competitive Tendering in Europe, *Built Environment*, 35(1), 43-58.

Alexandersson, G. and S. Hultén (2009), The Complexity of Market Structure – Prospects for On-the-track Competition in Sweden, in: Beck, A., W. Veeneman and D. van de Velde, *Beyond Competitive Tendering – Proceedings of the Eleventh International Conference on Competition and Ownership in Land Passenger Transport, Workshop 3*, NGInfra Foundation, Delft.

Alexandersson, G., S. Hultén, F. Fearnley and F. Longva (2009), Impact of Regulation on the Performance of Long Distance Transport Services: A Comparison of the Different Approaches in Sweden and Norway, in: Beck, A., W. Veeneman and D. van de Velde, *(op. cit.)*, NGInfra Foundation, Delft.

Augustin, K. and M. Walter (2009), Operator Changes through Competitive Tendering: Empirical Evidence from German Local Bus Transport. Dresden University of Technology, Chair of Energy Economics and Public Sector Management, Working Paper Transport Economics 17. www.tu-dresden.de/wwbwleeg/publications/wp_tr_17_augustin_walter_2009_operator_changes_competitive_tendering_bus_services.pdf , retrieved October 22, 2009.

Coelli, T. and S. Perelman (2000), Technical Efficiency of European Railways: A Distance Function Approach, *Applied Economics*, 32(15), 1967-1976.

Friederiszick, H., T. Gantumur, R. Jayaraman, L.-H. Röller and J. Weinmann (2009), Railway Alliances in EC Long-Distance Passenger Transport: A Competitive Assessment Post-Liberalization, 2010. ESMT White Paper No. WP-109-01, ESMT European School of Management and Technology.

Griffiths, T. (2009), On Rail Competition: The Impact of Open Access Entry on the Great Britain Rail Market, in: Beck, A., W. Veeneman and D. van de Velde (*op. cit.*), NGInfra Foundation, Delft.

Growitsch, C. and H. Wetzel (2009), Testing for Economies of Scope in European Railways: An Efficiency Analysis, *Journal of Transport Economics and Policy*, 43(1), 1-24.

Hirschhausen, C. v., J. Siegmann, A. Brenck, M. Holzhey, L. Hübner, B. Peter, T. Schulin and S. Schultz (2004), Synetra: Synergien zwischen Bahnnetz und –transport – Praxis, Probleme, Potentiale. Study prepared by the Workgroup for Economic and Infrastructure Policy and Fachgebiet Schienenfahrwege und Bahnbetrieb at Technische Universität Berlin, supported by the German Federal Ministry of Education and Research.

Holvald, T. (2009), Review of Railway Policy Reforms in Europe (2009), *Built Environment* 35(1), 24-42.

Holzhey, M., F. Berschin, I. Kühl and R. Naumann (2009), Wettbewerber-Report Eisenbahn 2008/2009. Report prepared by KCW GmbH on behalf of mofair e. V. and Netzwerk Privatbahnen-Vereinigung Europäischer Eisenbahngüterverkehrsunternehmen e. V., supported by Bundesarbeitsgemeinschaft der Aufgabenträger des SPNV e. V.

Kain, P. (2007), The Pitfalls in Competitive Tendering: Addressing the Risks Revealed by Experience in Australia and Britain, in: European Conference of Ministers of Transport, *Competitive Tendering of Rail Services*, OECD Publishing, Paris.

Link, H. (2004), Rail Infrastructure Charging and On-track Competition in Germany, *International Journal of Transport Management*, 2(1), 17-27.

Merkert, R. (2009a), Großbritannien: Wird der Zugbetrieb verstaatlicht? Eine aktuelle Kurzanalyse zur Franchiseproblematik im britischen Schienenpersonenverkehr, *Internationales Verkehrswesen*, 61(9), 320-322.

Merkert, R. (2009b), The Organisation of European Railways: A Transaction Cost Perspective. Dissertation, Institute for Transport Studies, University of Leeds.

Monopolkommission (2009) Bahn 2009: Wettbewerb erfordert Weichenstellung. http://www.monopolkommission.de/sg_55/s55_volltext.pdf, retrieved October 26.

Nash, C. and A. Smith (2007), Passenger Rail Franchising – British Experience, in: European Conference of Ministers of Transport, *Competitive Tendering of Rail Services*, OECD Publishing, Paris.

NS (2009) Annual Report 2008.

http://www.ns.nl/cs/Satellite?blobcol=urldata&blobheader=application%2Fpdf&blobheadername1=Content-Disposition&blobheadername2=MDT-Type&blobheadervalue1=attachment%3B+filename%3D577%2F52%2FVBAT_NS_Jaarverslag+EN+LR.pdf&blobheadervalue2=abinary%3B+charset%3DUTF-8&blobkey=id&blobtable=MungoBlobs&blobwhere=1208880232319&ssbinary=true, retrieved October 27, 2009.

ORR (2009), National rail trends 2009-10, Quarter 1, London.

Preston, J. (2008), A Review of Passenger Rail Franchising in Britain: 1996/1997 – 2006/2007. *Research in Transportation Economics*, 22(1), 71-77.

Preston, J., G. Whelan and M. Wardman (1999), An Analysis of the Potential for On-track competition in the British Passenger Rail Industry, *Journal of Transport Economics and Policy*, 33(1), 77-94.

Séguret, S. (2009), Is Competition on Track a Real Alternative to Competitive Tendering in the Railway Industry? Evidence from Germany, in: Beck, A., W. Veeneman and D. van de Velde, (*op. cit.*), NGInfra Foundation, Delft.

Smith, A., C. Nash and P. Wheat (2009), Passenger Rail Franchising in Britain: Has It Been a Success? *International Journal of Transport Economics*, XXXVI(1), 33-61.

Van de Velde, D., J. Jacobs and M. Stefanski (2009), Development of Railway Contracting for the National Passenger Rail Services in the Netherlands, in: Beck, A., W. Veeneman and D. van de Velde (*op. cit.*), NGInfra Foundation, Delft.

Vieregg, M. (2004), Schienenpersonenfernverkehr in Deutschland – Sollen weitere Neubaustrecken überhaupt noch realisiert werden? *Internationales Verkehrswesen*, 56(3), 72-77.

Walter, M., F. Haunerland and R. Moll (2009), Heavily Regulated But Promising Prospects: Entry in the German Express Coach Market. Dresden University of Technology, Chair of Energy Economics and Public Sector Management, Working Paper Transport Economics 16, www.tu-dresden.de/wwbwleeg/publications/wp_tr_16_walter_et_al_2009_express_coach_germany.pdf, retrieved August 25, 2009.

Yvrande-Billon, A. and C. Ménard (2005), Institutional Constraints and Organizational Changes: The Case of the British Rail Reform, *Journal of Economic Behavior & Organization*, 56(4), 675-699.

COMPETITION FOR LONG-DISTANCE PASSENGER RAIL SERVICES: THE EMERGING EVIDENCE

John PRESTON

University of Southampton
Southampton
United Kingdom

SUMMARY

ACKNOWLEDGEMENTS

The development of the PRAISE model and the Great Britain rail case study was undertaken at the University of Leeds, with a key role played by Gerard Whelan (now at MVA Consultancy). The Sweden rail case study was undertaken at the University of Oxford, with a key role played by Biao Huang (now at the Asian Development Bank). Numerous other colleagues contributed to this work, including Tim Griffiths of ORR, but the interpretations here (and any mistakes therein) are purely the author’s own.

1. INTRODUCTION

Railways were initially envisaged as open access facilities with head-on competition between service providers (Lardner, 1850). However, concerns about safety quickly resulted in railways being largely developed as vertically integrated monopolies at a route level but with significant competition between these route-based companies. Over time, competition from other modes reduced the scope for internal competition and led to the rationalisation of duplicated routes and the merger of railway companies. In most countries, long distance passenger rail services[1] became a state-owned monopoly but in recent years there has been renewed interest in competitive provision (see, for example, Gomez-Ibanez and de Rus, 2006).

Although route competition has remained a feature in countries such as Japan (Mizutani, 1994), on the tracks competition between passenger rail operators has been limited. However, in Great Britain, the 1993 Railways Act promised open access competition between rail operators. In the event, regulatory intervention heavily moderated competition up to 2002. Nonetheless, some open access competition has emerged in Britain with three passenger train operators having entered the market (Griffiths, 2009). There has been open access competition in passenger rail markets elsewhere – most notably, in Germany where open access has been permitted since 1999. There have been around ten instances of entry of which four were still operating in 2009, centred on Berlin (Séguret, 2009), but accounting for less than 1% of services.[2] The liberalisation of long distance passenger services has seen the incumbent operator Deutsche Bahn (DB) withdraw from secondary markets, with some 23 medium-sized cities losing long-distance train connections between 1999 and 2009. When permitted, niche competition has emerged in other rail markets, such as St Petersburg–Moscow in Russia (Dementiev, 2007). The Netherlands has had some experimentation with open access, most notably the ultimately ill-fated Lovers Rail services (1996-1999), with the Dutch Government subsequently favouring off-the track competition (van de Velde, 2009). Within the European Union (EU), open access for international passenger rail services, with domestic cabotage, will be implemented in 2010 (Directive 2007/58). Nash (2009) reports that, in preparation for this, the SNCF has set up a subsidiary, Nuovo Trasporti Viaggiatori, to operate in Italy, whilst TrenItalia is believed to be planning retaliatory action. Air France and Veolia have established a partnership, possibly with a view to competing with Thalys services, whilst DB are believed to be considering competing with Eurostar services. On a domestic level, Sweden is considering open access for its rail services in 2010-11 (Alexandersson, 2009)

Off-track competition, in the form of competitive tendering and franchising, is more common in the passenger rail industry than on-track competition (Thompson, 2006). In Europe, the pioneer was Sweden, where competitive tendering for local services began in 1990 and extended to regional services (many of which are long distance) in 1993, although key intercity services remain a commercial monopoly. This model has also been adopted in countries such as Denmark, Germany and the Netherlands and further afield in countries such as Kazakhstan (Sharipov, 2009). The EU's subsequent intention was for a widespread roll-out of competitive tendering but this met opposition from some Member States, and Regulation 1370/2007 merely requires clear and transparent contracts. In Latin America, urban and suburban services were privatised through concessions, with the Buenos Aires commuter network in Argentina being transferred to the private sector in 1992, as was the Rio de Janeiro Metro and commuter services (Flumitrens) in Brazil in 1997-8. These arrangements build on similar models in the United States, where commuter rail services have been contracted out in cities such as Boston, Baltimore/Washington, Chicago and Los Angeles (Preston *et al.*, 2001) and have been extended to other urban rail systems, most notably in Melbourne, Australia (Kain, 2006). However, contracting out of long-distance passenger services is relatively rare. In Argentina, it did not prove possible to find private operators for its long-distance services and 70% of such services were discontinued, with the remainder taken over by regional governments. The main exception is Great Britain, where all passenger services were franchised in 1996-7, with five out of 25 train operating companies being particularly focused on long-distance services (Cross Country, East Coast Mainline, Great Western Mainline, Midland Mainline and West Coast Main Line).

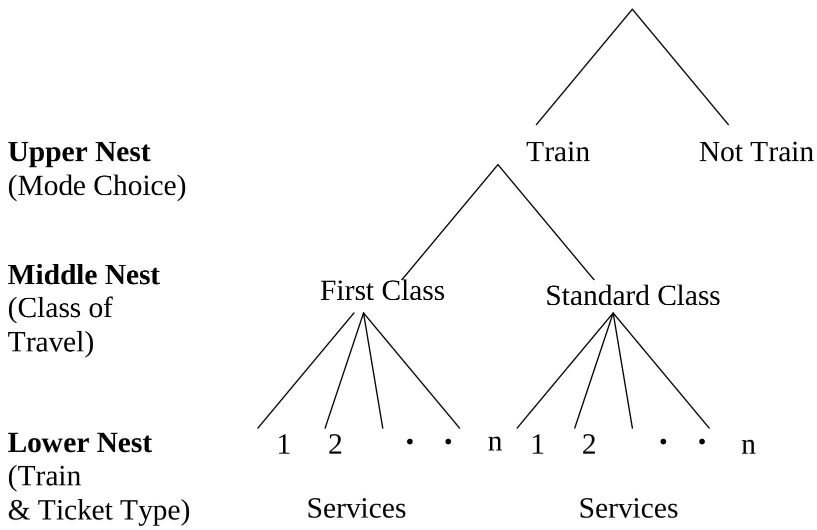
The aim of this paper is to review the emerging evidence on competition in long-distance passenger rail service. This draws on three bodies of evidence. In section 2, we examine the *ex-ante* evidence, from theoretical models based on Preston (2008a). In section 3, we examine the *ex-post* evidence on competition for the market, with particular emphasis on the East Coast Main Line franchise in Great Britain, drawing in part on Preston (2008b). Likewise, in section 4, we consider recent evidence on open access services that are competing in the market in Great Britain, drawing on Griffiths (2009). Finally, we shall draw some conclusions.

2. THEORETICAL MODELS OF RAIL COMPETITION

Rail competition, where it occurs, is likely to be small group in nature. Market demand is often too thin to support a large number of operators, whilst there may be some economies of scale and density that limit the optimum number of firms in rail markets (see, for example, Smith and Wheat, 2009). The relevant industry structure is therefore that of oligopoly competition. Classical models assume competition occurs either in the price dimension (Bertrand competition) or in the service dimension (Cournot competition). The conventional wisdom is that where capacity is difficult to increase (*e.g.* rail) competition will be of the Cournot type but where capacity can easily be increased (*e.g.* bus) competition will be of the Bertrand type (Quinet and Vickerman, 2004, p.263). However, this ignores demand side aspects. The urban rail market has turn-up-and-go characteristics which mean that passengers will tend to board the first train to arrive. Price competition is more effective in book-ahead markets such as long distance rail services. Indeed, price competition was a strong feature of the competition between British Coachways, National Express and British Rail in the early 1980s (see Douglas, 1987). However, Kreps and Scheinkman (1983) show that with appropriate quantity pre-commitment (which is likely to be the case in rail) Bertrand and Cournot competition can be equivalent.

Economic models of competition in rail have focused on the development of route based models in which the impacts of particular timetables (schedules) are examined and have some similarities with the dynamic schedule-based approaches developed by others (Wilson and Nuzzolo, 2004). An example is the PRAISE (PRivatisation of Rail SErvices) model (see Preston *et al.*, 1999, 2002). A similar modelling approach was adopted by SDG (2004) in modelling rail competition on the Brussels-Cologne and Madrid-Barcelona. The demand module of PRAISE assumes that individuals make their travel decisions at three linked stages (shown in Figure 1). At the first level (lower nest), the traveller's choice of service and ticket type is modelled, next the traveller's choice of class of travel is assessed in the middle nest and finally, the choice of travelling by train and not travelling by train is modelled in the upper nest. The model is therefore capable of distributing demand between trains and ticket types and allows for the overall rail market to expand or contract in response to fares and service level changes.

Figure 1**: Schematic representation of the PRAISE Demand Model**



The choice of service and ticket type on the outward and return legs of the journey are assessed in the lower nest of the model. For a given individual with a given set of tastes (attribute values) and preferred departure times for the outward and return legs of the journey, we can allocate a "utility weight" to each available train and ticket type combination. Choice probabilities are then estimated for the best nine return-trip combinations using a multinomial logit formula, where P_{ij} is the probability that individual i will choose service and ticket combination j, and U_j is a utility weight typically based on fare, adjustment time (i.e. the difference between when a person would ideally like to travel and the scheduled departure time), journey time, advanced purchase requirement and interchange, though it can include other rail attributes such as rolling stock quality. λ is a spread parameter that governs the sensitivity of choice between services and ticket combinations.

$$P_{ij} = \exp(\lambda U_j) \, / \sum_{j=1}^{9} \exp(\lambda U_j)$$

The middle nest of the model examines the choice between first and standard class travel. This is done by estimating a representative measure of utility for each class of service by way of the expected maximum utility (EMU).

$$EMU_{class} = \frac{1}{\lambda} \ln \sum_{j=1}^{9} \exp(\lambda U_j)$$

The choice between first and standard class of travel is then determined by the binary logit model, where θ is a scaling coefficient that determines the sensitivity of choice between first and standard class travel. Different θ values are estimated for different journey purposes.

$$P_{First} = \exp(\theta EMU_{First}) / (\exp(\theta EMU_{First}) + \exp(\theta EMU_{\text{Standard}}))$$

The final stage of the model represents the individual's choice between travelling by train and not travelling by train. This is done by estimating a representative value of rail travel for the individual (EMU_{train}) and allocating market shares using another binary logit model.

$$EMU_{Train} = \frac{1}{\theta} \ln(\exp(\theta EMU_{First}) + \exp(\theta EMU_{\text{Standard}})) \text{ and}$$

$$P_{Train} = \exp(\gamma EMU_{Train}) / (\exp(\gamma EMU_{Train}) + 1)\,.$$

Initial versions of the model involved setting the utility of not travelling by train equal to zero and estimating two separate γ values to restrict the fare elasticity of demand for business and non-business travel in Britain at -0.5 and -1.0 respectively (consistent with British Railway Board, 1990). In the Swedish application, elasticities of -0.4 for business travel, -0.6 for commuting and -0.9 for leisure were used (supplied by the state operator SJ). The British version of the model was based on a business value of time of 60 pence per minute and a non-business value of three pence per minutes (rebased to 2000 prices), based on local survey data (Preston *et al*., 1999). The Swedish version of the model was based on a business value of time of approximately 16 pence per minute and a non-business value of approximately eight pence per minute (again in 2000 prices) based on national values and the work of Rosenlind *et al.* (2001). Based on existing demand patterns, the model determines ideal departure times and the penalties for travelling earlier or later than the desired time. Changes in timetables will change the extent of these penalties. These ideal departure times are used to determine choice sets and reduce some of the concerns stemming from the independence of irrelevant alternatives property of multinomial logit models (Jansson and Mortazavi, 2000).

For a given route, the cost module is based on a fully accounted cost formulation which took the following general form:

TC = (1 + A) (aV + bVH + cVKM + dPKM)

Where:
TC = Total Cost
A = Administrative mark-up
V = Vehicles
VH = Vehicle Hours
VKM = Vehicle Kilometres
PKM = Passenger Kilometres.

Such a linear function is clearly a simplification of more complex relationships but has been widely used in the rail industry (Rosenlind *et al.*, 2001) and has some empirical support (Jörgensen and Preston, 2003). Parameters for the cost module were provided by the incumbent operators. A crucial difference relates to track access charging. In Great Britain, the track authority is a commercial enterprise (Railtrack from 1996 to 2002, Network Rail thereafter) and charges are based on the principle of full cost recovery. In Sweden, the track authority (Banverket) is a public body and charging is based on short run marginal costs. The upshot is that, at 2000 prices, track access charges were around GBP 5 per train-km in Great Britain compared to 65 pence per train-km in Sweden.

The appraisal module calculates profit as the difference between total revenue and total cost and calculates changes in consumer surplus using the rule of half. Change in welfare is simply the sum of the change in profits and in consumer surpluses.

Tables 1 and 2 summarize the findings of the PRAISE model in applications to a broadly hourly inter city service in Great Britain (Route GB1), with approximately 2 million passenger journeys per annum. This route links two major cities but has substantial commuting at either end of the route. It is assumed that both the existing (incumbent) operator and the new (entrant) operator use the same rolling stock so that the quality of service is the same and, with the same stopping patterns, the speed of services is also the same. In reality, it is likely that competition will occur with respect to the quality of service as well as with respect to the quantity of service and fares, but this would require detailed modelling of the rolling stock market.

Table 1**. Sample fringe competition results – Route GB1**

Model run	**Fares**	**Entrant service pattern**	**Inter-availability of tickets**	**Incumbent share (%)**	**Rail market growth (%)**
1	A*	1*	Yes	93.9	0.6
2	A*	1*	No	94.6	0.4
3	B*	1*	Yes	88.9	2.5
4	B*	1*	No	87.4	1.8
5	C*	1*	Yes	93.3	10.8
6	C*	1*	No	94.3	10.4
7	A*	2*	Yes	89.8	-2.6
8	A*	2*	No	89.6	-3.1
9	B*	2*	Yes	86.0	-0.3
10	B*	2*	No	84.3	-1.1
11	C*	2*	Yes	88.7	7.3
12	C*	2*	No	88.6	6.6

Notes:

1* Entrant provides two additional express services in the morning and evening peak periods in both directions of travel.

2* System is at capacity, the entrant replaces two of the incumbent's services in the morning and evening peak periods in both directions of travel with express services.

A* Entrant price matches incumbent's base fare levels

B* Entrant discounts fares by 20%

C* Both operators discount fares by 20%.

Table 1 examines the possible demand impacts of fringe competition. It indicates that two additional peak services provided by a new entrant may attract between 6% and 12% of the market and grow the market by between less than 1% and more than 10%, depending principally on whether there is fares competition or not. When the entrant replaces two of the incumbent's peak services, it can capture up to 15% of the market but the overall market size decreases slightly. This is because it is in the entrant's interest not to serve some intermediate stations previously served by the incumbent but an abstractive service of this type is unlikely to be in the public interest.

Table 2 indicates that with matching competition, in which the entrant provides the same service frequency as the incumbent, the entrant can capture between 45% and 57% of the market. However, the overall market will only grow by between 6% and 19%. Again, this is largely because the entrant will not serve some intermediate stations. However, the incumbent also has an advantage in that its existing timetable should have been designed to best match customers' preferred arrival times.

Table 2**. Sample head-on competition simulation results – Route GB1**

Model run	Fare incumbent	Fare entrant	Inter-availability of tickets	Incumbent share (%)	Rail market growth (%)
13	0	0	Yes	54.8	8.6
14	0	0	No	54.0	6.1
15	0	-10%	Yes	48.7	11.2
16	0	-10%	No	43.6	8.6
17	-10%	-10%	Yes	55.1	13.6
18	-10%	-10%	No	54.4	11.1
19	-10%	-20%	Yes	48.9	16.3
20	-10%	-20%	No	43.8	13.6
21	-20%	-20%	Yes	55.3	18.7
22	-20%	-20%	No	54.8	16.1

Note: Entrant matches service frequency of incumbent.

Similar work in Sweden modelled the effect of various competitive scenarios for two lines. The results are shown by Tables 3 and 4 which summarise the findings with respect to a high frequency inter city service, with an average service frequency of less than one hour (Route S1), and a low frequency inter city service, with an average service interval in excess of two hours (Route S2) respectively. Route S1 has approximately two million passengers per annum, with commuting at both ends of the route, whereas Route S2 has only around 0.25 million passengers per annum, with commuting at only one end of the route. Two service options are examined: the entrant matches the number of services provided by the incumbent or the entrant only runs one train in each direction in the peak periods (two trains for the high frequency service). This is referred to as fringe competition. With respect to fares it is assumed that the entrant matches the incumbent's fares or introduces 10% or 20% reductions across all ticket types. The incumbent either maintains existing fare levels or matches the entrant's fare reductions. It is assumed that tickets are not interavailable between operators.

Table 3 shows that for Route S1 if an entrant matches the incumbent's fares and services it gains a 53% market share. This is greater than 50% because the entrant can design a timetable to give particularly good coverage of the busiest times of day. In practice, the incumbent would adjust its existing departures in the light of the entrant's timetable, initiating an iterative process that might be expected to result in equal market shares. Fares competition from the entrant can have a dramatic effect on the incumbent's market share – reducing it from 47% to 6%. Fares competition has a greater impact on the high frequency route because the fare reductions more than compensate for the adjustment of schedules. Fringe competition from the entrant has a minimal impact, capturing 1% of the market without fare reductions, rising to 15% of the market with a 20% fare reduction. If the incumbent matches the entrant's fare reductions, it reduces the entrant's market share back to 1%. In this instance competition may not be academic. Both matching and fringe competition can be profitable for both parties.

Table 3. **Competition on a high-frequency inter city route S1**

Fare incumbent	Fare entrant	Service incumbent	Service entrant	Total patronage base=100	Incumbent market share	Entrant market share
As Now	Match	As Now	Match	112	47%	53%
As Now	-10%	As Now	Match	126	15%	85%
As Now	-20%	As Now	Match	139	6%	94%
-10%	-10%	As Now	Match	130	47%	53%
-20%	-20%	As Now	Match	144	47%	53%
As Now	Match	As Now	Fringe	101	99%	1%
As Now	-10%	As Now	Fringe	103	95%	5%
As Now	-20%	As Now	Fringe	105	85%	15%
-10%	-10%	As Now	Fringe	122	99%	1%
-20%	-20%	As Now	Fringe	136	99%	1%

Table 4 shows that for the low frequency service (S2) an entrant that matches the incumbent's fares and service levels can capture 56% of the market. This is greater than 50% for the same reasons as for Route S1, but in the low frequency case there are more gaps in the timetable at busy times of day for the entrant to fill. Fares competition from the entrant can reduce the incumbent's market share further from 44% to 30%. If the incumbent matches the entrant's fare cuts, it returns to obtaining a 44% market share. With fringe competition, the entrant can capture 23% of the market without fare cuts, rising to 31% with a 20% fare reduction. If the incumbent matches these fare cuts, the entrant's market share is reduced back to 23%. It should be noted that for such a low frequency route, competition may be largely academic as none of the scenarios examined revealed a profitable entry opportunity.

Table 4. **Competition on a low-frequency inter-city route S2**

Fare incumbent	Fare entrant	Service incumbent	Service entrant	Total patronage base=100	Incumbent market share	Entrant market share
As Now	Match	As Now	Match	122	44%	56%
As Now	-10%	As Now	Match	127	37%	63%
As Now	-20%	As Now	Match	133	30%	70%
-10%	-10%	As Now	Match	131	44%	56%
-20%	-20%	As Now	Match	140	44%	56%
As Now	Match	As Now	Fringe	108	77%	23%
As Now	-10%	As Now	Fringe	110	73%	27%
As Now	-20%	As Now	Fringe	112	69%	31%
-10%	-10%	As Now	Fringe	116	77%	23%
-20%	-20%	As Now	Fringe	125	77%	23%

PRAISE is not an equilibrium model. Instead it is a model that is used to assess the impact of a number of scenarios. An example for Route GB1 is given by Table 5.

This analysis suggests that matching competition is not feasible in most instances. However, Table 5 suggests that fringe competition may be feasible in certain circumstances (for example, if there is regulation to ensure interavailability of tickets – model run 5). However, in most cases welfare does not increase, with the exception of model run 11.

Route GB1 is paralleled by a slower Route GB1A, with end to end journey times one hour longer. It was found that if fares on Route GB1A were set at 50% of those of GB1, then Route GB1A could capture 25% of the end to end market. We were not able to undertake a welfare analysis of this scenario, as we did not have full demand and cost data for Route GB1A. However, this analysis suggests that route competition based on product differentiation may be possible and has been a feature of a number of origin and destination pairs, most notably between London and Birmingham.[3]

An example of the PRAISE model results for the Inter City Route S1 in Sweden is given by Table 6. It should be noted that this route is paralleled by the slower services of Route S1A, which has end to end journey times that are around an hour longer. Route S1A has around one million passengers per annum. This Table shows that, with a 20% cost reduction and no interavailable tickets (arguably the most likely competitive scenario), fringe entry (scenarios 68 to 72) is profit enhancing in that it encourages a shift from Route S1A services with low profit margins to Route S1 services with relatively high profit margins. Head-on competition (scenarios 63 to 67) reduces overall profits by up to 30%, although the Route S1 services remain profitable in total. The demand for Route S1 services, measured in terms of passengers, might increase by over 40% but the change in demand for Route S1 and S1A services combined is more modest (with a maximum growth of 12%). Consumers suffer disbenefits in some scenarios because the increases in service frequency are insufficient to compensate for the lack of integrated ticketing between Route 1A feeder services and Route 1 trunk services.

Table 5**. Example results from the PRAISE Model – Inter-city route GB1 (per day)**

Model run	**Fares**	**Entrant service pattern**	**Inter-availability of tickets**	**Incumbent profit (#)**	**Entrant profit**	**Consumer surplus change (Business)**	**Consumer surplus change (Non-business)**	**Welfare change**
1	A*	1*	Yes	30 815	1 267	1 528	82	-9 051
2	A*	1*	No	31 962	-847	891	82	-10 657
3	B*	1*	Yes	12 419	16 670	4 686	791	-8 178
4	B*	1*	No	17 799	10 379	3 510	512	-10 544
5	C*	1*	Yes	23 545	528	12 741	4 548	-1 383
6	C*	1*	No	25 017	-2 135	12 055	4 483	-3 326
7	A*	2*	Yes	29 591	11 381	-3 578	-1 046	-6 397
8	A*	2*	No	29 553	9 183	-4 603	-1 153	-9 765
9	B*	2*	Yes	20 050	18 888	446	-210	-3 570
10	B*	2*	No	22 158	14 700	-845	-507	-7 239
11	C*	2*	Yes	23 241	10 259	7 592	3 380	1 727
12	C*	2*	No	23 240	7 999	6 466	3 230	-1 810

Notes:
1* Entrant provides two additional express services in the morning and evening peak periods in both directions of travel.
2* System is at capacity, the entrant replaces two of the incumbent's services in the morning and evening peak periods in both directions of travel with express services.
A* Entrant price matches incumbent's base fare levels
B* Entrant discounts fares by 20%
C* Both operators discount fares by 20%.
Incumbent base profit GBP 42 ,746.

Table 6**. Example results from the PRAISE Model – Inter-city routes S1 and S1A**

Scenario	Fare Route 1A	Fare Route 1 -Inc	Fare Route 1 -Entrant	Service Route 1A	Service Route 1 -Inc	Service Route 1 -Entrant	Route 1 Pax change	Routes 1 & 1A Profit change	Routes 1 & 1A CS change *	Routes 1 & 1A Welfare Change *
63	As Now	As Now	Match	As Now	As Now	Match	12.3%	-26.0%	-8.6%	-34.6%
64	As Now	As Now	-10%	As Now	As Now	Match	25.5%	-22.7%	12.3%	-10.4%
65	As Now	As Now	-20%	As Now	As Now	Match	38.5%	-27.1%	42.6%	15.6%
66	As Now	-10%	-10%	As Now	As Now	Match	30.0%	-18.9%	20.8%	1.9%
67	As Now	-20%	-20%	As Now	As Now	Match	43.0%	-23.1%	54.4%	31.3%
68	As Now	As Now	Match	As Now	As Now	Fringe	1.6%	42.3%	-20.3%	22.0%
69	As Now	As Now	-10%	As Now	As Now	Fringe	2.5%	42.6%	-19.5%	23.1%
70	As Now	As Now	-20%	As Now	As Now	Fringe	4.9%	41.3%	-16.6%	24.7%
71	As Now	-10%	-10%	As Now	As Now	Fringe	21.8%	54.4%	7.1%	61.5%
72	As Now	-20%	-20%	As Now	As Now	Fringe	36.4%	53.0%	39.3%	92.2%

Note: Inc = Incumbent, Pax = Passenger, CS = Consumer Surplus, * Expressed relative to base profit and a base situation in which tickets are interavailable.

Our analysis suggests that with open entry, the most likely outcome is scenario 67, which involves head-on competition with fare cuts. This leads to an increase in welfare equivalent to 31% of base profits. It should be noted that if the incumbent is forewarned of entry it is likely to blockade such an opportunity by doubling frequency itself. Moreover, it should also be noted that a regulated monopolist in which service levels are reduced slightly but fares are cut by 20% could increase welfare by a greater amount, equivalent to 118% of base profits.

Overall, on Route S1, of the scenarios examined, unconstrained profit maximisation was found to be similar to the welfare maximising scenario. However, both situations require the Route S1A services to be subsidised. This suggests that Route S1 services operated as a regulated monopoly for high speed services may promote static efficiency, provided there is fringe competition from Route S1A conventional services in receipt of appropriate amounts of subsidy and inter-modal competition from car, coach and air. Also, there appears to be a strong welfare case for lower fares on Route S1 services compared to the current situation.

Further analysis indicated that, where tickets are not interavailable, it is still possible for two operators to be profitable with head-on competition but matching fare reductions of around 10% are more likely. With cost reductions, competition becomes more feasible but is still undesirable, although to a reduced degree. Although it is possible for two Route S1 operators to be profitable with head-on competition, even with interavailable tickets, the increase in welfare is only around one half of the maximum we have found. If tickets are not interavailable, the increase in welfare is only around a quarter of the maximum we have found. Welfare is maximised where fares are reduced by 20% and service levels are reduced slightly on route S1 whilst fare and service levels on route S1A are unchanged.

For the low frequency Inter City Route S2, in the base it is found that the service is loss making with a cost recovery ratio (expressed as a percentage) of around 60%. However, this is based on fully accounted costs where administration costs comprise 15% of total costs, whilst revenue calculations do not take into account contributory revenue elsewhere on the network and off train revenue. When these facts are taken into account we find that the service is close to break-even with current costs and will be profitable with the introduction of new rolling stock.

Overall, the modelling for route S2 indicates that competition is not feasible with current cost levels. Welfare is maximised when there are substantial fare reductions and modest service reductions. Losses are reduced by more than a third. By contrast, profit maximisation would involve substantial fare increases and service reductions that would lead to a halving of losses but an increase in welfare of only one sixth of the maximum found. With cost reductions of 20%, the profit maximising scenario and the welfare maximising scenario remain dissimilar in their welfare impacts, although the service can get close to break-even. If tickets are interavailable, there may be scope for some fringe competition on peak days (Fridays and Sundays when demand is double average weekday levels – see for example Jansson, 2001) but this reduces welfare.

It is possible to generalise the results of these computer simulations. A generic version of the PRAISE model was developed for the Strategic Rail Authority (Whelan, 2002) and meta-analysis of model runs has been undertaken to determine reaction functions. These results indicate that in Great Britain with prevailing track access rates, head-on competition is not commercially feasible, even if sufficient capacity was available. However, cream skimming entry with train movements focussing on the peak times and directions of travel and/or niche entry through various forms of product differentiation could be commercially feasible, particularly if there was regulation to ensure inter-availability of tickets. Moreover, competition would lead to service withdrawal from thinner markets (in this case lightly used intermediate stops) and a concentration on thick markets–a phenomenon also observed in the deregulated express coach market (Cross and Kilvington, 1985) and in the German passenger rail market (Séguret, *op cit*.).

By contrast, the work in Sweden indicated that with lower track access charges, head-on competition was commercially feasible on the busiest routes, although it might not be technically feasible because of capacity constraints. However, such competition was not desirable because it led to too much service, at too high fares, compared to the welfare maximising configuration which involved substantial fare reductions on the busiest route. An interesting feature was the importance of competition between parallel routes. If the slower route was subsidised so that fares and frequency were set at the welfare maximising level then a profit maximising monopolist on the fast route would probably produce at a fares-frequency combination that was close to the welfare maximum. Competition was not found to be feasible for thinner routes in Sweden.

The overall conclusion from models of this type is that competition in long distance rail markets, where it occurs, is not characterised by oligopoly (either of the Cournot or Bertrand type) but is likely to take the form of oligopolistic competition of the type described by Salop (1979) and Novshek (1980). This will involve too much service at too high fares, but also with spatial and temporal bunching.

The finding that competition in rail markets does not generally enhance welfare requires numerous qualifications. The first is that it is assumed that firms are already cost efficient. Where this is not the case, competition may be a powerful tool to promote cost efficiency. The second is that dynamic efficiency is ignored. There may be an argument that competition promotes innovation, particularly with respect to product differentiation, and this has not been taken this into account. A third, and related point, is that uniform pricing is assumed, at least for individual segments. Competition may particularly promote innovation in pricing, stimulated by technological developments in delivery channels such as the internet, smart cards and mobile telephony. As a result, modelling work is now focusing on intra-modal and even intra-firm competition between ticket types (Wardman and Toner, 2003).

3. OFF TRACK COMPETITION

It was noted in the introduction that off track competition for long distance rail service has been limited. In part, this may be because such services already face competition from car and coach for shorter distances and from air for longer distances. It also reflects that the case for subsidising long distance rail services is not strong. First best arguments for subsidisation related to user benefits increasing with service output (the Mohring effect) are limited for infrequent services where passengers time their arrival to match train departure times, whilst second best arguments related to the relief of road congestion are also diminished. As a result, there may be predilection for competition in the market for long distance services, as reflected by EC Directive 91/440. However, a combination of institutional inertia and limited commercial opportunities means that the development of such competition has also been limited.

The evidence of competition for the market in Great Britain is therefore relevant. Here, there have been three broad rounds of franchising (see also Preston, 2008b). The first round, organised by the Office of Passenger Rail Franchising, was undertaken between 1996 and 1997 and based in the main around seven year net cost contracts, with minimum service levels specified and around 50% of fares regulated. An important exception was the West Coast Main Line where a 15 year franchise was let, as the infrastructure was to undergo an upgrade to permit 200 km per hour tilting Pendolino trains, an upgrade which was only completed in 2008.

The second round, associated with the Strategic Rail Authority, saw some eight franchise re-let. Initially the focus was on longer franchise for 20 years in which the operator was given greater commercial freedom. In the event only two such franchises were let – for the urban services centred on Liverpool (Merseyrail) and for Chiltern Rail (which does include some long distance services from Birmingham (and beyond) to London). The rest of the re-let franchises were in response to the financial meltdown in the industry that resulted from the Hatfield accident in 2000 and the subsequent failure of Railtrack and some 13 of the 25 Train Operating Companies (Nash and Smith, 2006). Thompson (2006) notes that of these 13 failures only two were long distance operators whose holding company (Virgin Trains) had been affected by the delays and cost over-runs on the West Coast Main Line. Partly as a result of these franchise failures, there was a switch back to more tightly specified, shorter franchises.

The third phase of franchising – run by the Department for Transport (DfT) since 2005 – has seen ten further franchises re-let. A feature of this round is that the distinction between long distance intercity franchises and suburban and regional franchises has become blurred, with the Great Western incorporating the former Thames (London commuter services out of Paddington) and Wessex (regional services in the South West) franchises. Similarly, the Midland Main Line franchise was merged with some regional services to form East Midlands Trains. One feature of the third round is the cap and collar incentive regime which shares commercial risk between the franchisor and the franchisee. This typically means that after the first four years of the franchise contract have passed: 50% of any fares revenues in excess of 102% of the TOC’s original forecasts are shared with DfT; DfT makes a contribution equivalent to 50% of any revenue shortfall below 98% of the TOC’s original forecast; and for any short fall below 96%, DfT’s contribution increases to 80%. This revenue risk-sharing mechanism is intended to constrain overzealous bidding, something that was a particular feature towards the end of the first round (see Preston *et al.*, 2000). However, it may encourage backloading in which bids are more aggressive in later years when the risk sharing comes into force.

One initial concern about off track competition was that it may not prove to be very competitive (Preston, 1996). This has not proved to be the case given that the privatised bus companies have been heavily involved in bidding from the start, whilst interest from international organisations has grown so that currently organisations from France, Germany, Hong Kong and the Netherlands have stakes in franchised rail operators. In the first phase, there were an average 5.4 bids per franchise. This has reduced slightly so that there were 4.2 bids per franchise in the second phase and 3.8 bids per franchise in the third phase. There is some concern that high bidding costs (which are estimated at around GBP5 million per bidder) may be deterring entry.

Table 7. **The East-Coast franchise**

	Date Started	Expected Duration	PVNP 1st year (GBP m)	PVNP Final year (GBP m)
GNER	April 1996	7 years	65[1]	0
GNER	May 2005	10 years	(50)	(219)
National Express	Dec. 2007	7 ¼ years	7	(311)

PVNP = Present Value of Net Payments. Figure in brackets denote premiums paid.
[1]Out-turn.
Source. Preston and Root (1999) and www.dft.gov.uk.

An interesting case study is provided by the East Coast Franchise, the core of which is long-distance intercity services between London King's Cross and Leeds/Edinburgh. Table 7 gives some basic data. In the first round of franchising, the winning bid for this franchise was from Great North Eastern Railways (GNER), a subsidiary of the shipping company Sea Containers. This service required some GBP65 million of subsidy in the first year of operation falling to zero subsidy in the seventh year. Given the relatively good performance of GNER and uncertainties following Hatfield a two year extension was negotiated, prior to refranchising in 2004. The incumbent operator put in a robust bid which involved paying a premium of GBP50 million in the first year, rising to GBP219 million in the tenth year, indicating some backloading. However, the trade press indicated that the incumbent's bid was only a little higher than the second highest bid. This bid was accepted and GNER started operating its renewed franchise in May 2005. However, this bid was quickly overtaken by a series of events. GNER had not anticipated the upsurge in fuel costs that occurred in 2005/6, revenue was hit by the 7th July 2005 bombings in Central London and entry by an open access operator, Grand Central, would abstract some revenue from GNER, particularly at York. To confound matters, GNER's parent company was also in financial difficulties. It quickly became clear that GNER could not meet its premium payments and there was still three years before the cap and collar scheme came into force. In December 2006, GNER entered into a Management Agreement with the DfT, based on an incentive if revenue growth exceeds an agreed target. Almost immediately, the process of re-letting the franchise was begun.

The bids for these were submitted in June 2007 and the award announced in August. The winning bid came from the National Express Group, who began operations in December 2007. Again, the bid was a robust one. Although for the first year of operation a subsidy of GBP 7 million was required this would quickly convert into a premium of GBP 311 million some seven years later, again indicating backloading. There was some concern that National Express was buying-in work, given that it had lost a number of franchises (including Central, Midland Mainline and Scotrail) but the trade press also indicated that National Express was not the highest bidder. Once again, the bid was overtaken by events. In the light of the credit crunch, the 10% per revenue growth on which the bid was predicated was unlikely. In the light of this, and problems with the parent company, in July 2009, National Express East Coast announced that it would only be able to meet its contractual obligations up to the end of 2009. Mindful of evidence that re-negotiations would lead to cost increases of the order of 23-28% (Smith and Wheat, 2009), the Government fulfilled its earlier commitment not to negotiate and prepared to exercise its operator-of-last-resort powers, a role it had previously exercised for South East Trains (formerly operated by Connex) between 2003 and 2006. National Express East Coast will surrender a GBP 32 million performance bond and in combination with accumulated losses will face an exposure of some GBP 100 million. The Government is also minded to enforce cross-default provisions so that National Express will have to give up its other two (profitable) franchises. However, there is some uncertainty over the future of National Express as a whole, with at the time of writing, Stagecoach plc (the operator of two franchises and the part owner of another) considering a take-over bid.

One of the dangers of contracting-out, particularly in railways, is related to the gaming behaviour that can occur. In particular, there is the practice of low-balling in which bidders post an initial high bid in the belief that they can then re-negotiate or chisel on the offered level of quality. The performance regime for railways in Britain (with financial penalties for poor reliability and overcrowding) largely precludes the latter option. Re-negotiation is a high risk strategy and one that might involve a loss of reputation, but is predicated on at least three points. Firstly, the private sector is gambling that no Government could afford to let the railways (or a part of it) go bust. Secondly, in circumstances of a likely franchise failure, re-negotiations may be less costly (and speedier) than re-franchising. Thirdly, the private sector is assuming that in any re-negotiations it will exhibit better negotiation skills (and be able to devote more resources to this task) than the public sector. In so doing, it may be assisted by information asymmetries. There is some evidence that low-balling occurred in the first round of franchising, albeit it unsuccessfully in the case of Connex but perhaps with more success in the case of Virgin. Thompson (2006) notes that low-balling has been a feature of rail franchising elsewhere, particularly in Australia and Latin America. In the third round of franchising, low-balling does not seem to be effective, given the Government's firm stance on no renegotiations, implementation of cross-default provisions and recovery of a performance bond. However, the failure of the East Coast franchisee twice in three years is obviously a cause for concern and suggests that there are problems with the "winner's curse". Options might involve moving away from net subsidy to gross cost contract (as has occurred for the London Overground franchise) but this would weaken operator incentives to grow revenue, or considering flexible-length contracts which terminate once a franchisee has made its premium payments in PV terms – in effect a variant of the least present value-of-revenue approach advocated by Engel *et al.* (2001).

4. ON TRACK COMPETITION

As indicated above, open access competition in Britain has been moderated by the Office of Rail Regulation. In the first phase of moderation, open access competition was restricted to origin and destination pairs that constituted less than 0.2% of a franchisee' revenue – effectively limiting competition to where franchises overlapped (see Shaw, 2000). In the second phase, which operated up to 2002, franchisees could register revenue flows and could only face competition on 20% of registered flows but all unregistered flows would be open to competition. In the third phase, from 2002 onwards a more case by case approach has been adopted where services have to demonstrate that they are not primarily abstractive. It appears that the relevant threshold is that generated traffic needs to be at least 30% abstracted traffic (Griffiths, 2009). So far there have been three instances of open access competition, with a further case approved. These are Hull Trains, which has been operating services between Hull and London via the East Coast Main Line since 2000; Grand Central which has been operating services between London and Sunderland, also via the East Coast Main Line, since 2007; and Wrexham, Marylebone and Shropshire Railway, which has been operating services between Wrexham and London since 2008. In addition, Grand Northern has been licensed to provide services between Bradford and London, but has not yet started operation. Three open access proposals have been rejected: a Grand Central service between Preston and Newcastle via Manchester and Leeds; a Hull Trains service between Harrogate and London and a Platinum Trains service between Aberdeen and London. Currently non-franchised operations[4] account for 0.1% of passenger journeys, 0.6% of passenger revenue 0.8% of passenger kms revenue and 1% of train kms on the national network. (ORR, 2009).

Table 8. **Open access services, Summer 2009**

	Franchisee's trains per day	Open access trains per day	Franchisee super off peak return	Open access Off peak return
London – Hull	1 (19)	7	GBP85	GBP69
London –Sunderland	0 (23)	3	GBP105	GBP71

Table 8 shows some data for the two most established open access operators both of which are providing direct services to London from cities on the East Coast of England with populations of around 250 000 that have traditionally been poorly served by rail. The franchised operator in the main provides indirect services via Doncaster in the case of Hull and via Newcastle in the case of Sunderland. It can be seen that compared to these franchised services, the open access operator only provides 27% of service in the case of Hull and 12% in the case of Sunderland. However, headline fares for the open access operator are some 18% lower in the case of Hull and 32% lower in the case of Sunderland. This has resulted in large increases in demand. Rail travel between London and Hull has grown by some 60%, whilst on the uncompleted Grimsby to London route growth has only been around 10%. In terms of revenue, the first four Hull Trains services were estimated to have a generation to abstraction ratio of 0.7:1. Another feature of open access services is the high percentage of passengers on the main flows travelling on dedicated tickets – well above the 10% threshold used by the Competition Commission (2005) and in some case above 50%.

Table 9. **Economic benefit of open access services (GBP m)**

	Hull Trains		Grand Central	
	PV 5 years	PV 10 years	PV 5 years	PV 10 years
Economic benefit	47.3.	96.9	18.4	38.2
Net financial cost	29.1	45.4	15.5	24.3
Net Present Value	18.1	51.5	2.9	14.0
Benefit Cost Ratio	1.62	2.13	1.19	1.57

Source: MVA, 2009.

Table 9 shows that there appears to be a strong economic case for both the Hull Trains and Grand Central services, with a ten-year benefit : cost ratio in excess of 1.5 for both services.

5. CONCLUSIONS

Competition for long distance rail services remains relatively limited. On track competition, where it has occurred seems to focus on niche markets which the incumbent operator has neglected. However, modelling work indicates that if track access charges are based on short run marginal cost, head-on competition may be feasible for densely trafficked routes but not necessarily socially desirable, with a tendency to result in too much service, at too high fares. Where track access charges are based on fully allocated costs, competition may be more limited, even for densely trafficked routes, and this competition may have some cherry picking characteristics. Again, competition may be feasible (particularly if there are regulations enforcing interavailable ticketing) but not necessarily desirable. By contrast, analysis of the niche open access entry in Britain, based on marginal cost based track access charges, does appear socially desirable. An interesting question is whether the ratio of generated to abstracted traffic is a useful indicator. The most likely outcome for the heavily trafficked route in Sweden (S1- head-on competition, Table 6, Model Run 67) results in a ratio of 0.57, well in excess of the ORR's 0.3 threshold. By contrast, the most likely outcome on the heavily trafficked route in Great Britain (GB1–fringe competition, Table 5, model run 1) gives a ratio of only 0.10. With head-on competition and matched fares (Table 2, model run 13) this ratio increases to around 0.18. However, the ratio become difficult to interpret when there are matched fare cuts. For example, with fringe competition and fare cuts (model run 5, Table 5) generated traffic exceeds that abstracted by the entrant. However, this scenario results in an 11% reduction in total revenue and a welfare loss. Interestingly, for Table 5, model run 11 (fringe competition in which the entrant replaces the incumbent fro some services with matching fare cuts), the ratio is 0.6. This option is welfare enhancing despite a 14% reduction in total revenue, although this is partly due to the entrant cutting out some intermediate stops. Some of these results have echoes of the work undertaken by SDG (2004) that found that competition on European high speed rail routes was feasible, provided track access charges were based on marginal costs and provisions were made for interavailable tickets, but the case is not particularly robust.

Off track competition is relatively untested for long distance services, particularly those that are good commercial prospects, with the main evidence coming from Great Britain. Such a model has been able to attract sufficient numbers of bidders, has coincided with strong demand growth and can result in large premia being paid to the franchisor. However, such competition is vulnerable to the winner's curse which may be exposed by unexpected events (Hatfield, the 7/7 bombings, the credit crunch). Risk sharing mechanisms may reduce this exposure but do not remove it all together and alternative contractual models may be worth considering including flexible term contracts and Vickrey style second best auctions.

Where on track competition provides direct services to new markets, experience from Great Britain indicates this is commercially feasible and socially desirable, but capacity constraints on the main lines and at key terminals mean that such competition may be limited and there is the wider issue of whether these services are making the best use of limited capacity. There are indications from modelling work in both Britain and Sweden that route competition can be beneficial, but this will be limited by railway geography, although the scope for such competition will increase where new high speed lines are being constructed.

The overall impression is that the evidence in support of competition for long distance rail services, either in the market or for the market, is mixed. Indeed a commercial 'monopoly' may approximate a first-best solution if some conditions are met. First, this monopoly needs to face modal competition, particularly from deregulated coach and air markets. Secondly, where feasible this monopoly should face route competition. This may take the form of product differentiation, with the alternate route being slower but cheaper. Where there is sufficient capacity such differentiation may be provided on track, with express services competing with stopping services. It could be that the slower services are in receipt of subsidy, in which case they should be competitively tendered. Third, where possible there may be some benefits in terms of niche competition in which infrequent direct services compete with frequent indirect services. Of course, if these conditions are met then the commercial operator does not really have a monopoly, at least for significant parts of its market, although it may have some incumbency advantages. Where such conditions can not be met, then some competition for the market might be considered.

NOTES

1. We consider long-distance services as serving city pairs that are more than 50 miles (80 km) apart, although there may be intermediate stops.

2. These were the night ferry service from Berlin to Malmo, the InterConnex service between Leipzig and Rostock (via Berlin) and the Vogltand-Berlin and Harz-Berlin services.

3. Currently, Virgin Trains operate express services between Birmingham New Street and London Euston, with London Midland operating stopping services. Chiltern Trains operate stopping services between Birmingham Snow Hill and London Marylebone.

4. Also include Heathrow Express.

5. Now published by the Association of Train Operating Companies. Version 5 was released in 2009.

BIBLIOGRAPHY

Alexandersson, G. (2009), The Complexity of Market Structure – Prospects for On-the-Track Competition in Sweden. Presented to the 11th International Conference on Competition and Ownership in Land Passenger Transport. Delft, the Netherlands.

British Railways Board (BRB)[5] (1990), *Passenger Demand Forecasting Handbook*, Version 2, London: BRB.

Competition Commission (2005), *First Group/ICEC Merger Inquiry: Commentary on Issues Statement*, Competition Commission, London.

Cross, A.K. and R.P. Kilvington (1985), The Deregulation of Inter-City Coach Services in Britain, *Transport Reviews*, 5, pp. 225-245.

Dementiev, A. (2007), "Vertical Divestiture as a Competitive Strategy: The Case of Railway Passenger Transport Reform in Russia", Presented to the 10th International Conference on Ownership and Competition in Public Transport, Hamilton Island, Queensland.

Douglas, N. (1987), A Welfare Assessment of Transport Deregulation – the Case of the Express Bus Market, Aldershot: Gower.

Engel, E., R. Fischer and A. Galetovic (2001), "Least-Present-Value-of-Revenue Auctions and Highway Franchising", *Journal of Political Economy*, Vol. 109, pp. 993-1006.

Gomez-Ibanez, J. and G. de Rus (2006), Competition in the Railway Industry: an International Comparative Analysis, Edward Elgar, Cheltenham.

Griffiths, T. (2009), "On Rail Competition: The Impact of Open Access Entry on the Great Britain Rail Market", Presented to the 11th International Conference on Competition and Ownership in Land Passenger Transport, Delft, the Netherlands.

Jansson, K. and R. Mortazavi (2000), "Models for Public Transport Demand and Benefit Assessments", in: D. Hensher and K. Button (eds.), *Transport Modelling*, Pergamon, Oxford.

Jansson, J.O. (2001), "Efficient Modal Split", Proceedings of the 7th International Conference on Competition and Ownership in Land Passenger Transport, Molde, Norway.

Jörgensen, F. and J. Preston (2003), "Estimating Bus Operator's Short-run, Medium-term and Long-run Marginal Costs", *International Journal of Transport Economics*, Vol. 30, Issue 1, pp. 3-24.

Kain, P. (2006), "The Pitfalls in Competitive Tendering: Addressing the Risks Revealed by Experience in Australia and Britain", ECMT Workshop on Competitive Tendering of Rail Services, Paris.

Kreps, J. and J. Scheinkman (1983), "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes", *Bell Journal of Economics*, Vol. 14, Issue 2, pp. 326-337.

Lardner, D. (1850), *Railway Economy: A Treatise on the New Art of Transport*, Taylor, Walton and Maberly, London.

Mizutani, F. (1994), Japanese Urban Railways. A private-public comparison, Aldershot: Avebury.

MVA (2009), Assessment of Alternative Track Access Applications on the East Coast Main Line, MVA, London.

Nash, C. and A. Smith (2006), Passenger Rail Franchising – British Experience, ECMT Workshop on Competitive Tendering of Rail Services, Paris.

Nash, C. (2009), "European Rail Reform – The Next Steps", Presented to the 11th International Conference on Competition and Ownership in Land Passenger Transport. Delft, The Netherlands.

Novshek, W. (1980), "Equilibrium in Simple Spatial (or Differentiated Product) Models", *Journal of Economic Theory*, 22, pp. 313-323.

Office of Rail Regulation (2009), *National Rail Trends, Quarter 1, 2009/10,* ORR, London.

Preston, J. and A. Root (1999), Great Britain, in: van de Velde, D. (ed.), *Changing Trains. Railway reform and the role of competition: the case of six countries*, Ashgate, Aldershot.

Preston, J., G. Whelan and M. Wardman (1999), An Analysis of the Potential for On-track Competition in the British Passenger Rail Industry, *Journal of Transport Economics and Policy*, Vol. 33, Issue 1, pp. 77-94.

Preston, J.M., G. Whelan, C. Nash and M. Wardman (2000), The Franchising of Passenger Rail Services in Britain, *International Review of Applied Economics*, Vol. 14, Issue 1, pp. 99-112.

Preston, J., T. Holvad, N. Sykes and J. O'Reilly (2001), "Review of Empirical and Theoretical Evidence. Deliverable D2, Development of Market Models for Increased Competition in Railroad Passenger Traffic", Working Paper 906, Transport Studies Unit, University of Oxford.

Preston, J., T. Holvad and F. Rajé (2002), Track Access Charges and Rail Competition: A Comparative Analysis of Britain and Sweden, European Transport Conference, Cambridge.

Preston, J. (1996), The Economics of British Rail Privatization: An Assessment, *Transport Reviews*, Vol. 16, Issue 1, pp. 1-21.

Preston, J. (2008a), Competition in Transit Markets, *Research in Transportation Economics*, 23, 75-84.

Preston, J. (2008b), A Review of Passenger Rail Franchising in Britain: 1996/7-2006/7, *Research in Transportation Economics*, 22, pp. 71-77.

Quinet, E. and R. Vickerman (2004), *Principles of Transport Economics*, Cheltenham: Edward Elgar.

Rosenlind, S., G. Lind and G. Troche (2001), *LIME: Model for Capacity Utilisation and Profitability of a Railway Line*, Stockholm: Royal Institute of Technology.

Salop, S. (1979), Strategic Entry Deterrence, *American Economic Review*, 69, pp. 335-338.

Séguret, S. (2009), "Is Competition On-track a Real Alternative to Competitive Tendering in the Railway Industry? Evidence from Germany", Presented to the 11th International Conference on Competition and Ownership in Land Passenger Transport. Delft, The Netherlands.

Sharipov, T. (2009), "Results So Far and Prospects of Kazakhstan Passenger Rail Franchising", Presented to the 11th International Conference on Competition and Ownership in Land Passenger Transport. Delft, The Netherlands.

Shaw, J. (2000), Competition, Regulation and the Privatisation of British Rail, Ashgate, Aldershot.

Smith, A. and P. Wheat (2009), "The Effect of Franchising on Cost Efficiency: Evidence from the Passenger Rail Sector in Britain", presented to the 11th International Conference on Competition and Ownership in Land Passenger Transport. Delft, The Netherlands.

Steer Davies Gleave (2004), "EU Passenger Rail Liberalisation: Extended Impact Assessment. Report for DGTREN", European Commission, Brussels.

Thompson, L. (2006), "Competitive Tendering in Railways: What is Experience Telling Us?", ECMT Workshop on Competitive Tendering of Rail Services, Paris.

van de Velde, D. (2009), Development of Railway Contracting for the National Passenger Rail Services in The Netherlands", presented to the 11th International Conference on Competition and Ownership in Land Passenger Transport. Delft, The Netherlands.

Wardman, M.R. and J.P. Toner (2003), "Econometric Modelling of Competition Between Train Ticket Types", AET European Transport Conference, Strasbourg, France.

Whelan, G.A. (2002), "Analysing impact of changes in on-rail competition—model development". Unpublished Report to the Strategic Rail Authority, Institute for Transport Studies, University of Leeds.

Wilson, N. and A. Nuzzolo (2004), *Schedule-Based Dynamic Transit Modelling: Theory and Application*, Boston: Kluwer Academic Publishers.

Theme IV:

Transport System Interactions and Innovation

WHEN SHOULD WE PROVIDE SEPARATE AUTO AND TRUCK ROADWAYS?

Robert W. POOLE, Jr.

Reason Foundation
Los Angeles
California
United States

SUMMARY

1. INTRODUCTION

The concept of the general purpose (GP) lane has dominated modern highway thinking and practice in OECD countries, especially for limited-access highways such as inter-city motorways and urban expressways, whether tolled or non-tolled. This paper raises the question of whether, in some circumstances, specialized lanes for light vehicles (cars, vans and pickup trucks) and heavy vehicles (generally more than two axles) might be cost-effective.

The case for GP lanes appears to rest on two principal advantages: capacity and cost-savings. First, for road capacity in a single direction, the provision of two GP lanes permits somewhat higher throughput (vehicles/lane/hour) than two separate lanes. That is because with more than one lane, faster vehicles can pass slower-moving vehicles. This effect is less pronounced as the total number of lanes per direction increases, but even with four or five lanes in one direction (as on some Californian freeways), reserving one lane for specialized use subjects that lane to the problem of faster vehicles in that lane being unable to pass slow-moving vehicles—and hence that restricted lane is scored by traffic engineers as having lower capacity than the adjacent GP lanes that do permit lane-changing.[1] Special lanes for high-occupancy vehicles (HOVs) are sometimes opposed by traffic engineers for this reason, at least where only one such lane is provided per direction.[2]

The second argument for GP lanes concerns cost. Separate lanes are generally proposed for a subset of vehicles. In the United States today, the vehicle categories most often proposed for "managed lanes" are carpools (HOV lanes), buses (exclusive busways), toll-paying vehicles (HOT or Express Toll Lanes) or trucks (truck-only lanes). However, if the fraction of vehicles eligible to use the special lane is a significantly higher or lower percentage of the projected daily traffic than one lane's worth, the special lane may provide either too little or too much capacity for the designated subset of vehicles. The "lumpiness" of a lane's capacity means that, in general, the risk of building the wrong amount of capacity is less if all the lanes can be used by all types of vehicles—i.e. be operated as GP lanes.

Against this background of conventional wisdom, this paper will explore whether there are cases where, despite these factors, specialized lanes could make sense in coming decades. The next section provides a brief overview of exceptions to the standard GP lane practice, drawn from US experience. Next, the paper examines arguments for cars-only (actually light-vehicles only) roadways or lanes that have emerged in the transportation literature in recent years. This is followed by a comparable review of arguments that have been put forth in favour of truck-only lanes (or roadways). Following the cars-only and trucks-only discussions, the paper further explores the pros and cons of separate versus GP lanes, adding a more detailed consideration of vehicle operators' values of time. This is followed by a discussion of safety and environmental considerations that may be relevant in considering the creation of specialized lanes in coming decades.

