Simulating the impact of transport infrastructure investment on wages: a dynamic spatial panel model approach

Bernard Fingleton* and Nikodem Szumilo+

*Department of Land Economy, University of Cambridge, United Kingdom
+Department of Geography and Environment, London School of Economics

Abstract

This paper estimates the impact of a multi-billion pound investment in Britain's rail transport infrastructure, in the form of high-speed rail links, on wage levels across each of 347 districts of England and Wales. The impacts are based on a dynamic spatial panel model adaptation of standard urban economics based on employment density and commuting patterns. This allows estimation of these global impacts operating via improved commuting times. We demonstrate that while estimates of a traditional market potential approach with fixed effects are to some extent qualitatively and quantitatively similar to our predictions, our predictions allow more heterogeneous effects and give more accurate forecasts. The study finds that on average wages increase by around 2% as employment centres gain improved access to more skilled workers and as spillover effects become spatially more extensive. While most areas see modest positive effects, some locations are negatively affected, in the extreme case by as much as 7%.

Keywords: dynamic spatial panel model; simulation; prediction; transport infrastructure; commuting; wages; regional development.

JEL classification: C33; C53; J31; R11; R12; R4.

---

*Corresponding author. Bernard Fingleton. *E-mail:* bf100@cam.ac.uk

1 Introduction

As transportation plays an important role in determining the spatial distribution and productivity of economic activity, transportation innovations are a popular policy instrument for stimulating overall economic growth and reducing economic disparities across regions (Redding and Turner 2015). This is true for developed and developing countries alike as increasing spatial differences in economic performance are a universal concern. The impact of intra- and cross-region transportation improvements is a popular topic of research but most studies focus on particular investment projects and specific regions which makes their results difficult to generalize. There is also a tendency for new infrastructure to be targeted at connecting places that are already growing so identifying the contribution of transportation infrastructure is difficult. This paper develops a spatial model of wages that allows for spillovers of transportation innovations between locations and quantifies the impact of improving travel times. By focusing on improvements in travel times rather than on direct access to new infrastructure we can address the selection bias of locations that receive transportation investment. We allow for travel innovations to affect locations that do not receive new transportation infrastructure through improvements in indirect travel times and through spatial interactions of wages within a commutable area. We estimate the model using a dynamic spatial panel estimator as suggested by Baltagi et al. (2014, 2018) which allows us to simulate the spatial distribution of the impact of changes in travel times induced by transportation investment. In an attempt to avoid ad hoc assumptions and purely data-driven decisions, the model specification is based on a well-established urban economics theory and includes temporal and spatial lags which have a priori rationale as well as empirical support. This differentiates our approach from simpler methods used in the literature and we demonstrate the advantage of using our model by contrasting it with a simpler model. Our choice of a spatial moving average error process attempts to mitigate possible deleterious effects for estimation of lagged exogenous variables, and in our instrumentation strategy we seek to eliminate bias due to possible endogeneity. Likewise we opt for minimalist assumptions regarding the extent of spatial effects, thus allowing free choice within a commutable zone, rather than impose rigid assumptions embodied within one or other specific distance decay function. This is in the spirit of Simini et al(2012). An important feature of our forecasting approach is the focus on finding a dynamically stable, stationary prediction.

Unlike traditional approaches, we acknowledge that without satisfying the maximum absolute real eigenvalue criterion there is no long-run elasticity, and apply this restriction to our forecasts.

The specific context we study is the impact of transportation innovations on wages according to urban and regional economic models. The problem of inter-regional studies of wages is that most existing urban economic models are based on US cities in which commuting across urban areas has not been a prominent feature. In many other countries, however, commuting is a key determinant of local productivity (Fingleton 2003), which makes transportation innovations an important issue for local economic growth. Thus our model is built on the standard microfoundations but it allows for commuting between regions so that firms compete for the most productive workers while workers can choose between locations that make them most productive (and pay the highest wages). For example, one key issue of transportation innovation is the spatial distribution of the impact of connecting a large, dense and prosperous area with a less developed location (Chandra and Thompson 2000). The model used in this study allows interactions between wages and location-fixed effects to change with travel time which controls for the impact of being connected to a location with specific (observed or unobserved) characteristics. Indeed, using a realized project we compare our forecasts with estimates of a traditional market potential approach with fixed effects and show that while they are to some extent qualitatively and quantitatively similar, our model offers more heterogeneous effects and a better fit of actual post-improvement data.

As transportation technologies are undergoing large-scale changes, the question of their impact on the spatial distribution of economic activity remains open. The mainstream economic research focuses on the impact high speed rail has on access to the surrounding economic mass and shows that reducing transportation costs to new locations has positive economic consequences (Redding and Strum 2008). Although there is a large body of economic literature that considers the impact of transportation costs on trade (Duranton et al 2014, Redding and Turner 2015), relatively little is known about the impact of the movement of people on wages. While in a recent study Ahlfeldt and Feddersen (2018) consider the impact of high speed rail on agglomeration economies of both directly connected locations and their neighbours, we contribute to this literature by showing that the impact of high speed rail is not limited to locations where stations are located and their neighbours, but affects all districts. We also have a more narrow focus

and concentrate on the impact on wages rather than GDP. Concentrating on wages allows us to measure the benefits of high speed rail improvements attributable to gaining access to new workers. We develop a theoretical model of how wages change due to improvements in inter-district rail connections and link this effect directly to commuting. Small et al. (2007) estimate that the commuting time is around 7.5% of an average working day in most developed countries and Schafer (2000) estimates the average commuting time across 26 countries to be 73min (with a standard deviation of 12min). Aggregated, these numbers add up to a time cost of commuting of between 2.4 and 4.8% of GDP (although other methodologies suggest that this is a conservative estimate (Redding and Turner, 2015). It is, however, unclear how transportation technologies influence the impact of commuting on wages. For example workers with higher pay have longer commutes and it is unclear if this is because more productive workers are willing to travel more or because the most productive locations require the longest commutes (Schafer 2000). Critically, there is a positive correlation between the level of development of transportation infrastructure and average commuting distances (Redding and Turner 2015). This suggests that transportation innovations do not affect the time workers are willing to spend commuting but, as they are able to travel further within the same time, they affect the spatial distribution of where people live and work. These however are notoriously difficult to capture in traditional empirical models based on cross-sectional data as those choices and transportation improvements are highly endogenous. Researchers attempt to address this problem in different ways. For example, by instrumenting current transportation links with historical networks (Duranton and Turner 2012) or by using locations where a development was planned but abandoned as control groups (Baum-Snow 2007).

Our primary contribution is to overcome some of the empirical challenges by applying novel spatial econometric methods. The dynamic spatial panel estimator proposed by Baltagi et al. (2018) allows us to control for endogeneity in our spatial model. This includes all location sorting decisions of firms and workers who react to changes in wages and house prices within a commutable distance of their workplace. It also accounts for decisions of firms to change wages in response to competition in the labour market and changes in local productivity. The model assumes that workers can choose locations of work and residence freely as long as they are within a commutable distance from each other. We define this distance using travel time and show that the optimal upper limit is 85min. The impact of transportation innovations is

captured by allowing new locations to be added to the commutable area. All unobservable factors are treated as spatially dependent random effects which allows us to capture location-fixed effects and local shocks. We estimate the model on Local Authority Districts in England and Wales using annual data from 2008 to 2017. Our forecast of changes in travel times is based on a£56bn[1] rail improvement project planned by the UK government. In a one-step-ahead prediction we show that our results are accurate in predicting the spatial distribution of wages. We also test the model using historic improvements to England's rail network and show that wages forecasted by the model have a 0.9 correlation with actual wages. This shows that the model can be used not only to forecast the impact of planned developments but also to assess the impact of already completed projects.

Our results show that the planned improvement to the railway network will in the long run increase wages by an average of around 2%. However, the benefit is very uneven across space. Locations that will see the largest number of other areas enter into their 85min travel time radius will see the biggest benefit. Conversely, districts that are already relatively well-connected with the rest of the country do not see their wages changed. We also report that the majority of the increase in wages will not occur in locations that will have direct access to the new network as those are already relatively well-connected. Instead, their neighbours will be able to benefit from decreased indirect travel times.

The reminder of the paper is structured as follows. Section 2 develops the model. Section 3 outlines our data. Section 4 presents a preliminary analysis of our sample. Section 5 describes the estimator and section 6 presents estimation results. Our prediction methodology is described in section 7 and simulation outcomes are presented in section 8. Section 9 concludes by presenting final remarks

2 The Model

The model specification is based on the urban economics model outlined in the Appendix, but we extend this in two ways. First, we introduce time

[1]This is the official total cost for HS2, as set out in the Government's 2015 spending review. The actual cost may be different.

as indicated by the presence of t in equation (1). Second, we introduce the concept of labour efficiency units in place of the number of workers E given in the model (26). Thus N_t in equation (1) is the number of efficiency units at time t, which we define as the number of workers (E_t) per unit area multiplied by the level of technology they employ (A_t) at time t. Accordingly

$$\ln w_t = \ln \varphi + (\gamma - 1) \ln N_t \tag{1}$$

in which $\ln w_t$ is an n by 1 vector the log of the wage rate at time t, in which n denotes the number of districts, $N_t = E_t A_t, \varphi$ is a constant and γ a parameter with value dependent on the presence or absence of increasing returns. So if $\gamma > 1$, wages are increasing in the density of labour efficiency units, due to positive externalities associated with high density. If $\gamma < 1$, then negative externalities such as the effects of 'congestion', implicit in the labour efficiency units per unit area metric, offset positive externalities so that the net effect is diminishing returns.

Consequently,

$$\ln w_t = \ln \varphi + (\gamma - 1) \ln A_t + (\gamma - 1) \ln E_t \tag{2}$$

Also assume that

$$\ln A_t = \ln b_0 + b_1 \ln H_t \tag{3}$$

Equation(3) indicates that $\ln A_{it}$ depends on the skills of workers in and around district i, H_{it}. We use the fact that each location has a unique network of social and economic connections with its commutable area which is reflected in the local and regional intensity and direction of commuting flows as well as in the spatial distribution of employment and population density. Since workers are perfectly mobile within a given commuting radius, they choose where to work and live purely by what maximizes their utility. As a consequence, household location and commuting choices are driven by arbitrage in local labour and housing markets. We therefore assume that the n by 1 vector $\ln A_t$ depends on the number of people within commuting distance that have higher education, H_t. In order to calculate this number, for each district i with h_{it} residents with higher education we calculate the total residents with higher education living within 85 minutes[2] commute time of district i, including those living within district i, thus $H_t = D_N h_t$ where D_N is an n by n matrix of 1s and zeros. In this, 1 indicates that a pair of

[2]This commute time is subsequently supported by empirical analysis.

districts are within 85 minutes commuting time from each other, and zero indicates that they are not (including 1s on the leading diagonal).

The effect of employment density in district i on wages in district i reflects externalities, emanating from the variety of services and the level of congestion[3].

Combining this with equations(2) and (3) gives

$$\ln w_t = \ln(\varphi b_0) + (\gamma - 1) \ln E_t + (\gamma - 1) b_1 \ln H_t \tag{4}$$

However equation(4) ignores other important factors which also likely to influence $\ln w_t$,notably the effects of $\ln w_{t-1}, W_N \ln w_{t-1}$ and $W_N \ln w_t$, in other words the effects of wages in the previous period, both within district and within commuting distance, and contemporaneous effect of wage levels within commuting distance. These across-district interactions are governed by the matrix[4] W_N.

We assume that W_N is based the n by n matrix D_N of 1s and zeros, but with zeros on the leading diagonal. The resulting matrix is subsequently row normalized so that districts within 85 minutes of district i each have the same weight in the weighted average, to give W_N. We maintain the usual assumptions for a weights matrix, that it comprises fixed (non-stochastic) values and its row and column sums are uniformly bounded in absolute value, and maintain the same assumption for $B_N^{-1} = (I_N - \rho W_N)^{-1}$ (Elhorst, 2014, p. 99) in which I_N is an n by n identity matrix. Given estimation of parameter ρ in $B_N = (I_N - \rho W_N)$ (see equation (11)), a standard restriction found in the literature is that $e_{\min}^{-1} < \rho < e_{\max}^{-1}$ where e is the vector of purely real characteristic roots of W_N, and because of row-normalization, $e_{\max} = 1$,

[3]Note that we explicitly separate the impact of local density from the skills of locally employed workers. In this way our definition of density is different from elsewhere in the literature and needs to be interpreted in this context.

[4]Typically, one could moderate the spatial weighting matrix so that the commuting 'distance' around Manchester was less than London, but while distance may vary, perhaps because of infrastructure differences, it seems more likely that the upper bound of commuting time will be approximately the same. This could reflect physical and social limitations on regular commuting longer than 85 minutes each way, regardless of distance, for most people. Notwithstanding, what is important is the outcome, wages and how they evolve in the future. We have in the model individual effects $\mu_i, i = 1, ..., n$, which are unobserved individual (district) specific effects also affecting wages. These capture inter-district heterogeneity, for example a proclivity to commute further in one region rather another. These are spatially dependent according to an SMA process, so that groups of nearby districts will have similar values.

which avoids the singularity and ensures the inversion of B_N. However, the conditions needed to avoid non-stationarity are more complex in the dynamic spatial panel context, as discussed in more detail below.

