JOINT TRANSPORT RESEARCH CENTRE

Discussion Paper No. 2009-32
July 2010 (Revised version)

Estimating the agglomeration benefits of transport investments: Some tests for stability

Daniel J. GRAHAM
Imperial College, London
Kurt VAN DENDER
Joint Transport Research Centre of the OECD and the International Transport Forum

Estimating the agglomeration benefits of transport investments: some tests for stability

Daniel J. Graham[†] and Kurt Van Dender[§]

Imperial College London

Joint Transport Research Centre of the OECD and the International Transport Forum

Abstract

The case for including agglomeration benefits within transport appraisal rests on an assumed causality between access to economic mass and productivity. Such causality is justified by the theory of agglomeration, but is difficult to establish empirically because estimates may be subject to sources of bias from endogeneity and confounding. The paper shows that conventional panel methods used to address these problems are unreliable due to the highly persistent nature of accessibility measures. Adopting an alternative approach, by applying semiparametric techniques to restricted sub-samples of the data, we find considerable nonlinearity in the relationship between accessibility and productivity with no positive effect to be discerned over broad ranges of the data. A key conclusion is that we are unable to distinguish the role of accessibility from other potential explanations for productivity increases. For transport appraisal, this implies that the use of conventional point elasticity estimates could be highly misleading.

Keywords: Agglomeration, transport, causality, heterogeneity, confounding.

JEL classification: R42, R12.

[†]Author for correspondence: Centre for Transport Studies, Imperial College London, London, SW7 2AZ, UK, Tel: +44 20 7594 6088, Fax: +44 20 7594 6107, Email: d.j.graham@imperial.ac.uk

[§]Joint Transport Research Centre of the OECD and the International Transport Forum, Tel: +33 (0)1 45 24 95 98, Email: kurt.vandender@oecd.org

1 Introduction

Recent developments in the appraisal of transport infrastructure projects point to agglomeration economies as a potential source of wider economic impacts (WEIs) of transport investments (e.g. DfT 2005, Eddington 2006, Venables 2007, Graham 2007a). The key justification for this line of reasoning is that more or better transport infrastructure can boost agglomeration economies by reducing the generalized price of travel and thereby lowering interaction costs within the spatial economy. With such lower costs, agglomeration of economic activity increases, and this creates benefits through economies of scale. Such benefits are in theory additional to those captured in a standard CBA because they are sourced from increasing returns that are external to the firm and thus would not feature in the willingness-to-pay approach that underpins calculations of consumers surplus.

To obtain a measure of agglomeration economies, econometric analyses typically focus on estimating an elasticity of productivity with respect to firms' access to economic mass (cf. Rosenthal and Strange 2004, Melo et al. 2009, for reviews), where access is co-determined by transport facilities. For transport appraisal, the key challenge is to ensure that such estimates are appropriately adjusted for sources of confounding and endogeneity such that they capture the effect on productivity from changing accessibility alone. This is, however, difficult to achieve in practice and empirical tests of the theory of agglomeration still suffer from significant weaknesses. Whether the divergence between theory and the empirical representation of theory matters for applications in project appraisal, and how significant the problem may be, is the subject of this paper.

The paper demonstrates that conventional panel data methods to correct for endogeneity and confounding come up against serious difficulties when estimating the total factor productivity (TFP) effects of accessibility[1]. Measures of accessibility are highly persistent with the lack of temporal variation producing unreliable results from fixed effects and dynamic panel models. In principle, random effects models are also problematic since we suspect omitted variables which should be correlated with agglomeration. Adopting an alternative approach, the paper uses semiparametric techniques to look in greater detail at the nature of the accessibility-productivity relationship within restricted sub samples of the data. The

[1] Much of the empirical work on agglomeration uses wage based models. For the purpose of transport appraisal it is considered important to have an estimate of TFP effects rather than labour productivity alone since we expect changes in accessibility to affect the productivity of labour and capital (e.g. Graham 2007a Mare and Graham 2010).

results show considerable nonlinearity and also indicate that over broad ranges of accessibility there is often no positive effect to be discerned. For transport appraisal, this implies that the use of conventional point elasticity estimates could be highly misleading.

The paper is structured as follows. Section 2 provides a brief review of the case for including agglomeration benefits within transport appraisal and illustrates how agglomeration elasticities can be used to make the relevant calculations. It also explains the key challenges faced in obtaining a valid estimate of agglomeration economies. Section 3 provides a description of the data available for estimation and outlines our empirical approaches. Results are presented in section 4. Section 5 interprets the results and discusses implications for the assessment of agglomeration benefits within transport appraisal. Conclusions are drawn in the final section

2 Transport induced agglomeration effects: from theory to estimation

Agglomeration refers to the scale of 'economic mass' accessible to firms[2]. More agglomeration is a good thing in so far as an increase in scale generates positive externalities by reducing interaction costs in the spatial economy. For instance, agglomeration externalities are thought to arise from improved opportunities for labour market pooling, increased scope for industry specialization, greater efficiency in knowledge or technology sharing, and improved opportunities for input-output association[3]. Theory tells us that agglomeration economies will be manifest in tangible benefits such as lower average costs for firms and higher productivity. Thus, for some firm producing output y using a vector of inputs X and experiencing level of agglomeration A, we can define a general production function

$$y = f(X, A) \tag{1}$$

in which theory predicts that $\partial \log y / \partial \log A = \eta_{y,A}$ will be positive.

Venables (2007) develops a theoretical model that links agglomeration and transport provi-

[2] A distinction is sometimes made between intra-industry and inter-industry agglomeration, referred to respectively as localization and urbanization. It can be argued that the basic mechanisms underpinning the advantages derived from agglomeration are common to each (Duranton and Puga 2004). Since changes to the transport system affect accessibility in general, and not in some industry specific way, it is sufficient here to consider the concept of agglomeration in general as being derived from either broad class of externality.

[3] It is possible that the level of agglomeration can exceed some 'optimal' amount. See for instance Graham 2007b.

sion. The argument, as outlined in the introduction is straightforward: transport investment increases the access that firms have to economic mass, which, if agglomeration economies exist, induces a source of increasing returns that is not captured in a standard transport appraisal. Venables goes on to show that we can attain an estimate of the 'agglomeration benefits' of transport investment if we know: a) the change in access to economic mass that will result from making some transport intervention; and, b) the amount by which productivity will rise in response to an increase in agglomeration (i.e. $\eta_{y,A}$).

