UNIVERSITY OF LEEDS

Buses and the Economy II

Task 3 Report: Econometric Analysis

Daniel Johnson and Peter Mackie
Institute for Transport Studies
Marco Ercolani
University of Birmingham
June 2014

ITS

1 INTRODUCTION

A key research gap identified in the first phase of the work for Buses and Economic Growth (Mackie et al 2012) undertaken for Greener Journeys was whether there is a systematic variation in the level of employment at the local level with the quality of the bus network.

The work reported on here uses econometric models to address this issue and analyse the impact of changes and differences in bus accessibility on the labour market.

We analyse a panel dataset and a cross-sectional dataset of bus accessibility indicators, labour market indicators and socio demographic information to examine effect of differences in public transport (primarily bus) journey times on Local Authority District (LAD) areas' labour market outcomes. We examine different model specifications and discuss the use of statistical methods to establish the direction of causation in the relationship between bus accessibility and employment.

Our results add to the existing literature on labour supply elasticity/employment sensitivity within the spirit of the current WebTAG framework.

Section 2 discusses the background to this work, Section 3 the data used, 4 the methodology, 5 the results and Section 6 concludes.

2 BACKGROUND

The first order effects of better bus services emerge through improvements in travel times, reliability, comfort or fares. Currently the draft guidance of the DfT suggests Labour Supply Impacts should also be relevant to most schemes. The following text is taken from the guidance document TAG Unit A2.1 (DfT, 2014)

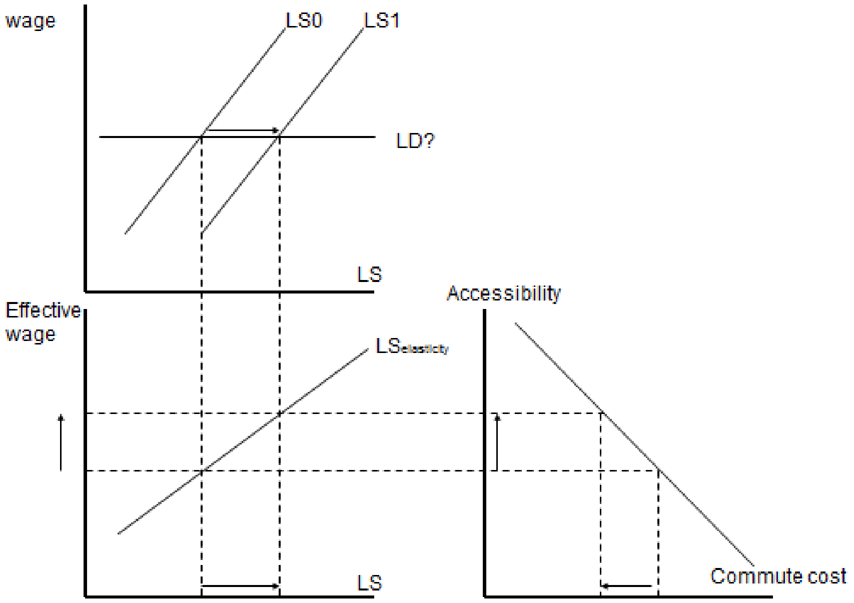
***Labour supply impacts**:*
"A change in transport costs alters the net financial return to individuals from employment. This is likely to affect the incentives of individuals to work, and therefore the numbers choosing to work and the overall amount of labour supplied in the economy. "

People at the margin of employment are faced with various employment opportunities. An improvement in bus accessibility will reduce the cost of entering employment through reductions in commuting costs, increasing the options available and the likelihood of finding work. The framework by which these labour supply impacts play out is as follows and illustrated in Figure 2-1:

- Changes in commuting costs impact on the wage net of commuting costs.
- Changes in net wages influence individuals' labour supply, the extent to which is determined by the labour supply elasticity.

We assume, following dialogue with the DfT on this issue, that the demand for labour is elastic at the going wage rate, gross of commuting costs.

Figure 2-1 Impact of Change in Commute Cost on Labour Supply



The modelling work carried out in this task aims to establish the underlying sensitivity of employment to changes in bus accessibility. This is a relatively difficult and unexplored area, characterised by complex modelling and onerous data requirements. To discover the appropriate mix of economics and geography, we adopt a multi-pronged approach. In order to estimate the impact of bus accessibility on employment we construct and analyse two datasets:

1. A panel dataset to examine impact of changes in accessibility over time and
2. A cross sectional dataset to look at the impact of differences between areas.