Typically throughout much of the spatial econometrics literature, spatial and temporal dependence involving the spatial lags $W_N \ln w_t$ and $W_N \ln w_{t-1}$ and the temporal lag $\ln w_{t-1}$ are ad hoc assumptions justified by empirical evidence. However, following the same arguments given by Baltagi et al. (2018), here we justify their inclusion by arguing that they are the logical outcome of assuming that, in theory, the distribution of wages tends towards an equilibrium. In other words, the contrast between local wages in district i and wages in its commutable area tends not to change over time because rigidities in the labour market cause unchanging demand for and supply of labour in each district unless disturbed by some outside force, such as infrastructral investment in transport systems causing an extension of the feasible commuting zone. We express this theoretical equilibrium as the difference between log wage and the average log wage within commutable distance remaining constant, so that

$$\ln w_t - W_N \ln w_t = \ln w_{t-1} - W_N \ln w_{t-1} \tag{5}$$

obviously holds with long run equilibrium such that $\ln w_t = \ln w_{t-1}$.However typically with real data we do not observe the long-run equilibrium, but a tendency for data to move towards equilibrium, we assume according to an autoregressive process

$$\ln w_t = \tau \ln w_{t-1} + \varsigma \tag{6}$$

where scalar $|\tau| < 1$.The assumption underpinning the long-run equilibrium is that if the process is undisturbed in the long run the vector of log wages reaches the equilibrium $\ln w_T = \frac{\varsigma}{1-\tau}$.Without loss of generality we introduce a scalar parameter ρ,and multiplying (6) by ρW_N, it follows that

$$\rho W_N \ln w_t = \tau \rho W_N \ln w_{t-1} + \rho W_N \varsigma$$

Subtracting $\rho W_N \ln w_t$ from $\ln w_t$ gives

$$(I_N - \rho W_N) \ln w_t = (\tau I_N - \rho \tau W_N) \ln w_{t-1} + (I_N - \rho W_N) \varsigma$$

hence

$$\ln w_t = B_N^{-1}\left(C_N \ln w_{t-1} + B_N \varsigma\right) \tag{7}$$

in which in which $C_N = (\tau I_N + \theta W_N)$ and $\theta = -\rho\tau$.Rearranging, and introducing other time-varying effects so that $B_N\varsigma_t = \pi_t + \varepsilon_t$ in which π_t represents the other observable effects identified in equation (4) and ε_t denotes unobservable effects, gives the more empirically sustainable expression

$$\ln w_t = \tau \ln w_{t-1} + \theta W_N \ln w_{t-1} + \rho W_N \ln w_t + \pi_t + \varepsilon_t \tag{8}$$

Expanding this by substituting equation(4) plus $\ln \overline{w}_t$ into (8) gives

$$\begin{aligned}
\ln w_t &= \tau \ln w_{t-1} + \theta W_N \ln w_{t-1} + \rho W_N \ln w_t + & (9)\\
&\quad \ln(\varphi b_0) + (\gamma - 1)\ln E_t + (\gamma - 1)\, b_1 \ln H_t + b_2 \ln \overline{w}_t + \varepsilon_t \\
\ln w_t &= (I_N - \rho W_N)^{-1}\left(\tau \ln w_{t-1} + \theta W_N \ln w_{t-1} + ...\right) & (10)\\
\ln w_t &= B_N^{-1}\left[\begin{array}{c} C_N \ln w_{t-1} + \ln(\varphi b_0) + (\gamma - 1)\ln E_t \\ +\,(\gamma - 1)\, b_1 \ln H_t + b_2 \ln \overline{w}_t + \varepsilon_t \end{array}\right] & (11)
\end{aligned}$$

The additional variable $\ln \overline{w}_t$ denotes the log of district-invariant mean wages, which captures the common effect of wage shocks across all districts, which is similar to the approach involving interest rates and share prices adopted in Fingleton, Fuerst and Szumilo (2018). Typically without controlling for the presence of common factors, the parameter ρ will tend to be biased upwards. In order to solve equation (11), and ensure that the system tends towards an equilibrium rather than explosive outcome, constraints are imposed on ρ, τ and θ. Hence, following Elhorst (2001,2014, p. 98), Parent and LeSage (2011, p. 478, 2012, p. 731) and Debarsy, Ertur and LeSage (2012, p. 162), we require that the largest characteristic root of $B_N^{-1}C_N$ is less than 1. This restriction maintains stationarity, dynamic stability and ensures convergence in wages to equilibrium levels[5].

The unobservables ε_t are treated as spatially dependent random effects so that, for simplicity of exposition, equation (11) becomes equation(12), with

[5]Figure 8 is an example of the dynamic evolution of wage levels given that the restriction is satisfied. Holding constant variables and parameters in prediction equation (21), the individual path for wages in each district is shown to converge to its own steady-state level. Figure 9 shows a typical outcome when the restriction is not satisfied. In this case future levels do not reach a steady state equilibrium, although the paths take by each district can be complex, depending on the estimated parameters.

x_{0t} an n by 1 vector of ones, $x_{1t} = \ln E_t$, $x_{2t} = \ln H_t$, $x_{3t} = \ln \overline{w}_t, x_t = [x_{0t}\, x_{1t}\ x_{2t}\ x_{3t}]$ and $\beta = [\ \beta_0\ \beta_1\ \beta_2\ \beta_3]$,hence

$$\ln w_t = B_N^{-1}\left[C_N \ln w_{t-1} + x_t\beta + \varepsilon_t\right] \tag{12}$$

in which

$$\varepsilon_t = u_t - \lambda M_N u_t = G_N u_t \tag{13}$$

where $G_N = (I_N - \lambda M_N)$,and $u_t = \mu + \nu_t$.Thus we have a compound spatial moving average (SMA) error process, building on Fingleton(2008), in which $\mu \sim iid(0, \sigma_\mu^2)$ is an unobservable time-invariant effect which accounts for heterogeneity across districts, and $\nu_t \sim iid(0, \sigma_\nu^2)$ is a remainder term. The essential characteristic of a moving average error process is that the unobservable elements have local rather than global interdependence, unlike an autoregressive process.

As explained by Fingleton, Le Gallo and Pirotte(2017), one justification for an SMA error specification is that 'it mitigates against the problem for instrumental variable estimation identified by Pace et al. (2012)'. In (cross-sectional) estimation involving two-stage least squares (2SLS), Kelejian and Prucha (1998, 1999) recommend that the instruments should comprise the 'exogenous' variables (X) and their spatial lags ($W_N X$), keeping the instrument set to a low order to avoid linear dependence and to retain full column rank for the matrix of instruments. In practice, as we explain in Section 6, we treat the variables X as predetermined, thus requiring temporal lags of the instruments X and $W_N X$ to satisfy moments conditions. Nonetheless, by including $W_N X$ among the set of explanatory variables, one could jeopardize the performance of the estimation procedure, as explained by Pace et al. (2012). Given that the spatial lags ($W_N X$) are among the set of regressors, then this requires the use of ($W_N^2 X, W_N^3 X$, . . .) as instruments, and these additional instruments could lead to a weak instrument problem. To avoid this, we adopt SMA errors as an alternative way to capture local spillovers which would otherwise be captured by incorporating spatial lags of the explanatory variables. In equation (13), we assume that $M_N = W_N$, restrict λ to its stationary bounds $e_{\min}^{-1} < \lambda < e_{\max}^{-1}$, and assume that the elements of x are uniformly bounded in absolute value.

3 Data

Our data come mainly from databases provided by the UK Office for National Statistics (ONS). Education data comes from the Annual Population Survey which is based on approximately 320,000 respondents and is the largest survey of households in the UK. In our dataset, to reflect education levels, the variable H_t is measured by the number of residents with NVQ level 4 or above qualifications. As defined by the ONS these usually are Degree and Higher Degree level qualifications or equivalent. This data set is based on the place of residence and persons are regarded as residents if they have lived at their current address for more than 6 months. Students are recorded as living with their parents. Employment data ($\widetilde{E}_t$) comes from the same source but is based on the place of work and full time male and female employees only, leading to E_t which is employment per hectare (ha) across all sectors.

Wages w_t are gross mean weekly wages in £ sterling for full-time male and female employees, whose earnings for the survey period were not affected by absence. The data is taken from the Annual Survey of Hours and Earnings which is based on a 1% sample of employee jobs provided by the HM Revenue and Customs (HMRC) for the Pay As You Earn (PAYE) tax system. These are adjusted for inflation using the Consumer Price Inflation time series dataset (2015 = 100) available from the UK's Office of National Statistics.

Table 1 : Descriptive statistics for key variables, $t = 2016$

	H_t		$\widetilde{E}_t$	ha	E_t	w_t
mean	39497.7		47302.6	43515	9.36328	591.6
median	29000		34000	23349.2	1.52613	573.4
st.dev.	31932.3		48252.5	62396.6	70.3935	87.63
min	5500		5000	289.78	0.0501895	446.4
max	230400		479000	518037	1280.28	1160.6

Note : H_t = number of residents with NVQ level 4 or above qualifications; $\widetilde{E}_t$ = employment level; ha = hectares; E_t = employment per hectare; w_t = gross mean weekly wages in £ sterling for full-time male and female employees.

Data for the estimation sample are collected for all 347 (n) Local Authority Districts in England and Wales (Isles of Scilly are excluded due to data limitations) on an annual basis starting in 2008 which is the earliest year for which data for these geographies is available. Thus data cover the period 2008 to 2017. Travel times and distances between areas are calculated for population centroids provided by the ONS. We use the quickest of drive times and public transportation as an indication of travel times between ar-

eas, calculated using Google maps. In our simulations we use the forecast of travel times after transportation improvements (HS2) given by the UK government's projections of direct improvements in travel times. If an indirect route between any two locations becomes faster than a direct connection the shorter of the two is used in our calculations. Historic travel times used for measuring the impact of realised improvements of the rail network (i.e. HS1) are approximated since this information is not available from Google maps and fastest indirect travel times cannot be calculated easily from current travel times. The approximation is based on simulating the impact the realised improvement in direct travel times would have on current connections (if it was implemented again) and slowing the affected routes down by the time they would gain from this hypothetical development.

In this study we show how the impact of transportation innovations can be estimated using the example of the HS2 (High Speed 2), which is a program of railway improvements planned by the UK government. The project is expected to cost £ 56bn and take 16 years to complete. The development is based on linking 4 major cities (see Figure 2) with a high-speed rail connections with additional cities being linked to the network over existing tracks. In our projections we use the forecast of travel times provided by the UK government (HS2 2013). This data lists changes in travel times between stations linked to HS2 directly and stations that will be connected via existing railways. We also allow for indirect travel and if an indirect route between any two locations becomes faster than a direct connection, the shorter of the two is used in our projections.

For the evaluation of a realised transportation project we use HS1 which is a high speed railway connection between London and the Channel Tunnel with three intermediate stations (see Figure 2). The project cost £5.8bn and has been completed in November 2007. Data used to apply our forecasting approach comes from the same sources as for the estimation sample although several districts changed their geographical boundaries in 2008 which makes data before this date inappropriate for spatial estimations. Note that the change has little effect on our forecasts of the impact of HS1 as there were no significant changes in the region affected by it.

Importantly, in our model improvements in travel times do not affect commuting patterns unless the change means that a destination that used to be out of reach becomes accessible. In 2016 the National Travel Survey reported the average commuting distance to be 10.9 miles and take 30 minutes. However, there is significant regional variation between commuting

times across different regions as London workers choose the longest commutes (42min), while the average for the South West is 25min. Although no data on standard deviation of these estimates is available, other reports show that only around 25% of commuters spent more than 30min traveling to work (McQuaid and Chen 2012) and around 4% reported commuting for longer than 60min (Chatzitheochari and Arber 2009). According to the 2011 census the 85min limit captures 95.14% of cross-district commuting population. This is shown by Figure 1, which gives the cumulative percentage of total cross-district commuting population captured by different (between districts) travel time limits. As shown in Figure 1 the 85min limit captures the majority of commuters but very few additional workers are captured by increasing this threshold. The marginal share of population captured by extending the threshold by an additional 5min from 70min and 80min is 0.73% and 0.71% respectively. Increasing the limit by 5min from 85min or 90min adds only 0.32% or 0.42% of extra commuters. This shows that the 85min limit is a point at which the marginal share of commuters captured by extending the travel limit falls substantially. Later we show that in our data the optimal assumption of the maximum commuting time is 85min. Therefore, we consider the benefit of transportation improvements to be measured by the number of districts that become accessible within 85min.

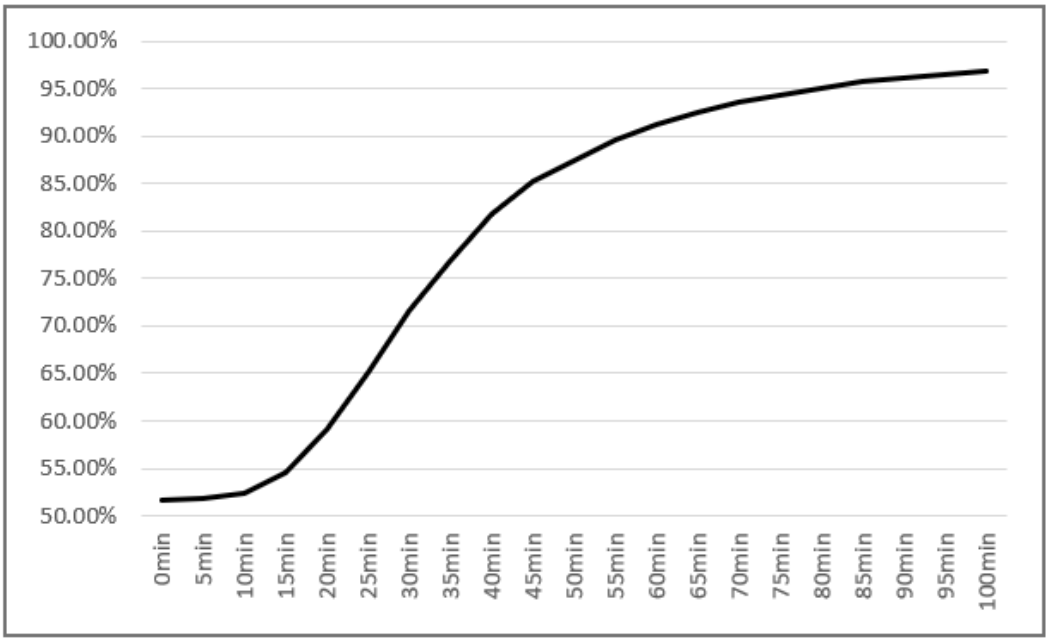

Figure 1 : Cumulative percentage of total cross-district commuting population by travel time.