The UK Department for Transport (DfT) have requested that agglomeration benefits be assessed as an additional component of transport scheme appraisal (e.g. DfT 2005). If we can estimate the level of accessibility after some transport intervention of size $b-a$ has been made, then the associated productivity change can in general be calculated using

$$\Delta y = y_b - y_a = \left[\left(\frac{A_b}{A_a} \right)^{\eta_{y,A}} - 1 \right] y_a. \tag{2}$$

with the unknown y_b given by the compound growth expression

$$y_b = y_a e^{\eta_{y,A}[\log A_b - \log A_a]}. \tag{3}$$

Given a point elasticity, this expression provides a reasonably consistent calculation for large changes in accessibility as for small. To evaluate (3) we therefore need an estimate of A_b and of the elasticity $\eta_{y,A}$. Methods for estimation of A_b are well developed since this value is required for conventional calculations of the value of travel time savings. The elasticity of productivity with respect to agglomeration is a relatively new parameter in the field of transport appraisal and its estimation forms the subject of this paper.

Venables (2007) sets out the conceptual basis for the assessment of agglomeration benefits providing a theoretical rather than empirical case. A key challenge lies in finding empirical estimates that can adequately represent the relationships defined in Venables' model. Several empirical studies conducted over the last 40 years have explored the relationship between agglomeration and productivity. Most of these have been concerned with the effects of agglomeration on manufacturing industries. Earlier studies used aggregate measures of city and industry size to represent urban and industrial agglomeration, but the preferred approach in more recent work is to proxy urban agglomeration with 'effective density' or 'market potential' measures calculated for small zones. For instance, for some zone i containing economic

activity E_i the market potential measure is

$$A_i = \frac{E_i}{d_i} + \sum_j \frac{E_j}{d_{ij}}, \tag{4}$$

where d_i is some measure of the internal distance of zone i and d_{ij} is the distance to zones j. It is therefore a distance discounted sum of economic activity from each zone and is directly relevant in the context of transport investment since it is an accessibility based measure[4]. The market potential measure is based on an assumed power decay of accessibility with distance where the exponent on is equal to 1. Recent papers by Graham et al. (2010) and Melo and Graham (2009) show that the value of the exponent may be different from one but tends to have little impact on the magnitude of aggregate estimates of agglomeration.

What do existing studies tells us about the effects of transport-induced agglomeration? The consensus is that urban scale has a positive and significant effect on productivity with agglomeration elasticities for manufacturing industries typically found to be somewhere between 0.02 and 0.10 (for reviews see Eberts and McMillen 1999, Rosenthal and Strange 2004, Melo et al. 2009). But these are estimates of the *general* effect of agglomeration as represented by city size, or more recently, by some measure of access to economic mass, and as such they do not provide direct evidence of transport effects since there is no explicit consideration of transport changes. The current approach for the assessment of WEBs in the UK is to simply assume that productivity changes induced through transport investment are more-or-less synonymous with those estimated using market potential type measures of agglomeration.

However, there are three key reasons to be caution on the direct application of market potential elasticities in the appraisal context. First, is that the effect of agglomeration on productivity is endogenous (Melo et al. 2009, Graham et al. 2010, Combes et al. 2010), which is clearly of first-order importance for the evaluation of transport investments since it seeks to capture only a unidirectional effect. Second, the relationship between accessibility and TFP is likely confounded. A commonly cited confounder is unobserved functional heterogeneity, the argument being that the distribution of this variable is skewed such that higher-productivity jobs tend to be found disproportionally in the most urbanised locations (e.g. Duranton and Puga 2005, Combes et al. 2008b). Confounding may also arise from output price distortions

[4]Note that it is possible to incorporate some explicit measure of transport accessibility in (4), such as travel time or generalized cost, but this poses severe problems of endogeneity in estimation because journey times tend to be simultaneously determined with productivity, due for instance, to the effect of congestion (e.g. Rice et al. 2006, Graham 2007b). Furthermore, in relation to appraisal, the DfT make the argument that since benefits to business and freight users from reduced congestion are already included in willingness to pay calculations distance based measures of accessibility are more appropriate than those based on travel times (DfT 2007).

because higher rents and other costs associated with more intense competition for resources may be reflected in output prices. In other words, some of the productivity-density gradient might simply be caused by higher prices in larger cities. Confounding is important because the case for assessing the agglomeration benefits of transport rests on an assumed causality between productivity and access to economic mass, but not necessarily with the unobserved confounders. Finally, there is an issue of nonlinearity in estimating agglomeration benefits from transport investments. When we consider the effects of transport-induced agglomeration we will typically be looking at interventions of different size, many of which will result in only relatively minor shifts in access to economic mass. So it is important to know whether small changes in accessibility have the same proportional effect on productivity as large changes do, or in other words, whether the elasticities are approximately constant over the sample.

The first two issues, endogeneity and confounding, have received some attention in the recent empirical literature (e.g.Ciccone and Hall 1996; Ciccone 2002,Rosenthal and Strange 2004,Combes et al. 2008a, Duranton and Puga 2004, and Rice et al. 2006), but have proven difficult to deal with in a satisfactory way. This is largely due to problems in finding relevant and exogenous instruments for agglomeration, but also to an absence of measures that can adequately represent any functional heterogeneity that is distributed systematically with levels of agglomeration. The third issue, concerning nonlinearity, has to our knowledge received no prior attention in empirical work other than in attempts to estimate diminishing returns to agglomeration through a quadratic specification of the agglomeration variable (e.g. Graham 2007b), which is, however, a quite separate issue.

Our objective in this paper is to test the stability of estimates of accessibility effects on TFP to treatment for endogeneity and confounding and to consider in some detail the issue of nonlinearity. The underlying question is whether these problems provide a significant barrier in estimating the agglomeration effects that may arise from transport investments. It is important to stress at the outset that we are interested in identifying the effect of *access to economic mass* on productivity since that is the attribute that we expect transport investments to change. The generation of agglomeration economies may involve a broader range of processes than changes in accessibility alone, and indeed some of the variables we class as confounders could be regarded as integral to the nature of agglomeration itself. To be clear, we do not regard accessibility as synonymous with agglomeration, but we focus on this specific aspect of the problem because that is where the links with transport investment are to

be found. Thus, throughout the remainder of the paper we refer to elasticities of TFP with respect to accessibility.

3 Data and models for estimation

Contemporary approaches to the estimation problems outlined above rely heavily on the advantages offered by conventional panel data methods. The fixed effects model provides a means of obtaining unbiased estimates in the presence of time-invariant confounding, dynamic panel GMM approaches can be used to address both time-invariant confounding and endogeneity, while the random effects model is reasonable to adopt when there is unobserved unit level heterogeneity that is uncorrelated with the covariates. In the context of agglomeration, the potential for dynamic effects also strongly support use of a panel approach.

