Productivity and employment impacts of agglomeration: evidence from transport improvements

Stephen Gibbons[ab], Teemu Lyytikäinen[ac], Henry Overman[a], Rosa Sanchis-Guarner[ad]

PRELIMINARY DRAFT – do not quote without permission

November 2010

Abstract

This paper estimates the effect of road transport infrastructure improvements on firm productivity and employment. The study uses micro longitudinal datasets on firms and employees in Britain, linked by detailed geographical location to road transport improvements that occurred between 1998 and 2003. We measure the extent to which new road infrastructure projects changed employment accessibility (or 'effective density') at locations close to the sites of the projects. We then estimate whether firms in locations that experienced large changes of this type showed productivity improvements relative to those that experienced smaller changes. We find evidence that total factor productivity improved more in places (postcode sectors) that experienced larger accessibility improvements due to transport improvements. We also find some evidence that improved accessibility increased total factor productivity within firms located in the same area before and after road improvements took place. We do not find evidence on significant effects on employment.

---

[a] London School of Economics & Spatial Economics Research Centre, London, United Kingdom.

[b] Corresponding author: Stephen Gibbons (email: S.Gibbons@Lse.ac.uk).

[c] The author wants to thank the *Yrjö Jahnsson* Foundation for financial support .

[d] The author acknowledges funding support from the Bank of Spain.

Any interpretations or opinions expressed in this report are those of the authors and do not necessarily reflect the views of the Department for Transport who contributed to the study.

1. Introduction

It has long been recognised that firms operating in large, dense and well integrated markets are more productive - that is they produce more output, from the same inputs, measured either in real or nominal terms (Marshall, 1920). Theory suggests two reasons for this, one operating at the macro level and one at the micro level. Firstly, at an aggregate macro level, the relationship could arise because of selection effects: in a world of heterogeneous firms, only the most productive firms survive in more integrated and competitive markets (Melitz 2003). Secondly, at the micro level, individual firms may benefit through 'sharing', 'matching' and 'learning' in larger economies (Duranton and Puga, 2004; Rosenthal and Strange, 2004). 'Sharing' refers to the sharing of indivisible facilities, intermediate suppliers, workers and consumers by firms, which reduces fixed costs, allows specialisation and allows firms to pool risks. 'Matching' benefits are usually discussed in terms of the benefits of having lots of workers in close proximity to employers, which means it is easier for different types of worker and different types of employer to find each other, and more productive matches occur at a faster rate. 'Learning' refers to the transfer of information, knowledge and skills.

This paper estimates the effect of road transport infrastructure improvements on firm productivity, and employment. The study uses a micro longitudinal datasets on firms and employees in Britain, linked by geographical location to road transport improvements that occurred between 1998 and 2003. We measure to what extent new road infrastructure projects changed employment accessibility (or 'effective density') at locations close to the sites of the projects. We then use statistical analyses to estimate whether firms in locations that experienced large changes of this type showed productivity improvements relative to those that experienced smaller changes.

Over the past 10-15 years, a large number of research studies have estimated the size of these 'agglomeration' effects, and found them to be positive in many different settings (see Melo et al, 2009; for a review). Usually, the estimates are based on the cross-sectional correlation between employment density and firm productivity (e.g. Graham 2006). Occasionally estimates are derived from the correlation between *changes* in employment density and *changes* in productivity (e.g. Ciccone and Hall, 1996).

It is only a small conceptual step from the observation that firms are more productive in more integrated and dense economies, to the conclusion that improvements in transport networks could improve productivity. Improvements to the road and rail networks bring firms closer to each other and firms closer to workers in terms of travel times and costs. This closer economic integration exposes firms to greater competition, improving productivity at the macro level, and improves the basis for agglomeration economies arising at the micro level.

Improvements in transportation infrastructure affect performance of firms and regions both directly and indirectly (Vickerman, 2007). On the one hand, it improves logistics and internal organisation of the firm. Transportation services enter the production function as inputs and change output and employment through labour/transportation substitution effects (Holl, 2006). On the other hand, improvements in transport infrastructure have what have been called "wider-economic benefits" (Graham, 2007).

Increased accessibility affects the equilibrium between the attraction (agglomeration) and the dispersion (increased competition and congestion) forces and therefore changes the spatial distribution of economic activities (Ottaviano, 2008). Depending on the initial level of transportation costs and on the changes induced by the transport improvements, the new equilibrium may imply higher economic concentration (in order to exploit increasing returns to scale and agglomeration economies) or higher economic dispersion (in order to benefit from lower competition and costs in the periphery and improved access to market).

Even if some authors have explicitly included the role of transportation into the spatial economics analysis (Combes and Lafoucarde, 2001; Puga, 2002; Behrens et at, 2004; Venables, 2007), there is still need to link further the transport and spatial economics literatures (Holl, 2006). Most of the empirical evidence of the effects of transport and infrastructure investment on economic outcomes has been provided at the macro-level and it has focused the impacts of investment in road and public infrastructure on several economic outcomes (Gramlich, 1994; Martin and Rogers, 1995; Boarnet, 1998; Chandra and Thompson, 2000; Gibbons and Machin, 2005; Michaels, 2008; Jiwattanakulpaisarn et al, 2010).

Only a handful of papers have studied the effect of increased accessibility on firms' outcomes, and they have mostly focused on the analysis of relocation (Coughlin and Segev, 2000; Holl, 2004a and 2004c) or firm birth (Holl, 2004b). Although using a different methodology from ours, Holtz-Eakin and Lovely (1996) find no effect of public infrastructure on aggregate productivity and a positive effect on the number and variety of plants. Our paper aims to contribute to this literature by providing new evidence on the links between increased accessibility and productivity and employment.

The central methodological approach of this paper is regression estimation of production and employment functions at the firm level and at postcode sector (about 10,000 sectors[1]) level using micro-data from the ONS Annual Respondents Database. We measure potential exposure to the

---

[1] The postal codes (postcodes) used in the United Kingdom provide very detailed geographical information. They are alphanumeric and usually correspond to a limited number of addresses or a single large delivery point. Postcode sectors codification eliminates the last two digits of the full postcode and so they gather between 15-20 postcodes on average. More information available at: http://www.ons.gov.uk/about-statistics/geography/products/geog-products-postcode/nspd/index.html

benefits from transport improvements by calculating road-network-based employment accessibility indices for firm locations, both before and after the transport improvements take place.

A source of bias with analysis of this kind is that transport schemes may be targeted at areas of high or low productivity, or increasing or declining productivity. Some authors have tried to overcome this problem by means of instrumenting the infrastructure network using "least-cost" paths or historical road maps (see Overman, 2010 for a discussion on this issue). However, as noted by Graham et al (2010), in their review of causal estimation of agglomeration effects on productivity, there is no consensus about the importance of the endogeneity bias.

In any case, in order to minimise it, we compare firms more or less exposed to the transport improvement, before and after the transport improvement occurred, using a difference-in-difference style methodology. This methodology removes bias induced by fixed productivity differences between locations that are more exposed to transport improvements and those that are less exposed. To improve on the comparability of areas in terms of differential trends in productivity, we progressively narrow the geographical scope of our samples to focus on firm locations that are close to the sites of transport improvement projects. This approach takes advantage of the fact that many of the areas that gain the most in terms of accessibility are very close to the sites of transport schemes, and benefit incidentally from policy targeted to improve the through-flow of traffic to locations further afield. Moreover, we use a instrumental variables strategy to ensure that the identification of the agglomeration effects on the firms and regions outcomes are solely based on changes in the road network.