Note : 0 to 10 min commute time identifies the approximate percentage of the population that works within their district of residence.

Figure 2 gives, for each district, the additional number of districts within commuting distance (85min travel time) in England and Wales after HS1 (Figure 2*a*) and HS2 (Figure 2*b*) are completed. As we show subsequently, the HS1 effect is relatively localised in the South and East of England. HS2 will impact some cities directly (for example Birmingham and London) but will have the biggest impact on travel times at locations that will not be linked to the network directly. This means that indirect train travel from remote areas will become much faster. This is consistent with the government's assumption that the key advantage of HS2 over the existing network will be its ability to connect into the existing railway connections. This effect is most noticeable for the North of England. HS2 will result in large nominal time savings for travellers in that region as their current journeys to other districts are very long.

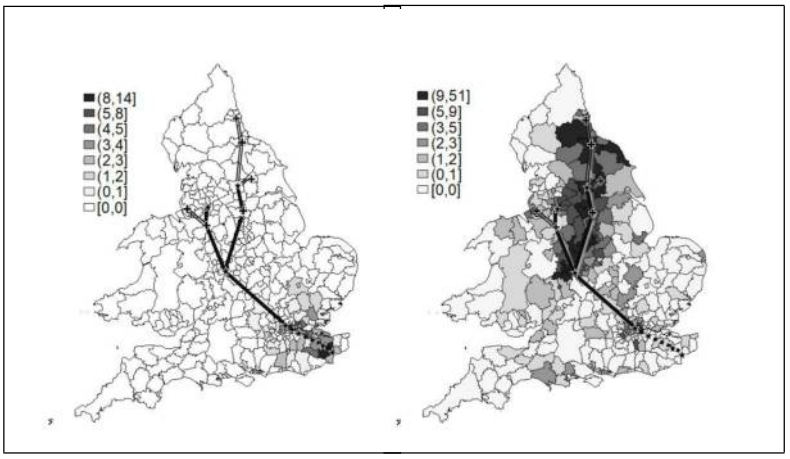

Figure 2a: Additional districts within commuting distance - HS1

Figure 2b : Additional districts within commuting distance - HS2

Figure 2: Additional districts within commuting distance after new infrastructure is developed

Note : The thick black line represents HS2 railway lines and the double lines represent existing track which will be linked into the HS2 network. The dashed line is the route of HS1. The stars represent the 4 target cities on the new lines of HS2 (or intermediate stations of HS1) and the crosses represent cities that will be operating trains that use the HS2 lines without changeovers.

4 Preliminary analysis

Initial evidence suggests that increasing employment density is related to wage levels, as indicated by Figure 3. In this we do not imply causation, because employment may be attracted to locations with high wages, as well as high wages being a consequence of externalities due to the density of employment. Notice that the implication in Figure 3 is that the relationship is positive, with the cross-sectional OLS regression of $\ln w_t$ on $\ln E_t$ giving an elasticity equal to 0.0465. The model used in Figure 3 fails to control for the other factors suggested by equation (11), but overall Figure 3 suggests that there are increasing returns to scale, wages do increase as the scale of the local economy increases. It turns out however that the interpretation of

the impact of employment density on wages is more nuanced than suggested by this preliminary analysis. As explained by Combes and Gobillon (2015), an endogeneity problem might occur at local level when the density variable is correlated to the error term either due to reverse causality or an omitted variable. When higher wages attract more workers, the reverse causality effect would cause the OLS estimate to be positively biased. Omitted variable bias can be a result of a factor that influences both wages and density such as local amenities. Amenities that increase productivity and attract new workers simultaneously (such as transport infrastructure) also introduces a positive bias into the OLS estimate of the impact of density while consumption amenities have the opposite effect. It is also possible that there is a bias introduced by worker heterogeneity and their preference for certain locations or types of work. Finally, there may be a bias resulting from the interaction of area heterogeneity with worker heterogeneity when the same worker would receive a different wage in different locations due to a factor that is correlated with density (Combes et al. 2011).

The relevance of location for wages is apparent from Figure 4. As the mean commuting time from a district to other districts increases, wages fall. We attempt to capture this via W_N as in equation(11).

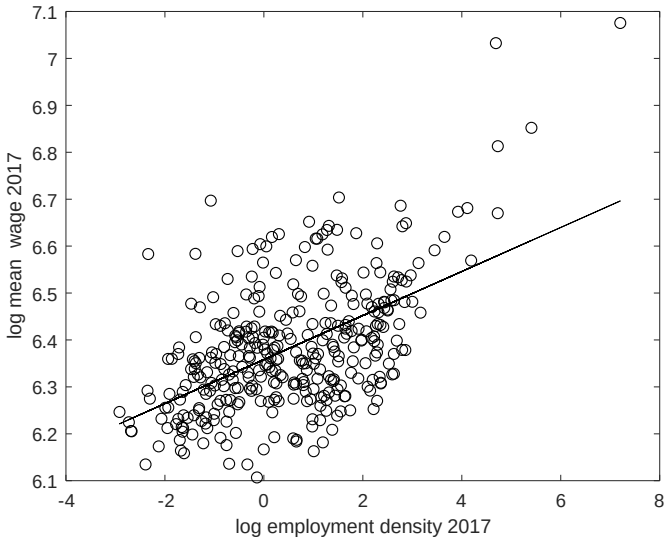

Figure 3 : Employment density versus wages

Note: $\ln \widehat{w}_t$ =6.3582+0.0469 $\ln E_t$, slope $t - ratio = 11.94$. $R^2 = 0.2925$

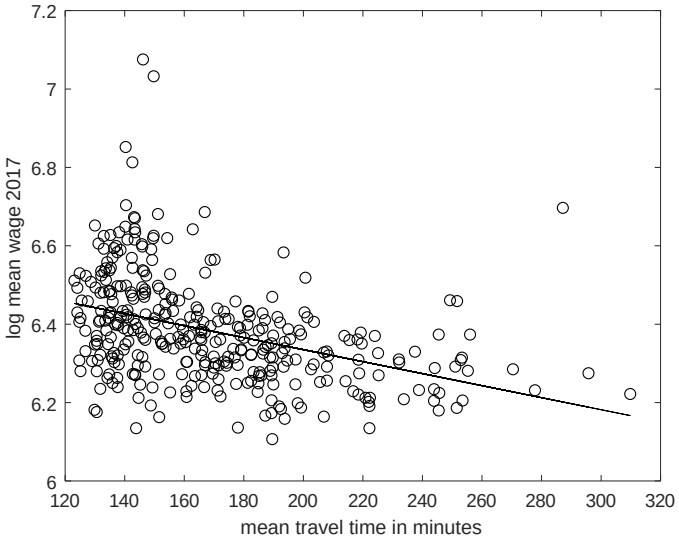

Figure 4 : Mean travel times (T) versus wages

Note : $\ln \widehat{w}_t = 6.6413 - 0.0015T$, slope $t - ratio = -7.98$. $R^2 = 0.1560$

4.1 Simpler rivals to the preferred specification

Instinctively one would hope that the model specification is as simple as possible. In this section we present an (over)simplified version of a standard approach to predicting the impact of transportation innovations. The first step is to estimate the impact of being 'treated' with new infrastructure on wages in the treated district and use the estimated value to predict the effect of future treatments. We therefore begin by estimating the impact of HS1 on wages in districts it affected. There is a large body of literature that focuses on estimating the impact of new transportation connections and it offers a variety of different methodologies attempting to recover the impact of being treated with new transport connections on different outcomes. Gibbons and Machin (2005) employ a simple difference in difference methodology in which having access to new infrastructure is a binominal dummy variable. Ahlfeldt and Feddersen (2017) offer a market potential approach in which the impact of new infrastructure is measured by the change it induces in the

market potential variable. However, few papers examine wages explicitly and while it may be tempting to apply methodologies used for other outcomes to examine wages, this strategy has several shortcomings.

To demonstrate this point, we follow a very basic approach to estimating treatment effects and add an additional variable to equation (4) to reflect its exposure to HS1. For each district, the variable reflects the number of new districts that enter its 85min commuting radius due to travel time improvements. It equals zero for years in which HS1 was not operating. This can be interpreted as a treatment variable with a measure of intensity or as a measure of changes in market potential (measured in the number of districts within the commuting distance). To focus on the immediate effect of the improvement, we limit our sample to the years 2006, 2007 and 2008, that is just one year after HS1 opens. This is partially due to the fact that we expect that the effect on the treated will vary over time (Ahlfeldt and Feddersen 2017) and partially because we expect that over time the directly affected districts (treated) will interact with their neighbours (non-treated) which means that the treatment effect spills over space and is difficult to estimate over time (Combes et al. 2008). In a fixed-effects panel estimation (with both district and year effects) clustering standard errors at district level[6] we find that the average impact of adding a new district on wages is an increase of 0.27% (standard deviation is 0.11%). Figure 5a demonstrates the spatial impact of HS1 estimated using this approach. Using this simple treatment effect parameter also facilitates our first attempt at predicting the impact of HS2, by multiplying it by the number of districts expected to be newly connected. In Figure 5b we see that the effect of HS1 was to increase wages by up to 3.10%, with the impact concentrated in the South East region of England[7]. The average across all affected districts is just 0.69%. The predicted impact of HS2 is more widespread and locally much more intense with the average impact on the treated districts of 1.78% but the highest increase of 19.5%[8].

[6]Meaning that they are robust to cross-sectional heteroskedasticity and within-district serial correlation.

[7]More precisely we work with the South East region (minus the New Forest and Isle of Wight districts) plus Cambridgeshire (minus Peterborough) and the county of Suffolk. Overall that gives 137 districts.

[8]Since HS1 affects a far smaller number of districts than HS2 the two frequency distributions 5b and 6b have different scales.

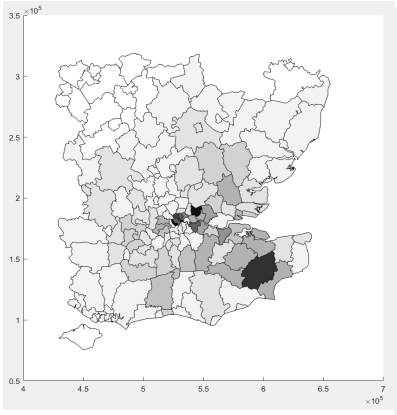

Figure 5a : The estimated impact of HS1

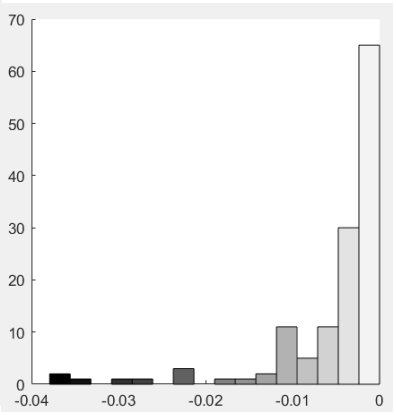

Figure 5b : Key to Figure 5a.

Figure 5 : The estimated impact of HS1 on wages using the simple model

Note : 5a maps the estimated % wage shortfall in each district had HS1 not occurred. Darker shading indicates bigger HS1 effect as demonstrated in 5b which represents the same data as a histogram.

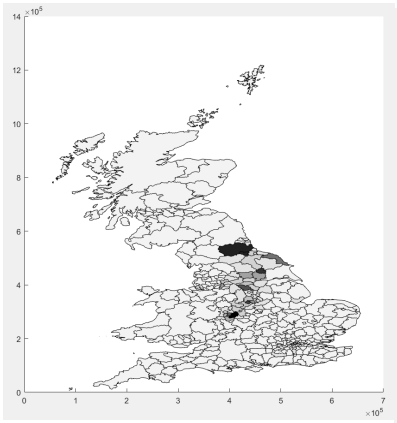

Figure 6a : The estimated impact of HS2

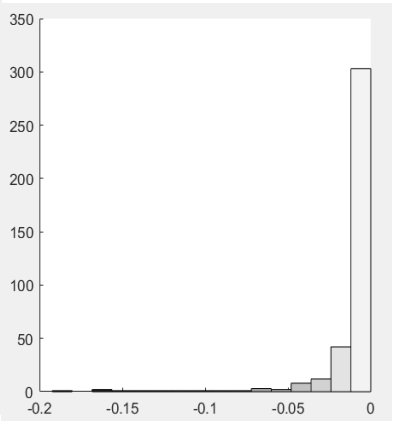

Figure 6b : Key to Figure 6a.

Figure 6 : The estimated impact of HS2 on wages using the simple model

Note : 6a maps the estimated % wage shortfall in each district without HS2. Darker shading indicates bigger HS2 effect as demonstrated in 6b which represents

the same data as a histogram.

