

Causal Influence for Ex-post Evaluation of Transport Interventions

13

Discussion Paper 2014 • 13

Daniel Graham
Imperial College London,
United Kingdom

Causal Influence for Ex-Post Evaluation of Transport Interventions

Discussion Paper No. 2014-13

*Prepared for the Roundtable:
**Ex-post Assessment of Transport Investments and Policy
Interventions***

(15-16 September 2014, OECD, Paris)

Daniel GRAHAM

Imperial College
London
United Kingdom

Revised October 2014

THE INTERNATIONAL TRANSPORT FORUM

The International Transport Forum at the OECD is an intergovernmental organisation with 54 member countries. It acts as a strategic think-tank, with the objective of helping shape the transport policy agenda on a global level and ensuring that it contributes to economic growth, environmental protection, social inclusion and the preservation of human life and well-being. The International Transport Forum organises an Annual Summit of Ministers along with leading representatives from industry, civil society and academia.

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 Members of the Forum are: Albania, Armenia, Australia, Austria, Azerbaijan, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Canada, Chile, the People's Republic of China, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, the Former Yugoslav Republic of Macedonia, Georgia, Germany, Greece, Hungary, Iceland, India, Ireland, Italy, Japan, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Mexico, the Republic of Moldova, Montenegro, the Netherlands, New Zealand, Norway, Poland, Portugal, Romania, the Russian Federation, Serbia, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, the United Kingdom and the United States.

The International Transport Forum's Research Centre gathers statistics and conducts co-operative research programmes addressing all modes of transport. Its findings are widely disseminated and support policymaking in member countries as well as contributing to the Annual Summit.

Discussion Papers

The International Transport Forum's Discussion Paper Series makes economic research, commissioned or carried out at its Research Centre, available to researchers and practitioners. The aim is to contribute to the understanding of the transport sector and to provide inputs to transport policy design.

ITF Discussion Papers should not be reported as representing the official views of the ITF or of its member countries. The opinions expressed and arguments employed are those of the authors.

Discussion Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the ITF works. Comments on Discussion Papers are welcomed, and may be sent to: International Transport Forum/OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

For further information on the Discussion Papers and other JTRC activities, please email: itf.contact@oecd.org

The Discussion Papers can be downloaded from: www.internationaltransportforum.org/jtrc/DiscussionPapers/jtrcpapers.html

The International Transport Forum's website is at: www.internationaltransportforum.org

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Table of Contents

Abstract	4
1. Introduction	4
2. Ex-post Appraisal as a Treatment Effect Estimation Problem.....	5
Challenges in estimating the causal effect of treatments on outcomes.....	6
Missing data	7
Non-random assignment and the problem of confounding.....	8
Identification of causal effects via the potential outcomes framework.....	9
3. Causal Methods for Treatment Effect Estimation.....	12
Treatment effect estimation under "ignorability"	12
Estimation given a non-ignorable treatment assignment	13
4. Applications.....	15
Ex-post evaluation of urban road capacity expansions in US cities	15
Ex-post evaluation of regional economic impacts of high-speed rail in Spain.....	20
5. Conclusions	22
Bibliography	23

London, October 2014

Abstract

This paper reviews methods that seek to draw causal inference from non-experimental data and shows how they can be applied to undertake ex-post evaluation of transport interventions. In particular, the paper discusses the underlying principles of techniques for treatment effect estimation with non-randomly assigned treatments. The aim of these techniques is to quantify changes that have occurred due to explicit intervention (or ‘treatment’). The paper argues that transport interventions are typically characterized by non-random assignment and that the key issues for successful ex-post evaluation involve identifying and adjusting for confounding factors. In contrast to conventional approaches for ex-ante appraisal, a major advantage of the statistical causal methods is that they can be applied without making strong a-priori theoretical assumptions. The paper provides empirical examples of the use of causal techniques to evaluate road network capacity expansions in US cities and High Speed Rail investments in Spain.

Keywords: Causal inference; treatment effect; propensity score; ex-post appraisal.

1. Introduction

Ex-ante transport appraisal has well established theoretical and empirical roots in the consumer surplus based calculation of Cost Benefit Analysis (CBA). Recent work on Wider Economic Benefits (WEBs) has extended ‘standard’ CBA to incorporate some key externalities and forms of imperfect competition, again based on clearly set out theoretical and empirical evidence. CBA provides a familiar and well understood approach that is routinely used by Civil Servants, transport professionals, and academics.

Considerably less attention has been paid to ex-post transport appraisal both in theory and in practice. This is presumably because we are generally more interested in predicting how our future investments will fare than in assessing how well we have allocated resources in the past. Yet if we want to obtain a solid understanding of the impacts that transport interventions will have, a good benchmark can be established by studying previous interventions and how they performed according to some defined metrics of interest.

One way of doing this is to simply re-run the ex-ante CBA calculations some time after the project has been completed using observed rather than predicted values. This can provide useful information both on the impacts of the project itself and on how well ex-ante CBA was able to predict the benefits and costs of the scheme. Such calculations are, however, still generated within the theoretical framework of CBA which make a number of quite restrictive simplifying assumptions. An alternative approach is to apply statistical models to data observed before and after transport interventions and attempt to estimate impacts that were caused by the intervention. Such statistical approaches, which we refer to as methods for causal inference, rely more on empirical methods than economic theory, but have their own assumptions and properties that must be met in order to generate valid causal inference.

In this paper we review statistical approaches that are routinely used across a range of scientific disciplines to infer cause-effect relationships from observational data. We argue that a causal inference framework is highly suitable for ex-post appraisal because it is specifically designed to estimate effects that arise when ‘treatments’ are non-randomly assigned, as is the case with most transport interventions. The key consequence of non-random assignment is that the effect of the treatment is ‘confounded’, implying that units in receipt of the treatment (or some particular dose of the treatment) may differ in systematic ways from units with an alternative treatment status. The objective of causal analysis is to estimate the average effect of the treatment (or intervention) net of confounding, or in other words to uncover the marginal causal effect. We refer to this as a treatment effect estimation problem and it is within this context that we discuss possible methods for ex-post appraisal.