A main contribution of our paper is to perform the analysis at the micro-level (firm) and at a very detailed geographical scale. As remarked by some authors (Moreno and Lopez-Bazo, 2007; Ottaviano, 2008), most of the impact of infrastructure investment on economic outcomes take place at the local level, so more evidence on the intra-regional and local effects using small geographies is needed (Holl, 2007). Although there exist some evidence on the effects of increased agglomeration on productivity for the UK (Rice et al, 2006; Graham, 2007b), our paper is, to our knowledge, the first one to assess these effects at the firm and the plant levels.

The main finding of the paper is that we can, in general, detect weak statistically significant productivity effects at the firm level. We do not find any evidence on positive employment effects at the postcode sector level. These effects are likely to be due to firm mobility and start-ups. We also find some indication that productivity measured at the postcode sector level may have increased, but these effects may be due to sorting of more productive firms to areas with higher effective density.

The rest of the paper is structured as follows. Section 2 presents the empirical methodology and explains the construction of the accessibility, productivity and employment measures. Section 3 describes the data used, while in section 4 the empirical results are presented and discussed. Finally, in section 5 we conclude.

2. Empirical methodology

We first describe the empirical strategies used and then discuss the measurement of effective density/accessibility.

2.1. Estimation of firm level productivity and employment effects

Our basic approach is to use regression analyses to estimate the average effect of changes in accessibility induced by transport changes on firms' and plants' employment and on firms' Total Factor Productivity (TFP). We use data on plants (local units – LU) and on firms to which these plants belong to (reporting units - RU). Economic data (value added, capital, employment, sector, number of plants) is available only at the RU level. For the LU only data on employment and year of birth is available[2].

We have detailed geographical information on the location of the plants (postcodes). A UK postcode unit corresponds to a limited number of addresses (around 14) or a single big delivery point. As explained in the introduction, we aggregate up to postcode sector level, which leaves us with around 10,000 geographical units. We use data for 7 years: 1998 to 2004.

For the estimation of employment and productivity effects at the firm level we use the Annual Respondents Database (henceforth ARD). This database is described in detail in Appendix A[3]. Employment regressions are estimated at the plant level while the productivity regressions are estimated at the firm level due to the lack of economic data at the plant level.

The underlying empirical model of these outcomes has a component A that depends on the economic mass/agglomeration/employment accessibility of the location j in which the firm i is sited (j will correspond to the postcode sector):

---

[2] We obtained the data on the year of birth of the plant from the Inter Departmental Business Register (IDBR) dataset which we merged to the ARD using plant reference numbers.

[3] In the original ARD data a firm can be singleton or have several plants. Plants are identified through the reference code and the postcode they report. For a given firm the reference code of its plant(s) may change over time due to administrative reasons. We cannot differentiate pure administrative changes in the reference code from genuine new plants in the same postcode sector. If a firm has a plant present in a given postcode sector over time, regardless of its reference code, we consider it a singleton. For the firms with several plants in the same postcode sector, we collapse these in a single plant. Therefore multi-plants are firms that have plants present in more than one postcode sector overtime, independently of the number of plants they have in that postcode sector.

$$y_{ijt} = \beta A_{jt} + x'_{ijt}\gamma + \mu_{ij} + \tau_t + \varepsilon_{ijt} \qquad (1)$$

Here, y denotes the outcome of interest (employment or value-added) for a plant/firm *i*, in location *j* at time *t*, *A* is a measure of employment accessibility along the transport network at location *j* (representing agglomeration, and described below), and *x* represents other firm inputs (labour and capital) and area characteristics.

The set of area characteristics includes distances to various transport facilities, distance to the closest road improvement opened during our period of analysis, employment rate, average age, proportion of population aged 16-74 with higher education and proportion of population living on social housing. This set of controls will be used across different specifications of the paper. The area characteristics are measured in one year and do not change over time in our data. For multi-plant firms, accessibility A and the control variables X are calculated as the average over postcode sectors in which the firm operates weighted by the firms' employment in each postcode sector.

The model includes firm fixed effects μ, which represents the unobserved time-invariant productivity component of firms[i]. The firm fixed effects μ depend partly on unobserved time invariant productivity advantages for all firms located at *j*. Year fixed effects τ represent general changes that influence all firms and locations in the study area in a given year (e.g. macro shocks). In most specifications we allow the year effects to vary by industry sector using six sectors. Both the dependent variable and the accessibility variable are in logs. Thus, parameter β is the elasticity of employment/productivity with respect to accessibility. Traditional estimates of agglomeration effects are usually based on OLS estimates of models like (1). These estimates are biased when unobserved firm effects μ are correlated with accessibility - if for example, and as seems likely, better transport connections and higher employment density have evolved in places with productive advantages.

A first step to eliminating these biases is to difference the data over time to eliminate fixed-over-time firm effects μ:

$$\Delta y_{ijt} = \beta \Delta A_{jt} + \Delta x'_{ijt}\gamma + \delta_t + \Delta\varepsilon_{ijt} \qquad (2)$$

In equation (2), Δ denotes time-differencing. We use first differencing to study the immediate effects as well as long-differences (from two-year to six-year differences) to allow for sluggish responses and anticipatory effects. This way the firm fixed effects disappear and the time-differenced estimates are robust to time-invariant firm heterogeneity that can be arbitrarily correlated with accessibility. The area characteristics are included as controls in the differenced equation even if they do not vary over time to control for different trends in productivity or employment growth in areas with different background characteristics. This formulation is a starting point for evaluating the effects of transport policy on firms, because transport improvements generate changes in *A* over time which can be exploited in the estimation of β. Note, that this estimation strategy ignores whether or not the specific firms or their employees or customers in fact make any use of the transport improvements that have been put in place. The productivity effects that are estimated are thus analogous to "intention to treat" estimates in the programme evaluation literature, and are the expected productivity changes for firms or areas exposed to the treatment (transport improvement).

In general, the change in accessibility ΔA_{jt} could come through changes in the spatial distribution of employment, or because of changes in the transport network. Changes in the road network may be correlated with relocation of employment across space, which may lead to bias in the estimation of employment and productivity effects of road improvements. We instrument changes in accessibility (due to changes in both employment and transport network) with a measure of ΔA_{jt} , which only picks up changes in the transport network. We calculate theaccessibility based on the pre-improvement spatial distribution of employment (year 1997). Weestimate (2) by two-stage least squares using this as an instrument for actual changes in accessibility (see discussion of the construction of A (section 2.5) for more detail).

For multi-plant firms, an additional issue is that changes in the accessibility variable may be driven by restructuring of the firm through plant births and deaths and changes in the employment shares of plants. In the firm level productivity regressions we use three different definitions for the panel identifier, which differ by the degree of restructuring allowed. 1) In one extreme, we define the firm identifier so that it refers to a firm that retains the same spatial structure over time, i.e. the firm has employment in the same set of *postcode sectors* in different years. If the set of postcode sectors in which the firm has employment changes, we define it as a new firm. This ensures that variation in accessibility is not driven by restructuring of the firm through death and birth of plants.

The disadvantage of the stable postcode sector structure is that we discard much of the observations, especially for larger firms operating in several areas. 2) As an intermediate case, we relax the definition for stable spatial structure somewhat by requiring the firm to operate in the same set of *postcode districts* (around 3000 districts) in stead of *postcode sectors* (around 10000 sectors). 3) Finally, we estimate the model allowing the spatial structure of the firm to change freely. In all three cases, we use the spatial structure of the firm in the first year it is observed, when calculating the instrument for accessibility. Thus, identification is based on changes in accessibility due to road improvements calculated keeping postcode sector employment shares constant.