Clearly, while the actual impact of HS1 and the predicted impact of HS2 are plausible, this estimation approach is amongst the simplest possible and falls short of the preferred estimator of Section 5. There a numerous alternative methods of estimating the treatment effect, but these too have limitations. Two common variations would be to use another definition of being 'treated' or chose a different measure of market potential. For example, one might replace the number of new districts with the percentage change in the number of accessible districts or (in the context of our theoretical model) with the number of new workers with higher education in those new districts. This choice is usually motivated by the theoretical model advocated in the specific context and although it clearly affects the obtained magnitude of the treatment effect and its functional form, we do not focus on this aspect as a major limitation. Instead, we argue that most similar measures of treatment effects will have several common shortcomings. First, the treatment effect is only measured in directly affected districts. Since neighbouring regions are economically dependent on each other, an increase in wages in one will affect the other (Combes et al. 2008). Regardless of which market potential change variables are used, the standard treatment effect approach ignores this relationship and provides a biased estimate. Our preferred estimator allows wages of neighbouring regions to interact so our results will be a combination of the treatment effect with the spatial spill over effect. The second shortcoming of using a simple treatment effects approach is that it only allows positive or negative treatment effects while it is possible that new transportation infrastructure could have a positive impact on some (for example on the directly connected districts) but a negative impact on others (for example on the neighbours of the directly connected districts). Our preferred estimator allows for positive and negative as well as direct and indirect effects of transport innovations. Note that the spatial spill over is especially problematic when the neighbours are used as the control. This leads to the last issue we highlight in this section; endogeneity of being treated. Since transportation innovations usually target specific locations, it is difficult to argue that the treatment is orthogonal to the outcome and most papers recognize that this is a considerable limitation. A natural solution to this problem would be to turn to instrumental variables but it is very difficult to find reliable instruments. We avoid the tricky problem of the appropriate choice of instruments in our preferred dynamic specification and

estimator, which is fully consistent with the dynamic theory and in which internal instruments are applied, in the spirit of Arellano and Bond (1991).

5 The Preferred Estimator

Pesaran(2015, Chapters 29 and 30) and Baltagi(2013, Chapter 13) provide comprehensive overviews of spatial panel econometrics highlighting the growing importance of this methodology for the applied econometrician. Our estimator, introduced by Baltagi et al(2018), which corresponds exactly to the specification of equations (12) and (13), is a new addition to the available methodology. The precursor to the estimator, as detailed in Baltagi et al. (2014), extends Arellano and Bond (1991) by introducing additional moments to take full account of the spatial dimension of the model (see also Bouayad-Agha and Védrine, 2010). However Baltagi et. al.(2018) introduce the spatial lag of the temporal lag of the dependent variable $W_N \ln w_{t-1}$ as an additional regressor, and allow moving average rather than spatial autoregressive compound error dependence. Because it is fully described in Baltagi et al(2018), we simply set out the broad structure of the method. The general (static) IV based methodology, which we can refer to as the GM spatial IV estimator, as described by Kelejian and Prucha(1999) and Kapoor et al(2007) and also discussed in a moving average panel data context in Fingleton(2008), has some computational advantages compared with ML estimation. These advantages relate to the robustness of the approach to distributional assumptions, compared with explicit distributional assumptions required under ML (although these issues may be resolved somewhat under quasi-ML), and to less strict assumptions regarding initial conditions (Bond, 2002, Hsiao,1986). Also the proposed approach is computationally feasible even with very large data sets. Additionally a major advantage is the way in which the IV approach can cater for endogeneity in its implicit form, which is not possible under ML. Thus we can include endogenous or predetermined right hand side variables additional to those involving the endogenous left hand side variable, namely $\ln w_{t-1}, W_N \ln w_{t-1}$and $W_N \ln w_t$.

In summary, the approach involves a four-stage parameter estimation procedure. Stage one is an initial step in which instrumental variables are used to obtain initial estimates of $\tau, \rho, \theta,\ \beta_1, \beta_2$,and β_3, from differenced data to eliminate the correlation between the individual effects μ and the

lagged dependent variable, which are then used to give estimated errors. The second stage involves using these estimated errors to obtain estimates of the parameters of the spatial moving average error process, namely λ, σ_μ^2 and σ_ν^2 using moments equations given in Fingleton(2008). The third stage involves computing preliminary one-stage consistent spatial GM estimates. The fourth introduces a robust version of the variance-covariance matrix to give the two-stage Spatial GM estimates of τ, ρ, θ and the βs.

Typically in many applications the regressors (x_1, x_2, x_3) are assumed to be exogenous. These then become part of the instrument set facilitating consistent estimation. Among the instrument set one also typically includes the dependent variable lagged by two periods, and we also include its spatial lag lagged by two periods. Assuming ν_{it} is serially uncorrelated and hence $E(\Delta\nu_{it}, \Delta\nu_{it-2}) = 0$, the moments equations will be satisfied. Arellano and Bond(1991) give a test ($m_2 = cov(\Delta\nu_{it}, \Delta\nu_{it-2})/s.e.$) for this assumption but evidently m_2 is only defined for series of at least length five and inference rests on a proof of asymptotic normality. Here we utilize the somewhat complex equations given by Arellano and Bond(1991) (see also the Stata manual), which are omitted to save space, to test for serial correlation. The outcome, based on differenced errors following stage one of the estimation procedure, is $m_2 = 0.0526$, which is not significant when referred to the $N(0,1)$ distribution, with upper tail probability equal to 0.3985. The method clearly indicate that there is no evidence of second order serial correlation[9].

Consequently, following Baltagi et al(2007), with E here denoting the expectation, we have

$$\begin{aligned} E\left(\ln w_{il}\Delta\nu_{it}\right) &= 0 \quad \forall i, l = 1, 2, ..., T-2; t = 3, 4, ...T \\ E(\sum_{i\neq j} w_{ij} \ln w_{il}\Delta\nu_{it}) &= 0 \quad \forall i, l = 1, 2, ..., T-2; t = 3, 4, ...T \end{aligned}$$

leading to matrices of instruments

$$Z_t = \begin{pmatrix} \ln w_1, ..., \ln w_{t-2}, W_N \ln w_1, ..., W_N \ln w_{t-2}, x_{11}, ..., x_{1T}, \\ x_{21}, ..., x_{2T}, ..., W_N x_{11}, ..., W_N x_{1T}, W_N x_{21}, ..., W_N x_{2T}, ... \end{pmatrix}$$

[9] Also we find that, using a randomization approach to obtain the empirical null distribution of $cov(\Delta\nu_{it}, \Delta\nu_{it-2})$,the observed covariance 0.0526 is exceeded by 38.4% of 1000 randomly generated covariances. Plotting $\Delta\nu_{it}$ against $\Delta\nu_{it-2}$ gives a random scatter with no obvious trend, supporting the inference of no second order serial correlation.

for $t = 3, ..., T$.The important point here is that since the regressors are exogenous, they are completely independent of the differenced errors and so at any time t the entire set from 1 to T can be included among the set of instruments and the moments equations remain satisfied.

Note however that exogeneity may be unrealistic, since 'strict exogeneity rules out any feedback from current or past shocks to current values of the variable, which is often not a natural restriction in the context of economic models relating several jointly determined variables'(Bond 2002). This is the case in this application, in which we wish to also take account of the possibility that employment density and worker skill both cause and are also responses to wage differences. One way to allow this is to treat all variables symmetrically, by also lagging the regressors so that the instrument set for individual i at time t becomes

$$Z_t = \begin{pmatrix} \ln w_1, ..., \ln w_{t-2,} W_N \ln w_1, ..., W_N \ln w_{t-2}, x_{11}, ..., x_{1t-2}, \\ x_{21}, ..., x_{2t-2}, ..., W_N x_{11}, ..., W_N x_{1t-2}, W_N x_{21}, ..., W_N x_{2t-2}, ... \end{pmatrix}$$

However in the context of our study it seems unreasonable to assume contemporaneous feedback from wages to the regressors. Instead we prefer a time lag, and so our preferred estimator assumes that the regressors are predetermined (see Bond, 2002, Pesaran, 2015). Predetermined regressors are contemporaneously uncorrelated, so that $corr(x_t, \nu_t) = 0$, but do depend on earlier shocks so that, for example, $corr(x_t, \nu_{t-1}) \neq 0$.This means that an adjustment to $\ln w$,which embodies ν, at time t does not have an instantaneous effect on employment density and worker skill at time t but takes effect at $t + 1$ and later. This allows an extension to the set of instruments compared with assuming endogeneity, enabling the inclusion of $x_{1t-1}, x_{2t-1}, x_{3t-1}, W_N x_{1t-1}, W_N x_{2t-1}$ and $W_N x_{3t-1}$ so that

$$Z_t = \begin{pmatrix} \ln w_1, ..., \ln w_{t-2,} W_N \ln w_1, ..., W_N \ln w_{t-2}, \\ x_{11}, ...x_{1t-2}, x_{1t-1}, x_{21}, ...x_{2t-2}, x_{2t-1}, ..., \\ W_N x_{11}, ...W_N x_{1t-2}, W_N x_{1t-1}, \\ W_N x_{21}, ...W_N x_{2t-2}, W_N x_{2t-1}, ... \end{pmatrix} \tag{14}$$

The Sargan-Hansen test of over-identifying restrictions gives a test statistic equal to 132.84. Evidently when referred to χ^2_{198},this indicates that the moments conditions implied by our dynamic spatial panel model with predetermined regressors are valid. However caution should be exercised,

partly because testing can be weakened by the presence of many moments conditions (Bowsher, 2002, Pesaran, 2015), resulting in low power.

6 Estimates

In Table 2 we give the estimated parameters of our estimator with instrument set given by equation(14) for $t = 3, ..., T$ and W_N based on a maximum commute time of 85 minutes. In Section 6.2 we give the outcome of one-step ahead prediction for different maximum commute times, which support the choice of 85 minutes.

Table 2: Estimates of equation (12)

parameter	parameter estimates	st. error	t ratio
τ	0.4650	0.03329	13.97
ρ	0.4798	0.1038	4.624
$\beta_1 = (\gamma - 1)$	-0.1331	0.02078	-6.408
β_2	0.04808	0.02807	1.713
β_3	0.5556	0.1121	4.957
θ	-0.4942	0.08354	-5.916
λ	-0.90683		
σ_μ^2	0.05179		
σ_ν^2	0.0018075		

Note : τ parameter for $\ln w_{t-1}$ = temporally lagged log wage; ρ parameter for $W_N \ln w_t$ = spatially lagged log wage; $\beta_1 = (\gamma - 1)$ parameter for $\ln E_t$ = log employment density; β_2 parameter for $\ln H_t$ = log qualified residents; $\beta_3$parameter for $\ln \overline{w}_t$ = log mean wage; θ parameter for $W_N \ln w_{t-1}$ = spatially temporally lagged log wage; λ = parameter for spatial dependence in moving average error process; σ_μ^2 = variance of unobserved time-invariant effect; σ_ν^2 =variance of unobserved remainder.

The Table 2 estimates are stationary and dynamically stable, as shown by the largest characteristic root of $A_N = B_N^{-1} C_N$ which is equal to 0.5582, and the stationary bounds for λ are $\widetilde{e}_{\min}^{-1} = -2.4318 < \lambda < \widetilde{e}_{\max}^{-1} = 1$,where given the SMA error process defined by equation(13), the negative value of $\widehat{\lambda}$ implies positive spatial dependence among the errors. Accordingly $G_N = (I_N - \lambda M_N)$ is non-singular with rank equal to size and condition number not excessively large. Thus we can invert G_N as required in equation (20).

In contrast to the parameter estimates $\widehat{\beta}_1$ and $\widehat{\beta}_2$ given in Table 2, the true derivatives, following Le Sage and Pace(2009) and Elhorst(2014, p.106) among others, are, in the short run for $\ln E_t$, equal to

$$\left[\begin{array}{c} \frac{\partial \ln w_{1t}}{\partial \ln E_{1t}} \cdots \frac{\partial \ln w_{1t}}{\partial \ln E_{nt}} \\ \cdots \\ \frac{\partial \ln w_{nt}}{\partial \ln E_{1t}} \cdots \frac{\partial \ln w_{nt}}{\partial \ln E_{nt}} \end{array} \right]_t = (B_N)^{-1} (\beta_1 I_N) \tag{15}$$

and the long run derivatives for $\ln E_t$ are

$$\left[\begin{array}{c} \frac{\partial \ln w_{1t}}{\partial \ln E_{1t}} \cdots \frac{\partial \ln w_{1t}}{\partial \ln E_{nt}} \\ \cdots \\ \frac{\partial \ln w_{nt}}{\partial \ln E_{1t}} \cdots \frac{\partial \ln w_{nt}}{\partial \ln E_{nt}} \end{array} \right] = [-C_N + B_N]^{-1} (\beta_1 I_N) \tag{16}$$

Similarly, the corresponding short run derivatives for $\ln H_t$ are

$$\left[\frac{\partial \ln w_t}{\partial \ln H_{1t}} \cdots \frac{\partial \ln w_t}{\partial \ln H_{nt}} \right]_t = (B_N)^{-1} (\beta_2 I_N) \tag{17}$$

and the long run derivatives for $\ln H_t$ are

$$\left[\frac{\partial \ln w_t}{\partial \ln H_{1t}} \cdots \frac{\partial \ln w_t}{\partial \ln H_{nt}} \right] = [-C_N + B_N]^{-1} (\beta_2 I_N) \tag{18}$$

These n by n matrices of derivatives provide measures of the direct, indirect and total effects of a unit increase in $\ln E$ (respectively $\ln H$). Thus, the direct short run effect of a unit increase at time t is, following LeSage and Pace(2009), equal to the mean of the leading diagonal of equation (15). The total effect is equal to the mean column sum of (15) and the indirect effect is equal to the total effect minus the direct effect. Likewise the long run effects of a persistent increase in $\ln E_t$ as t goes to infinity are based on the identical means of the n by n matrix (16). Also, the direct, indirect and total long and short run effects of a unit increase in $\ln H$ are given by the same calculations starting from equations (17) and (18).