The paper is structured as follows. Section 2 discusses ex-post appraisal as a treatment effect estimation problem within the potential outcomes framework for causal inference. It outlines the implication of non-random treatment assignment and the problem of confounding, and then introduces the key assumptions required for valid causal inference. Section 3 describes strategies for consistent treatment effect estimation. It discusses identification of causal effects under ‘ignorability’ via covariate adjustment, propensity score adjustment, and doubly robust methods. It then reviews two approaches that are commonly used when ignorability is not assumed to hold. Section 4 provides two empirical examples of the use of causal techniques for ex-post evaluation: one which evaluates the impacts of urban road network capacity expansions in the US and one which considers the regional economic impacts of High Speed Rail investments in Spain. Conclusions are then drawn in the final section.

2. Ex-post Appraisal as a Treatment Effect Estimation Problem

Our emphasis on causality in this paper arises from the conviction that transport policy is fundamentally concerned with cause-effect relationships. In the UK for example, the following concerns have been highly influential in decision making in recent years:

- What effect will fuel taxation have on transport emissions?
- By how much will traffic volumes reduce under congestion charging?
- How will travel demands change as standards of living rise?
- Will investment in transport infrastructure boost the productivity of the economy?
- How will investment options affect network performance?

Each of these issues involves a cause-effect relationship and the underlying goal of policy is to attempt to shape future outcomes via public intervention. For decision makers the question of interest is: what impact, or outcome, will proposed interventions have.

Ex-post evaluation can help answer this question. By applying statistical models to historic data we can attempt to capture the key relationships of interest and can seek to evaluate the effect of past interventions on defined outcomes. There are, however, two key problems we face in obtaining a causal interpretation from observed data. First, is that we observe only what has actually occurred, not what would have taken place had we intervened in a different way. Second, the interventions we make are rarely randomly assigned and non-random assignment obscures cause effect relationships. Taken

together, these two issues in effect mean that we do not have experimental evidence upon which to base policy decisions.

In this section we outline the potential outcomes framework for causal inference which can be used to obtain a causal interpretation of observational data in the absence of experimental conditions. We discuss the defining characteristics of this approach and demonstrate how it could be used to infer cause-effect relationships for transport interventions.

Challenges in estimating the causal effect of treatments on outcomes

There are three key components that require attention in analysing cause effect relationships from observed data: the intervention (or treatment) to be studied, the outcome of interest, and any relevant characteristics of the units of observation. For ex-post analysis we are fundamentally interested in the effect that a transport intervention (or some set of interventions) has on an outcome. We may wish to know what the outcome would have been had the intervention not been applied, or if some different intervention been applied.

Relevant outcomes of interest could relate to traffic conditions (i.e. speeds, flow, safety, congestion), economic characteristics (i.e. output, productivity, growth), mode share, environmental consequences, social concerns, and so on.

For analytical purposes an intervention in the transport system can be viewed as an observed realisation of random variables whose manipulation produces different outcomes. We refer to such random variables as ‘treatments’, defined in the broadest sense to encompass any ‘regime’ which can be manipulated to produce some effect. For instance, a treatment could involve the construction of a new link, the imposition of speed limits, changes in transport prices, changes in frequency or quality of service, allocation of subsidies, and so on. Treatment variables can be binary, multivalued, or continuous. Table 1 gives relevant examples of transport interventions classified as treatment variables.

Table 1. **Transport interventions classified as treatment variables**

binary	multi-valued	continuous
tolled / untolled route	frequency of service	network capacity
presence of speed camera	speed limit	length of segregated route
20mph zone designation	cars per train	density of intersections
peak / off-peak	no. of O-D routes	accessibility
pedestrianized / unpedestrianized	no. of network nodes	tax / subsidy rates

We are interested in the effect of the treatment on the outcome, but we also recognise that the units under study will likely not have homogeneous characteristics, and these may be relevant to the fundamental relationship of interest. Depending on the purpose of the analysis, and the available data, units could comprise particular transport schemes, network links, people, households, firms, geographical zones, cities etc.

We seek to estimate causal effects using data to represent these three components. We define $z_i = (y_i, d_i, x_i)$, $i = 1, \dots, n$, as a random vector of observed data where for the i -th unit of observation y_i denotes an outcome (or response), d_i the treatment (or exposure) received, and x_i a vector of pre-treatment covariates. As mentioned previously, the treatment can be binary (i.e. $D \in \{0, 1\}$); multivalued, in which dose d can take values in m categories $D \equiv (d_0, d_1, \dots, d_m)$; or continuous with dose d taking values in $D \subseteq \mathbb{R}$.

We want to estimate the effect that treatments have on outcomes. To do so we will draw on the potential outcomes framework for causal inference, which was first put forward for binary treatments in a series of papers in the 1970s by Rubin (e.g. Rubin 1973a,b, 1974, 1977, 1978), although Rubin acknowledges precursors to his approach in earlier works by Fisher (1935) and Neyman (1923). The potential outcomes framework defines the conditions under which we can estimate causal effects from observed data. These are two fundamental issues that shape the potential outcome approach.

Missing data

Ideally, we would calculate the effect of each treatment on an individual (or unit by unit) basis. Thus, for unit i and binary treatment $D \in \{0, 1\}$ we can define two potential outcomes: $Y_i(0)$ if $D_i = 0$, and $Y_i(1)$ if $D_i = 1$. The individual causal effect (ICE) of the treatment is then defined as:

$$\tau_i = [Y_i(1) - Y_i(0)].$$

For multivalued or continuous treatment we can define a potential outcome $Y_i(d)$ associated with each dose of treatment d , with $Y_i = \{Y_i(d) : d \in D\}$ denoting the full set of potential outcomes. The relevant ICEs would then be:

$$\tau_i = [Y_i(d) - Y_i(0)],$$

or the difference between the outcome given assignment to dose d and assignment to no treatment.

A key problem for causal inference, however, is that the data available for estimation reveal only actual outcomes not potential outcomes. For a binary treatment we observe:

$$Y_i = Y_i(1) I_1(D_i) + Y_i(0)[1 - I_1(D_i)],$$

where $I_1(D_i)$ is the indicator function for receiving the treatment, but we do not observe the joint density, $f(Y_i(0), Y_i(1))$, since the two outcomes never occur together. For multivalued or continuous treatments we observe only $Y_i(D_i)$, and outcomes at all other levels, $d \neq D_i$, are unobserved and we refer to these as counterfactual outcomes.

Thus, the problem we face is that the observed data do not provide enough information to evaluate ICEs because we do not observe the potential outcomes arising from treatment allocations that are contrary to fact. Holland (1986) refers to this as a fundamental identification problem of causal inference. A key insight of the potential outcomes approach is that if we focus on estimating average causal effects, rather than ICEs, then we do not have to observe all potential outcomes.