2. EXAMPLES OF SEPARATE LANES AND ROADWAYS

2.1. Cars-only parkways

The United States, in the first half of the twentieth century, developed a number of cars-only roadways. They were generally called "parkways" and were the country's first grade-separated and limited-access highways. The parkway phenomenon was especially prominent in the northeastern states and many of these parkways were developed as toll roads. Table 1 lists some examples, most of which are still in operation today, though nearly all without tolls. Parkways were generally built in suburban areas, sometimes in the flood plains of small rivers. They typically followed winding routes through forested areas and were often designed in part by landscape architects who sought to fit them into the existing landscape, minimizing cuts and fills and preserving as much of the treescape and waterways as possible (today this would be called "context-sensitive design"). They generally had low overhead clearances (e.g. 11-feet) aimed at reinforcing the policy of non-use by trucks, had short onramps (often with stop signs) and narrow lanes, typically 10-feet rather than today's US standard of 12-feet. Originally they were not equipped with breakdown shoulders or median barriers and were designed for speeds lower than today's limited-access highways.[3]

Table 1. **Representative US parkways**

State	**Name of parkway**
California	Arroyo Seco Parkway (later became Pasadena Freeway)
Connecticut	Merritt Parkway
	Wilbur Cross Parkway
Maryland	Baltimore-Washington Parkway
	Clara Barton Parkway
	Suitland Parkway
New Jersey	Garden State Parkway
New York	Bronx River Parkway
	Henry Hudson Parkway
	Hutchinson River Parkway
	Interboro Parkway
	Sawmill River Parkway
	Sprain Brook Parkway
	Taconic Parkway
Virginia	George Washington Parkway
	Mt. Vernon Parkway

Source: Peter Samuel, Note 3.

2.2. Separate truck lanes

A second example is the provision of separate truck lanes on major US highways. In most cases, these are provided as climbing lanes at locations where the highway's rather steep grade forces heavy trucks to slow considerably. To prevent these trucks holding up faster traffic, state transportation agencies often designate the right-most lane as a truck-only climbing lane. In a few cases, such as on I-5 north of Los Angeles, such truck climbing lanes are physically separate from the main roadway, taking a longer route to permit a somewhat less-steep grade.

One of the best-known examples of separated lanes is on a 45-mile section of the New Jersey Turnpike. For this "dual/dual" section, the Turnpike consists of four parallel roadways, each consisting of three 12-foot lanes. The inner roadways are designated cars-only, while the outer lanes are usable by cars and trucks. The Turnpike is heavily used by trucks, which account for about 12% of average daily traffic and about 34% of revenue. In 2008, the state proposed a USD2 billion project to extend the "dual-dual" configuration an additional 25 miles, including reconfiguration of seven interchanges.[4]

2.3. HOV and HOT lanes

The most common type of specialized lane in current US highway practice is the high-occupancy vehicle (HOV) lane, aimed at promoting carpooling. These lanes began to be added to urban freeways in the 1960s, originally as exclusive busways. The first one was added to the Shirley Highway (I-395) in northern Virginia, a commuter route to the Pentagon and Washington DC. However, although bus service on the new (reversible) lanes was popular, there was considerable unused capacity. Hence, in December 1973, vanpools and four-person carpool vehicles (HOV-4) were allowed to begin using the busway. After more than a decade of use under this policy, there was still unused capacity, while adjacent GP lanes had become highly congested during peak periods. So, in 1989, the minimum occupancy requirement was reduced to HOV-3.[5]

A similar evolution took place in Houston, where "transitways" were added to several key freeways starting in 1979. Initially, they were single-lane, reversible busways, but by the mid-1980s the existence of unused capacity led to opening these lanes first to vanpools, then HOV-4 and soon after, HOV-3 in 1985, and HOV-2 in 1986. In most other urban areas, carpool lanes became the freeway capacity addition of choice during the 1980s and 1990s, and nearly all such lane additions were designated for HOV-2 operations, where nearly all remain today.[6]

Because all but a handful of HOV lane projects (as they are now called) offer only a single lane in each direction, their performance in relieving traffic congestion has been criticized. On one hand, some studies suggest that most HOV lanes reduce overall freeway capacity compared with that additional lane being a GP lane, since most move fewer vehicles per lane per hour than the adjacent GP lanes and their single-lane configuration limits their speed to that of the slowest vehicles using them.[7] On the other hand, a few HOV lanes attract so much peak-period traffic that they become congested during peak periods and hence lose their intended time-saving advantage for carpoolers and buses.

Both phenomena—unused capacity and excessive use—have been cited as reasons to convert HOV lanes to HOT (high-occupancy toll) lanes. In the case of unused capacity, the rationale is to open up the HOV lane to those willing to pay a market-price toll in order to save time. In the case of HOV lanes that have been overcrowded, the rationale is that an increase in the occupancy requirement (generally to HOV-3) will create significant unused capacity, which can then be sold. Since 1993, when the original paper urging HOV to HOT conversions was published[8], such conversions have taken place for individual HOV lane facilities in Denver, Houston, Miami, Minneapolis, Salt Lake City, San Diego and Seattle.

A somewhat different case has more recently been made for adding a version of HOT lanes to congested freeways that do not already have HOV lanes. The prototype for this is the 91 Express Lanes project in Orange County, California. Space had been reserved in the median of this congested freeway for HOV lanes, but in the 1990s neither the state nor the county had funds available to build them. A private-sector proposal to finance, build and operate the lanes was put foward as express toll lanes was accepted by the state transport agency (Caltrans) with the proviso that discounts be offered to carpools of three or more people (HOV-3), and the project was financed and built on that basis.[9] Subsequently, private-sector proposals to add express toll or HOT lanes have been accepted in northern Virginia (I-495), Florida (I-595) and Texas (with the I-635 in Dallas and I- 820/SR 183 in Fort Worth). All of the private-sector projects thus far, like the original 91 Express Lanes, are two or more lanes in the peak direction, rather than single-lane facilities.

As of 2009, the US transportation community has generally accepted the term “managed lanes” to refer to all types of specialized (non-GP) lanes, though nearly all the literature using this term refers to lanes using some kind of pricing.

2.4. Truck only toll lanes and roads

This relatively new idea first arose in the 1990s. In 1995, under a Minnesota transportation public-private partnership law, a firm called Transportation Industries International proposed a privately financed (USD1.3 billion in 1996 dollars) trucks-only highway, mostly along the right of way of SR 2, from Winnipeg (in Saskatchewan, Canada) to Duluth, Minnesota.[10] To be built with heavy-duty pavement aimed at handling heavier trucks than those permitted on ordinary Interstate highways, it was intended to compete with freight railroads in carrying grain and lumber from Canada to the Great Lakes shipping port at Duluth and to Mississippi barge lines near St. Paul, Minnesota. Potential later extensions would have extended this “truckway” southeast to Chicago and points further east. The project was one of five submitted by private firms, all of which were ultimately rejected as either lacking sufficient local support or failing various benchmarks set by Minnesota DOT for financial and technical feasibility.

In the late 1990s, the Pennsylvania Turnpike—a very truck-intensive roadway—considered adopting the “dual/dual” configuration noted above on the New Jersey Turnpike. According to an interview with the Pennsylvania Turnpike’s research manager at the time, the idea was being considered for both safety and cost reasons. The former was to reduce the likelihood of car/truck accidents and the latter was based on the much higher pavement wear caused by heavy trucks. Since the Turnpike’s lanes were to be reconstructed, those designated as truck-only lanes could be built to handle even heavier loads than before, while those no longer serving heavy trucks could be rebuilt to lower-cost standards and would have much lower life-cycle cost.[11]

Those thoughts helped to generate the concept of Toll Truckways, introduced by the Reason Foundation in 2002. In 2000, the US Department of Transportation released a major truck size and weight study.[12] That report highlighted the potential productivity gains that could be realized if longer and heavier truck configurations (referred to generically as Longer Combination Vehicles—LCVs) could operate nationwide on limited-access roadways. However, the cost of upgrading that entire system to thicker pavements and stronger bridges was seen as a significant obstacle to bringing that about, as were unresolved concerns about the safety of automobiles on portions of the national network where traffic is far denser than in the mountainous western states where LCVs may legally operate in GP lanes on selected highways.

The 2002 Reason study proposed, instead, the addition of truck-only toll lanes to those Interstate highway routes that function as major truck corridors. The new lanes would be designed specifically for LCV-category trucks, would have separate on-ramps and off-ramps and would be separated from GP lanes by concrete barriers. They would charge tolls (electronically) to recover the cost of building and maintaining the lanes. LCVs would be allowed to operate in states from which they are currently banned, but only on the toll truckway lanes. Other trucks would have the option of using the truckways, if paying the toll offered enough value in terms of higher average speed, increased safety or other factors.[13]

The study modelled truck operations on a hypothetical Interstate highway corridor, testing a large number of scenarios assuming various fractions of truck traffic (including those newly induced to shift to LCV rigs) opting to use the truckway, and estimating the productivity gains from using the truckway. Those gains were quantified, using trucking industry data, and used to estimate possible toll rates for using the truckway. The analysis concluded that under a variety of scenarios, such truckways could break even or be revenue positive, though not necessarily at commercial rates of return on investment. Also quantified were savings in operation and maintenance costs to state DOTs from reduced wear and tear on the GP lanes, depending on the fraction of truck traffic shifted to the truckway lanes.

In 2007, the US Department of Transportation made grant funding available, on a competitive basis, under a new program called “Corridors of the Future”. One of the winning proposals was from a set of four state DOTs along 800 miles of the I-70 corridor, a major truck route from Kansas City on the west to Columbus, Ohio on the east. Their proposal was for a detailed feasibility study of adding LCV-capable truck-only lanes to I-70, as a possible alternative to doing the needed widening of that Interstate by adding GP lanes. The final environmental impact statement, completed in June 2009, selected the “Truck-Only Lanes Strategy” as the preferred alternative, compared with the “Widen Existing I-70 Strategy.”[14] And in 2008, the Montana DOT undertook a feasibility study on widening I-80 across that state, with toll truck lanes as one of the alternatives.

3. ARGUMENTS FOR CARS-ONLY LANES

3.1. Rethinking traditional design standards

What leads to the extremely high costs of urban expressways? Ng and Small, in a provocative 2008 paper, suggest that the US design standards that evolved in the 1950s for urban freeways lead to needlessly high cost per lane-mile.[15] The basic reference for these standards is the AASHTO Design Standards—Interstate System, produced by the American Association of State Highway and Transportation Officials (and most recently revised in 2005). Expressway design standards are based on two underlying assumptions. The first is that urban expressways must be designed for safe travel at high speeds. Second, they must be able to carry mixed traffic, including large trucks. However, if urban expressways are congested for much of the day, so that only a small fraction of their daily traffic can operate at high speed, Ng and Small ask if we should still design them to standards based on those high speeds. Furthermore, should all such expressways be designed to accommodate large trucks?

Ng and Small then explore the trade-offs involved in narrower lane and shoulder widths (which require lower design speeds). Specifically, they compare a 40-right of way, which would normally provide two 12-foot lanes and shoulders of six and ten feet, with an alternative configuration consisting of three 10-foot lanes plus shoulders of two and eight feet. Both configurations would have essentially the same construction cost, but the "narrow" configuration would have significantly more capacity, despite its lower design speed, under real-world conditions of serious congestion during long peak periods. Ng and Small present graphs showing travel times on regular versus narrow expressways for various levels of average daily traffic, illustrating a fairly wide range of traffic conditions under which the narrow expressway performs better, due to having greater capacity (but at the same construction cost as the regular expressway). They make a similar comparison between a regular urban arterial (with two 12-foot lanes) and a "narrow" arterial (with three 10-foot lanes)—both within the same 38-foot right of way. Their performance findings are similar to those for expressways. Ng and Small do not recommend that large trucks be allowed to operate on their proposed "narrow" expressways and arterials. These new types of roadways would be for light vehicles only.

3.2. Making use of unconventional rights of way

Another approach to adding needed highway capacity in urban areas is to seek out rights of way that were created for another purpose and use them for specialized roadways. If the mental model is a conventional expressway, these rights of way will generally be rejected as too narrow. Peter Samuel has suggested three such right of way categories:

– Underused railroads;
– Drainage channels;
– Power line corridors[16].

These days, underused or disused railroad rights of way in US urban areas are reflexively thought of only as corridors for commuter-rail or light-rail service. Yet those corridors may or may not be well-located for that purpose. An alternative is to use the corridor for a combination busway and HOT lane, providing both transit improvements and a higher-speed alternative for motorists. Railroad rights of way are typically 50 to 100-feet wide, enough to provide from four to eight "narrow" 10-foot lanes for light vehicles and buses. Samuel gives an example of a disused rail line in Los Angeles that would provide a shorter (ten mile) route from Los Angeles International Airport to downtown than the current nearly 15-mile freeway route. He also cites examples of two Texas urban toll roads, Houston's Westpark Tollway and the Dallas North Tollway, both built on former railroad line right of way. Another possible use for disused rail lines is urban truckways. Samuel cites possibilities in both Chicago and Brooklyn, New York, in which congestion caused by numerous trucks on regular city streets could be significantly relieved by converting little-used rail right of way to urban truckways[17].

Drainage channels in metro areas with arid climates could be the location of parkway-type roads sized for light vehicles (and possibly buses). One such project is in the planning stages along the flood plain of the Trinity River in Dallas, Texas. Others have been proposed for concrete-lined flood control channels of the Los Angeles River in Los Angeles and the Santa Ana River in Orange County, California. (One such roadway, a portion of Burbank Blvd., exists in the Sepulveda Dam Recreation Area of Los Angeles). Such roadways require access control so that they could be closed to traffic on those rare occasions when rainstorms would make them unusable as roadways due to the possibility of flash floods.

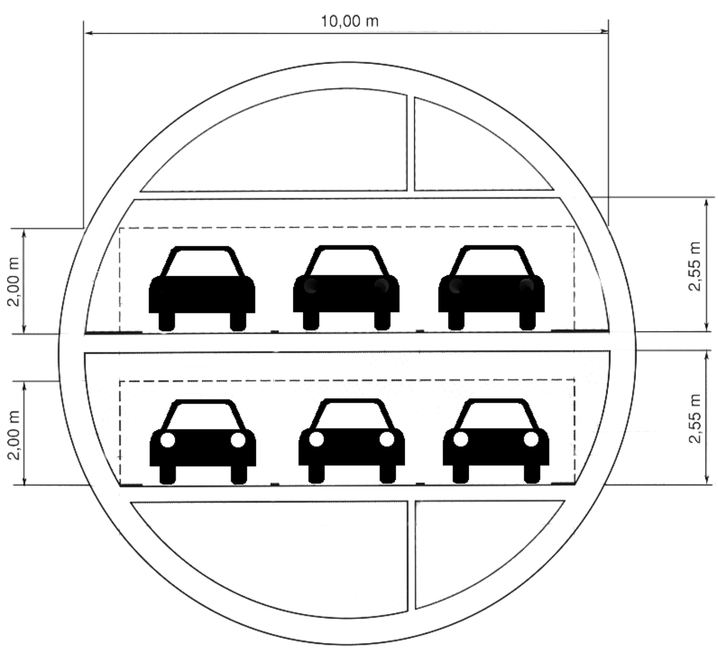
Power line corridors are sometimes wide enough for conventional expressways, but when limited to 50 to 100-feet, they would be better suited to specialized roadways, either for light vehicles only or for truckways. Samuel points to an example from the Maryland suburbs near the District of Columbia, in which a wide power line reservation was proposed as right of way to extend the I-95 expressway inside the Capitol Beltway, providing a new radial route to the nation's capital; that route would have extended about five miles, followed by a one-mile tunnel to permit it to connect with the existing I-395 near the Capitol Building. That project was defeated by local anti-highway opposition.

3.3. Retrofitting urban expressways

Besides having narrower lanes, expressways designed for light vehicles rather than heavy trucks need lower overhead clearance requirements. That opens up significant possibilities for adding capacity at less cost than conventional approaches.

An excellent European example is the missing link on the A86 Paris ring road. After several decades of opposition to a surface motorway through the Versailles area, toll road company Cofiroute made an unsolicited proposal to complete that 6.2 mile link as a deep-bore tunnel, financed entirely by congestion-priced toll revenues. Given this revenue constraint, Cofiroute needed to come up with an affordable design. By limiting the tunnel to light vehicles only, it was able to fit six 10-foot lanes into a double-deck configuration with an inside diameter of 34-feet. (Initial operations will use two lanes in each direction, with the third lane reserved as a breakdown lane). This basic concept appeared in Gerondeau's 1997 book[18], and is illustrated in Figure 1. Actually, the origins of the idea date back to at least 1988, when a private-sector proposal called for a network of toll-financed underground cars-only roadways in Paris named LASER.[19]

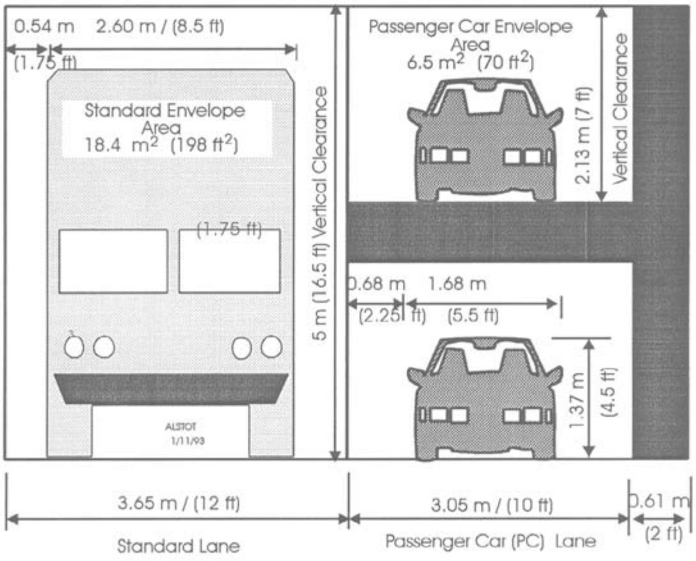
Figure 1. **Metroroute cross-section**



Source: Gerondeau, Note 18.

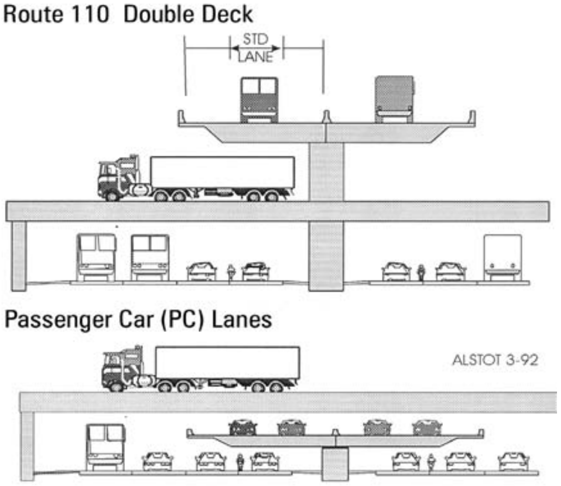
Reduced vertical clearance would also permit the addition of significant amounts of capacity to existing urban expressways without the need to acquire additional right of way. Figure 2 shows standard US roadway dimensions, illustrating that two lanes for light vehicles can be stacked, with ample vertical clearance, within the standard clearance height required for GP lanes able to accommodate large trucks. This provides an alternative to conventional double-decking approaches, such as that used to add an elevated busway/HOV lane on I-110 in Los Angeles.

Figure 2. **Standard US truck and car clearance dimensions**



Source: Gary Alstot, presentation to the American Society of Civil Engineers (March 1992).

Figure 3. **Double deck lanes *vs.* passenger car lanes**

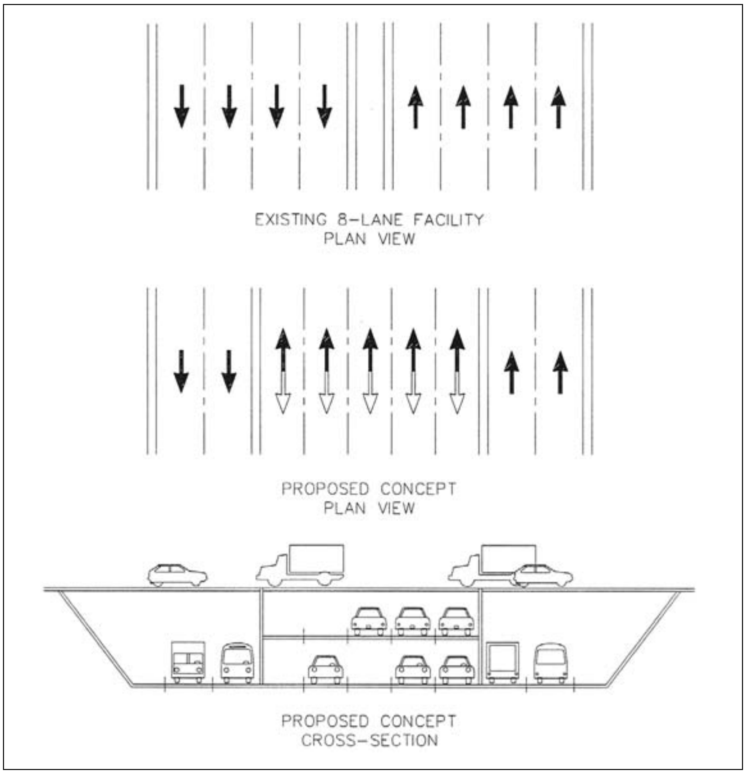


Source: Gary Alstot, presentation to ASCE (March 1992).

Figure 3 shows that if an auto-only second deck were acceptable, it could be built within the existing clearance height of the freeway.

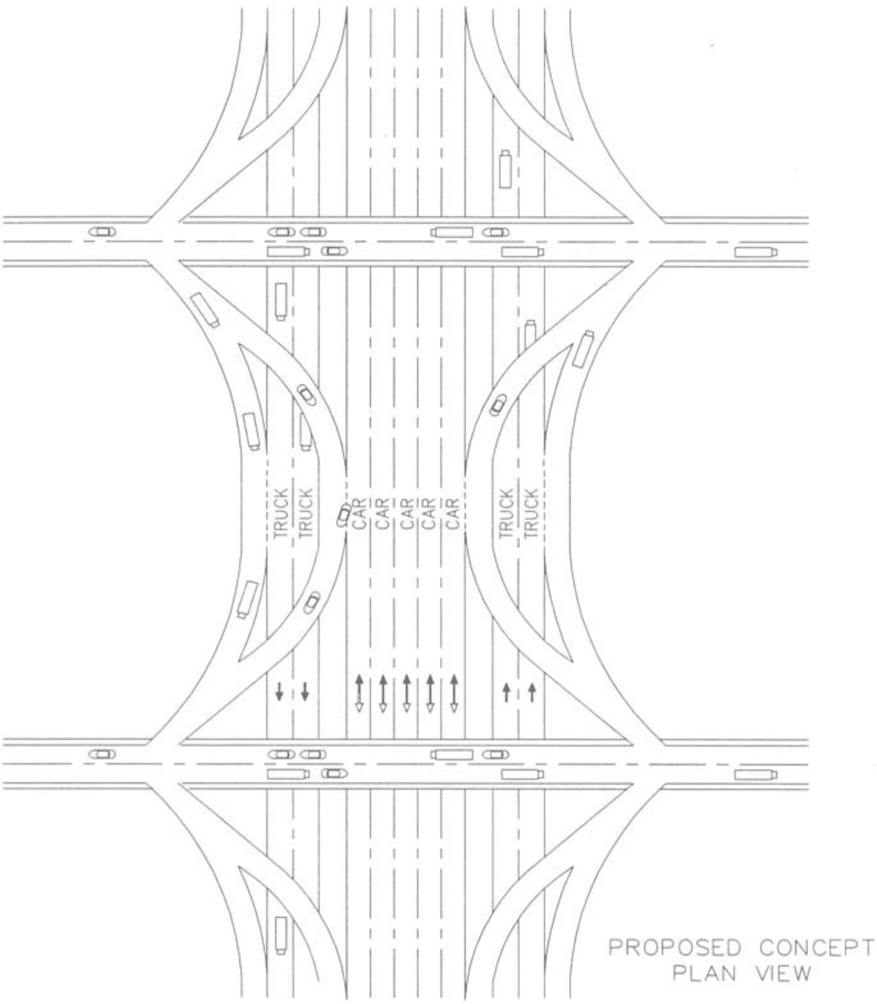
Civil engineer Joel Marcuson has taken these ideas further, envisioning how an eight-lane urban expressway could be reconfigured with cars-only lanes in its centre section, as shown in Figures 4 and 5. While these reconfigurations would be costly, they provide an alternative to the generally "politically impossible" prospect of condemning expensive urban land to add capacity by widening the expressway right of way.

Figure 4. **Reconfigured freeway using cars-only lanes**



Source: Joel Marcuson, Sverdrup, July 1995.

Figure 5. **Ramps to serve reconfigured freeway**



Source: Joel Marcuson, Sverdrup (July 1995).

3.4. Buses plus light vehicles

The A86 tunnel and the reduced clearance height designs shown previously all presume that only auto-size vehicles (cars, passenger minivans and pickup trucks) are allowed to use these non-GP lanes. There is a trade-off involved if these roadways are designed to also accommodate city buses. Clearance heights would have to be greater in this case, but not as high as needed for large trucks (for which the US standard is 16.5-feet). Moreover, ten-foot lanes might require the use of station-keeping technology for buses using these lanes.

US transit buses are typically ten feet and eight inches high, meaning a clearance height of 12 feet, rather than the seven feet shown in Figure 2. That would preclude the kind of double-decking shown in Figures 3 through 5. For tunnels, that clearance height would require a somewhat larger diameter. In both cases, the addition of buses to the vehicle mix would increase the costs, and this suggests that some applications might be limited strictly to light vehicles rather than including buses.

The feasibility of buses operating on narrow lanes has been demonstrated repeatedly on a small scale since the 1980s, via “curb-guided bus” technology. Under this approach, conventional city buses are equipped with small guide wheels that roll along an adjacent curb-side. A 2006 article cited 11 such systems in operation as of that time, with three more in the planning stages — in Australia, Germany, Japan and the United Kingdom.[20]

4. ARGUMENTS FOR TOLL TRUCK HIGHWAYS

4.1. Productivity gains

The primary rationale for toll truck lanes is productivity gains, due to being able to haul more freight payload per unit of fuel and driver cost. In the 2002 Reason study, the productivity analysis compared a hypothetical toll truckway permitting higher axle loads (weight per axle) than in either of two base cases, corresponding to current weight limits in various subsets of US states.[21] These cases were analyzed for two truck configurations: the common tractor/single-trailer rig with 18 wheels and the long double rig, with a tractor plus two long trailers and 34 total wheels. The toll truckway would produce the largest gains in states with lower axle-load limits and a maximum gross vehicle weight of 80 000 lbs., but there would also be significant gains in states that have more liberal limits. The higher-capacity trucks were found to be more economical for trips longer than about 50 miles.

The preceding analysis focused solely on productivity gains due to greater payload, and involved relatively long-haul (inter-city) corridors. A more recent toll truckway analysis looked into a high-capacity toll truckway to serve trucks in shorter-haul drayage service between the ports of Los Angeles and Long Beach and a large region of warehouses and distribution centres about 55 to 70 miles distant.[22] In this case, the source of productivity increases is two-fold. Current drayage rigs consist of a single 40-foot (or sometimes 53-foot) container on a chassis, hauled by a tractor. The proposed toll truckway would permit the operation of dual-container rigs, thereby doubling the payload of each drayage trip. However, in addition, since the existing freeways are heavily congested much of the day, a separate toll truckway would permit significantly higher speed, allowing for a larger number of “turns” per shift per driver.

Table 2 is reproduced from the 2005 report, and is based on freight rates and operation costs as of 2004. As can be seen, the estimated revenue gains from the combination of increased payload per trip and increased speed far outweighs the increase in costs of operating the larger and heavier rigs (in which “double-short” refers to two 20-foot containers, “triple-short” means three 20-foot containers, and “double-long” refers to two 40-foot containers). The bottom line of this analysis is that shippers would benefit from lower freight rates, truckers would gain additional revenue for overhead and profit and the toll truckway operator would be able to charge quite high tolls, ranging from USD0.61 to USD 1.83 per mile.

Table 2. **Urban toll truckway productivity**

	Col 2	**Col 3**	**Col 4**	**Col 5**	**Col 6**	**Col 7**
	Mixed freeway semi-trailer	**Mixed freeway double-short**	**Truckway semi-trailer**	**Truckway double-short**	**Truckway triple-short**	**Truckway double-long**
Payload	45 000 lbs	45 000 lbs	45 000 lbs	45 000 lbs	67 500 lbs	90 000 lbs
Metric tons	20t	20t	20t	20t	30t	40t
100 mile delivery – 2004 freight rates	$500	$500	$500	$500	$750	$1 000
Average speed on the road (mph)	38	38	60	60	60	60
Miles driven in 8-hr shift (6 hrs driving)	228	228	360	360	360	360
Revenue from 6 hrs payload at 2004 rates	$1 140	$1 140	$1 800	$1 800	$2 700	$3 600
Variable costs per shift	$684	$684	$684	$684	$1 007	$1 165
Available for overhead, profits, tolls	$456	$456	$1 116	$1 116	$1 693	$2 435
Extra earnings from using truckway/shift/ day			$660	$660	$1 237	$1 979
Assume the extra productivity split 3 ways			1/3 = $220	1/3 = $220	1/3 = $412	1/3 = $660
Shipper's savings on 100 mile delivery, $ & %			$61 12.2%	$61 12.2%	$76 15.2%	$91 18.3%
Additional for trucker for overhead & profit/day			$220 43%	$220 43%	$412 90%	$660 145%
Truck tollway – possible toll per mile			$0.61	$0.61	$1.15	$1.83

Source: Samuel, Note 22.

4.2. Operating and maintenance cost savings

Highway cost allocation studies have quantified the damage that heavy trucks do to pavements not specifically designed for such loads. Such damage is proportional to the third power of weight, so as the researchers Small, Winston and Evans note, “for all practical purposes, structural damage to roads is caused by trucks and buses, not by cars.”[23] Thus, to the extent that heavy truck traffic can be shifted from GP lanes to specialized truck-only lanes, highway owner-operators have the potential for considerable savings in operating and maintenance costs.

In the 2002 Reason toll truckways study, the authors made a rough estimate of these savings. They used the World Bank’s Highway Design and Maintenance model, which relates road usage to maintenance needs. In the case that was analyzed, only trucks of longer length and heavier weight than are currently allowed in a state would be required to use the new truck toll lanes; all other trucks could opt to use them if they judged the benefits (e.g. time savings, increased safety and better pavement condition) to be greater than the toll charged. The model calculated the GP lanes’ pavement conditions each year over a 50-year period, estimated maintenance and repaving needs and produced annual operations and maintenance costs for a range of truck-shift assumptions (ranging from 25% of total corridor truck traffic using the truck lanes to 100%). For the 100% case (which would apply if the law required all trucks to use the truck lanes), the annualized operation and maintenance cost savings on the GP lanes equaled 80% of the fuel tax revenue that would have been paid by the trucks had they remained on the highway’s GP lanes. (In this example, it was assumed that trucks using the new truck lanes would pay tolls *instead of* current fuel taxes). While at first glance this might appear to be a losing proposition to the highway owner, one must also take into account the avoided cost of adding a lane to the highway—i.e. the new lane in each direction would be paid for by the toll revenues, rather than by means of fuel taxes. Once that is taken into account, the highway owner comes out substantially ahead.[24]

5. HETEROGENEOUS VALUES OF TIME

5.1. Motorists’ values of time and reliability

Most transportation studies use a single value of time for motorists (or occasionally two different values, one for business travel, including commuting, and one for leisure/personal travel). Increasingly, however, researchers are finding that values of time vary greatly, depending on factors such as individual preferences, trip purpose, time of day and week, etc.

The complexity of commuters’ value of time has been studied in some detail in recent years in the United States, in connection with the introduction and use of HOT lanes and express toll lanes, where the price charged varies in proportion to demand. The variably priced facility that has been in operation the longest is the 91 Express Lanes, on SR 91, a congested freeway linking the bedroom communities of inland Riverside County (Calif.) with the employment centres in coastal Orange County. Small, Winston and Yan studied traveller behaviour in that corridor in some detail, and summarized their findings as follows: “We find that the users of SR 91 have high average values of travel time and travel-time reliability, and that the distributions of these values exhibit considerable dispersion[25].”

To illustrate the extent of heterogeneity in their sample of SR 91 corridor commuters, they found the median value of time (VOT) of Express Lane users to be USD 25.51, compared with USD 18.63 for the GP lane users. But the range of those values was very large: from a 5th percentile of USD 11.50

to a 95th percentile of USD 39.99/hour for Express Lanes users, and from USD 7.76 to USD 29.08/hour for GP lane users. And those were just the value of time figures. Also measured was the value of reliability (VOR), with median values of USD 23.78 for Express Lane users and USD 19.50 for GP lane users—and with even greater variability than shown for value of time. Moreover, their database is drawn from the A.M. peak period, whose toll levels (and hence presumably VOT and VOR) are considerably lower than those in the P.M. peak. As Small *et al*. sum up, motorists in this corridor "exhibit a wide range of preferences for speedy and reliable travel, as total heterogeneity in VOT and VOR is nearly equal to, or greater than, the corresponding median value. On average, express-lane users have higher values of travel time and reliability than do users of the [GP] lanes, as expected, but wide and overlapping ranges exist within these two groups, resulting from strong heterogeneity in preferences."

Small *et al.* use these findings to critique standard arguments for freeway congestion pricing, which would generally impose a uniform charge for all users of all lanes during peak periods, with lower or zero charges at other times of day. Using a demand model, they estimate the social welfare implications of policies such as HOV or HOT lanes alongside GP lanes, tolling all lanes, or charging different rates on premium and GP lanes. They conclude that some version of the latter (which they call a "two-route HOT" policy) is a reasonable compromise, providing some degree of peak-spreading and time-savings for all lanes on the expressway, but without greatly over-charging the majority whose VOT and VOR are lower than what needs to be charged to keep premium lanes uncongested during peak periods.

Douglass Lee has generally been critical of separate lanes such as HOV and HOT on the familiar grounds discussed in this paper – that overall capacity is less with multiclass lanes than with all GP lanes – while conceding that HOT lanes are generally an improvement over HOV lanes, since the former are more likely to operate at high throughput while avoiding hypercongestion.[26] In response to Small *et al.,* he argues that "the only way HOT lane[s] could be superior [to an all-GP lanes roadway] would be to charge prices on both lane classes, at least enough to keep both lane [types] at full capacity, but not identical flows." Lee also concludes that "the justification for more than one class of service requires that the preferences (value of travel time, or VOT) among users be very heterogeneous." While we thus far do not have detailed data on peak-period commuters' VOT and VOR from many urban areas, the detailed data from the SR 91 corridor at least suggests that such commuters have VOT and VOR far more heterogeneous than has traditionally been assumed.

5.2. VOT and VOR in urban trucking

Many studies of goods-movement use a single value of time, generally based on an assumed average value of time saved (*e.g.* by using a toll road), not explicitly taking into account the value of reliability. This unsophisticated approach is beginning to change, however, as further research is done. A study carried out by the American Transportation Research Institute and the Federal Highway Administration in 2005 measured travel times and delays in five Interstate highway corridors and used the data to derive both a travel time index (TTI) and a Buffer Index (BI). The former compares actual travel time with free-flow travel time, while the latter is a measure of travel-time variability.[27] That report also noted that "shippers and carriers value transit time at USD 25 to USD 200 per hour, depending on the product being carried. Unexpected delays can increase that value by 50 to 250 per cent."

A truck toll lane facility must be analyzed based on the types of goods movement most likely to be carried out on that facility. In the case of the proposed toll truck lanes in the Los Angeles region, as noted previously, their principal purpose would be the drayage of containers between the ports and the distribution centres and warehouses mostly located about 60 miles inland. In 2007, the Southern California Association of Governments (SCAG) prepared an analysis of that market, estimating both VOT and VOR for container drayage in that corridor.[28] The analysis estimated year-2030 values for both travel time index and buffer time index for the principal freeway routes that would be used if the toll truck lanes are not built. The combined VOT and VOR during peak periods was estimated at USD 73/hour for heavy drayage trucking.

Based on SCAG's travel demand models, truck speeds on the truck lanes were estimated to be up to three times as fast as would otherwise be the case in mixed-flow traffic on the freeway GP lanes. For three different destinations to/from the ports, the study produced the data shown in Table 3.

Table 3. **Los Angeles truck toll lane data, 2030 a.m. peak**

District		**Min. Saved**	**Hours**	**Value @ $73/hr**	**Toll Cost @ $.86/mi**	**Net Savings**	**Savings/ Toll**
Downtown	To	85	1.42	$103	$17	$86	5.1
	From	97	1.62	$118	$17	$101	5.9
Ontario	To	192	3.2	$233	$32	$201	6.3
	From	298	4.97	$361	$32	$329	10.3
Victorville	To	285	4.75	$345	$64	$281	4.4
	From	405	6.75	$490	$64	$426	6.7

Source: Killough, Note 28.

The numbers in Table 3 do not take into account either (a) additional productivity from an increased number of trips per driver per shift due to these time savings, or (b) higher value thanks to increased productivity from being able to haul multi-container rigs. As a point of comparison, SCAG estimates the construction cost of the truck toll lanes at USD 20 billion and the total project cost (including environmental mitigation) at over USD 30 billion. While the study did not estimate whether this mega-project could be financed solely based on toll revenues, the assumed USD 0.86/mile toll can be seen as far below what might be able to be charged, given the increased productivity gains of which the value is not included in the analysis summarized here.

6. SAFETY CONSIDERATIONS IN SEPARATION OF CARS FROM TRUCKS

One of the key issues that must be addressed in any consideration of separate lanes for cars and trucks (or, more accurately, light vehicles *vs.* heavy vehicles) is safety. We look first at empirical data regarding "narrow" roadway designs and then specifically at car-truck accidents. In addition, we consider trends that are likely to mean smaller automobiles in coming decades.

6.1. Safety data re "narrow" designs

In their paper making the case for "narrow" designs of expressways and arterials, Ng and Small provide an overview of recent research on the safety record of roadways with narrower lanes than current US AASHTO standards. The studies they examine focus on accidents involving injuries and fatalities on urban arterials and on expressways of four or more lanes.

They reviewed a number of studies, both before/after (e.g. narrowing the lane widths on certain freeways) and cross-sectional (comparing accident rates on narrow and conventional roadways in a given state). Their conclusion is as follows:

> *"[B]oth theoretical and empirical evidence linking road design to safety are ambiguous, although on balance they contain some indications that greater lane width and shoulder width may increase safety. Thus, we think it is an open question whether the 'narrow' road designs considered here would in fact reduce safety, but it is certainly a potential concern. Probably it would depend on factors that vary from case to case, especially the speeds chosen by drivers[29]."*

They go on to discuss design features that should accompany "narrow" designs, such as lower speed limits. They note the successful use in Germany and the Netherlands of variable speed limits, variable message signs, temporary shoulder use, and other techniques. Studies that use driving simulators and traffic simulation models, they report, find that speed limitation reduces average speed, speed variation, and lane-changing movements, all of which reduce accident rates. The US freeway operations community is currently exploring a number of these concepts under the rubric of "active traffic management."[30] Thus, active traffic management techniques offer an important complement to "narrow" roadway designs, to enhance their safety.

6.2. Car-truck accidents

Another factor in making "narrow" designs safer, as Ng and Small point out, is to limit such designs to light vehicles only – thereby avoiding car-truck accidents. They cite a study of the "dual/dual" sections on the New Jersey Turnpike which found that accidents are higher in the mixed-traffic lanes than in the autos-only lanes (which are otherwise identical in configuration) and that trucks are disproportionately involved in the accidents in the mixed-traffic lanes[31]. They cite another paper that uses an econometric model to conclude that overall accident rates are nearly four times as responsive to the amount of truck travel as the amount of car travel[32].

In the United States, about 4 800 large trucks are involved in fatal accidents per year (resulting in about 5 000 fatalities), and about 140 000 are involved in non-fatal crashes (resulting in about 90 000 injuries), according to the Federal Motor Carrier Safety Administration. FMCSA's Large Truck Crash Causation Study involved a sample of 963 large-truck crashes (involving 1 123 trucks and 959 other vehicles) during 2002-2003.[33] Of the total, 73% of the crashes involved a large truck colliding with at least one other vehicle; 50% of the total sample involved car-truck crashes. For this subset of crashes, the causation study assigned the "critical reason" for the crash to the truck in 44% of the cases, meaning that in 56% of them, the car was the critical reason for the crash. For truck-initiated crashes, the two most likely factors were brake problems and drivers either travelling too fast or being unfamiliar with the roadway. For passenger vehicle-initiated crashes, the most important factors were interruption of the traffic flow and unfamiliarity with the roadway. Interestingly, comparing these "associated factors" between truck-initiated and car-initiated crashes, several factors stood out in the car driver but not truck driver data: alcohol and drug use, fatigue and illness.

Since nearly half the car-truck crashes appear to be the "fault" of the truck, separation of car traffic from truck traffic would appear to have significant potential for reducing the deaths and injuries due to car-truck crashes.

6.3. Downsizing of automobiles

One other factor relating to car-truck accidents is the likely downsizing of automobiles in response to concerns over energy use and greenhouse gas emissions. In the United States, the Obama administration in Spring 2009 announced new federal Corporate Average Fuel Economy (CAFE) regulations for both cars and light trucks. The new requirement calls for new autos produced in 2016 to average 39 miles per gallon (compared with 27.5 today) and light trucks 30 mpg (*vs*. 22.5 today). Meeting those requirements is widely expected to require downsizing of new vehicles by 2016.

There is a definite correlation between vehicle size/weight and the seriousness of crashes, as measured by deaths and injuries. A 2002 National Research Council study on the impact of CAFE standards found that the vehicle downsizing that occurred in the 1970s and early 1980s due to the original CAFE standards appeared to have led to between 1 300 and 2 600 additional crash deaths in 1993. In recommending further increases in fuel economy of new vehicles, the NRC authors noted that there were alternative ways that fuel economy could be increased by vehicle manufacturers, and that even a scenario that involved further downsizing would likely involve considerably lower additional crash deaths than in the 1980s, due to the significant increase in safety features built into new vehicles in the intervening years. It concludes by saying "if an increase in fuel economy is effected by a system that encourages either down-weighting or the production and sale of more small cars, some additional traffic fatalities would be expected[34]."

Given the likely further downsizing of both cars and light trucks, the impact of crashes involving those vehicles and heavy trucks will almost certainly be more severe than has been the case historically. This provides a further reason for considering future roadway models that include facilities for light vehicles only.

7. ENVIRONMENTAL ISSUES

Some have argued against the provision of truck-only lanes as the wrong course to follow, on environmental grounds. One aspect of this argument is that since heavy trucks are largely powered by diesel engines, which are considered serious polluters in the United States, government policy should not be facilitating the expansion of goods movement by truck. On a larger scale, this argument calls for policy that aims to shift goods movement as much as possible from road to rail. While somewhat beyond the scope of this paper, these points cannot be ignored.

7.1. Greener trucks

Large-scale transportation infrastructure projects take a decade or more from initial studies to entry into service. Consequently, what is relevant in considering future truck-only toll projects is the truck fleet that will likely exist several decades from now (over the expected service life of the truckway), not the truck fleet of the past several decades. In the United States, new low-sulfur diesel fuel standards came into effect in 2006, to facilitate the requirement that all trucks sold after January 1, 2007 use of new low-emission diesel truck engines. A study by the American Transportation Research Institute, presented at the 2006 Transportation Research Board meeting, projected that by 2015, the US diesel truck fleet would produce 63% less particulate emissions and 53% less nitrogen oxides than the 2005 fleet.[35]

A second factor to consider is the positive impact of increased trucking productivity on truck emissions. A truck tractor hauling two long trailers hauls 100% more payload while using only about 60% more fuel. Thus, the emissions-intensity of goods movement is reduced considerably to the extent that the trucking industry adopts the more-productive longer combination vehicles. This point was confirmed by Cheryl Bynum of the US Environmental Protection Agency's SmartWay Transport Team in 2004. In response to a query by this author, she wrote:

> "*If a [trucking] fleet uses longer trailers and/or multiple trailers, total ton-miles are improved for that trip, and there are fewer trips. This also provides—in addition to the fuel and GHG savings—criteria pollutant savings. The actual environmental benefits depend upon the input the fleet enters into the FLEET Performance model, since it is specific to mileage, equipment type, mpg, and payload*".[36]

The EPA's FLEET model quantifies the fuel savings and emission reductions from various trucking company strategies.

7.2. Roads *vs.* rail

Recent years have seen a number of studies comparing the socio-economic costs of goods movement by rail and by truck. For example, a series of U.K. studies by the Commission for Integrated Transport found that the rail's environmental impacts were only about one-fourth those of road transport. Those impacts were then monetized and included in overall socio-economic benefit/cost analyses. While the rail projects tended to have Benefit/Cost ratios of less than 3:1, most of the goods-movement highway investment projects had BC ratios of 10:1 or higher.[37]

The reasons for this disparity stem from supply chain performance differences, which are the main driver of mode choice for shippers and receivers. Transit time, reliability and availability on-demand are what the market demands for most goods other than bulk commodities (for which rail has an overwhelming advantage). These points were quantified in an assessment of road vs. rail trade-offs for a study in South Africa. A possible truck-only toll road was compared with an expanded rail line over a 600 km route between Johannesburg and Durban. While both the fuel costs and CO_2 costs for the rail alternative were less than one-quarter those of the tollway, the additional (quantified) supply chain costs made rail nearly 50% more costly (and a large disparity would still exist even at double the current oil price). The bottom line was that a R30 billion investment could produce 72 million tons of economic capacity via the road alternative, but only 24 million tons of capacity (that would operate at a loss) in the rail alternative.[38]

France's Institut National de Recherche sur les Transports et Leur Securite (INRETS) is researching the most promising techniques for future goods movement in France and Europe in the 2030 time frame. According to a presentation given as part of a 2008 study tour in the United States, among the ideas they are exploring are automated trucking and "dedicated truck or goods train toll lanes."[39]

Automation and truck-only toll lanes have been studied by researchers at the PATH project at the University of California, Berkeley and at San Jose State University. Tsao and Botha of San Jose State have made a detailed proposal for dedicated, heavy-duty truck lanes equipped with a variety of high-tech aids to reduce driver workload and increase safety. An evolutionary path is aimed at bringing about what they call Segregated Electronic Road Trains (SERTs)—essentially a platoon system for trucks.[40] This could permit dramatically increased vehicle throughput, reducing the number of lanes required.

A somewhat similar proposal for an urban truck lane project was proposed for Chicago by researchers from PATH at UC Berkeley. Shladover *et al*. proposed a similar evolutionary path, initially building a two-lane (one lane per direction) urban truckway, of which the BC ratio was estimated at 3.6, based on truck travel time reductions and roadway congestion-relief benefits. When demand increases to the point where more capacity is needed, they propose adding platooning technology, which would double the truckway's capacity at significantly less cost than right of way and construction costs for adding two more lanes. The BC ratio for the second-phase truckway is estimated at 5.15.[41]

8. CONCLUSIONS

Despite the traditionally cited advantages of general purpose lanes, there is growing evidence that specialized lanes have a role to play in twenty-first century highways. Reduced lane widths and clearance heights would permit the addition of cars-only urban highway capacity in locations and configurations that have not been seriously considered, and at lower cost than conventional approaches to expanding expressways and arterials. In the inter-city market, specialized truck-only lanes could produce large productivity gains in goods movement, along with reduced environmental impact and significant safety benefits, due to the separation of large trucks from what will likely be smaller future automobiles. Separate lanes fit well into future urban road-pricing plans that take full account of the very large heterogeneity of values of time and reliability, for both individual motorists and trucking companies. Consequently, transportation planners should include consideration of at least the following types of non-GP lanes in their planning studies:

- Light-vehicle-only lanes and roadways, for both expressways and arterials in urban areas.
- Premium-priced and regular-priced lanes on urban expressways.
- Truck-only toll lanes in selected urban and inter-city corridors.
- Truck-only toll roads as alternatives to expanded rail lines in certain corridors.

REFERENCES

1. *Highway Capacity Manual*, Washington, DC: Transportation Research Board, 2000.

2. Pravin Variaya, "What We've Learned About Highway Congestion", *Access*, Vol. 27, 2005, pp. 2-9.

3. Peter Samuel, "How to 'Build Our Way Out of Congestion", Peter Samuel, Policy Study No. 250, Los Angeles: Reason Foundation, January 1999.

4. Peter Samuel, "State Senators Discuss Truck Lanes Concession on New Jersey Turnpike", TollRoadsNews.com, July 9, 2008.

5. Robert W. Poole, Jr. and Ted Balaker, "Virtual Exclusive Busways", Policy Study No. 337, Los Angeles: Reason Foundation, September 2005.

6. Ibid.

7. J. Kwon and P. Varaiya, "Effectiveness of California's High Occupancy Vehicle (HOV) System", *Transportation Research, Part C*, Vol. 16, No. 1, 2008, pp. 98-115.

8. Gordon J. Fielding and Daniel B. Klein, "High Occupancy/Toll Lanes", Policy Study No. 170, Los Angeles: Reason Foundation, November 1993.

9. Robert W. Poole, Jr., "Orange County's 91 Express Lanes", Policy Brief No. 39, Los Angeles: Reason Foundation, 2005.

10. Bill Reinhardt, "Minnesota DOT Gets Five Private Offers", Public Works Financing, November 1995.

11. Peter Samuel, note 3.

12. US Department of Transportation, "The US Department of Transportation's Comprehensive Truck Size and Weight Study", Washington, DC: Government Printing Office, August 2000.

13. Peter Samuel, Robert W. Poole, Jr., and Jose Holguin-Veras, "Toll Truckways: A New Path Toward Safer and More Efficient Freight Transportation", Policy Study No. 294, Los Angeles: Reason Foundation, June 2002.

14. "I-70 Final Supplemental EIS", June 2009 (www.improvei70.org).

15. Chen Feng Ng and Kenneth A. Small, "Tradeoffs Among Free-flow Speed, Capacity, Cost, and Environmental Factors in Highway Design", August 28, 2008 (www.economics.uci.edu/docs/2008-09/small-04.pdf)

16. Peter Samuel, "Innovative Roadway Design", Policy Study No. 348, Los Angeles: Reason Foundation, September 2006.

17. Ibid.

18. Christian Gerondeau, *Transport in Europe*, London: Artech House, 1997, Fig. 14, p. 340.

19. Peter Samuel, "French Low-Ceiling Tunnelways of Duplex A86 Comfortable to Drive", www.tollroadsnews.com/node/3906.

20. David Phillips, "An Update on Curb Guided Bus Technology and Deployment Trends", *Journal of Public Transportation*, 2006 (www.nctr.usf.edu/jpt/JPT%209-3S%20Phillips.pdf).

21. Samuel, Poole, and Holguin-Veras, note 13.

22. Robert W. Poole, Jr., Peter Samuel, and Brian F. Chase, "Building for the Future: Easing California's Transportation Crisis with Tolls and Public-Private Partnerships", Policy Study No. 324, Los Angeles: Reason Foundation, January 2005, pp. 19-25.

23. Kenneth A. Small, Clifford Winston, and Carol A. Evans, *Road Work: A New Highway Pricing & Investment Policy*, Washington, DC: The Brookings Institution, 1989, p. 11.

24. See Samuel, Poole, and Holguin-Veras, pp. 20-21 and 26-27.

25. Kenneth A. Small, Clifford Winston, and Jia Yan, "Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous Preferences in Policy Design", *Brookings-Wharton Papers in Urban Affairs*, 2006, pp. 53-96.

26. Douglass B. Lee, Jr., "Toward the Evaluation of Value Pricing", *Transportation Research Record No. 2079*, Transportation Research Board, 2008, pp. 71-78.

27. American Transportation Research Institute, "Measuring Travel Time in Freight-Significant Corridors", Office of Freight Management & Operations, Federal Highway Administration, FHWA-HOP-05-036, 2005.

28. Keith L. Killough, "Value Analysis of Truck Toll Lanes in Southern California", Southern California Association of Governments, Nov. 2007. (Available on the Transportation Research Board 2008 Annual Meeting CD-ROM)

29. Ng and Small, note 15, p. 21.

30. "Active Traffic Management: The Next Step in Congestion Management", Washington, DC: Federal Highway Administration, 2007. (http://international.fhwa.dot.gov/pubs/pl07012/index.cfm)

31. D. Lord, D. Middleton, and J. Whitacre, "Does Separating Trucks from Other Traffic Improve Safety?" *Transportation Research Record 1922*, 2005, pp. 156-166.

32. Lasse Fridstrom, "An Econometric Model of Car Ownership, Road Use, Accidents, and their Severity", TOI Report 457/1999, Oslo: Institute of Transport Economics, 1999. (www.toi.no/getfile.php/Publikasjoner/T%D8I%20rapporter/1999/457-1999/457-1999.pdf)

33. "The Large Truck Crash Causation Study", Washington, DC: Federal Motor Carrier Safety Administration, Publication No. FMCSA-RRA-07-017, July 2007. (www.fmcsa.dot.gov/ facts-research/research-technology/analysis/FMCSA-RRA-07-017.htm?=true.)

34. "Effectiveness and Impact of Corporate Average Fuel Economy Standards", Washington, DC: Transportation Research Board, National Research Council, 2002, p. 77.

35. Mike Tunnell, "Advanced Truck Engine Progress and Impacts on Highway Financing", presented at the Transportation Research Board 2006 Annual Meeting, American Transportation Research Institute, 2006.

36. Robert W. Poole, Jr. and Peter Samuel, "Corridors for Toll Truckways: Suggested Locations for Pilot Projects", Policy Study No. 315, Reason Foundation, February 2004, p. 6.

37. Andrew Marsay, "Sustainable Transport Infrastructure Investment", The Arup Journal, 3/2008.

38. Andrew Marsay, "A Case for Dedicated Freight Highways as an Alternative to Rail or Intermodal", presented at the annual general meeting of the Chartered Institute of Logistics and Transport, South Africa, 23 April 2008.

39. Jacques Beaumont, Jean-Marc Blosseville, Fabrice Hamelin, and Christophe Riper, "INNOFRET Technical Mission to the US on Innovative Truck, Rail, and Intermodal Goods Transport Systems", Paris: INRETS, presentation given in Los Angeles, Berkeley, Denver, and Washington, DC, Dec. 1-5, 2008.

40. Hsiao Shen Jacob Tsao and Jan Louis Botha, "An Automated Highway System Dedicated to Inter-City Trucking", San Jose State University Paper 2-2857, Transportation Research Board 2002 Annual Meeting.

41. Yafeng Yin, Mark A. Miller, and Steven E. Shladover, "Assessment of the Applicability of Cooperative Vehicle-Highway Automation Systems to Freight Movement in Chicago", presented at Transportation Research Board 2004 Annual Meeting. (http://database.path.berkeley.edu/reports/ index.cgi?reqtype=displayrecord&record=204)

DEDICATED LANES, TOLLS AND ITS TECHNOLOGY

Robin LINDSEY

University of Alberta
Edmonton

Canada

SUMMARY

ABSTRACT

The merits of separating cars and trucks have long been debated. Potential advantages include smoother traffic flows, lower accident rates, improved air quality and reduced maintenance and road infrastructure costs. Large trucks are often banned from urban roads and restricted to certain lanes on many highways but there are no dedicated truck facilities. However, truck-only lanes and truck tollways are now being actively studied. Tolls on cars and trucks are also becoming increasingly common and could be used to distribute car and truck traffic over road networks more efficiently.

This paper reviews the potential benefits from separating cars and trucks onto different lanes or roads while treating road infrastructure as given. U.S. studies of mixed traffic operations, lane restrictions and differential speed limits do not provide consistent evidence whether separating cars and trucks either facilitates traffic flows or reduces accident rates. Analysis with an economic model reveals that the potential benefits depend on the relative volumes of cars and trucks, capacity indivisibilities and the impedance and safety hazard that each vehicle type imposes. Differentiated tolls can support efficient allocations of cars and trucks between lanes. Lane access restrictions are much more limited in effectiveness. Toll lanes that are dedicated to either cars or trucks are a potentially attractive hybrid policy. Intelligent Transportation Systems (ITS) technology can help to improve safety and travel time reliability, and help drivers select between tolled and untolled routes.

1. INTRODUCTION

Most roads can be used by both cars and trucks even though these vehicles can differ greatly in size, weight, maneuverability and other characteristics [1]. Large trucks are often banned from urban roads and restricted to certain lanes on many highways, but there are no dedicated truck facilities. However, a number of truck-only lane and truck tollway projects have been proposed in the U.S. (Reich *et al.*, 2002; Federal Highway Administration, 2003; Transportation Research Board, 2003; Poole, 2007; Killough, 2008; GAO, 2008). Truck-only corridors between the U.S. and Canada, and truck-only road networks in Britain, Italy, and the Netherlands have also been studied.

Many proposed truck facilities would be tolled. Truck tolls are levied on 8,000 km. of U.S. roads, and over 20 European countries impose tolls on Heavy Goods Vehicles (Broaddus and Gertz, 2008). Tolls are imposed for various reasons including demand management, reduction of greenhouse gas emissions and revenue generation, but not specifically to separate light and heavy vehicles. Several potential advantages of separating cars and trucks have nevertheless been identified. Creating more homogeneous traffic flows could alleviate congestion by reducing the need for braking, accelerating, overtaking and changing lanes. Cars would suffer fewer delays from slow-moving trucks and less interference in fields of vision. Reducing congestion would also make travel times more predictable. Accident rates could fall, and fatalities could drop because of fewer crashes between light and heavy vehicles. Air quality would improve with higher and less variable speeds. Truck-only facilities can be designed to accommodate Long Combination Vehicles that exploit economies of vehicle size (Samuel *et al.*, 2002). And if truck traffic is gradually concentrated on dedicated facilities, other roads will require less maintenance and new roads designed for cars can be built to a lower standard (Holguín-Veras *et al.*, 2003).

Many considerations arise in designing dedicated facilities: car and truck traffic volumes, availability of uninterrupted rights-of-way, locations of entrances and exits, numbers of lanes and lane widths, pavement thickness, speed limits, services such as truck stops and refueling stations, and the possibility of allowing mixed use on some lanes such as High Occupancy Toll lanes (Chu and Meyer, 2008). Deciding whether to toll cars and/or trucks, and setting the levels of tolls by vehicle type, road segment, time of day and so on are also challenges.

This paper does not attempt to address all these design aspects. It concentrates on the benefits of vehicle separation using dedicated lanes or roads and/or tolls. Capital and operating costs of dedicated facilities, costs of levying and enforcing tolls and many other practical considerations are ignored. The analysis focuses on three questions. First, does vehicle separation enhance operations and safety? Second, is the unregulated equilibrium allocation of cars and trucks across lanes and roads optimal? Third, if the allocation is not optimal what is the best means of intervention?

Section 2 reviews the limited empirical evidence on the advantages and disadvantages of separating cars and trucks. Section 3 summarizes a study by De Palma *et al.* (2008) that assesses the benefits of vehicle separation and compares the effectiveness of lane access restrictions, differentiated car and truck tolls, and toll lanes for either cars or trucks. Section 4 follows up by addressing some important considerations left out of the model. Concluding remarks are made in Section 5.

2. MERITS OF VEHICLE SEPARATION: EMPIRICAL EVIDENCE

No dedicated truck lanes or roads yet exist, but studies of mixed traffic, truck lane restrictions and differential speed limits provide some evidence on the advantages and disadvantages of separating cars and trucks.

2.1. Mixed traffic

For several reasons trucks contribute more to congestion than do cars: they occupy more road space, they are slower to accelerate and decelerate and to negotiate turns, and they obscure vision more. A standard procedure to account for the greater impedance of trucks is to use a Passenger Car Equivalent (PCE). Typical PCE values are 1.5-2 for single-unit trucks and 2-3 for combination vehicles. Larger PCE values are used on hilly roads. The PCE factor has two limitations for assessing the merits of separating cars and trucks. One is that the impedance created by a vehicle may depend on the composition of vehicles in the traffic stream (Demarchi and Setti, 2003). Some studies have found that the PCE of trucks is an increasing function of the fraction of trucks (Yun *et al.*, 2005). A second, and more fundamental limitation is that while the PCE measures the overall impedance created by trucks it does not account for their separate effects on cars and trucks. These effects are not yet well understood (Peeta *et al.*, 2004).