Table 3 : Elasticities based on the Table 2 estimates					
		direct		indirect	total
E_t(employment density)	short run	-0.1344		-0.1215	-0.2559
E_t	long run	-0.2489		0.0065	-0.2423
H_t (qualified residents)	short run	0.0486		0.0439	0.0924
H_t	long run	0.0899		-0.0024	0.0875

Table 3 suggests the dominance of localised negative externalities (congestion etc.) over positive local externalities, with regard to the effect of E_t on wages, leading to a total short run elasticity of -0.2559. This is for a point in time increase in employment density in the local district. If the increase in employment density persists ad infinitum, then in the long run the total elasticity with respect to wages will converge to an equilibrium of -0.2423, implying that a 1% increase in density causes output to increase by only 0.76%, but as we show subsequently in Section 7.1, this amounts to a long run prediction based on assumptions about future values of variables and coefficients that may not hold in the long run. Without denying the probable existence of positive externalities relative to local density, it is evident that our data has thrown up countervailing negative density effects within districts, which we attribute to congestion, defined broadly as the limitations on economic actors working in a confined space, and the net effect of both is a negative elasticity. But while we estimate that the return to our measure of density is negative, we have a narrow definition of this variable that focuses on local effects. It also excludes any labour productivity effects captured by number of qualified people within a given district. Looked at more broadly, positive externalities with respect to labour come to the fore, for it is undeniable that mass of economic activity outside the local district is also important for wage levels within, and local economic activity *per se* is an insufficient basis for understanding the impact of economic density or mass. As local density increases, so does predicted wage. This comes through from our analysis in which we identify a positive effect due to the total number of qualified people (viz. workers) within commuting distance, denoted by H_t. Table 3 shows that the total elasticities with respect to H_t are positive.

7 Prediction

Following Chamberlain (1984), Sevestre and Trognon (1996), and Baltagi et. al.(2014, 2018), the linear predictor is

$$E\left[\ln w_t\right] = B_N^{-1}\left[C_N E\left[\ln w_{t-1}\right] + x_t\beta + G_N E\left[u_t\right]\right] \tag{19}$$

in which $E\left[.\right]$ denotes the expectation, so this can be seen to be identical to equation (12) but with expectations. With regard to the estimate of the time-invariant component of the error term μ , assuming a spatial moving average error process we commence with equation (12) rewritten thus

$$\begin{aligned} \varepsilon_t &= B_N \ln w_t - C_N \ln w_{t-1} - x_t\beta \\ G_N u_t &= B_N \ln w_t - C_N \ln w_{t-1} - x_t\beta \\ u_t &= \mu + \nu_t \\ \mu^{(t)} &= G_N^{-1}\left(B_N \ln w_t - C_N \ln w_{t-1} - x_t\beta\right) - \nu_t \end{aligned} \tag{20}$$

In order to obtain estimates $\widehat{\mu}^{(t)}$ we use estimates $\widehat{G}_N, \widehat{B}_N, \widehat{C}_N$ and $\widehat{\beta}$ and draw at random from $\nu_t \sim N(0, \widehat{\sigma}_\nu^2)$. We then take the mean over time of the $\widehat{\mu}^{(t)}s$ for $t = 2, ..., T$, subsequently scaling so that it has variance equal to $\widehat{\sigma}_\mu^2$,thus giving the estimate $\widehat{\mu}$ of the time-invariant error component. The outcome is the prediction equation

$$\ln \widehat{w}_t = \widehat{B}_N^{-1}\left[\widehat{C}_N \ln \widehat{w}_{t-1} + x_t\widehat{\beta} + \widehat{G}_N\widehat{\mu}\right] \tag{21}$$

7.1 total derivatives

It turns out that we can apply equation (21) to obtain the total short and long run derivatives, thus aiding our appreciation of what these mean. For the total short run effect relating to $\ln E_t = x_{1t}$, we calculate the difference between $\ln \widehat{w}_{ta}$ and $\ln \widehat{w}_{tb}$, where log wages increase from $\ln \widehat{w}_{tb}$ to $\ln \widehat{w}_{ta}$ across all n districts as a result of an increment Δx_{1t} in x_{1t} across all N districts, thus

$$\ln \widehat{w}_{ta} = \widehat{B}_N^{-1}\left[\widehat{C}_N \ln \widehat{w}_{t-1,a} + x_{0t}\,\widehat{\beta}_0 + (x_{1t}\ + \Delta x_{1t}\,)\widehat{\beta}_1 + x_{2t}\,\widehat{\beta}_2 + x_{3t}\,\widehat{\beta}_3 + \widehat{G}_N\widehat{\mu}\right] \tag{22}$$

and

$$\ln \widehat{w}_{tb} = \widehat{B}_N^{-1} \left[\widehat{C}_N \ln \widehat{w}_{t-1,b} + x_{0t} \widehat{\beta}_0 + x_{1t} \widehat{\beta}_1 + x_{2t} \widehat{\beta}_2 + x_{3t} \widehat{\beta}_3 + \widehat{G}_N \widehat{\mu} \right] \quad (23)$$

in which $\Delta x_{1t} = 1, \widehat{B}_N = (I_N - \widehat{\rho} W_N), \widehat{C}_N = (\widehat{\tau} + \widehat{\theta} W_N)$ and $\widehat{G}_N = \left(I_N - \widehat{\lambda} M_N \right)$.It follows that the derivative is the (average) change on log wages divided by the change in $\ln E_t = \Delta x_{1t} = 1$,hence using the Table 2 estimates,

$$\frac{\partial \ln \widehat{w}}{\partial x_1} = \frac{\sum_{i=1}^{N} (\ln \widehat{w}_{ita} - \ln \widehat{w}_{itb})}{n} = -0.2559 \quad (24)$$

which is what we have in Table 3. Likewise, in the long run[10], $\Delta x_{1t} = 1, t = 1, ..., T$ where T is a large number, in our case $T = 50$. For the long run derivatives, we iterate equations (22) and (23) over $t = 1, ..., T$, holding x_{1t} , x_{2t} and x_{3t} constant at their $t = 1$ (2016) levels as t goes to $T(= 50)$,and it is possible to show that

$$\frac{\partial \ln \widehat{w}}{\partial x_1} = \frac{\sum_{i=1}^{N} (\ln \widehat{w}_{iTa} - \ln \widehat{w}_{iTb})}{n} = -0.2423 \quad (25)$$

which again replicates the log run Table 2 estimate. Similar equations produce the same outcomes for the total derivatives relating to x_{2t} .

7.2 one-step ahead predictions

In order to support our preference for the model summarised by Table 2, a cross-validation strategy is employed to assess the performance of the competing estimators 'by comparing their predictive ability on data which have not been used in model estimation' (Anselin, 1988). We measure the predictive ability of our rival models via the root mean square error ($RMSE$), given by

[10]The assumption is that parameters and variables stay at the same level in the future, otherwise the equilibrium outcome for wages will go to a new equilibrium commensurate with the new levels.

$$RMSE = \sqrt{\sum_{i=1}^{n} \left(\ln w_{it+1} - \ln \widehat{w}_{it+1}\right)^2 / n}$$

Additionally, we employ mean absolute error (MAE), defined as

$$MAE = \sum_{i=1}^{n} \left|\ln w_{it+1} - \ln \widehat{w}_{it+1}\right| / n$$

which gives less weight to large forecast errors than RMSE, and Theil's U which is

$$U = \sqrt{\sum_{i=1}^{n} \left(\ln w_{it+1} - \ln \widehat{w}_{it+1}\right)^2 / \sum_{i=1}^{n} (\ln w_{it+1})^2}$$

In the latter, the numerator is essentially the $RMSE$, and the denominator is a way of making U independent of the units of measurement.

In order to predict one step ahead, and thus obtain $\ln \widehat{w}_{t+1}$, we apply equation (21) with data[11] for the year 2017 but use the parameter estimates based on the period up to 2016 thus giving a clear separation between data used for the estimates and data predicted. Table 4 gives the resulting forecast accuracy measures for different estimators based on different maximum commute times which then define different matrices W_N and M_N .

[11] Since the only 2017 data we have is for wages and E, the 2017 levels of H were obtained by projecting forward the 2016 levels using the annual growth rate over 2009 to 2016.

Table 4 : Forecast Accuracy and Theoretical Consistency	
commute < 50 min	
$RMSE, MAE, U, \theta_G$	$e^A_{\max}$
$0.096, 0.070, 0.015, -0.263$	0.800
commute < 60 min	
$RMSE, MAE, U, \theta_G$	$e^A_{\max}$
$0.051, 0.040, 0.008, -0.614$	1.173
commute < 70 min	
$RMSE, MAE, U, \theta_G$	$e^A_{\max}$
$0.052, 0.041, 0.008, 0.123$	4.965
commute < 80 min	
$RMSE, MAE, U, \theta_G$	$e^A_{\max}$
$0.047, 0.037, 0.007, -0.514$	0.687
commute < 85 min	
$RMSE, MAE, U, \theta_G$	$e^A_{\max}$
$0.044, 0.034, 0.007, -0.271$	0.558
commute < 90 min	
$RMSE, MAE, U, \theta_G$	$e^A_{\max}$
$0.050, 0.040, 0.008, -0.340$	0.516

Note : $RMSE$ = root mean square error; MAE = mean absolute error ; U = Theil's U; θ_G = theory consistency measure; $e^A_{\max}$ =maximum absolute real eigenvalueof $A_N = B_N^{-1}C_N$.

In changing maximum commute times hence W_N, the matrices $\widehat{G}_N$, $\widehat{B}_N$,and $\widehat{C}_N$ change, and the instruments in equation(14) also alter, so it is to be expected that the estimates obtained, and consequently the predictive ability of the model, can fluctuate somewhat. In Table 4 we report the $RMSE, MAE$ and U statistics and also the maximum absolute real eigenvalue of $A_N = B_N^{-1}C_N$, which we denote by $e^A_{\max}$. If $e^A_{\max}$ is equal to 1 or greater, then that indicates that the model estimates are dynamically unstable. This is evident from the 'long run total derivates' given by equation(25) and its counterpart for x_2, which taken on values very different from the result given by equations(16) and (18). In other words there is no fixed elasticity relating to our causal variables so it impossible to show the effect of changing density on wages. On the other hand, with $e^A_{\max} < 1$,equation(25) and equations(16) and (18) give identical outcomes. An additional consideration is the equality

$\theta = -\rho\tau$ deriving from the assumed equilibrium process given by equation(7). While our estimator does not permit strict satisfaction of this equality, we prefer estimates that approach this as closely as possible. We represent the closeness of this by $\theta_G = \theta + \rho\tau$ in Table 4. On the basis of $RMSE$, MAE, U and θ_G, a maximum commute time of 85 minutes, produces accurate one-step ahead predictions combined with the most appropriate parameter estimates leading to log-run stability. The predictive accuracy of our model improves up to a commute time of 85 minutes, which accounts for more than 95% of all commuters (see Figure 1). Commuting up to 90 minutes includes a few more extreme commuters but reduces the predictive performance of the model. Figure 7 shows the actual versus predicted log wages for one-step ahead assuming a commute times of up to 85 minutes. The predicted log wage up to 50 steps ahead, obtained holding employment density and worker skill constant at their 2016 levels, gives Figure 8, showing the paths of the 347 districts. By way of contrast, Figure 9 shows the lack of dynamic stability associated with assuming a maximum commute time of 60 minutes. Other commute times for which $e^A_{\max} > 1$ show similar unrealistic long term outcomes.