Estimation of (2) using within-firm changes in a panel of firms is only feasible using firms that exist, and appear in the data, both before and after the opening of the transport schemes that are being evaluated. This introduces sample selection issues. Firstly, firms that stay in the location of the transport scheme are likely to be those that can benefit most from it. In addition, there will be sampling-related reasons why some firms appear in our data in multiple years whilst others do not.

These caveats aside, estimation of β from the changes within firms over time provides the best guide to the micro-level agglomeration impacts of transport improvements for firms that experience the change in transport costs within an otherwise unchanged operating environment.

2.2. Estimation of 'aggregate' area-level productivity effects

We can also estimate the aggregate changes in productivity and employment that occur at the postcode sector level[4] (or other geographical level). As explained in the introduction, the advantage of using postcode sectors as the geographical units is that they are very small spatial units, so we can identify phenomena that would be unobservable at a higher geography level.

Implementing this approach relaxes the requirement that the same firms are observed before and after the transport improvement. The postcode-sector-level components can be estimated from the cross-section of firms in each year. Changes in these area level components can then be regressed on changes in accessibility at the area level.

Formally we use a two step approach. We first estimate a regression of the form:

$$y_{ijt} = x_{it}'\gamma + \eta_{jt} + \varepsilon_{ijt} \qquad (3)$$

using plant level data, where η_{jt} are area-by-year effects to be estimated, and x_{it} includes firm level characteristics. We recover estimates of the area-by-year effects $\hat{\eta}_{jt}$ and regress these on the accessibility variable A_{jt} after differencing to eliminate fixed-over-time area effects as in (2) i.e. we estimate

$$\Delta\hat{\eta}_{jt} = \beta\Delta A_{jt} + \Delta u_{jt} \qquad (4)$$

using a panel of postcode sectors. Again we use first differencing and long differences (up to six year difference). We also include the aforementioned time-invariant control variables in the estimation to control for different trends.

This two-step method provides estimates of the *aggregate* effects of the transport improvements at the geographical level, taking into account firm exit, entry and geographical relocation. In the

[4] The postal codes (postcodes) used in the United Kingdom provide very detailed geographical information. Postcode sectors codification eliminates the last two digits of the full postcode and so they gather between 15-20 postcodes on average. More information available at: http://www.ons.gov.uk/about-statistics/geography/products/geog-products-postcode/nspd/index.html

estimates below we use postcode sectors, but the principle could be applied at broader geographical scales.

2.3. *Targeting of transport policy*

Both the firm-level estimates from (2) and the area-level estimates from (3)-(4) will produce biased estimates of the productivity effects of transport improvements, if areas with increasing or declining productivity are those that experience the greatest accessibility changes. This implies that $\Delta\varepsilon_{ijt}$ in Equation (2) and Δu_{jt} in Equation (4) are correlated with ΔA_{jt}. The usual reason to suspect this kind of problem is the possibility that transport policy is endogenous to the productivity trends in the targeted locations, i.e. the decision to improve the transport network might be partly driven by productivity trends.

There is a limited amount we can feasibly do to eliminate this potential source of bias. Our main strategy is to focus the empirical analysis on places and firms that are close to the transport improvement sites. In the results section below we present estimates for samples within distances of 10km, 30km and 50km of the sites of improvement. In this way we are comparing closely neighbouring places that differ incrementally in terms of the change in accessibility they experience as a result of the road network improvements. We assume that these changes in accessibility close to transport schemes are an incidental by-product of the scheme rather than its intended outcome. The main changes in mean travel times and employment accessibility occur close to the end points of new road schemes, whereas they are typically intended to improve the flow of traffic between cities or areas further away from the improvement. There are also often long delays between commissioning and opening of road schemes, which will weaken any link between pre-existing local productivity trends and the decisions over where to site these projects.

We also control for some salient area-level variables in the regression, namely distances to nearest transport facilities - motorway junctions, rail stations, airports, ports (see Appendix C), and the straight-line distance to the transport scheme itself. In addition, we include background characteristics of the postcode sector derived from the 2001 census. The census controls include the share of residents with high qualifications, share of residents unemployed, share of residents in social housing, and the mean age of residents. Both the distance variables and the census variables are time in-variant. Hence, including them in the time-differenced equations (2) and (4), controls for the linear dependence of the trend of productivity growth on them.

2.4. *Pre-post analysis Top-bottom analysis*

The methods described above estimate the elasticity of productivity or employment with respect to accessibility by using a continuous treatment variable. The accessibility variables are likely to measure actual accessibility with error which may cause bias and inaccuracy in our estimates. In order to alleviate the potential measurement issues, we discretize the accessibility changes by dividing the sample into two based on the change in accessibility between 1998 and 2004. We then regress the log of accessibility change from 1998 to 2004 on a dummy that gets the value one for the firms/post codes belonging to the top 50% of accessibility change. The top-bottom analysis is only done for the six-year differenced model.

2.5. *Defining the accessibility index/effective density/agglomeration index*

The accessibility index used is identical in structure to market potential measures used in economic geography (e.g. Harris 1954), and the accessibility indices used more generally in the transport literature (e.g. Ahmed et al 2006, Vickerman et al 1999). This index is a measure of the economic mass accessible to a firm in a particular location, given the local transport network.

Consider a measure of economic activity or other variable of interest, such as employment l. For a firm in an origin location *j* at time *t*, an employment accessibility index A_{jt} is a weighted sum of employment in all destinations k that can be reached from origin *j* by incurring a transport cost c_{jkt} along some specified route between *j* and *k* (e.g. straight line distance, minimum cost route along a transport network). That is, the index has the structure

$$A_{jt} = \sum_{k \neq j} a(c_{jkt}) l_{kt} \qquad (5)$$

Note, that changes in *A* are partly driven by changes in employment in destinations k. This may lead to endogeneity problems in the estimation of the effect of transport improvements, if the road improvements are disproportionately allocated to places with growing (or declining) employment.

In the empirical work below, we , by fixing employment at the level at the first in year in our data sample (1997). We calculate accessibility with initial employment as

$$\hat{A}_{jt} = \sum_{k \neq j} a(c_{jkt}) l_k \qquad (6)$$

and use it as an instrument for actual accessibility.Fixing employment to 1997 level ensures that changes in the accessibility index (6) over time occur only as a result of changes in the costs c_{jkt} (e.g. travel time) and not changes in employment.

Note that, as a consequence, changes in accessibility calculated in this way may understate the true change in employment accessibility arising from transport improvements, if these changes induce shifts in employment towards more accessible areas. As it turns out, we do not find evidence of movements of this type. For multi-plant firms, accessibility is calculated as average over accessibility of its plants weighted by their employment shares. The instrument is calculated keeping these employment shares constant at the first year level.

In Equations (5) and (6) the value of the weight *a*(.) attached to any destination k is a decreasing function of the cost of reaching destination *k* from origin *j*. Potential weighting schemes include: 'cumulative opportunities' weights $a\left(c_{ijt}\right)=1$ if j is within a specified distance of i, zero otherwise; exponential weights $a\left(c_{ijt}\right)=\exp\left(-\alpha c_{ijt}\right)$; logistic weights $a\left(c_{ijt}\right)=\left[1+\exp\left(-\alpha c_{ijt}\right)\right]^{-1}$ or inverse cost weights $a\left(c_{ijt}\right)=c_{ijt}{}^{-\alpha}$. See Graham, Gibbons, Martin (2009) for further discussion of these indices. In line with common practice, we use the simple inverse cost weighting scheme $a\left(c_{ijt}\right)=c_{ijt}{}^{-1}$ in which the cost is the estimated travel time. The precise functional form of the distance decay function in these employment accessibility indices is likely to be of second-order importance.