In addition to congestion, trucks create safety hazards for other vehicles. Several truck characteristics suggest that these hazards are greater for cars than trucks. Long trucks have extensive blind spots and drivers may have difficulty seeing smaller vehicles beside and behind them. Trucks obscure a wider field of view for car drivers and the blockage is magnified when a column of trucks is traveling in the same lane. Trucks block sight of other vehicles as well as roadside and overhead signs – although the extent of this problem has not been studied (TRB, 2003). On bad roads and in bad weather trucks throw up water and debris that may cause vehicle damage and obscure vision. Trucks create obstacles and hazards when they lose their loads or blow a tire. And trucks with heavy axle loads cause road damage which, over time, may reduce safe driving speeds and increase wear and tear for cars.

Trucks also have features that enhance their safety. Advances in antilock brakes have improved truck braking performance, and on wet surfaces braking distances are comparable to those of cars (TRB, 2003). Because truck drivers sit higher off the road than car drivers they can see further and respond more quickly to safety hazards. Perhaps most important, many truck drivers are experienced professionals with strong incentives to drive safely.

Empirical evidence on car and truck accidents is varied and rather complex. Overall accident rates in the U.S. per 100 million vehicle-miles are lower for large trucks than cars, but fatal crash rates are higher and in collisions involving cars and trucks the car driver is much more likely to be killed (Adelakun and Cherry, 2009). Accidents involving trucks are more likely to be sideswipes and less likely to be truck-into-car rear-ends or run-off-the-road crashes (Golob and Regan, 2004). Road characteristics such as grades, lane widths, lateral sight distances and curves affect truck performance and accident hazards. Traffic volumes are also a factor. According to simulation models (e.g. Garber and Liu, 2007) accident rates per vehicle-mile increase with volume, but costs fall because of reduced accident severity. Lane changes per vehicle-mile – which are correlated with accident rates – increase with the fraction of trucks in the traffic volume up to about 25%, but drop beyond that (Siuhi and Mussa, 2007). Studies differ as to whether variance in speeds contributes to crash rates.

The behaviour of car drivers is affected by the presence of trucks in ways that can affect safety. There is some evidence that car drivers maintain longer headways when following trucks than cars (Yoo and Green, 1999). Car drivers are more inclined to overtake trucks than cars and to overtake them more quickly. Car drivers experience psychological discomfort from the presence of trucks - particularly in bad weather and at intermediate traffic volumes when both the probability and potential severity of collisions is elevated (Peeta *et al.*, 2004).

2.2. Lane restrictions

Large trucks are prohibited on many highways from using certain lanes. Most restrictions in the U.S. apply 24 hours a day to ease enforcement and driver compliance. Restrictions are sometimes voluntary, and many states do not attempt to enforce those that are mandatory (TRB, 2003).

Studies vary on how lane restrictions affect traffic operations. Using simulation software Rakha *et al.* (2005) concluded that providing separate lanes for trucks enhances performance as measured by speeds, fuel consumption and emissions. Not surprisingly, passenger vehicles benefit more from vehicle separation during peak hours when congestion is high (Siuhi and Mussa, 2007; Florida DOT, 2008) and on highway sections with extended upgrades. Lane restrictions are found to be more effective on highways with three or more lanes in each direction than on highways with only two (Stanley, 2009) and on freeways with limited access. Studies differ as to whether trucks should be restricted to the outer lane (Florida DOT, 2008) or inside lane (Adelakun and Cherry, 2009).

Before and after crash data are sometimes not available to assess the safety effects of truck lane restrictions at particular sites, and studies have employed microscopic computer simulations. Both simulation studies and studies with crash data have produced conflicting results and there is no clear evidence that truck lane restrictions have either positive or negative effects on safety (TRB, 2003).

2.3. Differential speed limits

Many highways have different speed limits for cars and trucks. The practice is controversial and arguments are made both for and against differential speed limits. Inferior maneuverability and braking capabilities of trucks militate in favour of lower speeds, at least in mixed traffic, although as noted above truck drivers tend to have superior vision and driving skills that enhance truck safety. Differential limits may increase speed variance and induce more frequent lane changes that increase the rate of car-into-truck rear-end collisions and sideswipes, but reduce other types of accidents such as truck-into-car rear-end collisions (Harkey and Mera, 1994). Evidence on the safety effects of differential speed limits is relatively sparse. There is little difference in overall accident rates or severity between U.S. states with uniform speed limits and differential limits although the types of collisions and the roles of cars and trucks appear to differ (TRB, 2003)

In summary, the evidence from U.S. studies of mixed traffic operations, lane restrictions and differential speed limits does not provide a clear indication whether separating cars and trucks either facilitates traffic flow or reduces accident rates. Findings vary with the composition and overall volume of traffic, type of road and terrain, whether dedicated lanes are located on inside or outside lanes and other factors.

3. MERITS OF VEHICLE SEPARATION: MODELING RESULTS

The potential benefits from separating cars and trucks were recently analyzed by De Palma *et al.* (2008) using a microeconomic model. This section provides a summary of the model and presents the analytical and numerical results of greatest interest.

The model features two vehicle types: light-duty vehicles ("*Lights*", denoted by subscript L) and heavy-duty vehicles ("*Heavies*", denoted by subscript H) [2]. Fixed numbers of trips by each vehicle type are made from a common origin to a common destination using one of two routes, Route 1 and Route 2, that can be either different roads or parallel traffic lanes of the same road. The total number of trips by type g is N_g, and the number of trips by type g on route r is N_{gr}. The total private cost of a trip by type g on route r is a linear increasing function of the numbers of vehicles of each type that take the same route:

$$C_r^L = \underbrace{F_r^L}_{(a)} + \underbrace{c_{Lr}^L N_{Lr}}_{(b)} + \underbrace{c_{Hr}^L N_{Hr}}_{(c)} + \underbrace{\tau_r^L}_{(d)},\ r = 1,2\,, \quad (1)$$

$$C_r^H = \underbrace{F_r^H}_{(a)} + \underbrace{c_{Lr}^H N_{Lr}}_{(b)} + \underbrace{c_{Hr}^H N_{Hr}}_{(c)} + \underbrace{\tau_r^H}_{(d)},\ r = 1,2\,. \quad (2)$$

Term (a) in eqns. (1) and (2) is the fixed cost of a trip. Term (b) is the additional cost imposed by *Lights* that use the same route and term (c) is the additional cost imposed by *Heavies*. Coefficients c_{Lr}^L and c_{Hr}^H, r = 1, 2, specify the external costs imposed by each vehicle on vehicles of the same type using the same route, and are called *own-cost coefficients*. Coefficients c_{Hr}^L and c_{Lr}^H specify the external costs imposed by each vehicle on vehicles of the other type, and are called *cross-cost coefficients*. Both the own-cost and the cross-cost coefficients depend on the capacities of the routes. Finally, term (d) is the toll (if any). It is assumed that tolls can be differentiated both by vehicle type and route.

In the absence of tolls or lane restrictions, drivers of *Lights* and *Heavies* are free to take either route. Three types of equilibrium usage configurations are possible: an *integrated* equilibrium in which both *Lights* and *Heavies* use each route, a *partially separated* equilibrium in which one type uses both routes and the other type uses only one route, and a *segregated* equilibrium in which each type uses only one route. Which of the three configurations prevails depends on the numbers of each vehicle type, N_L and N_H, and on the magnitudes of the cost coefficients, c_{hr}^g, which in turn depend on route capacities. Define $c_{h\bullet}^g \equiv c_{h1}^g + c_{h2}^g,\ g = L,H,\ h = L,H$. De Palma *et al.* (2008) show that a necessary condition for an *integrated* equilibrium is:

$$\frac{c_{H\bullet}^L}{c_{L\bullet}^L} < \frac{c_{H\bullet}^H}{c_{L\bullet}^H}. \quad (3)$$

Condition (3) stipulates that the ratio of the external cost imposed on a *Light* vehicle by a *Heavy* vehicle to the cost imposed on a *Light* vehicle by another *Light* vehicle ($c_{H\bullet}^L / c_{L\bullet}^L$) must be smaller than the corresponding ratio of costs imposed on a *Heavy* vehicle ($c_{H\bullet}^H / c_{L\bullet}^H$). If condition (3) holds, *Lights* prefer to travel on a route with *Heavies* and *Heavies* prefer to travel on a route with *Lights*. Condition (3) can be satisfied even if individual *Heavy* vehicles impose much higher congestion, accident and other external costs than do individual *Light* vehicles. What matters is the **relative** costs that *Light* and *Heavy* vehicles impose on each other.

Since travel demand is assumed fixed, the optimal allocation of drivers between routes is one that minimizes total social costs. Similar to the unregulated equilibrium, the optimum can be integrated, partially separated or segregated. However, the social cost of a trip by either vehicle type differs in two respects from the private cost given in eqns. (1) and (2). First, it excludes the toll because this is a transfer. Second, it includes the external costs of emissions, noise, vibration and pavement damage that are (mostly) borne by the population at large rather than by road users. For brevity these costs are referred to as *environmental* costs.

For several reasons the optimal allocation of vehicles to routes differs from the unregulated equilibrium. One is that drivers ignore the environmental costs of their trips, and another is that drivers are biased towards taking the route with lower fixed costs [3]. Since environmental and fixed costs are likely to be similar – if not identical – for lanes of the same road, these biases may not apply. However, the external costs reflected by the own-cost and cross-cost coefficients generally do not balance out between routes even for lanes of the same road. De Palma *et al.* (2008) show that a necessary condition for the optimal allocation to be integrated is:

$$c^L_{L\bullet} c^H_{H\bullet} > c^L_{H\bullet} c^H_{L\bullet} + \frac{1}{4}\left(c^L_{H\bullet} - c^H_{L\bullet}\right)^2 . \tag{4}$$

Rearranging Condition (3) the corresponding condition for the unregulated equilibrium to be integrated is:

$$c^L_{L\bullet} c^H_{H\bullet} > c^L_{H\bullet} c^H_{L\bullet} . \tag{5}$$

Since the quadratic term on the right-hand side of condition (4) is positive, condition (4) is more stringent than condition (5) and the optimal allocation may be partially separated or segregated even if the unregulated equilibrium is integrated. To see why, suppose that $c^L_{H\bullet} > c^H_{L\bullet}$. Heavies then impose higher external costs on Lights than Lights impose on Heavies. Heavies are therefore biased towards traveling with Lights and it is optimal to keep Heavies away from Lights. Similarly, if $c^L_{H\bullet} < c^H_{L\bullet}$ Lights are biased towards traveling with Heavies and it is advantageous to keep Lights away from Heavies. Unless the capacities of the routes are roughly commensurate with the numbers of Light and Heavy vehicles it is not efficient to eliminate conflicts between Lights and Heavies by segregating them, but partial separation is still warranted.

When the unregulated equilibrium distribution of *Light* and *Heavy* vehicles between the routes is not optimal various traffic control measures can be considered. Three such measures are considered here: lane access restrictions, tolls on both vehicle types and both lanes, and a toll lane restricted to one vehicle type. Because *Light* and *Heavy* vehicles impose different external and environmental costs on each route undifferentiated tolls are inadequate to support the optimum. But tolls that are differentiated by both vehicle type and route do suffice [4].

The scope for lane restrictions to support an efficient distribution of traffic turns out to be rather limited. If the optimum is segregated it can be supported simply by restricting each type to its designated route. But lane restrictions clearly don’t work if the optimum is integrated. Moreover, if the optimum is partially separated lane restrictions typically don’t work either because, while one vehicle type can be restricted to its designated route, the other vehicle type will not allocate itself between routes in optimal proportions. Indeed, if the unregulated equilibrium is integrated a lane restriction on one type can actually increase total travel costs [5].

The third policy instrument, a toll lane, entails dedicating one route to one vehicle type and levying a toll on it. A toll lane is therefore a hybrid of a lane restriction and a toll. A single toll lane cannot support the optimum if it is integrated. However, under certain conditions a single toll lane can decentralize the optimum if it is segregated or partially separated [6]. To consider one case, suppose the optimum is partially separated with only *Lights* using Route 1. Making Route 1 a toll lane for *Lights* supports the optimum if *Lights* are biased against using Route 2 (i.e., if $c^L_{H2} > c^H_{L2}$) because this bias can be offset by imposing a positive toll on *Lights* using Route 1. However, if *Lights* are biased against using Route 1 (i.e., if $c^L_{H2} > c^H_{L2}$) imposing a toll on Route 1 would only exacerbate the misallocation.

To examine more closely the potential benefits from lane separation, comprehensive tolls and toll lanes, De Palma *et al.* (2008) developed a specific version of the general model in which the external costs of travel are due to congestion and accidents: $c_{hr}^{g} = cong_{hr}^{g} + acc_{hr}^{g}$, $g = L, H$; $h = L, H$; $r = 1, 2$, where $cong_{hr}^{g}$ and acc_{hr}^{g} are congestion and accident cost coefficients respectively. The relative congestion costs imposed by *Heavies* on *Lights* and *Lights* on *Lights* are assumed to be:

$$\frac{cong_{Hr}^{L}}{cong_{Lr}^{L}} = \lambda_{H}^{L} PCE_{cong}, \; r = 1, 2, \qquad (6)$$

where PCE_{cong} is a generic PCE for congestion created by *Heavies* and $\lambda_{H}^{L} \geq 1$ is a scale factor to reflect the greater impedance that *Heavies* may imposed on *Lights* for reasons discussed in Section 2. The relative accident costs imposed by *Heavies* on *Lights* and *Lights* on *Lights* are given by an analogous expression:

$$\frac{acc_{Hr}^{L}}{acc_{Lr}^{L}} = \phi_{H}^{L} PCE_{acc}, \; r = 1, 2, \qquad (7)$$

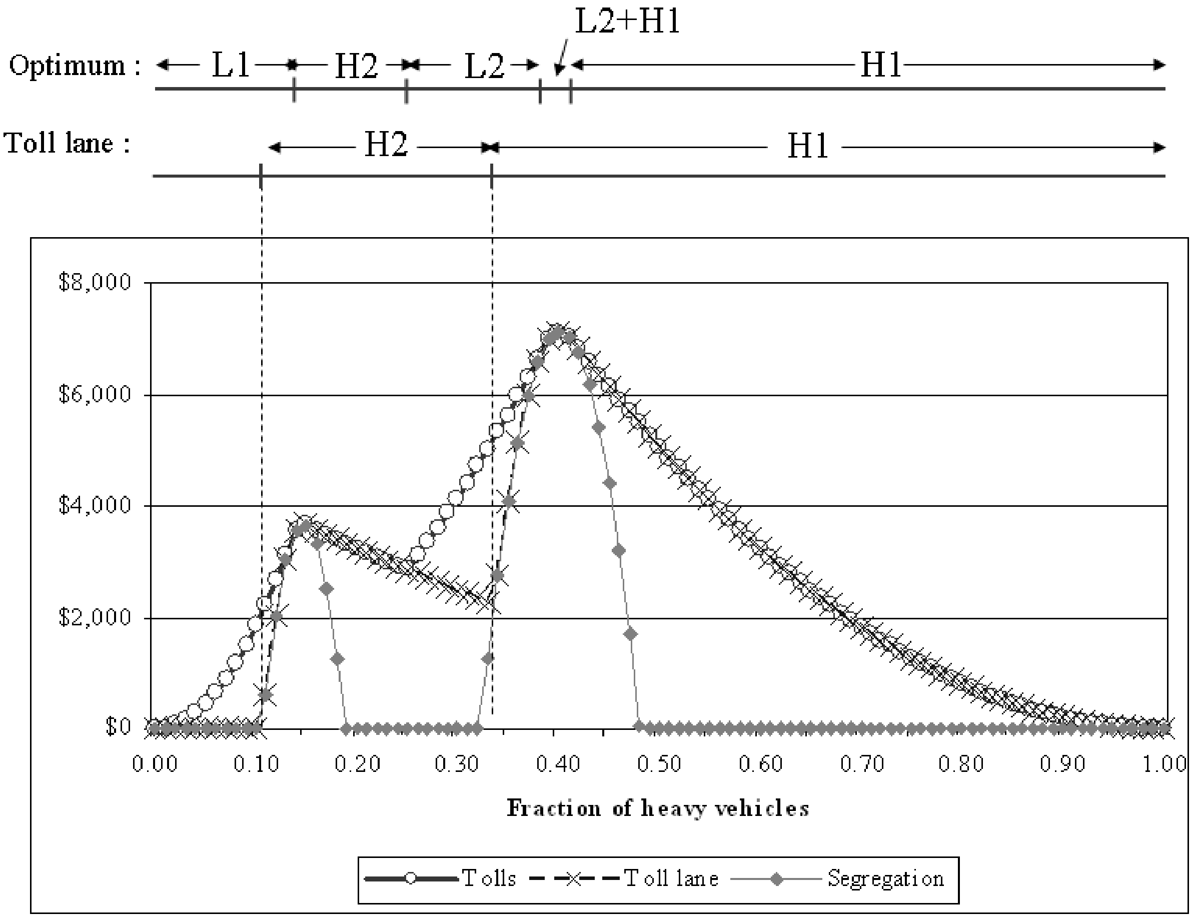
where PCE_{acc} is a generic PCE for accident costs created by *Heavies*, and $\phi_{H}^{L} \geq 1$ is a scale factor to account for the disproportionate safety hazard or fear that *Heavies* may impose on *Lights*. Base-case values of the parameters are given in Table 1. They describe a three-lane road with two lanes comprising Route 1 and the remaining lane comprising Route 2. The $cong_{Lr}^{L}$ and acc_{Lr}^{L} coefficients are calibrated so that in the unregulated equilibrium the marginal external congestion cost of a *Light* vehicle is about \$0.10/mile (€0.044/km) on each route, and the marginal external accident cost is about \$0.02/mile (€0.009/km) [7].

Table 1. **Base-case parameter values**

	Route 1	**Route 2**
Capacity	4 000 PCE /hour	2 000 PCE /hour
Speed limit	65 mph	65 mph
Length	32.5 miles	32.5 miles
Total trips	40,000	
Proportion of *Heavies*	Range 0-100%	
Travel costs	***Light* vehicles**	***Heavy* vehicles**
Fixed costs	\$0.194/mile	\$0.42/mile
Values of time	\$12/hour	Range
Congestion PCE for *Heavies* (PCE_{cong})		2
Relative impedance of *Lights* by *Heavies* (λ_{H}^{L})	1	
Accident PCE for *Heavies* (PCE_{acc})		0.75
Relative crash hazard for Lights (ϕ_{H}^{L})	1	
Relative cost of accident for *Heavies*		1
Environmental costs	\$0.0223/mile	\$0.2153/mile

Source: De Palma *et al.* (2008).

Figure 1. **Benefit from tolls, toll lane and segregation vs. fraction of *Heavy* vehicles**



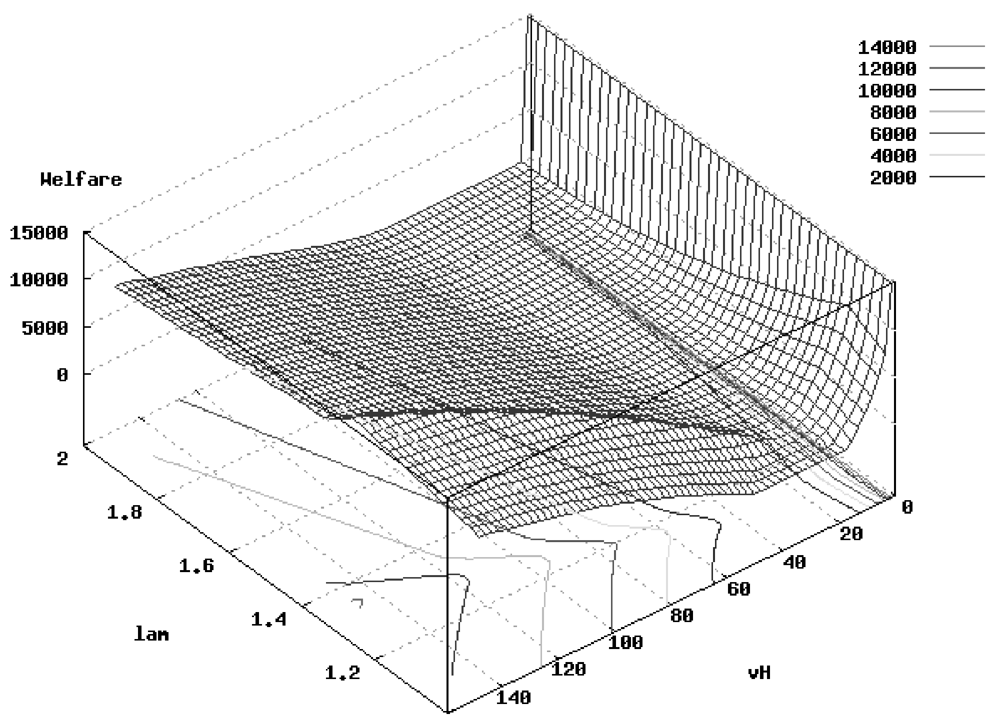
Source: Author's construction.

De Palma *et al.* (2008) compute unregulated equilibria and optima for a wide range of parameter values. Attention is restricted here to a few alternatives that illustrate the lessons of greatest policy interest. As a first alternative, the value of time (VOT) for *Heavies* is set to $75/hr (€53/hr). Condition (3) is satisfied and the unregulated equilibrium is integrated, but condition (4) is violated so that the optimum is not integrated. Since the VOT for *Heavies* is over six times the VOT of $12/hr for *Lights*, the main benefit from intervention is to reduce congestion delay for *Heavies* by allocating lots of road space to them.

As shown at the top of Figure 1 by the label "L1", if the fraction of *Heavies* in the traffic mix, call it *f*, is below about 0.14, Route 1 (with two thirds of total capacity) is dedicated to *Light* vehicles and *Heavies* are confined to Route 2. For $f \in (0.14, 0.24)$, Route 2 is dedicated to *Heavies* (H2) and *Lights* are confined to Route 2. For $f \in (0.24, 0.39)$, Route 2 is dedicated to *Lights* (L2). Within the narrow range $f \in (0.39, 0.41)$ segregation is optimal with Route 2 dedicated to *Lights*, and Route 1 dedicated to *Heavies* (L2+H1). Finally, for *f* > 0.41 Route 1 is dedicated to *Heavies* (H1) and *Lights* are confined to using the one lane on Route 2. As *f* varies from 0 to 1, the optimal allocation pattern of traffic to routes changes four times which highlights the importance of the proportion of *Heavy* vehicles in determining efficient use of road space [8]. As noted above, the optimum can be decentralized using differentiated tolls. The benefit, shown in Figure 1, exhibits a double peak with a local minimum at *f*=0.24 where *Heavies* are shifted from Route 2 to Route 1 and the allocation of vehicle types to routes is relatively less important.

The most effective toll lane configuration is identified just below the optimal pattern near the top of Figure 1. For *f*<0.10 there are two few *Heavies* to justify a toll lane. For $f \in (0.10, 0.33)$ a toll lane for *Heavies* on Route 2 is beneficial (H2), and for *f*>0.33 a toll lane on Route 1 (H1) is best. The toll lane configuration coincides with the optimal configuration over two subintervals of *f* and the toll lane supports the optimum over much of this range. By contrast, segregation is optimal only within the narrow range $f \in (0.39, 0.41)$ identified above, and segregation is beneficial (but not optimal) only for two relatively small intervals $f \in (0.11, 0.18)$ and $f \in (0.33, 0.48)$ within which the proportion of heavy vehicles is roughly commensurate with the capacity of either Route 2 or Route 1.

Figure 2. **Benefit from intervention *vs.* VOT for *Heavies* and relative impedance of *Lights* (20% *Heavies*)**



Source: De Palma *et al.* (2008).

Figure 2 shows the results of a second experiment in which the value of travel time for *Heavies* (denoted by v^H) and the relative impedance of *Heavies* (λ_H^L) are simultaneously varied while holding the fraction of *Heavies* fixed at 20%. The benefit from intervention is a roughly U-shaped function of v^H. For low values of v^H it is beneficial to keep *Heavies* away from *Lights*, whereas for large values of v^H (such as $75/hr. in the first experiment) it is advantageous to keep *Lights* away from *Heavies*. For intermediate values of v^H in the neighborhood of $25/hr. Condition 4 is satisfied and both the optimum and unregulated equilibrium are integrated. Since fixed trip costs and environmental costs are the same for the two routes in the numerical example, the equilibrium allocation of traffic is unbiased and there is no scope for beneficial intervention. This region shows up in Figure 2 where the surface is flat and the benefit is zero.

Turning attention to the effect of parameter λ_H^L it is apparent from the ridge on the surface in Figure 2 that the benefit of intervention is greatest when λ_H^L is slightly greater than one and falls off on either side. As λ_H^L begins to increase above 1 the cross-congestion-cost coefficients, $cong_{Hr}^L$, rise and so does the benefit from separating *Heavies* from *Lights*. But when λ_H^L exceeds a threshold value, the unregulated equilibrium becomes separated and moves closer to the optimal traffic allocation. This illustrates that the benefits from intervention depend on both the unregulated and optimal traffic allocation configurations. Varying parameter ϕ_H^L, the relative crash hazard for *Light* vehicles, has a similar inverse V-shaped effect on the benefits.

The model in De Palma *et al.* (2008) conveys at least two important policy lessons regarding vehicle separation. One is that lane-access or route-access restrictions are generally less effective than comprehensive tolls and may provide no benefit at all. The second and related lesson is that lane capacity indivisibilities make it difficult to allocate capacity between vehicle types in efficient proportions. Building dedicated truck facilities is cost-effective only if truck volumes are sufficiently high: a lesson that also applies to High Occupancy Vehicle (HOV) lanes for passenger transportation (Small, 1983; Dahlgren, 1998).

4. FURTHER CONSIDERATIONS

The model used in Section 3 is limited to driver's choice of lane or route on a corridor with fixed travel demand, and disregards various potentially important practical considerations such as travel time uncertainty and trip timing. Some of these factors are reviewed in this section.

4.1 Values of travel time, values of reliability and uses of information technology

Values of travel time

As the analysis in Section 3 makes clear, values of travel time for passenger and freight trips are key parameters in determining the benefits of separating cars and trucks. The VOT for automobile travel has been estimated in numerous studies [9]. It varies with trip purpose, vehicle occupancy, income and other factors. In order to do an accurate assessment of a specific project it is necessary to determine the proportions of trips made for business, commuting and leisure as well as the socioeconomic characteristics of the traveling population.

Valuation of travel time not been studied as thoroughly for freight transport as it has for passenger transport although its importance is now widely recognized. Truck VOTs depend on many factors: vehicle type and load, importance of punctual delivery, whether the truck is operated in-house or for-hire, truck drivers' wage rates and working hours, etc. VOTs tend to be higher for shippers than transporters and depend on whether receivers have an input into the scheduling of deliveries (Hensher and Puckett, 2008). VOT estimates range over an order of magnitude in developed countries and are highly skewed (Kawamura, 2000).

There has been a longstanding debate in the literature on whether VOT depends on trip duration and the size of travel time savings: relevant questions for trips on dedicated facilities which could range in length from a few kilometers to hundreds of kilometers on interstate or international road networks. In theory VOT could either rise with distance due to fatigue or boredom, or fall because trips many not be scheduled as tightly for long hauls [10]. The value attached to small travel time savings depends, *inter alia,* on whether the amount saved is enough to make an additional delivery during a driver's shift. This will vary from trucker to trucker and may average out in the aggregate.

Values of reliability

Variability in travel time is absent from the deterministic model used in Section 3. A study by Cambridge Systematics (2005) [11] identified the sources of highway congestion as bottlenecks (40%), traffic incidents (25%), workzones (10%), bad weather (15%), poor signal timing (5%), and special events & other sources (5%). Depending on the information available to travelers about incidents, weather and so on, between roughly one quarter and one half of congestion delays are unpredictable.

Although the literature on travel time reliability has advanced greatly in the last decade there are still no generally accepted monetary values for the value of reliability (VOR). In travel demand studies VOR is often estimated by the coefficient on the standard deviation of travel time and VOT by the coefficient on mean travel time. The reliability ratio, ρ, is defined by the ratio of VOR to VOT. If the coefficient of variation (*CV*) of travel time is assumed to be constant the effect of variability in travel time can be accounted for simply by scaling up the VOT by the factor $1+\rho * CV$ (Institute for Transport Studies, 2008, p.21). A problem with this approach is that *CV* tends to increase with congestion because congestion magnifies the effects of incidents and other disturbances. Another problem is that *CV* tends to decline with trip length (Arup, 2003). These findings would suggest that reliability accounts for a smaller portion of total trip costs on longer and less congested roads. To the extent that lane restrictions and/or tolls reduce congestion the unit value of the resulting travel time savings are reduced as well: an obvious complication for project and policy evaluation.

Uses of information technology

The cost of travel time unreliability depends on how well system operators can control travel conditions and on what drivers know when they make their travel decisions. Intelligent Transportation Systems (ITS) technology is advancing in both directions (TRB, 2003). Ramp metering is an established and relatively simple technology that alleviates congestion by controlling the inflow rate onto limited-access highways. Slowly changing variable speed limits help to smooth traffic flows and reduce the incidence of rear-end collisions. Dynamic truck restrictions involve the use of dynamic message signs and specialized ramp metering to direct large trucks onto certain traffic lanes, and operate in some ways similarly to conventional lane restrictions. ITS is also contributing to truck safety with warning systems for long downgrades and curves, and on-board collision avoidance systems.

As far as driver aids dynamic message signs have long been used to provide en route trip guidance. Pre-trip information is also becoming increasingly available by phone, on the Internet and at public places. ITS can also be used in conjunction with tolling to inform travelers about toll levels and travel times on tolled and untolled facilities (FHWA, 2009). In the future, drivers may be able to program onboard navigation aids to select a route with the shortest distance, shortest expected travel time or lowest expected generalized cost (Chorus and Timmermans, 2008).

4.2 Route choice

The model in Section 3 is limited to two routes or sets of lanes in the same travel corridor and the only choice for drivers is which route or lane to take. In many settings other routes will be available. A potential drawback of restricting trucks to certain lanes and/or levying high truck tolls is that truckers may divert to secondary roads or city streets that are not designed to handle heavy vehicles and where congestion, accident and environmental externalities are worse. Traffic diversion has not been a major problem for the German Heavy Goods Vehicle (HGV) toll because many potential alternate routes are closed to trucks (Broaddus and Gertz, 2008). And some freeways – such as many in Atlanta – have no good alternative routes (Chu and Meyer, 2008). However, traffic diversion has been a problem in some countries such as France. Setting tolls when substitute or complementary roads are not tolled is a classic problem in second-best pricing. It requires rather detailed information on travel demands and costs even on simple road networks and the consequences of setting tolls at nonoptimal levels can be severe [12].

4.3 Trip timing

The model in Section 3 is static and implicitly assumes that cars and trucks travel at the same time. To the extent that they can use the same roads at different times, dedicated lanes or facilities are not actually needed to separate them. Passenger and freight traffic flows do follow different daily and weekly time patterns (Rakha *et al.*, 2005) and truckers naturally prefer to avoid commuting periods (Fischer *et al.*, 2003). However, truckers are limited as to when they can travel. Hours of service regulations, maritime terminal operating hours, neighborhood curfews and union-negotiated hours of operation impose constraints.

Shippers' hours are another important constraint on truck delivery times (Vilain and Wolfrom, 2001). Just-in-time inventory management systems require that deliveries be made at certain times, and time-sensitive goods such as express services call for immediate delivery. Many receivers with traditional operating hours would incur additional labour costs to accept deliveries during off-hours, and since truckers often make deliveries to several businesses on a single tour the additional labour costs of off-hours would be magnified (Holguín-Veras, 2005). As a consequence, time-of-day price elasticities tend to be lower for trucks than for cars. For example, when the Port Authority of New York and New Jersey introduced a peak-period congestion charge in 2001, only 6 percent of truckers shifted operations to off-peak hours. Two thirds of the truckers who continued to drive during the peak cited shippers' hours as the reason for not switching (Congressional Budget Office, 2009). Tillema *et al.* (2008) report similar results from a survey of Dutch firms. This suggests that the scope for temporal separation of cars and trucks is rather limited.

4.4 Vehicle type, logistics and location choices

Most trucking firms would have little incentive to modify their vehicle fleets if truck-only facilities or tolls were established on a single travel corridor. For regional or national road networks, however, there may be substantial productivity gains from using large combination vehicles (Samuel *et al.*, 2002). The Swiss HGV tolling scheme, introduced in 2001, is levied on all roads. It has had a dramatic impacts on truck volumes and has induced a shift towards larger and heavier vehicles (Broaddus and Gertz, 2008). The German HGV charge, which varies with vehicle emissions class, has induced shifts towards environmentally friendly vehicles. It has also encouraged a sharp reduction in the proportion of empty trips. To the extent that tolls and future truck-only facilities succeed in reducing congestion delays freight companies may be able to make more deliveries per day with each vehicle and require fewer vehicles to conduct business (Hensher and Puckett, 2008).

Another possible long-run response of firms to the introduction of truck-only facilities and tolls is to change the locations of their businesses and transfer terminals. Such adjustments would, in turn, affect firms' accessibility to input suppliers, customers and employees and trigger further location shifts (Tillema *et al.*, 2008). Little is yet known about the potential magnitude of these shifts or their effects on truck flows over road networks (Roorda *et al.*, 2009). Nevertheless, as long as first-best conditions hold elsewhere these complications do not invalidate the analysis in Section 3 of given volumes of car and truck trips on a single corridor.

5. CONCLUSIONS

Truck-only lanes and roads have been proposed as a way to alleviate traffic congestion, enhance safety and reduce other external effects of traffic. This paper focuses on the potential benefits of separating cars and trucks while taking road infrastructure and operating costs as given. Because no truck-only facilities have yet been built there is no experience with their operational and safety benefits. However there is evidence on the advantages and disadvantages of separating cars and trucks from studies of mixed traffic, truck lane restrictions and differential speed limits. The evidence from U.S. studies is varied and suggests that the effects of separation are sensitive to car and truck traffic volumes, type of road and terrain, location of dedicated traffic lanes on multilane highways and other factors.

To examine when vehicle separation is beneficial a simple economic model is used in which car and truck drivers choose between two lanes or routes. Routes can differ in fixed, environmental and external (i.e. own- and cross-) costs and each difference can distort the unregulated equilibrium allocation of traffic between routes. If the external cost imposed by cars on trucks differs from the external cost imposed by trucks on cars intervention calls for partially separating or segregating cars and trucks. The optimal allocation can be decentralized using tolls that are differentiated by vehicle type and route. Lane access restrictions are less flexible and, because of capacity indivisibilities, may be unwarranted. For example, a dedicated truck lane is unlikely to be cost effective if trucks account for only a small fraction of total traffic. Toll lanes – a hybrid of lane restrictions and tolls – are generally more effective than access restrictions because they offer a continuous rather than discrete degree of control.

As road transport technology advances, and other changes occur, the economics of dedicated facilities may strengthen or weaken. In most developed countries truck traffic has been growing more rapidly than passenger traffic and this strengthens the economics of building new, dedicated truck facilities or reserving lanes on existing roads for heavy vehicles. However, continuing improvements in vehicle safety could lower accident rates and reduce the safety hazard posed by trucks on lighter vehicles. In the longer term, automated roads could dramatically increase road capacity and reduce both congestion and accidents [13].

A further consideration is that comprehensive road pricing for both cars and trucks may be introduced in the coming years. The German HGV charge uses satellite-based technology to toll heavy trucks on federal motorways and could be extended to other roads, lighter trucks and passenger vehicles. In 2008, the Dutch Parliament approved a national distance-based system of user charges for passenger and freight vehicles. The fee per kilometer will vary by time of day and with vehicle emissions. The technology would permit tolling by vehicle type, lane, time of day and current road conditions and would facilitate vehicle separation using tolls as suggested here.

NOTES

1. Passenger vehicles range from small electric cars to sports utility vehicles, vans and pickup trucks and vary widely in their characteristics as well. Freight vehicles vary even more. The generic terms “cars” and “trucks” are used here for ease of reference.

2. *Lights* and *Heavies* correspond to cars and trucks in the rest of this paper.

3. This bias is well known in the literature; see Verhoef *et al.* (1996).

4. See Proposition 3 in De Palma *et al.* (2008). This result remains valid if travel demand is elastic. However, tolls do not internalise all decisions, such as driving speed and weaving between lanes, and a role remains for speed limits and other traffic laws to control these facets of driver behaviour.

5. See De Palma *et al.* (2008), Proposition 5.

6. See De Palma *et al.* (2008), Propositions 6 and 7.

7. For details, see De Palma *et al.* (2008), Sections 3 and 4.

8. More complicated allocation patterns can occur; see De Palma *et al.* (2008).

9. For recent literature reviews, see Small and Verhoef (2007) and Intervistas Consulting Inc. (2008).

10. For freight transport, average VOTs may increase with distance because a greater fraction of trucks have two drivers.

11. This information is taken from Congressional Budget Office (2009, Figure 1-1).

12. See Small and Verhoef (2007, Section 4.2).

13. To the extent that automated roads would operate more effectively with homogeneous vehicles, this reinforces rather than weakens the argument for vehicle separation.

BIBLIOGRAPHY

Adelakun, A. and C. Cherry (2009), "Exploring truck driver perceptions and preferences: Congestion and conflict, managed lanes, and tolls", 88th Annual Meeting of the Transportation Research Board, Washington, D.C. Conference CD Paper No. 09-1324.

Arup (2003), Frameworks for modeling the variability of journey times on the highway network. London: Arup. [Cited in Institute for Transport Studies (2008, p.51).]

Broaddus, A. and C. Gertz (2008), "Tolling heavy goods vehicles: Overview of European practice and lessons from German experience", Transportation Research Record 2066, 106-113.

Cambridge Systematics (2005), Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation, prepared for the Department of Transportation, Federal Highway Administration, September 1.

Chorus, C.G. and H.J.P. Timmermans (2008), "Personal Intelligent Travel Assistants", forthcoming in A. de Palma, R. Lindsey, E. Quinet and R. Vickerman, eds., Handbook in Transport Economics, Edward Elgar.

Chu, H-C. and M.D. Meyer (2008), "Screening process for identifying potential truck-only toll lanes in a metropolitan area", Transportation Research Record 206, 79-89.

Congressional Budget Office (2009), Using pricing to reduce traffic congestion, March.

Dahlgren, J. (1998), "High occupancy vehicle lanes: Not always more effective than general purpose lanes", Transportation Research Part A 32(2), 99-114.

Demarchi, S.H. and J. R. Setti (2003), "Limitations of passenger- car equivalent derivation for traffic streams with more than one truck type", Transportation Research Record 1852, Paper No. 03-3496, 96-104.

De Palma, A., M. Kilani and R. Lindsey (2008), "The merits of separating cars and trucks", Journal of Urban Economics 64(2), 340–361.

Federal Highway Administration (2009), Technologies That Complement Congestion Pricing - A Primer (http://www.ops.fhwa.dot.gov/publications/fhwahop08043/cp_prim3_00.htm, accessed April 21, 2009).

Fischer, M.J., D.N. Ahanotu and J.M. Waliszewski (2003), "Planning truck-only lanes: Emerging lessons from the Southern California Experience", 82nd Annual Meeting of the Transportation Research Board, Conference CD Paper No. 002048.

Florida Department of Transportation (2008), Evaluating the effectiveness of various truck lane restriction practices in Florida: Phase II, BD543-10 (1/08) (http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_TE/FDOT_BD543_10.pdf, accessed February 20, 2009)

Garber, N.J. and Q. Liu (2007), "Identifying impact of truck-lane restriction strategies on safety by using simulation", 86th Annual Meeting of the Transportation Research Board, Washington, D.C., Conference CD Paper No. 07-0885.

Golob, T.F. and A.C. Regan (2004) "Traffic conditions and truck accidents on urban freeways", UCI-ITS-WP-04-3, July (http://www.its.uci.edu/its/publications/its.html, accessed May 30, 2008).

Government Accountability Office (2008), Approaches to Mitigate Freight Congestion, GAO-09-163R, Washington D.C., November 20. (http://www.gao.gov/new.items/d09163r.pdf, accessed November 25, 2008).

Harkey, D.L. and R. Mera (1994), Safety impacts of different speed limits on cars and trucks: Final Report, Report No. FHWA-RD-93-161, Federal Highway Administration.

Hensher, D. and S. Puckett (2008), "Behavioural responses of freight transporters and shippers to road-user charging schemes: An empirical assessment", in E.T. Verhoef, M. Bliemer, L. Steg and B. van Wee, eds., Pricing in Road Transport: A Multi-disciplinary Perspective, Cheltenham, UK; Northampton, MA: Edward Elgar, pp. 29-62.

Holguín-Veras, J. (2007), "Necessary conditions for off-hour deliveries and the effectiveness of urban freight road pricing and alternative financial policies in competitive markets", Transportation Research Part A, doi:10.1016/j.tra.2007.10.008.

Holguín-Veras, J., D. Sackey, S. Hussain and V. Ochieng (2003), "Economic and financial feasibility of truck toll lanes", Transportation Research Record 1833, 66-72.

Institute for Transport Studies, University of Leeds, Imperial College and John Bates Services (2008), Multimodal Travel Time Variability: Final Report, November 11, 2008 (http://www.dft.gov.uk/pgr/economics/rdg/jtv/, accessed April 28, 2009).

InterVISTAS Consulting Inc. (2008), Value of Time and Reliability for Local Trips in Canada, Final Report, Prepared for Transport Canada, March 2008.

Kawamura, K. (2000), "Perceived value of time for truck operators", Transportation Research Record 1725, 31-36.

Killough, K. (2008), "Value analysis of truck toll lanes in Southern California", 87th Annual Meeting of the Transportation Research Board, Washington, D.C. Conference CD Paper No. 08-0140.

Peeta, S., W. Zhou and P. Zhang (2004), "Modeling and mitigating of car-truck intersections on freeways", Transportation Research Record 1899, 117-126.

Poole, R.W. (2007), Miami Toll Truckway: Preliminary Feasibility Study, Reason Foundation Policy Study 365, November 1, 2007 (http://reason.org/areas/topic/319.html, accessed June 13, 2009).

Rakha, H., A.M. Flintsch, K. Ahn, I. El-Shawarby and M. Arafeh (2005), "Evaluating alternative truck management strategies along I-81", Transportation Research Record 1925, 76-86.

Reich, S., J. Davis, M. Catalá, A. Ferraro and S. Concas (2002), The potential for reserved truck lanes and truckways in Florida, Center for Urban Transportation Research, University of South Florida.

Roorda, M., R. Cavalcante, H. Abdelgawad, B. Abdulhai and C. Woudsma (2009), "Exclusive truck facilities in the Toronto area: Rationale and model development", Proceedings of the 44th Annual Conference of the Canadian Transportation Research Forum, Victoria, May 24-27.

Samuel, P., R.W. Poole Jr. and J. Holguín-Veras (2002), Toll truckways: A new path toward safer and more efficient freight transportation, Reason Public Policy Institute (http://www.rppi.org).

Siuhi, S. and R. Mussa (2007), "Simulation analysis of truck-restricted and high-occupancy vehicle lanes", Transportation Research Record 2012, 127-133.

Small, K.A. (1983), "Bus priority and congestion pricing on urban expressways", in T.E. Keeler, ed., Research in Transportation Economics 1, JAI Press, Greenwich, Conn., pp. 27-74.

Small, K.A., and E.T. Verhoef (2007), The Economics of Urban Transportation, Second Edition, Routledge, London and New York.

Stanley, J. (2009), "Passenger car equivalents of trucks under lane restriction and differential speed limit policies on four lane freeways", thesis submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College, May (http://etd.lsu.edu/docs/available/etd-01212009-023639/unrestricted/stanleythesis.pdf, accessed May 23, 2009).

Tillema, T., B. van Wee, J. Rouwendal and J. van Ommeren (2008), "Firms: Changes in trip patterns, product prices, locations and in the human resource policy due to road pricing", in E.T. Verhoef, M. Bliemer, L. Steg and B. van Wee, eds., Pricing in Road Transport: A Multi-disciplinary Perspective, Cheltenham, UK; Northampton, MA: Edward Elgar, pp. 106-127.

Transportation Research Board (2003), Freight capacity for the 21st century.

Verhoef, E.T., P. Nijkamp and P. Rietveld (1996), "Second-best congestion pricing", Journal of Urban Economics 40(3), 279-302.

Vilain, P. and P. Wolfrom (2001), "Value pricing and freight traffic: Issues and industry constraints in shifting from peak to off-peak movements", Transportation Research Record 1707, 64-72.

Yoo, H. and P. Green (1999), "Driver behavior while following cars, trucks, and buses", www.umich.edu/~driving/publications/UMTRI-99-14.pdf, accessed May 30, 2009).

Yun, S., W.W. White, D.R. Lamb and Y. Wu (2005), "Accounting for the impact of heavy truck traffic in volume/delay functions within transportation planning models", 84th Annual Meeting of the Transportation Research Board, Conference CD Paper No. 05-2163.

THE INFORMED AND ORIENTED TRANSPORT SYSTEM USER

Peter ZIMMERMANN

Federal Ministry of Transport, Building and Urban Affairs
Bonn
Germany

SUMMARY

EXECUTIVE SUMMARY

The German Ministry of Transport, Building and Urban Affairs has set up a vision which shall improve the situation of the transport system user (in Germany) with respect to more and better transport system information and services, less stress, improved availability and increased safety. In order to make use of all resources of telematics applications an intensive co-ordination with the Ministry of Economics and Technology (transport research programme) was undertaken and the project “the informed and oriented transport system user” was launched to improve the business cases for information services.

The idea is to provide better access to online traffic data and therefore create a so-called “Meta Data Platform”. The Meta-Data-Platform shall enable data exchange of various sources, shall provide common interfaces and protocols, shall allow for exchange of geographical data of different formats and enable b2b processes between service and content provider organizations. Traffic data shall become more reliable, be of higher quality and enable access to real time data. State authorities for controlling traffic, broadcasting stations for traffic warning news, service providers for individual route recommendations and content owners can make use of the system. The services of the meta data platform can be used as a virtual internet portal. Centralized services for judgment of data quality transfer of different interfaces and protocols and transfer of different geographical formats will be offered separately for example for service providers.

Despite having intermodal transport in mind, there will be separate platforms for road and public transport. The realization has been splitted into a bunch of individual projects. Some of the projects have just been started. For some of them call for tenders are presently prepared. Development will take approximately four years. At the end of the development there will be test fields used for validation. Validation will be done with model services.

1. INTRODUCTION

1.1. The idea and objectives

In 2006 the German Government started the so-called high tech initiative. Federal ministry of transport, building and urban affairs (BMVBS) decided to support among others modern and innovative transport technologies and services. Securing mobility of persons and goods is vital for success and competitiveness of a modern society. Technologies and business models have been investigated quite often. However – especially for road traffic – the market for individual commercial services did not develop as expected. For this reason the ministry decided to push telematics applications and to lower barriers. The idea of a meta data platform was created. It should contribute to save resources and environment. It should make transport safer, more flowingly, more efficient and more undisturbed. Free access to information on all modes of transport should be enabled. Another argument was to secure mobility in a society which is growing older and the quality of traffic date should be improved.

The motivation was to improve the situation of transport systems users with respect to better and more reliable information, orientation and on the long run real time dynamic data. Comfort during travelling shall be improved. Individual planning shall be possible and b2b-business shall be enabled. Middle of 2006 the project was initiated. From the beginning both individual and public transport have been considered. The project is not a single project but a bunch of separate projects.

Potential users and experts have been involved in the definition of requirements. The platform should provide access to online traffic data by means of a virtual central internet portal without any discrimination, enable exchange of data from different partners by means of standardized interfaces, and enable access to data for service providers, for public authorities for collective traffic control, for broadcasting stations for traffic warning news. It should serve as link between service providers and content owners and should not affect the competition between service providers. It shall be the starting point for better travel information.

Reference to the German situation does not mean that no cross boarder transport is considered. International standards have been analysed. The envisaged solution makes use of the international practice and development.

1.2. Co-ordination and co-operation between the Federal Ministries of Transport, Building & Urban Affairs and Economics & Technology

From the beginning not only potential users and experts have been involved in the definition process. There are a lot of research and development projects for transportation technology supported and funded by the German federal ministry of economics and technology (BMWi). Therefore a co-ordination between both ministries was absolutely vital.

The burden was/is shared between the two ministries. Ministry of Transport, Building and Urban Affairs focuses, e.g., on:

- Metadataplatform for road transport
- Ownership and user rights for public transport
- Quality of traffic data for public transport
- Further development of superregional timetable information and
- Electronic fare management.

Ministry of Economics and Technology focuses, e.g., on:

- Quality of traffic data for road transport
- Routing in rail traffic
- In-door-routing in interchange stations
- Requirements/improvements of/for handicapped people.

During the next years additional specific research and development projects will be defined if there are further missing links detected which have to be closed for the focused informed and oriented transport system user.

1.3. Previous studies

In 2007 six studies have been carried out to analyse the state of the art and to identify existing missing links (most important results, see Chapter 3). These studies did also reflect the situation outside Germany. The reports are only available in German language, namely:

- Detection in road traffic (stationary and mobile)
  Florian Weichenmeier, Jan Körber, Thomas Meyer, Michael Ortgiese, Florian Schimandl, Ralf Weiss, Arnold Zwicker
  Forschungsbericht FE-Nr. 63.0002/2007/ im Auftrag des BMVBS
  Vorbereitende Begleituntersuchungen zur Metaplattform hier: Detektionsverfahren im Straßenverkehr, 2007
- Methods and practice for georeferencing
  Michael Ortgiese, Jan Körber, Andreas Schmid, Timo Hoffmann, Fritz Busch, Florian Schimandl
  Forschungsbericht FE-Nr. 63.0004/2007/ im Auftrag des BMVBS
  Vorbereitende Begleituntersuchungen zur Metaplattform hier: Georeferenzierungsverfahren für die Lokalisierung von Verkehrsnetzen, Verkehrsdaten und –informationen, 2007
- Existing data models, technical interfaces and protocols in road transport
  Josef Kaltwasser, Tilo Schön
  Forschungsbericht FE-Nr. 63.0005/2007 im Auftrag des BMVBS
  Vorbereitende Begleituntersuchungen zur Metaplattform hier: Analyse der technischen Rahmenbedingungen für die Integration von Verkehrsdaten in einer Metadatenplattform für Verkehrsinformation, 2007
- Methods and procedures to define quality of traffic data in road transport
  Michael Poschmann, Heribert Kirschfink
  Forschungsbericht FE-Nr. 63.0003/2007(im Auftrag des BMVBS
  Vorbereitende Begleituntersuchungen zur Metaplattform hier: Qualität, Qualitätsstufen und deren Kategorisierung, 2007
- Analysis and survey of existing information platforms in road traffic
  Steffi Klinghammer, Jan Kätker, Wilke Reints, Heiko Jentsch, Olaf Carsten Schiewe
  Forschungsbericht FE-Nr. 63.0006/2007 im Auftrag des BMVBS
  Vorbereitende Begleituntersuchungen zur Metaplattform hier: Bestandsaufnahme abgeschlossener und laufender Projekte zu Verkehrsinformationsplattformen für den Straßenverkehr, 2007
- Analysis and survey of existing timetable information platforms in public transport
  Volker Sustrate, Reinhold Pohl, Michael N. Wahlster, Bert Lange, Jürgen Roß, Jörg Janeke
  Forschungsbericht FE-Nr. 63.0007/2007 im Auftrag des BMVBS
  Vorbereitende Begleituntersuchungen zur Metaplattform hier: Bestandsaufnahme abgeschlossener und laufender Projekte zu Verkehrsinformationsplattformen für den öffentlichen Verkehr, 2007

The results of these projects have been included in the vision of the informed and oriented transport system user. Improvement in Germany shall be reached but international tendencies and developments shall not be precluded.

1.4. Relevant traffic information platforms

In Germany, some superregional platforms exist. The meta data platform shall be the link between service an content providers and shall not act as a service provider. However, these existing superregional platforms offer traffic information services. Examples are Mobile I-in-Rhineland, Ruhrpilot, Traffic management centre Berlin, MoBIN in Baden-Württemberg, Traffic information agency in Bavaria, Traffic information in North-Rhine-Westphalia and Traffic management centre Hesse.

European activities are, e.g., CENTRICO, CORVETTE, EUROROADS, CERTI, Streetwise, Travel Information Highway Great Britain, National data warehouse the Netherlands, VIKING Northern Europe and eMotion. EU-SPIRIT has to be mentioned with respect to public transport.

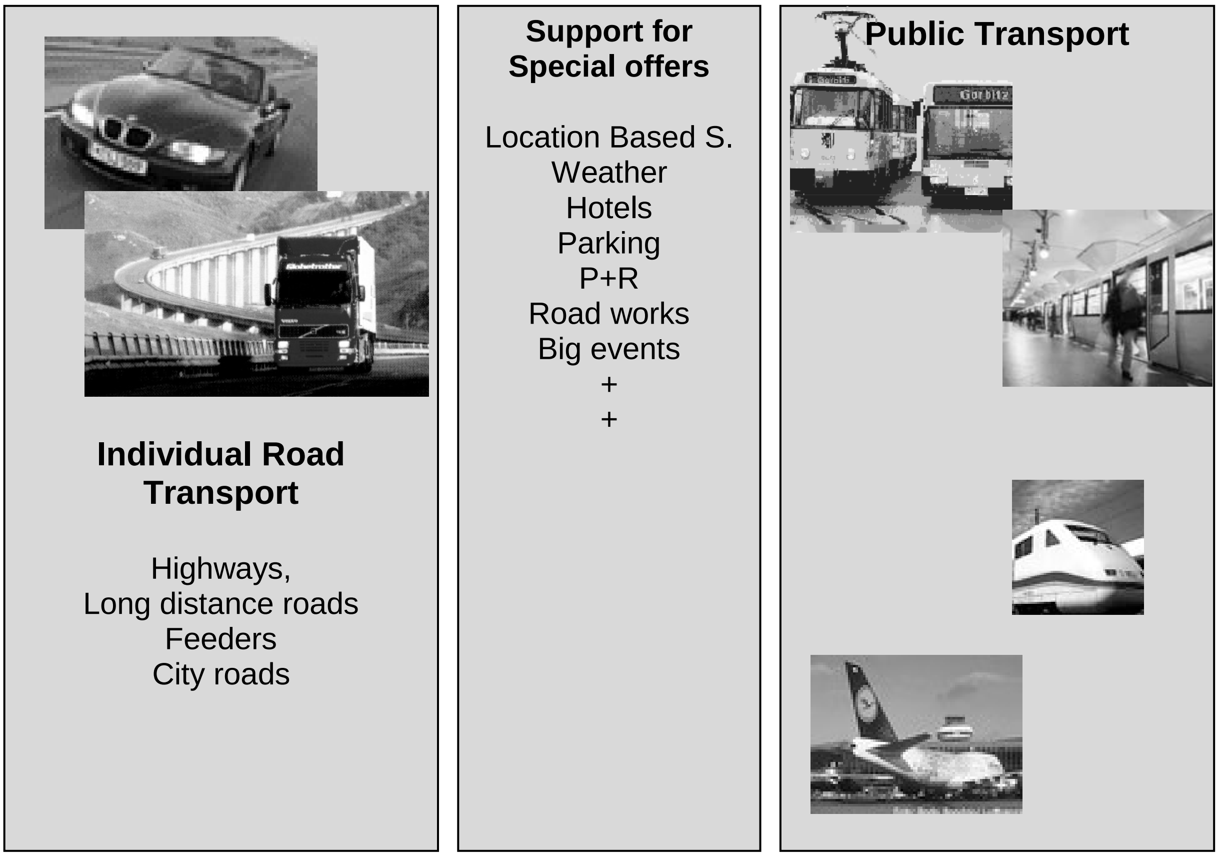
2. SCOPE OF THE PROJECT

2.1. Intermodal transport

The situation is quite different for individual road and public transport. From the beginning the ministry insisted to reflect intermodal transport mode. This is in line with the ground rules of German transport politics. The goal is to reduce individual road traffic and to strengthen public transport.

The next chart illustrates the basic situation. Most probably the available traffic data for individual road transport, public transport and services on demand will not be hosted on a common platform. But if an enterprise wants to offer intermodal transport services to customers the architecture of the metadataplatform shall not exclude this. By the way metadataplatform shall not be a physical platform which stores and manages traffic data (see chapter 4). Individual road transport shall enclose both passenger traffic and freight traffic in the (on the long run) entire road network (long distance and cities). Public transport has – also on the long run – to cover passenger traffic with busses, passenger traffic with tramways, light rail transit systems, metros, passenger traffic with railways and air traffic as well as freight rail traffic. If there is a need for dedicated support services in the context of travelling this information shall also be handled. Some ideas among others are locations based services, weather situation, hotel accommodation, parking, park and ride, road works and event management.

Figure 1. **Metadataplatform shall support intermodal travelling**

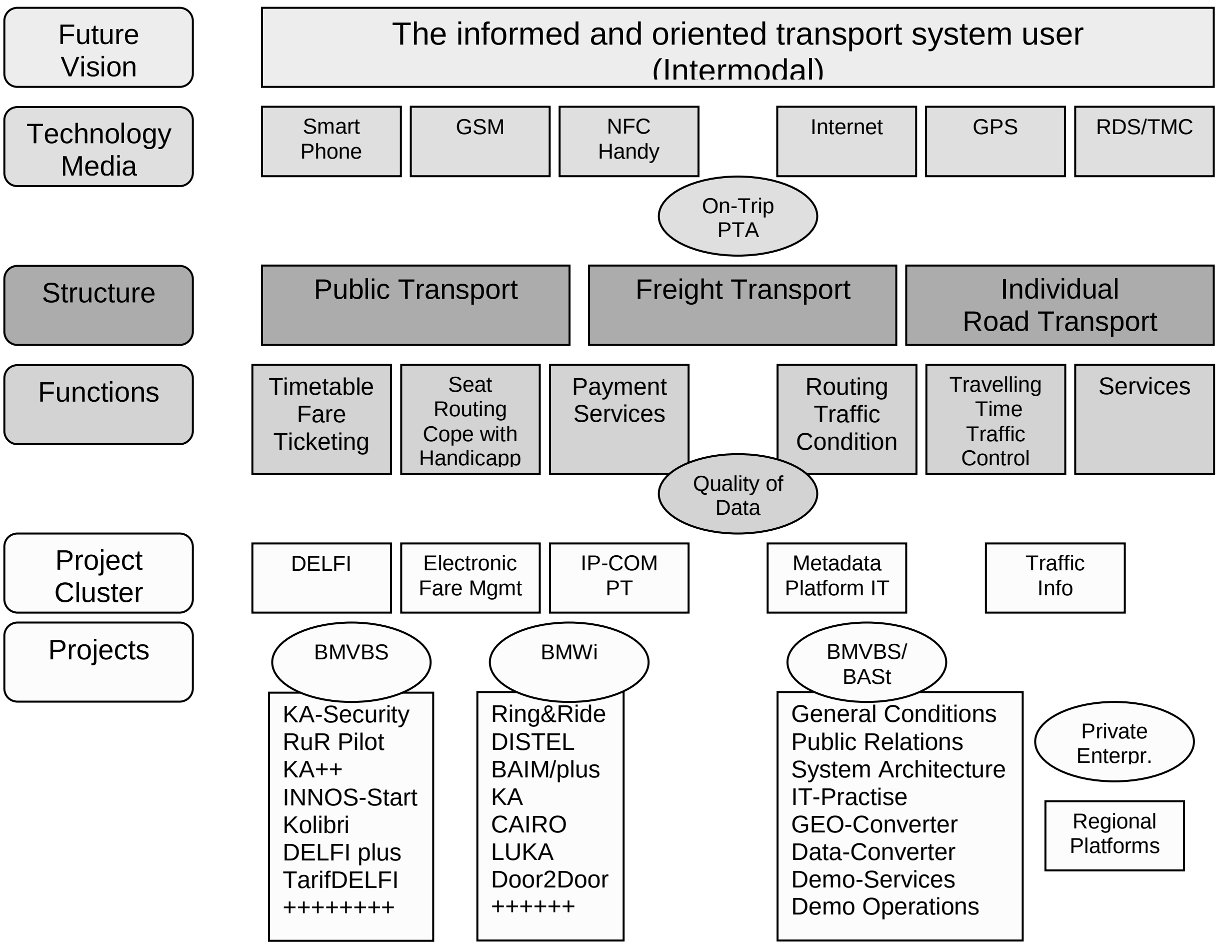


2.2. The approach

The future vision is oriented towards the transport system user who is informed with real time information and uses tools for orientation. The approach shall lower barriers which obstruct usage of all modes of transport for all sorts of people. The following chart illustrates the approach. Technologies and media shall cover present and future spectrum: ranging from mobile to stationary devices, ranging from point-to-point communication to broadcasting, ranging from pre-trip to on-trip. Structure ranges from public to freight and to individual road transport. At least the following functions shall be addressed: timetable, fares and ticketing, reservation, routing, barriers, payment services for public transport. For individual road transport dynamic and static routing, traffic conditions, travelling time, traffic control, traffic warning services are to be considered. The (up to now) missing link of quality of data along the value chain is highly important and has to be paid particular attention.