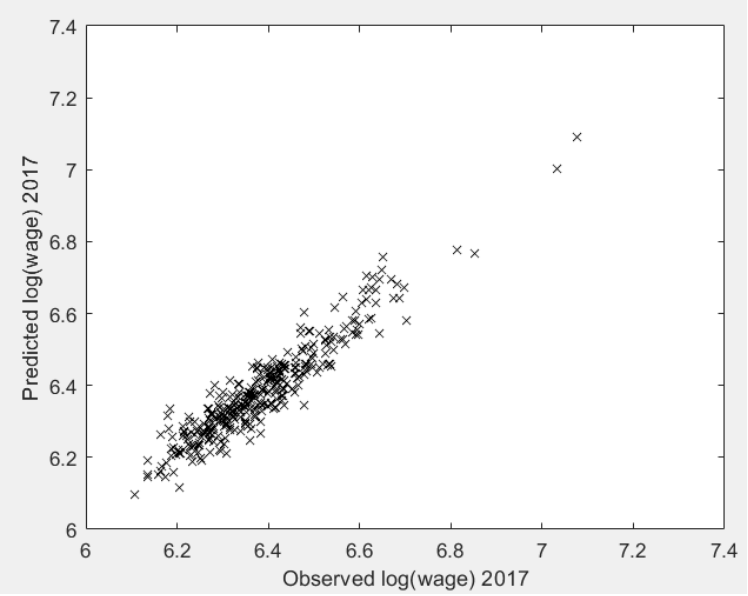

Figure 7 : Predictions from Table 2 estimates

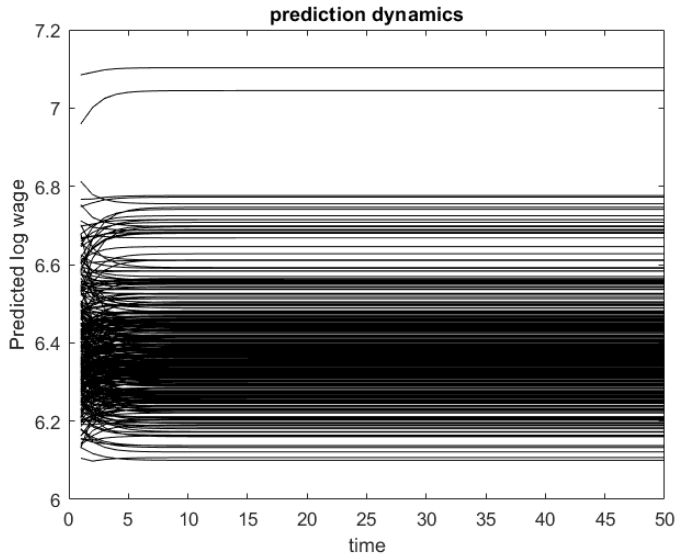

Figure 8 : Dynamics assuming maximum commute time is 85 minutes

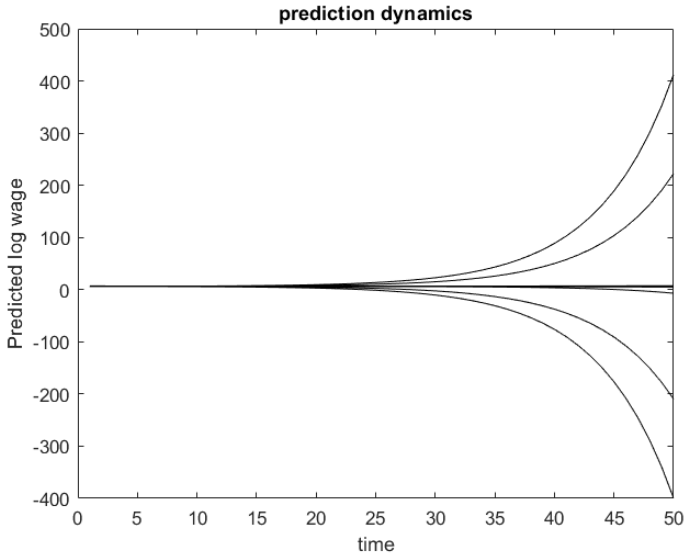

Figure 9 : Dynamics assuming maximum commute time is 60 minutes

8 The impact of HS1

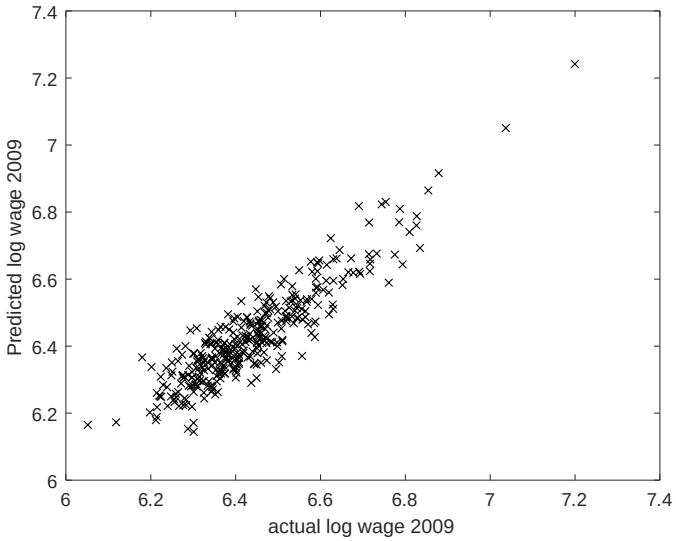

Figure 10 : Actual and predicted log wage 2009

In order to show that the HS1 effect predicted by the preferred model is more realistic than the prediction under the simpler model, we consider the difference in predicted wage before and after HS1 under both models. Figure 11 shows that despite significant heterogeneity in how actual wages changed between 2007 and 2009 predictions of the preferred model are positively correlated with (both positive and negative) actual changes. This is not the case for the simple model which is negatively related to the overall change in actual wage (as is evident from the slope of the best fit line and the equation given in Figure 12). The negative correlation is not surprising as the simple model predicts that the change in most areas will be zero and Figure 13 shows that after adjusting for this fact, the simple model performs better. When predictions of the two models are compared (Figure 14) it is clear that the main differences are due to the simple model predicting no negative effects and only effects in the directly affected areas. From Figure 14 it is clear that the lower predictive power of the simple model comes mainly from these drawbacks. This speaks to the forecasting ability of the dynamic panel

model and shows that it can produce reliable forecasts at least of comparable accuracy as simple estimates of treatment effects using actual data.

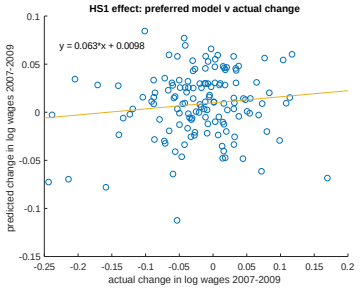

Figure 11 : % change in wages 2007 to 2009 in each district in South East England versus predicted % change due to HS1 given by the preferred model

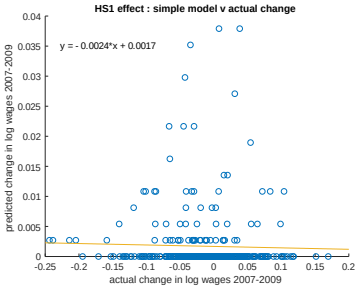

Figure 12 : % change in wages 2007 to 2009 versus predicted % change due to HS1 given by the simple model, including districts with zero predicted change.

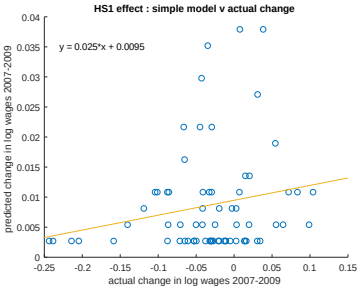

Figure 13 : % change in wages 2007 to 2009 versus predicted % change due to HS1 given by the simple model, having eliminated districts with zero predicted change

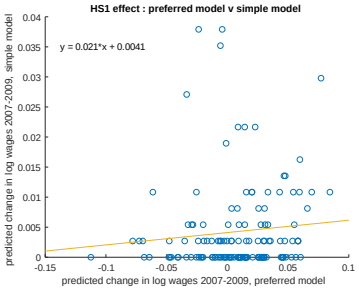

Figure 14 : Predicted % change in wages 2007 to 2009 due to HS1 under preferred model versus predicted % change according to simple model

Although the predictions of the two models are correlated, the preferred model offers significantly more heterogeneity in the predicted effect. This includes not only allowing for negative effects but also for spatial spillovers. Therefore to map HS1 effects estimated using the preferred model we assume a maximum travel time of 85 minutes, leading to $\widetilde{W}_N$ $\left(\widetilde{M}_N\right)$ consistent with pre-HS1 travel times. As[12] with W_N,we again assume that $\widetilde{W}_N$ is based an n by n matrix of 1s and zeros $\widetilde{D}_N$. In this, 1 indicates that a pair of districts are within 85 minutes commuting time from each other, and zero indicates that they are not within 85 minutes travel time of each other. The matrix, that is $\widetilde{D}_N$ but with zeros on the leading diagonal, is subsequently row normalized to give $\widetilde{W}_N$ with non-zero values in row i indicating districts within 85 minutes of district i, with each non-zero cell of row i containing the same value. The weights matrix $\widetilde{W}_N$ leads to $\widetilde{G}_N, \widetilde{B}_N$,and $\widetilde{C}_N$, where $\widetilde{B}_N = \left(I_N - \widehat{\rho}\widetilde{W}_N\right), \widetilde{C}_N = (\widehat{\tau}I_N + \widehat{\theta}\widetilde{W}_N)$ and $\widetilde{G}_N = \left(I_N - \widehat{\lambda}\widetilde{M}_N\right)$, which togther with data for 2007 are used in the pre-HS1 prediction equation(21). Post-HS1 (2009) predicted wages are obtained using post-HS1 travel times giving $\widehat{G}_N, \widehat{B}_N$,and $\widehat{C}_N$ combined with 2009 data, with parameter estimates again taken from Table 2. Figure 15 shows the differences between before and after HS1 predicted wage levels in England and Wales. To demonstrate how estimates of the two models differ, in Figure 16 we present the difference between the effects estimated using the simple model (Figure 5) and the preferred model (Figure 15). Although, there is no clear spatial pattern to the differences, the two models disagree on which areas benefited from HS1. Indeed, the spatial pattern of changes predicted by the preferred model is a better fit for actual changes in wages (presented in Figure 17) than estimates of the simple model.

[12] We could envisage our approach within the context of convex combinations of different types of connectivity structures (Debarsy and LeSage, 2018) with both matrices matrices W_N and $\widetilde{W}_N$ embodied simultaneously within the model specification, but with weights 1 and zero according to which of the two matrices was relevant. However, trying to apply this is beyond the scope of the current work.

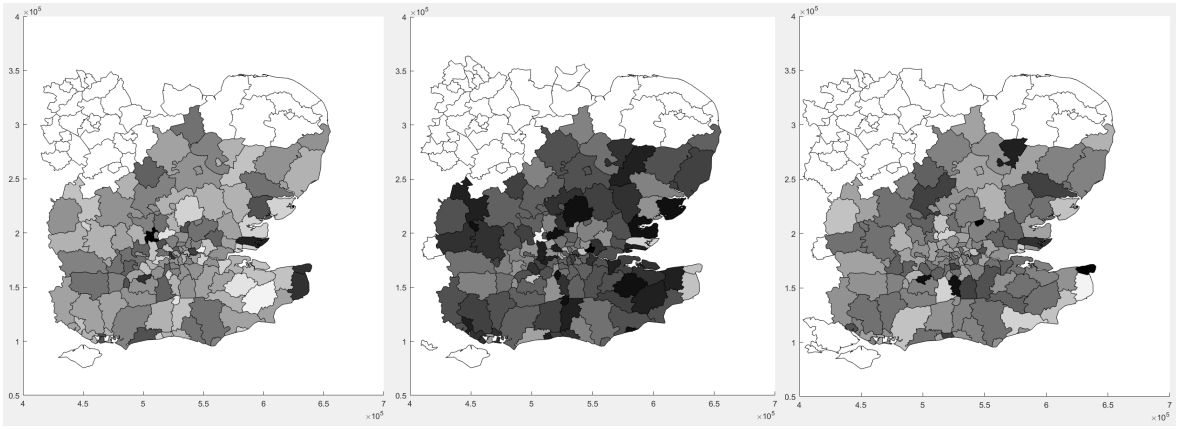

Figure 15a : Estimated impact of HS1 on wages in South East England in 2009 according to preferred model.

Figure 16a : Difference between predicted HS1 effect under the simple model and predicted HS1 % wage increases under the preferred model.

Figure 17a : % change in wages over the period 2007-2009 in districts in the South East Region of England plus the county of Suffolk and Southern Cambridgeshire.

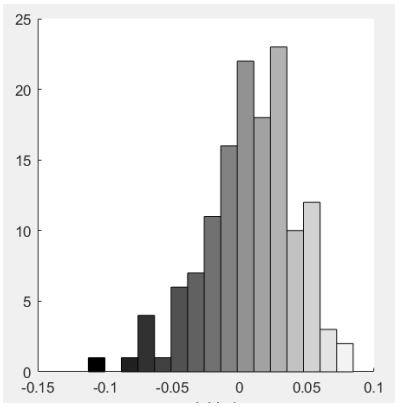

Figure 15b : Key to Figure 15a. Estimated impact of HS1 according to preferred model versus number of districts.

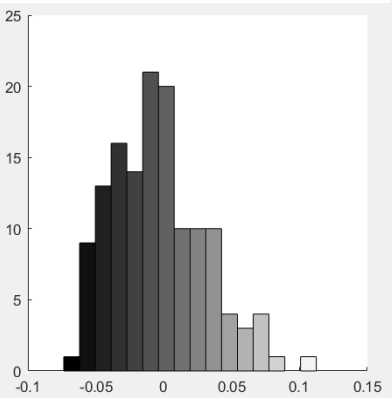

Figure 16b : Key to Figure 16a. Difference in predicted HS1 effect according to simple and preferred models, versus number of districts

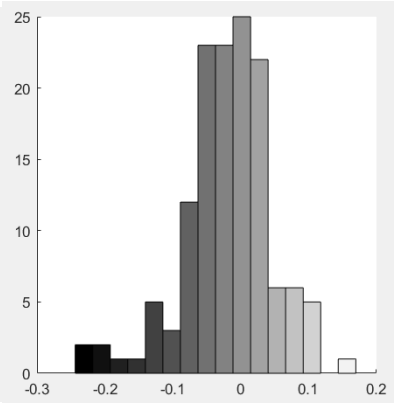

Figure 17b : key to Figure 17a. Actual % change in wages in South East England versus number of districts

Note : Fig. 15a -Lighter shading indicates bigger impact. Fig. 16a - Darker shading indicates districts with higher % wage increases under the preferred model. Fig. 17a - Dark shading indicates wage reduction.

9 Simulation outcomes for HS2

Compared with the HS1 impact analysis, we do not have 'actual' wages to compare against a counterfactual. Nevertheless the methodology is similar, we simply compare two outcomes based on alternative assumptions regarding travel times. One is embodied in a counterfactual model which presumes that HS2 will not happen. The second is a prediction of wage levels based on a model with travel times assuming the existence of HS2. To give more detail, we first obtain predicted equilibrium wage levels (obtained by iterating equation (23) for $t = 1, ..., 50$) as shown by Figure 8, using the Table 2 estimates and W_N and M_N, hence $\widehat{G}_N, \widehat{B}_N$,and $\widehat{C}_N$,based on a maximum commute time of 85 minutes. These predictions are compared with predicted

wage levels assuming a post-HS2 impact on travel times, with again a maximum of 85 minutes, which gives $\widetilde{W}_N$ $\left(\widetilde{M}_N\right)$. The difference, for $t = 50$, is our preferred prediction of HS2 effects. In Figures 18 and 19 we compare these HS2 effect predictions with predictions given by the simple model and demonstrate that while there are similarities, the simple model does not allow for the possibility of negative impacts. Finally, a detailed map of our preferred forecasts is given in Figure 20,with the dynamics leading to this equilibrium outcome given by Figure 21. These highlight the diversity of outcomes across all districts, many of which have no local HS2 treatment. While the majority of districts have near-zero impacts, they are not exactly zero, falling as low as -7%.