The production data available for estimation comprise information on output, labour and capital inputs for a panel of UK firms in 2 digit sectors over the period 1995 to 2004, and are describe in full in Graham (2007b). To represent accessibility we calculate market potential measures (see equation 4) for small zones, Post Code Sectors (PCSs)[5], in which the firms are located using employment as our measure of economic activity. In constructing these measures the distance between zones, d_{ij}, is calculated using the x and y coordinates of zone centroids and the internal distances, d_i, are approximated by taking the average of the distances between each pair of full postcodes that are contained within that PCS.

The data allow us to estimate our models separately for different industry groups. Table 1 below shows a breakdown of the sample of firms by industry, listing the number of firms and the total number of observations. To allow for a concise presentation of results we estimate using samples aggregated into five industry groups: manufacturing (SIC 14-40), construction (SIC 45), wholesale & retail (SIC 50-52), transport & communications (SIC 60-64), and business services (SIC 65-75). Within each industry group we allow for unobserved heterogeneity associated with distinct industrial activities by including a set of dummy variables corresponding to 2 digit industries.

To model the effect of accessibility on TFP we specify a dynamic production function for the

[5] There are 11,344 postcode sectors defined in our data for which there are extensive detailed employment data that allow us to construct measures of access to economic mass.

Table 1: **Data description: firms and no. of observations by industry.**

	firms	obs
manufacturing (SIC 15-40)	4,661	35,686
construction (SIC 45)	1,472	10,442
wholesale & retail (SIC 50-55)	3,545	25,956
transport & communications (SIC 60-64)	1,081	7,834
business services (SIC 65-75)	6,909	48,353
Total	17,668	128,2713

ith firm ($i = 1, ..., N$) producing output Y at time t ($t = 1, ..., T$),

$$\log Y_{it} = \beta_L \log L_{it} + \beta_K \log K_{it} + \beta_A \log A_{it} + \zeta_t + f_i + \varepsilon_{it}. \tag{5}$$

The firm uses labour (L) and capital (K) inputs and is located in an environment with a level of access to economic mass measured by A. The term ζ_t is a time specific effect that allows for unobserved shocks which are common across firms and f_i represents unobserved individual time-invariant heterogeneity. We introduce dynamics by specifying a potentially autoregressive productivity shock $\varepsilon_{it} = \rho\varepsilon_{it-1} + \nu_{it}$, with $|\rho| < 1$ and $\nu_{it} \sim IID\left(0, \sigma^2\right)$ representing serially uncorrelated white noise error.

To address the problems outlined in section 2 we estimate equation (5) using a number of standard panel approaches. First, assuming no sources of confounding or endogeneity we apply a feasible GLS estimator which allows for serial autocorrelation and random individual heterogeneity. Estimates obtained using this approach are based on the most restrictive assumptions and provide a sort of base case against which we can compare our other results. Second, we adjust for sources of time-invariant confounding using a fixed effects estimator but again assume unidirectional causality. Third, following Blundell and Bond (2000), we allow for both unobserved confounding and potential endogeneity, of accessibility and of the production function itself (i.e. endogenous inputs), and specify (5) in the form of an ADL(1,1) dynamic model

$$\begin{aligned} \log Y_{it} = \rho \log Y_{it-1} + \beta_L \log L_{it} - \rho\beta_L \log L_{it-1} + \beta_K \log K_{it} - \rho\beta_K \log K_{it-1} \\ +\beta_A \log A_{it} - \rho\beta_A \log A_{it-1} + (\zeta_t - \rho\zeta_{t-1}) + f_i (1 - \rho) + \varepsilon_{it}. \end{aligned} \tag{6}$$

Estimation of (9) is by means of dynamic panel Generalized Method of Moments (GMM) in which the time series nature of the data is used to derive a set of instruments which are assumed

correlated with the covariates but orthogonal to the errors, allowing us to then define and solve a set of moment conditions which will be satisfied at the true value of the parameters to be estimated (e.g. Arellano and Bond 1991; Arellano and Bover 1995; Blundell and Bond 1998). When the data available for estimation are highly persistent, as is characteristic of our data, the so called *system* GMM estimator, compromising both first-differenced and levels equations has been shown to offer much increased efficiency and less finite sample bias compared to the difference GMM estimator alone (e.g. Arellano and Bover 1995; Blundell and Bond 1998; Blundell and Bond 2000; Blundell et al. 2000)[6].

In addition to these standard panel techniques we also adopt an alternative approach which uses semiparametric estimators with sample restrictions. To investigate the potential for nonlinearities in the relationship between accessibility and TFP we estimate by fitting a semi-parametric linear additive mixed model (SPLAM) (see Rupert et al. 2003). The SPLAM model has the general form

$$Y_{it} = X_{it}^{'}\beta + \sum_{k=1}^{K} f_k(X_{it}) + \epsilon_{it}, \tag{7}$$

where $X_{it}^{'}\beta$ is the parametric component of the model and where the f_k, $k = 1, ..., K$, are basis functions of the covariates modelled semiparametrically. The model allows for various specifications of residual error covariance.

For our production function analysis we model the relationship between output and factor inputs parametrically (but flexibly) using a second-order response surface (i.e. a 'translog' specification), but introduce the accessibility variable non-parametrically. We allow for firm level random effects and an AR(1) error for serial correlation. Thus, the estimating equation is

$$\begin{aligned}\log Y_{it} =& \beta_L \log L_{it} + \beta_K \log K_{it} + \frac{1}{2}\beta_{LL}(\log L_{it})^2 + \frac{1}{2}\beta_{KK}(\log K_{it})^2 + \\ & \beta_{LK} \log L_{it} \cdot \log K_{it} + \sum_{k=1}^{K} f_k(\log A_{it}) + \zeta_t + f_i + \varepsilon_{it}\end{aligned} \tag{8}$$

where ζ_t is a time specific effect, f_i is a random firm level error component, and ε_{it} is an autoregressive productivity shock such that $\varepsilon_{it} = \rho\varepsilon_{it-1} + \nu_{it}$ with $\nu_{it} \sim IID(0, \sigma^2)$. Two digit dummies are also included. The models are estimated using the `gamm` package in `R` using Restricted Maximum Likelihood (REML).

[6] A good summary of this literature is given in Baltagi (2005). For a full discussion of GMM in the context of dynamic panel models see Arellano and Bover (1995), Arellano and Honore (2001), Blundell and Bond (1998), Blundell and Bond (2000), Blundell et al. (2000).