Both of these approaches will yield sensitivities of employment to changes in bus accessibility. The first approach assumes that changes in bus accessibility impact on employment. The second approach assumes that, if we can control for other factors influencing employment in an area, we can identify the sensitivity of employment to changes in bus accessibility by looking at whether remaining employment differentials between areas can be attributed to differences in levels of bus accessibility.

We seek to establish whether there is some consistency in the estimates from these two very different approaches to serve as a robustness check on the results.

3 DATA

3.1 Accessibility Indices

For measures of accessibility, we used DfT derived accessibility indicators (DfT, 2012) for journey times to employment areas calculated by public transport/walking and by car. The public transport/walking variable primarily captures bus travel times as the main public transport (PT) mode but also includes rail. The travel time indicators measure the time taken for users to reach the nearest employment centre by mode of transport (public transport/walking, cycle and car). The calculation of these travel times is rather complex as described below.

Employment destinations are determined using the number of jobs in a Lower Super Output Area (LSOA) made up of several output areas (OA) which are the smallest spatial areas used in the Census. For our purposes we are focusing on the nearest employment site (with separate measures calculated for employment areas with <500, between 500 & 5000 and above 5000 jobs). The employment data are assumed to be located at the centroid of the Output Area that has its population centroid nearest to the population centroid of the LSOA as a whole. The employment destination is assumed to be located at the population centroid of this OA.

The public transport network is represented by the National Public Transport Data Repository (NPTDR, www.nptdr.org.uk) – which is a snapshot for a single week of public transport access points (e.g. bus stops) and timetables for England. These data are then processed to create single public transport access point locations for each OA. The destination is assumed to be accessed through the public transport access node for that OA. Public transport travel times from OAs with no public transport access points are calculated by adding the walk time from the relevant OA population centroid to the nearest public transport access node. It is assumed that all residents of each OA start at the population weighted centroid and walk to the bus stop.

It is assumed that people in an output area can access points outside their area but within the MSOA, thus creating a number of possible access points.

Travel times are calculated for journeys between these access points and employment destinations and captured in a journey time matrix. The public transport travel time matrix is constructed using travel time paths out from each public transport access node starting with the nodes with the highest frequencies of public transport services and working down to the nodes with the lowest frequencies. The ten shortest travel times from each Origin (i.e. Output Area) to large employment sites are retained. For each of the ten shortest public transport routes, the public transport times are calculated for 23 half hourly slots for incoming and outgoing trips (i.e. 46 travel times per destination). Each of the 46 travel times is then weighted by service and slot to give a representative travel time.

The following assumptions outlined in Table 3-1 were used when calculating times on each route. The travel times estimated for each MSOA and LAD are population weighted versions of the derived OA values.

Table 3-1: Elements of Journey time measures

Journey time element	Assumption
Door to boarding public transport	Minimum time: 5 mins
Door to public transport stop/node	Maximum distance: 1.2 miles (24 mins)
Waiting time at bus stop/station/etc	Maximum time: 20 mins
Maximum interchanges	3
Interchange time	Minimum 10 mins

Car journeys are assumed to start at the population centroid. The journey connection is made directly from the road and footpath network to the destination point as specified by the co-ordinates. To create the car journey matrices, all roads other than alleys and local streets are used in the analysis. A similar approach to that used for public transport is adopted by building a MSOA level matrix across England and infilling this to each OA using the local road network.

We experimented with two measures of bus accessibility data over the five year period 2007-2011. These bus accessibility datasets measure average travel times (including wait, walk and interchanges) in Local Authority Districts to nearest large employment centre. One measure is for employment centres of 500 to 5000 individuals and the other for 5000 or more individuals.

Initially only the 500 to 5000 travel area data were available to carry out the analysis. However, closer scrutiny revealed that 2007 travel data had more variance in and had been computed with a different methodology. Without 2007, the analysis on the 500-5000 measures performed poorly. This was because there is very little variance in these bus travel times, with most being clustered around 9 minutes and most car travel times at 5 minutes which is the lower-bound for these measures.

The results in this report are based on the 5000-plus travel area measures for 2008 to 2011 as we believe these are a better representation of concentrated areas of employment, and as such performed better in the estimates. These data were derived for us by the DfT[1] during the project but the resulting delays gave us less time for the analysis. The inconsistency in the 2007 travel data remained in the 5000-plus dataset.