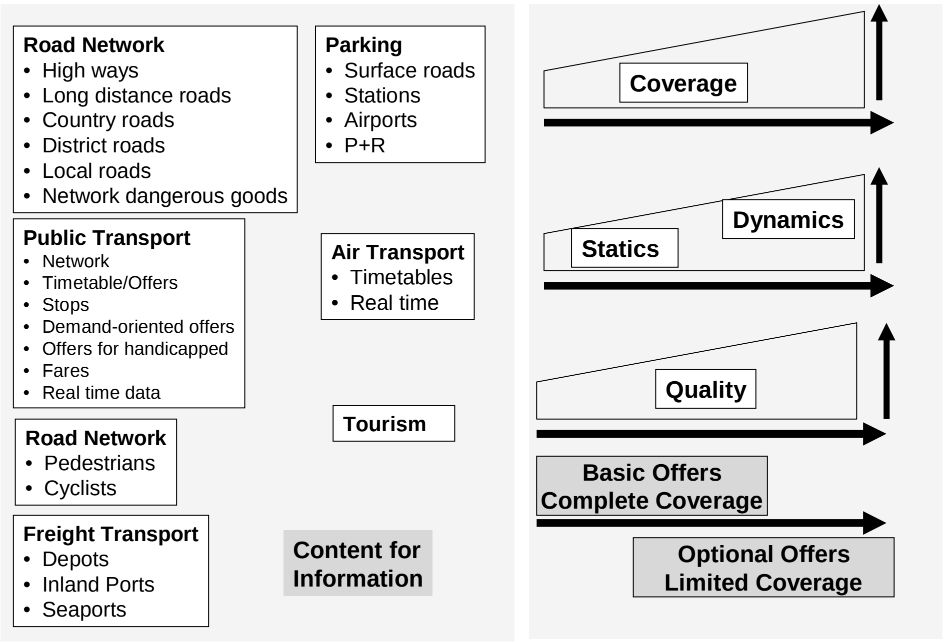
The lower part of the chart shows German project clusters/discrete projects (German project titles are not translated) which are presently in progress or planned. The responsibility of both ministries is indicated.

Figure 2. **The realisation of the vision needs technology, functions and discrete projects**



The following chart shows on the left hand side which information content will be needed. Road network also with consideration of hazardous goods. Public passenger transport from timetables and real time information, covering all present and future services. Pedestrians and cyclists shall be served. On the right hand side of the chart is demonstrated that we envisage a continuously growing depth of information and quality. The geographical coverage will grow continuously. Static information will be expanded to dynamic data. Whereas quality of data is presently almost unknown, special attention will be paid in order to make information transparent and reliable for potential customers and users. The aim for basic services is a complete geographical coverage, whereas optional services could also be geographically limited.

Figure 3. **The envisaged information content covers all aspects of transport. The variety of information will grow continuously**



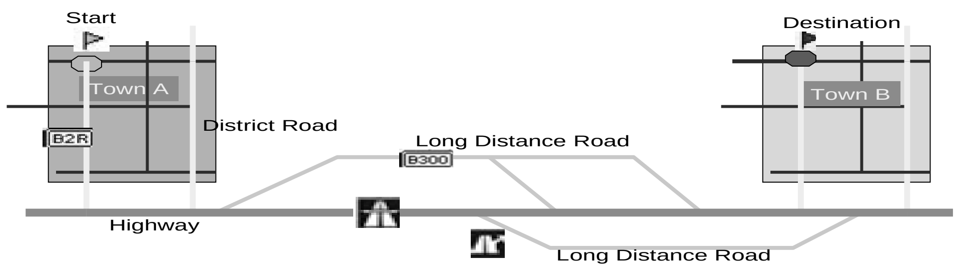
3. MISSING LINKS AND CONTRIBUTION OF METADATAPLATFORM

In order to develop a vision which is as complete as feasible, the missing links have been elaborated on the first run. Missing links are in relation to the German situation and identified separately for road and public transport.

3.1. Missing links in road transport

Let us discuss the need for improvement on the basis of a model travelling chain (see next chart). Start is in a town A with a classified road network. Access to highway is possible by district roads. Long distance roads are the network for deviations. Destination is town B with a similar road network like town A.

Figure 4. **Travelling from town A to town B**



What are the results of accidents, traffic jams and road works? The following chart describes these situations.

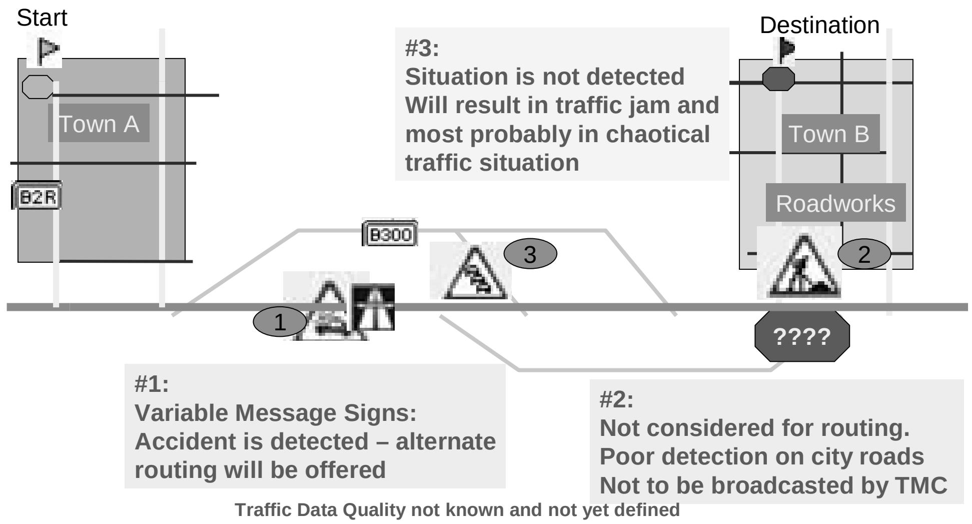
Case 1: Accident on the highway in an area with variable message signs. The accident is recorded by detectors. Variable message signs will recommend an alternative rout via long distance roads. Cars with dynamic navigation get a recommendation for an alternative route via traffic message channel (TMC).

Case 2: Road works on the road network in town B (destination). In cities the detection is poor and road works are not recorded. Therefore no alternative route will be recommended via TMC (travel message channel). The driver has to search on his own for an alternative route.

Case 3: Additional to the accident on the highway (case 1) a severe traffic jam happens. As only a few long distance roads have detection, this situation is not recorded. No alternative route is recommended. The driver is either stuck or he has to search on his own for an alternative route.

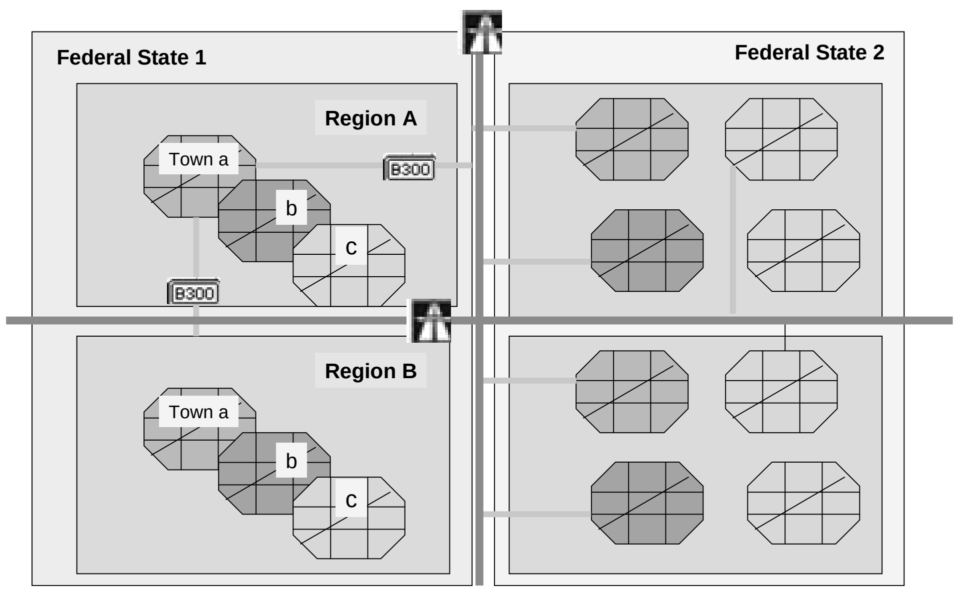
This description shows that poor detection of the road network is one of the most important missing links for managing road traffic. Of course there are new methods of detection available – like floating car data or floating phone data. However improvement will take time and is not guaranteed for all applications.

Figure 5. **Potential critical situations travelling from A to B**



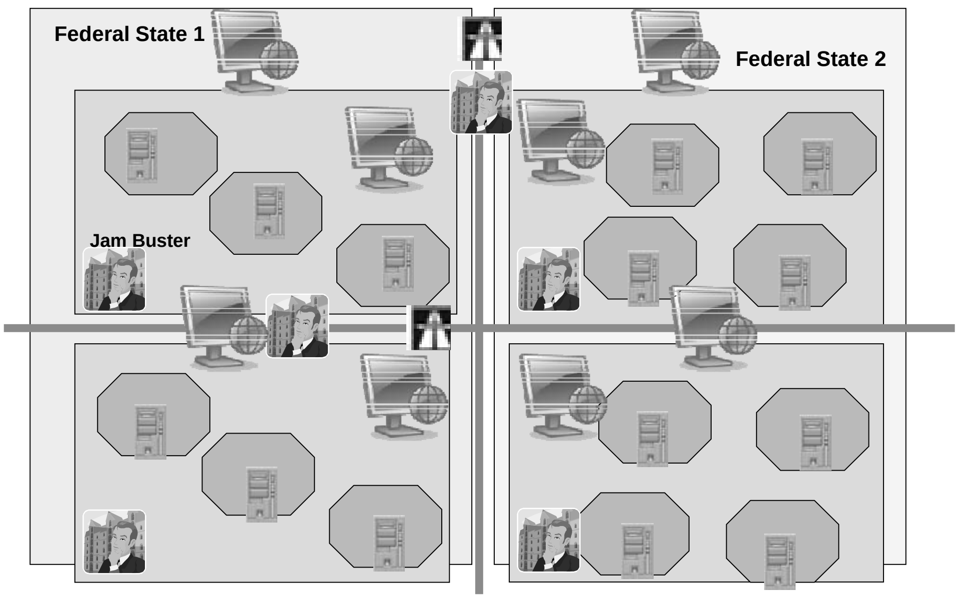
Besides insufficient detection on roads there is another weakness. The following chart shows a model case. Two federal states are connected by highways. Both states have different regions. Different cities are located in the regions.

Figure 6. **Exemplary geographical situation**



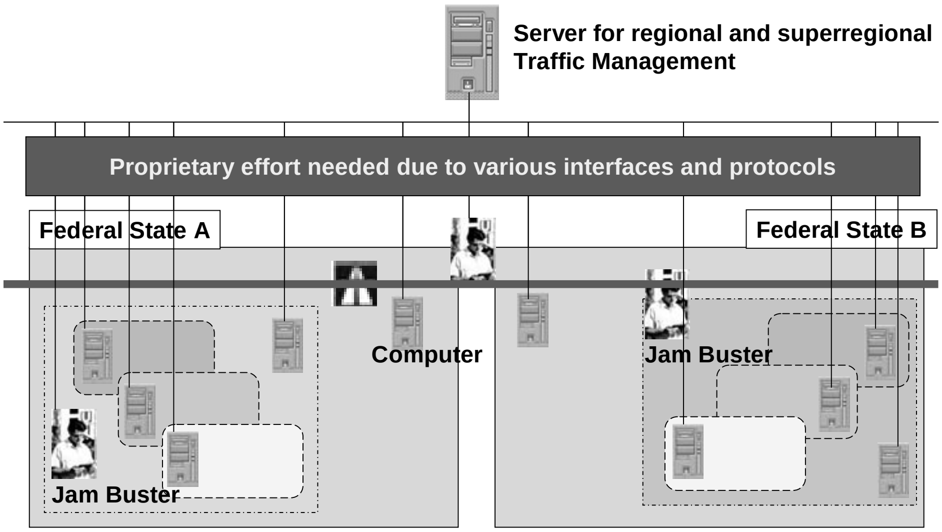
Any of these independent units (state, region and town) operates computers for traffic management. Both states operate computers along there sections of highway. If there are roadside detectors actual traffic situation is known in most cases with geographical reference. Broadcasting companies and German Automobile Association ADAC have established a network of jam busters. Traffic warning news is broadcasted via traffic message channel TMC with geographical reference.

Figure 7. **Computers for traffic management are available and jam-busters report actual traffic situations**



At first glance, it seems to be an excellent starting point. However this conclusion is wrong. If a service provider wants to establish a routing service for this model geography a considerable proprietary effort is needed due to various technical interfaces and protocols which are used by the different content owners and service providers. This is partly due to different manufacturers supplying the computers/software. Sometimes also different generations of computers and data models are used which are not compatible. In any case proprietary effort is inefficient. Meta data platform will facilitate access to different data sources.

Figure 8. **Interfaces and protocols are incompatible in most cases**



In summary, the most important missing links for an effective management of road traffic are: Detection covers only specific areas of the road network and therefore does not allow efficient and real time transport management. Different methods of georeferencing are used.

Exchange of data from different formats requires proprietary effort. Different technical interfaces and data models are used for different applications and regions. This also precludes simple access to different data sources without additional effort. Last but not least quality of traffic data is not yet defined and quite often unknown. Service providers can therefore not offer reliable and overall coverage traffic information.

3.2. Contribution of Metadataplatform Road Transport

Metadataplatform shall enable access to various data sources such as:

- Sensors/Detection
- Traffic Computers of cities
- Controls of Traffic Lights
- Regional Information Platforms
- Traffic Computers of Federal States
- Private Content Providers
- Computerized Information Systems for Road works

Metadataplatform shall enable business cases and provide central services for processing different methods of georeferencing. It shall improve models for data transfer and provide algorithms for translation of existing standards. Models and benchmarking for describing data quality will be developed and can be implemented by service and content providers.

Meta data platform is mainly focused on superregional exchange of traffic data.

3.3. Potential users of metadatenplatform road transport are:

- Service providers which offer individual services to their customers
- Administration for improvement of traffic management
- Broadcasting stations for traffic warning news
- Traffic planners for their daily field of actions
- Town planning institutions for their daily tasks
- Environmental institutions which can get access to a variety of data sources
- Freight transport and logistics for route planning and scheduling

3.4. Missing links in public transport

Door-2-Door timetable information is needed for superregional travelling. In Germany 84 out of 99 investigated regional transport entities offer door-2-door timetable information.

Up-to-date information is getting increasing importance from the customers point of view. This requires access to mobile information. Mobile services are limited to a few large transport entities. Presently fare cannot be determined automatically by internet services of the transport entities for superregional travelling. The project Tarif-DELFI shall close this missing link. For handicapped people only strictly limited service are available. This situation has to be improved considerably. Also for public transport no common definition of data quality is available. Timetable information does not give an indication for quality.

Electronic fare management is presently being introduced with model field tests. Coverage in all areas has to be reached.

Special attention has to be given to properties and user rights in public transport.

All rights belong to transport entities. Rights of regional timetables are increasingly regulated. Exchange of superregional data is quite often not regulated. However the DELFI Consortium (16 federal states plus Deutsche Bahn AG) has regulated properties and user rights. The federal states and Deutsche Bahn agree to exchange their timetable information on a regular basis. In order to improve this situation BMVBS has initiated a relevant analysis to do the investigation in more detail.

3.5. Contribution of Metadataplatform Public Transport

Metadataplatform Public Transport is and will not be a single project but a bunch of separate projects. Those projects will be outlined in chapter 5 (Next Steps). Attention will be given to the co-ordination of these projects in order to reach the vision of the informed and oriented transport system user.

3.6. Potential users of the metadatenplatform public transport are:

- Customers of public transport via service providers (including air traffic) to match with customer needs.
- Companies which operate and plan/manage public transport. This is especially true for superregional travelling or operating integrated transport systems.
- Traffic planners for their daily field of actions
- City planning institutions for their daily tasks
- All other service providers.

4. ARCHITECTURE OF METADATAPLATFORM ROAD TRAFFIC

4.1. Metadataplatform – a virtual Internet portal

Metadataplatform will represent the link between content owners and service providers. It will enable a more efficient dialogue and co-operation between both parties. As mentioned before service providers can be private and public ones.

Metadataplatform will be a virtual internet platform which offers a link list, brokerage services and establishes fixed rules for data exchange, registration, property rights, liability etc. In the course of the project business models will be developed and assessed.

Centralised services for conversion of various georeferences and data standards will be developed as well as algorithms which define quality of traffic data. This will not necessarily be integral part of the metadataplatform.

In order to use the metadataplatform following steps have to be performed:

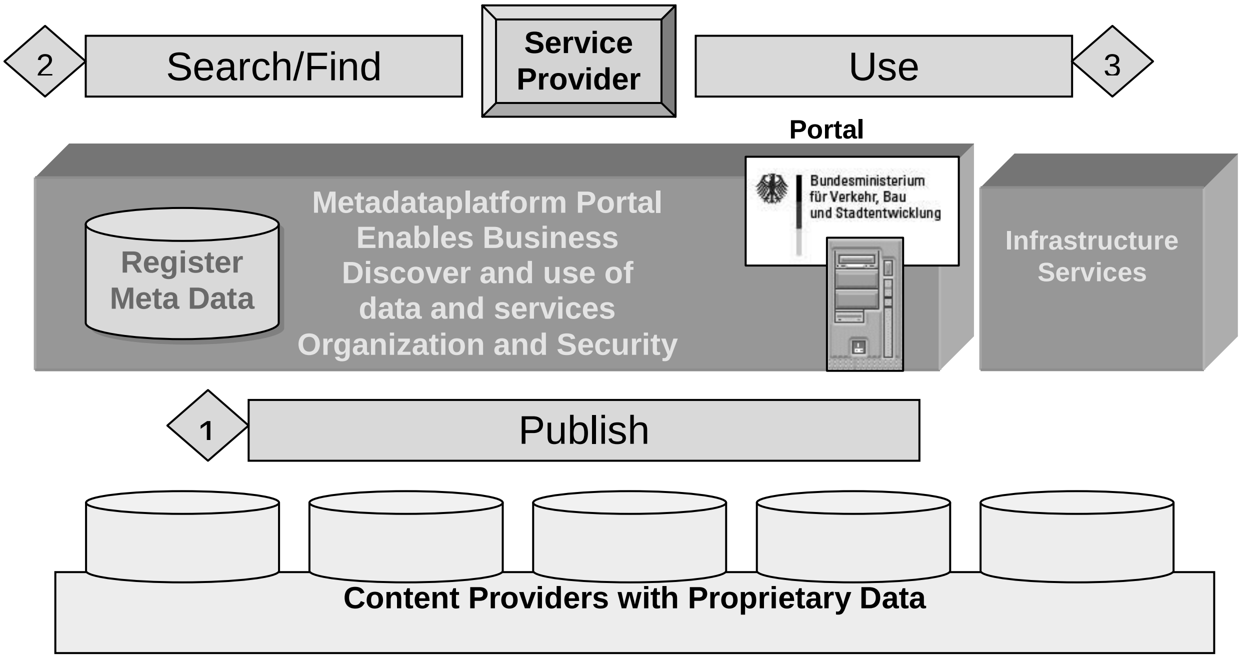
Step1: Content owners/providers register in the metadataplatform and describe (type, origin, location, periods of update etc.) their data including quality indication

Step2: Service providers which want to offer services to their customers also register and can search and find required content.

Step 3: Service provider and content provider come in direct contact with each other under the rules described above. No content provider can use the data of the content owner without accepting common rules.

The present situation (without meta data platform) would require that service providers have to negotiate and contract separately with all content owners.

Figure 9. **Metadataplatform as virtual Internet portal**

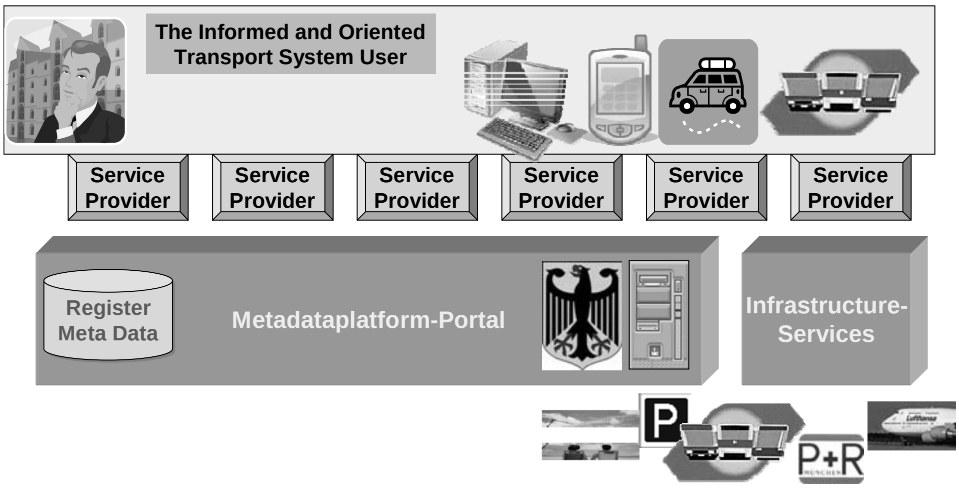


4.2. Metadataplatform enables business for content and service providers

As shown in section 4.1, metadataplatform has content providers and service providers as customers. It will make b2b-business easier and therefore has the goal of enabling new business models and optimize existing ones.

Everything shall be servicing the informed and oriented transport system user. Easier access to data sources will allow for dynamic and high quality information. If demand exists for add-on services they can also be included e.g. weather conditions (actual conditions and forecast), parking situation (location, free capacity and tendency), and park and ride capacities for public transport (location, free capacity and tendency). In case of road transport timetable information and real time situation of public transport can be offered and also timetable information and real time situation of airlines and airports may be required.

Figure 10. **Metadataplatform shall act as business enabler**



5. NEXT STEPS

5.1. Metadataplatform Road Transport

Middle of the year 2008 the federal highway institute (BASt) has taken the general project management of metadataplatform road transport. Several projects have been defined. Call for tenders have been issued and up to now partly evaluated. Some contracts have been awarded and work has been initiated. The projects are:

- Project control
- Public relations
- Proof of concept – system architecture
- Organizational and legal framework
- IT realization
- Converters for georeferencing
- Converters for data interfaces and protocols
- Pilot services for concept development proof
- Field tests

In the course of the development there might be a need for additional developments.

A project called traffic IQ (information quality) has also been started. It is financed and controlled by the ministry of economics and technology in close contact to the ministry of transport, building and urban affairs. Algorithms and benchmarking criteria will be developed for the entire value chain. Cities of Leipzig, Frankfurt, Düsseldorf and Nurnberg will be test areas. Highways of southern Bavaria and Baden-Württemberg are also included.

The overall period of development and verification is expected to be four years.

5.2. Metadataplatform Public Transport

The development projects for metadataplatform public transport will be shared among the ministries of transport, building and urban affairs (BMVBW) and economics and technology (BMWi). Key projects are e. g.:

- Further development of German timetable information system (DELFI)
- Further development of electronic fare management modules (in total 8 projects)
- Initiation of TarifDELFI (automated fare information across different transport entities via internet)
- Analysis of ownership and user rights
- Initiation of data quality modules for public transport
- Definition of a door2door-programme and initiation of developments
- Indoor routing in great interchange stations
- Initiation of internet protocol communications for public transport
- Further development of modules which make public transport more attractive for handicapped persons

The overall period of development and verification is expected to be four years.

POTENTIAL ECONOMIC IMPACTS OF TECHNOLOGICAL AND ORGANIZATIONAL INNOVATIONS IN INTERMODAL ACCESS TO MAJOR PASSENGER TERMINALS

Francisco J. TAPIADOR

Universidad de Castilla-Castilla la La Mancha

Jordi MARTÍ-HENNEBERG

Universidad de Lleida

Spain

SUMMARY

ACKNOWLEDGEMENTS

This work was partially funded by research grants CGL2006-03611 and PAI06-0101 (innovation and knowledge-transfer packages); the I3 and Ramon y Cajal programs, and the FP-005 Water, Road and Rail 2005SGR01089 project (European Science Foundation).

ABSTRACT

This report deals with the potential economic impacts of innovations such as smart ticketing and instantaneous access to rail and modal connection information schedules. First, the qualitative role of TOIs (technological and organizational innovations) is explored within the framework of intermodality. Secondly, a simple, quantitative, parametric model is described. The model is then used to analyze the impact of TOIs on rail demand, accessibility and passenger welfare under the assumption of bounded rationality. Providing that the model captures the major processes in play, the results will show the potential effects of policy choices and technological innovations both on their own and in a combined form, thus enabling discussion of their relative merits and synergies. An analysis of quantitative results shows that the effect is positive, highly non-linear, and prone to cumulative effects due to far-reaching impacts related, for instance, to the economics of climate change.

1. INTRODUCTION

The High Speed Train (HST) is steadily becoming a flexible and convenient mode compared with alternatives such as the private car, bus and plane. Swift boarding; the possibility to work and/or have meetings while in transit; and the centrality of most of the HST stations in Europe, have all helped to increase the number of business trips involving train transport in the last decade. HSTs are also used for commuting in several European contexts (notably in Spain, where services connect Madrid to Segovia, Toledo, Guadalajara and Zaragoza; and Barcelona to Lleida and Zaragoza). The captive market represented by the current growth in commuter traffic sees new users making rational choices that offset escalating property prices in central locations with cheaper living costs in satellite cities that lie within a reasonable range.

Empirical evidence shows that white collar workers and those in the advanced tertiary sector account for the majority of weekly trips (Figure 1). This market is attracted by ticket discounts for bulk purchases, flexible fares, and reliability. Another, more reduced, business market is insensitive to price. Occasional travellers preferring rail over plane favour stress-free trips to the increasing annoyances associated with air travel, and centrally located, urban rail stations to peripheral terminals that are often a long way outside city centres. Considering all these factors, intermodality has a definite influence on leaning frequent users towards the train and limiting car use to what has popularly been labelled as “the last mile”: the connecting trip from the last public transport mode to either home or work. Improved intermodality is widely seen as one of the major factors that can be used to promote widespread public transportation in Europe.

It is generally accepted that the success of the intermodal model largely depends on whether or not public transport is perceived as efficient and on how seamless the modal shift can be made (UITP, 2003). Within this framework, technological and organizational innovations (TOIs) may have a profound impact on intermodality. Newly-available technologies, including high-tech phones with internet access, combined with real time information on timetables and the possibility to make remote purchases of tickets at the last minute, reduce impedances in the rail business. Thus, TOIs increase efficiency on both the rail operator and user sides.

2. INTERMODALITY AND ACCESSIBILITY IN EUROPE

Improved intermodality is one of the cornerstones of a sustainable transport policy. One of the reasons for the widespread use of private cars throughout Europe is their ability to provide door-to-door transport despite problems associated with traffic congestion and the lack of parking spaces in most urban regions. Diseconomies associated with the use of private cars include: injuries and death due to road accidents; unproductive travel time due to accidents and traffic congestion; a dependence on non-renewable sources of energy; and damage and other negative effects associated with environmental pollution (Jakob *et al*., 2006). One way to palliate these effects would be to promote hybrid or electric cars. Another strategy would involve promoting a modal shift from the use of private cars to public transport. The basic idea would be to persuade travellers to only use cars on trips between their homes and public transport, instead of driving all the way to their final destination. There is growing recognition of the fact that sustainable mobility implies inter-connecting transport systems that must provide a door-to-door service (European Commission, 1999). In this respect, the intelligent planning of intermodality offers a means of increasing the sustainability of interurban passenger transport systems: the better that these resources can be combined and co-ordinated in an integrated manner, the greater the sustainability of the whole transportation system (European Commission, 2001).

The main nodal points in the intermodal networks of present day Europe are the European high speed train stations (HSTS). While the impedances in the rail network itself are related to environmental or physical constraints, such as slopes and the volume of rail traffic, and are difficult to overcome, friction resulting from the suboptimal intermodality of high speed train stations has much more of a planning component. It has already been shown that there is a clear hierarchy of stations with status being linked to their respective roles within the regional system and with strong constraints that prevent some stations from performing optimally and as truly intermodal nodes (Tapiador *et al*., 2009). In this context, TOIs may help to smooth out passenger flows.

An in-depth study that was carried out by the Task Force of the Transport Intermodality group highlighted modal imbalance in the EU transport system and identified obstacles that prevent the development of user-oriented door-to-door intermodal transport services. In that work, transfer point efficiency and the efficiency of intermodal networks were identified as two of six areas of major interest for advancing research into intermodality. A lack of information and the impossibility to investigate the way in which some services were organised were amongst other relevant factors. Alternative methods and tools for assessing potential modal shifts have been described by Tsamboulas *et al.* (2007). These include complete policy action plans that could be useful for decision makers.

Regarding accessibility, there is a clear connection between improved intermodality and increased accessibility. Accessibility is defined here as the ease with which an individual can reach or access a specific place, infrastructure, amenity, or job opportunity, or generally to participate in activities. The more accessible the activity is, the fewer travel barriers and less travel friction need to be overcome to reach or access it. This term is also used to specifically refer to the ease with which the disabled can use transit or transportation facilities. The difference between the two meanings lies in the fact that what can be generally seen as a cause of friction within the system (for example, a staircase at a two-level exchange) may represent a barrier for disabled people (if there is no lift available).

Accessibility is of great economic and social significance in the field of transport economics and policy and this has been recognised by the European Spatial Development Perspective (European Commission, 1999), which states that improving the accessibility of Europe's regions is considered necessary for improving their competitive position and also the competitiveness of Europe as a whole. Accessibility influences the advantage of one location over others. For the USA, Kuby *et al.* (2004) examined the importance of accessibility (among other factors) in terms of light-rail station boardings, which they found to be significant. Estimates of accessibility have therefore been used to assess the advantages that households and firms derive from the existence, and use, of local transport infrastructure. It is supposed that areas with better access to points supplying input materials and offering markets will, *ceteris paribus*, be more productive, more competitive and more successful than those whose locations are more remote (Spiekermann, 2005).

3. PASSENGER PROFILING

Modelling the effects of TOIs in HSTS requires an indication of the composition and behaviour of the users. Passenger profiling from passenger surveys, such as that described by Burckhart et al. (2008) for the Madrid-Barcelona line, is a useful way of feeding a parametric model with empirical information for case studies (Figure 1). The modal share offers an important way to quantify how TOIs may affect travel. For instance, underground and conventional rail users have less need for real time information as they can rely on stable timetables and generally have established habits and routines; but private car, bus and taxi users may prefer rail to other alternatives if timetables, ticketing and access information is promptly available anywhere and at any time. The relative proportion of each mode depends on the station in question (Figure 2) and this constitutes an obstacle to proposing any kind of comprehensive quantitative model that would be valid for every location. Instead, the model has to be of the parametric type and allow adaptation and the incorporation of up-to-date data when this is available.

The reason people travel is also relevant when constructing a model. In the Burckhart *et al*. study (2008), most of the trips were work-related with a predominance of professional business trips (Figure 3). The modal break-up, such as that shown in Table 1, is a key input if we are to derive results that will be useful for planning because the effects of TOIs are modulated by cross-relationships between the transport mode in question and the reason for travelling.

4. FACTORS AFFECTED BY TECHNOLOGICAL AND ORGANISATIONAL INNOVATIONS (TOI)

TOIs have both direct and induced effects on rail transport welfare. Direct effects refer to those that have a simple functional relationship with TOIs. The function itself can be either linear or non-linear. Induced effects are those motivated by other variables and/or those that have resulted from the internal dynamics of the model.

4.1 Direct effects

4.1.1 Increased intermodality

Intermodality is defined by the EC (2004) as "a characteristic of a transport system that allows at least two different modes to be used in an integrated manner in a door-to-door transport chain". TOIs favour such seamless journeys by reducing transit times and associated uncertainties. Precise information on parking space and/or remote bookings of crowded car parks, the ability to reschedule trips combining several modes (if for instance a meeting ends sooner than expected), and new tools to cope with delays due to traffic jams, all help users to organize their travel both from and to the HSTS.

4.1.2 Policy priorities

TOIs permit access to deals devised to fill demand gaps. Intelligent pricing, targeting specific markets (last-minute or early-bird travellers; pensioners or students); time slots (late trains) or seasons (summer doldrums) are now remotely accessible for a range of potential customers.

4.1.3 Timetable and ticketing information

Instant access to timetable information relating to several different modes helps to match travel plans. Price information favours consumers making rational choices on trip mode and scheduling. Flexible fares and a sensible and user-friendly interface reduce impedances in the purchasing process. On-line ticket purchases and smart ticketing for public transportation increase both the number of transactions made and passenger welfare. An example of good practice is provided by the Swiss system, where timetables are sensibly matched to minimize dead time and, at the same time, ensuring modal connections. Thus, for instance, transitions between rail and postal buses are coordinated so that passengers can reach remote locations without excessive waiting.

4.1.4 Modal connection and accessibility

Regarding organizational innovations, it is important to provide a smooth modal transition. Apart from obvious measures such as ensuring full accessibility for every passenger, connections should be clearly indicated to avoid confusion. This is also applicable to on-line interfaces.

4.1.5 Greenhouse gas (GHG) emissions

Increased passenger traffic in HSTS directly reduces road congestion and carbon emissions. For comparison, the respective per capita CO2 emissions associated with a 100 km trip are: about 13 kg for a small car; 26 kg for a large car, and 6 kg for a rail trip. In terms of international carbon emissions, travelling by rail offers substantial savings in carbon emissions. The current price per tonne of CO2 is about € 12.

4.2. Induced effects

4.2.1 Station carrying capacity

The carrying capacity of a station is increased if waiting times are reduced, as an increased passenger flux permits more clients to use the same space at different times. Optimal passenger use in an HSTS is achieved when passengers can easily change modes without either delays or rushing, and can also make economic transactions in the (short) time between transfers. This avoids crowding, discomfort, stress, and risks and helps to create a perception of rail travel as a pleasant experience.

4.2.2 Average stay

Increasing the time spent at the HSTS reduces both perceived quality of life and productivity. The potential effect on shop sales, and thus on rents is not linear: whereas a certain amount of spare time spent at the station makes some travellers buy goods, behaviour is parabolic after a certain threshold time (which varies according to the HSTS). This effect adds to the discomfort of a long wait and increases the tendency for passengers to avoid the station in question in the future. The sharing and dissemination of passengers' negative impressions also generates diseconomies. It is well known that some HSTS are perceived as comfortable and friendly, while others are regarded as uncomfortable and confusing, etc. Being located 'in the middle of nowhere' or at peripheral locations plus presenting an infuriating lack of information on connections or travel alternatives creates a very poor impression of the intermodality of some European HSTS.

5. QUANTITATIVE INSIGHT: A PARAMETRIC MODEL

To gain an insight into how TOIs may help the economics of rail transport, it is useful to construct a quantitative model that takes into account the factors presented above. Whereas other approaches, such as Data Envelopment Analysis (DEA), have been applied to transportation modelling (e.g. Tapiador et al. 2008), most of the techniques referred to in recently published literature require empirical data, which are not readily available in this case.

The model described here is dynamic and simulates the structure of the problem in a schematic way so the complexities of the system do not render the problem impossible to analyse. The aim of the model is to simulate –rather than predict- the effects of changes in the different parameters. Models used to perform such sensitivity analyses have proved useful in several other fields, including climate change. As this model does not include empirical data, it is called a parameterized model. The results are projections under a prescribed scenario and the conclusions must be understood as estimates of the potential effects of changes in the parameters.

Figure 4 illustrates the different variables and relationships. Behind this graphical layout lies the mathematical modelling of the problem. The model assumes the existence of a captive market (commuters) and a new market yet to be attracted. TOIs affect both markets and their effects are modulated by independent policy priorities. These may include strategic decisions taken outside the rail business, such as those serving potential corporate interests in joint flight-rail ventures. The modal split is considered a social feature and is therefore an independent variable in this model. It directly affects the new market by providing new users and also affects greenhouse gas (GHG) emissions by helping to take cars off the road.

The analysis used in the model is in the form of time-varying coupled differential equations. The model is run for full annual periods with slightly-varying initial conditions resulting in a large ensemble of trajectories. This procedure is deemed to account for sensitivity to initial conditions in dynamical systems. The resulting ensemble is then averaged to provide the mean behaviour of the system, which is the variable used to extract policy conclusions. The spread of the ensemble members is comparatively smaller than the internal variability of the model.

The variables in the model are related through a variety of linear and non-linear functions. The actual shape of every function is derived from observations, for some cases, and from hypothesis, for those cases for which no empirical evidence can be easily extracted. Thus for instance, the new market variable is modelled as a function of technological innovations linear function with support in [0,1], and modal spilt linear function also within the [0,1] domain; both modulated by a seasonal pattern function. Other variables such as overcrowding effects are considered as non-linear, and modelled as such. Thus, a normalized sigmoid function is used for agglomeration diseconomies as it is assumed that after a threshold the negative effects stabilize.

The accessibility variables used in the mode are as follows. The *station capacity* variable encapsulates accessibility and intermodality variables such as intermodal entropy and intermodal integral time (Tapiador et al. 2009). *Modal split* is considered as a separate effect as it is affected by demand fluctuations. Regarding TOIs, timetable and ticketing information and possibilities for on-line purchasing are normalized in the model. These factors affect both the passenger market and HSTS operations by reducing confusion and crowding (Lam et al. 1999). This variable also depends on the carrying capacity of the HSTS in question and on the average stay, which is also dependent on TOIs. Reducing the average length of stay is deemed to slightly reduce passenger spending at the HSTS. This is, however, a simplification, as businesses would tend to react to shorter stays by adapting their offer. Even so, the overall effect would be relatively small within the scope of the inter-annual modelling carried out.

Seasonal patterns in new and captive markets are also considered to model holidays and working days without lunch breaks, which are characteristic of the summer routines of Spanish state employees. Split shifts modify the behaviour of the model, but only have a limited effect on daily aggregations. The weekly pattern is explicitly accounted for in the model by including a stochastic component. The seasonal pattern is modelled as piece-wise.

Even models as simple as the one described provide a wealth of information in the form of sensibility graphs, scenarios and possible parameterizations of the variables involved. The results must be analysed with an eye to the problem in hand. The crux of modelling is to achieve a delicate balance between the beneficial effects of attracting more passengers and the potential diseconomies associated with doing so.

Among many other possible effects that can be explored is the coupling between phase-shift cycles and TOIs, and the probable saturation of the market. Applying this analysis to real cases would require a precise knowledge of the time-evolution of the variables involved. This could be achieved by surveys or by undertaking dedicated studies at key HSTS.

6. RESULTS AND CONCLUSIONS

Given the stochastic component of weekly and seasonal cycles, ensemble simulations were carried out to characterize the mean behaviour of the system. The rationale of ensemble techniques is to palliate sensitivity to initial-conditions (SIC) by running a model under numerous slightly-different initial conditions. Since nonlinear dynamical systems are highly sensitive to these conditions, the runs will provide a set of different forecasts, no matter how close the initial conditions are to each other (Tapiador and Gallardo 2006). The results of the simulations then exhibit the effect of TOIs for a variety of likely cycles, with the mean values embedding the actual dynamics. Sensitivity analyses using the Jacobian of the (linearized) model can then be used to trace back the effects of every choice.

Runs were performed for several combinations of potential use of TOIs. Considering the input data, the results can be regarded as the application of a set of policies aimed at increasing passenger comfort and welfare in a canonical case. Empirical evidence shows that differences in accessibility and HST demand are highly correlated. Those stations with low accessibility and intermodality are those less used, whereas well-connected stations present high passenger traffic. The relationship, however, is not simple as the physical size of some HSTS was designed considering expected traffic. In the model presented here, accessibility, intermodality and the size of TOIs impacts in relation to the generalised cost of travel is highly dependent on the assumptions made, but some general conclusions can be derived. By elaborating the results in qualitative, policy terms, several issues arise.

The implementation cost of TIs is several orders of magnitude lower than the costs in infrastructure and in improving accessibility and intermodality. Passengers using TIs are by definition connected, and act as nodal points of innovations. Organizational innovations are also comparatively cheaper, and their effects multiplicative. Thus, and recognising that the model presented here cannot provide a quantitative estimate of such impact, the benefits of TOIs for rail travel are non-linear. One euro inverted in TOIs (including the effect of advertising) is likely to produce a larger effect on demand than the same euro put into other branches of the business. Figure 5 shows the result of simulation for a year after TOIs are implemented. It is observed a slightly-exponential growth of passenger traffic, which is indicative of accumulative effects.

Everything else the same, the net effect of an increase or improvement in the TOIs is positive for welfare. The effects on the modal shift show a cumulative effect on the economics of climate change. Synergies appear when satisfied customers spread the benefits of rail travel. Non-linearities within the model yield reinforcing feedbacks that suggest that one of the most efficient actions would be to favour the modal split. This can be achieved by a number of actions, including: providing a free Kiss&Ride drive through; reducing or removing parking fees for commuters; and setting special fares for business trips (such a combined park and rail ticket). Another measure would be to allocate free, dedicated parking spaces for electric cars, and to set up solar-powered charging bays for them. This organizational innovation would help to further reduce carbon emissions, improve intermodality, and reduce road congestion.

Effects not considered by the model include measures that would benefit urban economies. It is dysfunctional to have empty parking spaces at rail stations alongside commuters' cars packing the neighbouring streets. Such action would only transfer costs to municipal authorities and neighbours, with the associated risks and increased insurance claims. Reducing congestion in the areas around HSRS improves their centrality and makes it possible to control their gentrification. The benefits of TOIs in reducing passenger stress and discomfort, in promoting new values (balanced family/work time), and in increasing personal safety and security (no money involved in phone transactions, nine times lower accident risk associated with travelling by rail as opposed to by car, etc.) shows the importance of TOIs for the future of interurban passenger transport.

Figure and Table captions

Figure 1. Modal split at the Madrid-Atocha HSTS (data from Menéndez *et al.*, 2006)

Figure 2. Modal split at the HSTS analyzed in Burckhart *et al.*, 2008

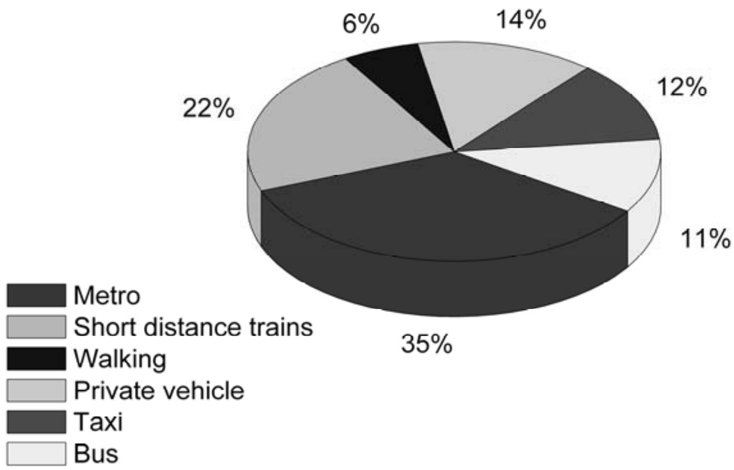
Figure 3. Access times to the HSTS analyzed in Burckhart *et al.*, 2008

Figure 4. Conceptual view of the quantitative model

Figure 5. Estimated evolution of welfare in a prescribed scenario. The x-axis represents the day of the year after implementing TOIs; the y-axis indicates the evolution of passenger traffic (arbitrary units).

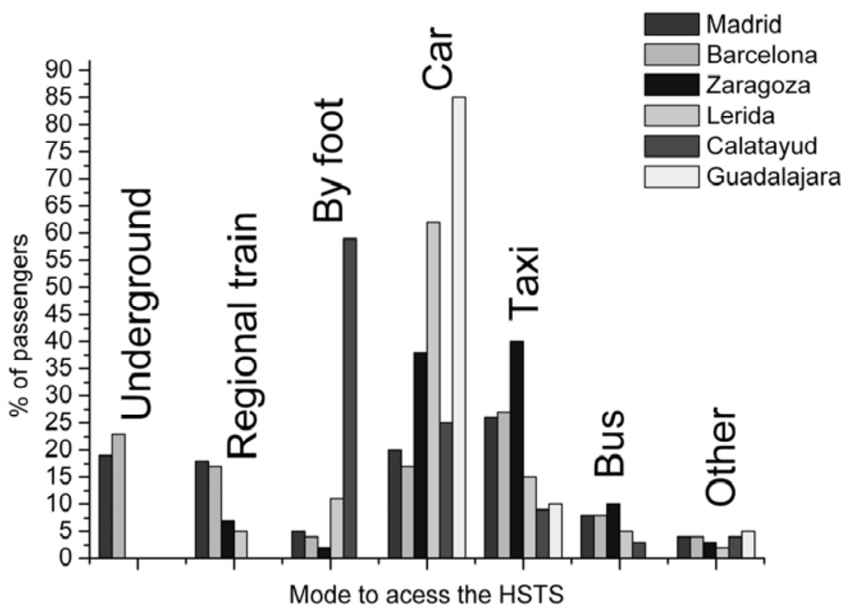
Table 1. Cross relationships between mode and reason for travelling (expressed as a %, from Burckhart *et al.*, 2008)

Figure 1. **Modal split at the Madrid-Atocha HSTS**



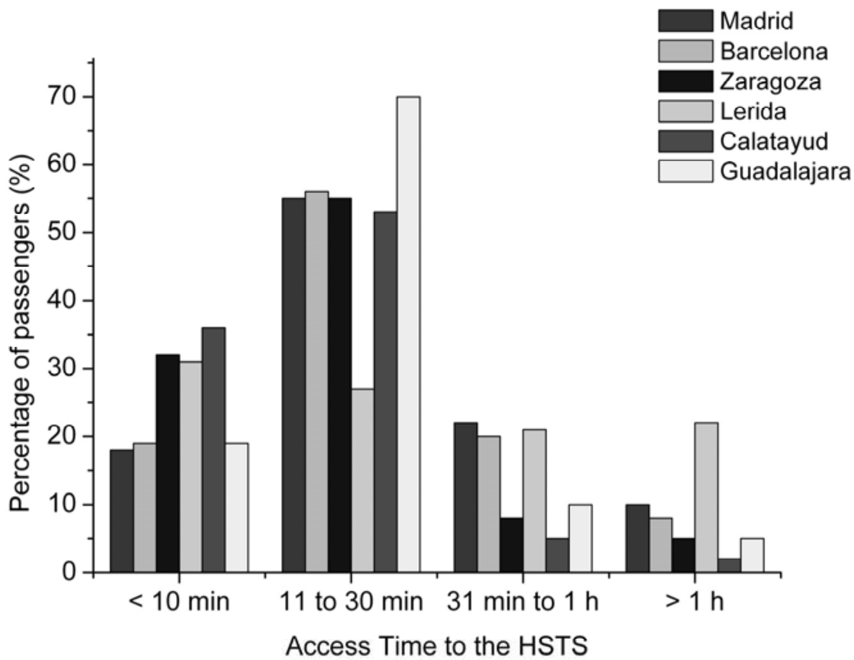
Source: Source: Menéndez *et al.* 2006.

Figure 2. **Modal split at the HSTS**



Source*:* Analyzed in Burckhart *et al.* 2008.

Figure 3. **Access times to the HSTS**



***Source*:** Analyzed in Burckhart *et al.*, 2008..

Figure 4. **Conceptual view of the quantitative model**

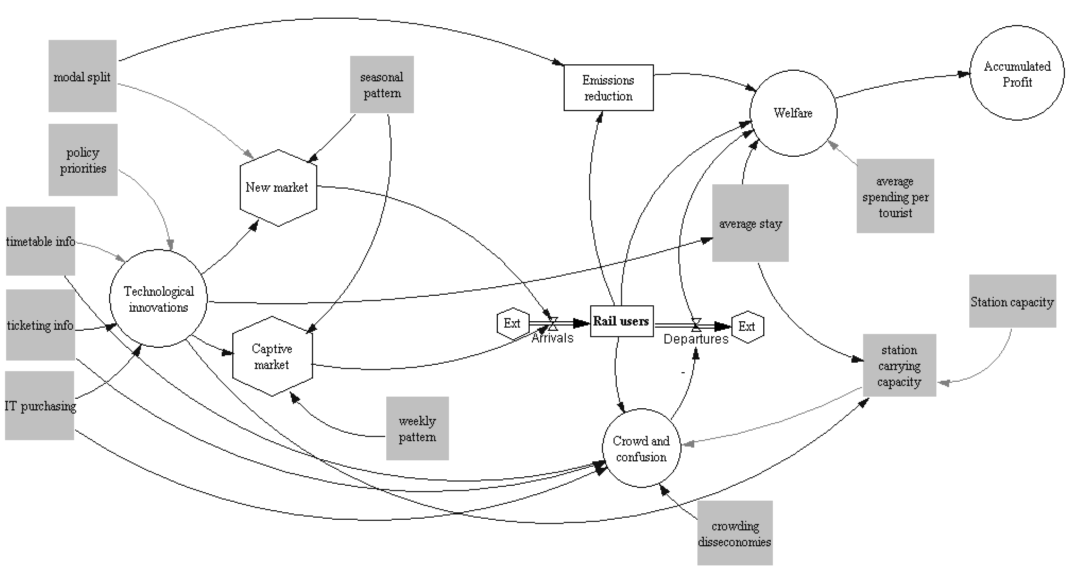


Figure 5. **Estimated evolution of the welfare in a prescribed scenario**
The x-axis represents the day of the year;
the y-axis indicates the evolution of passenger traffic (arbitrary units).

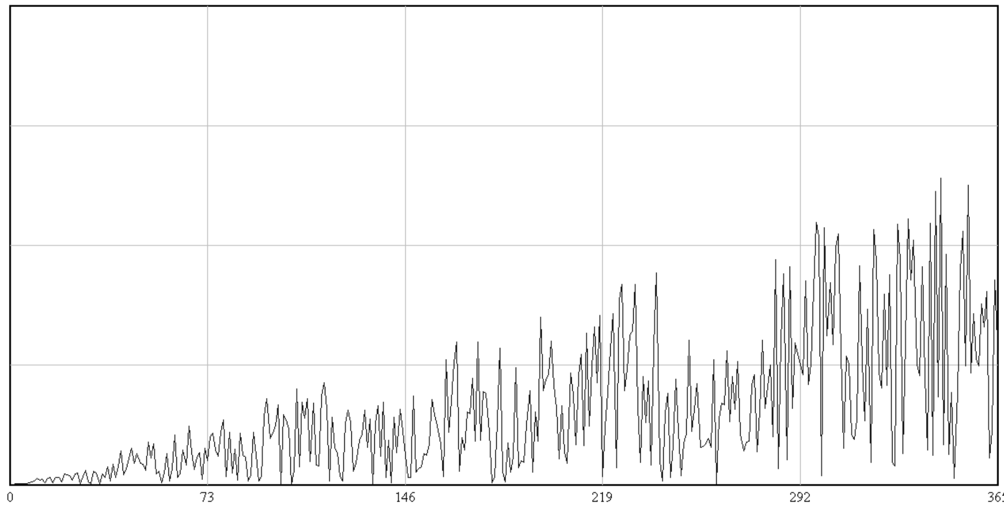


Table 1. **Cross-relationships between mode and reason for travelling**
(expressed as a %)

	Trip reason							
	Work	**Business**	**Tourism**	**Other**	**Family-related**	**Educat.**	**Health**	**Sum**
Car	7.1	15.7	4.0	1.0	5.3	0.9	0.4	34.5
Taxi	6.3	12.3	3.8	0.8	4.5	0.4	0.3	28.5
Bus	0.9	1.7	1.1	0.3	2.0	0.2	0.2	6.4
Other	0.7	1.6	0.5	0.2	0.4	0.1	0.0	3.5
Reg. Train	2.0	2.7	2.0	0.6	3.0	0.2	0.1	10.5
Underground	1.3	2.8	1.7	0.5	2.2	0.3	0.1	8.9
Walking	2.0	2.8	1.0	0.2	1.4	0.2	0.1	7.7
Sum	20.3	39.6	14.1	3.5	18.9	2.4	1.2	100

Source: Burckhart *et al.*, 2008.

BIBLIOGRAPHY

Burckhart, K., J. Martí-Henneberg and F.J. Tapiador (2008), Cambio de hábitos y transformaciones territoriales en los corredores de alta velocidad ferroviaria. Resultados de una encuesta de viajeros en la línea Madrid-Barcelona. *Scripta Nova*, Vol. XII, 270.

European Commission (1999), Task Force ‘‘Transport Intermodality”. <http://cordis.europa.eu/transport/src/taskforce/home.html>

European Commission (2001), COST 340, Towards a European Intermodal Transport Network: Lessons from History. <http://www.cordis.lu/cost-transport/src/cost-340.htm>

European Commission (2004), *Towards passenger intermodality in the EU*. Report 1, Dortmund.

Jakob, A., J.L. Craig and G. Fisher (2006), Transport cost analysis: a case study of the total costs of private and public transport in Auckland, *Environmental Science & Policy*, 9, 55–66.

Kuby, M., A. Barranda and C. Upchurch (2004), Factors influencing light-rail station boardings in the United States, *Transportation Research Part A: Policy and Practice*, 38 (3), 223–247.

Menéndez, J.M., J.M. Coronado, B. Guirao, F.J. Rodríguez, C. Ribalaygua, A. Rivas and J.M. Ureña, (2006), *Diseño, dimensión óptima y emplazamiento de estaciones de alta velocidad en ciudades de tamaño pequeño*. Cuadernos de Ingeniería y Territorio, 7.

Lam, W.H.K., C.-Y. Cheung and C.F. Lam (1999), A study of crowding effects at the Hong Kong light rail transit stations, *Transportation Research Part A: Policy and Practice*, 33 (5), 401-415.

Spiekermann, K. (2005), Transport and accessibility in Europe. Interreg North Sea Programme & North Sea Commission, in: Joint Annual Conference, Vlissingen, 15 June.

Tapiador, F.J., A. Mateos and J. Martí-Henneberg (2008), The Geographical Efficiency of Spain’s Regional Airports: a Quantitative Analysis, *Journal of Air Transport Management*, 14, 205-212.

Tapiador, F.J. and C. Gallardo (2006), Entropy-Based Member Selection in a GCM Ensemble Forecasting, *Geophysical Research Letters*, Vol. 33, L02804.

Tapiador, F.J., K. Burckhart and J. Martí-Henneberg (2009), Characterizing European High-Speed Train Stations using Intermodal Time and Entropy Metrics, *Transportation Research Part A: Policy and Practice*, 43, 197-208.

Tsamboulas, D., H. Vrenken and A.M. Lekka (2007), Assessment of a transport policy potential for intermodal mode shift on a European scale. *Transportation Research Part A: Policy and Practice* 41 (8), 715-733.

UITP (1994), *Location and design of interchanges – rail, bus and car*, Brussels.

UITP (2003), *A one-stop approach to mobility: the challenge of integration*, Brussels.

Vägverket (2004), *Trafik för en attraktiv stad*, Stockholm, p. 212.

Theme V:

Sustainable Interurban Mobility

ENVIRONMENTAL ASPECTS OF INTER-CITY PASSENGER TRANSPORT

Per KAGESON

Nature Associates
STOCKHOLM

Sweden

SUMMARY

1. INTRODUCTION

Many governments in different parts of the world are investing in high speed rail. Some of them do so thinking that it will be an important part of climate change mitigation. Intercity traffic over medium distances is particularly interesting in the environmental context as it constitutes the only transport segment where aircraft, trains, coaches and cars naturally compete for market shares.

This report calculates the effect on emissions from building a new high speed link that connects two major cities located 500 km apart. It assumes that emissions from new vehicles and aircraft in 2025 can be used as a proxy for the emissions during a 50 year investment depreciation period. The emissions from the marginal production of electricity, used by rail and electric vehicles, are estimated to amount on average to 530 gram per kWh for the entire period. Fuels used by road vehicles are assumed to be on average 80 percent fossil and 20 per cent renewable (with a 65% carbon efficiency in the latter case).

Traffic on the new line after a few years is assumed to consist to 20 per cent of journeys diverted from aviation, 20 per cent diverted from cars, 5 per cent from long-distance coaches, and 30 per cent from pre-existing trains. The remaining 25 per cent is new generated traffic. Under these assumptions would the investment result in a net reduction of CO_2-emissions of about 9,000 tons per one million one-way trips. Assuming 10 million single journeys per year, the total reduction would be 90,000 tons.

When the price of CO_2 is $40 per ton, the socio-economic benefit of the reduction would amount to $3.6 million, which is very little in the context of high speed rail. The sensitivity analysis shows that alternative assumptions do not significantly change the outcome. One may also have to consider the impact on climate change from building the new line. Construction emissions for a line of this length may amount to several million tons of CO_2.

There is no cause to prohibit investment in high speed rail on environmental grounds so long as the carbon gains made in traffic balances the emissions caused during construction. However, marketing high speed rail as a part of the solution to climate change is clearly wrong. Investment in infrastructure for modal shift should only be considered when traffic volumes are high enough to carry the cost. The principal benefits of high speed rail are time savings, additional capacity and generated traffic, not a reduction of greenhouse gases.

2. BACKGROUND

Intercity passenger transport is growing fast to meet demand for mobility from private citizens and the business community. A shift to fast modes of transport makes it possible to travel longer annual distances within restrained time budgets. Aviation and high speed trains are the fastest among modes. Although high speed comes at the price of negative environmental impact, many environmentalists claim, along with the companies and interest organizations of the rail sector, that high speed trains are environmentally benign and should be allowed to form an important part of climate change mitigation.

De Rus and Nash (2007 take another view; “Decisions to invest in high speed rail have not always been based on sound economic analysis. A mix of arguments, besides time savings – strategic considerations, environmental effects, regional development and so forth – have often been used with inadequate evidence to support them.”

Intercity traffic over distances between 400 and 600 km is particularly interesting in the environmental context as it constitutes the only transport segment where aircraft, trains, coaches and cars naturally compete for market shares. Among the parameters that influence modal choice are price, travel time, frequency, comfort and personal safety. Environmental considerations may also play a role although rather few appear to be willing to make any larger sacrifice in terms of cost in order to contribute to a better environment.

The objective of this paper is to analyze whether the difference in environmental impact between passenger transport modes is large enough to justify government investment in modal shift. As investment in new infrastructure usually has to be written-off over 40 to 60 years, the perspective in this paper is long-term. Limiting the analyses to current differences in environmental impact between cars, buses, trains and aircraft would clearly be wrong.

A distinction is made in this paper between fast passenger trains and high speed trains. The former travel at a maximum speed of 150-200 km/h, while the latter are capable of top speeds in excess of 250 km/h. Average speeds, however, may be lower due to track restraints.

3. INTERMODAL COMPETITION

The author assumes that few people find it acceptable to travel between cities located 400 to 600 kilometers from each other at average speeds below 90-100 km/h when infrastructure that allows such speeds or more exists. Therefore the potential environmental benefits from travelling at average speeds below 100 km/h are disregarded in this paper.

The willingness-to-pay for high speed varies among people and is closely connected to income (or having someone else pay ones bill). The fact that some customers prefer low-speed intercity trains to fast trains or high-speed trains is a sign of this. This means that investment in rail for high-speed trains will have only an insignificant effect, if any, on those customers who currently prefer local trains (with many stops) to existing fast trains. Presumably the effect on those who today prefer to travel by car will also be small, although perhaps not insignificant. The reasons for taking the time to travel by car may be reduced cost (e.g. several people travelling in one car) or that the traveler needs a car on his arrival to the destination.

Table 1 shows the time that it takes to travel from city center to city center by different modes. It is assumed that the air passenger on average has to devote a total of 70 minutes on ground level connections to and from airports and needs to check in 30 minutes earlier than a train passenger. Passenger who check-in luggage may need an additional 10 minutes. Aircrafts are assumed to spend 10 minutes on waiting and taxing. People travelling by car are assumed to need a 30 minutes break for a fast meal when the distance is 600 km.

From the table it is evident that conventional trains can compete with air traffic up to a distance of a little less 400 km. However, at 500 and 600 km it takes a high speed train to beat aviation. The fact that some people nevertheless choose to travel by aviation may be explained by several factors among them frequency of connections, price and personal preferences.