Average causal estimands of interest include Average Potential Outcomes (APOs) and Average Treatment Effects (ATEs). For binary treatments the APOs are:

$$\mu(1) = E[Y_i(1)] \text{ and } \mu(0) = E[Y_i(0)],$$

and the ATE is defined as:

$$\tau(1) = \mu(1) - \mu(0).$$

For continuous and multi-valued treatments the APO under treatment level d is denoted:

$$\mu(d) = E[Y_i(d)],$$

and the ATE is:

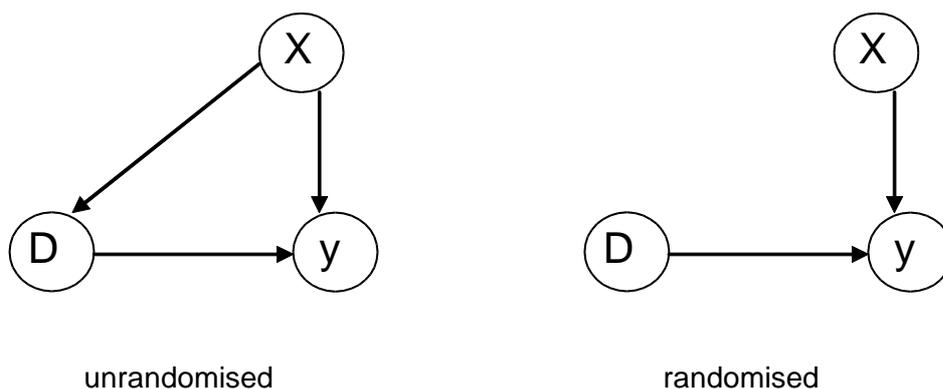
$$\tau(d) = \mu(d) - \mu(0).$$

Other causal estimands can include ATEs on the treated, quantile effects, ATEs for a variety of sub-populations, ATEs conditional on covariates, and causal odds and risk ratios. In this paper the primary concern is with APOs and ATEs as defined above.

Non-random assignment and the problem of confounding

The conditions under which we can use the observed data to estimate APOs and ATEs depend crucially on whether the treatment is assigned randomly or not. Figure 1 below shows a graphical comparison of randomised and non-randomisation treatment assignments.

Figure 1. **Directed acyclic graph of observational data with randomisation and non-randomisation of treatment assignment**



Under a randomised assignment unit characteristics X have no influence on the treatment received (i.e. on D). Consequently, potential outcomes are *unconditionally independent* of the treatment assignment mechanism. For binary treatments randomization implies:

$$Y_i(0), Y_i(1) \perp I_1(D_i)$$

and for multivalued or continuous treatments:

$$Y_i(d) \perp I_d(D_i) \text{ for all } d \in D,$$

where $I_d(D_i)$ is the indicator function for receiving dose d of the treatment. Under a random assignment there are no systematic differences in characteristics between treated or controlled units, or in the case of multivalued and continuous treatments, between units receiving different doses of the treatment. Consequently, we can treat the unobserved potential outcomes much like data that are missing at random and consistent estimators of ATEs for binary and multivalued or continuous treatments can then be formed as:

$$\hat{t}(1) = \frac{\sum Y_i \cdot I_1(D_i)}{\sum I_1(D_i)} - \frac{\sum Y_i \cdot (1 - I_1(D_i))}{\sum (1 - I_1(D_i))}$$

and:

$$\hat{t}(d) = \frac{\sum Y_i \cdot I_d(D_i)}{\sum I_d(D_i)} - \frac{\sum Y_i \cdot I_0(D_i)}{\sum I_0(D_i)}$$

Under non-randomisation, however, allocation of the treatment depends on a set of covariates X which are themselves important in determining outcome Y . Thus, some part of the association between the treatment and the outcome could be attributed to X rather than D . Under these circumstance we refer to X as confounders and note that simple comparisons of mean responses across different treatment groups (as in equations 1 and 2) will not in general reveal a ‘causal’ effect because mean outcomes across treated and control units will differ regardless of treatment status.

Identification of causal effects via the potential outcomes framework

While non-random assignment has consequence for causal estimation, consistent estimates of APOs and ATEs can still be obtained under the potential outcome framework, but adjustment for confounding has to be made. In this subsection of the paper, we define the conditions under which causal estimates can be identified in the presence of confounding.

There are three key assumptions required for valid APO and ATE estimation in the presence of confounding within the potential outcomes framework. These are as follows.

1. **Conditional independence** -- the potential outcomes for unit i should be conditionally independent of the treatment assignment given a (sufficient) set of observed covariates X_i . For binary treatments the assumption requires that:

$$Y_i(0), Y_i(1) \perp I_1(D_i) | X_i, \tag{3}$$

and for multivalued or continuous treatments Imbens (2000) and Hirano and Imbens (2004) introduce the concept of weak conditional independence which can be stated as:

$$Y_i(d) \perp I_d(D_i) | X_i \text{ for all } d \in D. \tag{4}$$

The key difference between the binary and non-binary assumptions is that in the latter conditional independence is required to hold for each value of the treatment (i.e. pairwise), but not joint independence of all potential outcomes.

The conditional independence assumptions essentially require that, conditional on some set of pre-treatment covariates, assignment to treatment does not depend on the outcome. If X_i is sufficient for this to hold then we can in effect mimic, for observational data, the assignment that would occur in a randomised control trial in which the treatment is allocated independently of pre-treatment characteristics.

2. **Common support** - the support of the conditional distribution of X_i given a particular treatment status should overlap with that of X_i given any other treatment status. For binary treatments this requires that the probability of assignment to the treatment lies strictly between zero and one

$$0 < \Pr(I_i(D_i) = 1 | X_i = x) < 1, \forall x. \quad (5)$$

For multivalued or continuous treatments we require common support by treatment status in the covariate distributions within some region of dose $C \subseteq D$. A sufficient condition is that for any subset of C , say $A \subseteq C$,

$$\Pr(D_i \in A | X_i = x) > 0, \forall x \quad (6)$$

The intuition behind the common support, or overlap, assumption is that if some sub-populations observed in X_i have zero probability of receiving (or not receiving) a treatment, then it does not make sense in these cases to talk of a treatment effect since the counterfactual does not exist in the observed data.