2.6. Effects on accessibility arising from transport improvements

This type of accessibility/market potential/effective density index can be applied to the study of productivity improvements arising from transport projects, assuming the costs c_{jkt} are calculated using routing along the transport network. This works, because transport improvements change the structure of costs c_{jkt} along the transport network and the structure of costs along routes from *j* to potential destinations *k*. This in turn changes the accessibility index (implying more agglomeration).

For example, consider a transport improvement 'treatment' that involves a journey time reduction on a road link between two nodes *p* and *q*. This treatment will have a first order effect on the costs of the least-cost route between *j* and *k* if:

a) the least-cost route between *j* and *k* passes along the link *p-q* in both the pre and post-improvement periods, such that the transport improvement reduces the cost of the journey along *p-q* and brings employment at destination *k* 'closer' to origin *j* in cost terms.

b) the least-cost route between *j* and *k* bypasses link *p-q* in the pre-improvement period, but switches to use the link *p-q* in the post-improvement period because of the reduction in costs; again this brings employment at destination k 'closer' to origin j in cost terms

There are also 'second order' effects arising when:

c) the least cost route between *j* and *k* bypasses link *p-q* in both the pre-improvement and post-improvement periods. However, journeys between other origin and destination pairs have switched to using the link *p-q*, which reduces congestion on the alternative links in the network used by the routing between *j* and *k*; again this brings employment at destination *k* 'closer' to origin *j* in cost terms.

In the empirical work below we focus only on the first order effects of type a) and b) arising from new transport infrastructure. We have to ignore second order effects of type c) because our road transport network data does not allow us to observe changes in travel time induced by changes in congestion occurring as a result of transport improvements (we have no information on traffic flows observed prior to the improvements).

Changes in cost of all these types imply changes in the accessibility index A_{jt} (i.e. a change in agglomeration). The amount of change in the accessibility index at a location *j* depends on the likelihood that a route between *j* and *k* uses the improved link *p-q*. The idea in our method is to use the changes in the accessibility index at each location *j* to estimate the extent to which firms in location j are "potentially" affected by the transport improvements under investigation. In turn, this change in potential accessibility enters into our productivity regressions, as described in Section 2.1 above.

3. Data setup and sample selection

3.1. IDBR/BSD data

The first component in the construction of the employment accessibility index *A* is a source of geographically localised employment data. We use data from the Interdepartmental Business Register/ Business Structure Database (IDBR/BSD) held by ONS. This dataset holds information on the population of businesses in the UK, with information on employment and industrial classification. In this dataset the plants are referred to as the 'Local Units', which belong to different 'Reporting Units' (firms). The IDBR/BSD records plant location down to the full postcode level, although for our analysis we aggregate employment to postcode sectors (roughly 10,000 postcode sectors). We combine this dataset (linked by reporting unit) with the Annual Respondents database (ARD) which includes more detailed information on economic variables (as explained below in Section 3.3).

When we construct accessibility index *A* in equation (7) we use employment in a base year: 1997. Fixing the spatial distribution of employment to this base period ensures that changes in *A* are attributable (by construction) only to changes in the transport costs caused by the selected transport schemes, and not to employment relocation (which may be endogenous to increased accessibility).

3.2. NTM road network data and O-D matrix construction

The second component in the employment accessibility index is an origin-destination (O-D) matrix containing the costs c_{jkt} (journey time) between each origin and destination. This matrix is required for both the pre-improvement and post-improvement periods.

We use data on traffic speeds and flows on a generalised primary road GIS network for Great Britain provided by the Department for Transportation (DfT). Traffic flows are modelled from traffic flow census data using the FORGE component of the National Transport Model. We construct the road network for year 1997 – 2003 (end of the year) by using the 2003 network provided by the DfT and information on the opening years of road links.

Using this generalised traffic network, we use the network analysis algorithms in ESRI ArcGIS to compute least-cost (minimum journey time) routes between each origin postcode j and destination postcode k in years 1997 - 2003.

When computing the O-D matrix we apply a limit of 80 minutes drive time in computing the matrix. The estimation sample for production and employment effects is subsequently limited to a maximum radius of 50km of the schemes.

We use journey times in the non-busy direction averaged over all time periods between Monday-Friday 08:00 and 18:00. We focus on non-busy travel directions because the busy travel directions are, in principle, more sensitive to changes in congestion induced by new travel links.

Therefore, the accessibility index *A* is calculated using the postcode-sector employment data from IDBR/BSD, and the postcode-sector-to-postcode-sector O-D travel times in the pre and post periods using equation (7).

It should be noted that the network is highly generalised. Journeys via the minor road network are not modelled. Forbidden turns and one way systems are not modelled. All link intersections are treated as junctions. The changes in accessibility must therefore be regarded as approximate.

3.3. Estimation samples from the Annual Respondents Database (ARD)

For the productivity regressions we use the Annual Respondents database (ARD) which includes information needed to measure value added and capital at the firm level. For the employment regressions in Section 4.2 we use Local Units (plants) as the unit of observation.

In the first part of the production function analysis in Section 4.3 we consider only Reporting Units that appear in at least two years. In our main specification, the unit of analysis are Reporting Units (firms) that have employment in the same set of postcode sectors in the different years. We treat firms with changes in their postcode sector structure as new observations, since measured changes in accessibility would be due to relocation, not the transport schemes. However, we also estimate models using stable postcode district structure (in stead of stable postcode sector structure) and unrestricted spatial structure.

Implementation of the methods set out in Section 2.1 for productivity analysis are made more complicated by the fact that the Annual Respondents Database does not provide productivity or other economic data (apart from employment) at the plant level. Instead, economic data on value added and capital is observed at Reporting Unit level, which corresponds to one or more Local Units (plants). We assign to a Reporting Unit (firm) the average of the accessibility index of its component Local Units (plants). We do this using weights that correspond to a Local Unit's share in its Reporting Unit's total employment. We use employment at the first year a Local Unit is recorded in the data, to avoid problems induced by firms endogenously shifting employment between Local Units in response to productivity changes,

In the second part of the production function analysis explained in Section 2.1 and presented in Section 2.2, we construct a panel of postcode sectors and estimate postcode-sector level productivity changes. In this analysis we use all ARD Reporting Units and Local Units, regardless of how many times they appear in the ARD sample over the period.

The Reporting Unit/Local Unit data structure has to be accommodated in this area-level analysis too. The fact that accessibility is relevant at plant level implies that Equation (3) should be estimated at Local Unit level. But the fact that economic data is available only at firm Reporting Unit level implies that Equation (3) should be estimated at Reporting Unit level. There are no perfect ways of resolving this conflict.

The first method we use is to apportion all the economic variables (value-added, capital) to Local Unit level based on Local Unit employment (on which we have data) and estimate Equation (3) at Local Unit level. Allocating outputs and capital in proportion to employment assumes constant returns to scale.

The second method we use is to estimate (3) at Local Unit level, with Reporting Unit economic variables assigned equally to each Local Unit. However, the regression is weighted using the Local Unit's share in the employment of the Reporting Unit to which it belongs. This means we put more weight on plants that take a greater share in Reporting Unit output, and do not discard the information in multi-plant firms. The implicit assumption in this method is that Local Unit production is related to Reporting Unit production by a multiplicative structure[5]. This implies that log of value-

[5] That is, assuming the local unit production function has the Cobb-Douglas form, the production structure is:

added, and log of capital are allocated to Local Units in proportion to the Local Unit's employment share.

Both the above methods have the drawback that estimation does not take into account productivity differences between Local Units within Reporting Units. Hence imputed value-added will be too low in plants in high productivity areas, and it will be too high in low-productivity areas. This could clearly be important, when it comes to investigating the effects of local area productivity changes induced by transport improvements.