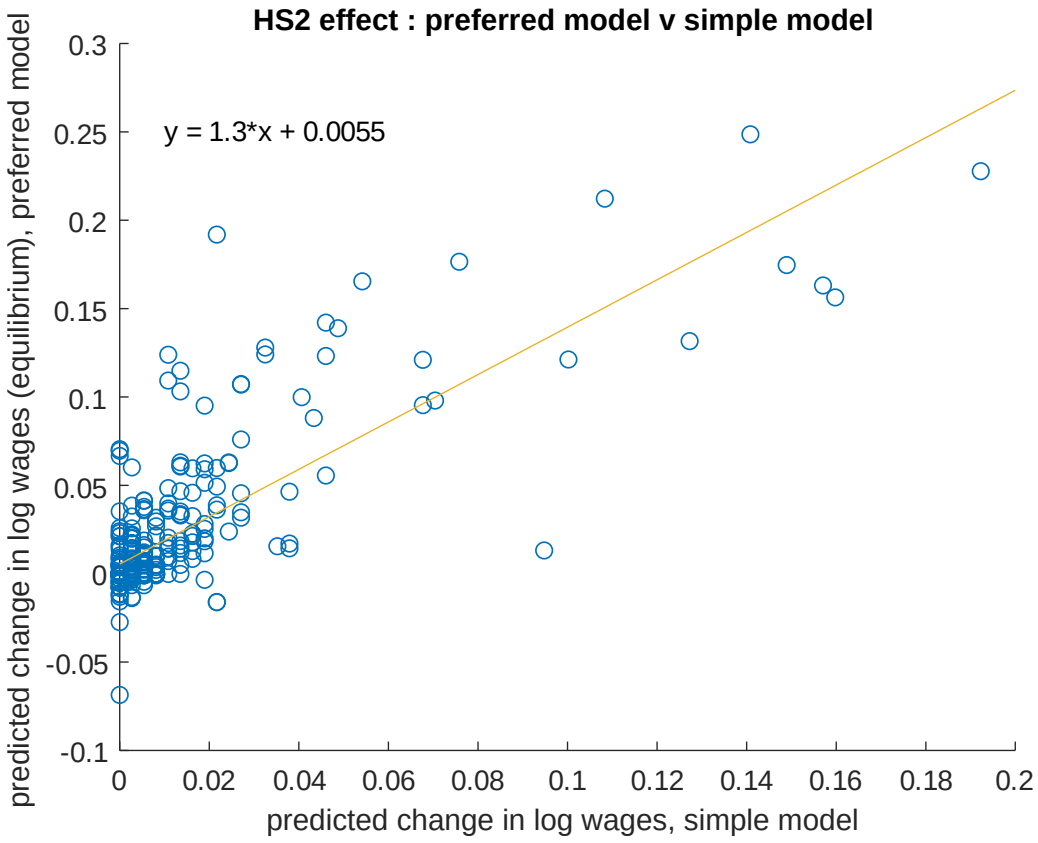

Figure 18 : Predicted % change in wages due to HS2 according to simple model and preferred model

Note : $R^2 = 0.6196$

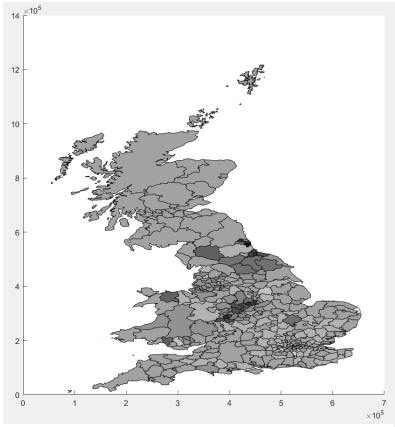

Figure 19a : Difference between HS2 predictions

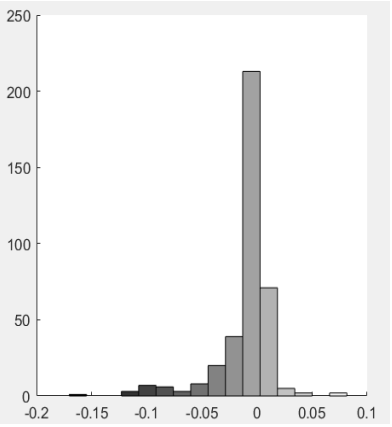

Figure 19b : Difference versus number of districts

Figure 19 : Difference between predicted HS2 effect under the simple model and predicted HS2 effect under the preferred model

Note : Darker shading indicates districts with higher % wage increases under the preferred model.

The mean effect is to increase equilibrium wages by almost 2%, but that conceals a diverse range of outcomes, from a maximum difference of 24.9% (Derby) to a minimum of -7%(Great Yarmouth). As with W_N,we again assume that $\widetilde{W}_N$ is based an n by n matrix of 1s and zeros $\widetilde{D}_N$. In this, 1 indicates that a pair of districts are within 85 minutes commuting time from each other, taking into account the change in travel times due to HS2, and zero indicates that they are not within 85 minutes travel time of each other. The resulting weights matrix $\widetilde{W}_N$ leads to $\widetilde{G}_N, \widetilde{B}_N$,and $\widetilde{C}_N$, which are used in the prediction equation(21) in place of $\widehat{G}_N, \widehat{B}_N$,and $\widehat{C}_N$. We again hold employment density (E_t), worker skill (h_t) and mean wages ($\overline{w}_t$) constant at their 2016 levels, but in this case the improved travel times lead to additional skilled workers within commuting distance, as given by $\widetilde{H}_t = \widetilde{D}_N h_t$.

One assumption that could not be captured by our estimation approach is that the 85 minute travel time will remain the optimal limit of commuting distance after HS2 is completed. It is possible that transportation innovations could affect the maximum time workers are willing to spend traveling.

However, there are several reasons to believe that this is not an issue for our predictions. First, our estimation of travel times includes high speed trains, thus our estimates capture the choice workers make between fast trains and other transport modes. Since our maximum travel time is informed by analysing revealed preferences over time and because the overall choice set of transportation modes does not change, the 85 minute limit is likely to be the same after HS2 is developed. This assumption is consistent with the finding of a report by The Royal Academy of Engineering (RAE 2015) that changes in congestion did not affect the maximum amount of time people are willing to spend driving to work over the last 20 years. It is also consistent with the figures reported by the UK Department for Transportation, which show that average travel times in the UK have not changed significantly over the last three decades (DFT 2016). Finally, we show that using the 85min travel time limit works well in predicting the impact of HS1. This shows that maximum travel times are unlikely to change over time and even if this assumption is violated, the changes are small and do not affect model predictions significantly.

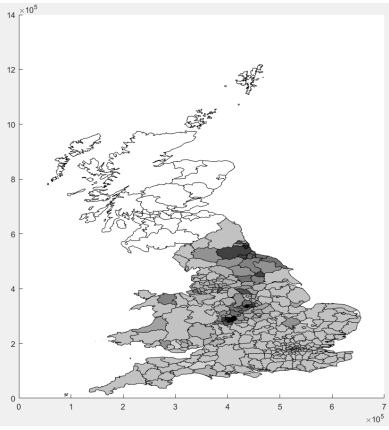

Figure 20a : Impact of HS2 according to the preferred model.

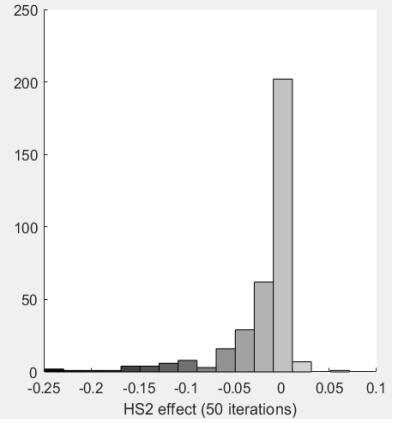

Figure 20b : Key to Figure 20a. % wage shortfall without HS2

Figure 20 : Impact of HS2

Note : Fig. 20a, districts with darker shading have largest % wage increases

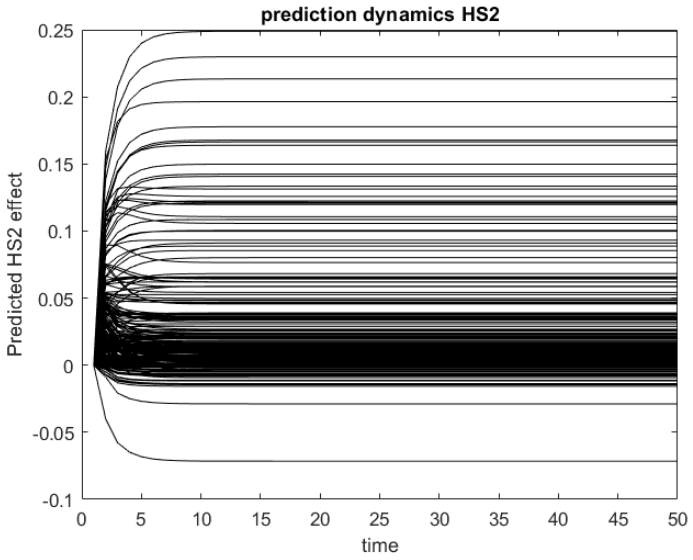

Figure 21 : HS2 prediction dynamics in accordance with Figures 20a and 20b

Note : The evolution of the HS2 effect towards equilibrium according to the preferred model. This ranges from 24.9% to -7%.

10 Conclusion

The paper presents a method of forecasting the impact of transportation innovations on wages using a spatial version of a traditional urban economics model and a prediction equation based on a dynamic spatial panel estimator. In particular, we focus on predicting the spatial distribution of the impact of a high speed rail improvement programme (HS2) planned by the UK government based on the projected changes in travel times under HS2. Our preliminary data analysis shows that wage levels are correlated with mean travel times to other districts, indicating that links to other locations are an important determinant of local wages. We capture this effect by including commuting into our model explaining spatial and temporal wage variation. We find that optimal results for our preferred model are obtained assuming commutable distance to be no more than 85 minutes travel time. We test this model by estimating the impact of a realised high speed rail (HS1) improvement project, comparing the estimated impact given by our model

with a simpler approach devoid of the commuting induced spatial interactions embodied in our model. The results show that our model performs comparatively better than the simpler approach, allowing a greater diversity of outcome, with effects transmitting over a wide range of districts not directly benefitting from investment under the HS1 project.

In our HS2 impact forecasts based on this model, we find that HS2 will on average positively increase wages but that the impact will be spatially uneven. The districts that will benefit most will be the ones projected to see the biggest reduction in commute times, due either to direct access to new stations or to the indirect effect of better commute times even where there is no direct access. One major concern evident in the literature is the negative impact of transport innovation on some places which lose their most productive and better paid workers who move to jobs in districts benefitting directly from an improvement in transport infrastructure, causing wages to fall rather than increase. However, the broadly positive benefits embodied in our analysis reflects the multi-facetted spatial dependence embodied within our simulation model. Wages across districts are mutually interdependent even if these other districts are not the direct or even indirect beneficiaries of a new HS2 station. Moreover, under our model a district's wage depends on its own wage in the previous period, and on temporally lagged wages within commuting distance. Via the economic theory underpinning our model, wages also depend on the density of employment locally within each district, and on the mass of educated workers within commuting distance. Additionally, wages are a consequence of unobserved effects embodied within the compound errors of the model, which are also spatially dependent according to commuting distance. The outcome is that linking an underperforming district with a successful does not adversely affect wages in the low wage district. In fact, our model indicates that the opposite is true. The projected positive impact of HS2 is driven by access to new workers, indeed, the biggest projected positive HS2 impacts are for districts at the lower end of the wage spectrum. This is because low wage districts tend to be relatively remote and see the greatest benefit from enhanced connectivity. Locations that will benefit the most from the new infrastructure will be the ones that currently have the poorest links to the rest of the economy. This has important implications for transport investment as an instrument of policy aimed at reducing spatial disparities in economic performance. Typically firms in such districts will gain access to workers with different skills. Generally, transportation innovation allows workers to travel further and firms to

benefit from a larger labour pool. This has been shown in other studies to have a positive impact on productivity of both firms and workers (Duranton and Turner, 2012).

One caveat regarding this interpretation is that we assume that our estimates obtained using current transportation times remain stable as transportation improvements are implemented. Although this assumption is impossible to test directly, there is international evidence (Redding and Turner, 2015) that travel time budgets remain constant over time. If workers value their time equally before and after changes to transportation networks, our estimates of their choices should also remain unaffected.

Finally, it is important to emphasize that while our aggregate analysis provides a plausible methodology to prediction at the aggregate level, it is not intended to be a full or complete account of the determinants of wage levels and the role played by economic density. We appreciate that a fuller picture of this can be provided by analysis at the micro-level, especially taking into account worker-specific heterogeneity (Combes et. al., 2012, Doran and Fingleton, 2015). As an example, Fingleton and Longhi(2013) combine employment density and market potential at the aggregate level with individual level variables (gender, occupational status, family status and age) to explain individual residence-based wage levels, as provided by the British Household Panel Survey.

11 Appendix : The theory underpinning equation (1)

In this section we set out a well-established theoretical basis for our model given as equation (1), as given by Abdel-Rahman and Fujita(1990), Ciccone and Hall(1996) and Fujita and Thisse(2002), among others. In this we partly follow the exposition given by Fingleton(2003).