3. **Stable unit treatment values** - the relationship between observed and potential out-comes must comply with the Stable Unit Treatment Value Assumption (SUTVA) (e.g. Rubin 1978, 1980, 1986, 1990), which requires that the observed response under a given treatment allocation is equivalent to the potential response under that treatment allocation. For binary treatments we require that:

$$Y_i = I_i(D_i)Y_i(1) + (1 - I_i(D_i))Y_i(0) \quad (7)$$

for all $i = 1, \dots, N$. For multivalued or continuous treatments we require:

$$Y_i \equiv I_d(D_i)Y_i(d) \quad (8)$$

for all $d \in D$, for all $Y_i(d) \in Y_i$, and for $i = 1, \dots, N$.

The SUTVA requires that the outcome for each unit be independent of the treatment status of other units, or in other words, there should be no interference in treatment effects across units (Cox 1958). It also implies that there are no different versions of the treatment. The no-interference assumption is generally satisfied when the units are physically distinct and have no means of contact. Violations of the assumption can occur when proximity of units allows for contact and this presents a particular concern for transport applications.

The three assumptions defined above, which are together referred to by Rosenbaum and Rubin (1983) as strong ignorability, allow for identifiability of causal effects from observational data. In the case of binary treatments the ATE can be derived as:

$$\tau = E_i(Y_i(1) - Y_i(0)) = E_X [E_i(Y_i(1)|X_i = x) - E_i(Y_i(0)|X_i = x)] \quad (9a)$$

$$= E_X [E_i(Y_i(1)|X_i = x, I_1(D_i) = 1) - E_i(Y_i(0)|X_i = x, I_1(D_i) = 0)] \quad (9b)$$

$$= E_X [E_i(Y_i|X_i = x, I_1(D_i) = 1) - E_i(Y_i|X_i = x, I_1(D_i) = 0)]. \quad (9c)$$

Conditional independence justifies the equality of (9a) and (9b), the SUTVA allows the substitution of observed for potential outcomes to give (9c), and overlap ensures that the population ATE in (9c) is estimable since there are units in both the treated and untreated groups. Note that the ATE is defined as an expectation over covariates X . If we do not take this expectation, but instead simply use the integrand, we obtain an estimate of the causal effect of D within strata of X . In other words, we get the conditional treatment effect, that is the average treatment effect for units with characteristics $X = x$. By integrating X out of this distribution we get the average causal intervention distribution.

For continuous or multivalued treatments the APO under a given dose $D = d$, $\mu(d) = E[Y_i(d)]$, or the dose-response function, can be derived as,

$$E[Y_i(d)] = E_X [E(Y_i(d)|X_i)] = E_X [E(Y_i(d)|I_d(D_i), X_i)] = E_X [E(Y_i|I_d(D_i), X_i)], \quad (10)$$

where the second equality follows from conditional independence, the third from the SUTVA, and the overlap assumption ensures that the APO is estimable since there are comparable units across treatment levels.

3. Causal Methods for Treatment Effect Estimation

The literature on methods for causal estimation is vast and growing at a rapid rate. Consequently, a truly comprehensive review of the field is outside the scope of this paper and in fact would make little contribution since excellent up-to-date reviews already exist such as those by Hernan and Robins (2012), Imbens and Wooldridge (2009), Tsiatis (2006) and van der Laan and Robins (2003). Instead, we outline the general principles under which the construction of estimators proceeds. We do so first for methods that assume ignorability and then we consider two popular approaches that are used when ignorability is thought to be violated.

Treatment effect estimation under "ignorability"

Using the notation of Tsiatis and Davidian (2007), we define joint densities of the observed data of the form:

$$f_Z(z) = f_{Y|D,X}(y|d, x)f_{D|X}(d|x)f_X(x).$$

When ignorability is assumed to hold, estimation of APOs and ATEs generally proceeds in one of the following ways:

- a) **Direct covariate adjustment** - leave $f_{D|X}(d|x)$ and $f_X(x)$ unspecified and posit a model for $E[Y_i|D_i, X_i]$; the expectation of the conditional density of the response given treatment and covariates. This is typically achieved via an outcome regression (OR) model such as a Generalized Linear Model (GLM), a Generalized Linear Mixed Model (GLMM), a Generalized Additive Mixed Model (GAMM), or other spline-based approach. ATEs can be estimated directly from these OR models. This regression approach is commonly used in transport analyses.
- b) **Propensity Score adjustment** - leave $f_{Y|X}(y|x)$ and $f_X(x)$ unspecified but assume a model for $f_{D|X}(d|x)$ and use these to form Propensity Scores (PS), which measure the probability of assignment to treatment given the set of observed pre-treatment covariates. An important result, due to Rosenbaum and Rubin (1983), is that the conditional independence assumption (i.e. equations 3 and 4) can be restated by replacing the covariate vector X_i with the scalar PS. Rosenbaum and Rubin (1983) proved this result in the case of binary treatments and Imbens (2000) and Hirano and Imbens (2004) generalise the PS to cover the case of multivalued and continuous treatments.

The PS are to be used to form a number of different nonparametric and semiparametric estimators, via weighting, matching, stratification, blocking and regression (for details, see Imbens and Wooldridge, 2009). A key advantage in using the PS is that it avoids the need to condition on a potentially high dimensional covariate vector and it is this dimension-reducing property that allows for effective implementation of flexible estimators. Another advantage of the PS is that it is highly effective in isolating the region of common support, a task that is difficult using multiple covariates (for discussion, see Joffe and Rosenbaum, 1999).

- c) **Doubly Robust estimation** - leave $f(x)$ unspecified but assume both an OR model and a PS model and form an estimator that combines both models. This is usually achieved by weighting or augmenting the OR model with covariates derived by inverting the PSs. The key feature of doubly-robust estimators is that APO and ATE estimates are consistent and asymptotically normal when either the OR or the PS model are correctly specified, but we do not require both models to be correct (e.g. Robins, 2000; Robins *et al.*, 2000; Robins and Rotnitzky, 2001; van der Laan and Robins, 2003; Lunceford and Davidian, 2004; Bang and Robins, 2005; Kang and Schafer, 2007). The rationale for doubly-robust estimation is that the analyst effectively has two chances at getting the model specification right.

Estimation given a non-ignorable treatment assignment

The validity of the estimation approaches discussed in the previous subsection requires us to maintain that ignorability holds. When this assumption is no longer tenable we have to look at other approaches. In this subsection we review two popular estimators that are used when ignorability is not assumed: differences-in-differences and instrumental variables.