The SPLAM model is useful for identifying nonlinearities in the accessibility-TFP relationship but it does not implicitly correct for endogeneity and confounding. To address confounding we adopt an approach similar to that used in the causal inference (or programme evaluation) literature which applies semiparametric estimation but with sample restrictions imposed to introduce balance in confounding covariates (for a review of this literature see Imbens and Wooldridge 2009). A key principle in the estimation of 'treatment effects' is that identification of the true effect is made more credible in samples containing units that can be assumed broadly homogeneous in confounding covariates. The standard way of achieving covariate balance is through propensity score methods (e.g. Rosenbaum and Rubin 1983, Joffe and Rosenbaum 1999, Rosenbaum 1999 and Rubin 2006). However, in order to estimate the propensity score, the confounders need to be observable. This is typically not possible in the case of agglomeration, which is why a combination of fixed effects and IV is commonly used (see Combes et al. 2010 for a summary of these problems). In the absence of observed confounders we use sample restrictions according to an urban 'area type' classification provided by the UK Department of Transport (see table 2).

Table 2: Classification of area types in Britain.

		area type
1	National centre	Central London
2		Inner London
3		Outer London
4	Regional centres	Inner Conurbation
5		Outer Conurbation
6	Sub-Regional centres	Urban Big (pop > 250,000)
7		Urban Large (pop > 100,000)
8		Urban Medium (pop > 25,000)

The DfT classification allocates small zones of the country to each of the categories shown in table 2. The regional centres, or conurbations, offer an interesting case study for estimation because we would expect them to be broadly homogeneous in terms of confounding covariates such as functional heterogeneity and output prices, but they also contain a considerable range of variance in accessibility. Other interesting categories include urban big and urban large, both of which may contain a more diverse range of areas than the conurbations, but again will retain sufficient variance in accessibility. Comparing results for these three sub-samples,

along with those for the full sample, will allow us to test whether a positive accessibility effect on TFP is to be universally found and to explore any nonlinearities in greater detail.

The logic of this approach, therefore, is that since we believe the confounding factors vary in some way with the level of urbanisation then by estimating the accessibility effect within sub-samples we can eliminate extrema in the data and thus hopefully reduce the influence of confounding factors on productivity. Clearly, we do not know, and indeed do not even assume, that confounding will be eliminated within our sub-samples. However, we can still test a simple and fundamental proposition: that if accessibility does induce higher TFP ceteris paribus then we should identify positive effects within any stratum of the data that contains sufficient variance in accessibility.

4 Results

4.1 Panel models

Table 3 below shows production function estimates obtained using a feasible GLS random effects estimator with AR(1) errors (FGLS-RE), system GMM (sys-GMM), and feasible GLS but with firm level fixed effects (FGLS-FE). For some industries FGLS-FE model failed and in these case, we replace the fixed effects estimator with one based on first differences (FGLS-FD) which, like the FE estimator, is also consistent under the assumption of unobserved correlated individual effects and uses within firm variation for parameter estimation. We also found that the ADL(1,1) GMM specification given in (6) suffered from multicollinearity and so we instead opted for the ADL(1,0) partial adjustment model

$$\log Y_{it} = \rho \log Y_{it-1} + \beta_L \log L_{it} + \beta_K \log K_{it} + \beta_A \log A_{it} + \zeta_t + f_i + \varepsilon_{it}, \tag{9}$$

which is a more parsimonious specification that still allows us to distinguish short from long run effects. Table 3 shows results for all sectors of the economy pooled and for five industry groups; manufacturing, construction, wholesale & retail, transport & communications, and business services. All models are estimated with a set of dummy variables at the 2 digit industry level.

Table 3: Production function estimates

	all industries			manufacturing			construction		
	FGLS-RE	sys-GMM	FGLS-FE	FGLS-RE	sys-GMM	FGLS-FD	FGLS-RE	sys-GMM	FGLS-FE
$\log Y_{t-1}$	-	0.670**		-	0.301***	-	-	0.451**	-
	-	0.03		-	(0.029)	-	-	(0.058)	-
$\log L_t$	0.700**	0.269**	0.693**	0.674**	0.470**	0.717**	0.649**	0.468**	0.703**
	(0.003)	0.022	(0.004)	(0.005)	(0.035)	(0.006)	(0.010)	(0.060)	(0.014)
$\log K_t$	0.360**	0.131**	0.285**	0.361**	0.304**	0.285**	0.460**	0.213**	0.401**
	(0.002)	0.018	(0.003)	(0.004)	(0.030)	(0.006)	(0.009)	(0.052)	(0.015)
$\log A_t$	0.105**	0.045**	0.058	0.077**	0.061**	-0.132	0.095**	0.134**	-0.16266
	(0.008)	0.009	(0.042)	(0.013)	(0.024)	(0.082)	(0.034)	(0.051)	(0.200)
LR η_L	0.700	0.815	0.693	0.674	0.672	0.717	0.649	0.852	0.703
LR η_K	0.361	0.397	0.285	0.361	0.435	0.285	0.46	0.388	0.401
LR η_A	0.105	0.136	-	0.077	0.087	-	0.095	0.244	-
R^2	0.870	-	0.797	0.89	-	0.43	0.87	-	0.83
Baltagi-Wu LBI	1.650	-	1.644	1.59	-	2.44	1.811	-	1.811
AR(1)	-	0.000	-	-	0.000	-	-	0.000	-
AR(2)	-	0.069	-	-	0.475	-	-	0.022	-
Hansen	-	0.000	-	-	0.000	-	-	0.000	-
N	133,461	47,072	115,073	35,686	30,525	30,525	10,442	8,752	8,970

Notes: Numbers in parentheses are standard errors; FGLS-RE - random effects with AR(1) errors, sys-GMM - system GMM, FGLS-FE - fixed effects with AR(1) errors, FGLS-FD - first differences with AR(1) errors;** - significant at 1%, * - significant at 5%; LR η_L, LR η_K and LR η_A are the long run elasticities of labour, capital and accessibility; LBI is the Baltagi-Wu test for serial autocorrelation; AR(1) and AR(2) are the Arrelano and Bond tests for first-order and second-order serial autocorrelation.