4. Descriptive statistics and results

4.1. Accessibility indices:

We compute employment accessibility for each postcode sector for years 1998-2004. In each year, we take account of the new road transport infrastructure that was opened in the previous year when we compute the accessibility indices from the road transport network. The empirical analysis is thus based on the changing geographical pattern of accessibility as the road transport system evolved over the period 1997-2003. As shown in table 1 we have around 11,000 postcode sector every year, which give us around 80,000 observations of the accessibility measures during the period of analysis. We also report the number of postcode sectors within each of the distance bands (10, 30 and 50 kilometres) we will use in the main analysis. About 88% of postcode sectors are within 50 kilometres from a road improvement opened during the period of analysis.

Table 1: *Number of postcode sectors by distance band*

	All	10 kms	30 kms	50 kms
Annual	11262	3529	8141	9977
Total	78834	24703	56987	69839

Tables 2 and 3 summarise the changes in the log of accessibility (inverse travel-time-weighted employment sum) arising from the transport schemes in 1998-2003. In table 2 we calculate the accessibility allowing both the employment and the road infrastructure to vary. In table 3 the accessibility measure is calculated fixing employment at 1997, so all the variation in accessibility

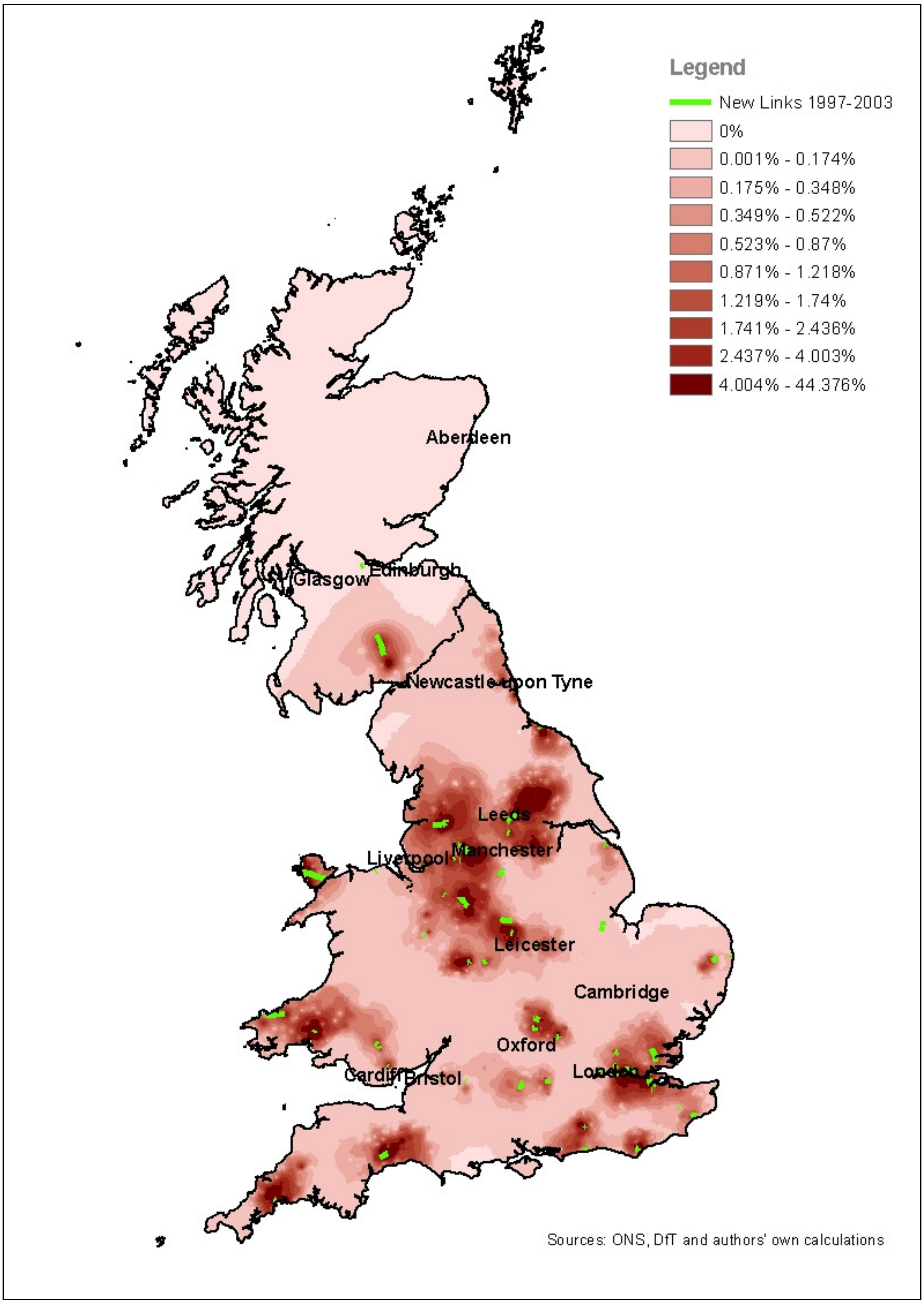
$$\ln y_{lu} = \pi_{lu} \ln y_{ru} = \pi_{lu}\beta_0 + \beta_1\pi_{lu} \ln k_{ru} + \beta_2\pi_{lu} \ln l_{ru}$$

$$\sum_{lu \in ru} \pi_{lu} = 1$$

$$\ln y_{ru} = \sum_{lu \in ru} \ln y_{lu} = \beta_0 + \beta_1 \ln k_{ru} + \beta_2 \ln l_{ru}$$

comes from changes in the road network. In both tables, panel A reports annual changes in accessibility while panel B shows long differences between 1998 and 2004.

Figure 1: Changes in postcode-sector employment accessibility resulting from transport improvements



The results in panel A of Table 3 show that the annual changes in employment accessibility induced by these road improvements were on average very small. However, the 99th percentile shows that some pcsects saw large accessibility changes. For example, for the changes within the 10 km distance band the mean is only 0.3% and the 99th percentileis 6.3%. As we expand the sample away from the sites of the schemes, the mean and highest ranked changes in accessibility all tend to fall. Within 50km of the schemes mean accessibility change is 0.15% and 99th percentile is 3.5%.

Panel B of Table 3 shows that when we sum up all the accessibility changes over the period, mean accessibility change is 1.9% within 10km of the schemes and 0.93% percent within 50km.

Table 2: *Changes in postcode-sector employment accessibility resulting from transport improvements – economic mass is annual employment*

Panel 2.A: *Year to year changes 1998-2004*

	Mean	Std. Dev.	1st pctile	99th pctile	Obs
All	-0.003	0.0563	-0.216	0.0977	65196
10 kms	-0.0026	0.0558	-0.216	0.1019	21054
30 kms	-0.0023	0.0541	-0.214	0.0929	48102
50 kms	-0.0025	0.055	-0.216	0.0935	58350

Panel 2.B: *Changes from 1998-2004*

	Mean	Std. Dev.	1st pctile	99th pctile	Obs
All	-0.018	0.0881	-0.293	0.1743	10866
10 kms	-0.0158	0.0817	-0.293	0.2129	3509
30 kms	-0.0141	0.0751	-0.245	0.1607	8017
50 kms	-0.0147	0.0756	-0.234	0.1605	9725

Table 3: *Changes in postcode-sector employment accessibility resulting from transport improvements – economic mass is employment in 1997*

Panel 3.A: *Year to year changes 1998-2004*

	Mean	Std. Dev.	1st pctile	99th pctile	Obs
All	0.0014	0.0098	0.000	0.0287	65196
10 kms	0.0031	0.0163	0.000	0.0626	21054
30 kms	0.0018	0.0113	0.000	0.0371	48102
50 kms	0.0015	0.0103	0.000	0.0315	58350

Panel 3.B: *Changes from 1998-2004*

	Mean	Std. Dev.	1st pctile	99th pctile	Obs
All	0.0083	0.0238	0.000	0.1002	10866
10 kms	0.0186	0.038	0.000	0.1787	3509

30 kms	0.011	0.0271	0.000	0.1123	8017
50 kms	0.0093	0.025	0.000	0.1045	9725

The changes in (log) accessibility of panel 3.B are mapped in Figure 1. It shows the geographical distribution of the changes in accessibility induced only by road improvements. We also depict the new road links constructed between 1997 and 2003. It is seen that areas close to the new schemes saw greater accessibility increases.