- a) **Differences-in-differences** - a common problem in identifying causal effects is that there may be unobserved differences between the treated and untreated units which affect potential outcomes and are also influential in treatment assignment. In addition, we may suspect that temporal trends affect the outcome variable due to events that are unrelated to the treatment.

The differences-in-differences (DID) estimator addresses such potential sources of bias by using information for both treated and control groups in both pre- and post-treatment periods. The DID estimator approximates:

$$\tau_{\text{DID}} = \{E [Y_i(1)|D = 1] - E [Y_i(1)|D = 0]\} - \{E [Y_i(0)|D = 1] - E [Y_i(0)|D = 0]\} \quad (11)$$

The ‘double-differencing’ of the DID estimator removes two potential sources of bias. Firstly, it eliminates biases in second period comparisons between the treated and control groups that could arise from time invariant characteristics. Secondly, it corrects for time-varying biases in comparisons over time for the treated group that could be attributable to time trends unrelated to the treatment.

It is important to note two potential limitations with the DID approach. First, it relies on the strong identifying assumption that the average outcomes for the treated and control groups would have followed parallel paths over time in the absence of the treatment.

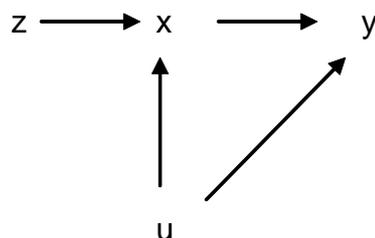
Second, the model is sensitive to error specification, and in particular, it has been shown that the existence of correlation within groups or over time periods can adversely affect the performance of the DID estimator (Bertrand *et al.*, 2004).

- b) **Instrumental Variables (IV)** - the IV estimator is well known and widely used and for that reason we do not provide an extensive review here. The key principles of IV estimation are:

1. Find a set of instruments which are exogenous and highly correlated with the covariates.
2. Use the instruments to enforce orthogonality between the error term and an instrument transformed design matrix.

The relationships assumed in IV estimation are shown graphically below in the context of the linear regression model $y = X\beta + u$ with instrument matrix Z .

Figure 2. Relationships in instrumental variables estimation



The defining characteristics of the IV model are that: changes in z are associated with changes in x , but do not lead to changes in y other than through x ; z is causally associated with x but definitely not with u ; z would not be found in the regression model for y .

A common method used to obtain IV estimates is two-stage Least Squares (2SLS):

1. Regress each column of X on the instrument matrix Z .
2. Regress y on the predicted values from the first stage.

IV can be used to establish causal effects under a non-ignorable treatment assignment and is particularly useful when endogeneity via bi-directionality is present. However, it is crucial that the two key assumptions of exogeneity and relevance are met, and in practice such instruments can be hard to find. When instruments are only weakly correlated with the endogenous regressors, or when the instruments themselves are correlated with the error term, IV estimation can produce biased and inconsistent estimates. This problem is further confounded by the fact that the available diagnostic statistics do not provide a full proof means for detecting an inadequate instrument specification. To quote Hahn and Hausman (2003), even using standard tests for instrument validity “the researcher may estimate ‘bad results’ and not be aware of the outcome” (p 118). In addition it is also worth noting that the IV estimator can be much less efficient than OLS.

4. Applications

In this section we describe two applications of the treatment effect approach for ex-post evaluation of transport interventions. The first relates to an ex-post evaluation of urban road capacity expansions in the US. The objective is to estimate ATEs from road capacity expansion in relation to induced traffic demand, traffic densities and productivity. The study uses a PS based methodology for dose-response estimation for continuous treatments proposed by Graham et al. (2014). The second application considers the regional economic impacts of high speed rail investment in Spain using a DID estimator.

Ex-post evaluation of urban road capacity expansions in US cities

Objective

The objective of this study is to use available longitudinal data to assess the impacts of urban road network capacity expansions on traffic volumes, traffic densities and productivity.

Method

The study uses a GPS based regression methodology to control for confounding and estimate APOs and ATEs. The APOs of interest are defined by $\mu(d) = E[Y_{it}(d)]$ and the ATEs by $\tau(d) = E[Y_{it}(d)] - E[Y_{it}(0)]$, where d is some dose of capacity expansion, i indexes units and t indexes time. Calculations are made for several doses of interest and a dose response curve is formed using penalised spline regression.

As mentioned above, the GPS is defined by $\pi(D_{it} = d | X_{it}; \hat{\alpha})$ and for valid causal inference we require conditional independence and common support:

$$Y_{it}(d) \perp D_{it} = d | X_{it} \text{ and } \Pr(D_{it} \in A | X_{it} = x_{it}) > 0 \forall x_{it}, A \subseteq C$$

where C is a region of common support (e.g. Hirano and Imbens, 2004).

Consistency requires that X_{it} is sufficient to represent confounding. This is, however, effectively an untestable assumption. In longitudinal applications we often assume the existence of unobserved covariates U_i or W_i , say, which could enter our causal model such that $D_{it} = f(X_{it}, U_i, W_i)$ and $Y_{it} \perp D_{it} = d | X_{it}, U_i$. Thus, U_i is a time-invariant unobserved confounding covariate while W_i is a time-invariant unobserved non-confounding covariate. To address this issue, Graham *et al.* (2014) specify a longitudinal mixed (LM) model for the GPS:

$$D_{it} = X_{it}^T \theta_1 + b_i + H_{i,t-p}^{yT} \theta_2 + \varepsilon_{it} \text{ with } b_i \sim N(0, \sigma_b^2)$$

which, in addition to observed time-varying confounders X_{it} conditions on unobserved unit level effects b_i and some lag of the response variable $H_{i,t-p}^y$. Thus the paper proposes a GPS approach to ATE estimation which allows for measured time-varying confounding, unobserved time invariant confounding, and bi-directionality between response and treatment.

Analytical results and simulations presented in the paper show that under given conditions a LMGPS approach will yield unbiased estimates of the dose-response function, but more extensive conditioning can adversely affect efficiency and can render the task of finding overlap in support of the covariate distribution more challenging.