Table 3: (continued):Production function estimates

	wholesale & retail			transport & communications			business services		
	FGLS-RE	sys-GMM	FGLS-FD	FGLS-RE	sys-GMM	FGLS-FE	FGLS-RE	sys-GMM	FGLS-FD
$\log Y_{t-1}$	-	0.498**	-	-	0.527**	-	-	0.376**	-
	-	(0.035)	-	-	(0.052)	-	-	(0.019)	-
$\log L_t$	0.710**	0.293**	0.736**	0.725**	0.278**	0.740**	0.713**	0.447**	0.681**
	(0.006)	(0.037)	(0.008)	(0.012)	(0.061)	(0.048)	(0.005)	(0.023)	(0.008)
$\log K_t$	0.364**	0.403**	0.260**	0.286**	0.246**	0.266**	0.361**	0.275**	0.318**
	(0.005)	(0.038)	(0.006)	(0.010)	(0.048)	(0.032)	(0.004)	(0.019)	(0.005)
$\log A_t$	0.064**	0.009	0.025	0.116**	0.086*	0.345	0.127**	0.095**	0.04
	(0.016)	(0.023)	(0.101)	(0.035)	(0.038)	(0.380)	(0.013)	(0.016)	(0.104)
LR η_L	0.710	0.584	0.736	0.725	0.588	0.740	0.713	0.716	0.681
LR η_K	0.364	0.803	0.26	0.286	0.520	0.266	0.361	0.441	0.318
LR η_A	0.064	-	-	0.116	0.182	-	0.127	0.152	-
R^2	0.87	-	0.4145	0.83	-	0.36	0.84	-	0.29
Baltagi-Wu LBI	1.578	-	2.426	1.499	-	-	1.675	-	2.526
AR(1)	-	0.000	-	-	0.000	-	-	0.000	-
AR(2)	-	0.167	-	-	0.880	-	-	0.121	-
Hansen	-	0.000	-	-	0.000	-	-	0.000	-
N	25,956	21,891	21,891	7,834	6,627	6,627	48,353	40,323	40,323

Notes: Numbers in parentheses are standard errors; FGLS-RE - random effects with AR(1) errors, sys-GMM - system GMM, FGLS-FE - fixed effects with AR(1) errors, FGLS-FD - first differences with AR(1) errors; ** - significant at 1%, * - significant at 5%; LR η_L, LR η_K and LR η_A are the long run elasticities of labour, capital and accessibility; LBI is the Baltagi-Wu test for serial autocorrelation; AR(1) and AR(2) are the Arrelano and Bond tests for first-order and second-order serial autocorrelation.

We focus first on results from the FGLS-RE models, which provide the base against which we compare models that incorporate some explicit treatment for endogeneity and confounding, or that make use of different sampling variance in accessibility. The R^2 values for the FGLS-RE models indicate reasonably high degrees of explanatory power with all values falling in the range 0.8 to 0.9. The Baltagi-Wu locally best invariant (LBI) statistic for serial autocorrelation, which is a suitable diagnostic for unbalanced panels (Baltagi and Wu 1999), rejects the null hypothesis of no first-order serial autocorrelation for all models. FGLS-RE estimates relating to the output elasticities and returns to scale (RTS) are broadly similar across industries and indicate labour share in the range of two-thirds to three-quarters, and constant or slightly increasing RTS.

For the economy as a whole and for all five industry groups we estimate positive and significant accessibility economies using the FGLS-RE estimator as follows: all industries (0.105), manufacturing (0.077), construction (0.095), wholesale & retail (0.064), transport & communications (0.116), and business services (0.127). Thus, we find evidence of substantial accessibility economies consistent with the orders of magnitude typically found in the agglomeration literature with the largest effects for business services (see for example Melo et al. 2009).

The rationale for the use of the dynamic panel GMM specifications is to provide instrumentation for endogenous regressors and adjust for confounding. The key diagnostic statistics are the tests for first and second order serial autocorrelation and the Hansen test for overidentifying restrictions. In all cases the GMM models shown in table pass the Arellano-Bond tests, AR(1) and AR(2), for autocorrelation in the errors of the levels equations. None of the models, however, pass the Hansen test of overidentifying restrictions. The failure of the models to pass the Hansen test indicates that the instrument matrix may not be not truly exogenous. The results on accessibility do however appear plausible and are similar in the long run, though somewhat larger, to the FGLS-RE estimates for manufacturing (0.087), transport & communications (0.182), and business services (0.152).

This could be taken as evidence that the influence of endogeneity is small for these sectors, but equally, it could be indicative of weak instruments which tend to give estimates that are biased in the same direction as least squares[7]. There are in fact other problems which

[7] Bound et al. (1995) provide some good examples of the problems associated with the use of inappropriate instruments.

question the validity of dynamic panel GMM results. For instance, we find that estimates of the autoregressive parameters are smaller than those obtained using pooled OLS. This suggests biased estimates which can result when the instruments are only weakly correlated with the endogenous regressors, or when the instruments themselves are not orthogonal. Furthermore, we also find that the estimates are highly sensitive to any changes in the lag structure used for instrumentation. A key problem with dynamic panel GMM specifications in our context is that the data on production and accessibility are so highly persistent, in fact nearing unit root in AR(1) specifications as shown in table 4 below, that the problem of weak instrumentation can become extreme.

For this reason, and given the failure of the models to pass the Hansen test, it is not possible to draw any substantive conclusions on the role of endogeneity from the GMM models. An alternative approach could be to address the endogeneity by constructing exogenous instruments. In the existing literature instruments commonly proposed include long lags on population density (e.g. Ciccone and Hall 1996; Mion 2004; Mion and Naticchioni 2005; Hanson 2005; Rice et al. 2006; Combes et al. 2007) or even geological features (e.g. Rosenthal and Strange 2005 and Combes et al. 2008a). Generally, the role of reverse causality is found to be small. For our data it would be hard to defend such instruments as either relevant or exogenous, and in fact the literature gives little convincing guidance on their validity[8]. Given the difficulties faced in constructing valid instruments, and the failure of dynamic panel models to offer a plausible alternative, our belief is that it will not be easy to identify the true role of reverse causality in future empirical work.

We next turn to results from the FGLS-FE and FGLS-FD models. These are the models which provide consistent estimation under the assumptions of unobserved firm level heterogeneity which is correlated with the regressors (confounding), and which also draw on within group rather than between group variance. The first point of interest from these results is that estimates relating to the elasticities of labour and capital, and therefore to RTS, are very similar to those obtained using the FGLS-RE specification. Labour shares range from 0.68 to 0.74 with estimates of RTS all very close to 1.0, though in general slightly less than the FGLS-RE estimates. The second and key point of interest, however, is that we find no evidence of accessibility economies for any of the five industries listed in the table. Thus, conditional

[8] A key problem here is that the commonly used diagnostic test for instrument exogeneity, the Hansen test, has poor finite sample properties (see for example Andersen and Sorensen 1996 and Bowsher 2002). To quote Hahn and Hausman (2003), even using the standard tests for instrument validity "the researcher may estimate 'bad results' and not be aware of the outcome" (p 118).

Table 4: AR(1) specification for accessibility, OLS and sys-GMM estimates.

	OLS	sys-GMM
manufacturing	1.002	0.997
construction	1.005	0.996
wholesale & retail	0.999	0.985
transport & communications	1.002	0.996
business services	1.000	0.998
all industries	1.000	0.995

on individual firm effects, there is no evidence that changes in accessibility over time have affected productivity for firms. Accessibility economies effectively disappear.