Table 1. **Travel time between city centers (point-to-point) by different modes of passenger transport, accounting for access and waiting times, and for cars at 600 km stopping for a short meal**

Mode	Average speed, km/h	Distance city center to city center		
		400 km	500 km	600 km
Passenger car	100	4:00	5:00	6:30
Coach	85	4:43	5:53	7:34
Fast train	150	2:40	3:20	4:00
High speed train	280	1:26	1:47	2:09
Aircraft	800	2:20	2:28	2:35

As most central stations are located down-town, an advantage of travelling by rail is that the journey takes you from city center to city center. However, all passengers to not have a down-town location as their point of departure or a final destination that is located in or close to a city center. For them the total travel-time may be shorter by a combination of aviation and a rental car or a taxi. A shift from fast to high-speed train may or may not make a difference for this type of customer. The difference in travel time is small already when travelling from city center to city center.

4. ENVIRONMENTAL ASPECTS

Transport affects the environment in numerous ways. The most important parameters in the context of intercity passenger traffic are exhaust emissions (NO_x, SO_x, PM and VOC), noise, and climate change. Land-use, including intrusion and barrier effects, may also be important.

Calculating the environmental effects of different ways of moving people must in the context of infrastructure planning and investment take account of the anticipated technological development during the depreciation period and the step-wise enforcement of more stringent environment standards. The correct way of doing this would be to calculate the cost of emissions year-by-year and to translate future costs to present day value by a discount rate. In a world of fast technological development, the outcome to a large extent depends on the length of the depreciation period and the choice of discount rate. A long period in combination with a low discount rate (e.g. 60 years and 2 or 3%) gives a high weight to future, more environmentally benign technologies, while a high discount rate, say 4-6 per cent, means that the results are much influenced by the current, relatively large, differences between the modes. The general expectation among experts is that these differences will diminish over-time as all modes need to become cleaner and more energy efficient.

However, no expert can tell us what new vehicles and engines will look like 30 or 50 years ahead. At best they can forecast with some degree of accuracy what designs and engines that will dominate the production of new vessels and vehicles 10 or 15 years from now. Given that trains and aircraft tend to become 25 to 30 years before being scrapped, most vessels produced in, say, 2025 will still be used in 2045, which is 36 years from now. However, by then these vessels will only make up a small part of the total fleet. The life of cars and buses are shorter but newly produced passenger cars may on average become 15-20 years old before being replaced. Electric vehicles, that have very durable engines and transmission systems, may in future become even older.

One way around the problem with unknown future technologies and the choice of discount rate may be to base the evaluation of the long-term environmental performance of the different modes on what might be the best available technology in 2025, 16 years from now. The assumption would then be that these technologies will dominate transport at mid-term of the depreciation period and may be taken as a proxy for the environmental impact of a mode over an entire period of 50-60 years. In the following sections this simplified method is used for producing a rough picture of the long-term differences in environmental impact per passenger kilometer.

5. TAILPIPE EMISSIONS

Regulated exhaust emissions occur from all types of internal combustion engines as well as from power plants that use fossil fuels or biofuel. The maximum permissible tailpipe emissions from cars and buses have been drastically reduced over the last 15 years and will continue to decline. By 2025, new vehicles may be expected to emit so little that the aggregated impact from the entire new fleet is negligible. However, as cars and buses have operating lives of 15-20 years, it will take until about 2035 for the total fleet to be clean. By then the share of electric road vehicles and plug-in electric hybrid cars may also be substantial.

The electricity used by trains, and in the future by a growing number of cars, is marginally produced in coal-fired power plants, and in most countries such plants dominate the grid. Some power plants still emit huge quantities of sulphur and NO_x. Several European power plants, most of them located in Eastern Europe currently emit more than 100,000 tons of SO_2 at thermal capacities ranging from 800 to 12,000 MW, and a number of plants, most of them British, emit more than 20,000 tons of NO_x per year (Entec, 2008). This means that the worst emit more than 20 gram SO_2 and 3 gram NO_x per kilowatt-hour produced. However, by 2025 such power plants will either have been decommissioned or have had to clean up their operations. Thus in the longer term also the regulated emissions from power plants will have been reduced to sustainable levels.

The assumption here is therefore that the remaining tailpipe emissions, if any, as well as those originating from power production can be disregarded in a long-term comparison between the land-based modes.

For aviation, the situation is more complicated. The emissions of NO_x from aircraft are a major long-term concern, however, primarily because of their contribution to climate change. Emissions of NO_x from airplanes will therefore be addressed in the below section on greenhouse gases.

6. NOISE

Problems associated with noise from vehicles and vessels are sight-specific. It is therefore difficult or impossible to calculate average noise costs for different modes. However, a few general observations can be made. Intercity journeys by car or bus usually takes place on motorways or other high-standard roads that allow speeds of 90 km/h or more. At such speeds the tire-to-surface noise dominates over engine noise, which means electrification of road vehicles will have limited impact on equivalent noise levels. On the other hand, motorways and other major roads are often built as to avoid crossing through minor towns and other settlements. That means fewer people are victims of such road noise compared with noise from railway lines which for historical reasons were often designed to go through the heart of towns. However, new high speed lines may also avoid crossing through smaller towns where no stop is made anyway.

A 50 per cent reduction of external noise from trains and aircraft appears to be technologically possible. Additional improvement can be achieved by using noise-absorbing road surface materials or install absorbents close to the railway track. Shielding by noise-protection walls may significantly reduce the impact but only relatively close to the barrier. People living further away will be affected by the diffuse background noise that barriers cannot stop. Where aviation is concerned, the only way of shielding is by improved insulation, particularly of windows.

The noise created by large carriers amount to less per passenger kilometer compared with an equally high sound from a smaller vehicle or vessel. Trains that can seat hundreds of passenger therefore create less noise per passenger kilometer than cars even when making much more noise when passing. However, in the road environment noise is dominated by trucks. The marginal contribution of an additional car to an already busy road is small.

The conclusion is that the social marginal cost caused by traffic noise cannot be included in a generalized comparison of the different modes. A shift from one mode to another may increase or decrease the impact on human health depending on local circumstances.

7. LAND USE AND LANDSCAPE

The use of land and the impact on landscape is also to a large extent site-specific. However, some general observations can be made.

Aviation, for natural reasons, consumes much less land per passenger kilometer than other passenger transport modes. An additional flight generally does not cause any extra damage in this sense while growing traffic volumes may after a while require an additional runway or even a new airport.

Intercity traffic by car, bus or traditional intercity trains share infrastructure with vehicles bound for other destinations and to some extent with local traffic. The marginal impact on land-use is usually zero. It is only when congestion calls for additional infrastructure capacity to be built that more intercity traffic will make a difference. If new capacity is created simply by adding a new lane or track, the marginal effect on land-use is limited and no new barrier is formed.

Introducing high-speed trains where no previous infrastructure for such trains exists requires a new railway especially designed for this type of traffic. High speed traffic requires a layout with large radius curves and limited gradients.[1] The horizontal curve radius must be at least 5.5 kilometers to accommodate speeds of 300 km/h, and should ideally not be less than 7 kilometers (UIC, 2008a).[2] For these reasons high speed tracks are often built in new corridors although partial location to existing railway or motorway corridors is sometimes possible. This means new land is occupied and new barriers are created.

A potential side-effect of building a new line for high speed trains is that more room is left for other types of train on the pre-existing rail infrastructure. Proponents of high speed rail often claim that the creation of new corridors makes an expansion of goods transport by train possible in otherwise congested railway systems. An indirect effect of this may be that a shift from road to track will reduce the overall environmental impact of freight transport. However, for this to happen there must be a latent demand for transport by rail that could previously not be met for lack of capacity.

Theoretically, rail because of the high capacity of trains require much less land for a given number of passenger than roads (although buses require less space than cars). However, to be able to make maximum use of this advantage all trains should run at the same speed and stop at the same stations. Mixing fast and slow trains with each other, and passenger trains with freight trains, may significantly reduce the actual capacity of a railway corridor.

8. CLIMATE CHANGE

The transport sector's contribution to climate change appears to be the only environmental parameter of great concern in a long-term perspective. The remaining part of this paper will therefore be devoted to the question of whether modal shift in intercity passenger transport would do the climate any good. It starts with providing current data and assumptions concerning the future energy-efficiency of the various modes. It then discusses the issue of how to calculate the short-term marginal effect on greenhouse gas emissions, and finally goes on to analyze the impact on actual emissions of load factors.

Calculating carbon emissions well-to-wheel is a complicated matter. In this short paper the comparison between emissions, direct or indirect, from rail, road and aviation are based on tank-to-wheels for diesel and gasoline cars and on fuel-to-electricity for electric trains and cars. This means that for road and aviation fuels, the extraction, refining and delivery of fuels to gas stations have been omitted, and for electricity, the extraction and transport of coal to the power stations as well as any grid-losses. In both cases these emissions amount to 10-15 per cent of the well-to-wheel emissions.

8.1. Electric trains

Modern fast passenger trains travelling at a medium speed of 150 km/h use 0.031 to 0.045 kWh per seat kilometer (Lukaszewicz and Andersson, 2006), while high speed trains at service speeds of around 250 km/h require 0.041 to 0.065 kWh per seat kilometer (RSSB, 2007). The Japanese Shinkansen 700 consumes as little as 0.029 kWh due to a wide-body and large train length which results in more seats per length meter and a very high total number of seats. As the gauge in many countries, notably in most of Europe, does not allow wide-bodies that can seat 2+3 passengers, the following sections focus on trains that seat 2+2 passengers.

Passenger trains need energy for:

- accelerating up to speed;
- overcoming resistance to movement;

– climbing gradients;
– powering control systems;
– lighting, heating, cooling and ventilating the carriages.

The energy needed for accelerating up to speed is defined by the weight (mass) of the train and the speed. This kinetic energy increases with the square of the speed and so does the aerodynamic resistance, i.e. the drag needed to push the train through the air (UIC, 2008b). Therefore moving the train at 300 km/h will require four times the energy needed for a trip in 150 km/h (all else equal).

The rail sector is committed to reducing the average electricity consumption by different types of trains by investing in new technologies and by making operations more fuel-efficient. CER, the Community of European Railway and Infrastructure Companies has committed to an overall reduction of 30 per cent in CO_2 emissions per passenger and ton kilometer between 1990 and 2020. The companies will make use of new or improved technologies as well as “Eco-driving”, active traffic management and efficient timetabling (UIC, 2008b).

Some railway companies have already achieved reductions of this magnitude. In the UK the specific primary energy consumption in passenger rail transport fell by 25 per cent between 1995 and 2006, and Deutsche Bahn reports reductions by one third between 1990 and 2007 for both freight and regional passenger traffic (UIC and CER, 2008). However, the primary energy consumption in passenger long distance traffic was not reduced at all, presumably reflecting a shift to higher average speed.

Actual consumption per seat kilometer depends on:

– Train length;
– Number of seats per length meter;
– Aerodynamics;
– Weight;
– Tunnel length and tunnel diameter;
– Average speed and top speed;
– Number of stops and accelerations/decelerations due to changing speed limits;
– Engine efficiency and degree of regenerative braking.

Reducing air drag is the single most important measure for cutting electricity consumption at high speeds. Nose and tail need to be adequately streamlined. Bogie shielding, removal or shielding of roof-based equipment and closing inter-car gaps are other measure of importance. Much of this is cost-effective already at speeds between 150 or 200 km/h and should be demanded by any cost-conscious operator and delivered by all train manufacturers. However, the fact that energy consumption (all else equal) rises dramatically with speed indicates that there may be some very expensive measures that pay-off only in high speed trains.

The impact of tunnels on aerodynamic resistance depends on the narrowness of the tunnel area and is larger for single-track than for double-track tunnels. If tunnels make up 10 per cent of the distance of a high speed line they may increase overall air drag by around 8 per cent and the overall energy consumption by 5 per cent.

The number of seats per length meter is also highly relevant for energy consumption (at any speed).Travelling at high speed over medium distances may allow the operator to save some space by substituting a bistro or restaurant car by trolley catering.

As shown in Table 2, the estimates of the difference in energy consumption between conventional intercity trains and high speed trains vary considerably. High speed trains are said to require anything from 9 to 150 per cent more energy per seat km.

Table 2. **Literature statements on the difference in energy consumption between conventional and high speed rail**

Source	Unit	IC	HSR	Difference, %
van Essen et al. (2003)	MJ/seat km	0.22	0.53	+141
Kemp (2004)#	Litre/seat km	46 (225 km/h)	88 (350 km/h)	+91
Rail White Paper (UK, 2007)	Energy/seat km	200 km/h	350 km/h	+90
Kemp 2007 (Figure 27)		200 km/h	300 km/h	+45
Network Rail (2009)	g CO_2/seat km	11.7	12.8	+9
Network Rail (2009) *	kWh/seat km	0.028 (200 km/h)	0.033 (300 km/h)	+18
Lukaszewicz and ¤ Andersson (2009)	kWh/seat km	0.022 (180 km/h) Θ	0.027-0.031 (250 km/h),	+32

Approximate figure, taken from graph
* Future trains, Hitachi Super Express vs. Alstom AGV (both 650 passengers)
¤ Future high speed train
Θ Evert Andersson, personal communication.

This range is clearly much too big. According to Network Rail (2009), running resistance accounts for 68 per cent of the energy consumed by an intercity train, while inertia and comfort functions make up respectively 10 and 22 per cent. The two latter will not change much with increasing speed for a point-to-point service. As running resistance at speeds above 200 km/h is dominated by drag, which increases with approximately the square of the speed, it seems reasonable to assume that increasing the service speed from 200 to 300 km/h should raise electricity consumption by about 85 per cent, all else equal. This is close to figures given by Zängl (1993), who says that a German ICE running at 300 km/h (constant speed) would require 83 per cent more energy per seat km than when the same train travels at 200 km/h.

The author of this paper has searched, in vain, the literature for marginal cost curves for technical measures that improve the fuel-efficiency of passenger trains, and he has in addition consulted experts without getting an answer to his question concerning how much more could be done on high speed trains compared to new conventional trains. However, the marginal cost curve for reducing air drag most likely is rather flat.[3]

The difference between what can be done to reduce the consumption of a high speed train compared to conventional fast trains will be a result of the former being able to accept a higher marginal abatement cost and/or of a difference in train length or seats per length meter. However, the latter, assuming that operators attempt to maximise profits, would only be a result of substituting a restaurant by a trolley service, assuming that passengers can content themselves with the latter when the travelling time is short. Where train length is concerned, operators can choose to meet the higher demand for the high speed train compared to traditional services either by adding cars or by offering more frequent services (or possibly a combination).

Based on the above argumentation, it is assumed that the difference in seat km energy consumption between future 200 km/h IC trains and 300 km/h high speed trains may be in the order of 40 to 50 per cent, assuming the same number of intermediate stops. The assumption is that most of the reduction compared with the all-else-equal case comes from making the high speed train longer than the conventional train. In the sections below it is assumed that trains with such top speeds will on average run at respectively 150 and 280 km/h due to track restraints.

Diverting traffic from an existing line where the service speed is 150 km/h to a new high speed line that allows an average speed of 280 km/h would make energy use per seat km increase by at least 60 per cent (assuming use of modern technologies in both cases). The effect on air drag from 10 per cent tunneling is also considered in this choice of mark-up for high speed electricity consumption. In the calculations below, the new 2025 fast train is assumed to consume 0.018 kWh per seat km (150 km/h), while the high speed trains uses 0.029 kWh (280 km/h).

The issue of how electricity consumption affects CO_2 emissions is discussed in a later section.

8.2. Aviation

Modern commercial aircraft use on average between 0.029 and 0.039 liters per seat km.[4] Short-range aircraft, however, burn significantly more than long distance flights.

Substantial improvements in airframe design and engine efficiency as well as wide-scale use of composite materials to reduce weight is expected to take place in the next few decades. "Clean Sky" is a R&D program under the European Commission's Seventh Research Framework Program. According to the Advisory Council for Aeronautical Research in Europe (ACARE), the greening of air transport means developing technologies to reduce the environmental impact of aviation with the aim of halving the amount of CO_2 emitted by air transport, cutting specific emissions NO_x by 80 per cent. Reducing soot, water vapor and particulates emissions will also be tackled. These targets are to be reached in 2020.

The International Air Transport Association (IATA) is much less optimistic but believes that it should be possible to reach 1.5 per cent average annual improvement in fuel efficiency to 2020.[5] The difference between ACARE and IATA can probably be explained, at least partially, by ACARE's focus on new technologies. IATA's target appears to concern fleet averages.

Based on Easyjet's Corporate Responsibility Report 2006, ATOC (2009) estimates that flying a new Airbus A319 causes CO_2 emissions of about 115 g per seat km for a 300 km flight and 85 g for a flight twice that long. This corresponds to 0.046 and 0.034 liter of kerosene respectively. Boeing claims that the new 7E7 "Dreamliner" will require only 0.017 liters per seat km, while Airbus says the A380 will consume less than three liters per 100 passenger km (RSSB, 2007). Assuming a 70 per cent load factor, the latter means 0.021 liter per seat km. However, these figures are for long flights. One might have to multiply by 1.5 to arrive at figures that correspond to the fuel consumption of short-hauls (RSSB, 2007)

ATOC (2009), referring to work done for the British Committee on Climate Change, thinks that CO_2 emissions from short-haul flights can be reduced by 35 per cent by 2025 compared with 2006, resulting in 62 g CO_2 per seat km. This corresponds to 0.025 liters per seat km and is the value that will be used in this report. It should be recognized that shifting from jets to turbo-props would reduce fuel consumption further, however, at the price of lower speed.

Aircraft emit several other gases and substances that contribute to global warming, among them NO_x, water vapor and particles that form ozone and contrails and may contribute to the formation of cirrus clouds. To take account of this contribution, the aggregated impact of aviation is sometimes calculated by multiplying the radiative forcing of the aircraft's CO_2 emission by a factor of 1.5-2.5. An earlier study by the IPCC (1999) even suggested a factor of 2.7. However, technological development will reduce these emissions substantially until 2025, where NO_x is concerned probably more than CO_2. Of relevance in the context of short-distance flights is also that the aircraft cruises at high altitude during a relatively short part of its journey and often does not reach the tropopause at all. Therefore it makes sense to use a relatively low factor. Econ (2008) suggests factor 1.3 but this report will use factor 1.5.

8.3. Passenger cars

New cars sold in Europe in 2008 on average emitted 154 gram CO_2 per kilometer when driven according to the official EU test cycle. Emissions in real traffic are probably higher, in particular in areas plagued by congestion. An EU directive limits emissions from new cars of average size (weight) in 2015 to 130 g/km and indicates that the limit may be as low as 95 g/km by 2020. The assumption here is thus that new fossil-fuelled cars, including electric hybrids, may on average emit 85 gram in 2025 when driven as prescribed by the European test cycle. However, emissions in intercity traffic may differ a bit from those resulting from the test.

Speed has a large impact on fuel consumption, not only in trains, but in any type of vehicle or vessel. Tests made by the Swedish National Road Administration (2000-2001 models) indicate that emissions at constant speed are 30 per cent higher at 110 km/h compared to 70 km/h. Nevertheless, today´s cars emit more in the urban part of the test-cycle than in the part that represents driving in rural areas. This, however, may change when most cars are equipped with start-stop functions and systems for regeneration of braking energy. The use of full hybrids will have a much greater impact on fuel consumption in urban driving than on the highway. In the long-term the difference in average fuel consumption between the urban and the rural part of the test cycle is likely to be small.

In addition, one should be aware that the rural part of the cycle does not include much of motorway speeds. On the other hand, modern future cars will be equipped with cruise control and other assists that help the driver to keep the speed constant, thereby avoiding the efficiency losses associated with variations in speed. Therefore the assumption here is that the average new fossil-fuelled car will in 2025 emit 105 gram CO_2 per km when driven on non-congested motorways where the speed-limit is 120 km/h and the average speed is around 110 km/h. This represents a level 24 per cent above the assumed emission limit when new cars in 2025 are driven according to the test cycle and equals 21 gram per seat km.

Plug-in electric hybrids and all-electric battery cars that are charged from the grid may become common by 2025. Provided that air drag and rolling resistance are similar to those of the fossil-fuelled cars of the same vehicle generation, one may expect these electric vehicles to consume on average around 0.15 kWh per km when driven according to the European test cycle (King, 2007, and Hacker et al, 2009. Under motorway conditions the specific consumption may increase to 0.19 or 0.20 kWh per km. Thus, this paper assumes that the average consumption when used for intercity journeys will be 0.2 kWh per vehicle km and 0.04 kWh per seat km. The issue of how grid electricity affects CO_2 emissions is discussed in a later section.

8.4. Long-distance buses

Megabus, a British company, reports an average fuel consumption of 0.577 liter per km for its double-deck Stagecoach Megabus, presumably under rather mixed traffic conditions (no details provided).[6] This equals 0.0063 liters per seat km. Norges Naturvernforbund (2008), based on data from Volvo, says that an annual average for the Volvo 9700 when used in intercity traffic is 0.28 liters per vehicle km. This corresponds to 0.0054 liters per seat km.

The current long-term potential for improvement appears to be in the order of 25 per cent.[7] This translates into 0.21 liters per vehicle km for a new bus in 2025 when travelling on motorways with little variation in speed and making few stops. This means 0.0040 liters per seat km and 10.5 gram CO_2 per seat km. Eco-driving may reduce fuel consumption further but is not accounted for here.

8.5. Emissions indirectly caused by electricity consumption

When comparing the impact of different modes, studies often use average power production emission figures for electric trains assuming that they run on that mix. This may be correct when wanting to illustrate the actual impact of traffic during a given (historical) year. However, when the task is to analyze the consequences of investments made to facilitate modal shift, it is necessary to base the assessment on the marginal effects on production and emissions from growing or declining demand.

Growing demand for electricity may coincide with a growing number of windmills and other carbon-free means of power production, but in most countries and regions coal-fired condensing stations will remain the marginal form of production for the foreseeable future. That means under normal conditions that any change in demand will either increase or decrease the use of coal or lignite. A successful and large scale introduction of Carbon Capture and Storage (CCS) may in the long term change this, but it is currently impossible to know to what extent this method will be used by 2025. In the absence of CCS that covers all fossil-fuelled power stations connected to a grid, any increase in demand will, in the short to medium term, slow down the replacement of coal by more environmentally benign sources of electricity.

In this context it is also necessary to note the outcome of shrinking demand for electricity on emissions of CO_2, regardless of whether demand is declining as a result of a recession or because of energy efficiency improvements. Short term, the power plant with the highest variable production cost would be the unit to close first. This will, especially under emissions trading, normally be plants that use lignite or hard coal. Wind mills and hydro power stations would not reduce production in a situation of diminishing demand. For this reason the European Union's Directive on Energy End Use Efficiency and Energy Services (2006/32/EC) recommends that the effect of electricity efficiency improvements should be multiplied by 2.5 to reflect the reduction in primary energy consumption.[8] It would be very strange, indeed, to use marginal figures when demand is shrinking, and average figures in situations of increasing demand. However, this is what the rail industry sometimes does.

From the above it is obvious that in a system-perspective a shift from aircraft, cars and buses to electric trains would reduce demand for kerosene, diesel and gasoline and increase the use of coal and gas.

Some argue that the introduction of carbon dioxide emissions trading means that taking the marginal effect into account has become obsolete. The emissions will not be allowed to exceed the cap no matter how much demand for electricity increases. The only effect would be that scarcity will make the price of allowances rise. This way of arguing would be reasonable if the cap was set to reflect the final target of cutting the carbon emissions of Annex 1 countries by 80 or 90 per cent below their 1990-levels. However, the caps under discussion in Europe and the United States are intermediate targets for 2020 that are only the first steps on a long journey to sustainability.

If scarcity leads to a fast increase in the equilibrium price of carbon, there is reason to fear that politicians may deviate from their current long-term plans. A high or fast rising price may make them hesitate about the future of the cap-and-trade systems, and the caps of the next stage might be set higher than would have been the case at a lower price of carbon (WWF, 2009).

Another difficulty in the context of emissions trading is that the European Emissions Trading System (EU ETS) covers CO_2 emissions from power production but not emissions caused by cars and buses. Aviation will be included from 2012. In contrast, the proposed American scheme (The Waxman-Markey Bill)[9] covers not only fossil energy used in power production but also emissions from fuels delivered to any mode of transportation (by an up-stream approach). As all emissions from transportation are subject to the scheme, one could argue that any expansion of demand for road fuels could not alter the cap. If so there would be no need to worry about gas-guzzling SUVs. However, also in this case there is an evident risk that a high price will prevent future politicians from following the route outlined in the bill, which says emissions should be cut by 80 per cent by 2050 from 2005 levels.

A claim that cap-and-trade systems reduce the marginal climate effect to zero can under no circumstances be used on only one mode of transport (rail), and if used on all, there would be no ground for the rail sector to claim that high speed rail has an environmental advantage over road and aviation as the effect of modal shift on greenhouse gases would by definition always be zero. The conclusion here is therefore that the marginal long-term effect on greenhouse gases is the most suitable way of comparing the environmental performance of transport by different modes.

In power production the marginal emissions may differ somewhat depending on whether hard coal or lignite is used and whether coal is sometimes replaced by natural gas as the marginal source of production. In some systems other means of production may temporarily replace coal, for instance during periods of low demand or periods of high production in hydro power stations. The efficiency of marginally used coal-fired power plants may also differ from time to time and from grid to grid.

A successful climate change mitigation policy will require that coal-fired power plants are gradually phased out or alternatively equipped with CCS. In this paper the use of lignite is assumed to have been terminated by 2025 (or equipped with CCS). In the even longer term, natural gas may have replaced hard coal or all hard coal-fired plants may have installed CCS. In a successful climate change policy, aiming at 80 per cent reduction by 2050, hard coal (without CCS) may have been phased out by 2035 and replaced by renewable energy, nuclear power and natural gas. The latter would then be the new marginal production fuel.

In order to reflect emissions during an entire depreciation period of 50 years for investment in new rail infrastructure one might assume hard coal to be the marginal production fuel during the first two decades and natural gas during the last three. Electricity produced in a coal-fired condensing station with 40 per cent production efficiency gives rise to an emission of about 800 gram CO_2 per kWh electricity produced, while natural gas used in a plant with 58 per cent efficiency emits 350 gram per kWh electricity. This gives an average of 530 gram per kWh for the entire period. This figure disregards the fact that one may have good reasons to give higher weight to emissions in the near future compared to those produced 30 or 40 years from now.

Alternative estimates may be plausible and should, of course, be applied equally on both electric trains and road vehicles that use grid electricity.

8.6. Marginal effects of increasing demand for fossil fuels

Carbon free fuels and low-carbon fuels are also in limited supply. The global potential for producing first and second generation biofuels is much smaller than current demand for conventional fuels. This means any change in demand for road fuels will either increase or decrease the use of fossil fuels such as gasoline and diesel. Aviation is even more dependent on petroleum-based fuels. As fossil fuels are more easily substituted in other applications, they are under emissions trading (or equal taxation) bound to be the last to be phased out.

However, the European Community requires all Member States to deliver 10 per cent of the demand for road fuels in 2020 in the form of biofuels or electricity. Most States are expected to respond by making the oil companies market diesel or gasoline that is blended with renewable fuels such as biodiesel or ethanol. This low-blend may over time increase somewhat. The assumption here is therefore that a mandatory blend of 20 per cent biofuel and 80 per cent diesel or gasoline is the typical marginal road fuel during the depreciation period. The below calculation is based on the assumption that the biofuels will reduce well-to-wheel emissions by 65 per cent (i.e. 0.65 x 0.2 = 0.13). The fact that some cars and buses may run to 100 per cent on biofuel is disregarded as focus is on the marginal effect on a system level.

In the longer term there is an obvious risk, from a climate perspective, that scarcity pricing of oil products will provide incentive for ventures in production of un-conventional oils from tar sands and shales. The Economist reports a continuing high investment in Canadian tar sand despite the financial crises.[10]

However, in countries and states where the government has adopted regulations that force the oil companies to reduce the overall carbon intensity of the entire production chain, unconventional oil is very unlikely to get a foothold. Gradually lowering the caps of trading schemes will also make it difficult for these fuels to enter the market. They are therefore disregarded in the context of this paper.

8.7. Future emission factors

Table 3 summarizes the results of the previous sections of this report and provides estimates for the direct or indirect emissions of greenhouse gases from new vehicles and vessels by 2025.

Table 3. **CO_2 emissions from new vehicles/vessels in intercity traffic 2025. Gram/seat-km**

Mode	**Emission**
Cars with internal combustion engines	18.3
All electric cars	21.2
Long-distance buses	10.5
Fast trains (150 km/h)	9.5
High speed trains (280 km/h)	15.4
Short range aircraft	93.8

It may be surprising that the conventional car emits less than the electric car. This is due to the assumption that it runs on 20 per cent biofuels. With 100 per cent diesel/gasoline the emission would be 21 gram/seat km.

The reader should be aware that the above figures and the calculations carried out below disregard emissions to and from airports and train stations.

8.8. Load factors

To be able to compare passenger modes with each other one has to take account of differences in load factors. Today, on average 45-70 per cent of the seats are occupied in intercity trains with no or few stops at intermediate stations. Traditional airlines appear to have cabin factors around 70 per cent. Where high speed rail is concerned, Network Rail (2009) reports load factors for 12 different services in four countries, ranging from 42 to 88 per cent, with a median value of 64. Regional trains and long-distance buses that stop at many stations have difficulties filling vehicles over the entire distance. They seldom reach load factors above 50 per cent (de Rus and Nash, 2007) although Swebus, in fierce competition with regional and fast trains, reports 56 per cent (2008) for its Stockholm-Gothenburg service.

However, new strategies are gathering ground. Low-cost airlines achieve high occupancy rates by varying their prices, and traditional airlines and train operators are gradually learning how to improve yield management. The below calculations are based on the assumption that the average load factor in 2025 is 80 per cent for regional aircraft, 75 per cent for high speed trains, 65 per cent for conventional intercity trains and 55 per cent for long-distance coaches.[11] The conventional intercity train is assumed (in the absence of a high speed service) to make the same number of intermediate stops as the high speed train. The lower load factor compared to high speed is explained by the fact that the latter will attract more passengers, which will partly be used for offering more frequent services (see above) but will also to some extent improve the average passenger density.

The differing load factors for conventional trains (150 km/h) and high speed trains (280 km/h) means that the latter would only consume 41 per cent more energy per pkm, down from a difference of 60 per cent per seat km.

In many countries, on average only 1.2-1.5 people occupy the five seats of a passenger car. However, the occupancy rate is higher for long-distance journeys than for daily commuting and other local trips. Colleagues tend to share cars when traveling to far away business meetings, and families often choose cars before trains and aircraft when on holiday trips. Studies of the effect on car travel of introducing high speed rail use average car occupancy rates in the range of 1.5-2.2 persons (CCAP &

CNT, 2006, Econ 2008a, ATOC, 2009). Thus, this paper assumes that on average 2.0 persons travel together in cars on intercity journeys (= 40% occupancy rate).

Table 4 shows the emissions from intercity journeys by new vessels and vehicles in 2025 when account has been taken for the average load factor.

Table 4. **CO_2 emissions from new vehicles/vessels in intercity traffic 2025. Gram/pkm**

Mode	**Emissions**
Cars with internal combustion engines	45.8
All electric cars	53.0
Long-distance buses	19.1
Fast trains	14.6
High speed trains	20.6
Short range aircraft	117.2

8.9. Effects on greenhouse gas emissions from modal shift

Politicians all over the world invest public money in new transport infrastructure and to the extent that they do so in order to cut emissions of greenhouse gases they favor investment in rail, including high speed lines. Making road transport shift from cars to buses would also cut emissions but does not require new infrastructure. So it is really about rail. When considering the effects on emissions of greenhouse gases from investing in new rail one must take account of the following changes that may result from the investment:

- Diminishing emissions due to shifts from aviation and road transport to rail;
- Increasing emissions due to higher train speeds;
- Additional emissions resulting from new journeys that are generated by new and faster services;
- The use of free capacity on pre-existing tracks when a new line is built.

The last issue concerns the opportunities of reducing emissions by using the old tracks for extended freight transport and regional passenger traffic and will not be discussed here but in a later section of report.

It may be worth observing that the generation of new journeys means that passengers now spend their time and money on something that in the absence of high speed rail they would have used for some other purpose. However, there is no way of knowing what they might have done and what the carbon intensity of that activity would have been. It is therefore disregarded.

The relative importance of the first three factors will depend on the circumstances in each case. A high speed line that replaces a very inefficient line that only allows average speeds of 80 km will attract more new traffic than a line that is a supplement to or replacement for an existing service in, say, 150 km/h.

The high speed line between Madrid and Seville (471 km) attracted 5.6 million travelers in 2000, seven years after it was opened. The market share for trains rose from 14 per cent in 1991 to 54 per cent, while aviation shrank from 11 to 4 per cent, cars from 60 to 34 and buses from 15 to 8 per cent

(Nelldal et al, 2003). However, at the same time the number of journeys grew by an annual average of approximately 5 per cent, and as a consequence, the number of trips by aviation, car and bus declined much less. The figures indicate that opening the new line must have attracted a great deal of new journeys that would not in its absence have taken place. High speed in combination with modest prices allows people to make trips that they under previous circumstances would not have contemplated, for instance seeing friends more often, having meetings instead of talking over the phone or attending football matches.

In the case of Madrid-Seville, travel time by train was cut by half, from 5 to 2.5 hours (UIC, 2008). Investment in high speed rail does not always result in such reduction. High speed lines between Stockholm and Gothenburg (460 km) and Stockholm and Malmoe (615 km), Sweden's three largest cities, are expected to cut travel times by respectively 27 and 44 per cent (from 2:45 to 2:00 hours in the case of Stockholm-Gothenburg). An investigation on behalf of the Swedish government (UOH, 2009) nevertheless thinks that the investment (a combination of the two new lines) would result in:

- 7.7 billion more pkm by rail;
- 1.6 billion pkm less by aviation;
- 3.1 billion pkm less by car;
- 0.1 billion pkm less by bus.

This means deviated traffic would amount to 4.8 billion pkm per year, and that the high speed links would generate 2.9 billion pkm in the form of new intercity trips. The latter constitute 38 per cent of the total expected increase in passenger transport by rail. However, the high figure for traffic diverted from cars could be put in question. Why would so many prefer rail because of higher speed when the existing rail services between the three cities are already substantially faster than the cars?

In a study for the Norwegian Government, VWI (2007) estimates that building a high-speed link between Oslo and Trondheim (464 km) would result in the changes shown in Table 5. Interestingly in this case, travel diverted from cars to rail is expected be quite small, and the share that consists of new traffic is also smaller despite a more substantial travel time reduction than in the Swedish case.

Table 5. **Diverted and generated traffic after the completion of a high speed link between Oslo and Trondheim**

	Share of total traffic, %
Diverted from aviation to train	38
Diverted from car to train	7
Diverted from bus to train	6
Total shift	**51**
Generated new traffic by train	26
Total increase in train traffic	**77**
Pre-existing train passengers	23
Total passenger traffic by train	**100**

CCAP and CNT (2006) expect diversion from cars to on average account for 47 per cent of total traffic on 12 future American high speed lines, with aviation contributing only 19 per cent and current trains 20 per cent. Such high figures for cars is presumably explained by the fact that cars have a dominating position over medium distances in the United States and that rail services in some cases do not really exist.

To be able to calculate how modal shift may affect emissions of greenhouse gases it is necessary to make an assumption concerning the split between cars with internal combustion engines and electric cars and plug-in hybrids. As the methodology of this paper is based on the emissions from new vehicles in 2025, one must base the assumption on how new sales in 2025 may be divided on different types of engines. It is assumed that traditional cars (including electric hybrids that cannot plug-in) make up 40 per cent of the market, plug-in electric vehicles 40 per cent and all-electric cars 20 per cent. The plug-in cars are assumed to run half of their annual mileage on grid electricity but as they can store only a limited amount of electricity on board, one must assume that only a tenth of an intercity journey by such a car is in electric mode. Presumably many of the battery cars will be bought for local and regional use rather than for long-distance journeys (some families will own more than one car). It thus makes sense to suppose that these cars will represent less than 20 per cent of the intercity traffic. We assume that the share is 10 per cent. That means that the share of grid electricity will only be 14 per cent for new cars in 2025 and that the average emission per pkm will be 46.8 gram.

In order to get an idea of how investment in new rail infrastructure may affect emissions of greenhouse gases a fictitious example is shown in table 6. It shows the result on emissions some years after the inauguration of a new high speed line when the service speed increases from 150 km/h on an existing line to 280 km/h. The example is based on the assumption that traffic consists of 30 per cent pre-existent train travelers, 20 per cent diverted from aviation, 20 per cent from cars and 5 per cent from buses. The remaining 25 per cent are assumed to be newly generated. The table shows the result per 1 million one-way trips between cities located 500 km apart.

Table 6. **An example of changes in direct and indirect greenhouse gas emissions from opening a 500 km high speed link that replaces an existing passenger train service**

Ton $CO_{2equiv.}$ per 1 million one-way trips

	Share of total traffic, %	**Effect on emissions, ton $CO_{2equiv.}$**
Diverted from aviation to high-speed train	20	- 9,660
Diverted from car to high-speed train	20	- 2,620
Diverted from bus to high-speed train	5	+ 38
Total shift	**45**	
Generated new traffic by high-speed train	25	+ 2,575
Total increase in train traffic	**70**	
Pre-existing train passengers	30	+ 900
Total passenger traffic by train	**100**	**- 8,767**

A reduction of about 9,000 tons per 500 one-way trips is not much of improvement. It would not even add much to climate change mitigation in a case where the total number of annual (one-way) journeys was 10 or 20 million.

The calculations behind table 6 disregard the fact that in some countries the existing rail infrastructure can accommodate higher, wider or longer trains than normal in other parts of the world. For instance, in Japan and Scandinavia the distance between the tracks is wide enough to allow for wide-body cars, which is generally not the case in Britain and continental Europe. However, reducing high speed rail energy by 15 or 20 per cent would not change the overall picture much.

9. VALUING CO_2

The positive effect on climate change of investing in modal shift depends to some extent on the value put on carbon.

Provided that cap-and-trade systems become a favored method for limiting the emissions of greenhouse gases and that these systems are linked to each other, there will in future be a (more or less) global price on CO_2. The economic value of achieving net reductions by investing in modal shift will thus depend on the future price of carbon. Depending on the stringency of the caps and technological development the long-term price may fall anywhere in the range of $30-80 per ton. However, by 2025 it is less likely to be much higher than $40-50 per ton CO_2. That means that the socio-economic benefit from reducing emissions as indicated in Table 6 would only amount to $7.0-8.8 million when total annual traffic amounts to 20 million journeys per annum.

10. SENSITIVITY ANALYSIS

Varying some of the main parameters 10 per cent up or down does not provide results that differ enough from the main calculation to justify a differing conclusion. Not even combining several optimistic assumptions in favor of high speed rail makes much difference. Reducing the marginal climate effect from electricity consumption by half and raising the share of total traffic that is diverted from aviation to 30 per cent (and reducing the car share to 10%) would in combination reduce emissions for every one million trips to 16,167 ton CO_2. Assuming very high figures for air traffic diversion does not make sense. It is not possible to replace more than 100 per cent of aviation and in most cases airlines will be able to keep 20 or 30 per cent of their customers. Assuming fewer generated journeys means emissions will fall somewhat, but at the same time an important part of the market for high speed rail disappears.

Not even disregarding completely the environmental impact of electricity consumption (as the Swedish Rail Administration wants it) would reduce emissions per one million single journeys by more than 16,000 to 20,000 ton, depending on whether the share diverted from aviation is set at 20 or 30 per cent. Multiplying the highest figure by 10 or 20 to get the annual contribution from a high speed line does not add up to more than approximately 0.2 to 0.4 million tons per year.

Calculating the emissions year by year throughout the entire depreciation period of the new infrastructure and discounting the cost to present day net value might produce a differing result, especially at a high discount rate. The marginal CO_2 emissions from electricity production will probably be higher in the short term but the price of carbon will, on the other hand, be lower. Using a shadow price on carbon may argue in favor of a rather high discount rate as it is essential to start cutting emissions soon in order to avoid the earth from warming up too much. This rationale means that the environmental benefit of high speed rail diminishes compared with the above example.

The calculations have been based on tank-to-wheel emissions and overhead wires-to-wheels, plus the emissions from electricity production (disregarding emissions from coal and gas extraction and grid losses). A well-to-wheel approach would not have produced significantly different results, not even in a case when the well-tank/well-to-wire emissions were twice as large for fuels than for electricity.

An important parameter that was disregarded in the above calculations is the impact on short-term emissions of building a new railway line. Building a 500 km long high speed line may cause emissions of several million ton $CO_{2eqv.}$ (Norges Naturvernforbund, 2008, and Network rail, 2009), and even if these emissions are balanced by reduced overall emissions in the longer term, they do have a short-term impact on the atmospheric concentration of greenhouse gases. There is thus an obvious risk that investment in high speed rail will add to the difficulties of keeping the atmospheric content of greenhouse gases at a level that prevents the mean global temperature from exceeding its pre-industrial level by more than 2 degrees Celsius. From a climate point of view it might be better to up-grade existing lines and to try to make people use modern telecommunications rather than investing lots of money in making us travel more.

Make your own calculation

The author of this paper has made his best to provide detailed information concerning all of the assumptions on which his calculations and conclusions are based. This allows the reader to vary the assumptions according to his or her own beliefs and make his/her own calculation.

11. FREEING SPACE FOR FREIGHT TRANSPORT - THE SWEDISH CASE

Building a completely new high speed line means traffic will be diverted from the pre-existing rail network. Thus capacity on those tracks can be used for other types of trains, provided, of course, that demand for such services exist. The situation is often complex and the optimal solution may differ greatly from case to case. The Swedish case discussed below should just be seen as one example.

In Sweden, freeing capacity for freight transport is a major argument for constructing new high speed lines between Stockholm and respectively Gothenburg and Malmoe. However, before taking the step to invest in these high speed rail lines, there is cause to investigate whether capacity problems in rail freight transport can be overcome by other and less expensive measures. Improved signaling systems and investment in passing siding may increase substantially the capacity of an existing track (Nilsson and Pydokke, 2009).

In the Swedish case, part of the congestion on the Stockholm-Gothenburg line is caused by containers being transported across the country from the Port of Gothenburg. Most of the containerized goods transported to and from the greater Stockholm area travel via Gothenburg, despite the fact that most of it comes from or is destined to far-off places like China. The Port of Stockholm in now investing in a new container port, located close to the open sea, in order to compete for this market. Hutchinson Port Holdings will run the terminal in connection with its operations in Rotterdam.

In addition a greater part of goods entering or leaving northern Sweden could use short-sea shipping. Improving the Swedish rail infrastructure at high cost does not make much sense so long as the freight trains cannot continue through Denmark to destinations on the European continent. There will be limited rail capacity over the Sound and Fehmarn Belt even after the completion of the Fehmarn Belt Bridge (Rødby-Puttgarten) in 2018.

However, one restraint on short-sea shipping is that the government enforces fairway dues on all ships calling at Swedish ports, and that these fees recover not only the short-term marginal costs but also the fixed infrastructure costs. Freight trains, on the other side, enjoy Europe’s lowest track fees that do not even cover the short-term marginal cost, much less the costs associated with expanding the infrastructure. Sweden could level the playing field by enforcing the same principle of liability on all modes. This implies raising the track fee for trains and introducing kilometre-charging on heavy goods vehicles, which several Member States of the EU have already done or are in the process of doing.

The Swedish example does not apply to other countries or regions, unless they have similar conditions. In other parts of the world other alternative solutions may be more relevant, for instance increasing the use of inland-waterways and/or pipelines. “Gigaliners” fuelled by grid electricity might be an option in regions where the motorways are not crowded. Electrifying a motorway would not involve excessive cost or high emissions of CO_2.

12. CONCLUSIONS

There is no cause to prohibit investment in high speed rail on environmental grounds so long as the carbon gains made in traffic balances the emissions caused during construction. The rail sector, however, often claims that investment in rail infrastructure will bring large environmental benefits (Banverket, 2008, UNIFE 2008, UIC 2008). Independent research, on the other hand, concludes that these benefits are not so important (de Rus, 2008, WSP and KTH Järnvägsgruppen, 2008, Nilsson and Pydokke, 2009). The results of this report support the latter view.

Investment in high speed rail cannot be expected to contribute much to climate change mitigation. Investment in conventional fast trains may in some circumstances be significantly more beneficial. It may be time for many environmentalists to reconsider their attitude to high speed rail. While in some cases calling for huge investment in high speed rail, the environmental organizations want speed restrictions for road vehicles to be tightened, aircraft to be designed for lower speeds and ship operators to involve in slow-steaming.

The cost of building high speed lines is high, €9-40 million per km according to de Rus (2008), and 12-30 million according to UIC (2008). de Rus puts the average cost at €18 million. Huge traffic volumes appear to be the only way to recover these costs. The principal benefits of high speed rail are time savings, additional capacity and generated traffic. Wider economic benefits may also be important, however, difficult to estimate. The strongest case for high speed rail is where traffic volumes are high (de Rus and Nash, 2007).

"Only under exceptional circumstances (a combination of low construction costs plus high time savings) could a new HSR line be justified with a level of patronage below 6 million passengers per annum in the opening year; with typical construction costs and timer savings, a minimum figure of 9 million passengers per annum is likely to be needed" (European Commission, 2008).

The conclusion of this paper is that investment in high speed rail is under most circumstances likely to reduce greenhouse gases from traffic compared to a situation when the line was not built. The reduction, though, is small and it may take decades for it to compensate for the emissions caused by construction. However, where capacity restraints and large transport volumes justify investment in high speed rail this will not cause overall emissions to rise.

In cases where anticipated journey volumes are low it is not only difficult to justify the investment in economical terms, but it may also be hard to defend the project from an environmental point of view as it will take too long for traffic to offset the emissions caused by building the line. Under such circumstances it may be better to upgrade an existing line to accommodate for somewhat higher speeds as this would minimize emissions from construction and cut emissions from train traffic compared to high speed rail.

NOTES

1. However, high-speed trains can tolerate somewhat steeper gradients than conventional trains.

2. For speeds of 200 km/h, a curve radius of 2.5 km is sufficient (and 3.5 ideal).

3. Evert Andersson, Royal Institute of Technology, Stockholm, personal communication, and Network Rail (2009), p. 9.

4. http://sasgroup.net (Airbus A321-200, A319-100, Boeing 737-400/500/600/700/800, and MD90).

5. *Airlines present climate change proposals to heads of governments*. Press release 22 September 2009.

6. http://www.megabus.com

7. Edward Jobson, AB Volvo, personal communication.

8. 2.5 equals an electricity efficiency of 40%, which is normal in coal-fired condensing power stations.

9. The American Clean Energy and Security Act of 2009.

10. *The Economist*, 5th September 2009.

11. RSSB assumes 60% for point-to-point services.

BIBLIOGRAPHY

ATOC (2009), *Energy Consumption and CO_2 Impacts of High Speed Rail: ATOC analysis for Greengauge 21*. The Association of Train Operating Companies, London (by Richard Davies and Leigh Thompson).

Banverket (2008), *Svenska Höghastighetsbanor*. Banverket Rapport 2008-005-30, Borlänge, Sweden. (National Rail Administration)

CCAP and CNT (2006), *High Speed Rail and Greenhouse Gas Emissions in the U.S.*. Center for Clean Air Policy, and Center for Neighborhood Technology.

CCC (2008). Building a low-carbon economy – the UK's contribution to tackling climate change. The Committee on Climate Change, London.

DfT (2007), *Delivering a Sustainable Railway*. Department for Transport, London.

de Rus, G. and Nash C. (2007), *In What Circumstances is Investment in High Speed Rail Worthwhile?* Institute of Transport Studies, University of Leeds. Working Paper 590.

de Rus, G. (2008), *The Economic Effects of High Speed Rail Investment*. University of Las Palmas and OECD Joint Research Centre.

ECON (2008a), *Klimaeffekter av højhastighetstog*. Rapport 2008-101, Oslo.

ECON (2008b), Nytte-kostnadsanalyse av højhastighetstog i Norge. Rapport 2008-154, Oslo.

Entec (2008), Evaluation of the Member States' emission inventories 2004-2006 for LCPs under the LCP Directive (2001/80EC). Final report by Entec UK Ltd. Consultancy for the European Commission, DG Environment.

European Commission, DG Regional Policy (2008), Guide to Cost-Benefit Analysis of Investment Projects.

Hacker, F., Harthan, R., Matthes, F. and Zimmer, W. (2009), Environmental impacts and impact on the electricity market of large scale introduction of electric cars in Europe. Critical Review of Literature. ETC/ACC Technical Paper, European Topic Centre on Air and Climate Change.

IPCC (1999), *Aviation and the Global Atmosphere*. Intergovernmental Panel on Climate Change, Cambridge University Press.

Kemp, R. (2004), *Transport energy consumption. A discussion paper*. Lancaster University, 10 September 2004.

King, J. (2007), The King Review of low-carbon cars. Part 1: The potential for CO2 reduction. London.

Lukaszewicz, P. and Andersson, E. (2006), *Energy consumption and related air pollution for Scandinavian electric passenger trains.* Report KTH/AVE 2006:46, KTH Rail Group, Royal Institute of Technology, Stockholm.

Lukaszewicz, P. and Andersson, E. (2009), *Green Train energy consumption. Estimations on high-speed rail operations*, KTH Rail Group, Royal Institute of Technology, Stockholm.

Nelldal, B.L., Troche, G. and Jansson, K. (2003), *Europakorridoren. Ett bredband för fysiska transporter*. KTH Rail Group, Royal Institute of Technology, Stockholm

Network Rail (2009), Comparing environmental impact of conventional and high speed rail. London.

Nilsson, J.E. and Pyddoke, R. (2009), *Höghastighetsjärnvägar – ett klimatpolitiskt stickspår*. Rapport till Expertgruppen för miljöstudier 2009:3. Ministry of Finance, Stockholm.

Norges Naturvernforbund (2008), Energi- og klimakonsekvenser av modern transportsystemer. Effekter ved bygging av høyhastighetsbaner i Norge. Rapport 3/2008.

RSSB (2007), *Traction energy metrics*. Rail Safety & Standard Board, London (by Roger Kemp).

UIC (2008a), *High speed rail. Fast track to sustainable mobility*. Union Internationale des Chemins de fer, Paris.

UIC (2008b), *Process, Power, people. Energy Efficiency for Railway Managers*. Union Internationale des Chemin de fer, Paris.

UIC and CER (2008), *Rail Transport and Environment. Facts & Figures*, Paris and Brussels, November.

UNIFE (2008), *More rail = Less* CO_2. The European Rail Industry.

UOH (2009), *Höghastighetsbanor – ett samhällsbygge för stärkt utveckling och konkurrenskraft*. Betänkande av Utredningen om höghastighetsbanor. SOU 2009:74 (Ministry of Enterprise, Stockholm)

van Essen, H., Bello, O., Dings J. and van den Brink R. (2003), *To shift or not to shift, that´s the question*. CE Delft and RIVM.

VWI (2007), Feasibility Study Concerning High Speed Rail Lines in Norway, Report Phase 2. Verkehrswissenschaftliche Institut, Stuttgart

WSP and KTH Järnvägsgruppen (2008), *Höghastighetståg – affärsmässighet och samhällsnytta*. Slutrapport 2008. WSP Analys & Strategi, and KTH Järnvägsgruppen, Stockholm.

WWF (2009), Auswirkungen von Elektroautos auf den Kratwerkspark und die CO_2-Emissionen in Deutschland, World Wide Fund for Nature, Frankfurt.

Zängl, W. (1993), Die Geisterbahn. Das Dilemma der Hochgeschwindigkeitzüge. Raben Verlag, München.

THE ECONOMICS OF CO_2 EMISSIONS TRADING FOR AVIATION

Peter MORRELL

Cranfield University
Cranfield
United Kingdom

SUMMARY

1. INTRODUCTION

There has been a growing interest in the environmental impact of aviation, both in terms of noise and aircraft engine emissions. Discussions have included both mitigation measures and methods of internalisation of these environmental costs also described as the principle of polluter pays.

This paper focuses on CO_2 emissions from aircraft engines, which have both local and climate change implications, and where the emphasis of most recent discussions has centred. These have taken place at an international, regional and local level: The Kyoto Protocol addresses measures to limit and/or reduce the emissions of greenhouse gases in the transport sector in Article 2(1), and in Article 2(2) directs the Annex I (developed) countries to pursue these goals through the International Civil Aviation Organization (ICAO) with regard to international aviation. The standing Committee on Aviation Environmental Protection (CAEP) of the International Civil Aviation Organisation (ICAO) was asked to investigate proposals for emissions trading, in addition to ICAO's role in setting international standards for engine emissions. CAEP recommended and ICAO accepted the endorsement of an emissions trading scheme (ETS) for international aviation, the establishment of an open voluntary aviation ETS, and the provision of guidance to contracting states on the incorporation of international aviation into domestic schemes. The guidance has been issued but so far no voluntary scheme has been established. Regional schemes such as the EU ETS were welcomed but only if the countries of all participating airlines were in agreement. The EU aviation ETS was thus not supported by ICAO.

At the regional level, the European Commission's 2006 proposals for the inclusion of aviation in the EU Emissions Trading Scheme (ETS) were finally adopted in 2008 in amended form, most of the details of how it would be introduced in its Directive 2008/101/EC that was published in January 2009. The EU has also sets limits on local air quality that affects emission levels around airports, especially from NOx. Thus in Europe, aviation is likely before too long to be required to control or pay for both its local and climate change impacts of aircraft engine emissions. Up to now this has only been subject to longer term changes through increasingly stringent ICAO standards for NOx applied to new aircraft engines during the landing and take-off cycle (but not cruise). Air transport has also been taxed at the country level, with both domestic and international flights included, although these are usually based on passenger numbers with no incentives for reduced emissions.

At the local level, a few airports have introduced emissions charges, and local air quality has become an important issue in airport expansion applications (e.g. London Heathrow).

The pollutants considered as the main ones emitted from aircraft movements (Woodmansey and Patterson, 1994) are CO_2, PM, SO_2, NO_x and HC. The first, CO_2, has lower unit social cost than the others, but the total amount emitted is far larger (especially for the cruise part of the flight). Social costs are defined as the damage to human health, vegetation, buildings and climate change. Their valuation is discussed in Mayeres *et al.* (1996) and Perl *et al.* (1997). The other pollutants account for a lower weight of emissions but have higher unit social costs. CO_2 is estimated to have the longest life (50 to 100 years) followed by methane (8-10 years), with NOx lasting only a number of days or weeks. However, the global warming impact from aviation is compounded by the emissions of NO_x

and water vapour in the upper atmosphere, the latter sometimes leading to contrails and cirrus cloud formation (these effects are summarised in Annex 2 of European Commission, 2005). This is difficult to deal with through an ETS and it is intended to address it through other measures, one of which (standards) is discussed below.

Europe is the region of the world with the greatest pressure to reduce emissions, and it is also the region where almost all of the countries have ratified the Kyoto Convention. The EU has also pushed for the inclusion of environmental impacts in the EU/US aviation bilateral agreement. Growing concern is also evident in other world regions, reflected in the work programme of ICAO referred to above.

The first section of this paper will discuss policy options for aviation in the light of the post-Kyoto pressures for action at an international level. An increasing priority is evident with the lead up to Kyoto 2 in Copenhagen at the end of 2009. Emissions trading will then be examined in the light of the inclusion of aviation into the EU scheme, focusing particularly on the method of allocation and possible distortions that a regional scheme such as this might produce. Next the likely impact of the EU ETS will be assessed in terms of costs, followed by a discussion of potential price strategies and their effect on demand.

2. THE POLICY OPTIONS FOR AVIATION

The first policy option might be some sort of rationing or upper limit on the number of flights operated. This would be almost impossible to administer fairly at an international level. For an individual country it could be implemented at the level of an airport, and this is effectively already done on a selective basis at certain airports. Runway movement constraints and conditions imposed on building new runways has the effect of limiting hourly and annual aircraft movements, although more passengers can be carried by using larger (and more fuel efficient) aircraft.

The second option is to setting stricter standards for new aircraft and engines. There is an existing framework for this at the international level (ICAO) although existing standards are considered to give little incentive to speed up the application of technology to reduce emissions. The standards are for new aircraft and cover only NOx, CO and Hydrocarbons during the landing and take-off cycle up to 915 metres in altitude. These are recommended standards which still need to be incorporated into national legislation. Standards have been set so as not to force the early retirement of aircraft from fleets. A major problem here is the economic life of aircraft and the high costs relative to emissions saved involved in early retirement (Morrell and Dray, 2009).

A third way is by replacing fossil-based fuels with so-called “drop-in” biofuels which offer very low greenhouse gas emissions. This option is thought by some airlines to be a solution by itself, but the Group on International Aviation and Climate Change (GIACC) set up by ICAO proposed a basket of measures which included both biofuels and economic/market-based measures (ICAO, 2009).

A fourth path is through voluntary targets. These have been introduced by many airlines, airline trade associations and aircraft and component manufacturers (e.g. through the targets for new aircraft in ACARE, 2002). Airlines usually set targets at between 1-2% a year improvement in fuel efficiency;

given longer term traffic growth of 4-5% this does nothing to cap or reduce emissions, but does slow their growth.

None of the above options use market mechanisms to offer incentives for emissions reduction, and neither do they incorporate the principle of polluter pays. This can be achieved either by capping emissions at a given level and allowing entities to buy and sell emissions permits according to whether they are above or below the cap. These are called Emissions Trading Schemes (ETS). Depending on how much of the cap is allocated without cost, entities will pay for their pollution by acquiring additional permits at the "market price".

As governments prepare for the UN climate change meeting in Copenhagen (December 2009), the world airline trade association, IATA, has proposed a global sectoral approach for aviation in the successor to the Kyoto Protocol (IATA, 2009). Under such an approach, aviation's emissions would be capped and accounted for globally, not by state. IATA would work with ICAO to ensure compliance. Airlines would get carbon credits for every "environmental" payment, whether in taxes, charges or ETS payments to avoid double counting. It should be noted that there is a number of sizeable airlines that are not members of IATA, and ICAO at present has no power to enforce such a scheme.

The alternative is a fuel or emissions taxes. These would need to be coordinated globally, and would give a long-term signal enabling investment in fuel efficiency. The level of tax would need to decided, but this could be lowered with increased market fuel prices and vice versa, to improve incentives. Implementation would have to be at a country level and there would be a high chance of the revenues collected not being directed to environmental projects. However, the real problem would be to remove a clause in international Air Services Agreements that forbids any kind of fuel taxes or levies. These form the basis for international route rights, although their importance has declined with the growth of open aviation areas such as the European Aviation Area.

Emissions trading schemes are becoming more widespread with the EU ETS started in 2005 and a similar scheme (RGGI) for seven states in the northeast of the US, capping CO_2 emissions at their 1990 level. Voluntary schemes have been launched domestically in the UK (British Airways participated) and Japan. More recently both the Australian and New Zealand governments are planning to introduce a cap and trade scheme to control industrial emissions including those from domestic aviation. These schemes are similar to the EU one, except they have an initial cap on the carbon price paid, exclude international air transport and, for domestic flights, issue permits to fuel suppliers rather than airlines. New Zealand gives domestic aviation a two year delay before it is included.

In the US, the Waxman-Markey Amendment was narrowly passed in the House of Representatives and will shortly be debated by the Senate. This bill would introduce a cap and trade scheme but only applied to ground based emitters. The bill deals with aviation through stricter standards on new aircraft from 2013.

Introducing an aircraft engine emissions trading scheme has been studied and discussed since the 1990s, both at the world and regional level. Considerable analysis has been applied to an international scheme by ICAO's CAEP mentioned above. The reluctance of the US to be involved and other problems has meant that actually introducing such a scheme is unlikely for the foreseeable future at a world level.

In Europe, the focus switched from a preference for emissions charges and taxes (European Commission, 1999) to emissions trading as the best way forward. A European scheme is now almost certain to become reality for air transport in 2012. The European Commission published a earlier study on economic incentives to mitigate greenhouse gas emissions from air transport in 2002 (CE Delft, 2002). Their analysis was limited to two policy options: an environmental charge and a Performance Standard Incentive (PSI). The latter was based on emissions per unit of output, but did not address the allocation of permits or trading. In a later study, it concluded that the most effective way of meeting its policy objectives were:

– Emissions trading;
– Emissions charges.

Both the above are economic instruments that would lead to the internalisation of the cost of climate change. Each could, in principle, be designed to achieve the same level of emissions reduction (European Commission, 2005).

GAO (2009) argues that an emissions tax is “generally a more economically efficient policy tool to address greenhouse gas emissions than other policies, including a cap-and-trade program... .” This is because “...it would better balance the social benefits and costs associated with emissions reductions.” However, both the tax and cap need to be set at the right level and an aviation tax would have to be agreed internationally. The ETS also has the potential to achieve emissions reductions at the lowest cost or impact on GDP. But as with a tax the cap could be set at the wrong level, although in this case the efficiency loss would be greater than setting the tax at the wrong level (Stern, 2006). Those industries with lower abatement costs such as coal fired power stations would sell allowance to those with much higher marginal costs such as air transport. Aviation has already exploited many of its lower abatement cost options as a result of its dependence on highly priced kerosene, and is thus likely to need to purchase allowance from others to expand. This underlines the importance of an aviation scheme with open trading.