The algorithm for ATE estimation under the LMGPS approach is as follows:

1. Estimate $f_{D|X,U}(D_{it}|X_{it}, b_i; \alpha)$ using a mixed model;
2. Use $\hat{\alpha}$, with some appropriate density function, to calculate LMGPSs for observed $\pi(D_{it}|X_{it}, b_i; \hat{\alpha})$ and unobserved $\pi(d|X_{it}, b_i; \hat{\alpha})$ treatments;

3. Isolate a common support region

$$\Pr(D_{it} \in A|X_{it}, b_i) > 0 \forall X_{it}, A \subseteq C;$$

4. Estimate $E[Y_{it}|D_{it}, \pi(D_{it}|X_{it}, b_i; \hat{\alpha})]$ using a flexible model;
5. Average over predicted values from 4., evaluated at dose d , to obtain a point estimate of the APO at d : $\mu(d)$;
6. Repeat for all dose of interest, form the dose-response curve, and estimate ATEs:

$$\tau(d) = \mu(d) - \mu(0)$$

7. Use a single (block) bootstrap re-sampling scheme over 1. to 6. to obtain standard errors.

Data

The data available for estimation are taken from the Texas Transportation Institute (TTI) urban mobility data, which describe traffic conditions for 101 US cities over the 1982-2007.

Figure 3. Cities in Texas Transportation Institute urban mobility data



Responses: Annual proportional change in demand (vmt), network performance (delay per vmt), and productivity (average wage).

Treatment: Annual proportional change in network lane miles.

Pre-treatment covariates (confounders):

- Lagged responses: to capture reverse causality;
- Congestion & traffic volume: measured by delay and vmt;
- Network scale & mix: network length, mix of freeway / arterial;
- Traffic mix: volume on freeway / arterial;
- Mode characteristics: public transport patronage, state fuel price;
- Economy: productivity, income and economic structure;
- Employment and population distribution and growth.

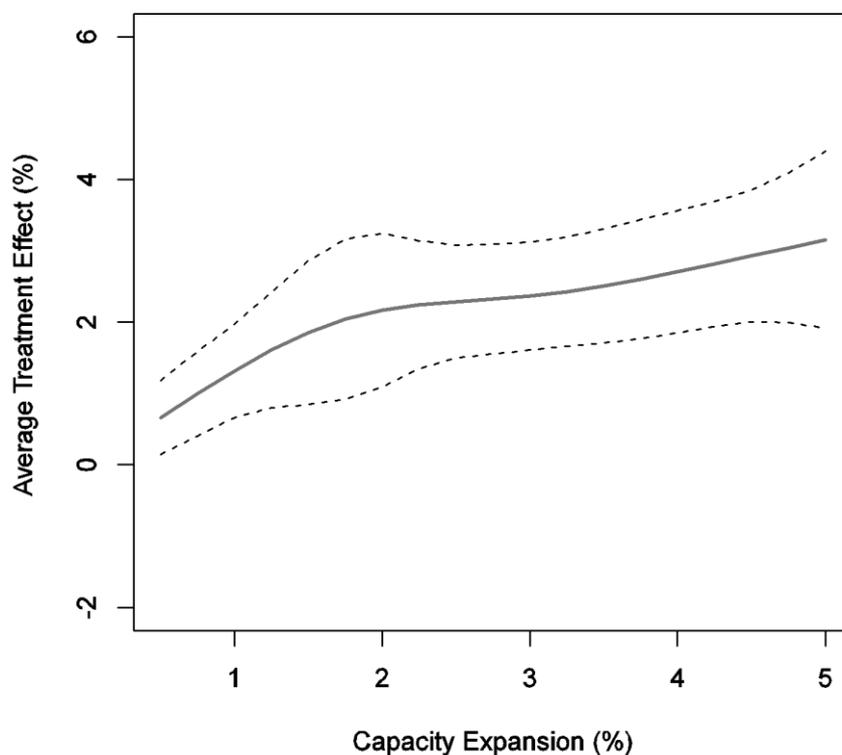
Unobserved (unknown) confounders: zone / area /region characteristics, road network design, activity/travel behaviour.

Results

The results for our three responses are shown graphically below. In each figure the y-axis shows the ATE and the x-axis shows the corresponding dose of capacity expansion.

First we consider results for traffic volumes as measured by vmt.

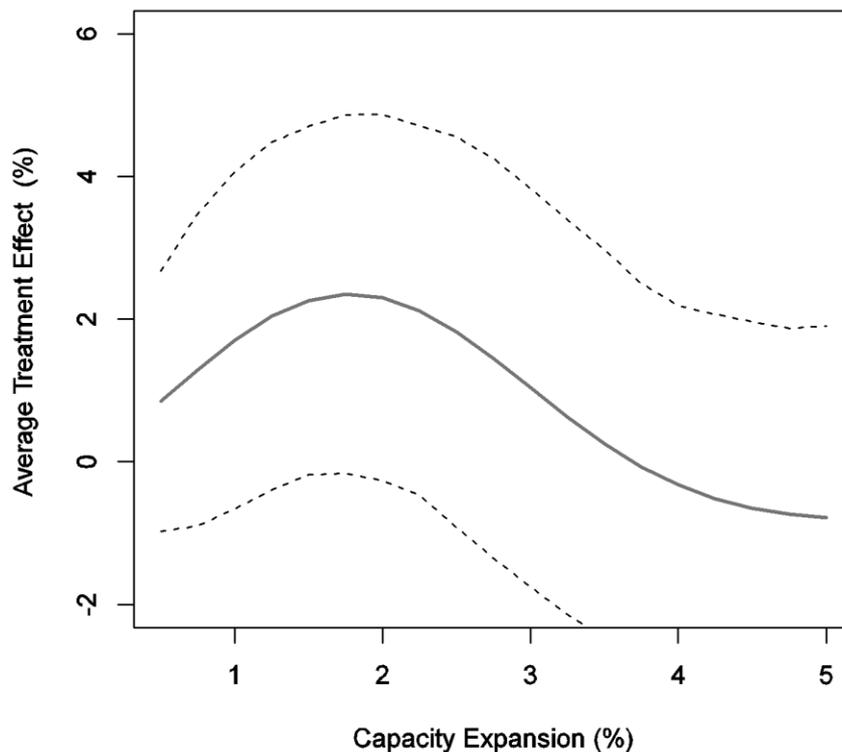
Figure 4. Dose-response relationship for traffic volumes



The results show evidence of induced demand over the range of dose having adjusted for confounding. The ATE is growing faster than capacity for doses of up to 2% increase in capacity. On average we find that a 10% increase in lane miles is associated with a 9% increase in vmt net of ‘natural growth’ (estimated 1.4% per annum). As a consequence of this, we find that capacity expansions in the range considered have not in general reduced traffic density (i.e. the traffic volume to capacity ratio).

Next, we look at impacts on network performance as measured by delay per vmt.