[1] We acknowledge, in particular, the help of Rachel Moyce from the DfT in co-ordinating the processing of this data.

These measures are not weighted to represent generalised journey time or able to be broken down to their component elements of walk, wait and in vehicle time to facilitate estimation of GJT. The framework of the current WebTAG guidance in TAG Unit A2.1 (DfT, 2014) is one in which changes in generalised cost of commuting stimulate a labour market response. Clearly this involves GJT and Fares, so is not a perfect fit. In Task 2 we show how our unweighted accessibility measures and the associated employment elasticities can be converted to be compatible with GJT framework.

3.2 Panel Data

Due to the availability of the accessibility data discussed in section 3.1, we had four years of data upon which to construct our panel.

3.2.1 Level of disaggregation

The level of disaggregation chosen for the panel data was the 324 local authority districts (LAD) in England, excluding City of London and Isles of Scilly, over the years 2008-2011. Six further observations were dropped due to missing values for the covariate on elementary occupations. This gave 1,290 observations on which to conduct our analysis.

3.2.2 Dependent variables

Dependent variables were based on the employment rate as published as part of the Annual Population Statistics series by the ONS on the NOMIS Official Labour Market Statistics website (www.nomisweb.co.uk) and we included the following at the LAD level:

- Official employment level for 16 to 64 year olds.
- Official employment rate for 16 to 24 year olds.[2]

The official employment rate is calculated as E/(E+U+I), where E represents the numbers in employment, U unemployment and I represents the numbers economically inactive. We analysed the youth employment rate separately as this group has particularly high dependence on the bus network.

3.2.3 Rural/Urban Stratification

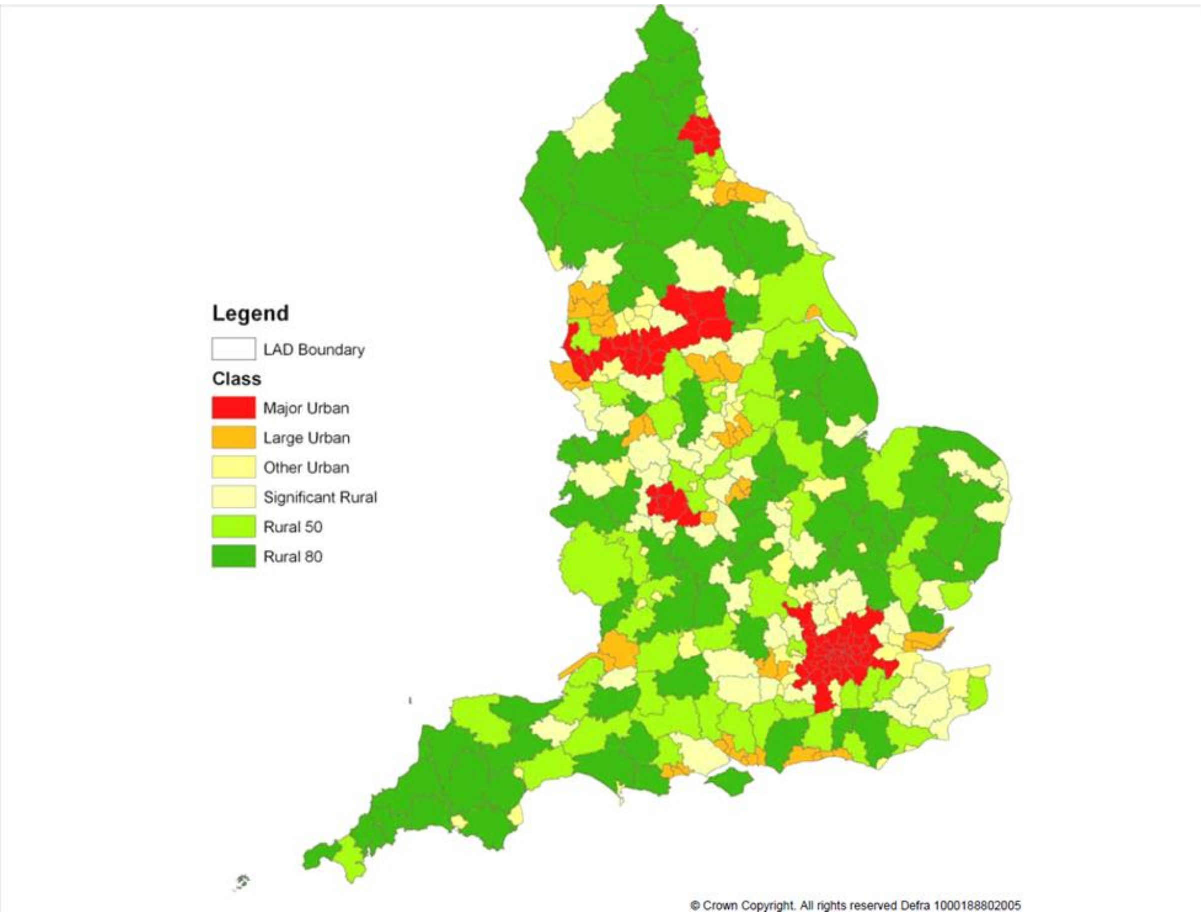
An important element of our analysis was to understand whether the sensitivity of employment to bus travel time varied by area type. In order to undertake this segmentation of area types we referred to the typology outlined in the Defra Classification of Local Authority Districts and Unitary Authorities in England (DEFRA, 2009). Table 3-2 below shows the DEFRA classifications and our final grouping. Figure 3-1 shows the spatial distribution of these areas.