4.2. Employment

The first regression results, presented in Table 4 are Local Unit (plant) level regressions of (log) employment on (log) accessibility using the ARD data. As explained in the methodology, we define a LU as the presence of a RU in a given postcode sector, regardless of the number of plants it has in that postcode sector. Table 4 contains the coefficient on employment, its standard error and number of observations for 7 different specifications and for the three distance bands.

All models include sector specific year fixed effects and controls. We show results by OLS, by first differences and instrumental variables regression. In the IV regressions, we instrument the actual accessibility changes with changes in accessibility keeping employment fixed at 1997 level. This way, we only use the variation in accessibility changes coming from road improvements and not from employment relocation. Excluded instruments F-test is always very high and above 10 indicating that our instrument is strong. We report IV-results for 1, 2, 4 and 6 year differences.

***Table 4**. Employment regressions for plants – panel 1998-2004*

Dependent variable: log of plant employment

		10km	30km	50km
OLS	Coefficient	0.0406***	0.0654***	0.0826***
	Std error	0.0047	0.0032	0.0028
	N	396610	827705	1024708
First differences	Coefficient	-0.008	-0.0004	0.0049
	Std error	0.0241	0.0165	0.0145
	N	196817	408384	505048
First differences - IV	Coefficient	-0.0115	-0.0052	-0.0109
	Std error	0.066	0.0665	0.067
	N	196817	408384	505048
2 year differences - IV	Coefficient	-0.0056	0.0384	0.0219
	Std error	0.0931	0.0919	0.0926
	N	123996	256300	316886
4 year differences - IV	Coefficient	-0.0924	-0.041	-0.1035
	Std error	0.1025	0.1001	0.101

	N	52962	109315	135176
6 year differences - IV	Coefficient	-0.0599	0.0242	-0.0023
	Std error	0.1857	0.1601	0.1614
	N	13243	27284	33775
6 year - top/bottom	Coefficient	-0.0057	-0.0094	-0.0102
	Std error	0.0147	0.0101	0.0094
	N	13243	27284	33775

All regressions include controls and sector specific time fixed-effects

Standard errors are clustered at postcode sector level

* p<0.10, ** p<0.05, *** p<0.01

OLS estimates suggests positive effects of accessibility on employment, but once we include plant fixed-effects through differencing we find no effect. The coefficient is statistically insignificant also in the IV specifications suggesting that improved accessibility does not induce firms in the area to increase their employment. The results are, however, not conclusive as the estimates become imprecise in the IV specifications.The last model of table 4 reports results on employment using long differenced data and a discrete treatment variable. The sample includes postcode sectors that belonged to the top half or bottom half of accessibility changes between 1998 and 2004 due to road improvements. The estimates are the coefficients of a dummy variable for belonging to the top half. As before, we do not find statistically significant employment effects.

4.3. *Total Factor Productivity*

In this section we report results on the effect of accessibility on productivity both at the firm level and at the aggregate (postcode sector) level. They correspond to specifications explained in sections 2.1 and 2.2.

Results from the Reporting Unit production function regressions are reported in Table 5. All models control for capital and labour and include sector specific year variables as well as area characteristics. The sample is based on a panel of Reporting Units with a stable postcode sector structure. We use the firms which are present in the same set of postcode over the period of analysis. The dependent variable is value-added and the table reports only coefficient on employment accessibility, its standard error and the number of observations used in each model. All variables are in natural logarithms, so the coefficients show the elasticity of value-added with respect to accessibility. Because the regressions control for capital and labour, the coefficient estimates correspond to the effect of accessibility on Total Factor Productivity. All specifications include controls. As in table 4, we show results for OLS, first differences and instrumental variables regression. We also show the results for regressions with the discrete treatment.

OLS results show elasticities around 15%. First difference results suggest negative effects, but once we instrument the estimates become imprecise. However, if we compare the firms located in the top half of the postcode sectors with higher changes with those located in the bottom half we find positive and statistically significant estimates with a 30 km distance band.

Table 5. *TFP regressions using stable postcode sector for RUs*

Dependent variable: log of firm value added

		10km	30km	50km
OLS	Coefficient	0.1536***	0.1428***	0.1452***
	Std errors	0.0083	0.0054	0.0047
	N	33447	78899	98935
First differences	Coefficient	-0.1246*	-0.0891*	-0.0637
	Std errors	0.071	0.047	0.0391
	N	17010	40296	50557
First differences - IV	Coefficient	0.0439	-0.033	-0.076
	Std errors	0.1235	0.1198	0.1194
	N	17010	40301	50562
Two differences - IV	Coefficient	-0.0272	0.0441	-0.0041
	Std errors	0.1762	0.1766	0.1688
	N	6310	15271	19147
Four differences - IV	Coefficient	0.3559	0.2482	0.2154
	Std errors	0.261	0.2501	0.2337
	N	2780	6661	8418
Long differences - IV	Coefficient	0.5269	0.3824	0.1934
	Std errors	0.4328	0.3508	0.3267
	N	736	1649	2102
Long differences - top/bottom	Coefficient	0.0387	0.0684**	0.0266
	Std errors	0.0524	0.0321	0.0299
	N	736	1650	2103

All regressions include controls and time fixed-effects

Robust standard errors and clustered at postcode sector level

* p<0.10, ** p<0.05, *** p<0.01

Tables 6 to 8 show the results for the two-step methodology explained in section 2.2. We first estimate postcode-sector productivity effects for every year using all Local Units from all Reporting Units, regardless of their location. We then regress these estimated productivity effects on the accessibility variable, whilst controlling for postcode-sector fixed effects (through time-differencing) and other area-level control variables.

Table 6 shows the results for the 1st step, where we calculate the TFP residuals at the postcode sector level. Economic data is available at the RU level while postcode sector is available at the LU level. Therefore, in order to calculate the postcode-level residual we need to run the productivity regressions at the LU level. In column 1 of Table 6 we apportion value added and capital of the firm to the plant based on the plant's share of firms' employment. In column 2 of table 6 we apportion them equally and then we weight the regressions by employment of the local unit. Our dependent variable is the log of real value added, which we regress on the inputs (capital and labour) and on some plant/firm characteristics (age, age to the square, multi-plant/singleton). We include postcode sector*year fixed effects. All coefficients are highly significant and the explanatory power of the models is very high. We save the postcode sector*year fixed effects to be used as our dependent variable in the 2nd step.