A study by CE Delft (2005) for the European Commission examined concepts for amending Directive 2003/87/EC to address the full climate change impact of aviation through emissions trading. The study concluded that aircraft operators would be the best entity upon which to base the system, with allocation decided at the EU, rather than individual member state level. It also came out in favour of including only CO_2, at least initially. It looked at the possibility of restricting the scope to intra-EU flights, as well as including all flights to/from EU airports. This and other main recommendations of the study formed the basis for the scheme that was adopted.

As to the allowance allocation method, the study reported that “auctioning appears to be the most attractive option for allocation.” This was because it was the most efficient method, it treated new entrants and incumbents equally, it provided credit to airlines taking early action and it involved the issuing authority with lower data requirements. On the other hand it imposes a greater financial burden on the industry. Their second-best option was benchmarking, and the least attractive was grandfathering, although either of these could be combined with the other method.

3. EMISSIONS TRADING AND AIR TRANSPORT

3.1. The principles of the EU aviation ETS

Following the proposal at Kyoto, ICAO's Committee for Aviation (CAEP) considered and evaluated measures to reduce aviation emissions including the possible introduction of an ETS. It concluded that fuel taxes were impossible to introduce and encouraged regional emissions trading initiatives (subject to third country agreement). Thus nothing was likely on a global scale. In the meantime, the EU moved ahead with the incorporation of aviation into their existing ETS that was implemented for other ground based polluting industries from 2005.

The EU Directive for aviation was finally introduced in January 2009, and its provisions are expected to be incorporated into the legislation of each member country by the end of the year (European Parliament and Council, 2009):

- Includes aviation in the existing scheme for greenhouse gas emission allowance trading;
- First year 2012;
- All flights to/from European Community airport;
- Various exemptions including smaller aircraft, military, training and rescue flights;
- Greenhouse gases cover only CO_2;
- Cap based on actual emissions averaged across calendar years 2004, 2005 and 2006;
- Cap set at 97% of baseline in 2012, and 95% for 2013 to 2020;
- Emissions allocation based on benchmark;
- Initially 15% of allowance to be auctioned;
- Provisions for free allowance to be given to start-up airlines (with no operations in 2010) and those whose Revenue Tonne-kilometres (RTKs) are growing by more than 18% pa.

Some details were still to be finalised, such as the method of auctioning and the percentage of auctioning in subsequent years. The baseline 2004-06 cap is expected to be published later in 2009, and the actual amounts allocated to airlines will have to await the 2010 shares of RTKs.

From the time of publication of the European Commission's first proposal (2006) and the emergence of the Directive, there was considerable industry lobbying and studies, and the stronger role of the European Parliament is also reflected in the outcome. The latter proposed that the Commission's original proposal of a 100% cap was reduced to 90%, with all flights included from 2011 (not just the intra-EU flights in the original). The European Parliament Green Party was advocating 100% auctioning, with the Parliament settling on 25%. This crucial variable was initially set at 15% for 2012 but left open for 2013 to 2020, presumably dependent on how other industries in the scheme are treated. Given the state of the economy in general and the air transport industry in particular it would not be surprising to see little change in the auctioning share.

Taking British Airways as an example, 85% of its 2004-06 aircraft emissions of around 16 million tonnes of CO_2 would be worth €544 million at a CO_2 price of €40 per tonne. This gives an average of €16 per passenger, many of which are on long-haul sectors. New entrants might be deterred in a limited way by this free allocation. However, a fund will be established both for new entrants and those airlines growing by more than 18% a year. The Directive states that 3% of the total allocation of allowances shall be reserved for such applications, with a maximum of 1 million allowances per airline. Since there are unlikely to be any fast growing airlines, all or most of this should be available to start-ups, with the upper limit allowing the new entrant up to between 2-5 million passengers a year, depending on business model and length of haul.

3.2. Benchmarking

There are two different approaches to the allocation of the free allowances: grandfathering and benchmarking. The former gives airlines allowances in proportion to their emissions in the base year or years, while the latter seeks to reward those airlines that have already take steps to reduce their emissions through investment or improved operations. Benchmarking penalises those airlines that are less efficient than the "average" and rewards those that do better. The "average" can be formulated in different ways.

Benchmarking using a traffic rather than capacity metric has the advantage of rewarding airlines that have already introduced efficient aircraft, and those that achieve higher efficiency than their competitors. It is thus favoured by airlines that have high passenger load factors, e.g. Low-Cost Carriers or LCCs (Frontier Economics, 2006).

Benchmarking involves the determination of a baseline efficiency measure, say RTKs per tonne CO_2, fixing an overall CO_2 cap, and allocating CO_2 allowances depending on an airline's share of RTKs. This was EU aviation ETS approach:

$$RTK_{total} = \sum_{i=1}^{n} RTK_i \quad (1)$$

$$E_{total} = \sum_{i=1}^{n} E_i \quad (2)$$

$$A_i = \frac{(E_{total})}{RTK_{total}} * RTK_i \quad (3)$$

where,

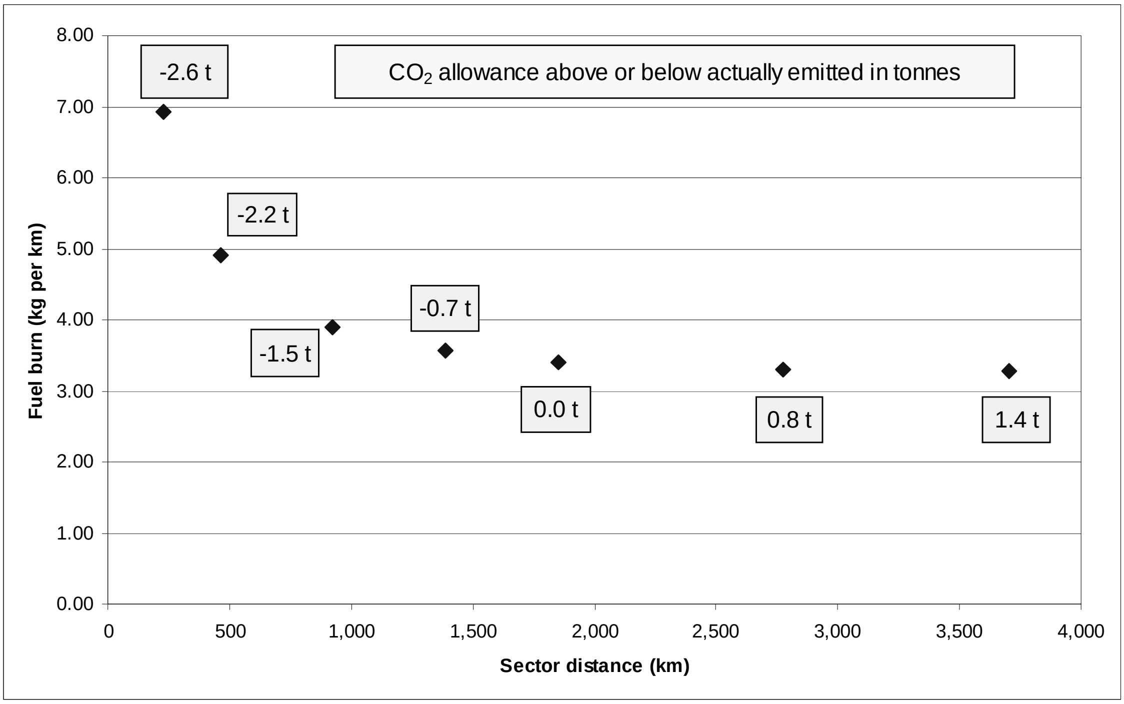
n	=	number of airlines taking part;
RTK_{total}	=	Total RTKs in the reference year (calendar 2010) for those taking part;
RTK_i	=	Total RTKs performed by the airline i in 2010;
E_{total}	=	Emissions assigned to all airlines in the base period 2004-06 (average);
E_i	=	Emissions assigned to airline i in the base period times 97% (less amounts reserved for new entrants and fast growers) in the first year and 95% subsequently;
A_i	=	Emission allowances assigned to each airline for each of the years 2012 to 2020.

First, this method puts a smaller burden on those airlines operating with high load factors and over longer sectors. Second, those airlines flying shorter sectors would tend to be penalised, although Sentance and Pulles (2005) argue that this would encourage passengers to take less polluting forms of transport such as rail. The latter distortions could be addressed in alternative benchmark approaches, but with increased complexity (Morrell, 2007). Other distortions are addressed in Faber *et al.* (2007).

Figure 1 shows a hypothetical example of the difference in allocation using the EU ETS proposed method of benchmarking. The average fuel efficiency used in the allocation (assuming the base and reference year emissions are the same) is likely to reflect a relatively long sector length, given the inclusion of routes to/from non-EU countries. Taking 1 000 nm or 1 852 km as the average, operators of identical aircraft types could get 1.4 tonnes of free CO_2 allowance more than it actually emitted over its longer than average sector length or 2.6 tonnes less than it emitted. A similar relationship would apply to the latest technology aircraft of this size (B737-700) and equivalent Airbus types (e.g. the A320 family). It should be added that for routes of this traffic density a more fuel efficient aircraft would not be currently available.

If these allowance shortfalls are monetarised using a CO_2 price of €40 per tonne, the extra costs incurred by the 230 km operator would be €103 per flight or less than one € per passenger.

Figure 1. **Impact of benchmarking on B737-400 flight with hypothetical average at 1 850 km sector length**



The use of RTKs rather than ATKs might be considered to favour low-cost carriers (LCCs) at the expense of network carriers. LCCs would favour the RTK metric which would inflate their share of the reference RTK total used for allocation relative to the network, lower load factor airline. However, the cost of additional allowance required by the LCC would be a higher share of its average ticket price. Furthermore, the network airline would have fewer passengers to pass on the cost to, but more passengers that were less price-sensitive and the cost would be a lower percentage of the average ticket price. The network carrier is making a choice to offer fewer seats and operate at a lower load factor to encourage higher yielding (less price-sensitive) passengers.

3.3. Distortions from the EU ETS scope

Air travel markets are often served on a multiple sector basis, especially longer haul ones. Such markets cannot always be operated non-stop, but a one-stop service can be attractive in terms of price, timing, earning frequent flyer awards etc. An example given by EU carrier Finnair (Ihamäki, 2009) is the New York/Delhi market:

- New York - Helsinki - Delhi (11 821 km) served by Finnair;
- New York - Dubai - Delhi (13 229 km) served by Emirates Airlines.

There is no non-stop flight serving this market. The two sectors operated by Finnair would emit an estimated 294 tonnes of CO_2 while the Emirates flights 326 tonnes. Finnair would have to submit an equivalent amount of allowances under the EU ETS, while both Emirates sectors would be outside the ETS scope. Taking €40 per tonne CO_2 would result in Finnair paying €11 740 or €43 per passenger.

It should be added that the Finnair is serving the New York/Delhi market using more fuel efficient sector lengths. Fuel burn per kilometre flown generally declines up to around 4 000 km to 6 000 km in length and then starts to increase due to the additional fuel required to carry the larger fuel load (Peeters *et al.*, 2005). This is more pronounced and at the lower end of this range for flights with very high load factors, as is often the case today. One estimate suggests that serving the market with one long non-stop flight might add 4% to total fuel burn, allowing for the landing and taking off at the intermediate stop (Green, 2002).[1] Thus in the above example an additional 13 tonnes of CO_2 is emitted due to this effect (+4%), the remainder due to the longer overall distance flow (+1 408 km).

On the other hand, Emirates would burn an extra 32.2 tonnes of jet kerosene, or 3 292 US gallons from its indirect routing. At the peak mid-2008 price of USD 4 per US gallon, this would mean extra costs of USD 13 168 or USD 48 (€34) per passenger. Thus Finnair's ETS cost disadvantage is offset to some extent by the extra fuel cost incurred by Emirates, assuming the high fuel prices experienced in 2008. Other flight-time-related costs such as aircraft and engine maintenance would also be higher for Emirates.

An alternative and probably more common pattern would be to locate the sixth freedom traffic hub to the East of the EU rather than within it (as per the last example). These are hubs such as Dubai, Mumbai, Singapore, and Bangkok that can attract traffic between Australia, the Far East and to a lesser extent Africa and the EU. An example that does not require much extra flying is London/Singapore:

- London - Singapore (10 851 km) served by British Airways non-stop;
- London - Dubai - Singapore (11 304 km) served by Emirates Airlines via its hub.

The two sectors operated by Emirates would emit an estimated 415 tonnes of CO_2 using a B747-400 aircraft while the non-stop British Airways B747-400 would emit 387 tonnes. Emirates would have to submit allowances only for its first sector under the EU ETS, or 200 tonnes. British Airways' entire flight would be within the ETS scope, with allowances required for 387 tonnes of CO_2. Taking €40 per tonne CO_2 would result in an Emirates' London/Singapore passenger paying an ETS charge of €24 and a British Airways' passenger €43.[2] In this example, Emirates would consume an additional 8.9 tonnes of fuel, which would approximately cancel the difference at fuel prices of USD 4 per US gallon of jet fuel (an additional fuel charge of USD 32 or €23 per passenger). In this case the one-stop route provides the most fuel efficient mode of operation.

Longer routes such as between the EU and Australia might confer an even larger advantage on sixth freedom carriers located outside Europe along this "Kangaroo" route. However, no airline can operate non-stop and much would depend on whether the EU ETS applied to the initial destination outside the EU as would appear to be the case.

The question arises as to whether EU carriers could overcome this disadvantage by making use of their own or their partners' non-EU hubs. Setting up their own hubs outside the EU is at present severely restricted by their lack of traffic rights under third country Air Services Agreements. Most major network airlines are members of strategic alliances and could make use of such hubs through code sharing or joint ventures.

An example of this is the Hamburg/Los Angeles market which Lufthansa currently serves via its Frankfurt hub. This involves a two sector operation but with the economies of scope that are available from combining other markets through the hub (e.g. Berlin/Los Angeles, Bremen/Los Angeles etc). A non-stop Hamburg/Los Angeles flight has very limited feed at both ends of the market, and unlikely to be economic. The relative viability of the non-stop flight would only be marginally improved from the saving of ETS allowance costs as a result of more direct flying without the extra take-off and landing.

Lufthansa's major alliance partner in the US is United Airlines, and it could operate a joint service, say, Hamburg-Washington DC-Los Angeles with the first sector operated by Lufthansa (and subject to ETS) and the second sector by United (not subject to ETS) with its own code. The Washington-Hamburg DC flight might be operated by a reasonably fuel efficient aircraft since it would benefit from feed traffic from the Americas to/from Hamburg, but this is already available to United without ETS or Lufthansa involvement. Lufthansa will have limited feed to provide from the Hamburg end, and thus United gain little from such co-operation with the EU carrier.

Overall, EU carriers' increased use of non-EU hubs operated by alliance partners will not be much of a solution for them, given that the net cost incentive will be small. It would weaken their own strategic position and probably reduce the number of viable long-haul flights that they could operate, with limited alliance benefits. Any attempt by the EU to try to levy a charge on the non-EU sectors connecting to flights to/from EU airports would achieve little environmental gain in return for a serious diplomatic backlash.

Finally, since rail is not included in the ETS there is the potential for some change in the distortion between the two modes, especially high speed rail (apart from that stemming from other taxes or subsidies). Little research has been done on any likely impact but a recent study of the effect of the Dutch Government's tax on air tickets of €11.25 per departing European passenger estimated only a "slight shift to car and train" (Jorritsma, 2009). The tax was subsequently withdrawn.

4. THE APPLICATION OF THE EU AVIATION ETS

4.1. Allowance costs

The final cost of acquiring the necessary allowances for the first year of the scheme will not be known until the end of 2012, when airlines have a last chance to purchase them in the market. Even the initial free allowance cannot be estimated until 2010 RTK traffic has been reported.

Table 1 shows the range of possible impacts of ETS allowance costs on air fares and profits. The earlier studies assumed that only departing flights form EU airports would be included. Even if 100% of the cost of allowances is passed on, the impact on an intra-EU flight is unlikely to exceed €5 per passenger at what are historically relatively high market prices of CO_2. Long-haul passengers could pay up to €40 on these assumptions, but this attributes none of the costs to the cargo shippers (see below).

The European Commission commented that at an allowance price of €30 “these ticket price increases are modest. Their modesty is also demonstrated by the very limited impact they have on reducing forecasted demand ...” (European Commission, 2006).

The allowance prices assumed are generally based on past market prices. However, studies have suggested that air transport is likely to be a purchaser of allowances given its growth rate and its marginal cost of abatement. This and a tighter scheme for ground based emitters could push up the market price of CO_2 to well above the €30-40 assumed above. Consultants Green Aviation forecast prices in the range of €30-50 in the 2012-2013 timeframe.[3] Airlines can purchase CO_2 emissions derivatives well in advance of the first year in which allowance for their own emissions needs to be found (2012). There will therefore be winners and losers in such trading activity. Auction prices for European Aviation Allowances (EUAAs) are unlikely to go above the market or futures prices for European Allowances (EUAs) at the time, since the former can only be used by other airlines.

Table 1. **Summary of previous EU aviation ETS impact studies**

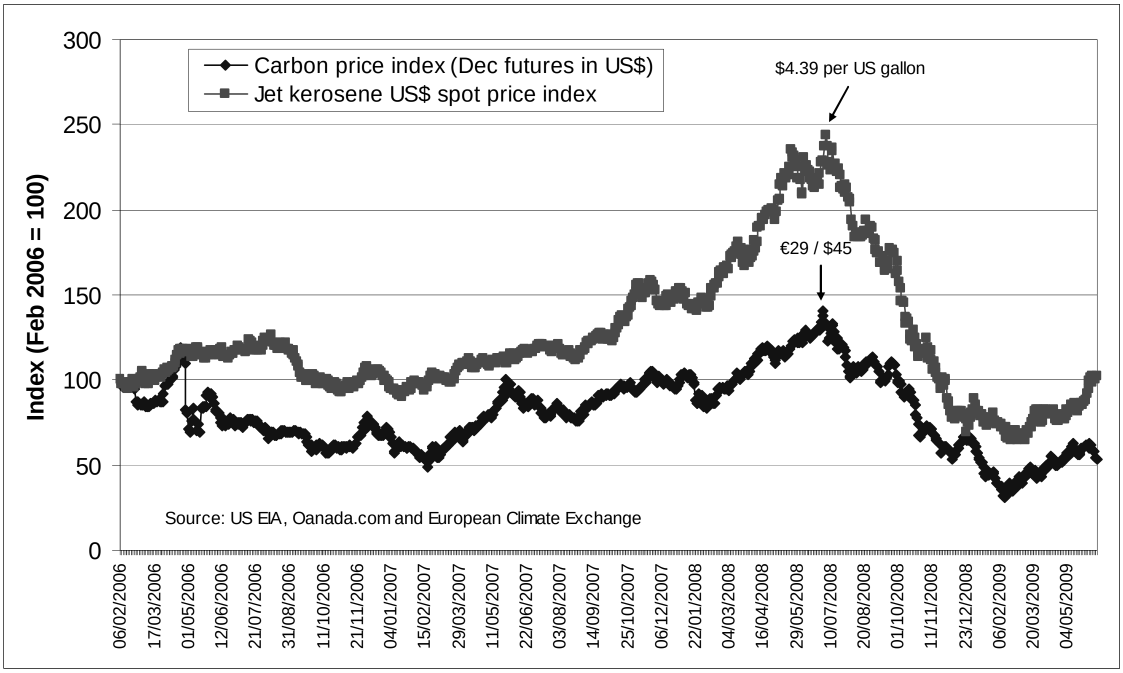
	Short-haul	Medium-haul	Long-haul
European Commission (2006)[1]			
€ per return ticket impact:			
Allowance price: € 6 per tonne	0.90	1.80	7.90
Allowance price: € 30 per tonne	4.60	9.00	39.60
CE Delft (2005)[1)]			
€ per return ticket impact:[2)]			
Allowance price: € 10 per tonne	0.20	0.40	1.00
Allowance price: € 30 per tonne	0.70	1.30	2.90
Ernst & Young - York Aviation (2007)			
€ per return ticket impact:	*Low-cost*		
Allowance price: € 30 per tonne	0.81		
Average one-way fare €	47.44		
Percent increase in fare	1.7%		
Change in demand (elasticity -1.5)	-2.6%		
UK Defra (2008)			
Impact on airline profits			
Price elasticity 1.1-1.3	8-18%		9-20%
Price elasticity: 0.6-0.7			15-20%
Frontier Economics for ELFAA (2006)			
€ per return ticket impact:	*Low cost*	*Full service*	
Allowance price: € 27 per tonne	2.72		5% of av.fare
Allowance price: € 40 per tonne	4.00		8% of av.fare
Change in demand (elasticity -1.5)			
Allowance price: € 27 per tonne	-7.5%	-2.0%	
Allowance price: € 40 per tonne	-12.0%	-3.0%	
Merrill Lynch (2008)			
€ per return ticket impact:	*Low cost*	*Full service*	
Allowance price: € 30 per tonne	1.54	3.52	

1) ETS scope restricted to only departing flights from EU airports.

2) Assuming that the 100% free allowance is not valued and passed on in higher fares.

As Figure 2 shows, there is quite a strong correlation between oil and carbon prices. Electricity generators have a large influence on the carbon market price, and when the price of oil is high they switch to “dirtier” coal which needs to be covered by a greater number of allowances which are likely to have to be purchased in the market. This means that airlines could be faced by increased volatility of combined fuel and fuel emissions costs, some of which will be smoothed out by hedging. It should be noted that the CO_2 prices quoted here are for EUAs. These are used by ground-based emitters in the scheme but can be purchased in the market and submitted by airlines. Airlines are allocated EUAAs which can only be submitted by airlines and not the other emitters. There is likely to be a spot and futures market for EUAAs but with much less trading and liquidity than the ones for EUAs and other instruments.

Figure 2. **Jet kerosene spot and carbon futures (EUA) market prices**



The above studies generally concluded that the impact on LCCs would be lower per passenger, but higher in terms of percentage reduction in traffic. This conclusion is arrived at using similar elasticities and the higher share of ETS surcharge in relation to their average fare, which could be as low as €40-50.

None of the above studies passed on the allowance cost for each flight to the cargo shippers, even though they account for a sizeable part of the payload on long-haul flights. The European Commission's analysis (based on CE Delft, 2005) took a B777-300 for their long-haul assessment and passed on the estimated long-haul allocation of €9 422 per flight to the 238 passengers carried (340 seats at a passenger load factor of 70%), giving €39.60 per passenger at a CO_2 price of €30 (see Table 1). These should be doubled to allow for the fact that the scheme covered both arrivals and departures at EU airports (which would only affect flights between EU and non-EU airports). However, the long-range B777-300 can carry up to just over 25 tonnes of cargo in the lower deck compartment (depending on cargo density), or around 40% of total payload. If only 60% of ETS costs are allocated to passengers rather than the full 100%, the impact on the long-haul ticket price would be €47 rather than just under €79. This should be viewed against airline fuel surcharges which in June 2008 reached a peak of £218 (€275) per long-haul flights of more than nine hours on British Airways.[4]

The Ernst & Young and York Aviation study (2007) was commissioned by airline trade associations covering all business models. It concluded that a large part of the ETS costs would have to be absorbed in reduced profits, with network airline operating margins reduced from 4% to 2.4% (for a CO_2 price of €30 per tonne), passing on around 35% of ETS costs to passengers. Low cost carriers would face margins reduced from 15% to 11.1%, passing on 30% of costs.

4.2. The impact on airline pricing

The costs incurred by airlines as a result of the EU ETS can either be absorbed in reduced profits, passed on to the consumer in higher fares and rates or a combination of the two. Profits could also be enhanced by passing on the value of free allowance received (opportunity costs), or passing on over 100% of the costs incurred. Previous studies do not always make it clear whether they also include these opportunity costs, and it is assumed they do not where they are not specifically identified.

One consideration is the possible marketing advantages of including the ETS charge as a separate add-on to the fare. This might be attractive to some passengers in confirming that the polluter is paying (and to the airline in withdrawing its voluntary offset mechanism). In this case, the non-EU airline gets a clearer signal from its competitor and can include a similar charge but reduce the underlying fare accordingly.[5]

The European Commission's impact assessment assumed that airlines would be able to pass on all of the allowance costs incurred. This was based on CE Delft (2005) assumptions:

- All of the extra costs of ETS allowance would be passed on in markets subject to the ETS;
- Cross-subsidisation between services subject to ETS and those outside it would not occur because this would imply raising fares in non-ETS markets to offset fare reductions in ETS markets; if this increased profits it should be done regardless of the EU ETS;
- There is no empirical evidence either way on the pass through of the opportunity costs, so their evaluation included both approach.

IATA (2007) assumed that 75% of the ETS allowance cost would be passed through to higher fares or a CO_2 surcharge. Merrill Lynch expects that operators will try to pass on "as much as possible of the cost of emissions allowances to customers."

UK Defra (2007) concluded that "the rate of cost pass through is likely to be around 100%, for aviation as a whole, with variations by sub-market". The variations ranged "from 90% to 120% for most aviation services." This was based on a largely theoretical analysis by a consultancy, Vivid Economics, depending on the nature of demand, competition, and whether firms seek to maximise profits, market share or sales.

Some studies differentiated between flights to/from congested airports, where none of the additional costs would be passed on, and uncongested airports where all of the costs would be added to fares (Oxera, 2003). Frontier Economics (2006) calculated a differential impact on low-cost and network carrier, suggesting that not all of any increase in costs due to ETS would be passed on in higher prices by the LCC: "the impact of ETS on aviation prices in general and in any particular market would in practice depend on the elasticity of demand (and supply) in the relevant market." This study assumed a price elasticity of demand of -0.8 for the network carrier's short/medium haul network, and -1.5 for the LCC. This gave a 2-3% reduction in demand for the network carrier as a result of passing through all of the €4 per passenger ETS cost, but a 7.5-12% drop for the LCC. The network carrier could pass on all of the €4 per passenger increase with no fall in revenues, but the LCC would suffer a revenue decline of 2.5-4% if it did so. This analysis ignores the more price elastic passengers carried by airlines such as British Airways (a large part of its non-premium leisure passengers which accounted for 58% of total traffic in 2007[6]), and the less price sensitive traffic on LCCs (e.g. foreign property owners).

Other studies took the more realistic line that ETS costs might be passed on to certain markets or market segments. This is cross-subsidisation which occurs where an airline uses profits it makes in one market or market segment where it has market power to support low prices in other markets or segments which are subject to greater competition. Markets are usually defined on a city or airport pair basis. But they could also be the various market segments travelling on the same city-pair, often simplified to premium and economy passengers. Cargo is another segment carried on the same flight but often disregarded.

Premium passengers are generally thought to be price inelastic and economy price elastic, although there are sub-groups within each category that behave differently. Increasing premium prices and reducing economy fares would thus be expected to increase revenues, other things being equal. Some commentators think that this has been exploited to the full and that premium or business passengers are becoming more price elastic. Airlines are also keen to increase their share of premium passengers on competitive routes because they are generally more profitable (apart from those travelling first class). In Europe, this is likely to take the form of discounting premium transfer passengers (those connecting at their hubs) but not non-stop markets to/from their hubs.

In the context of the EU ETS, the routes in question will involve all airlines (EU and non-EU) incurring additional costs from the need to purchase emissions permits. These costs would lead initially to lower profits. All carriers could pass on the additional costs to the passengers in higher fares in the same way as fuel surcharges, but in highly competitive markets they may prefer to absorb the costs in lower profits or take steps to reduce other costs (such as labour) further to compensate. In this context, airlines appeared much more successful in reducing non-fuel costs during periods of very high fuel prices. Reduced profits would also lead to a higher cost of borrowing, less ability to invest in more fuel efficient aircraft and more competitive products. This would reduce their ability to compete with non-EU carriers in the future. Non-EU carriers could take a hit on profitability much more easily, since the markets in question will probably account for a small part of their total revenues. They could also much more easily absorb the costs across the rest of their network.

Some of the previous studies have discussed “profit maximisation” and an “equilibrium situation” but this is likely to be an oversimplification, and in reality airlines are responding to many changes in both demand and supply as the date of departure of the flight approaches. In the short term airlines tend to try to maximise revenues, with costs relatively fixed. This amounts to profit maximisation but on a dynamic and network basis. Each market’s revenues are spread over a number of sectors such that profit maximisation can only be viewed on a network basis; this offers considerable scope for cross-subsidisation that has nothing to do with ETS (e.g. short-haul feeder routes from profits from long-hauls).

The market segment that this is likely to focus on is the premium traffic, since the marginal revenue gained from attracting these passengers far exceeds marginal costs. However, price is only one of a number of important factors governing premium traffic purchase decisions, the ticket for which is usually purchased by the company rather than the individual (Brons *et al*, 2002). Others include:

- Frequent flyer programmes;
- Corporate agent and travel manager incentives;
- Product features (flight timings, service levels, frequency, etc.).

The last is difficult to adjust on a shorter term basis, and one carrier may have a marked advantage that is already reflected in market share and yield. The first two factors are also very important and give the home carrier a built-in advantage that small price changes would not easily shift (e.g. British Airways in the markets with UK origin or destination). This applies to home market sales, and explains why premium sales in adjacent markets (British Airways' sales in, for example, Germany connecting with their long-haul flights to/from London) are much less dependent on the first two bullet points above and easier to attract. Thus cross price elasticity in the non-stop home markets is relatively low and in the multi-stop (hub-feed) markets much higher.

This example needs to be expanded to include non-EU carriers. They will be competing in the non-stop flights to/from EU carrier hubs, but efforts to attract home market sales will be limited for the above reasons. The home carrier might also defend its premium point-to-point passengers by allocating more of the flight's ETS costs to other segments. On the other hand the non-EU carrier will be able to cross-subsidise in all multi-stop markets travelling on the flight between its hub and the EU carrier hub, and also the non-stop market sold in its home country, although this may be quite small (e.g. Dubai and Singapore).

4.3. Price elasticities of demand

Previous studies have tried to estimate the impact of price increases or ETS surcharges on demand. Some have gone a step further in attempting to gauge the supply response and resultant changes in profitability. Price elasticities have been determined in past studies using econometric techniques over given historical periods of time. These have encompassed periods of economic growth and downturns. The estimates are shown separately for business travel and leisure travel, since these would be expected to show different reactions to price increases or reductions. They are often based on business and economy class or cabin passengers and this is used as a proxy for purpose of trip data that is not reported on any regular basis.

UK Defra (2007) highlighted the range of elasticities determined in previous studies, while Brons *et al.* (2002) distilled some key findings from a survey of 37 studies and examined the impact on the estimates of such variables as class of travel, distance, and level of income. Omitting income from the estimation resulted in an overestimation of price sensitivity. This seems to be supported by a more rigorous pass through of fuel surcharges by airlines during periods of strong economic growth.

Long-haul markets might be expected to show less price sensitivity since there are fewer substitutes and this was apparent from the Brons *et al* data. On the other hand Defra (2007) concluded that there is "no evidence that long and short haul flights have different price elasticities."

Some feel that LCC passengers should be treated differently from leisure traffic in general; this was the view of the Defra study, but Frontier Economics (2006) disagreed. Some LCCs carry up to 20% of passengers on business trips in contrast to European charter flights which have almost none. LCCs differ from network carrier short-haul flights that also carry a mix of business and leisure passengers in having only one fare available at any point in time. This means that they cannot take advantage of price differentiation based on the difference in price elasticities confirmed by previous studies.

The impact on demand resulting from ETS induced price increased will vary depending on the elasticity used. Most previous estimates of the impact on demand are small and insufficient to prevent aviation emissions from continuing to rise in the future. For this reason, an open trading scheme is crucial in allowing aviation to pay for emissions reductions in more polluting industries or to encourage alternative technology energy. Anger *et al* (2008) concluded that 100% pass through of aviation allowances would result in its emissions being 7.5% lower in 2020 than without the ETS.

5. SUMMARY AND CONCLUSIONS

The failure to get agreement for a global Emissions Trading Scheme for air transport through ICAO led the EU to finalise its own scheme. An emissions or fuel tax is ruled out given the hundreds of international aviation agreements that would have to be re-negotiated. Thus the first international aviation ETS will start in 2012 and be applied to all flights arriving in and departing from EU airports. Most of the details of the scheme have now been published, although not yet totally incorporated into EU member states' legislation. The most important decision remaining is whether the percentage of allowance that is auctioned will increase from 2013 onwards and, if so, by how much.

For the first year of operation it is now possible to make good guesses as to how it might impact various airlines, although the baseline emissions total for all airlines in 2004-06 has not yet been published, and the traffic data for the benchmarking (2010) can only be forecast.

Most studies of the impact of the EU ETS on airlines show a modest increase in cost per passenger even assuming all their allowance value is passed on in full. This cost is well below recent fuel cost surcharges, and may have a limited impact on air traffic growth. The degree to which these costs are passed on and which market segment takes the brunt of this will depend on the position in the economic cycle and the pricing strategies of the airlines involved. Given that many airlines take a network-wide approach to pricing it will also depend on the size of the costs in relation to their system-wide revenues. Almost none of the previous studies assumed any pass through of ETS costs to cargo shippers, even though they can account for almost 40% of payload on long-haul flights.

The assumption on the cost of acquiring additional EUAAs or EUAs though auctioning or in the market has tended to be based on past market trends determined by the existing ground based emitters. These may increase significantly as a result of a tighter scheme for ground based emitters and the addition of airlines as net purchasers.

The approach to passing on ETS costs may be similar to fuel surcharges, which network airlines showed as a separate add-on (although some LCCs absorbed them in the underlying fares offered). This might be attractive to some passengers in confirming that the polluter is paying (and to the airline in withdrawing its voluntary offset mechanism). Previous studies also looked at the likely impact on demand of possible price increases. As expected there was a large range of elasticities used, and differing views on the differential impact on leisure versus business, long-haul versus short-haul and LCC versus other types of airline. None of them considered the economic context in which the airlines find themselves, or ETS cost increases acting as a driver to reduce other non-fuel costs.

Any scheme that uses benchmarking for allocation of free allowance will produce some distortions and the EU approach tends to favour longer haul carriers and LCCs. The regional coverage of the EU scheme penalises EU hub carriers and favours those with hubs outside the EU but this impact on not large. The extra fuel needed to carry passengers on indirect routings via non-EU hubs may more than outweigh any ETS costs avoided, and the EU carrier could market its competing service as the more environmentally friendly.

NOTES

1. The gain would be far higher if the long-haul aircraft were designed for a maximum range of, say, 7 500 km, since weight would be saved from lighter structures.
2. Both based on a passenger load of 360 in identical B747-400 aircraft.
3. http://www.greenaviation.org/, 24 June 2009.
4. British Airways Press Release, 29 May 2008.
5. As was the case with fuel surcharges.
6. British Airways Investor Day 2008, www.bashares.com

BIBLIOGRAPHY

ACARE (2002), Strategic research agenda, Vol. 2: The challenge of the environment. October, www.acare4europe.org.

Anger, A., P. Allen, J. Rubin and J. Köhler (2008), Air Transport in the European Union Emissions Trading Scheme. study for the Omega Project, December.

Brons, M., E. Pels, P. Nijkamp and P. Rietveld (2002), Price elasticities of demand for passenger air travel: a meta-analysis, *Journal of Air Transport Management*, 2.

CE Delft (2002), Economic incentives to mitigate greenhouse gas emissions from air transport in Europe. Study for the European Commission, DG TREN, July.

CE Delft (2005), Giving wings to emission trading, a study for the European Commission. DG TREN, July, 51-122.

Ernst & Young, York Aviation (2007), Analysis of the EC proposal to include aviation activities in the Emissions Trading Scheme. 1 June.

European Commission (1999), Air transport and the environment: towards meeting the challenges of sustainable development, COM(1999) 640 final, 1 December.

European Commission (2005), Working Document SEC(2005) 1184, 27 September.

European Commission (2006), Working Document SEC(2006) 1684, 20 December.

European Parliament and Council (2009), Directive 2008/101/EC amending Directive 2003/87/EC so as to include aviation activities in the scheme for greenhouse gas emission allowance trading within the Community, 19 November.

Faber, J., G. Vreede and D. Lee (2007), The impacts of the use of different benchmarking methodologies on the initial allocation of Emissions Trading Scheme permits to airlines. Final Report to DfT and the Environment Agency, UK, July.

Frontier Economics (2006), Economic consideration of extending the EU ETS to include aviation, Report for ELFAA, March, 67-74.

GAO (2009), Aviation and Climate Change. Report to Congressional Committees. US Government Accountability Office, GAO-09-554, June.

Green, J.E. (2002), *The Aeronautical Journal*, 106, February.

Ihamäki, K. (2009), EU ETS: how ready are the airlines? Presentation to GreenAviation Airline ETS Masterclass, London, 1 June.

International Air Transport Association (IATA) (2007), Financial impact of extending the EU ETS to airlines, 9 January, www.iata.org/economics

IATA (2009), Fact sheet: Global sectoral approach to addressing aviation CO_2 emissions. http://www.iata.org/pressroom/facts_figures/fact_sheets/emissions-approach.htm, viewed on 23 June 2009.

ICAO (2009), GIACC develops action plan to tackle international aviation emissions. International Civil Aviation Organization Press Release PIO 07/09, 3 June.

Jorritsma, P. (2009), Substitution opportunities of High-Speed Train for air transport, *Aerlines*, http://www.aerlines.nl/, 28 July.

Mayeres, I., S. Ochelen and S. Proost (1996), The marginal external costs of urban transport, *Transportation Research Part 1D*, 111-130.

Merrill Lynch (2008), Aviation in EU ETS; an incentive for efficiency, 8 September.

Morrell, P. (2007), An evaluation of possible EU air transport emissions trading scheme allocation methods, *Energy Policy*, 35, 5562-5570.

Morrell, P. and L. Dray (2009), Environmental aspects of fleet turnover, retirement and life cycle. Part of UK Omega Project, Final Report, March.

Oxera (2003), Assessment of the financial impact on airlines of integration into the EU greenhouse gas emissions trading scheme.

Peeters, P.M., J. Middel and A. Hoolhorst (2005), Fuel efficiency of commercial aircraft: an overview of historical and future trends. NLR-CR-2005-669, National Aerospace Research Laboratory NLR, November.

Perl, A., J. Patterson and M. Perez (1997), Pricing aircraft emissions at Lyon-Satolas Airport. *Transportation Research 2D*, 89-105.

Sentance, A. and H. Pulles (2005), The initial allocation of permits at the beginning of each year (benchmarked allocation). Discussion paper for Working Group 5 of the Committee on Aviation Environmental Protection (CAEP), ICAO CAEP 5 – WG5 WP5-5/3.

Stern, N. (2006), The economics of climate change: the Stern Review, HM Treasury, London.

UK Defra (2007), A study to estimate ticket price changes for aviation in the EU ETS. A report for the UK Department for Environment, Food and Rural Affairs Defra) and the Department for Transport, November. London.

UK Defra (2008), A study to estimate the impacts of emissions trading on profits in aviation. A report for the UK Department for Environment, Food and Rural Affairs Defra) and the Department for Transport, January. London.

Woodmansey, B.G. and J.G. Patterson (1994), New Methodology for Modelling Annual-Aircraft Emissions at Airports, *Journal of Transportation Engineering*, 120, 339-357.

THE CONTRIBUTION OF STRATEGIC ENVIRONMENTAL ASSESSMENT TO TRANSPORT POLICY GOVERNANCE

Rodrigo JILIBERTO HERRERA

TAU Consultora Ambiental
Madrid – Santiago de Chili - Bogota
Spain - Chili - Colombia

SUMMARY

1. INTRODUCTION

Strategic Environmental Assessment (SEA), understood as a practice that aims to incorporate the environmental dimension into strategic decisions such as policies, plans and programmes, already has a substantial tradition. According to Dalal-Clayton and Sadler, the formalization phase of this instrument began in the early 1990s, and its international dissemination can be said to have started in 2001 (Dalal-Clayton, Sadler, 2005).

The first books and special editions of international journals devoted to this topic date from the first half of the 1990s. European Directive 2001/42/EC on the Assessment of the Effects of Certain Plans and Programmes on the Environment, better known as the Strategic Environmental Assessment Directive, is clearly a milestone, because, since 2004, it has required the 27 members of the European Union to submit a long list of plans and programmes to an SEA procedure (EU, 2001).

In keeping with that general trend, SEA has been increasingly applied to policies, plans and programmes in the transport sector. Over 10 years ago in 1998, the European Conference of Transport Ministers (ECMT) published an initial volume on the topic of SEA and the transport sector (ECMT, 1998).

In 1999, the OECD and the ECMT organised a joint conference on SEA, which produced a publication entitled *Strategic Environmental Assessment for Transport* (ECMT, 2000), partly revising the earlier volume and adding the conclusions of the conference. In those years the European Environment Agency also published results of what it called the “Spatial and Ecological Assessment of the European Transport Network (ETN)” (EEA, 1998) an exercise carried out just at the right moment to understand the contribution that SEA could make towards a strategic European effort such as ETN. In 2000, the European Commission published the results of the study on the application of SEA specifically in the transport sector (EC, 2000), and in 2001 it published another on the use of SEA in transport corridors (EC, 2001).

Consequently, and to follow up work in this area, in 2001 a major workshop was held in Finland, sponsored by that country's Environment Ministry. The conference emblem of “Transport Planning: Does the influence of Strategic Environmental Assessment/Integrated Assessment Reach Decision Making?” betrayed the SEA community's early concerns about the instrument's effectiveness (Furman, Hildén, 2001).

The literature also contains many and varied reports of SEAs applied to specific transport plans and initiatives such as transport corridors (EC, 2001; Finnish Ministry of the Environment, 2001); and the first volume devoted specifically to the topic of SEA and transport planning and land use was published in 2002 (Fischer, 2002).

Since then, it can be said that the use of SEA in the transport sector has become widespread, not only in Europe, but also among OECD countries and in Asia, and to some extent in Latin America too (World Bank, 2006). This process has been accompanied by an expansion of practical applications of SEA in the transport sector (Dalal-Clayton and Sadler 2005), and the publication of various specific guides to SEA the transport sector (EC, 2005; Department of Transport, 2004).

At the turn of the new millennium, therefore, the initial work on SEA and transport are eliciting work to evaluate the interest, specifics and feasibility of applying this new tool when formulating transport policy; and a positive appreciation is consolidating of its use and relevance as a tool to support decision-making in this sector.

This initial positive assessment is responsible for increasing use of SEA in the design of transport plans and programmes, and a wide ranging analytical toolkit has been developed to adapt to the specifics of the relation between transport planning and the environment – in terms of its main environmental effects, the scales of planning work, the diversity of planning models and the typology of strategic transport decisions. Rather than considering the singularity or specific nature of SEA as applied to decision-making on transport policy, subsequent developments have sought to facilitate and promote the use of this tool by disseminating specific cases or producing guides.

This relatively strong development of SEA in the transport sector does not, however, mean that it is free from controversy and ambiguities, because, as shown in the literature (Dalal-Clayton and Sadler, 2005) and by the international SEA community (Wallington *et al.*, 2007, 2008), there is still an ongoing debate on key aspects of SEA, including the definition of its basic objectives.

Accordingly, its application to the transport sector is also not free from sometimes substantive problems, since the issues under discussion in land-use planning are the same as in a transport plan. Past practical experience of SEA in the transport sector, together with that to be gained in the coming years, will likely be judged in the light of the solutions they provide to the conceptual problems that are still unresolved in SEA; and it is perhaps too early to classify them as good or less good practices.

In this context, it probably does not make much sense to perform a more in-depth analysis of the various methodologies proposed for the SEA of transport decisions, with their respective phases and steps, or in the analytical tools used in numerous specific SEAs, such as geographic information systems (GIS), expert panels, linkages with land use planning and transport system modelling, among others, because it has already been clearly demonstrated in practice that an SEA procedure can be applied to a transport plan or programme.

What might be more interesting is to consider the result of those exercises, i.e. SEA's contribution to effectively improving transport policies. This raises two very closely related issues, the first of which is the instrument's effectiveness, already discussed at the aforementioned workshop in Finland in 2001, i.e. its supposed capacity to affect the decision-making process. The second issue involves clarifying basic conceptual issues relating to SEA's nature and ultimate objectives, because the type of contribution that SEA can be expected to make to better strategic decision-making, the effectiveness of which is being questioned, is heavily dependent on this. In other words, it is idle to question the effectiveness of SEA in improving or influencing strategic decisions, if the ultimate direction of such influence, which is defined by the nature and objectives of the instrument, is still a matter for discussion.

Practical experience of SEA clearly suggests, and several studies show (Finnish Ministry of the Environment, 2001; EC, 2009), that SEA always influences the decision-making process. It is almost inconceivable that it could be otherwise, because by implementing an evaluation process that interacts with the central decision-making process, the mere formal act of evaluation alters it.

That does not mean that SEA is efficient, however. To identify whether or not the influence obtained is what was being sought requires clarity of aims, and these are still under debate. So much so, that one frequently reads in the literature statements like “SEA is best described as an evolving family of tools” (World Bank, 2005, p. 1), or “SEA can be described as a family of approaches which use a variety of tools, rather than a single, fixed and prescriptive approach" (OECD, 2007, p. 15), or a longer quote from Sadler, "SEA is understood to be a generic process or approach that encompasses a family of instruments, which may have different names and features but are functionally related by common aim of integrating environmental considerations into the higher levels of decision-making" (Sadler, 2008, p. 14). All of this shows that it is hard to know nowadays what SEA really is and what its precise aims are, and hence whether it is effective in achieving them.

In any event, the ultimate aim of SEA is to contribute to environmental improvement, and in many cases also to the sustainability (Partidário, 2007; Wallington *et al.*, 2007), of strategic decisions, including those in the transport sector. This requires ongoing research, through practice and beyond, to establish a theoretical and conceptual framework that provides a solid underpinning for SEA intervention in the development of public policies. This will make it possible to say rather more about its contribution to improving public policies than that it involves a "variety of tools applied to those public policies."

Accordingly, this article proposes a frame of reference for SEA to clarify the real contribution this tool can make to the effectiveness of public policy generally and transport policies in particular. This obviously means better incorporation of their environmental and sustainability dimensions, as well stronger public policy governance, particularly in the transport sector.

2. SEA: AN EVOLVING WORK IN PROGRESS

The SEA literature has insistently drawn attention to the lack of a precise definition for SEA and its objectives (Brown, Therivel, 2000; World Bank, 2005; Dalal-Clayton and Sadler, 2005; Wallington *et al.*, 2007). In practice this has meant that SEA appears in a very wide range of forms, causing it to be viewed as "an evolving family of tools".

This situation has also spawned various schemes for classifying the SEAs undertaken in practice, depending either on the objectives, approaches and techniques used (Partidário, 2000; Dalal-Clayton and Sadler, 2005; Sadler, 2008; Bina, 2008), or else on the conceptual frameworks used to classify the different varieties of SEA (Wallington *et al.*, 2008).

The interesting thing is that this process of conceptual evolution, which began in the second half of the 1990s, is still continuing. In other words, both the conceptual evolution of SEA and the schemes or models that aim to classify the varieties of applications are changing through time, and continue to evolve. In some cases scholars have put forward certain definitions and schemes, only later to suggest different ones. See, for example, the change of SEA types suggested by Dalal-Clayton and Sadler in 2005, and modified by Sadler in 2008 (Dalal-Clayton and Sadler, 2005; Sadler, 2008).

This suggests, firstly, that it is hard to conceptually summarize the complex practice of SEA; and secondly that there is an urgent need for conceptualisation to give direction to that practice. Having said that, it is worth noting that in addition to evolution and diversification there has also been also [1]progress; in other words its evolution shows a line of development, a direction.

The earliest definitions of SEA were strongly rooted in the concepts of project environmental impact assessment (EIA) (Dalal-Clayton and Sadler, 2005; Bina, 2007). As been noted elsewhere (Jiliberto, 2007) many of the early definitions of SEA (Therivel *et al.*, 1992; Therivel and Partidário, 1996), including that of Sadler and Verheem, often cited in the literature, put analysis of the environmental consequences of decisions at the heart of SEA. "*The strategic environmental assessment is the systemic process of studying and anticipating the environmental consequences of proposed initiatives at high level decision-making. The purpose of the process is to incorporate the environmental criteria from the beginning, as an element of decision in all the sectors and degrees of planning, placing it at the same level as the economical and social criteria* (Sadler and Verheem, 1996)."

As time has passed, the positivist consequentialism of SEA has been eroded.[2] For example, it is interesting to observe the evolution shown by specific authors, such as Verheem, who in 1996 initially argued that SEA had the aim of anticipating the environmental consequences of decisions, but in 2000 claimed that the role of SEA was to "*... strengthen the role of environmental issues in the strategic decision* (Verheem and Yonk, 2000)", and then in 2005 argued that "*SEA is a tool for including environmental consideration into policies, plans and programmes at the earliest stages of decision making* (Ahmed, Mercier, Verheem, 2005)."

This same evolution can be detected in other widely published authors from the SEA community, such as Partidário, who as early as 1996 argued with Therivel that the aim of SEA was to incorporate the environmental effects in policies, plans and programmes (Partidário and Therivel, 1996), but now says that "Strategic Environmental Assessment (SEA) is an impact assessment tool that is strategic in nature and has the objective of facilitating environmental integration and the assessment of the opportunities and risks of strategic actions in a sustainable development framework." (Partidário, 2007, p. 12), positing the need to develop what she refers to a strategic SEA.

At heart, this evolution involves a weakening of the environmental impact concept as the core of SEA, as much more bland or vague concepts take centre stage, such as environmental effects, environmental issues, environmental aspects, and so forth. This firstly shows how difficult it is to make operational use of a positivist-consequentialist concept in the environmental assessment of policies, plans and programmes. It also confirms the difficulty, demonstrated in SEA practice, of gaining a reasonable understanding of the consequences of decisions and making those consequences the focus of improving decision design. Highly illuminating in this regard is Bina's appraisal of the model changes operating in SEA. "*An important aspect of this change is the demotion of prediction and evaluation (…) in favour of a wider range of activities* (Bina, 2008, p. 114)."

This entails a sharp change of direction in the rationale on which the assessment is founded, since it is now impossible to base it strictly on a substantive, objective rationale, founded on knowledge of the likely material consequences of decisions.

Later definitions have gradually stressed that the aim of SEA is neither exclusively nor primarily to incorporate the consequences of decisions into decision-making processes, but to improve those processes themselves, clearly from an environmental perspective (Brown and Therivel, 2000; Jiliberto, 2002 and 2004; Caratti *et al.*, 2004, Bina, 2007; UNDP, 2004). The World Bank definition of SEA as "a participative approach to place the environmental and social aspects in the centre of the decision-making process and to influence in the development planning, the decision-making and the implementation processes at a strategic level" (Mercier, 2004; World Bank, 2005) is relevant in this regard.

This conceptual evolution shows that SEA is no longer primarily seen as a tool whose main function is to provide technically based information on the material environmental consequences of a decision (Bina, 2007). Obviously this has not occurred by chance, but reflects the dysfunctionality involved in turning the effective, albeit simple, technical and procedural baggage of project environmental impact assessment, towards evaluating complex sociopolitical-technical entities, such as policies, plans and programmes.

One consequence of this initial break has been the emergence of a strong methodological trend that believes SEA should focus on the decision-making process (Caratti *et al.*, 2004; Partidário, 2007; Kornov, Thissen, 2000). This envisages SEA more as a tool of strategic and proactive interaction with the decision-making process, rather than being used to report on its generally negative environmental consequences (Bina, 2007, 2008).

The break with the substantive-positivist-consequentialist rationale of SEA gave rise to a proposal for an evaluation based on a procedural rationale. If it is impossible to improve the decision on the basis of its positively identified consequences, then it can be done by improving the procedure through which the decision is reached. The substantive rationale underlying the assessment is thus replaced by a procedural rationale. This was essentially the very conscious proposal of the European ANSEA research project (Caratti *et al.*, 2004; Dalkmann *et al.*, 2004).

Despite the conceptual rigour and abundant empirical material provided by the proposal for an SEA centred on the decision-making process, and the fact that it has not been consistently contested or criticized in the literature (Nilsson *et al.*, 2009), in practice this has not prevented the conventional environmental-impact-centred view of SEA from continuing to represent mainstream practice. This is largely explained by the ambivalent message emitted by most recent legislation on the subject, the evolution of which has failed to keep pace with the conceptual debate. Other influences include methodological inertia and cultural practices, both in terms of environmental assessment and planning, which change slowly, and because new approaches need to be tested in practice.

None of this means, however, that the evolution of SEA is haphazard or directionless, and merely a proliferation of possible content alternatives. The first step in SEA evolution involves overcoming the consequentialist evaluation paradigm focused on the positive effects and outcome of the decision, to move towards an evaluation focusing on the decision-making process and improving its quality from an environmental standpoint.

But the process has not stopped there. SEA centred on the decision-making process, with decision-making at its centre, is forced to delve into the complex world of decision theory and policy analysis, since it must show the extent to which strategic decision-making can be consistent with environmental assessment methodology, founded ultimately on objective or procedural rationality criteria. This is particularly critical bearing in mind that decision theory and policy analysis generally posit the opposite, namely the absence of a rational decision-making model in public policies, which tend to be dominated by models or rationales that are outside the archetypal substantive rational model (Kornov, Thissen, 2000; Dalkmann, Nilssen, 2001).

The results of this theoretical research, and SEA practice itself, have led to another change of course in the conceptualisation of SEA. At the heart of this turning point is a growing conviction that the strategic environmental assessment model, even when centred on decision-making, is based on a technical rationality paradigm (objective or procedural) that is inconsistent with the decision-making nature of strategic decision processes.

The basic line of argument is that both the SEA model based on environmental impacts, and that centred on the decision-making process, assume that the decision process being evaluated has a number of properties that in fact it lacks: substantive rationality, in the sense of adapting ends and means; procedural rationality, in the sense of following a rationally grounded set of steps; and a rational subject, in the sense of an identifiable entity that assumes those rationalities (Wallington *et al.*, 2008; World Bank, 2005). Both extensive proven experience in policy analysis, and much of SEA practice, would confirm that those assumptions are not valid (Kornov and Thissen, 2000).

Both SEA models are ultimately based on the assumption that the mere contribution of technical-rational information would have a positive influence on the decision-making process, environmentally speaking (Bina, 2008). This view simplifies the real nature of strategic decision-making processes, which are sometimes affected not only by complex settings, but also significant levels of uncertainty, value conflicts, power relations that are mostly asymmetric, negotiations, networks, political culture, not forgetting the interplay of political forces that occurs between stakeholders in the evaluation framework.

Alongside this conceptual evolution, SEA practice has identified the important role played by context in the quality of the results obtained from an SEA process. As early as 2001 the workshop in Finland on SEA in the transport sector systemized a number of contextual factors that were decisive for the effectiveness of SEA in influencing decision processes; i.e. the success of SEA depends heavily on the setting in which it occurs. There have since been more elaborate studies on this point (Hilding-Rydevik and Bjarnadóttir**,** 2007; Wallington *et al.*, 2008; Fischer, 2005); and it is also present in the insistent warnings made in SEA guides regarding the supposed uniqueness of each SEA, determined by a supposedly unique context (Jiliberto and Bonilla, 2009; Wallington *et al.*, 2007; Ahmed, Mercier, Verheem, 2005; Verheem, 2000) calling for flexible design.

“Context” is understood here in a broad sense, ranging from the type, scale, and function of the evaluated decision, through the political-institutional setting and its priorities, the power structure and its rules, and the planning culture, to the deliberative or more technocratic tradition of planning itself.

Concern for context reflects the singular fact that in SEA, unlike many other public policy formulation support tools, contextual factors are so important that in practice they determine what each SEA can become as a tool supporting the formulation of a strategic decision. In fact, therefore, contextual factors cease to be a backdrop but become elements of SEA itself.

A very reasonable explanation of the importance of context in applying SEA is that the dysfunctionality of using technical-rational evaluation models in decision-making contexts that do not behave according to those rationales, appears as an over-determination of context, simply because the medium in which the tool is being applied is not consistent with it.

For the assessor who thinks his tool is appropriate, this dysfunctionality does not appear as shortcoming of the tool itself, but as a “confused” feature of the setting which makes its application complex. What happens in fact is that the non-rational decision-making rationales of the setting are imposed over the desired rationalisation of the technical-rational evaluation model, thereby preventing it from adopting a standard universal model.

These two considerations, which have emerged in the most recent SEA literature, need to be taken seriously, because it would be a big mistake to believe that SEA can change the decision-making rationales that dominate public policy-making processes. Instead the tool needs to be thought and rethought to adapt it creatively to its application setting and reinterpret it in its function.

The new twist in the discussion on SEA suggests, therefore, that SEA models based on the impact concept, and those focused on the decision, are founded on a technical-rational evaluation paradigm (substantive or procedural rationality) which would have to be overcome to be functional and effective. This involves not only moving from evaluating a product to assessing a process, but also overcoming the technical rational evaluation model aimed basically at providing rationally grounded information for decision-making. And, if it is consistent with the analysis undertaken, to propose a tool that is consistent with the decision-making rationales of the setting in which it is being applied.

3. GOVERNANCE AND SEA

At this point, one can consider the relation between SEA and governance, and what this tool can contribute to the governance of transport policies.

Governance is a relatively new concept that tends to be understood in various ways, so at least a minimum reference is needed to be able to use it. The United Nations offers the following definition: “Governance is the system of values, policies and institutions by which a society manages its economic, political and social affairs through interactions within and among the state, civil society and private sector. It is the way a society organises itself to make and implement decisions–achieving mutual understanding, agreement and action. It comprises the mechanisms and processes for citizens and groups to articulate their interests, mediate their differences and exercise their legal rights and obligations. It is the rules, institutions and practices that set limits and provide incentives for individuals, organisations and firms. Governance, including its social, political and economic dimensions, operates at every level of human enterprise, be it the household, village, municipality, nation, region or globe.” (UNDP, 2000)

A more concise definition that relates directly to the topic we are dealing with is the following: Governance “…is about how governments and other social organisations interact, how they relate to citizens, and how decisions are taken in a complex world. Thus governance is a process whereby societies or organisations make their important decisions, determine whom they involve in the process and how they render account” (Graham, Amos and Plumptre, 2003).

At the heart of governance is the way human groups take strategic decisions about the direction of development and each individual's roles in it, and how these are implemented and held accountable. The central component of governance is clearly decision-making on strategic aspects of development of the human group.

Viewed in this way, it can be said that SEA is at the heart of the governance of our current societies; in particular the governance of public policies, and naturally the governance of transport policies. This is because SEA is simply a tool to support strategic decision-making, whose ultimate purpose is to adequately incorporate the environmental values that society holds at a given time into strategic decision-making (Jiliberto, 2002).

In other words, SEA occupies the same space as governance (i.e. strategic decision-making), and its purpose is fully consistent with the principles of good governance. SEA helps to improve the legitimacy of strategic decisions and broadens the range of actors participating in them; it promotes a strategic view in recognising society's environmental values; and it helps to improve the quality and accountability of their decisions, while respecting the current legal framework and equality for all parties in disputes – all of which are recognised principles of good governance (UNDP, 2000). As noted in the European White Paper on Governance, “Five principles underpin good governance and the changes proposed in this White Paper: Openness, participation, accountability, effectiveness and coherence” (EC, 2001). SEA can make a specific contribution to each of these.

Clearly, SEA does not encompass the whole of governance, because, while nowadays it tends to incorporate more values than just environmental ones, assuming a sustainability perspective, in principle it is limited to these; and in particular, because its contribution to best governance practices is made chiefly at the time of evaluating the decision, and makes that evaluation much more thorough.

SEA has potential to improve governance because it systematically questions the environmental quality of strategic decisions. By analysing the consistency of a decision's strategic choices, transport decisions can be questioned in terms of their effective contribution to a sustainable transport model; or strategic transport alternatives can be examined, such as demand management, incentives for public transport, non-motorised forms of transport; or how the decision favours intermodality, or the capacity to coordinate transport planning with urban development and land management.

SEA favours opening up strategic decision-making to a number of social stakeholders who were traditionally excluded from such processes, by systematically building civil society participation mechanisms into the evaluation process.

This is no different in the case of transport policies, although the special complexity of transport policy governance needs to be recognised, since it is permeated by a dense network of interests, institutions, sectoral policies and processes.

This potential of SEA to help strengthen public policy governance depends on how SEA is understood and applied, as commented on in previous chapters.

Experience in this regard is ambiguous or ambivalent, since no specific evaluation has been made on the subject. Nonetheless, the excessive importance of the setting in which SEA is applied, as noted above, suggests the difficulty that SEA has faced in adequately adapting to the processes through which sector policy governance takes place.