[2] The employment level for 16 to 24 year olds was not available from ONS.

Table 3-2: Rural and Urban Stratifications used in analysis

DEFRA Classification	No. Districts	DEFRA Definition	Final Stratification
Major Urban	76	100k people or 50 percent of their population in an urban area with a population of more than 750,000	**Split into Dense Urban (ie Metropolitan) and London**
Large Urban	**45**	50k people *or* 50 percent of their population in one of 17 urban areas with a population between 250,000 and 750,000	**Other Urban**
Other Urban	**55**	<37,000 people *or* less than 26 percent of their population in rural settlements and larger market towns	
Significant Rural	**53**	More than 37,000 people *and* more than 26 percent of their population in rural settlements and larger market towns	**Rural**
Rural-50	**52**	districts with at least 50 percent but less than 80 percent of their population in rural settlements and larger market towns	
Rural-80	**73**	districts with at least 80 percent of their population in rural settlements and larger market towns; there are 73 districts in this group	

Figure 3-1: Map of Urban/Rural Classifications of LAD areas



3.2.4 Other co-variates

We use the following additional covariates in our model, identified for each LAD in each year:

- Population density. This is often used as a measure of agglomeration in estimating productivity impacts of different urban sizes/densities. Agglomeration economies have been the main focus of attention in the literature on the wider economic impact of transport investments (see for example SACTRA, 1999 and Venables, 2007). This literature postulates that numerous linkages between economic agents, brought closer together by the transport improvement, generate externalities which, collectively and at a localised level, give rise to aggregate increasing returns or agglomeration economies. By including this measure, we attempt to control for impacts on employment through any productivity impacts which may occur through changes in density.

- Average GCSE and NVQ attainment. These are used to control for changes in the skills base of a particular area over time; these are measured by the proportion of the working catchment with GCSE attainment of grades A-C and NVQ3+ or higher respectively.

- Gender mix and ethnic mix[3] of an area. Changes in these reflect changes in the mix of the labour supply which will impact on overall employment and might be proxies for other population and labour market characteristics.[4]
- Mix of public and elementary occupations within employment. These measures control for the structure of employment in each area. Some more deprived areas are reliant on public sector employment and may have higher concentrations of employment in elementary (lower skilled) occupations.

Descriptive Statistics of these variables are provided in Table 3-3

Table 3-3: Descriptive Statistics of Panel Data:

Variable	Mean	s.d.	Min	Max	Comment
# Employed, aged 16-64	80624	67404	15000	673000	Numbers employed
% Employed, aged 16-64	72.52	5.50	54.90	88.20	Employment rate [=E/(E+U+I)] aged 16-64
% Employed, aged 16-24	55.13	11.47	19.20	100.00	Employment rate [=E/(E+U+I)] aged 16-24
Bus T.T.	37.17	22.39	8.09	119.44	Bus travel times to nearest large employment area
Bus T.T. * London	20.31	7.44	8.09	45.88	128 observations
Bus T.T. * Dense Urban	20.96	5.15	13.21	45.03	152 observations
Bus T.T. * Other Urban	27.30	17.30	11.03	109.97	388 observations
Bus T.T. * Rural	50.64	21.66	15.31	119.44	628 observations
Car T.T.	10.97	8.09	5.00	87.08	Car travel times to nearest large employment area
# Populat., aged 16-64	114534	105978	19084	1068254	Population aged 16-64
# per Sq. Km	1626	2315	24	13885	Population density
% Ethnic min.	9.74	11.82	0.14	75.00	% of ethnic minority - aged 16-64
% Male	49.74	1.03	46.40	53.20	% of aged 16-64 who are male
% NVQ3+ aged 16-64	49.98	8.21	27.20	76.30	% with NVQ3+ - aged 16-64
% GCSE+ aged 16-64	23.47	5.79	5.50	41.40	% with GCSE grades A-C or equivalent
% Public Employees	23.36	5.02	8.60	43.90	% of those in employment working in Public sector
% Elementary Occup.	10.91	3.52	2.80	25.00	% of those in employment in elementary occupations

[3] A number of LADs had missing cases for ethnic composition. We filled in the remaining missing cases by using mean values we obtained from "ONS Neighbourhood Statistics" at www.neighbourhood.statistics.gov.uk, [accessed 24/6/2013]

[4] Other variables were included in earlier analysis, including proportion of those working in lower ranked occupations, but these were dropped due to data inconsistencies.

3.3 Cross-Sectional Data

3.3.1 Level of disaggregation

This dataset utilises the 2011 UK Census data at the Mid level super output area (MSOA), giving 6786 observations on social and labour market measures for England, matched again to bus accessibility data from the DfT. MSOAs are constructed within local authority/UA boundaries to contain populations between 5,000 and 15,000 individuals, and between 2,000 and 6,000 household units.

3.3.2 Dependent variable

Dependent variables were based on the employment measures available from NOMIS for the 2011 Census (www.nomisweb.co.uk) at the MSOA level:

- Employment level for 16 to 74 year olds.
- Employment level for 16 to 24 year olds

Again we are interested in looking at the youth employment rate separately given their high dependence on the bus network.

3.3.3 Rural/Urban Stratification

To segment our results we use the same rural urban measures as were used at the LAD level as outlined in 3.2.3

3.3.4 Other co-variates

This data set allows us to investigate the relationship between spatial differences in bus (and car) accessibility and differences in employment rates, controlling for other localised factors such as population level, and the profile of the population in terms of the percentage of car availability, males, ethnic minorities, and those with English as a first language. Other variables experimented with included degree and no qualifications, lone parents, multiple deprivation scores and social class. We had to omit these variables because they were highly correlated with one another and with some of the variables we did retain.[5]

Descriptive statistics for the dependent variables and co-variates (including the ones not used in the final specifications of the models reported here) are shown in Table 3-4 We also identify a set of constants for each Local Authority District which control for wider local factors (ie at the LAD level rather than the MSOA) which may affect employment.

[5] Correlation coefficients between collinear variables exceeded 0.8.

Table 3-4: Descriptive Statistics of Cross-Sectional Data

Variable	Mean	SD	Min	Max	Comment
# Employed	3551.39	859.63	1183	8830	Numbers employed aged 16-74
# Employed aged 16-24	467.78	175.66	124	3131	Numbers employed aged 16-24
Bus T.T.	33.55	24.88	5	120	
Car T.T.	11.4	9.02	5	120	
Popul. aged 16-74	5712.16	1232.15	1622	15222	Population aged 16-74
Popul. aged 16-24	915.96	559.37	145	9896	Population aged 16-24
% No Car Availability	25.03	14.88	2.92	82.49	% of population with no access to a car
% Ethnic minority	13.43	17.81	0.4	94.4	% of ethnic minority of all residents
% Econ active males	53.27	2.23	46.92	68.78	% of those economically active who are male
% English 1st lang.	92.61	9.61	38.42	99.73	% of all residents with English as a first language
Population/SqKm	*33.04*	*34.76*	*0.1*	*247.2*	*Population density (all residents)*
% Qualif. level 4+	*27.03*	*11.4*	*4.8*	*71.4*	*% of all residents with qualifications degree level or higher*
% Social Class D-E	*25.12*	*11.55*	*4.17*	*64.88*	*% of all residents in lowest social class*
% Deprived	*0.51*	*0.48*	*0*	*6.04*	*% of all residents in deprived (across 4 or more measures of deprivation)*
% No qualifications	*22.79*	*7.9*	*2.2*	*50.7*	*% of all residents with no qualifications*

Note: Variables in italics were not used in final models.