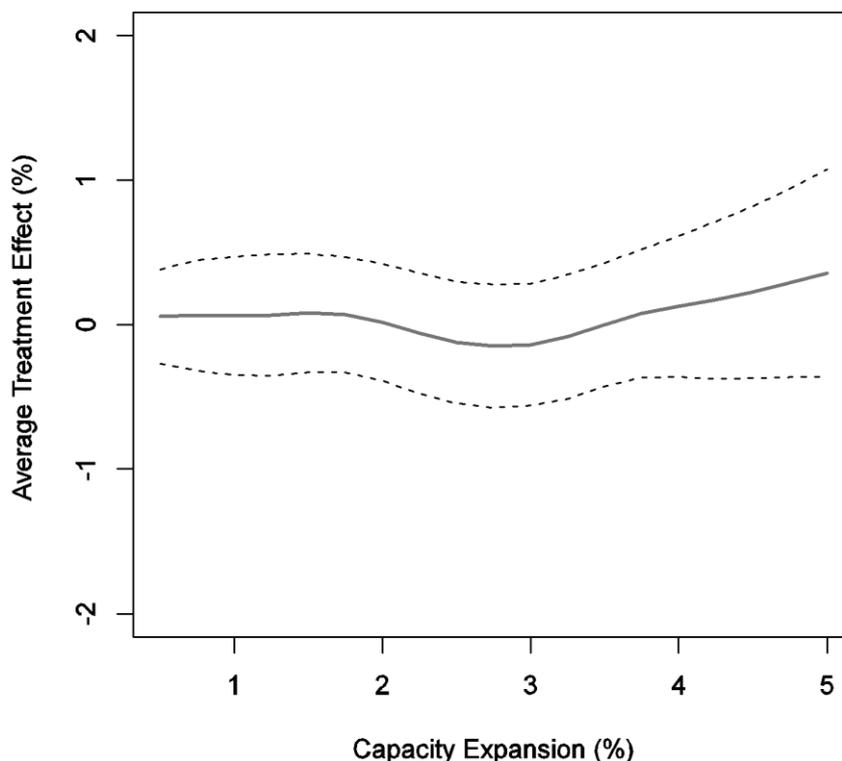
Figure 5. Dose-response relationship for network performance (delay per vmt)



The results indicate that capacity expansions have not ameliorated urban congestion. The average road user has not experienced change in delay from capacity expansions as there has been no statistically significant effects on delay per vmt, and we find this to be the case even for large capacity expansions. In fact, due to natural growth congestion has worsened by approximately 3% per annum and because there is now more traffic total urban delay increases over the range of dose.

Finally, we look at effects on productivity as represented by the average urban wage rate.

Figure 6. Dose-response relationship for productivity (average wage)



The results indicate that urban road network expansions have not induced higher productivity. If we run a 'naive' regressions of productivity on treatment we do find a positive association between capacity growth and wages, but we do not find significant ATEs having adjusted for confounding and isolated a region of common support.

Thus, our causal analysis finds that urban road network expansions have induced demand but have not ameliorated congestion or raised productivity. These results do not imply that there are no economic benefits from road capacity expansions *per se*. The results are specific to marginal changes on mature congested urban networks. While capacity expansions have allowed for increased mobility, in the sense that there is more traffic, network generalised costs have not improved and total urban delay has risen. The scale (increased traffic) effect does not appear to have influence productivity (either +ve or -ve).

Ex-post evaluation of regional economic impacts of high-speed rail in Spain

Objective

Between 2000 and 2010 the Spanish Government carried out the largest high-speed rail construction programme in Europe such that by 2011 the Spanish HSR network had become the largest in Europe, exceeding France and Germany. By 2020, it is planned that 90% of the country population will live within a 50km radius to the nearest high-speed rail station. Investments in high-speed rail projects are

frequently justified on the basis of projected positive effects on regional and national economic growth. In this application, a DID analysis is under-taken to study impacts on economic output arising from construction of the Madrid-Barcelona high speed rail corridor. This is PhD work which is still in progress, and is presented here as an illustration of ex-post evaluation rather than as a definitive statement on the impacts of HSR in Spain.

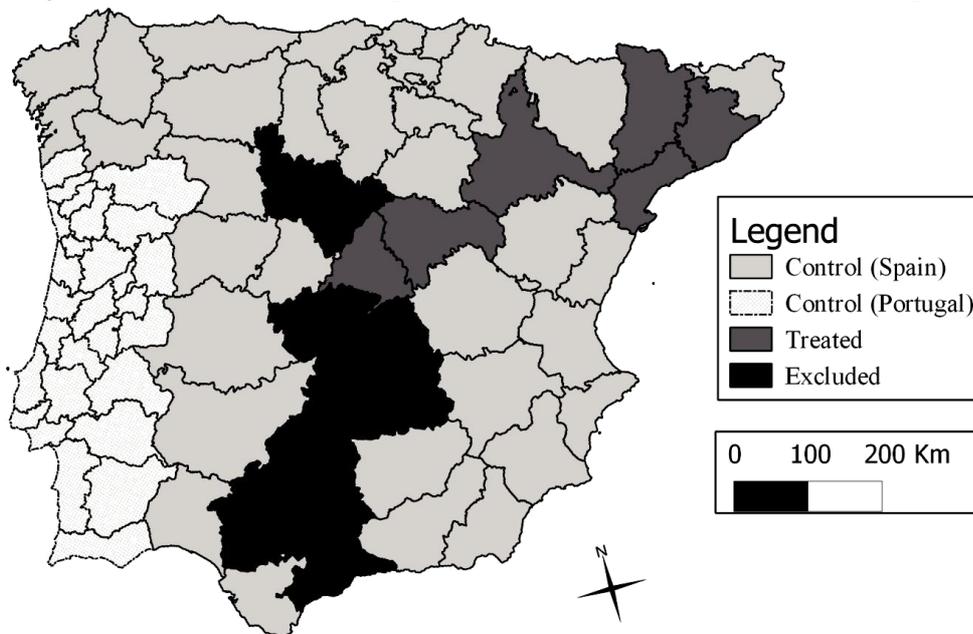
Method

To assess the effect of the HSR corridor linking Madrid with Barcelona we divide the Spanish economy into 47 peninsular provinces and treat access to HSR as a binary treatment. Provinces receiving/not receiving a HSR connection are called treated/untreated and provide the basis for the DID analysis. These are shown in Figure 7. We use GVA per capita (GVAPc) as the outcome variable in province i at time t . We run a regression of this response on the year variable, a binary variable for each of the groups (east, west and north) and the three interaction terms between year and the three potential control groups north, east and west (year*east, year*west and year*north respectively) where the base case is the group of treated provinces. We also estimate the same equation adding two covariates to account for the economic structure of the different provinces: share of jobs in manufacturing, energy and construction sector (share industry) and share of jobs in the services sector (share services).