The technical-rational SEA models applied mostly thus far, each with its own specific features, do not make it easy to develop their governance potential. This is basically because they force the central actors of the procedure – e.g. the promoter of the policy, plan or programme being evaluated, on the one hand, and the environmental authority on the other – to adopt opposing strategic positions on SEA from the outset, which obviously makes it hard to generate a dialogue that would favour the sector governance process.

Moreover, it is precisely the supposedly technical-rational basis of the evaluation that is responsible for this, since by taking for granted that the evaluation has an “objective” foundation and that its role is only to reveal, the parties adopt entrenched positions based on a principle, in response to the threat that such “revealed objective information” may be against their interests.

Thus, each party feels *a priori* threatened by the supposed technical-objective, and hence irrefutable, arguments that the other party may raise. Before the process begins, this generates a strategic positioning that is clearly contrary to the rationales of good governance.

This phenomenon is greater precisely because of the absence of a grounded technical-rational basis for the evaluation, since it clearly leaves a lot of room for discretion. A good example of this is the outcome of the application of the European SEA Directive, which has placed an environmental report at the centre of a procedure whose key objective is to generate supposedly "objective and technically based" information on the potential environmental impacts of the plans and programmes evaluated (EC, 2001)

Experience of SEA shows that real technical difficulties have meant that the potential environmental impacts of plans and programmes evaluated have gradually ceased to hold the centre ground in SEA (Bina 2007). On this point, it is worth paying special attention to the conclusion reached by the European Commission's evaluation of the application of the European SEA Directive in terms of predicting impacts: "The lack of methodology to predict impacts has been mentioned as a key problem" (EC, 2009). Thus the regulation's demand to focus the procedure on an assumption that is hard to fulfil objectively, heightens the risk for each party that the other will raise arbitrary arguments in the SEA process that will be difficult to refute in a "reasoned" debate. This favours strategic positioning and mistrust rather than dialogue and co-operation.

4. A REINTERPRETATION OF SEA IN THE INSTITUTIONAL AND GOVERNANCE SETTING

Just as the criticism of the environmental-impact-centred SEA model put the decision-making process at the core of the evaluation, improvement of which became the goal of SEA, the critique of the technical-rational models of SEA again changes the locus of the evaluation; and the new locus is none other than dialogue and negotiation (Wallington *et al.*, 2007).

The conceptual shift of SEA has followed an almost scholarly logic. The initial SEA models were based on a rationality that decision theory would classify as substantive, and pertaining to the economic concept of rational choice (Dalkmann, Nilsson, 2001). The goal of SEA is to provide information on the environmental consequences of alternative choices, which will then be used to optimally adjust means to ends.

Given that this is “technically” impossible, then a decision-based SEA model is proposed with an essentially procedural rationale (Simon, 1987). If it is not feasible to find an optimal solution, then what needs to be improved is the process that generates the solution, whatever that may be (Dalkmann *et al.*, 2004).

As the limits of both approaches were clearly revealed through the antibodies generated by technical-rational approaches in the decision-making setting that are inconsistent with those rationality models, alternative approaches to SEA are put forward that show that the contribution of this instrument to better decision-making requires a focus on aspects such as deliberation, dialogue, negotiation, cooperation, institutionality and governance (Wallington *et al.*, 2007; Bina, 2007; World Bank, 2005). In other words, it is proposed to move from an SEA based on a substantive or procedural rationale to one based on a deliberative rationale (Habermas, 1997). The significance of this rationality proposal is that the decision is good because it is the outcome of a deliberation process, given that it was impossible to improve it based on objective information and/or procedural prescriptions.

The logics of deliberation, dialogue and negotiation are much closer to the rationality models that actually govern the strategic decision-making process. In this way, SEA would be able to more smoothly match the process or object being evaluated, by creating the real possibility of influencing strategic decision-making processes.

Nonetheless, when SEA is deprived of substantive or procedural content, a vacuum is created in terms of what it is intended to achieve, either specifically or substantively. If SEA is no longer a matter of incorporating data on the consequences of policy choices, or improving the process through which such choices are generated and selected, then what is its purpose? Clearly one cannot claim that setting up a dialogue mechanism will produce nothing very specific, except for the dialogue *per se*. So, is the aim of SEA to talk about incorporating the environmental dimension into strategic decisions, without worrying about the outcome of the dialogue process?

Those who have called for SEA to move in a deliberative direction (Wallington *et al.*, 2007) answer this question by arguing that the ultimate purpose of SEA would instead be to induce a learning process that enables decision-making processes to be gradually permeated by a ecological rationale (Bina, 2007).

In other words, individual SEAs would not have specific instrumental objectives, but the SEA process as a whole would aim to catalyze learning, the subject of which is the broad socio-institutional governance space in which a society's strategic decisions are taken. Each unique SEA has one meta-objective, at most.

This solution to the dilemma of the purpose of SEA in a deliberative model raises two clear problems. The first is having to show that no other instrument is better at achieving the meta-objective, by directly addressing the supposed under-representation of the ecological rationale in the strategic decision-making process. In terms of public policy efficiency it is hard to argue that, to achieve the objective of introducing ecological rationale logics in strategic decision-making processes, one does not use an instrument that acts directly on the desired objective, e.g. the process of providing training for entities responsible for the decisions in question.

The second problem with this solution to the dilemma of the purpose of a deliberative SEA model is having to show the effectiveness of the deliberative activity as such in increasing the level of ecological rationality in the strategic decision-making process. Moreover, backing an instrument whose effectiveness in achieving the meta-objective is practically indeterminate is also debatable in public policy terms.

All of this suggests that, while it is reasonable to argue that a deliberative SEA model, and also one of a technical-rational nature, whether substantive or procedural, helps to introduce an ecological rationality assumption into the strategic decision-making process, this should be seen as a by-product – a positive externality created by the process that tends towards a substantive achievement that should be direct and perfectly verifiable whenever the SEA tool is applied.

A move towards identifying the purpose of SEA in a deliberative model requires returning simply to the original purposes of the instrument, i.e. to incorporate the environmental dimension into strategic decisions. The substantive or procedural rationality approaches to SEA directly identify the supposedly incremental dimension of strategic decisions, and thus clearly specify the substantive direct objective of each SEA, namely to ensure that what is identified as the strategic environmental dimension is incorporated into the decision.

What SEA practice has precisely shown is its relative ineffectiveness in achieving those objectives; and for that reason it has evolved towards a deliberative model. But what the need for a deliberative model calls into question is not the substantive content of incorporating the environmental dimension into strategic decisions, but how to do so. In principle, a technical-rational approach would not be consistent with the rationales of the political settings in which these decisions are made.

Nonetheless, this does not mean *a priori* that what technical-rational models put on the table as the environmental dimension of strategic decisions is not in fact the environmental dimension of those decisions. Moreover, it is entirely reasonable to believe that the environmental dimension of strategic decisions may only be expressible in technical-rational terms.

The problem lies exclusively in the fact that, given the rationale that dominates strategic decision-making, a procedure for incorporating that environmental dimension through an evaluation process that prioritises technical-rational aspects over and above deliberative ones is unthinkable, since the logic of that setting is precisely to negotiate between many and varied discourses and interests, all of which have some technical-rational foundation.

It is therefore not true to say that the technical-rational discourse has no place in SEA. What cannot happen is that the incorporation of the environmental dimension becomes confused with the presumptive existence of a uniquely valid rational technical discourse. Hence the importance of deliberation as the hub, not of the content of SEA, but of its rationale. In the deliberative SEA model, one can claim that it is rational, and therefore good practice, to incorporate an environmental dimension into a strategic decision, as the outcome of open dialogue between different technical-rational discourses on the strategic environmental dimension of the decision in question.

Accordingly, each SEA based on a deliberative model does not have a meta-objective, but a highly practical and verifiable one, namely, reaching consensus on what, technically and rationally, the decision's key stakeholders understand by incorporating the environmental dimension into it.

This raises the possibility that the environmental dimension of a strategic decision is indeterminate, although the possibility of defining it in a rule-governed dialogue and negotiation process is not.

SEA would in practice become an institutional mechanism for clarifying the responsibility pertaining to strategic public decisions on what tends to be the institutional mandate of environmental conservation and protection and the promotion of sustainable development.

That responsibility is clearly complex and diffuse, and, above all, ultimately has an institutional scope since its identification assigns specific institutional responsibilities. Moreover, the mechanism through which it is clarified and established as another piece of the institutional machinery of environmental management is equally complex, and governed first and foremost by a deliberative logic that forms the basis of our societies' institutional arrangements, and secondly, by a technical and rationally grounded deliberation, which is the another of the pillars on which the modern institutional framework is based (Faludi, 1987).

SEA would thus have an institutional objective of defining, on a case-by-case basis, the responsibility of each evaluated strategic decision in upholding the constitutional principle of protecting the environment and promoting sustainable development. This is a clear goal that can be verified for each case. To achieve it, a rule-based institutional mechanism is defined, founded on dialogue and negotiation, which must provide a technically and rationally grounded (i.e. not arbitrary) result, giving the decision maker guidance on what it means, in the specific case, to incorporate the environmental dimension into the decision, and thus fulfil its institutional responsibility.

This view of SEA speaks more than any other to the concept of governance, and becomes a mechanism for promoting the principles of good governance, since it encourages practices of dialogue, transparency, consensus, mutual respect, by genuinely fostering the development of shared visions of the sustainability of key sector decisions.

5. THE ENVIRONMENTAL DIMENSION OF STRATEGIC DECISIONS

Nonetheless, the environmental dimension of strategic decisions remains at the heart of SEA. As noted above, this cannot necessarily be captured by a single technical-rational representation that could therefore be classified as objective; and this situation gives rise to a deliberative SEA model.

That does not prevent a proposal being made on the specific nature of the environmental dimension, however; although, in the context of the discourse developed thus far, it can only be seen as an approximation whose sole virtue is to structure a definition of the environmental dimension of strategic decisions for the purpose of debate.

Firstly, it is worth clarifying that the concept of the environmental dimension of a decision at the time of its environmental assessment is understood as any aspect that is environmentally relevant at the time the decision is taken.

These do not have to be substantive environmental aspects only. For example, it is environmentally relevant to consider alternatives, even though that aspect may not be considered environmentally substantive. Secondly, not all the environmental aspects of a decision have to be relevant at the time of its environmental assessment, but only those that form part of the decision at that time. For example, a project's environmental management systems do not form part of its environmental assessment.

In other words, the environmental dimension of a decision being evaluated is not one of its universally objective characteristics, but aspects pertaining to it that are functional at the phase of the decision in question, in this case its evaluation.

Moreover, SEA needs to be targeted on the strategic environmental dimension of strategic decisions (Bina, 2007; Partidário, 2007). What makes SEA strategic is not that it evaluates strategic decisions, but that it focuses on the strategic aspects of the evaluated decision (Jiliberto, 2007).

This is consistent with the strategic nature of the decision being evaluated; but it is also consistent with the phase of the decision in which the evaluation is made. This generally occurs at an early stage in a long decision process, which ranges from strategic phases until what is strategically decided upon becomes an effective intervention that could have a material effect on the environment.

As this is an early phase, the decision handles aggregate information at a low level of detail, so it is inappropriate for SEA to target the “non-strategic” environmental dimension of the decision. Not doing so is a leading cause of the shortcomings in impact prediction displayed by SEA (EC, 2009).

Lastly, this is consistent with the ex-ante nature of SEA, which should begin before the decision starts to take shape, as it were; this makes it impossible to focus on operational and design aspects, or its ultimate environmental effects.

The environmental dimension of strategic decisions, like their environmental assessment process, is unique and complex, and necessarily differs from project environmental assessment.

This uniqueness and complexity reflects the fact that the environmental dimension of strategic decisions is three-dimensional, because their evaluation process, as we have been explaining, also is. It has a substantive dimension, as well as a procedural one and a deliberative one.

The first dimension is the substantive one. This dimension of the evaluation concerns how it takes account of the consequence of the decision for the "environmental state of things" that the decision is intended to affect. The substantive dimension thus relates to how one expects the evaluation to cause an improvement in that environmental state of things. In SEA, this dimension has traditionally been seen in a positivist-consequentialist way, i.e. using the concept of environmental impact or effect, similarly to how it was treated in project environmental impact assessments (EU, 2001).

Secondly there is a procedural dimension. This is a new evaluation dimension arising in the case of SEA as an *ex ante* or process evaluation; SEA does not have a product to evaluate, but a process to contribute to. That makes the procedural aspects of the decision-making process important elements of the evaluation, given its environmental reach (Caratti *et al.*, 2004). SEA now not only considers the substantive consequential aspects of the decision, but also its construction and its environmental scope. Merely instrumental aspects such as the information used, the alternatives considered, definition of objectives, etc., now become part of the environmental dimension of the decision being evaluated.

The third dimension is the deliberative one. This new dimension of environmental assessment stems from the structural indeterminacy of the two previous dimensions and the institutional political setting in which the evaluation takes place. As the evaluation has an institutional function in a context of diverse technical-rational discourses on the substantive and procedural dimensions of the decision being evaluated, its ultimate environmental dimension will be the outcome of the unrestricted deliberation process that SEA facilitates. Its contribution to the environmental governance and sustainability of policies will thus depend on the quality of the deliberative processes.

According to the latest SEA developments, any SEA should include these three dimensions and consider how to take each one into account.

5.1. The substantive environmental dimension of strategic decisions

Clearly the most conflictive of these three dimensions is the substantive one, since it engages deeply held beliefs about the contribution that SEA can make to ensuring the decision improves the "environmental state of things" that it is intended to affect. For that reason it warrants a more in-depth treatment. Obviously, there is no formal consensus in referring to this dimension as substantive, since it is an ad hoc classification.

In the European SEA Directive it is clear that the substantive dimension of SEA relates to the concept of impact, since the aim is to ensure that the decision's environmental impact improves/conserves the "environmental state of things" that it is intended to affect (EU, 2001). As noted above, this approach can be classified as positive consequentialism. If the decision has known positive material effects on the environment, then the decision improves the "environmental state of things" that it affects.

In the case of the "strategic SEA" proposal put forward by Partidário, what is substantive is the contribution of SEA to the construction of the decision, i.e. the definition of policy strategies that give rise to SEA's contribution to ensuring that the decision improves the "environmental state of things" that it is intended to affect (Partidário, 2007). By helping to identify new and more sustainable decision strategies, a decision is generated that improves the "environmental state of things" affected by it. In this case, the contribution made by SEA is not based on any type of consequentialism.

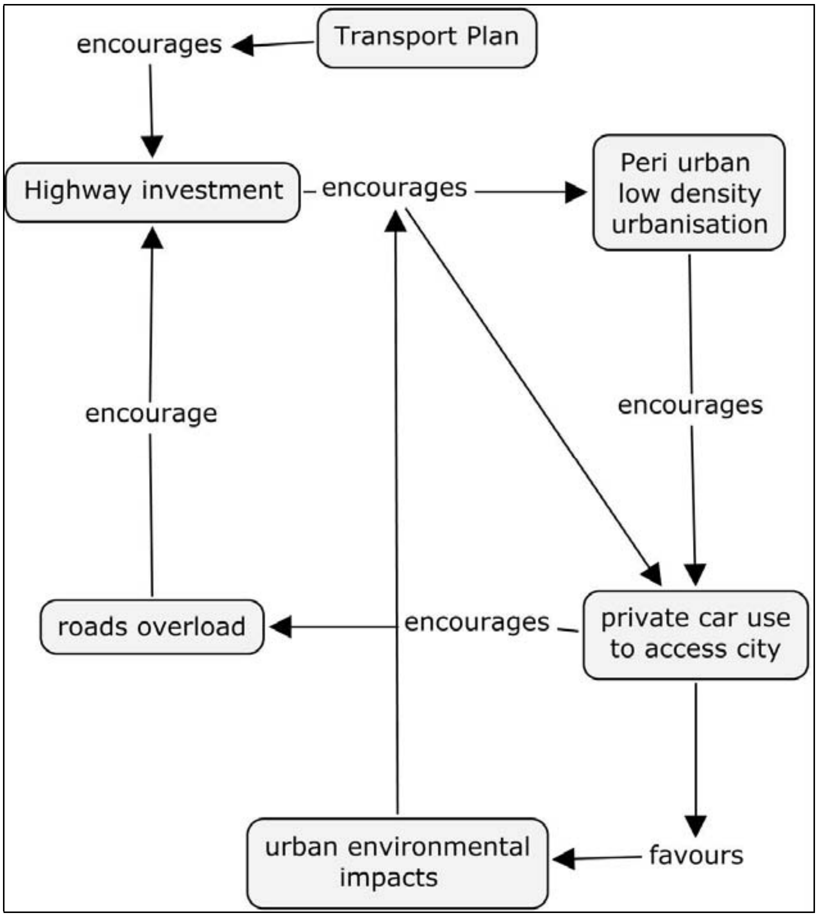
In the implicit proposal contained in an SEA guide developed by the South African Environmental Affairs and Tourism Department, SEA is expected to help the decision improve the "environmental state of things" that it affects, by helping to identify opportunities and constraints on sustainability at a strategic level, and verifying whether those constraints and opportunities, along with other recommendations, are taken into account in the policy options considered (Audouin, Lochner, 2000). Here again, the contribution of SEA is not based on any type of consequentialism.

The European research project ANSEA developed a proposal for a procedural SEA model, based on a radical critique of the belief that the impact concept could become the hub of the substantive environmental dimension of strategic decisions (Caratti *et al.*, 2004; Dalkman *et al.*, 2004). That procedural proposal did not explicitly consider how an SEA developed under those assumptions would improve the environmental state of things that the evaluated decision aims to affect. But this did not mean that that dimension was absent, because the implication was that the application of environmentally relevant procedural criteria would lead the decision, by itself, to generate an option that improved or conserved that "environmental state of things". Here once more, the contribution of SEA is not based on any type of consequentialism.

Since then, part of the ANSEA project team has developed a proposal for re-interpreting the substantive environmental dimension of strategic decisions, by providing a systemic description of it (Jiliberto, 2007). In this approach to the substantive environmental dimension, the relevant issue is not whether the sector activities involved in a strategic decision, such as road building or the transportation of passengers and freight, will generate greater or lesser environmental impact in the future, but whether the policy, plan or programme has taken account of the structural pattern that explains how the environmental profile of the sector as a whole is produced and reproduced.

At the centre of this pattern, which is referred to as systemic because it is recursive, are "sector environmental dynamics". An example of a sector environmental dynamic in the hypothetical domain of transport planning might be the so-called "vicious circle of infrastructures". This can be described schematically as follows: investment in roads to provide better access to the outlying areas of metropolitan cities encourages low-density urban development; this stimulates private car use which impacts on the urban environment by increasing traffic flows in the city, and also causes vehicle congestion and overloads highway infrastructures, which again generates the need to build new roads or to increase the capacities of existing ones, thereby producing a new incentive for urban development of the outlying urban area, thus giving rise to a new cycle. This is shown graphically in Figure 1.

Figure 1. **Sectoral Environmental Dynamic 1**

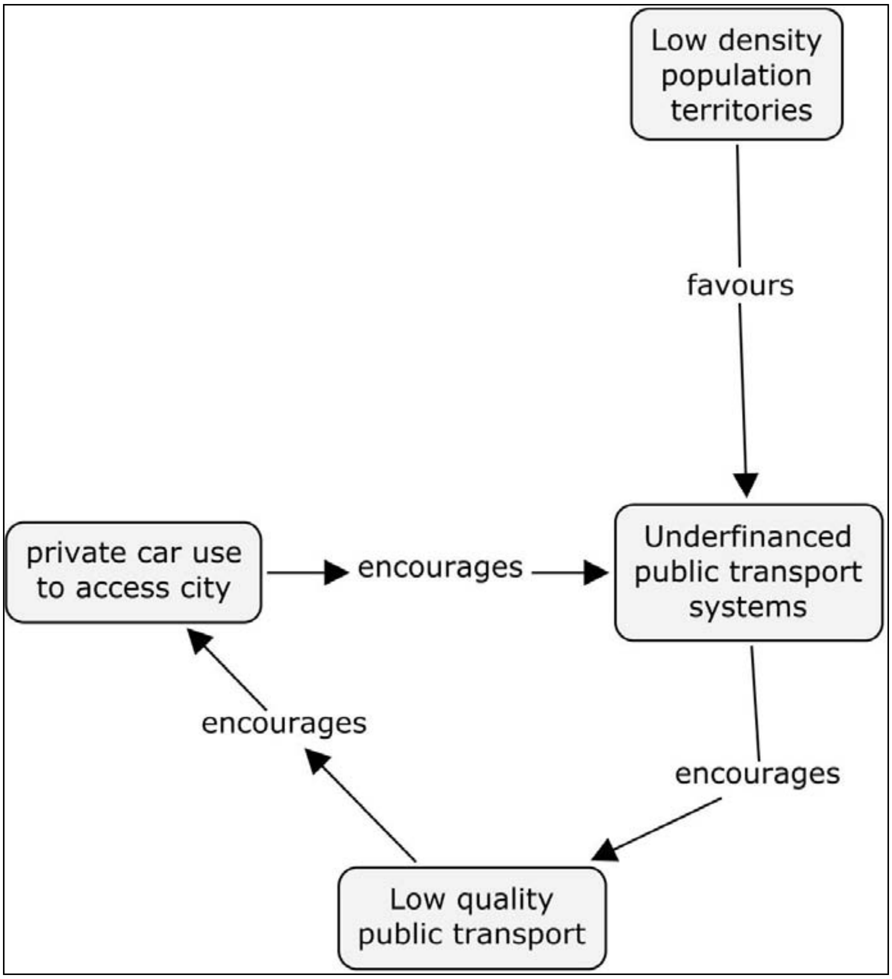


Source: Jiliberto and Bonilla, 2009.

This figure shows that the strategic effect of a strategic transport decision is systemic, irrespective of the scale of the decision or its setting, or the information available, etc.

A sector environmental dynamic is not detached from the other elements of sector policy; on the contrary, it interacts with a large number of them. A second dynamic in the same policy setting, which can be defined as the weakness of public transport, helps to clarify this aspect. This dynamic can be described illustratively and schematically as follows: low-density outlying districts favour poorly financed public transport systems; this results in the provision of low-quality public transport services, which encourages private car use and leads to under-financing of public transport. This dynamic is illustrated in Figure 2.

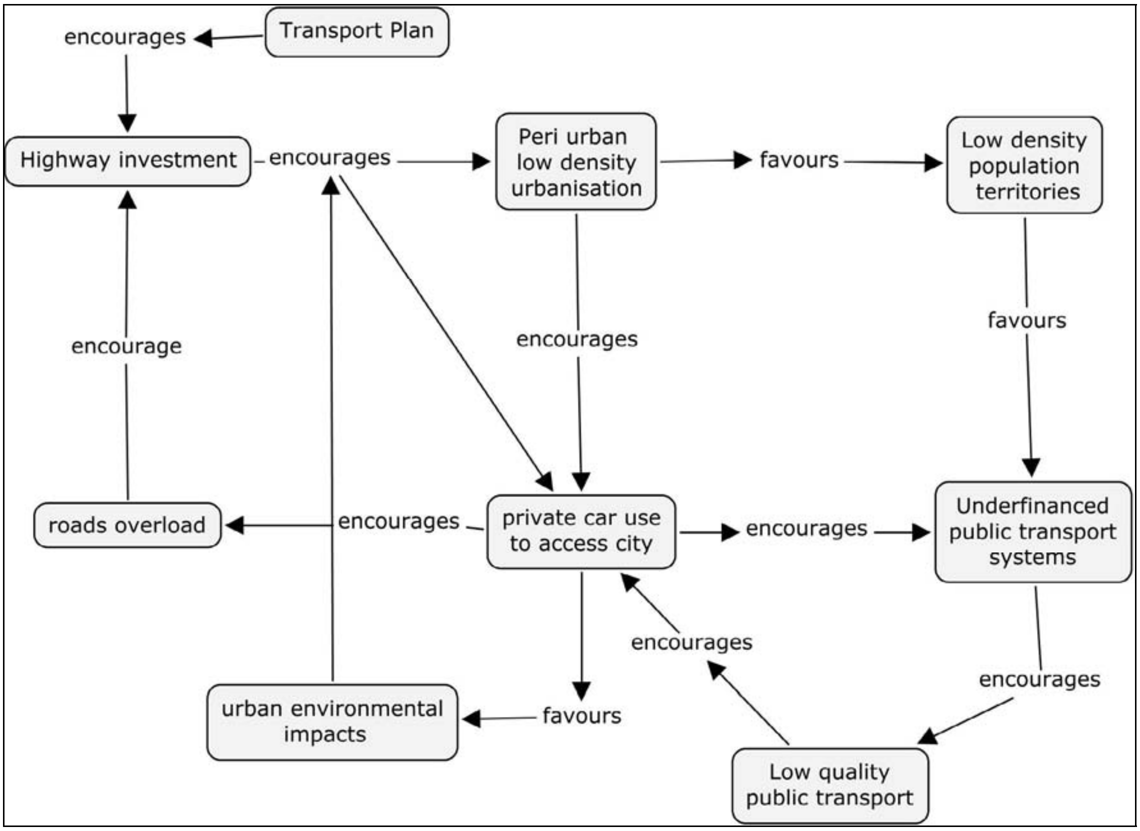
Figure 2. **Sectoral Environmental Dynamic 2**



Source: Jiliberto and Bonilla, 2009.

The two sector environmental dynamics interact with each other, giving rise to a more complex entity that we refer to as the sector environmental system. A simplified version of the sector environmental system that is relevant to a planning process, transport in this case, is shown in Figure 3, which unifies sector environmental dynamics 1 and 2.

Figure 3. **Sector Environmental System**



Source: Jiliberto and Bonilla, 2009.

Figure 3 shows what should be understood as the substantive environmental dimension of the plan or programme being evaluated, which, initially constituted by sector environmental dynamics, can be described as a more complex system. Consequently, the substantive environmental dimension of strategic decisions is systemic by nature, and understanding it is fundamental for assessing the environmental scope of the strategic aspects that the decision is attempting to promote. In this case, the substantive environmental dimension of strategic decisions is based on a consequentialism, since the contribution that SEA is supposed to make to the decision stems from understanding and taking account of the decision's consequences for the behaviour of the systemic pattern described. It is not a matter of referring to the ultimate material consequences of the decision for the environment, but its strategic reproduction pattern. This could be described as a strategic consequentialism.

It is worth noting that understanding this structural pattern is also a proactive tool for designing environmentally sustainable policy alternatives; and in this respect it is consistent with other SEA proposals such as the one promoted by Partidário (2007).

In fact, the substantive environmental dimension of strategic decisions assumes that any SEA implicitly or explicitly incorporates a mental model of how SEA can materially imply an improvement in an "environmental state of things" that the evaluated decision is affecting. Even in purely deliberative models, the improvement that SEA can imply becomes a social learning process, which should foster environmentally sustainable decisions by helping to incorporate an ecological rationale in decision making (Bina, 2007, 2008).

A large part of any SEA process involves clarifying how this substantive environmental dimension of the evaluated decision is visualised. Clearly, there is no universal model, partly because what Bina calls the "impact assessment mindset" (Bina, 2007) still persists, which strongly affects the chances for open conceptual exchange. The wide variety of evaluation situations also share responsibility: a strategic transport plan is very different from a transport corridor, or a highway. Differences in content, scales of work, information available and many other factors make it very difficult to standardize an approach to the substantive environmental dimension of the decision.

5.2. The procedural environmental dimension of strategic decisions

The possibility of describing the substantive environmental dimension of strategic decisions, albeit in a less-than-universally valid way, does not obviate the need to identify its procedural environmental dimension. This is because, as noted above, SEA is dealing with a process, not a product, and therefore can and should help enhance the environmental quality of the decision, by improving its procedures.

In general, these improvements involve seeking to ensure that the decision-making process follows codes of good decision-making practice, at least in the sense established in the state-of-the-art. The European SEA Directive (EU, 2001) is a good example of procedural recommendations because it relates SEA to fulfilment by the evaluated plan or programme of the following requirements:

- The potential significant environmental effects are evaluated;
- The alternatives are identified, described, and evaluated using common and environmental criteria;
- Consultations are held with other administrations and with the social stakeholders involved;
- Appropriate environmental data is used;
- Environmental goals are identified;
- Mitigating measures are defined;
- A monitoring system is defined.

Curiously, a recent evaluation of the application of the European SEA Directive, performed by the European Commission itself, gives the highest ratings to aspects such as procedural gains and improvements in planning processes (EU, 2009).

The European research project ANSEA makes a very detailed description of procedural criteria for decision-making, which involve an environmental improvement of the decision-making process (Caratti *et al.*, 2004).

The procedural environmental dimension of strategic decisions attracts little debate or attention, although procedural criteria form part of many SEA approaches. This may partly be because there is a degree of consensus on the need to do strategic decision-making well, which fosters an assumption that SEA obviously promotes the incorporation of good decision-making practices in the processes being evaluated.

Another explanation for the lack of discussion and concern on this issue is the predominance of positivist consequentialism in environmental assessment, or of the substantive environmental dimension, which underlies the entire contribution that SEA can make to the decision-making process by incorporating information on its hypothetical environmental consequences. This means that SEA focuses on the substantive dimension of effects and ignores the procedural aspects of the decision-making process.

Unfortunately, this situation has prevented SEA from making an in-depth environmental analysis of strategic decision-making processes, by restricting its critical capacity to processes that generally display methodological and technical weaknesses that result in a distorted consideration of their environmental dimension.

The experience of many SEA practitioners, and that of the author of this article, is that a very high percentage of the environmental weaknesses in policies, plans and programmes do not stem from consideration or otherwise of the environmental consequences, nor even from environmental aspects of any type, but from poorly structured and low-quality decision-making processes that prevent a strategic view being taken of the environmental dimension of the policy domain being affected.

Strategic environmental assessment needs more than just a good description of the substantive environmental dimension of the evaluated decision; clarity is also required as to the procedures and methodologies used in the decision-making process to ensure that this substantive dimension is not only incorporated but also understood.

It can be argued that procedural criteria, which crystallise the procedural environmental dimension of SEA, favour inclusion of the substantive environmental dimension in decision-making. This idea is well articulated in the guide to SEA published by the United Nations Economic Commission for Latin America and the Caribbean (Jiliberto and Bonilla, 2009).

Any SEA must therefore ask which procedural considerations are relevant in this decision-making process to satisfactorily incorporate what is deemed substantively relevant.

5.3. The deliberative environmental dimension of strategic decisions

Lastly, the deliberative environmental dimension of strategic decisions involves recognising the plurality and diversity of possible interpretations of the substantive and procedural dimensions, as well as recognising a political-institutional environment that does not depend on technical-rational rationales and discourses. Accordingly, the deliberative dimension lays SEA open to indeterminacy and support for governance as the hub for incorporating environmental dimension into strategic decisions.

Very little work has been done in this area, particularly, as explained above, because this dimension arises from the currently emerging modality of SEA. As argued in this article, the function of SEA is to generate institutional consensus on the diffuse responsibility pertaining to strategic decisions for conserving and improving the environment. This basically involves reaching consensus on the substantive and procedural environmental dimensions of the decision being evaluated, which should form the hub of the dialogue and negotiation process.

Any SEA needs to take progressive and flexible account of the triple environmental dimension of strategic decisions.

6. GOVERNANCE AND THE ENVIRONMENTAL DIMENSION OF STRATEGIC DECISIONS IN THE TRANSPORT SECTOR

This institutional, deliberative and strategic interpretation of SEA clearly has real potential to contribute to the environmental governance of transport policies; firstly, because of the deliberative nature conferred on the instrument, which places it at the centre of governance processes and rationales; and secondly, because it focuses the deliberation on the responsibility of strategic decisions to protect and conserve the environment in its strategic aspects – precisely where the nucleus of governance exists, in the structural aspects of collective life.

A deliberative interpretation of SEA thus favours this process. Nonetheless, the dialogue and negotiation to be undertaken need an environmental content that is understood and accepted by the parties. This is not yet the case with strategic transport decisions.

Strategic decisions in the transport sector are clearly highly varied. The SEA manual for the transport sector, produced by the European Union's BEACON project, classifies such decisions in several places to systemise their strategic evaluation to some extent. It defines policy decisions first of all, then classifiable decisions at the level of transport plans, corridors and networks, and lastly programming decisions (EC, 2005). Other classification exercises adopt similar schemes (Fischer, 2006), but they all contain a wide range of different situations.

The alternatives considered at each scale of decision-making differ significantly in the degree to which they are materialised. In policy decisions, for example, the relevant options concern the modal split, and the management of transport demand and pricing, among other issues. In decisions on transport corridors or networks, the options may have a higher level of materialisation and be associated with alternative routes or design aspects.

In environmental-impact-based SEA models, this means that descriptions of the decision's effect on the environment as a concrete physical entity, differ considerably at each level. Alternatives at the level of policy, and sometimes plans, are hard to relate to the physical environment, whereas in programmes and in certain plans this is more plausible.

This has made it methodologically impossible to generate a single discourse to describe what the substantive environmental dimension of strategic transport decisions really is, because, ultimately, the nature of the effect is highly variable and impossible to standardize. Thus, the contribution of SEA guides applied to transport, in this respect goes no further than listing, casuistically but as comprehensively as possible, what was environmentally important in individual SEAs, whether as an objective or an environmental effect, the indicators used, methodologies applied (EC, 2005; Department of Transport, 2004).

This has made it very difficult to agree on a common language to refer to the substantive environmental dimension of transport decisions, which would facilitate dialogue, negotiation and ultimately the environmental governance of key sector decisions. Instead, the environmental dimension of strategic transport decisions has become been dissolved in a sea of specifics such as indicators, the use of geographic information systems, and a series of supposed methodological and secondary technical aspects. Moreover, even efforts aimed at systemising this casuistic, to anchor the environmental dimension of decisions in aspects that materially effect the environment, have been unable to avoid producing a matrix to rank the casuistic in which the environmental dimension of transport decisions is immersed (Fischer, 2006).

The potential contribution of SEA to the environmental governance of transport policies involves overcoming this paradigm, which submerges the substantive environmental dimension of transport decisions in a sea of singularities. Only a common and generalisable language describing the substantive content of that environmental dimension can enable a constructive dialogue capable of generating shared visions of the strategic challenges facing society in terms of transport system sustainability, and thus strengthen the sector's environmental governance. Otherwise, SEA will tend to be seen as a bureaucratic requirement, the boundaries of which lend themselves as attractors for pressure and dispute.

For that purpose, the following paragraphs make a contribution to designing a generic framework for interpreting the substantive environmental dimension of strategic transport decisions. This is put forward naturally as a meta-interpretation that needs to be specified for each decision.

The map shown in Figure 4 is a generic proposal for understanding the substantive environmental dimension of strategic transport decisions, on the understanding that it deals with its *strategic* substantive dimension. In other words, it assumes that SEA focuses on the strategic aspects of the decision, rather than on its operational aspects. This is not arbitrary, since what is being decided, permanently in the case of a strategic decision, is its strategic core. Operational aspects clearly play a functional role, but a much weaker one, and they may vary considerably in the future depending on the behaviour of the multiple contextual variables in which the policy is implemented.

This proposal postulates that any strategic decision in the transport domain has an environmental strategic content, which generically can be defined illustratively as shown in Figure 4.

The illustration in fact represents an analytical approach to describing the policy situation facing any strategic transport decision in which its environmental aspects have been incorporated. It does so based on an interpretation that draws on several assumptions that need clarifying:

- The map is not a description of the physical relations generated by a transport system in providing services, but of the elements and relations of its policy situation from an environmental standpoint, i.e. the physical transport system confronted by policy alternatives, its institutional setting, stakeholders, environmental externalities, among other things – all of which form part of the policy reality that arises as a result of providing those transport services, and affect the way in which it is done.

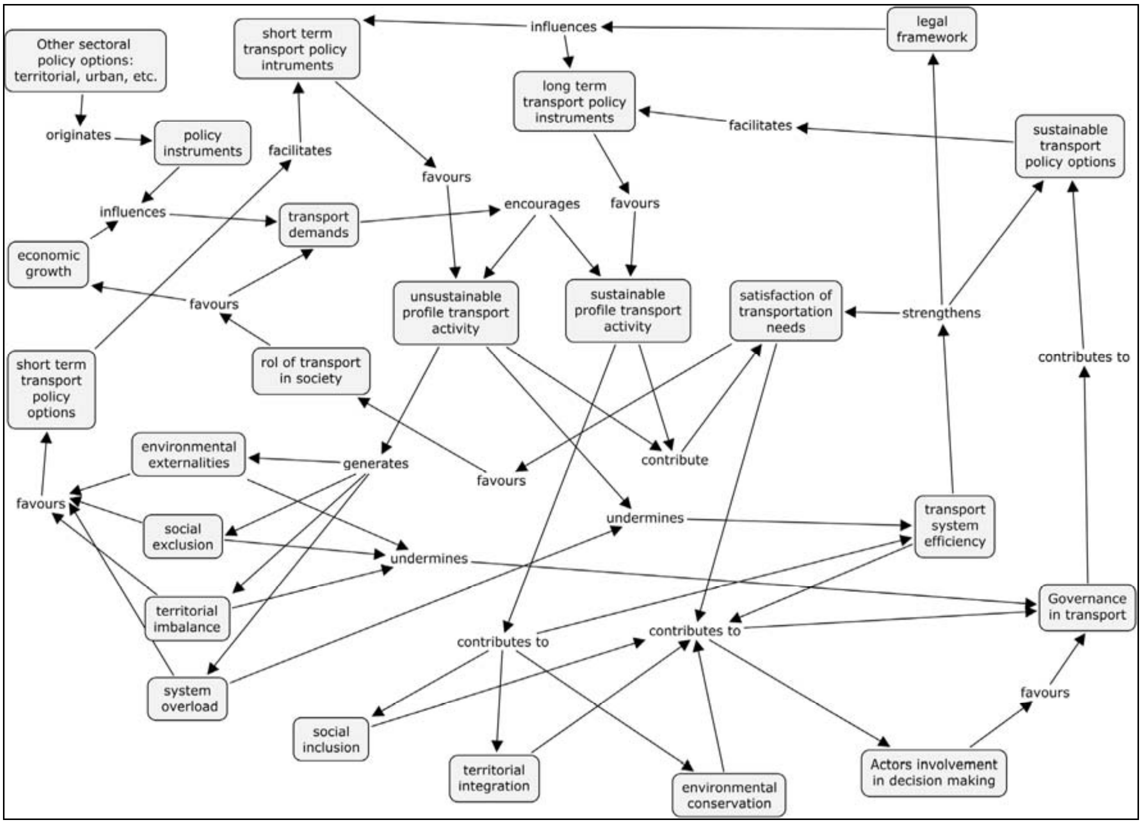
- The map is not an “objective” description of what it aims to describe, but heuristic. There is probably no universal description of what it is trying to describe. In any event, its function is not to facilitate a more or less universal or general scheme, but to illustrate an approach or an analytical rationale.

- The map describes the current policy situation facing strategic transport decisions, because any policy scenario is always contingent. And it assumes that the current transport policy scenario is necessarily a transition from an unsustainable transport model to a sustainable one. If this is not valid in any context, it will be necessary to start from a policy assumption that is.

– The map tries to show how the environmental dimension of strategic decision stems from a systematic pattern of articulation between the elements of the system, not exclusively the efficient elements of the system, i.e. those that physically produce detectable effects.

– The map does not illustrate a specific situation, but a generic analytical structure that needs to be specified for each decision. It describes a logical order of analysis, an analytical structure, which can reveal the environmental dimension of a transport policy situation when applied to a specific case. Once this analytical structure has been specified in a given case, one can say that the map describes a situation.

Figure 4. **Substantive environmental dimension of strategic transport decisions**



Source: the author.

Figure 4 aims to describe how the environmental profile of the policy domain framing a transport decision is the result of a structural feedback pattern. Accordingly, the substantive environmental question that underlies any strategic transport decision concerns the current status of this structural pattern and how it will alter the evaluated decision. In the specific context described in the illustration, the answer to this question will indicate whether the decision favours a pattern change in favour of transport sustainability, or not.

The map contained in Figure 4 is composed of several elements, each of which represents an aggregate that can be broken down into many components. The specific domain of each element needs to be determined when analysing each specific decision. The specific relations between the elements proposed in the map are also not universal, but respond to the generic discourse of how to systemically or structurally understand the environmental dimension in the transport policy domain. Accordingly, this needs to be contextualised according to the specific policy setting of each decision.

The meaning of each element is easy to understand since it is expressed in ordinary language. Perhaps two elements need explanation to understand the map better. One is the concept of transport activity with a sustainability or non-sustainability profile. This aggregate aims to abstractly identify the specific way in which a combination of transport activities provides services; and it includes all types of activities associated with transport, or its life cycle, from the building of infrastructures, to actual transport logistics, etc. The simple assumption underlying this element is that in each specific situation it is possible, based on the specific transport policy, international proposals, or state-of-the-art, to identify a combination of those activities that jointly constitute a sustainable model for that specific situation, and another combination that produces an unsustainable model, generally identified as the trend transport model.

The other group of elements to be explained consists of short- and long-term policies and instruments. These refer to the fact that in each specific case it is possible to identify a set of policies and instruments that are capable of generating structural changes that improve the sustainability of the transport activity (intermodality, demand management, traffic calming, incentives for public transport, internalisation of the social cost of transport, among other things); and another, which by relating more to short-term situations, tends to strengthen the transport trend model (solution of traffic congestion, parking problems, unimodal investment, subsidy for private vehicle use, among others).

The basic description provided by the map is as follows:

- Transport activities, whether building a road or transporting passengers and goods, may have an environmental profile of sustainability or non-sustainability in any specific setting; and their material effects will diverge, depending on this, towards environmental conservation or the generation of externalities, territorial integration or disintegration, etc.

- Any of these modalities of carrying out transport activities generate the satisfaction of transportation needs, promoting the role of transport in society. In contrast, the unsustainable model generates system overload which renders it less efficient.

- The pressures generated by the undesired effects of transport activities that have an unsustainable profile favour short-term transport policies to deal with urgencies of various types. This facilitates the use of short-term instruments and produces very quick results, feeding back into transport activities with a non-sustainability profile.

- In contrast, transport activities with a sustainable profile generate positive environmental, territorial and social externalities that favour transport policy governance and thus facilitate long-term transport policies. These lead to the use of long-term instruments, which once again provide incentives for activities with a sustainable profile. In addition, those positive effects encourage the involvement of sector stakeholders and thus strengthen sector governance.

- Apart from this, each transport modality has effects on the transport system as a whole. Firstly, transport activities with an unsustainable profile reduce systemic efficiency, which results in less satisfaction of transportation needs and diminishes the role of transport in society and its contribution to economic growth. This in turn affects the demand for transportation, which affects the two transport modalities, sustainable and unsustainable. In contrast, transport activities with a sustainability profile promote the efficiency of the transport system and ultimately increase the demand for transportation in a sustainable model.

– As noted above, policy alternatives can support one or other type of transport activity, the more sustainable or the less sustainable ones, through the type of instruments used, or the way in which a specific instrument is applied. Investment infrastructure is one such instrument. If it is used predominantly to build high-speed motorways, it will be favouring unsustainable transport activities; if it is used to balance the modal split of a system at a given point in time, it will be contributing to more sustainable activities. Policy tools are also constrained by the legal framework, which in turn is strengthened by the efficiency of the system. Each typology of instrument strengthens a different modality of transport activity, sustainable or unsustainable, and these feed back into the use of the instruments themselves. Long-term policy options are strengthened by sector governance, which in turn is strengthened by the positive externalities of the system and weakened by the negative ones, just as it is strengthened by the efficiency of the transport system. Governance thus favours long-term policy and also favoured by it at the end of the loop.

– Lastly, other sector policies use tools to materialise their strategic options, which affect transport demand and also transport activities more directly.

As shown in Figure 5, the system describing the policy situation that contains the substantive environmental dimension of strategic transport decisions can be broadly divided into three overlapping subsystems, such that a given element can be in two different subsystems.

One is the transport system as such, which is at the centre of the system. This is a simplified description of how these elements form a feedback loop, stemming from transport demand, as it were, which is the subject of feedback from the functioning of the system itself.

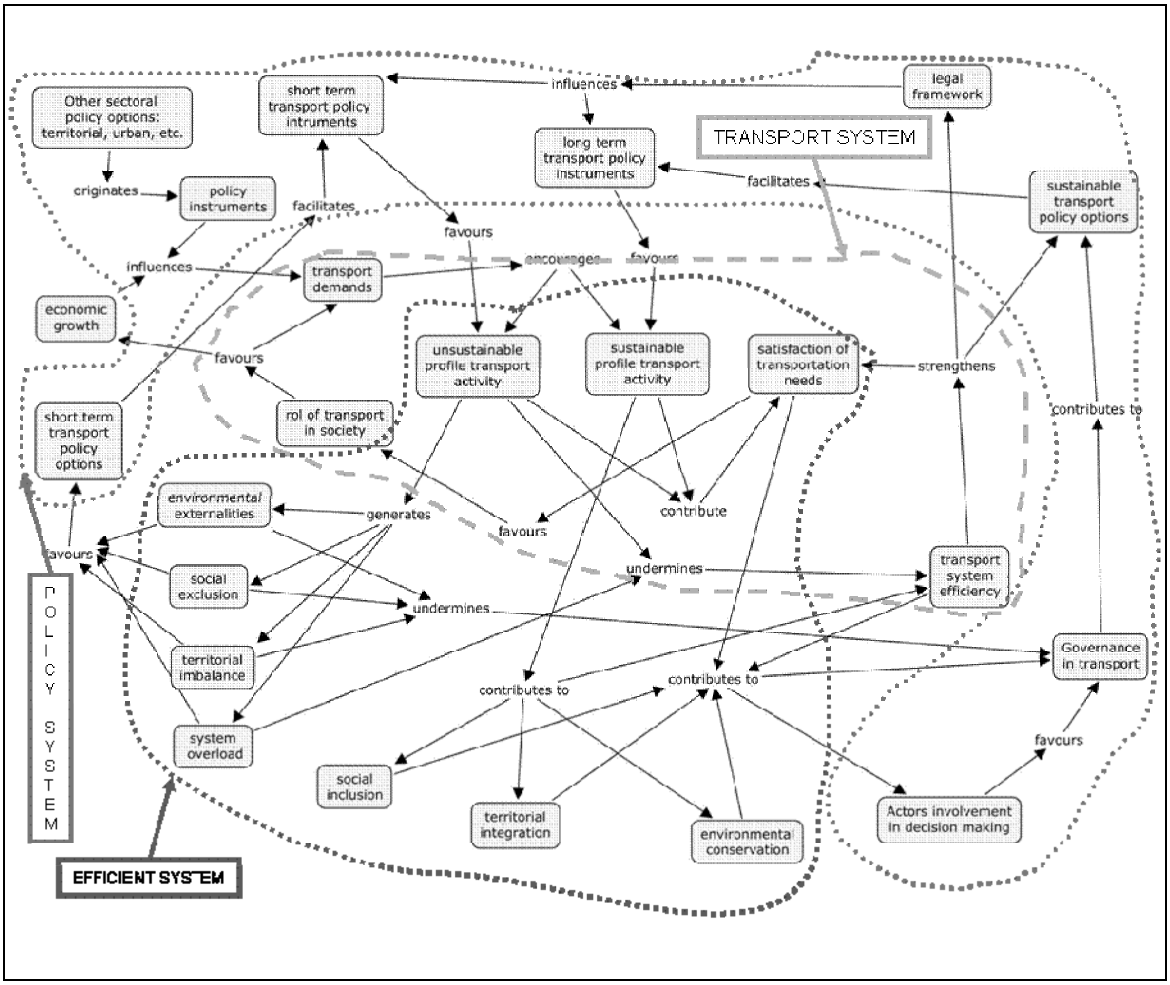
The second subsystem is the efficient one; i.e. comprising the elements that generate an effective physical action in the form of transport services, together with environmental, social and territorial ones.

Lastly, there is the policy subsystem, which encompasses all policy and institutional elements that are relevant to the system.

This classification aims to show that the description of a relevant policy situation to describe the substantive environmental dimension of any strategic transport decision openly combines elements from, in this case, three analytical domains, and that the combination makes it possible to understand the strategic environmental issues at stake in each substantive decision.

This description needs to be contextualised for each strategic transport decision. In every case, it will be necessary to identify the specific sub-elements of each of the elements contained here; and it will be necessary to determine the specific way in which they inter-relate, since nothing shown in the two illustrations is permanent. What is permanent, however, is the systemic and crosscutting way of understanding the substantive environmental dimension of strategic decisions – what this involves, ranging from an efficient system that explains the material effects of the policy situation, through to the policy system that determines how the other subsystem operates, i.e. transport, which is the engine of the efficient system, as it were.

Figure 5. **Subsystems of the environmental dimension of strategic transport decisions**



Source: the author.

The evaluated decision itself can affect one or more of these elements; so the description of the system as a whole needs to be adapted to be able to capture the new decision's influence on the system. That decision might be a national transport plan; and naturally, Figure 5, as such, could provide a basis for describing the policy situation of the plan as a whole, since a transport plan generally has a globalising aim. But it could also involve a plan relating to transportation logistic services only, in which case each element needs to be adapted to the reality of the policy and the dimensions of a logistics plan for transportation services. The decision might concern a transport corridor, so the description should relate to the transport system containing the corridor, to understand its own policy situation and thus derive its substantive environmental dimension.

As noted above, this is not being claimed as the only possible description of the policy situation that elucidates the substantive environmental dimension of transport decisions. On the other hand, only a description similar to this one makes it possible to locate the substantive environmental dimension of transport decisions in the strategic setting of relevance to SEA.

It is also argued that the development of a specific strategic language to represent the substantive environmental dimension of transport decisions is a *sine qua non* for generating constructive dialogue and negotiation in the SEA framework that effectively strengthens the environmental governance of transport policy.

As noted above, the substantive environmental dimension is the most complex of the three dimensions of the environmental dimension of strategic decisions explained in the previous chapter, and for that reason has been further developed in this one. Clearly, the procedural dimension would require similar treatment, although, as noted, at first sight this is less conflictive and possibly not such a high priority. In the context of deliberative and governance strengthening processes, it is also important to generate common discourse on the scope and specific content of that procedural dimension. While progress has been made on this subject (Caratti *et al.*., 2004), further reflection is clearly needed, particularly in relation to its application in a deliberative SEA model.

CONCLUSIONS

SEA now has considerable experience, including as applied to transport policies, plans, and programmes. One of its most salient features is its theoretical, methodological, and practical diversity, which until now has been assumed merely as a characteristic of SEA. Nonetheless, this characteristic can also reveal dysfunctionality between the theoretical model of evaluation and the context in which it is applied: e.g. a model based on technical-rational premises, and an institutional political setting governed by political rationales that diverge widely from the substantive rationality paradigm.

This would explain the rapid evolution in the conceptualisation of the instrument, from a highly technical one based on the concept of environmental impact, passing through an SEA proposal centred on the decision itself, to a more deliberative one based on dialogue in negotiation processes, which is currently emerging. In other words, that dysfunctionality generates a cognitive shift that has been opening up new horizons for SEA.

This conceptual reflection, however, has not translated mechanically into the practice of SEA, which remains tied to the ‘impact assessment mindset’, partly, given the form that SEA legislation has taken, because all new development requires time to move from conceptual reflection to practice.

SEA is at the centre of sector policy governance, particularly in the case of transport policies, since it concerns management of the community's strategic decisions, how to improve them, and how to make them more consistent with prevailing values and purposes.

Nonetheless, the technical-rational models of SEA do not encourage SEA to deploy all its governance potential. By assuming the existence of an “objective” foundation for a complex and diffuse entity, such as the environmental dimension of strategic decisions, technical-rational models encourage key actors to adopt strategic positions, thereby obstructing open and transparent dialogue.

Both to benefit environmental governance and to ensure the consistency of SEA itself, it is necessary to move towards more deliberative models, as the literature has been suggesting. A deliberative SEA model starts from the assumption that the institutional political settings in which SEA is implemented are spaces of dialogue and negotiation for a diversity of technical-rational discourses. This is particularly true in the case of SEA, where it is impossible to claim the existence of a uniquely possible technical-rational discourse on what the environmental dimension of a strategic decision really is.

Accordingly, while all SEA must have a technical-rational foundation, this needs to be determined in a process of negotiation and dialogue in which several possible discourses on the same topic are discussed.

Similarly, a deliberative model of SEA needs to understand the functional purpose of SEA at the institutional and governance levels. SEA would basically have an institutional function of determining a complex and confusing aspect of public policies: their responsibility in upholding the constitutional premise of protecting and conserving environment and favouring sustainable development. Although the technical-rational scope of the environmental dimension of a strategic decision is indeterminate *a* *priori*, what is not is the fact that each SEA can define what that means in each case, thus becoming institutionalised.

A deliberative SEA model does not deny the need for a technical-rational description of the environmental dimension of strategic decisions; it merely recognises that there is no unique and universal one. Accordingly, it is reasonable and necessary to continue thinking about what it means to incorporate the environmental dimension into strategic decisions. Accordingly, it is proposed to view this as a complex entity consisting of three dimensions, substantive, procedural and deliberative.

The substantive dimension is possibly the most complex of all, since it concerns how SEA is understood to improve the "environmental state of things" that the evaluated decision aims to affect. In simple terms, the substantive dimension answers the question why is SEA good for the environment? Some analysts will say because it minimises impact, others because, by improving the decision process environmentally, an environmentally superior proposal is generated; still others will say because it helps in a diffuse way to incorporate an ecological rationale into decision-making processes, which at some point in time will result in more environmentally sustainable proposals.

Generating a common discourse on that substantive environmental dimension is necessary to enable SEA to be applied more consistently, and to make progress in environmental and sector governance processes.

In the case of transport policies and their strategic decisions, whether these involve policies, plans or programmes, the aim is to understand the substantive environmental dimension from the strategic and systemic standpoint. In this case what SEA brings to the "environmental state of things" is that it helps the evaluated decision to internalise the structural pattern explaining the environmental profile of the decision's specific policy situation. SEA places a systemic-structural description of the policy situation in which the evaluated decision is immersed at the centre of the evaluation; and it attempts to determine whether the decision favours a transport sustainability profile or not, assuming transport sustainability is a policy aim, obviously. It does this at a strategic level, provided by its systemic description, and not at the level of the material changes actually caused by the decision, which are not the focus of an SEA.

NOTE

1. In this article, positivist consequentialism is understood as the analysis that understands that the consequences of decisions or acts can only sensibly be valued in terms of their positive (i.e. materially discernible) consequences.

BIBLIOGRAPHY

Ahmed, K., J.R. Mercier and R. Vermeem (2005), "Strategic Environmental Assessment - Concept and Practice", *Environment Strategy*, No. 14, Environment Department, World Bank, Washington.

Audouin, M. and P. Lochner (eds.) (2000), *SEA in South Africa - Guideline Document*, CSIR and the Department of Environmental Affairs and Tourism (DEAT), Pretoria.

Bina, O. (2008), "Strategic Environmental Assessment", in: Jordan, A. and A. Lenschow, *Innovation in Environmental Policy? Integrating environment for sustainability* Edward Elgar Publishing Ltd, Cheltenham, UK.

Bina, O. (2007), "A critical review of the dominant lines of argumentation on the need for strategic environmental assessment", *Environmental Impact Assessment Review*, 27.

Brown, A.L. and R. Therivel (2000), "Principles to guide strategic environmental assessment methodology", in: *Impact Assessment and Project Appraisal*, Vol. 18, No. 3, September 2000, 183-190.

Caratti, P., H. Dalkmann and R. Jiliberto (2004), *Analysing Strategic Environmental Assessment*, Edward Elgar Publisher, Cheltenham, Great Britain.

Dalal-Clayton, B. and B. Sadler (2005), Strategic Environmental Assessment, a source book and reference guide to international experience, Earthscan, London, Great Britain.

Dalkmann, H., D. Bongart and R. Jiliberto (2004), "Analytical strategic environmental assessment (ANSEA), developing a new approach to SEA", *Environmental Impact Assessment Review*, 24(4).

Dalkmann, H. and M. Nilsson (2001), "Decision making and strategic environmental assessment", *Journal of Environmental Assessment Policy and Management* (3).

Department of Transport (2004), TAG Unit 2.11: Strategic Environmental Assessment for Transport Plans and Programmes, UK.

European Commission (EC) (2009), DG Environment, Study concerning the report on the application and effectiveness of the SEA Directive (2001/42/EC), Final Report, Brussels.

EC (2005), DG for Energy and Transport, *BEACON*, *SEA Manual*, Brussels.

EC (2001), *La Gobernanza Europea*, *Un Libro Blanco*, *Official Journal* No. 287, 12 October, pp. 0001-0029, Brussels.

EC (2001), European Commission, DG Environment, Strategic Environmental Assessment of Transport Corridors: Lessons learned comparing the methods of five Member States, Brussels.

EU (2001), Directive 2001/42/EC on the Assessment of the Effects of Certain Plans and Programmes on the Environment, European Commission and Parliament, Brussels.

EC (2000), European Commission, DG Environment, Strategic Environmental Assessment in the Transport Sector: An overview of legislation and practice in EU Member States, Brussels.

European Conference of Ministers of Transport (ECMT) (1998), *Strategic Environmental Assessment in the Transport Sector*, OECD Publications, Paris.

ECMT (2000), Strategic Environmental Assessment for Transport, OECD Publications, Paris.

EEA (1998), European Environmental Agency, "Spatial and Ecological Assessment of the TEN: Demonstration of Indicators and GIS Methods", *Environmental Issues Series*, No. 11, Copenhagen.

Faludi, A. (1987), *A decision centred view of environmental planning*, Pergamon Press, Exeter, Great Britain.

Fischer, T.B. (2006), "SEA and transport planning: towards a generic framework for evaluating practice and developing guidance", *Impact Assessment and Project Appraisal*, 24 (3).

Fischer, T.B. (2005), "Having an impact? – Context elements for effective SEA application in transport policy, plan and programme making", *Journal of Environmental Assessment Policy and Management.*

Fischer, T.B. (2002), Strategic Environmental Assessment in Transport and Land-use Planning, Earthscan, London.

Furman, E. and M. Hildén (eds.) (2001), Transport Planning: does the influence of strategic environmental assessment/integrated assessment reach decision-making? Final Report of a workshop, Finnish Ministry of Environment, Helsinki.

Graham, J., B. Amos and T. Plumptre (2003), "Principles for Good Governance in the 21st Century", *Policy Brief No. 15*, Institute on Governance, Ottawa, Canada.

Habermas, J. (2003), Teoría de la acción comunicativa, Racionalidad de la acción y racionalización social, Taurus, Spain.

Hilding-Rydevik, T. and H. Bjarnadóttir (2007), "Context awareness and sensitivity in SEA implementation", *Environmental Impact Assessment Review*, 27.

Jiliberto, R. and M. Bonilla (2009), *Guía de Evaluación Ambiental Estratégica*, Economic Commission for Latin America and the Caribbean (ECLAC), United Nations, Santiago, Chile.

Jiliberto, R. (2007), "The need to transform the environmental assessment paradigm", *Journal of Environmental Assessment Policy and Management*, Vol. 9, No. 2, June 2007.

Jiliberto, R. (2004), "Setting the ground for a new SEA approach", in: Caratti, P., H. Dalkmann and R. Jiliberto (eds.), *Analysing Strategic Environmental Assessment*, Edward Elgar Publisher, Cheltenham, Great Britain.

Jiliberto, R. (2002), "Decision-related Environment Values as object of analysis of Strategic Environment Assessment", *Impact Assessment and Project Appraisal (IAPA)*, Vol. 20, No. 1, March, Great Britain, pp. 61-70.

Kornov, L. and W. Thissen (2000), "Rationality in decision- and policy-making: implications for strategic environmental assessment", *Impact Assessment and Project Appraisal*, Vol. 18, No. 3, September.

Mercier, J. and K. Ahmed (2004), EIA and SEA at the World Bank. Paper presented to the proceedings of the 8th Intergovernmental Policy Forum on Environmental Assessment, Vancouver, Canada.

Nilsson, M., H. Wiklund, G. Finnveden, D. Jonsson, K. Lundberg, S. Tyskeng and O. Wallgren (2009), "Analytical framework and tool kit for SEA follow-up", *Environmental Impact Assessment Review*, 29.

OECD (2007), La Evaluación Ambiental Estratégica Una guía de buenas prácticas en la Cooperación para el Desarrollo, OECD Publications, Paris, France.

Partidário, M. (2007), *Strategic Environmental Assessment Good Practices Guide*, Portuguese Environment Agency, Lisbon.

Partidário, M. and R. Clark (eds.) (2000), *Perspectives on Strategic Environmental Assessment*, Boca Raton, FL: CRC-Lewis.

Sadler, B. (2008), "International Trends and Developments in SEA Process and Practice", in: E.W.K. Au, K.C. Lam, T. Zhu and M. Partidário (eds.), *International Experience on Strategic Environmental Assessment*, Center of Strategic Environmental Assessment for China, the Chinese University of Hong Kong, Hong Kong Institute of Environmental Impact Assessment, Research Centre for Strategic Environmental Assessment, Nankai University, Beijing, China.