Two possible factors that might explain this stark contrast between the FGLS-RE and FGLS-FE(FD) models are the treatment, or lack thereof, of unobserved heterogeneity (i.e. confounders); or the differences in sample variance between model specifications (i.e. between and within variance versus only within variance). It is worth noting that the mean annual change in the accessibility variable is only 1.25% (with a standard deviation of 2.3%) so the range of sample variance over time is actually very small. And this demonstrates the key problem with conventional panel methods: the highly persistent nature of accessibility implies a trade-off between correction for unobserved heterogeneity and the retention of sampling variance. While we want to eliminate confounding by differencing or by imposing time invariant effects, this can result in a lack of variance due to the near unit root properties of the accessibility variable. So from the statistical results it becomes difficult to know which estimates are closer to the 'true' effect rendering conventional panel approaches highly unreliable.

4.2 Semiparametric analysis

As described in section 3, to supplement the panel models we also estimate SPLAM models for the full sample of firms from all areas of the country and for restricted sub-samples of firms from three area type categories: conurbations, urban big (population $> 250{,}000$), and urban large (population $> 100{,}000$). All models are estimated by REML with a normally distributed random intercept at the firm level and an AR(1) error process. Figure 1 below shows kernel

density estimates of the distribution of (log) accessibility for all areas, conurbations, urban big and urban large.

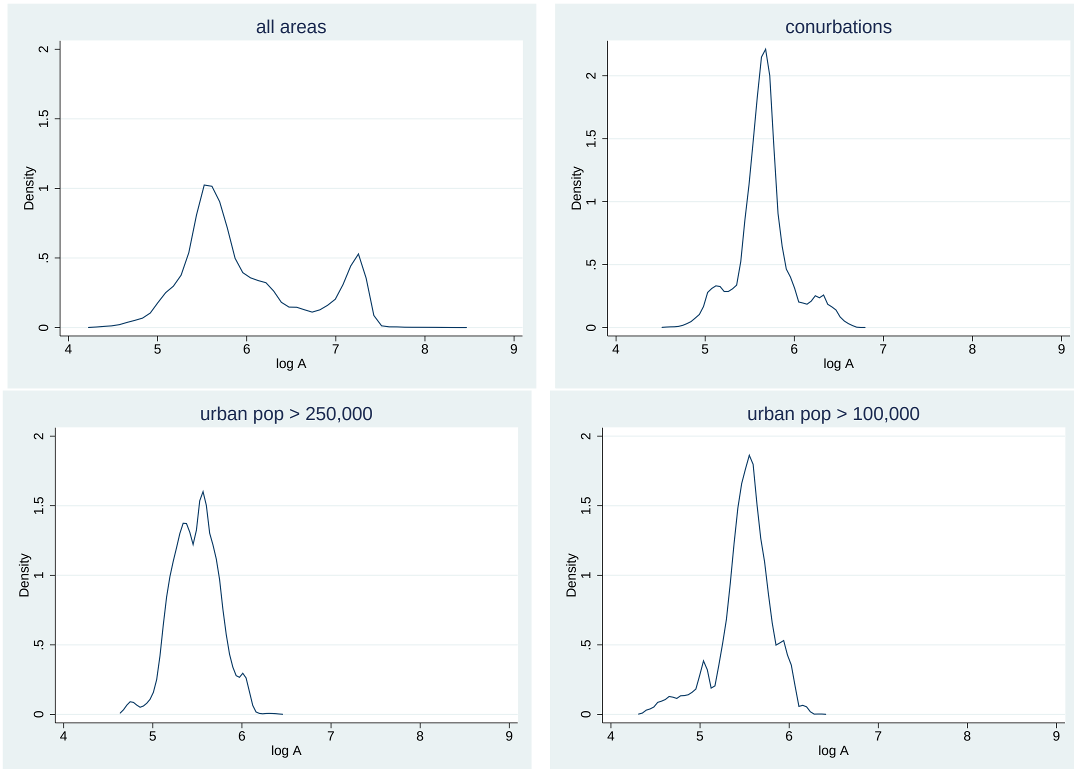

Figure 1: Kernal density estimates for $\log A$: all areas, conurbations, urban big (>250,000 pop) and urban large (>100,000 pop).

The figure shows that substantial variation in accessibility is retained within our sub-samples. Note the clear influence of London on the distribution in the upper left panel, causing a second peak in the distribution around $\log A = 7.3$. Britain's conurbations have maximum values of $\log A$ around 6.5, while the urban big and urban large samples are located in zones with ranges of $\log A$ between 5 and 6 and 4 and 6 respectively.

The basic production function estimates from the SPLAM models for factor inputs and RTS are very similar to those presented earlier and are nor repeated here. The key issue of interest from the semiparametric analysis is the fit of TFP to the accessibility variable. These results are shown graphically in figure 2. The x-axis of the charts measures $\log A$ and the y-axis shows the fit of $\log A$ to TFP, or in other words, it shows the shape of the non-parametric function $\sum_{k=1}^{K} f_k(\log A)$. The dashed lines show plots at 2 standard errors above and below the estimate of the smooth curve.