***Table 6**. 1st-step regressions: calculation of TFP postcode sector-year residuals by methodology*

Dependent variable: Log of local unit/RU-weighted real value added

	1		2
Log of local unit	0.6385***	***Log of reporting unit***	0.6650***
employment	0.0015	***employment***	0.0038
Log of local unit	0.3344***	***Log of reporting unit***	0.2915***
weighted capital	0.0011	***weighted capital***	0.0023
Age of the local unit	0.0042***	***Average age of plants of the***	0.0088***
	0.0006	***reporting unit***	0.0012
Age of the local unit	-0.0002***	***Average age of plants of the***	-0.0002***
to the square	0.000	***reporting unit to the square***	0.000
Belongs to a single	0.2045***	***It is a single***	-0.0735***
plant firm	0.0039	***plant firm***	0.0063
Constant	2.0075***	***Constant***	2.4214***
	0.005		0.0135
Absorbing indicators	pcsect_year	***Absorbing indicators***	pcsect_year
Weights	No	***Weights***	Employment
N	1016436	***N***	1073274
R2	0.850	***R2***	0.895

Standard errors are clustered at the postcode sector level

*** p<0.10, ** p<0.05, *** p<0.01**

Tables 7 and 8 show the results of the 2nd step by the two different methodologies. Both sets of results are very similar. We find weak but significant effects using OLS. First differences and IV results are very imprecise. As for the firm level TFP regressions, we do find statistically significant positive effects when we compare the top/bottom changes in accessibility.

Table 7. *2nd-step regressions: effect of accessibility on aggregate TFP – residuals from 6.1*

Dependent variable: Postcode sector - year TFP residual

		10km	30km	50km
OLS	Coefficient	0.0226***	0.0225***	0.0258***
	Std errors	0.0086	0.0057	0.0049
	N	19689	43617	53629
First differences	Coefficient	-0.1393**	0.0583	0.0603
	Std errors	0.0652	0.0424	0.0373
	N	16789	37056	45591
First differences - IV	Coefficient	-0.1386	-0.0736	-0.1116
	Std errors	0.1357	0.1348	0.1356
	N	16747	36880	45376
Two differences - IV	Coefficient	-0.2702	-0.2101	-0.2315
	Std errors	0.1654	0.1595	0.1598
	N	13924	30600	37666
Four differences - IV	Coefficient	-0.1134	-0.0607	-0.1117
	Std errors	0.1329	0.1292	0.1301
	N	8350	18333	22568
Long differences - IV	Coefficient	-0.0789	0.0035	-0.0888
	Std errors	0.1954	0.1769	0.1759
	N	2791	6127	7537
Long differences - top/bottom	Coefficient	0.1207*	0.4491***	0.6850***
	Std errors	0.0654	0.082	0.1459
	N	2791	6127	7537

All regressions include controls and time fixed-effects

Robust standard errors and clustered at postcode sector level

* p<0.10, ** p<0.05, *** p<0.01

Table 8. *2nd-step regressions: effect of accessibility on aggregate TFP - residuals from 6.2*

Dependent variable: Postcode sector - year TFP residual

		10km	30km	50km
OLS	Coefficient	0.1226***	0.1287***	0.1231***
	Std errors	0.015	0.0094	0.0079
	N	19720	43762	53806
First differences	Coefficient	-0.2645**	-0.1115	-0.0876
	Std errors	0.1133	0.0758	0.067
	N	16815	37162	45723
First differences - IV	Coefficient	-0.1028	-0.0428	-0.1262
	Std errors	0.3654	0.3578	0.3574
	N	16788	37065	45603

Two differences - IV	Coefficient	-0.0757	-0.0711	-0.1205
	Std errors	0.3325	0.3192	0.319
	N	13967	30785	37890
Four differences - IV	Coefficient	-0.0434	0.063	-0.0336
	Std errors	0.2413	0.2293	0.229
	N	8371	18421	22672
Long differences - IV	Coefficient	0.2296	0.3636	0.22
	Std errors	0.3195	0.2824	0.2839
	N	2793	6133	7549
Long differences - top/bottom	Coefficient	0.123	0.3890***	0.2935
	Std errors	0.1318	0.1452	0.2263
	N	2781	6077	7481

All regressions include controls and time fixed-effects

Robust standard errors and clustered at postcode sector level

* p<0.10, ** p<0.05, *** p<0.01

5. Conclusions

This paper uses a novel methodology in order to asses the productivity and employment effects from transport improvements at a very detailed geographic scale. We construct employment accessibility changes at the postcode-sector level using simple GIS network analysis combined with the register of local business employment (IDBR). These accessibility indices are linked to local units (plants) in the Annual Respondents Database (ARD) for the analysis of employment and productivity.

Overall, when we examine the combined effect of all major road transport improvements between 1998 and 2003 with firm level data focusing on firms and plants that remain in situ before and after the opening of new stretches of road, we find insignificant effects on the employment of firms. When we examine total factor productivity of firms with stable spatial structure, we find some indication that firms with higher accessibility changes due to road improvements saw larger productivity increases. Also postcode sector level productivity regressions allowing for start-ups and closures, give some support to the idea that productivity increased accessibility caused by a better road network may lead to productivity gains.

However, our results are insignificant in many specifications. Hence, the results do not sit comfortably with common-sense notions of the benefits of transport improvements and the evident willingness to pay for travel time savings that have been demonstrated in the transport evaluation literature. However, it should be noted that we are not measuring productivity improvements of firms which necessarily make direct use of improved transport links, but rather the exposure of firms to accessibility improvements. Our methods will only detect the direct benefits of travel time savings to the extent that these are correlated with accessibility changes.

Some extensions to the methods and refinements will be considered in future analysis. Future work will evaluate the structure of the accessibility indices e.g. using residential population and sector specific employment rather than overall employment and experimenting with other assumptions about transport costs and travel speeds on the road network.

References

Behrens, K., Lamorgese, A., Ottaviano, G.I.P., and Tabuchi, T. (2004). "Testing the home market effect in a multi-country world: The theory", Centre for Economic Policy Research, Discussion Paper No. 4468.

Boarnet, M. (1998), "Spillovers and the Locational Effects of Public Infrastructure", Journal of Regional Science, Volume 38 Issue 3, Pages 381 – 400

Chandra, A and Thompson, E, (2000). "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system", Regional Science and Urban Economics, Elsevier, vol. 30(4), pages 457-490, July.

Combes, P.P and Lafourcade, M. (2001), "Transport Cost Decline and Regional Inequalities: Evidence from France", CEPR Discussion Papers 2894, C.E.P.R. Discussion Papers.

Coughlin CC and Segev E, (2000). "Location Determinants of New Foreign-Owned Manufacturing Plants", Journal of Regional Science, Blackwell Publishing, vol. 40(2), pages 323-351.

Duranton, G. and D. Puga (2004). "Micro-foundations of Urban Agglomeration Economies" in J. F. Thisse and J. V. Henderson (eds.) Handbook of Urban and Regional Economics, Volume 4. North Holland.

Gibbons, S & Machin, S, (2005)."Valuing rail access using transport innovations," Journal of Urban Economics, Elsevier, vol. 57(1), pages 148-169, January.

Graham D. J. (2006). "Wider economic benefits of transport improvements: link between agglomeration and productivity", Stage 2 report. DfT: London.

Graham D. J. (2007a). "Agglomeration economies and transport investments", OECD/ITF Joint Transport Research Centre Discussion Papers 2007/11, OECD, International Transport Forum.

Graham, D. J. (2008). "Agglomeration, productivity and transport investment". Journal of Transport Economics and Policy 41(3) pp. 317-343(27)

Graham, D. J., Melo, P. S., Jiwattanakulpaisarn, P and Noland, R. B., (2010). "Testing for causality between productivity and agglomeration economies", Journal of Regional Science 44 (10) 1-17.