Results

The results indicate that predictions of a positive impact on the economic performance of regions receiving HSR have not taken place, at least in the short to medium term. In the case of the Madrid-Barcelona HSR corridor, our results show that there are no significant differences in the pattern of regional economic growth before and after the HSR corridor between the treated and untreated provinces.

Figure 7. Control and treated provinces for the north-east corridor of Spain



5. Conclusions

In this paper we have reviewed methods that seek to draw causal inference from observed data and have shown how they can be applied to undertake ex-post evaluation of transport projects. We argue that a causal inference framework based on potential outcomes is highly suitable for ex-post appraisal because it is specifically designed for instances in which ‘treatments’ are non-randomly assigned and experimentation is not possible, circumstances that characterise the allocation of transport interventions. The methods we review have been used extensively for casual analysis across a range of scientific disciplines, but to our knowledge have received little attention in transport analyses. We provide two applications of ex-post evaluation based on causal techniques: one which evaluates the impacts of urban road network capacity expansions in the US, and one which considers the regional economic impacts of High Speed Rail investments in Spain.

A major advantage of the statistical causal methods is that we can use them to analyse the impact of interventions without making strong a-priori theoretical assumptions about underlying economic behaviour, as is required in ex-ante or ex-post CBA. However, valid causal inference from observational data has its own set of rather stringent assumptions, which in many instances may not hold in the available data.

Bibliography

- Bang, H. and J.M. Robins (2005), Doubly robust estimation in missing data and causal inference models, *Biometrics*, 61, 962-972.
- Bertrand, M., E. Duflo and S. Mullainathan (2004), How much should we trust difference-in-differences estimates? *Quarterly Journal of Economics*, 119 (1), 249-275.
- Cox, D.R. (1958), *Planning of Experiments*, London: John Wiley & Sons.
- Fisher, R.A. (1935), *The Design of Experiments*, Edinburgh: Oliver and Boyd.
- Graham, D.J., E.J. McCoy and D.A. Stephens (2014), Quantifying causal effects of road network capacity expansions on traffic volume and density via a mixed model propensity score estimator. *Journal of the American Statistical Association*, DOI:10.1080/01621459.2014.956871.
- Hahn, J. and J. Hausman (2003), Weak instruments: diagnosis and cures in empirical economics, *American Economic Review*, 93, 118-125.
- Hernán, M.A. and J.M. Robins (2012), *Causal Inference. Monographs on Statistics and Applied Probability*, London: CRC Press.
- Hirano, K. and G.W. Imbens (2004), The propensity score with continuous treatments, in: A. Gelman and X.L. Meng (Eds.), *Applied Bayesian modeling and causal inference from incomplete data perspectives*, pp. 73-84. New York: Wiley.
- Holland, P.W. (1986), Statistics and causal inference, *Journal of the American Statistical Association*, 81 (396), 945-970.
- Imbens, G.W. (2000), The role of the propensity score in estimating dose-response functions, *Biometrika*, 87 (3), 706-710.
- Imbens, G.W. and J.M. Wooldridge (2009), Recent developments in the econometrics of program evaluation, *Journal of Economic Literature*, 47 (1), 5-86.
- Joffe, M.M. and P.R. Rosenbaum (1999), Propensity scores, *American Journal of Epidemiology*, 150 (4), 327-333.
- Kang, J.D.Y. and J.L. Schafer (2007), Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data, *Statistical Science*, 22 (4), 523-539.

- Lunceford, J.K. and M. Davidian (2004), Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study, *Statistics in Medicine*, 23, 2937-2960.
- Neyman, J. (1923), On the application of probability theory to agricultural experiments. Essay on principles, Section 9. *Statistical Science*, 5 (4), 465-480. Translated in 1990.
- Robins, J.M. (2000), Robust estimation in sequentially ignorable missing data and causal inference models, in: *Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*, pp. 6-10. Alexandria, VA: American Statistical Association.
- Robins, J.M. and A. Rotnitzky (2001), Comment on “Inference for semiparametric models: some questions and an answer”, *Statistical Sinica*, 11, 920-936.
- Robins, J.M., A. Rotnitzky and M.J. van der Laan (2000), Comment on the Murphy and Van der Vaart article “On profile likelihood”, *Journal of the American Statistical Association*, 95, 431-435.
- Rosenbaum, P.R. and D.B. Rubin (1983), The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70 (1), 41-55.
- Rubin, D.B. (1973a), Matching to remove bias in observational studies, *Biometrics*, 29, 159-183.
- Rubin, D.B. (1973b), The use of matched sampling and regression adjustments to remove bias in observational studies, *Biometrics*, 29, 185-203.
- Rubin, D.B. (1974), Estimating causal effects of treatments in randomized and non-randomized studies, *Journal of Educational Psychology*, 66 (5), 688-701.
- Rubin, D.B. (1977), Assignment to treatment group on the basis of a covariate, *Journal of Educational Statistics*, 2 (1), 1-26.
- Rubin, D.B. (1978), Bayesian inference for causal effects: the role of randomization, *Annals of Statistics*, 6 (1), 34-58.
- Rubin, D.B. (1980), Comment on ‘Randomization analysis of experimental data in the Fisher randomization test’, by Basu, *Journal of the American Statistical Association*, 75 (371), 591-593.
- Rubin, D.B. (1986), Comment: which ifs have causal answers? *Journal of the American Statistical Association*, 81 (396), 961-962.
- Rubin, D.B. (1990), Neyman (1923) and causal inference in experiments and observational studies, *Statistical Science*, 5 (4), 472-480.
- Tsiatis, A.A. (2006), *Semiparametric theory and missing data*, Berlin: Springer.
- Tsiatis, A.A. and M. Davidian (2007), Comment: Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data, *Statistical Science*, 22 (4), 569-573.
- van der Laan, M. and J.M. Robins (2003), *Unified methods for censored longitudinal data and causality*, Berlin: Springer.

International Transport Forum

2 rue André Pascal

75775 Paris Cedex 16

itf.contact@oecd.org

www.internationaltransportforum.org