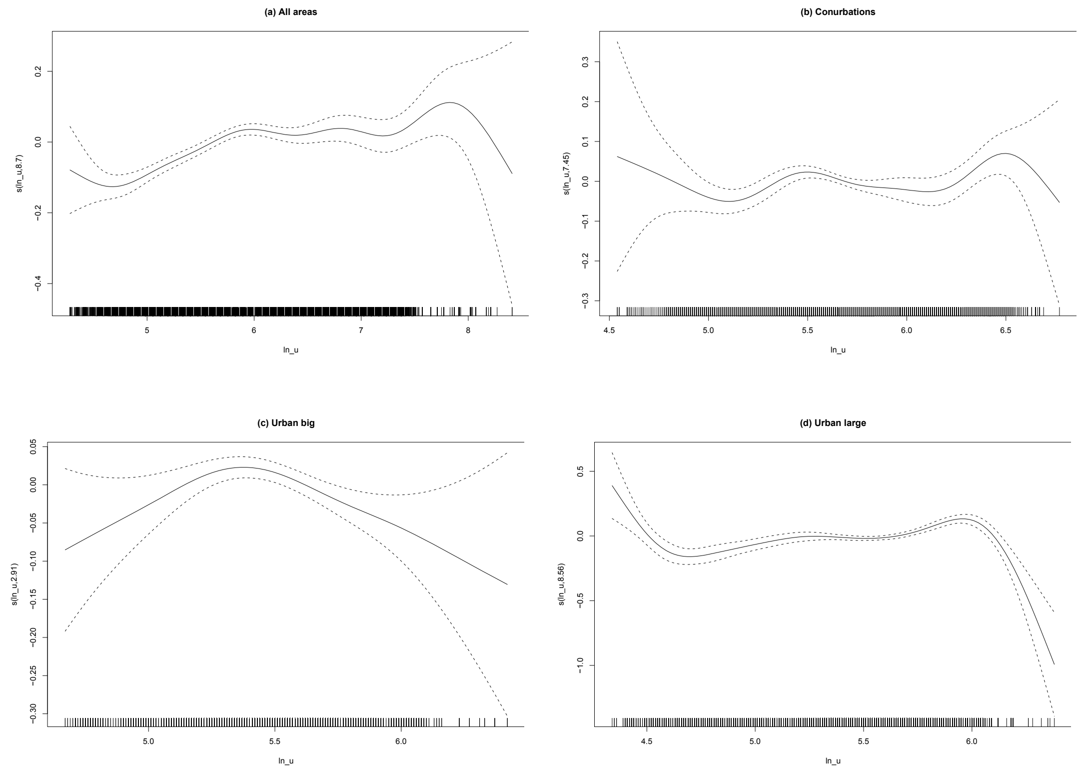

Figure 2: Semiparametric estimate of effects of $\log A$ on TFP: (a) all areas, (b) conurbations, (c) urban big (>250,000 pop), (d)urban large (>100,000 pop).

Turning first to the upper left panel, which shows the relationship between accessibility and productivity across all areas, we can see the estimated curve is increasing in places indicating a positive relationship, but it also exhibits considerable nonlinearity with no discernible positive accessibility effects over substantial ranges of $\log A$. Note also that for firms with the lowest and highest values of $\log A$ the curve indicates a negative relationship, but in fact the standard errors are large around these portions of the curve so no solid conclusion can be drawn. The effect of London firms on the productivity gradient is clearly shown by the jump in the curve as $\log A$ takes values between 7 and 8.

Panel (b) looks in more detail at the relationship between accessibility and productivity for firms located in the British conurbations. This chart shows no clear positive effect over the range of $\log A$ considered, which is in fact substantial with values ranging from 4.5 to over 6.5, over half the range included in the sample of all firms. The indication is that *within* the conurbations, greater market access is not systematically associated with increased TFP. The corresponding parametric estimate for this restricted sub-sample also shows no evidence of

accessibility economies with an insignificant estimated elasticity of 0.041 (0.038).

A similar finding of no clear positive effect from accessibility on TFP is shown in panel (c) for the urban big sub-sample. The curve appears to have an approximate quadratic shape, but in fact the standard errors are large along the length of the curve indicating essentially that there is no significant relationship to be found in the data. The corresponding parametric estimate support this view with an estimated elasticity of -0.026 (0.060). Finally, panel (d) shows the relationship for firms located in areas classed as urban large. The curve again shows considerable nonlinearity but is increasing over the majority of the range of $\log A$ and the corresponding parametric estimate support a positive relationship with an estimated elasticity of 0.114 (0.039).

Thus, from the semiparametric analysis we get some mixed evidence about the effect of accessibility on TFP. Overall there is a positive gradient as indicated by the parametric models, but the accessibility effect appears to be highly nonlinear with no discernible effect over vast ranges of the data. If accessibility does have a positive effect on TFP, the effect appears to be discontinuous and non universal.

5 Using agglomeration estimates in transport appraisal

There are two important implications of our results for the use of elasticity estimates in transport appraisal.

First, is that on the basis of the available evidence we are unable to distinguish the positive TFP effect of accessibility from other potential explanations for productivity increases. We cannot pin down the productivity response that will arise from a change in accessibility *alone* using conventional panel methods and the semiparametric analysis shows no accessibility effect within sub-samples of the data. Note that the basic production function parameters are robust to model specification but the accessibility estimates are not. One key implication of the different results presented above is that the observed accessibility effect in the full sample parametric models is capturing something other than simply access to economic mass alone.

Second, the models question whether small changes in accessibility will actually have any discernible productivity impact. Even large transport investments, tend to bring about relative

modest changes in accessibility. For instance, in assessing the benefits of a major mainline rail infrastructure project for central London which would cost around £16 billion, the UK Department for Transport (DfT) estimates changes in employment densities from the scheme of 1.8%, 5.9% and 0.6% in financial, business services and 'other sectors', respectively[9]. These are very small changes and we do not currently know much about the productivity benefits that might result from shifts of this order of magnitude. The FE and semiparametric estimates given in this paper suggest they may have little impact.

Of course these remarks must not be construed as suggesting that agglomeration economies do not exist or are weak, but should instead be understood in relation to the measures of agglomeration being used. The empirical work in this paper, and in the wider empirical literature, uses *effective density* or *market potential* measures which we cannot assume synonymous with changes in the theoretical concept of agglomeration. It is possible that other measures could be devised that are more sensitive to any behavioural differences that result from changes in agglomeration. Drawing on the well developed theory of the microfoundations of agglomeration, future empirical research should seek to find how we can best capture the mechanisms underpinning agglomeration externalities. By identifying the actual sources and their relative effects on productivity, we will obtain a better understanding of how improvements in transport accessibility might offer advantages for the performance of the spatial economy.

6 Conclusions

Current thinking on the wider economic benefits of transport investments draws on the theory of increasing returns to urban scale to argue for the existence of agglomeration benefits. These benefits are referred to as 'wider' or 'additional' because they are believed to be extraneous under the conventional consumers surplus approach which assumes constant returns and perfect markets. To assess the magnitude of agglomeration benefits we need compelling evidence that there is in fact a causal process running from improvements in accessibility to increased productivity.

In this paper, we identify three key issues to be addressed in approaching such a causal interpretation: reverse causality, confounding, and nonlinearity. We examine the relationship between access to economic mass and productivity using conventional panel estimation

[9]The full methodology and a background to Crossrail can be found in DfT 2005.

techniques that instrument for potential endogeneity and that adjust for unobserved time-invariant heterogeneity. We also adopt an alternative approach which applies semiparametric techniques to restricted sub-samples of the data. A key aim of the paper is to illustrate the analytical difficulties faced in estimating agglomeration economies and to assess whether these problems provide a significant barrier in estimating the agglomeration effects that may arise from transport investments.

The results show a high degree of sensitivity to treatment for unobserved heterogeneity, however conventional panel methods produce unreliable results due to the highly persistent nature of accessibility measures. The semiparametric analysis shows considerable nonlinearity in the relationship between accessibility and productivity with no positive effect to be discerned over broad ranges of the data. A key conclusion is that we are unable to distinguish the role of accessibility from other potential explanations for productivity increases. For transport appraisal, this implies that the use of conventional point elasticity estimates could be highly misleading.