Graham, D. J., Gibbons, S. and Martin, R. (2009). "Transport investments and the distance decay of agglomeration benefits", draft report for the Department of Transport

Harris, C. (1954).”Market as a Factor in the Localization of Industry in the United States”, Annals of the Association of American Geographers, 44 (4) 315-348

Holl, A. (2004a). “Manufacturing location and impacts of road transport infrastructure: Empirical evidence from Spain”, Regional Science and Urban Economics 34 (3): 341-363.

Holl, A. (2004b). “Transport infrastructure, agglomeration economies, and firm birth”. Empirical evidence from Portugal’, Journal of Regional Science, 44 (4): 693-712.

Holl, A. (2004c). “Start-Ups and Relocations: Manufacturing Plant Location in Portugal”, Papers in Regional Science, 83 (4): 649-668.

Holl, A. (2006). “A review of the firm-level role of transport infrastructure with implications for transport project evaluation”, Journal of Planning Literature 21 (1): 3-14.

Holl, A. (2007). “20 years of accessibility improvements. The case of the Spanish motorway building programme”, Journal of Transport Geography 15: 286-297.

Holtz-Eakin, D and Lovely, M E., (1996). "Scale economies, returns to variety, and the productivity of public infrastructure," Regional Science and Urban Economics, Elsevier, vol. 26(2), pages 105-123, April.

Jiwattanakulpaisarn, P, Noland, R B. and Graham, D J., (2010). "Causal linkages between highways and sector-level employment," Transportation Research Part A: Policy and Practice, Elsevier, vol. 44(4), pages 265-280, May.

Marshall A. (1920), “Principles of Economics”

Martin, P and Rogers, CA, (1995). "Industrial location and public infrastructure," Journal of International Economics, Elsevier, vol. 39(3-4), pages 335-351, November.

Melitz M. (2003). “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”, Econometrica 71, 1695-1725.

Melo, P. S., Graham, D .J. and Noland, R. B., (2009). “A meta-analysis of estimates of urban agglomeration economies”, Regional Science and Urban Economics, Elsevier, vol. 39(3), pages 332-342, May.

Michaels, G (2008) “The effect of trade on the demand for skill: evidence from the Interstate Highway System”. Review of economics and statistics, 90 (4). pp. 683-701.

Moreno, R. and Lopez-Bazo, E.(2007). "Returns to Local and Transport Infrastructure under Regional Spillovers", International Regional Science Review, Vol. 30, No. 1, 47-71

Puga D., (2002). "European regional policies in light of recent location theories," Journal of Economic Geography, Oxford University Press, vol. 2(4), pages 373-406, October.

Puga D., (2008). "Agglomeration and cross-border infrastructure," EIB Papers 9/2008, European Investment Bank, Economic and Financial Studies.

Ottaviano I. P., G., (2008). "Infrastructure and economic geography: An overview of theory and evidence," EIB Papers 6/2008, European Investment Bank, Economic and Financial Studies

Overman H.G. (2010) "'GIS a Job': What Use Geographical Information Systems in Spatial Economics?", Journal of Regional Science, Blackwell Publishing, vol. 50(1), pages 165-180.

Rice, R, Venables, A J. and Patacchini, E, (2006). "Spatial determinants of productivity: Analysis for the regions of Great Britain," Regional Science and Urban Economics, Elsevier, vol. 36(6), pages 727-752, November.

Rosenthal, S. S. & Strange, William C., (2004). "Evidence on the nature and sources of agglomeration economies," Handbook of Regional and Urban Economics, in: J. V. Henderson & J. F. Thisse (ed.), Handbook of Regional and Urban Economics, edition 1, volume 4, chapter 49, pages 2119-2171 Elsevier

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, Volume 41, Part 2, pp. 173–188

Vickerman R., Spiekermann K, and Wegener M. (1999). "Accessibility and Economic Development in Europe", Regional Studies 3(1) 1-15

Vickerman R., (2007). "Recent Evolution of Research into the Wider Economic Benefits of Transport Infrastructure Investments," OECD/ITF Joint Transport Research Centre Discussion Papers 2007/9, OECD, International Transport Forum.

Appendix A: Data

As discussed in the text, our source of firm data is the Annual Respondents Database (ARD). The Interdepartmental Business Register/Business Structure Database (IDBR/BSD) is used to construct measures of employment accessibility. Addresses of businesses in the IDBR/BSD are compiled using a combination of tax and VAT records. The IDBR has been operating since 1994 but service sector firms are only included from 1997 onwards. Measures of skills are taken from the Labour Force Survey (LFS). The source for transport networks is described in the text.

Annual Respondents Database

Value added and capital stock need to be deflated and to do this we use sectoral value added and inputs deflators from EU KLEMS "Growth and Productivity Accounts: March 2008 Release". Capital is measured with the perpetual inventory method. Labour is number of employees.

The most disaggregated level at which the ARD provides information on value added is the Reporting Unit level. We refer to Reporting Units as firms. Reporting Units can operate with one or more Local Units. We refer to these Local Units as plants and refer to firms with only one plant as singletons; firms with more than one plant as multi-plant firms. For each plant, we know the sector of activity, employment and exact location (post code).

Reporting Unit balance sheet information comes from the Annual Business Inquiry where each year only a fraction of the active Reporting Units is sampled. The first two columns of the table below (A.1) report for the years 1997-2004 the number of unique firms in the population and the number of unique firms sampled. The third reports the number of unique reporting units that can be used to estimate the production function. We need the firm to have in that year non missing value added, employment and capital. Then we need to have information on Local Units, from which the firm location is derived.

Table A1: Unique Reporting Units in the ARD

	Pop.	Sampled	Estimation sample
Total	2,993,514	273,831	182,794

Appendix B: Additional transport and skills control variables

Skills variables are derived from the Labour Force Survey, aggregated to Local Authority District level according to place of survey respondents residence. We use the proportion of the working age population with NVQ Level 4+, qualifications (which includes HNDs, First Degrees, Higher Degrees and similar qualifications). To obtain skills measures that vary by postcode-sector level, we use a similar procedure to that described above for urbanisation. The effective skilled share at Level 4+ for a given postcode sector *i* in a given year is calculated by averaging the contemporaneous skilled share in nearby Local Authorities, using an inverse-distance weighting sequence. This sequence applies a weight of $d_{ij}^{-1}\sum_{j} d_{ij}^{-1}$ to the skilled share in Local Authority *j*, where d_{ij} is the distance between the centroids of postcode sector *i* and Local Authority *j*. Note that these weights sum to 1. Postcode sector *i* is therefore assigned a weighted *average* of the skilled share in neighbouring Local Authorities, with nearest LAs receiving higher weights than those further away. The equation for the Level 4+ skilled share is thus: $S_i = \sum_{j\in D}\left(L4share_j \times d_{ij}^{-1} \times \sum_{j\in D} d_{ij}^{-1} \right)$ where *D* is the set of Local Authorities within 20km of postcode sector *i*. A plant (local unit) in the ARD is assigned skilled shares for the postcode sector in which it is located. Again, Appendix B details the treatment of multi-plant firms.

Some regression specifications include straight line distances to transport infrastructure nodes. The set of transport infrastructure nodes were derived from various sources: postcodes of passenger airports with international links obtained from http://www.ukairportguide.co.uk/; full access motorway junctions extracted from Ordnance Survey Strategic mapping data; rail station postcodes provided by the Department of Transport; sea port postcodes from UK Major Ports Group (http://www.ukmajorports.org.uk/). All these transport nodes were converted to point features in GIS software (ArcGIS), and straight line distances computed from each GB postcode sector to the nearest of each type of transport node feature (airport, Motorway junction, rail station, port). The figure below shows the locations of these transport nodes for each type of transport. These transport variables clearly best capture the ease of access to the transport links between rather than within locations.