Sadler, B. and R. Verheem (1996), *Strategic Environmental Assessment - Status, Challenges and Future Directions*, Ministry of Housing, Spatial Planning and the Environment of the Netherlands, The Hague, The Netherlands.

Scott, P. and P. Marsden (2003), *Development of Strategic Environmental Assessment (SEA) Methodologies for Plans and Programmes in Ireland* (2001-DS-EEP-2/5). Prepared for the Environmental Protection Agency by ERM Environmental Resources Management Ireland Limited, Wexford, Ireland.

Simon, H. (1997), *Administrative Behaviour* (Free Press, New York).

Therivel, R., E. Wilson, S. Thompson, D. Heaney and D. Pritchard (1992), *Strategic environmental assessment*, Earthscan.

Therivel, R. and M. Partidário (1996), *The practice of strategic environmental assessment*, Earthscan.

Verheem, R. and J. Tonk (2000), “Strategic environmental assessment: one concept, multiple forms”, *Impact Assessment and Project Appraisal*, Vol. 18, No. 3, September, 177-182.

UNDP (2004), *UNDP Environmental Mainstreaming Strategy*, United Nations Development Programme, Bureau for Development Policy, UNDP.

UNDP (2000), *Strategy Note on Governance for Human Development*, United Nations Development Programme, UNDP.

Wallington, T., O. Bina and W. Thissen (2008), “Strategic environmental assessment Theory and Research: an analysis of discourse”, in: Sadler, B., R. Aschermann, J. Dusik, T.B. Fischer, M. Partidário and R. Verheem (eds.), *Handbook of Strategic Environmental Assessment*.

Wallington, T., O. Bina and W. Thissen (2007), “Theorising strategic environmental assessment; Fresh perspectives and future challenges”, *Environmental Impact Assessment Review*, 27.

World Bank (2005), Integrating environmental considerations in Policy formulation, Lessons from policy-based SEA experience, Environment Department, World Bank, Washington DC.

World Bank (2006), Environmental Impact Assessment Regulations and Strategic Environmental Assessment Requirements, Practices and lessons learned in East and Southeast Asia, Environment and Social Development Department, World Bank, Washington DC.

DOES STRATEGIC ENVIRONMENTAL ASSESSMENT CHANGE OUTCOMES?

Maria R. PARTIDÁRIO

Instituto Superior Técnico (IST)
Lisbon
Portugal

SUMMARY

1. INTRODUCTION

Strategic Environmental Assessment (SEA) is going through times of turbulence. Initially introduced to help improve environmental performance in development decision-making, and overcome the inability of Environmental Impact Assessment (EIA) to deal with the complex decision frameworks that support development projects, it has subsequently been interpreted in multiple ways, now translating into various forms and applications.

The current diversity of understanding around SEA is possibly related to three important aspects relevant to this paper: a) the political attitude in relation to forms of environmental interference in the decision processes; b) the emergence of a strategic decision-making culture, its meaning and positioning in the decision-making spectrum, that anticipates and differentiates from a project decision-making culture; c) the ownership of SEA by different disciplinary fields that have molded SEA as a function of their professional backgrounds.

Some authors consider this diversity to be enlightening regarding the potential of SEA, and one of its major features that can be encouraged. A few still see this multiplicity as a diversion away from SEA's key purpose of extending EIA to other levels of decision-making. A third group sees this as part of SEA growing pains, where accumulated knowledge and experience will lead SEA to evolve to a matured, full-fleshed and effective instrument with clear and coherent functions and forms.

For a number of years, the author has discussed these understandings of SEA and has argued that there is no point in re-inventing EIA in the shape of SEA. The need to consider SEA in a strategic sense has been suggested on a number of occasions (Partidário, 1999, 2000, 2005a and b, 2006, 2007, 2008). This means that SEA should not be a subsequent form of EIA that develops studies to assess the impacts of policies, plans and programmes. Instead, SEA must be an instrument that performs a fundamental new attitude in strategic development processes, establishing a relationship with the decision-making process, with a fresh and constructive look, centred in the strategic dimensions of the decisions to be taken. Increasingly, this strategic understanding of SEA is also being advocated by several colleagues (Kornov and Thissen, 2000; Wallington, 2002; Bina, 2003; Notteboom, 2006; Cherp *et al*, 2007; Wallington, Bina and Thissen, 2007), each of them arguing from their own valuable professional points of view, suggesting ways of approaching SEA under different theoretical paradigms.

This paper addresses the advocacy role that SEA can strategically play towards more sustainable and environmental decision-making and how this can be achieved. It discusses the required conditions for this performance and also the frustration of SEA when such conditions are absent or insufficient. This paper shares experience with the case of a Strategic Environmental Assessment on the location of the new international airport in Lisbon; how SEA made a difference to infrastructure development decisions and the conditions that were met to make it possible.

2. SEA – AN ADVOCACY ROLE FOR BETTER ENVIRONMENTAL AND SUSTAINABLE DECISION-MAKING

Strategic Environmental Assessment is a decision support instrument that was found necessary, in the early days, to upstream environmental and social issues into higher levels of decision-making, improving the policy and planning decision contexts for the development of project's Environmental Impact Assessment. It seems clear that SEA advocacy role for better environmental and sustainable decision-making has been at the genesis of the instrument. But advocacy is played in different forms, depending on society priorities, political and organisational cultures, on developed knowledge and applied advocacy methods. This has led to different interpretations on how SEA should shape and deliver its expected role.

Over the years, many forms of SEA have been founded, mostly on projects' EIA-based approaches, others on policy science and decision-making systems or on spatial planning approaches (Dalal-Clayton and Sadler, 2005). For those familiar with the various methods and approaches to policy-making and planning around the world, it is easy to understand how differently SEA correspondently shapes if it gets molded to the respective system to which it will apply.

The European Directive 2001/42/EC, commonly known as the SEA Directive and frequently assumed as a world standard approach to SEA, particularly when seconded by the China EIA law section on Plan SEA, but also because it was one of the first legal frameworks established in the world, represents however only one of the several interpretations of SEA. By and large, it stands as a rather limited form of SEA if we expect SEA to deliver as a strategic-based instrument. The European Directive priority target are plans and programs that set the framework for future development consent of projects that require an EIA [Art. 3, No. 2, (a)], which determines a project decision culture. In practice, the fulfillment of the directive requirements is mostly being interpreted as an enlarged EIA.

Other more strategic interpretations of SEA have been evolving, looking at an SEA that proactively assists the shaping and the design of strategies. This requires a mutual molding process of SEA and strategies formation, working through problem perception and policy design to flexibly respond to problems, with SEA assisting policy and planning to formulate and discuss strategic alternative options, and then to help decision in choosing and implementing those strategies that better recognize environmental and sustainability priorities.

While a strategic approach to SEA looks into the capacity of SEA to influence decisional contexts and the formulation of strategic initiatives, whether policies, plans or programs, there is still a quite strong line of SEA approaches based on a rationalistic attitude that tends to design SEA to perform as a standard sequence of activities, inspired in the EIA process and centred on the preparation of an SEA report that culminates in the key purpose of informing and validating a final decision.

The concept of environment has also been the reason for advancing alternative instruments that eventually compete with SEA. Because of the often limited understanding of the term "environment", when associated only to earth issues, integrated impact assessment (UNEP, 2005 and 2009) as well as sustainability assessments (Pope *et al*., 2004; Gibson *et al.*, 2005) have evolved as instruments that aim to ensure the inter-linkages between the social, physical-ecological and economic systems. Integrated or sustainability assessments however are currently used at any scale, from projects to the policy range of decisions.

Institutional approaches to SEA have also been recently advanced (Ahmed and Sanchez-Triana, 2008) which are very much supported by institutional and organizational learning principles and practices, directed by capacity-building priorities, ranging from project-based logics to more strategic logics of assessment.

What is argued here, and has been argued in previous occasions, is that in order to be effective and responsive to decision needs, SEA must offer flexibility and cannot be formatted as a streamlined sequence of standard activities such as EIA. The concept of framework of activities that enable SEA to become flexible, diversified and tailor-made to the decision-making processes has been suggested before (Partidário, 2005b). SEA has the potential to help decision-makers to identify options that meet sustainability aims, looking for risks and opportunities of proposed strategic actions, also providing for an early warning of cumulative, synergetic and indirect effects, and large-scale impacts. In order to do this, however, SEA must understand and address the complexity of strategic processes and be able to provide advice in a timely and pragmatic fashion.

Increasingly SEA major key role can be argued to be that of facilitating decision-making by involving key actors, enabling dialogues towards mutual understanding, ensuring a long-term and large scale perspectives when considering development options. When addressing the complex nature of strategic decision-making, SEA cannot be limited to a technical assessment, and consequent advice on proposed options, but it must be well embedded in the strategic decision-making context to be able to influence decision-making performance (Partidário, 2005a). SEA should not be about controlling decisions. SEA should be about demonstrating the competing advantage of taking into account big-picture environmental issues to enable sustainable decision-making.

3. WHAT IS NEEDED FOR SEA TO ACT STRATEGICALLY?

A strategic-based model for SEA was proposed in Partidário (2006). It was later adapted to Guidance for SEA, with the purpose of meeting European Directive requirements, and was published by the Portuguese Environment Agency (Partidário, 2007a). Since then the methodology laid out in the guidance has been generally followed in Portugal, although often not fully meeting its principles and conditions for success.

The proposed approach of strategic-based SEA is conceived as a decision-centred instrument, that is driven by the dynamics of the decision process and which is focused on assessing strategic processes, rather than plans or programs. It aims at integrating environmental issues in a sustainability context, taking SEA as a strategic facilitator of sustainability processes. A decision-centred SEA means that SEA is flexible and tailor-made to each decision process, conceived as a framework of key elements that need to be strategically positioned to enable SEA to play its decision support role and to ensure the added-value of SEA to decision-making (Partidário, 2000).

This concept and its supportive methodology represent an innovative approach in relation to traditional practices of SEA. Key drivers in the strategic-based model to SEA (Partidário, 2008) include:

- Follow strategic thinking, as opposed to project thinking;
- Work with processes, not with outcomes;
- Allow and promote early engagement, community participatory planning, use and enhance communication skills;
- Use dialogue, persuasion and negotiation as techniques throughout the entire process;
- Focus on long-term objectives and the strategy to achieve them;
- Ensure a long term view, but taking short-term action following few priority objectives;
- Be strategic but not descriptive – use clusters of themes [Critical Factors for Decision-making (CFD)], and perform a quick and sharp diagnosis;
- Apply integrative holism – CFD are integrated dimensions;
- Adopt a large picture, sustainability approach;
- Be very focused and pragmatic about the assessment;
- Rather than predicting impacts, help to think about future pathways for sustainability;
- Be a facilitator of decision-making – enable flexibility and continuity, follow the decision cycle;
- Change terminology to adopt a strategic oriented terminology.

In this model, the purpose of SEA is to help understand and appropriately address a problem, and to find environmentally, and sustainable, viable options to achieve objectives. It is based on policy processes, generation of knowledge, networks of actors, inter-sectoral co-operation and governance. The adopted approach recognizes three main functions in SEA:

1. *Integration* of environmental and sustainability issues in strategic processes;
2. *Assessment* of opportunities and risks of strategic options;
3. *Validation* of the assessment of strategic processes and outcomes;

and suggests a general format to enable a strategic performance:

- Establish a framework of institutional governance and participation, and recognise different perspectives;
- Build a strategic reference framework (SRF) – working for a sustainable future and development objectives and creating an assessment benchmark;
- Identify Critical Factors for Decision-Making (CFD) – priorities setting exercise, generating clusters based on the fundamental strategic issues (SI) for development, the relevant environmental factors (EF) and the macropolicy framework defined by the SRF;
- Analyse trends, not moments. The strategic context is identified, based on an analysis of trends. What matters is a dynamic analysis, not a static analysis;
- Conduct sectoral studies that perform an analysis of the CFD, and the assessment, to provide information to the decision-maker;
- Analyse strategies and assess strategic options for different future scenarios;
- Prioritise and explore plausible options that enable choice, foreseeing and avoiding risks and exploring opportunities;
- Produce as many issues notes, comments and short reports as necessary, depending on the opportunities created by decision windows;
- Propose guidelines that drive possible pathways, avoid the mitigation paradigm;

– Strongly support the strategy life-cycle with a follow-up process that ensures: design, assessment, monitoring – integrating in the strategic process of decision-making.

A new lexicon for SEA was suggested in (Partidário, 2007) to help enhance a strategic culture in impact assessment.

Table 1. **Proposed new lexicon to create strategic thinking in SEA**

In SEA strategic model use:	**In traditional terminology**
Critical factors (Clusters)	Scoping
Decision windows (in strategic process)	Planning phases
Strategic issues Drivers of change Context data	Baseline
Strategic Options	Alternatives
Opportunities and risks	Impacts
Guidelines (planning, management)	Mitigation measures

So what should be expected from SEA as a strategic approach? What may be required to ensure a strategic performance of SEA?

The point that has been made here is that for SEA to perform more strategically it must fulfill a set of functions and assume a consequent form. Emphasis is on the strategic role of SEA in influencing decision-making through the integration of relevant "big picture" environmental issues at the core of strategic decisions to help identify pathways for sustainability. Which means that SEA need to act strategically in relation to why doing, who to engage, what to consider and when to influence decision-making.

4. WHEN SEA BECOMES IRRELEVANT – NEGATIVE EXPERIENCES ON GOVERNANCE AND THE DECISION-MAKING PROCESS

In order to be relevant to decision-making, SEA needs to target decision concerns and priorities and bring an added-value. Decision-making has to recognize SEA as an ally, an approach that can bring benefits and not just a waste of time. The advocacy role of SEA in mainstreaming environmental and social issues in decision-making has been discussed above, recognizing that it can be done in many different ways. Three approach categories can be identified:

A *marginal approach* is when SEA becomes an end in itself, in other words, SEA is conducted to be a perfect exercise of baseline studies that perform analysis and diagnosis of environmental and social issues and assess the effects of proposals, following a standard streamline of formal activities, and culminating in a fat formal report. The purpose is normally the preparation of comprehensive studies that can provide the best image of the situation that contextualizes the decision in environmental and social terms. Often the outcome of such SEA becomes irrelevant to decision-making because a lot of work is done, much effort and resources are used, it is quite time-consuming but it is not focused into what decision-making actually needs to know.

A *compliance approach* is when SEA is mainly a mechanism of control of compliance with the existing legislation and policy requirements. In this approach, what is laid out in the legislation is taken as the road map for SEA. The priority is to fulfill each item identified in legal terms, and it even happens that adjectives or other sentence connecting expressions used in the legislation become formal names for types of SEA. For example, in the UK Appropriate Assessments became a formal type of SEA, to specifically address the Habitats Directive requirements, only because the legislation says that "...the assessment should be appropriate...", and guidance has been issued. Once could wonder if other types of SEA do not need to be appropriate!

Finally a *constructive approach* is when what is relevant for decision-making becomes central in SEA, so that SEA single purpose is to help drive strategies towards better environmental and sustainability integration in a constructive way. The priority here is to understand and analyze decision needs and priorities and design the SEA to respond to decision-making. In this approach SEA must be shapeless, so that it can be molded to each decision case. It needs to be highly flexible, agile, focused on the issues that bring an added-value to decision-making, that will help decision to be taken in a more environmental and sustainable way. SEA develops to identify few but highly relevant themes for decision-making, works with alternative strategic options that can show alternative pathways for sustainability. The outcomes of the SEA are embedded in the decision process, several inputs are made throughout the decision cycle at key moments when such input can actually be used and make a difference to decision-making.

These three categories are not just virtual, they are defined based on existing experience with SEA. A systematic review of the SEA experience worldwide would show that the Marginal and Compliance approach are, by and large, the most common SEA approaches. Recent experiences point towards the constructive approach, but fewer cases can be identified. Often there will be combinations of these different approaches, particularly when we want to use SEA to make the best possible decision case – the constructive approach – but, at the same time, we need to comply with legal requirements, such as those imposed by the European Directive and subsequent national legislation – the compliance approach.

Consequences of the wider use of the marginal and compliance approaches are often responsible for frustrations with the application of SEA, in view of the constructive approach. One of the major frustrations is that SEA is still very much seen as an EIA applied at upper levels of decision-making, such as policies, plans and programs, carrying all the burdens and limitations created around the practice of EIA over its 40 years lifetime. This is when SEA is increasingly reactive to decision intentions, is dominated by extensive baseline descriptions, it provides very little analysis and even less advice to decision-making, it offers short-term view of effects, is report-driven and is becoming a necessary "industry" process to reach permits. All these are the opposite of what SEA should do. Other frustrations include:

Concerning governance:

– Limited participation and diversity of view points – institutional or mono-oriented assessments, often with public environmental institutions playing the drivers role where SEA is legally enforced;
– Limited influence in the decision-making process, originating parallel, non convergent, decision and SEA processes and mutual tensions that eventually bring limited benefits to the environment and to the society;
– In Europe, the legal requirements focus on effects assessment, mitigation measures and deliver of environmental report determine a strong EIA proxies, and consequent project culture in the assessment, which ultimately influence strongly the expectations of authorities as to the structure and detail resulting from SEA.

Concerning the decision process:

– At strategic decision levels SEA is seen as the environment weak link – policies, plans and programs will carry on their initial purposes and intentions and will "staple" the environmental report for purposes of legal compliance.
– Legal requirements for the demonstration of effects towards mitigation measures hinder the capacity of SEA to be more constructive, innovative and tailor-made.
– The practice shows that SEA is often centred in the production and delivery of an environmental report. This limits the decision flexibility to work with several short SEA reports that rather than bringing a demonstration of impacts should be bringing useful inputs to decision-making, to think about strategic pathways that would avoid future problems. SEA should act as the walking stick, that helps decisions to be made, rather than as a barrier that steps across the decision way.

5. A SHORT SUCCESS STORY ABOUT SEA'S CAPACITY TO CHANGE DECISIONS

The story of the decision process on the Lisbon new international airport provides a good example of the capacity of SEA to strategically influence decision-making. After 40 years of multiple attempts and struggled debate around alternative location sites, the decision on its final location was made. Yet, once the decision was made, it was suddenly changed because of a strategic insight into other relevant, long-term strategic issues that were not considered before in a systematic and transparent way. A better option, that had never been considered before, was found and eventually politically chosen. SEA played a role that changed a 40-year decision, in less than one year.

Lisbon Airport has been operating in its current location, at Portela-Lisbon, since 1942. At the time built on the city outskirts, the airport was surrounded by urban expansion in the following two decades. The relocation of this infrastructure was on the table for the first time in 1969. At that time five alternative sites were identified, all located in the south bank of the Tagus River. An initial study was completed in 1971, selecting an area of over 6 500 hectares in Rio Frio, about 40 kilometres south of Lisbon, where a four parallel runway airport would be constructed. The economic and political context in Portugal, however, changed significantly in the 1970s, following the first oil crisis and the Portuguese political process. The airport was not a national priority anymore and this all process was put on hold.

The issue was re-opened in 1982 and a comprehensive study analysed 12 alternative locations. The study concluded on a new better location at Ota, 40 kilometres north of Lisbon, on the right bank of the Tagus River and opposite to Rio Frio, earlier identified and located on the left side of the river. Again, the process was slowed down for political reasons but it was reopened in 1990 after the integration of Portugal in the European Community.

During the following eight years, several studies were developed for these two sites concerning the economical and operational feasibility. Finally, in 1998-99 environmental impact assessment studies looked separately at Ota and Rio Frio site locations, with a pure project perspective. Again Ota was selected as the site for the construction of the new airport of Lisbon, a site that appeared to meet both environmental and economic objectives. Government decision arguments were based on the natural sensitivity of the Rio Frio site, which would involve the destruction of more than 50 000 cork trees, a protected species and habitat in Portugal, and the fact that Rio Frio occupies an ecological corridor that spans between the Tagus and Sado rivers.

However, the issues were not closed here. A national debate started then mainly because of the high costs of construction at Ota due to environmental problems, partly derived from the hydrological and topographical complexity of the site. At the same time increasing tourism and urban development pressures in the southern bank of the river were challenging the ecological sensitivity of the area that was saved from the location of the airport at Rio Frio. Lisbon surroundings within a 50-kilometre perimeter were definitely changing, and ecological concerns were increasingly intense. Despite all the debate, decision was maintained and in 2005 the government took the final decision to build the airport at Ota. The detailed project design continued and the EIA was started.

Some people and organisations, however, were uneasy about this decision, in particular the business sector. In the first semester of 2007, when the EIA process was half way through, a study sponsored by the Confederation of the Portuguese Industry (CIP) (IDAD, 2007) screened the surrounding area up to 50 kilometres, centred in Lisbon, for possible strategic locations. Rather than first identifying sites and then checking on their adequacy for the purpose, this study searched strategically for the best locations that would support the airport from various view points: international connections, regional development of Lisbon metropolitan area, relevance for tourism and industrial development, ecological sensitivity, physical features and infrastructures, population and mobility.

The objective of the CIP study was to show that it was possible to identify new feasible sites, applying the same assumptions with new technological tools and recent environmental data, and did not pretend to discuss if the previous decision was right or wrong (Coutinho and Partidário, 2008). It was the first time such open territorial search was undertaken. With the support of GIS, the CIP study identified a new site – Campo de Tiro de Alcochete (CTA), a shooting range, a military facility that had never been considered in previous studies. This site would avoid, based on a three months fast-track study, many of the problems pointed out for Ota, particularly those that represented a higher economic burden and which could undermine the long-term feasibility of the investment.

Once the study was completed, it was presented to the government and access was opened to the public through the internet a week later, right after the government announced that the previous decision was suspended. During the development of the study a high degree of confidentiality was established (Coutinho and Partidário, 2008) to avoid leakage to the press and preventing additional political pressures. Yet, once finished transparency of results was ensured. The report was focused on the few decision factors that supported the previous decision. The language used was accessible and the methodology used avoided complex models, based on simple technical approaches. This allowed a rapid understanding of the outcomes of the study by the general public.

A week after the CIP study was delivered, the Minister of Public Works (MOPTC) announced in Parliament the suspension of the Ota decision and that a strategic comparative study between Ota and CTA would be commissioned. During this process, negotiations occurred at the top level which included the President of Republic, Prime-Minister and the Portuguese Air Force. New strategic issues had been brought up to the negotiation table, the previous decision had been challenged with a new strategic logic.

The government commissioned the National Laboratory of Civil Engineering (LNEC) to develop a strategic comparative assessment between Ota and CTA (Figure 1). A team of over 50 experts was put together, nearly 40 of them under the coordination of the President of LNEC, with a double mandate: first to check on the technical (physical and engineering) feasibility of CTA to support the construction of an international airport. Secondly to conduct a comparative assessment of Ota and CTA alternative locations, driven by strategic objectives concerning the role of the international airport for the sustainable development of the Lisbon region, and of Portugal at a global level.

The LNEC adopted a strategic assessment methodology (LNEC, 2007) which was constrained by the following facts: 1) the government had commissioned the study to deliver results within a six-month period; 2) there were many details on the project design for Ota site, and on the Ota location as well, but no project design details for the second location at CTA, or any site studies; 3) the intention of the government was only to get the necessary information that could support a decision on the best strategic location for the new airport, based on two alternative locations. Such a study should provide the arguments to justify the need to change, or not, the previously taken decision.

The author was contracted as a consultant to LNEC to lead the methodological approach. The whole methodology was designed to perform strategically and provide answers in a short period. Not much time could be allocated to baseline studies. We only had six months to deliver our advice to the government. A highly pragmatic and focused approach to SEA was adopted. Following a strategic-based SEA methodology developed by Partidário (2007a), an assessment framework was developed around seven critical factors for decision-making (CFD):

1. Safety for air navigation and transportation;
2. Natural resources and risks;
3. Biodiversity and nature conservation;
4. Accessibility;
5. Spatial planning;
6. Social and economic competitiveness; and
7. Financial feasibility.

Each of these CFD adopted environmental, social and economic assessment criteria and indicators that ensured the consideration of the key decision factors. The study also included a cost-benefit analysis, that shared some indicators with the SEA, but which ultimately concluded on the equivalence of both locations from an economic standpoint. Multiple meetings were convened involving the whole team as one group, as well as in thematic groups. Much interaction was enabled across the team through these meetings and the final result was reasonably integrated.

Eventually, this study identified CTA as the preferred location and the advice was forwarded to the government that the location at CTA offered comparative advantages to that of Ota. This advice was adopted by the government leading to a radical change in decision. Final decision was made at the end of the second semester in 2007.

Later, during subsequent debates, the government would choose the first critical factor: Safety for air navigation and transportation, as the determinant factor for decision. This was rather paradoxical since that had been one of the major criticisms to the Ota location, but it had never been considered or put on the decision table before!

How is that SEA made a difference to this process? How did it change the outcomes? Firstly, the whole assessment was narrowed down to a few key decision factors and experts involved were constantly asked to be focused and to keep the essential aspects that would enable the comparative assessment. Seven critical factors for decision-making have driven the whole assessment. The outcomes were presented according to that framework, which was very easy to perceive and to communicate.

Secondly, the entry point for SEA was an important issue. The CIP initiative to screen out for a better location, indicating that a new site for the construction of the airport had been identified at CTA, created a political opportunity to use SEA strategically. CTA study outcomes alerted the government attention to a possible way out to a public conflict that the government was faced with, and which kept the government under a tremendous public pressure.

Thirdly, the outcomes of both CIP and LNEC studies pointed to a similar conclusion, while there was total independency between these two studies and institutions. CIP is a private NGO and represented the vested interests of the private sector. LNEC is a public research institution and was commissioned by the government to develop robust technical assessments. There were no pressures or influence on the LNEC study to try to meet the CIP study outcomes. Much on the contrary, the media and other public forces made all sorts of speculative comments that included both possibilities: that the

LNEC study would want to meet the CIP study results, as well as that the LNEC study would want to maintain the government previous decisions.

Fourthly, the SEA was conducted with a sustainable development orientation. All aspects, from physical to ecological, social and cultural, as well as economic, institutional and political, were brought together aiming at a conciliation of interests to the possible extent. Guidelines for follow-up emerged as pathways to be followed. Many impact assessment principles were met (IAIA, 1999): transparent, scientific rigorous, systematic, useful, practical. The SEA good practice criteria (IAIA, 2002) were also almost all met: integration, sustainability-led, focused, accountable.

The methodological approach was based on the following premises:

1. The object of assessment was clearly defined: it was not the airport infrastructure that was being assessed but its strategic location regarding national and regional overall development.
2. The assessment was pragmatically conducted around the seven mentioned critical factors for decision-making. The team was asked not to undertake long dissertations on their area of expertise but to concentrate on explanatory indicators that could reveal the critical aspects. And to be as robust as possible with the available data, within the time frame.
3. A strong interdisciplinary context was ensured across the team. The achievements of each team were closely followed and cross-sectoral interactions were frequent.
4. The long-term perspective was ensured in various ways, the development of scenarios playing a crucial role. These have determined the strategic discussions that influenced many choices made throughout the process concerning the purpose and plausibility of the airport.
5. Guidelines were prepared to orient future actions, rather than mitigation measures that would assume that nothing else could be done except introducing additional measures to minimize physical, or political, unavoidable effects.
6. Even though many pages were written (close to 1 000 in total) and complex methods were used (in many different themes involving modelling and complex calculations), the final report was written in a simple form, albeit longer than desired, but easy to understand. The final outcomes, presented through the seven critical factors for decision-making, were very easy to perceive by the government and very easy to communicate to the general public to support the government change of decision.
7. Indicators used in the assessment were given different weights by the experts while doing the assessment, but the critical factors for decision-making were all equally weighted. This has raised some criticism amongst the public when the results were known: the business sector wanted the economic competitiveness to be more important, the environmental NGOs wanted to have biodiversity to be more important and the municipalities wanted the spatial planning to be more weighted. It was good that no weighting was introduced, it would have been impossible to satisfy all vested interests.

It was indeed a political opportunity to use SEA and to show how useful it can be to assist strategic decision-making. The government got a sound advice at the end of six months, it was clear and the change of decision was easy to justify. Even the Prime Minister would talk about the critical factors for decision-making in his speeches:

> The **PM spoke today**….the decision …for the *Campo de Tiro de Alcochete* (CTA) is supported in "**four of seven critical factors for decision-making**" indicated in the LNEC report: safety, efficiency and capacity of air traffic operations; sustainability of natural resources; economic development compatibility and financial assessment. He underlined "**the report was very clear**" and that its conclusions expose that both sites were viable and sustainable, but **the choice** for CTA is the one that the government favors for safety and operational reasons…it is also safe from an environmental perspective"
> (Público, **10.01.2008 -** http://economia.publico.clix.pt/noticia.aspx?id=1316214)

Yet, there were obvious drawbacks. While well-integrated and quite robust in its conclusions, it dealt with multiple scales and a wide range of perspectives. It engaged strategic-based studies as well as site specific studies, which mixed up the long-term and short-term views, the large and the site-specific scales. This generated much confusion as to the expected outcomes since some aspects of the study pointed towards a more EIA based analysis. While the study was insufficient regarding the consideration of certain aspects that required broader scales, it added too much information on detailed aspects that were not essential.

These multiple scales and details, however, generated another problem – the different expectations and misunderstanding amongst the public and institutions as too what was really the scale and scope of the analysis. The pressure created by the media exacerbated the public reactions and generated a number of expectations, stories and false alarms that created a vicious perception against the robustness of the study and the legitimacy of its conclusions. Ultimately, it created the idea that this was no more than a social and political construction and diversion created by the government, which is entirely absurd, particularly considering the earlier resistance of the government to accept a new location for the airport.

The whole decision process was weakened by the tensions created over the years, particularly over the last 10 to 20 years. This has led to the need to develop complex studies in a short period of time, which also generated tensions within the teams. The existing and exacerbated tensions required greater confidentiality around the Study, which determined weak public and institutional engagement throughout the process and less iterations than desired.

The fact is, however, that after the first shock wave determined by the sudden change, there was a general feeling of acceptance amongst the public. There were angry reactions from municipalities at Ota area of influence, and from other members of the public, such as environmentalist groups who did not want the airport anyway. Nevertheless, despite the whole conflicting debate that was created, the majority of the public, based on the public consultation results, considered the new location to be better, namely due to the safety aspects, one of the critical factors considered in the study. Not surprisingly, and as mentioned above, safety was also the factor used by the Minister of Public Works as the key argument to justify the government decision.

6. FINAL REMARKS

The question that I was asked to address in this paper was: Does SEA change outcomes? My answer is simple: yes it can! But for that to happen, SEA needs to avoid the hard way that we know of EIA experience: a sequence of difficulties, time and resource consuming, a barrier to efficiency, an industry imposed by over-watchful environmental powers. For SEA to be able to change outcomes, it needs to cut links with EIA practice and develop a full-fleshed capacity to act as a strategic instrument to facilitate decision-making.

As argued above and on other occasions, SEA needs to be a decision-making support instrument in its own right, it needs to be wished by decision-makers. SEA must find the right way that will enable reaching the core of decision-making and deliver inputs that are useful and practical, efficient and cost-effective.

The case that was presented on the SEA of the new international airport of Lisbon stands as a significant case-study of the success of strategic approaches in environmental assessment. It did change the outcomes. And that happened because:

1. SEA was highly focused on the decision that was needed to be made. The decision that was offered to SEA was not about whether or not to do a new airport, but it was about where and how to do the airport. Many criticisms have been made to this SEA, namely by environmental NGOs, because it did not question the need for the airport. However, two things have to be made clear. Firstly, the decision was about the best strategic location for the airport, not if an airport was needed. Secondly, part of the arguments developed to strategically discuss the most plausible location, as well as many consultations carried out with stakeholders, demonstrated that the airport was needed.

2. SEA was highly focused on the critical factors that could make a difference to decision-making. Seven critical factors for decision-making were identified and acknowledged as important by the decision-makers and other stakeholders involved. This allowed a much more structured development of studies that contributed to the SEA. Outcomes were clear and to the point. We had only six months to do the study, we could not afford to spend time with marginal issues. The bulk of the material collected went into appendices, the main report only kept the cream of the assessment. It made it easier for decision-making to justify the decision.

3. The methodology adopted to look for a location, and which was developed by the CIP study, was a strategic screen out of the potential territory surrounding Lisbon within 50 kilometres. Several criteria were followed in this first exercise, including economic, infrastructural, social, ecological, combining several policy and physical drivers. Never, in the previous 40 years, had this been done this way.

4. Finally, another reason to make this a success story is that the outcomes of SEA met general public concerns and responded to many questions people had been raising over the years. These concerns were exactly related to the critical factors for decision-making that were

identified. Of course, SEA did not resolve all the problems and there are still different views concerning even the reason for a new airport. However, tensions have decreased significantly and if it were not because of the global economic crisis, this decision could be said to have reached a consolidated and generally accepted stage.

The robustness of strategic studies findings, engaging stakeholders, ensured a mix of evidence support with societal acceptance. Yet, above all, SEA success lies upon the importance of adopting communicative capacities closer to the politicians, being less concerned with the analytical and technocratic forms of environmental assessment. It is by adjusting the speech and forwarding the right messages in a short period, in a very precise and consistent way that one increases the success in hitting the core of decision-making.

BIBLIOGRAPHY

Ahmed, K. and E. Sánchez-Triana (eds.) (2008), *Strategic environmental assessment for policies: an instrument for good governance*. World Bank.

Bina, O. (2007), A critical review of the dominant lines of argumentation on the need for strategic environmental assessment, *Environmental Impact Assessment Review*, 27, pp. 585-606.

Cherp, A.; A. Watt and V. Vinichenko (2007), SEA and strategy formation theories – from three Ps to five Ps, *Environmental Impact Assessment Review*, 27, pp. 624-644.

Coutinho, M. and M.R. Partidário (2008), The new Lisbon international airport: the history of a decision-making process, *Proceedings of the 28th Annual Conference of the International Association for Impact Assessment*, Perth.

Dalal Clayton, D.B. and B. Sadler (1998), *Strategic Environmental Assessment and Developing Countries*, IIED, London.

Gibson, R.B., S. Hassan, S. Holtz, J. Tansey and G. Whitelaw (2005), *Sustainability Assessment: Criteria, Processes and Applications*, 240 pp., Earthscan Publications, London.

IAIA (International Association for Impact Assessment) (2002), *SEA performance criteria*, Special Publication Series No.1 (www.iaia.org).

IAIA and IEA (Institute of Environmental Assessment) (1999), Principles of Environmental Impact Assessment Best Practice (www.iaia.org).

IDAD (2007), Localizações Alternativas para o Novo Aeroporto de Lisboa, EEP 04.07-07/17, Aveiro.

Kørnøv, L. and W.A.H. Thissen (2000), Rationality in decision- and policy-making: implications for strategic environmental assessment, *Impact Assessment and Project Appraisal*, Vol. 18, No. 3, Beech Tree Publishing, UK, pp. 191-200.

LNEC (2007), Avaliação Ambiental Estratégica do Estudo para Análise Técnica Comparada das Alternativas de Localização do Novo Aeroporto de Lisboa na Zona da Ota e na zona do Campo de Tiro de Alcochete – Relatório Ambiental, LNEC, Lisboa.

Nooteboom, S. (2006), Adaptive networks. The governance for sustainable development, Erasmus University, Rotterdam/DHV.

Partidário, M.R. (1999), Strategic Environmental Assessment – principles and potential, Ch. 4, in: Petts, Judith (ed.), *Handbook on Environmental Impact Assessment*, Blackwell, London, pp. 60-73.

Partidário, M.R. (2000), Elements of Strategic Environmental Assessment, *Environmental Impact Assessment Review*, 20, pp. 647-663.

Partidário, M.R. (2001), What makes an SEA an SEA?, 21st Annual Conference of the International Association for Impact Assessment, Cartagena, Colombia.

Partidário, M.R. (2004), Designing SEA to fit decision-making. 24th Annual Conference of the International Association for Impact Assessment, Vancouver, Canada.

Partidário, M.R. (2005), (a) Future challenges of strategic environmental assessment. III Congreso Nacional de Evaluación de Impacto Ambiental, Pamplona: Associación Española de Evaluación de Impacto Ambiental.

Partidário, M.R. (2005), (b) The contribution of strategic impact assessment to planning evaluation, in: D Miller and D. Patassini (eds.), *Accounting for Non-market Values in Planning Evaluation*, Ch. 9, Aldershot: Ashgate Publishing, pp. 153-164.

Partidário, M.R. (2006), Metodologia de base estratégica para AAE – uma proposta, Proceedings of the National Conference on Impact Assessment, APAI, 18-20 October.

Partidário, M.R. (2007), Scales and associated data – what is enough for SEA needs?, *Environmental Impact Assessment Review*, 27, pp. 460-478.

Partidário, M.R. (2007), (a): Strategic Environmental Assessment Good Practice Guidance – methodological guidance. Agência Portuguesa do Ambiente, Lisboa. www.iambiente.pt.

Partidário, M.R. (2008), Strategic-based model for SEA based on Critical Factors for Decision Making (CFD). Proceedings of the 28th Annual Conference of the International Association for Impact Assessment, Perth.

Pope, J., D. Annandale and A. Morrison-Saunders (2004), Conceptualising sustainability assessment, *Environmental Impact Assessment Review*, 24, pp. 595-616.

UNEP (2005), *Integrated Assessment and Planning for Sustainable Development: Key features, steps, and tools*, UNEP/ETB, Geneva.

UNEP (2009), *Integrated Assessment: Guidance for mainstreaming sustainability into policymaking*, UNEP, Geneva.

Wallington, T. (2002), Civic Environmental Pragmatism – a dialogical framework for Strategic Environmental Assessment, PhD Thesis, Murdoch University, Murdoch, Australia.

Wallington, T., O. Bina and W. Thissen (2007), Theorising strategic environmental assessment: Fresh perspectives and future challenges, *Environmental Impact Assessment Review*, 27, pp. 569-584.

FINAL SESSION

1. THE FINAL SESSION PANEL

The final session of the Symposium was a panel discussion between **Cristina Narbona Ruiz,** Ambassador to the OECD, **Paolo Costa**, President of the Venice Port Authority, and **Chris Nash**, Professor at Leeds University, chaired by **Francesc Robusté**. The main points of the discussion are summarized below, followed by the concluding remarks of Francesc Robusté.

Cristina Narbona Ruiz underlined the role of the OECD in developing the new economic model needed to respond to the present economic crisis.

1. Sustainability is the inescapable focus of the changes needed in the economic model:

– We are facing a paradigm crisis: "economic growth at any cost" equals, in fact, high short-term benefits for a minority, but risks and undesirable effects in the medium- and long-term for the majority of citizens and ecosystems.

– Sustainability means perennial progress for all citizens on the planet. It requires intelligence and responsibility, primarily in the wealthiest countries.

– We have to move towards a new "economic rationality". This implies internalisation of environmental and social costs, comprehensive and long-term analysis, consistency among policies and principles of "good governance", both in the public and private domains.

2. The OECD is contributing to the paradigm shift, and this is reflected in its new maxim, "Towards a stronger, cleaner, fairer economy"

– The OECD's work on sustainability is wide-ranging but relatively little-known. It comprises committees and periodic publications on environmental policy, climate change, water, biodiversity, waste, energy, etc., and transport, most of which are elaborated in conjunction with the International Transport Forum. The OECD is also working on indicators of progress that go beyond measuring Gross Domestic Product (GDP).

– The OECD "Green Growth Strategy" is also a vehicle for overcoming the crisis, to avoid merely returning to where we were.

– Key OECD messages on climate change affecting the transport sector are:

- The cost of action to reduce emissions is much less than the cost of inaction.
- The later we act, the more costly will be the measures to be taken.
- The most immediate step should be to gradually eliminate fossil fuel subsidies (a measure agreed upon by the G20, in response to a proposal by the OECD and the International Energy Agency, which requires periodical reports on compliance).
- Other measures should include carbon markets (the more countries and sectors, the better the results), CO_2 taxes, regulation, public support for research, development and

innovation (RDI) in renewable energies and energy efficiency, and public support for reforestation. At the same time, citizen information and rising awareness are playing a major role.

3. Further reflections on sustainable mobility are arising from discussions on the "Green Growth Strategy":

– Sustainability means more than just a "low-carbon economy". The use of land and natural resources must be retargeted to promote biodiversity, conservation and restoration.

– New technologies are essential but insufficient. Demand needs to be redirected towards reducing the need for mobility. This underlines the role of ICTs in the new approaches to urban development and land management.

– Any analysis of the impact of transport should include the whole cycle of production, use of infrastructures and means of transport, and not just the CO_2 balance. Special attention must be paid, for example, to geological studies and impacts on biodiversity.

– Less-polluting vehicles and public transport need to be accessible by low-income citizens.

– The future of sustainable mobility depends on developing a new culture that emphasizes the usage value of means of transport (the service provided) rather than their ownership, i.e. moving towards the so-called "functionality economy".

Paolo Costa's intervention focused on three main questions related to investment in high-speed railways.

1. Why did EU policy focus on the high-speed railways?

– In the mid-1970s, the European Economic Community started to discuss creating a Common European infrastructure policy to give physical support to the achievement of the common internal market and its four freedoms. The policy adopted was to support the construction or the improvement of cross-border infrastructures.

– In the middle of the 1980s, industry strongly supported the political consensus to set up a common infrastructure policy, with shared competency between EU member states. It was included in the Maastricht Treaty (title XII, articles 128-129 in the original version) under the name, Trans-European Network.

– This policy aimed to achieve economic integration, growth and economic cohesion. At first, this network consisted of motorways, railways and high-speed railways. High-speed rail was chosen to replicate the Japanese and French experience, and to answer a growing demand for faster and more reliable interconnections between the main European cities. Initially, air and maritime transport were not included in this policy.

– Jacques Delors' paper on growth and cohesion added two new reasons to invest in the Trans-European Network-Transport (TEN-T) and especially high-speed rail: to boost technological development and competitiveness. Thus, 15 cross-border priority projects have been selected for implementation, most of them high-speed rail links.

– In 2001, the White Paper on EU Transport Policy added a new objective to be achieved through TEN-T: the reduction of carbonisation by investing in railways to alter the modal split both for freight and passengers. For passenger transport, high-speed rail was considered the most important tool to substitute for road and air transport on medium- and long-distance trips.

– The enlargement of the EU led to a revision of TEN-T, adding new priority projects to be quickly implemented and financed. The EU member states and the private sector have so far invested 130 billion euros in the TEN-T, with an estimated total cost of 345 billion euros to 2020.

2. Are high-speed railways delivering sustainable transport?

The second issue to be discussed is whether high-speed rail projects are the right option, both in terms of economic and environmental sustainability, and thus whether the EC should continue with this policy rather than focusing on other options. The preliminary answer could be positive if, in co-ordination with other EU transport policy objectives, high-speed rail will induce an appropriate diversion of traffic from road and air. This implies the following:

– The key point for decarbonisation is that the greenhouse gases emitted during the construction phase of a high-speed line must be compensated for by emissions avoided by operating services on the line. The magnitude of the latter is difficult to assess, because traffic predictions and positive externalities must be carefully estimated.

– In particular, network effects need to be taken into account, not limiting consideration of the impact of a high-speed line investment to services on the line itself when it is part of a high-speed network. Assessment should encompass the whole multimodal network concerned, and include effects such as liberation of conventional railway capacity by opening a new high-speed line.

– Indirect effects are extremely difficult to assess, and have been ignored in a number of recent assessments. This is not satisfactory, as these effects can nevertheless be large.

3. The role of transport nodes in facilitating co-modality

The third issue for TEN-T policy and the development of high-speed rail is the role of multimodal interurban passenger nodes (ports, stations and airports) as pillars for achieving sustainability:

– The main point is that an improvement of accessibility for road and air transport to high-speed rail networks, through multimodal terminals, enhances the competitiveness of railways. This will divert traffic to the railways, thus reducing greenhouse gas emissions. One example is the importance of a high-speed rail station at Venice Airport, both for the environment and for the competitiveness of rail services.

Chris Nash underlined that:

– Reducing greenhouse gas emissions from the transport sector requires major change; much depends on developing low-carbon fuels, such as second generation biofuels, or much greater production of nuclear electricity, and this will take a long time.

– High-speed rail and public transport more generally makes only a small contribution to reducing CO_2 emissions and, by extension, TEN-T policies can only make a marginal contribution to the environmental pillar of sustainability. Moreover, much of the traffic on high-speed rail is traffic diverted from conventional rail, which is associated with lower greenhouse gas emissions.

There are missing links in the high-speed rail network, and filling them would generate network benefits. However, many of the TEN-T projects identified at present are not part of the “core network” and correspond rather to national priorities.

– The EU plays an important role in driving an international approach to transport policymaking, and this is essential for freight transport due to its characteristics:

  - An important part of freight transport is international, over long distances;
  - Freight can be diverted from road transport to rail;
  - There is relatively little generated traffic as a result of investment in new infrastructure in the case of freight.

– In the framework of the TEN-T, freight traffic should therefore be given high priority.

– Much of the funding for high-speed rail projects comes from EU Cohesion Funds. Is funding megaprojects the most effective way of promoting cohesion? High-speed rail is suited to serving markets with very large populations, but a focus on upgrading conventional rail could achieve more overall. The high-speed rail system in Europe is characterised by high costs, a low level of interoperability and technical complexity. At the same time, a consistent approach to pricing for the use of infrastructure has yet to be found. Policies that focus on cost recovery through high access charges are bound to produce socially suboptimal use of available infrastructure, and moderating infrastructure charges could do more for cohesion than investment in more megaprojects.

– A package of reforms to make a significant contribution to the sustainability of the European transport system would include: moderating charges for the use of rail infrastructure; upgrading some rail links; and promoting road pricing for demand management, with more internalisation of environmental and congestion costs.

2. CONCLUSIONS BY FRANCESC ROBUSTÉ

The Madrid Symposium provided a wide-ranging overview of themes concerning the future of interurban passenger transport. Francesc Robusté's concluding remarks summarise the 10 main topics as they were addressed by the papers presented and the debates that followed.

Sustainability

We are facing strategic challenges in which sustainability is a necessary condition for economic development. In this connection, we should give priority to renewable energies as well as energy-efficient and low-emission industries. It is also very unlikely that we will return to "business as usual", and future growth will be more moderate than in the past.

There is an increasing need to measure progress in terms other than GDP and to include welfare and human capital as complements to services and the production of goods. In addition, the social aspects of mobility mean that we must guarantee people's right to access transport.

In the 1970s the increase of capacity and road productivity was the obvious solution to the ever-expanding demand for infrastructure. However, this sent the public the message that "you do not need to change your habits." Today, this attitude will diminish in developed countries with the growth of environmental awareness. Nevertheless, establishing a new paradigm will require a cultural change in which demand management and optimisation are given priority over the development of infrastructure. In this process of change, ITS will likely play a major role.

Spatial patterns

We are also facing changes in the economics of geography, for the economic territory of the 21st century is being re-shaped and even re-defined. Transport costs still shape the spatial pattern of economic activities. Moreover, we are not yet facing "the death of distance", nor do we live in a "flat world". Proximity is still very important. However, geography in general has been re-scaled from local proximity (neighbourhoods, local cities) to global proximity.

Increasing returns drive agglomeration. Many large cities are changing in structure from a single Central Business District into much larger polycentric cities with. The general trend seems to be toward a "city of cities", i.e. a network of cities that are in competition but are also in co-ordination with each other. We will probably face "comperation" (competition with cooperation) between "megalopolises" and also between different modes of transport.

A trade-off has to be made between spatial coverage and the critical mass of demand when defining a public transport network such as HSR, which is usually defined as rail service faster than 250 kilometres per hour. This trade-off is causing a shift away from slogans such as "no provincial capital more than 3 hours away from the country's capital", towards "no provincial capital with more than X thousand inhabitants more than Y minutes away from an HSR station."

Strategic *vs.* operational decisions

Infrastructure planning models based on a "demand-serving" principle are increasingly complex and dynamic, requiring a great deal of data on behavioural insights – research in this area is still in progress – as well as a high level of training for modellers and administrators.

Some regional planners propose infrastructure planning models based on a "demand-shaping" principle, also called "supply models", in which transport infrastructure shapes the economic territory. There is a need to quantify transport "supply models" that go beyond demand. While they may ensure territorial structuring and social cohesion, investment decisions need to respond to economic logic that considers the critical mass of demand related to investment, operation costs and the timing of implementation.

Empirical evidence shows that both individuals and governments may make strategic decisions irrespective of their economic rationality. While all investments of public money should be submitted to cost-benefit analysis (CBA), some special transport projects such as HSR may contain hidden attributes for users beyond considerations of time, cost, comfort, reliability, etc., in a way that is similar to how public money is allocated to the arts and education, for example. During the Madrid Symposium, after recognising the need for economic rationality, the floor raised the issue of whether economic rationality should predominate in all transport infrastructure investments. Strategic proposals made by the floor included development of the European Union, the "need for TEN-T network" and "Strategic Environmental Assessments".

Some decisions regarding the implementation of HSR networks are strategic rather than operational and should consider all the effects of the projects for all stakeholders involved, including users, non-users, operators, competition modes, industry and society. In addition, CBA should be supplemented by input-output tables and multi-criteria analysis, such as those described in the Railway Project Appraisal Guidelines of the European Community and European Investment Bank (www.railpag.org).

A new quantitative methodology for assessing strategic decisions is needed. For example, most multi-criteria analysis methods rely implicitly or explicitly on a utility function that ranks alternatives. However, methods like ELECTRE define "over-classification" relationships with thresholds, such as "we will not do this project if it does not meet a minimum demand or maximum environmental impacts or costs," and they are thus more suitable for effective strategic decision-making.

Efficiency

Minimising the "social cost" of transport modes (including cost, time and the monetary value of externalities) might favour faster modes with high time savings values (see "Time and speed"). In addition, short-run marginal cost should be used for pricing, but long-run marginal cost should be used for investment coverage. This is especially critical in HSR, conventional rail systems and maritime transport. Moreover, a rational approach should link investment, demand and pricing because generalised prices depend on cost, time and government investment and pricing decisions.

Railways have evolved into vertically integrated national monopolies with some route competition. On-track competition (the subject of current policy momentum) and off-track competition are problematic and are exacerbated by capacity scarcity and network effects.

Railways need a re-engineering process similar to that of low-cost airlines several decades ago, and policy makers and administration civil servants should play a leading role in this process. If they do not, and if administrations fail to devise incentives that are efficient (which could include shorter concessions periods, *bonus* and *malus* incentives, etc.), this adjustment will be a long, drawn-out process.

No allocation of public money should be made without a CBA, because of the opportunity cost of public funds. HSR systems may be profitable in the long run but the question is whether they are today. HSR systems need a minimum of 6 million passengers during the first year of use in order to be feasible. This demand threshold may increase for higher unit costs and longer corridors up to 9 or 12 million passengers the first year. (See "Strategic vs. operational decisions").

HSR systems are useful for medium distances (150 km to 600 km), provided that there is suitable geography as well as construction and operation efficiencies and large demand in the corridor. However, there are other variables besides infrastructure capacity, costs and demand that determine the efficiency. For example, a European Commission project dealing with freight transport by rail, FERRMED (www.ferrmed.com), provides some standards that constitute a real re-engineering exercise.

Public Service Obligations (PSO), as established in Europe, might make the system less efficient but they are aimed at serving social goals.

Motorists pay their external costs in off-peak time rather than during peak time periods. A question was raised by the floor regarding the possible reduction of subsidies for commuter costs because commuters are increasingly making a trade-off between unit land prices and household size and their transport costs.

Although the Pareto optimum is a mixture of pricing and regulation, pricing is a powerful tool for regulating demand, making the system more efficient and generating revenues in times of increasing demand for public resources.

Time and speed

Demand for speed is directly linked to GDP and individual welfare and is continuing to rise. We all behave with an increasing awareness of our travelling time and a desire to reduce it so as to fit more tasks and activities into our extra time.

"Most people don't know where they are going, but they want to get there as fast as possible." We are facing "more distant, faster and more frequent trips, for a shorter stay", i.e. something similar to promiscuity in mobility patterns. At the same time, this immature perspective is being challenged by the recent development of trends such as "praising slowness", tantric philosophy and eco-mobility attitudes.

The contributors raised the question "do we need faster modes?" Speed incurs costs: economic (e.g. Concorde), social (accidents...) and environmental (emissions and climate change). Speed also has non-linear effects, but it has a linear proximity role, which is to overcome distances (see "Sustainability").

Pricing

High-occupancy tolled lanes (HOT lanes) and other value pricing schemes may provide possible ways for some countries to partially introduce road pricing. Although the Symposium did not address freight issues, the implementation of the *Eurovignette* (perhaps after the current economic recession) requires a European perspective. The freight transport sector needs the assistance of institutions and consumers to transfer the extra cost to manufacturers, since the income shortage will probably limit the "tax neutrality" principle. We therefore need to progressively apply road pricing to light vehicles.

In this sense, it is useful to have a European policy with a coherent, harmonious and stable model for infrastructure financing and service, provided it is based on scientific principles. Pricing schemes generate revenues that can benefit society as a whole, but attention needs to be given to social acceptability, since the more efficient the pricing scheme is, the more likely that it will be unpopular.

There is also an opportunity for the yield management of long-distance public transport modes as it has been successfully applied in the air transport sector.

HSR has very high access charges in some European countries, and the higher they are, the lower the likelihood of competition (see "Deregulation").

Deregulation

If railways had been invented later in history, they would probably have found their natural place in the transport system and would have had the right "chromosomes" for market and customer oriented services. While an evolutionary leap in this sense is possible (which would require a philosophy of low-cost service), their current position still has great political influence. As it was said by the floor, "free-market" is not the right expression for railways.

It is simpler to regulate intercity bus services than railways and urban bus services. In fact, bus deregulation can be viewed as a sub-problem of railway deregulation. Some participants raised the issue of generalised "government failure" when dealing with public transport deregulation. The longstanding protection of railways has restricted the development of more flexible modes. Sometimes versatility, which is an advantage in the short run, can be a disadvantage in the long run: "Rigidity is powerful vs. versatility is beautiful".

For long-distance bus services in Europe (scheduled passenger transport for voyages over 50 km with published timetables and open to everyone), competitive tendering does not seem to be the preferred choice. Deregulation has shown that it can work and the markets seem to remain sufficiently competitive, both in an intermodal and intramodal sense.

While Directive 2007/58 introduces open access for international services with domestic cabotage from 1 January 2010, private operators will still see the situation from a financing perspective and will continue to demand a secure environment for their investment.

Decarbonisation

The decarbonisation principle should be applied to all modes of transport. The rail sector continues to claim that investment in rail infrastructure will bring large environmental benefits. However, these benefits are not empirically significant. In fact, contributors from the floor claimed that, from an emissions standpoint, investment in conventional fast trains may be significantly more beneficial than HSR.

Strategic Environmental Assessment is a practice that aims at incorporating the environmental dimension into strategic decisions. "Green Growth Strategy" has a broader meaning than the "Low Carbon Economy" and some forgotten opportunity costs such as losing bio-diversity will have to be evaluated.

Innovation and the future

Technological innovation appears to have been slow in the transport sector, with the brakes on innovation mainly being the behaviour of stakeholders and the failure to deregulate and foster competition. Transport system users will want real-time, comprehensive information systems, efficient government responses to incidents and realistic forecasts. In fact, it is likely that in the future, administrations will include a guarantee of mobility services as a part of the "service charter" for their citizens.

In the future, we can look forward to seeing current emissions and energy problems partially solved, but other externalities such as safety and congestion problems are likely to remain with us for as long as we move atoms at a certain speed.

Although the intermodality concept is very old, its implementation needs to be boosted. HSR stations should be viewed as public transport interchanges that play a functional, informational and economic role. HSR linking a hub airport may create new accessibility that will probably affect regional airports.

In some countries, dedicated truck lanes can be implemented on some motorways, provided that demand is high and the environmental costs are not too high to allow this kind of transport supply segregation. Exclusive truck lanes could favour the operation of megatrucks (see "Sustainability").

Competition will likely face smart regulation schemes that improve efficiency, incentivise performance and ensure social and territorial cohesion and PSOs. There is a great need for innovation in procedures and system management and ITS can also have an important role to play.

Maglev may extend the effect of HSR systems due to its higher speed, but its CO_2 emission efficiency is one-third that of HSR. It is unlikely that it will be massively implemented in Europe because of the huge investments recently made in HSR, for when technological changes are implemented in practice, they must also follow CBA feasibility rules.

Research and policies

In the relationship between policy-making and research findings, urgent matters are often given priority over more crucial issues. Nevertheless, there must be greater interaction between researchers, policy-makers and professionals so that researchers can tackle the key topics that policy-makers want to address and so that policy-making can be based on science.

A general impression is that the civil administration should "do its homework" in the transport field. "Politicians must be brave," as the floor remarked. The mobility sector in Europe is powerfully influenced by the public sector. Public officials need to have the proper training in the contributions of new research so that they can appreciate more sophisticated models and the importance of adequate data for sound decision-making.

Despite an initial aversion to pricing, it has been shown to be an efficient mechanism for regulating demand and raising funds, particularly during times when there is likely to be increased pressure on the availability of public money. A coherent, harmonious and stable model for transport infrastructure provision and service management, based on scientific grounds, could help to clarify the "rules of the game" both for private operators and the various countries. In this framework, public-private partnership schemes could be devised by quantitatively regulating genuine win-win positions.

A cultural change in attitudes towards infrastructure management and sustainable growth is required, which involves raising awareness through information and education. The current period of crisis and excess infrastructure capacity provide the opportunity to design a road map for this soft but extremely important change.

Lastly, more research and innovation are needed, with cross-fertilisation between economists, engineers, geographers, demographers, sociologists, psychologists, environmentalists, planners, etc., since their analytical methods are often complementary. Translating the findings of research into transport policy is critical to ensuring the competitiveness of regions, cities and companies and enabling a better use of public resources. Some countries are devoting a small percentage of their transport infrastructure investment budgets to research, with successful results.

ALSO AVAILABLE

Biofuels: Linking Support to Performance. Series ITF – Round Table 138 (2008)
(75 2008 02 1 P) ISBN 978-92-82-10179-7

Oil Dependence: Is Transport Running out at Affordable Fuel? Series ITF – Round Table 139 (2008)
(74 2008 03 1 P) ISBN 978-92-82-10121-6

The wider Economic Benefits of Transport: Macro-, Meso- and Micro-Economic Transport Planning and Investment Tools. Series ITF – Round Table 140 (2008)
(74 2008 04 1 P) ISBN 978-92-821-0160-5

17th International Symposium on Transport Economics and Policy: Benefiting from Globalisation – Transport Sector Contribution and Policy Challenges (2008)
(74 2008 01 1 P) ISBN 978-92-821-0168-1

Privatisation and Regulation of Urban Transit Systems. Series ITF – Round Table 141 (2008)
(74 2008 06 1 P) ISBN 978-92-821-0199-5

The Cost and Effectiveness of Policies to Reduce Vehicle Emissions. Series ITF – Round Table 142 (2008)
(74 2009 01 1 P) ISBN 978-92-821-0212-1

Port Competition and Hinterland Connections. Series ITF – Round Table 143 (2009)
(74 2009 02 1 P) ISBN 978-92-821-0224-4

Terrorism and International Transport: Towards Risk-based Security Policy. Series ITF – Round Table 144 (2009)
(74 2009 03 1 P) ISBN 978-92-821-0231-2

Competitive Interaction between Airports, Airlines and High-Speed Rail. Series ITF – Round Table 145 (2009)
(74 2009 05 1 P) ISBN 978-92-821-0245-9

Integration and Competition between Transport and Logistics Businesses. Series ITF – Round Table 146 (2010)
(74 2010 01 1 P) ISBN 978-92-821-0259-6

To register for information by email about new OECD publications: www.oecd.org/OECDdirect
For orders on line: www.oecd.org/bookshop
For further information about ITF: www.internationaltransportforum.org

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(74 2010 02 1 P) ISBN 978-92-821-0265-7 – No. 57351 2010

The future for interurban passenger transport

Bringing citizens closer together

Economic growth, trade and the concentration of population in large cities will intensify demand for interurban transport services. Concurrently, the need to manage environmental impacts effectively will increase. How successful we are in coping with demand will depend on our ability to innovate, to manage congestion, and to improve the quality of transport services. Technological and regulatory innovation will shape the future of transport.

The Symposium brought together leading transport researchers from around the world to explore the future for interurban passenger transport. A first set of papers investigates what drives demand for interurban passenger transport and infers how it may evolve in the future. The remaining papers investigate transport policy issues that emerge as key challenges: when to invest in high-speed rail, how to regulate to ensure efficient operation, how to assign infrastructure to different types of users, and how to control transport's environmental footprint by managing modal split and improving modal performance.

www.internationaltransportforum.org

OECDpublishing
www.oecd.org/publishing

(74 2010 02 1 P) ISBN 978-92-821-0265-7

9 789282 102657