A key issue that would improve our understanding of the relationship between access to economic mass and productivity concerns the relative roles of different sources of agglomeration. Empirical evidence on the role of sources, and their relative productivity effects, would provide a useful test of the theory and improve our understanding of the mechanisms that actually drive agglomeration economies. For transport applications, this is particularly important because the ease of making different types of trips matches well to sources. Thus, economies associated with labour markets will be affected by the efficiency of commuting trips, knowledge spillovers economies by the ease of business travel, and externalities of input-output association by provision for freight movement. By identifying the actual sources and their relative effects on productivity, we will obtain a better understanding of how improvements in transport accessibility might offer advantages for the performance of the spatial economy.

References

Andersen, T. G. and B. E. Sorensen (1996). GMM estimation of a stochastic volatility model: a monte carlo study. *Journal of Business and Economic Statistics 14*, 328–352.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error component models. *Journal of Econometrics 68*, 29–51.

Arellano, M. and B. Honore (2001). *Panel data models: some recent developments*, Chapter in J Heckman and E Leamer (eds) Handbook of Econometrics Vol. 5. Amsterdam: North Holland.

Baltagi, B. and P. Wu (1999). Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory 15*, 814–823.

Baltagi, B. H. (2005). *Econometric analysis of panel data*. Chichester: Wiley.

Blundell, R., S. Bond, and F. Windmeijer (2000). *Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator*, Chapter in B Baltagi (ed) Advances in Econometrics Vol. 15 Nonstationary panels, panel cointegration, and dynamic panels. Amsterdam: JAI Elsevier.

Blundell, R. W. and S. R. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*, 115–143.

Blundell, R. W. and S. R. Bond (2000). GMM estimation with persistent panel data: an application to production functions. *Econometric Reviews 19*, 321–340.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*, 443–450.

Bowsher, C. G. (2002). On testing overidentifying restrictions in dynamic panel data models. *Economics Letters 77*, 211–220.

Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review 46*, 213–227.

Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *American Economic Review 86*, 54–70.

Combes, P. P., G. Duranton, and L. Gobillon (2007). Spatial wage disparities: Sorting matters! *Journal of Urban Economics (in press)*.

Combes, P. P., G. Duranton, and L. Gobillon (2010). The identification of agglomeration economies. *Working Paper, University of Toronto*.

Combes, P. P., G. Duranton, L. Gobillon, and S. Roux (2008a). Estimating agglomeration economies with history, geology, and worker effects. *Working Paper, University of Toronto*.

Combes, P. P., G. Duranton, L. Gobillon, and S. Roux (2008b). Spatial wage disparities: Sorting matters! *Journal of Urban Economics 63*, 723–742.

DfT (2005). *Transport, wider economic benefits and impacts on GDP*. London: HMSO.

DfT (2007). *The additionality of Wider Economic Benefits in transport appraisal*. London: HMSO.

Duranton, G. and D. Puga (2004). *Microfoundations of urban agglomeration economies*, Chapter in Henderson JV and Thisse JF (eds) Handbook of Regional and Urban Economics, Volume 4. Amsterdam: Elsevier.

Duranton, G. and D. Puga (2005). From sectoral to functional urban specialization. *Journal of Urban Economics 57*, 343–370.

Eberts, R. W. and D. P. McMillen (1999). *Agglomeration economies and urban public infrastructure*, Chapter in HP Cheshire and E S Mills (eds) Handbook of regional and urban economics, Volume III. New York: North Holland.

Eddington, R. (2006). *The Eddington Transport study: transports role in sustaining the UKs productivity and competitiveness*. London: HM Treasury.

Graham, D. J. (2007a). Agglomeration, productivity and transport investment. *Journal of Transport Economics and Policy 41*, 1–27.

Graham, D. J. (2007b). Variable returns to agglomeration and the effect of road traffic congestion. *Journal of Urban Economics 62*, 103–120.

Graham, D. J., P. Melo, P. Jiwattanakulpaisarn, and R. Noland (2010). Testing for causality between productivity and agglomeration economies. *Journal of Regional Science*.

Hahn, J. and J. Hausman (2003). Weak instruments: diagnosis and cures in empirical economics. *American Economic Review 93*, 118–125.

Hanson, G. (2005). Market potential, increasing returns and geographic concentration. *Journal of International Economics 67*(1), 1–24.

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.

Mare, D. C. and D. J. Graham (2010). Agglomeration elasticities and firm heterogeneity. *SERC Discussion Papers* (No SERCDP0043). http://www.spatialeconomics.ac.uk/textonly/SERC/publications/download/sercdp0043.pdf.

Melo, P. and D. J. Graham (2009). Agglomeration economies and labour productivity: evidence from longitudinal worker data for gb's travel-to-work areas. *SERC Discussion Papers* (No SERCDP0031). http://www.spatialeconomics.ac.uk/textonly/SERC/publications/download/sercdp0031.pdf.

Melo, P., D. J. Graham, and R. B. Noland (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics (forthcoming)*.

Mion, G. (2004). Spatial externalities and empirical analysis: the case of Italy. *Journal of Urban Economics 56*, 97–118.

Mion, G. and P. Naticchioni (2005). Urbanization externalities, market potential and spatial sorting of skills and firms. Technical report, CEPR Discussion Papers 5172. Centre for Economic Performance, London School of Economics and Political Science.

Rice, P., A. J. Venables, and E. Patacchini (2006). Spatial determinants of productivity: analysis for the regions of great britain. *Regional Science and Urban Economics 36*, 727–752.

Rosenbaum, P. R. (1999). Propensity score. In P. Armitage and T. Colton (Eds.), *Encyclopedia of biostatistics*, pp. 3551–3555. New York: John Wiley.

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.

Rosenthal, S. and W. Strange (2004). *Evidence on the nature and sources of agglomeration economies*, Chapter in Henderson JV and Thisse JF (eds) Handbook of Regional and Urban Economics, Volume 4. Amsterdam: Elsevier.

Rosenthal, S. and W. Strange (2005). The attenuation of human capital spillovers: a manhattan skyline approach. *Working Paper*.

Rubin, D. B. (2006). Using propensity scores to help design observational studies: applications to the tobacco litigation. In D. Hantula (Ed.), *Advances in social and organizational psychology*, pp. 41–59. Mahwah, NJ: Erlbaum.

Rupert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Cambridge: Cambridge University Press.

Venables, A. J. (2007). Evaluating urban transport improvements: cost-benefit analysis in the presence of agglomeration and income taxation. *Journal of Transport Economics and Policy 41*(2), 173–188.