The equilibrium wage derives from the existence of equilibrium in the 'services' sector of the economy, that is the sector under monopolistic competition. The equilibrium for a service firm (m) considered at the micro-level is a profit maximising equilibrium, so at this point (with free entry etc) profits are driven to 0. Denote profits by π, output by i_m,wages by w and labour by $ai_m + s$, where fixed parameter s is fixed costs and a is variable cost. Also

the profit maximising price is $p_m = wa\mu$, which is equal to marginal cost wa times mark up μ, since we don't have perfect competition, and μ is a measure of monopoly power. The profit maximising price is given by the derivative $\partial\pi/\partial p_m = 0$ of the service profit function π with respect to price p_m, where $i_m = kp_m^{-(\mu/(\mu-1))}$ so that $\pi = p_m kp_m^{-(\mu/(\mu-1))} - w\left(akp_m^{-(\mu/(\mu-1))} + s\right)$. Given $\pi = 0$,

$$\begin{aligned} wa\mu i_m &= wai_m + ws \\ ws &= (\mu - 1)\, wai_m \\ i_m &= \frac{s}{a(\mu-1)} \end{aligned}$$

So i_m is the equilibrium level of output, denoted from now on by i. All service firms are identical, each with equilibrium labour force $L = ai+s$.Given $(1-\beta)$ is the share of the workforce in the service sector, then total service employment is $(1-\beta)E$ where E is total labour. Dividing total services employment by the employment size of each firm (L) gives the number of service firms

$$x = \frac{(1-\beta)E}{ai+s}$$

Hence

$$\begin{aligned} Q &= \left[M^\beta I^{1-\beta}\right]^\alpha L^{1-\alpha} = \left[M^\beta I^{1-\beta}\right]^\alpha \text{ industry production function, with } L(\text{land}) = 1 \\ I &= x^\mu i \qquad \text{services production function} \\ Q &= \left[M^\beta x^{\mu-\beta\mu} i^{1-\beta}\right]^\alpha \\ x &= \frac{(1-\beta)E}{ai+s} \quad \text{number of service firms at equilibrium} \\ M &= \beta E \\ Q &= E^{\alpha(\beta+\mu-\beta\mu)}\left[\beta^\beta (ai+s)^{\mu(\beta-1)} i^{1-\beta}(1-\beta)^{-\mu(\beta-1)}\right]^\alpha \\ Q &= E^{\alpha[1+(1-\beta)(\mu-1)]}\phi \\ Q &= \phi E^\gamma \\ \ln Q &= \ln\phi + \gamma \ln E \end{aligned}$$

With $\gamma > 1$,there are increasing returns to density. But with small α so that $\gamma < 1$, the effect of 'congestion' is so severe that it completely overturns

any tendency to increasing returns. Increasing density is not accompanied by a commensurate increase in output.

From the marginal product of labour, $w = \partial Q/\partial M = \beta\alpha Q/M$ in which $M = \beta E$ is industry labour, so that the total wage bill as a share of Q is $wE/Q = \alpha$, hence $\ln w = \ln\alpha + \ln Q - \ln E$ and $\ln w = \ln\alpha + [\ln\phi + \gamma \ln E] - \ln E$ so that

$$\ln w = \ln\varphi + (\gamma - 1)\ln E \tag{26}$$

So equilibrium wage is an outcome of the marginal product of labour theory and the theory giving the equilibrium output for service firms. Thus equation (26) provides the basis for equation (1).

12 References

Abdel-Rahman, H., & Fujita, M. (1990). Product variety, Marshallian externalities, and city sizes. Journal of regional science, 30(2), 165-183.

Ahlfeldt, G. M., & Feddersen, A. (2018). From periphery to core: measuring agglomeration effects using high-speed rail. Journal of Economic Geography, 18(2), 355-390.

Anselin, L. (1988). Spatial Econometrics: Methods and Models Kluwer Academic Publishers Dordrecht.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The review of economic studies, 58(2), 277-297.

Baltagi, B H. (2013). Econometric Analysis of Panel Data, 5th Edition. Wiley.

Baltagi, B. H., Fingleton, B., & Pirotte, A. (2014). Estimating and forecasting with a dynamic spatial panel data model. Oxford Bulletin of Economics and Statistics, 76(1), 112-138.

Baltagi, B. H., Fingleton, B., & Pirotte, A. (2018). A Time-Space Dynamic Panel Data Model with Spatial Moving Average Errors. Regional Science and Urban Economics, xxx, 1–19.

Baltagi, B. H., Song, S. H., Jung, B. C., & Koh, W. (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data. Journal of Econometrics, 140(1), 5-51.

Baum-Snow, N. (2007). Did highways cause suburbanization? The Quarterly Journal of Economics, 122(2), 775-805.

Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. Portuguese economic journal, 1(2), 141-162.

Bouayad-Agha, S., & Védrine, L. (2010). Estimation strategies for a spatial dynamic panel using GMM. A new approach to the convergence issue of European regions. Spatial Economic Analysis, 5(2), 205-227.

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

Chamberlain, G. (1984). Panel data. Handbook of econometrics, 2, 1247-1318.

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

Chatzitheochari, S., & Arber, S. (2009). Lack of sleep, work and the long hours culture: evidence from the UK Time Use Survey. Work, Employment and Society, 23(1), 30-48.

Ciccone, A., & Hall, R. E. (1996). Productivity and the density of economic activity. The American Economic Review, 86(1), 54.

Combes, Pierre-Philippe. The empirics of economic geography: How to draw policy implications? Review of World Economics, 147(3):567–592, 2011.

Combes, P. P., Duranton, G., & Gobillon, L. (2008). Spatial wage disparities: Sorting matters!. Journal of Urban Economics, 63(2), 723-742.

Combes, P. P., & Gobillon, L. (2015). The empirics of agglomeration economies. In Handbook of regional and urban economics (Vol. 5, pp. 247-348). Elsevier.

Combes, P. P., Duranton, G., & Gobillon, L. (2012). The Costs of Agglomeration: Land Prices in French Cities (No. 7027). Institute for the Study of Labor (IZA).

Debarsy, N., Ertur, C., & LeSage, J. P. (2012). Interpreting dynamic space–time panel data models. Statistical Methodology, 9(1-2), 158-171.

Debarsy, N., & LeSage, J. (2018). Flexible dependence modeling using convex combinations of different types of connectivity structures. Regional Science and Urban Economics, 69, 48-68.

DFT - Department for Transportation (2013), High Speed Rail: Investing in Britain's Future - Phase Two, the Route to Leeds, Manchester and Be-yond.

Available at: https://www.gov.uk/government/publications/high-speed-rail-investing-in-britains-future-decisions-and-next-steps

DFT - Department for Transportation (2016), Road Use Statistics Great Britain 2016.

Available at: https://www.licencebureau.co.uk/wp-content/uploads/road-use-statistics.pdf

Doran, J., & Fingleton, B. (2015). Resilience from the micro-perspective. Cambridge Journal of Regions, Economy and Society 8 (2), 205-223.

Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. Review of Economic Studies, 79(4), 1407-1440.

Duranton, G., Morrow, P. M., & Turner, M. A. (2014). Roads and Trade: Evidence from the US. Review of Economic Studies, 81(2), 681-724.

Elhorst, J. P. (2001). Dynamic models in space and time. Geographical Analysis, 33(2), 119-140.

Elhorst, J. P. (2014). Linear spatial dependence models for cross-section data. In Spatial Econometrics. Springer, Berlin, Heidelberg.

Fingleton, B. (2003). Increasing returns: evidence from local wage rates in Great Britain. Oxford Economic Papers, 55(4), 716-739.

Fingleton, B. (2006). The new economic geography versus urban economics : an evaluation using local wage rates in Great Britain. Oxford Economic Papers, 58, 501-530.

Fingleton, B., Fuerst, F., & Szumilo, N. (2018). Housing affordability: Is new local supply the key?. Environment and Planning A: Economy and Space, 0308518X18798372.

Fingleton, B. (2008). A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. Spatial Economic Analysis, 3(1), 27-44.

Fingleton, B., & López-Bazo, E. (2003). Explaining the distribution of manufacturing productivity in the EU regions. In European Regional Growth (pp. 375-409). Springer, Berlin, Heidelberg.

Fingleton, B., Longhi, S. (2013). The effects of agglomeration on wages: evidence from the micro-level. Journal of Regional Science, 53, 443-463.

Fingleton, B., Le Gallo, J., & Pirotte, A. (2018). A multidimensional spatial lag panel data model with spatial moving average nested random effects errors. Empirical Economics, 55,113-146.

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

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

HS2 (2013), Journey times.

Available at:http://assets.hs2.org.uk/sites/default/files/consulation_library/pdf/P2C37_Journey%20times%20and%20frequencies%20LOW.pdf

Gibbons, S., & Machin, S. (2005). Valuing rail access using transport innovations. Journal of urban Economics, 57(1), 148-169.

Harris, C. (1954). The market as a factor in the localization of industry in the United States. Annals of The Association of American Geographers, 64, 315–48.

Hsiao, C. (2003). Analysis of panel data, 2nd. Cambridge University Press, Cambridge.

Kapoor, M., Kelejian, H. H., & Prucha, I. R. (2007). Panel data models with spatially correlated error components. Journal of econometrics, 140(1), 97-130.

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. The Journal of Real Estate Finance and Economics, 17(1), 99-121

Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. International economic review, 40(2), 509-533.

Kelley Pace, R., LeSage, J. P., & Zhu, S. (2012). Spatial dependence in regressors and its effect on performance of likelihood-based and instrumental variable estimators. In 30th Anniversary Edition (pp. 257-295). Emerald Group Publishing Limited.

LeSage, J. P., & Pace, R. K. (2009). Introduction to Spatial Econometrics (Statistics, textbooks and monographs). CRC Press.

Matyas, L., & Sevestre, P. (1996). The econometrics of panel data. A Handbook of the Theory with Applications.

McQuaid, R. W., & Chen, T. (2012). Commuting times–The role of gender, children and part-time work. Research in transportation economics, 34(1), 66-73.

Mutl, J., & Pfaffermayr, M. (2011). The Hausman Test in a Cliff and Ord Panel Model. The Econometrics Journal, 14(1), 48–76.

Parent, O., & LeSage, J. P. (2010). A spatial dynamic panel model with random effects applied to commuting times. Transportation Research Part B: Methodological, 44(5), 633-645.

Parent, O., & LeSage, J. P. (2011). A space–time filter for panel data models containing random effects. Computational Statistics & Data Analysis, 55(1), 475-490.

Parent, O., & LeSage, J. P. (2012). Spatial dynamic panel data models with random effects. Regional Science and Urban Economics, 42(4), 727-738.

Pesaran. M. H. (2015). Time Series and Panel Data Econometrics. Oxford University Press.

RAE - The Royal Academy of Engineering (2015), The transport congestion challenge getting the most out of the UK's road and rail networks.

Available at: https://www.raeng.org.uk/RAE/media/Publications/Reports/The-Transport-Congestion-Challenge.pdf

Redding, S. J., & Sturm, D. M. (2008). The costs of remoteness: Evidence from German division and reunification. American Economic Review, 98(5), 1766-97.

Redding, S. J., & Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. In Handbook of regional and urban economics (Vol. 5, pp. 1339-1398). Elsevier.

Schafer, A. (2000). Regularities in travel demand: an international perspective. Journal of transportation and statistics, 3(3), 1-31.

Simini F., González, M. C., Maritan, A., & Barabási, A. (2012). A universal model for mobility and migration patterns. Nature, 484, pp. 96–100.

Small, K. A., Verhoef, E. T., & Lindsey, R. (2007). The economics of urban transportation. Routledge.

Online Appendix

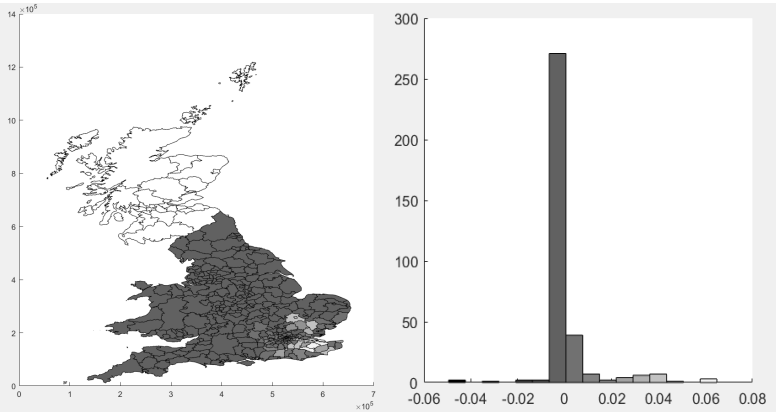

Figure 1a : Impact of HS1 on wages in 2009

Figure 1b : Key to Figure 1a

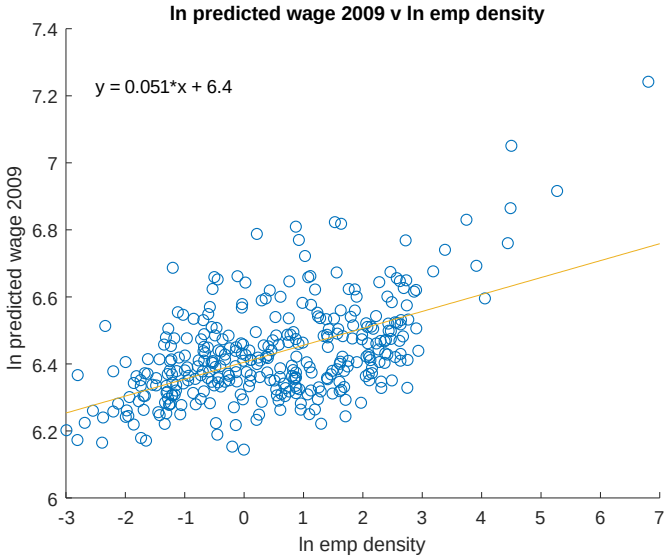

Figure 2 : The relationship between local employment density and predicted wage under the preferred model