4 METHODOLOGY

4.1 Introduction

Our analysis on labour market outcomes is conducted on 2 datasets that by their nature require different estimation techniques.

Firstly, we estimate a model using panel data. The panel dataset is measured across LAD areas and over 2008-2011, giving 1290 observations. The nature of this data is that we cannot observe all the variables which determine employment in a given area. These variables include, for example, natural resources (eg a coalfield), local geography (eg ports may improve employment prospects in coastal areas), the presence of large historical employers in an area (eg Boots in Nottingham) and any agglomeration economies arising from high levels of concentration of economic activity in an area (as found in London). A Fixed Effects approach is a standard way of controlling for any time-invariant unobserved characteristics, in this case of an area through the estimation of an area constant. We found that alternative estimation approaches from OLS and Random Effects were inappropriate in this context[6].

Secondly, we estimate a cross sectional model utilising 2011 UK Census data at the Mid level super output area (MSOA), giving 6786 observations on social and labour market measures for England, matched again to bus accessibility data from the DfT. This data set allows us to investigate the relationship between spatial differences in bus (and car) accessibility and differences in employment rates, controlling for other localised factors (such as population, car availability, qualifications, occupations etc). In estimating this model we again include fixed effects constants for the LADs to emulate the control for the same LAD fixed effects as in the panel regressions.

4.2 Endogeneity and Instrumental Variable Analysis

Another important issue within this analysis is to understand if accessibility and employment are endogenous – could levels of bus accessibility be lower in areas of higher employment because of higher levels of access to other modes of transport, or do better levels of bus accessibility lead to increases in employment? In other words which direction is the causality in the relationship between the two?

To identify causality in a relationship, we need an exogenous variable - that is, a variable which is not related to any of the other variables, unobserved and observed. There are potentially an infinite number of unobserved variables which could cause endogeneity. If unobserved variables are time invariant then a Fixed Effects approach can control for them. If this unobserved heterogeneity changes over time then a different approach is required. For instance a historically deprived area with persistently lower levels of employment may

[6] OLS models cannot control for unobserved differences between areas, resulting in poor levels of fit and counterintuitive signs on the estimated parameters. Random Effects models are statistically more efficient than Fixed Effects models, giving more precise coefficient estimates but are based on the assumption that the unobserved time-invariant variables are not correlated with any observed variables. We tested for the appropriateness of Random Effects models using a Hausman test, which suggested Fixed Effects were better suited (more consistent) to the data.

have higher levels of bus services because of higher demand from the population. In as much as this effect persists over time a Fixed Effects approach can deal with it and provide unbiased estimates. But Fixed effects will not control for any endogeneity arising from changes in employment causing changes in demand for bus services.

We investigate this issue of endogeneity further through the use of instrumental variable (IV) approaches to control for the endogeneity between bus accessibility and employment. Successful IV estimation requires the identification of at least one variable (instrument) which influences bus accessibility but is uncorrelated with the error term in the equation that explains employment. I.e. is only correlated with employment through its impact on accessibility. IV estimation comes in many forms and we use the 2 stage-least squares (2SLS) variant of it. In the first stage an instrumenting regression is estimated where the endogenous variable (bus accessibility) is modelled using the instrument(s) and all the explanatory variables from the second stage. In the second stage an instrumented regression is estimated the 'fitted' or predicted variable(s) from the first stage is used as one of the explanatory variables in place of its actual counterpart.

Good instruments in non-experimental datasets such as ours, i.e. not derived from the selective treatment of particular units in order to establish clear control groups, are notoriously difficult to identify. This is due to the inter-relationship (endogeneity) between most observable socio-economic variables.

There are two obvious requirements for a good instrumental variable. Firstly, the instrument must be highly correlated with the endogenous explanatory variable it is instrumenting (bus accessibility). Secondly, the instrument must have a very low correlation with the residual error from the second stage regression (on employment).[7] These two requirements are referred to as instrument relevance and instrument exogeneity.

We cannot test this second requirement, instrument exogeneity, other than appealing to economic intuition. Given a selected instrument (or instruments) we can however test for instrument relevance by comparing the OLS and 2SLS estimates for employment to determine whether the differences are significant. If they differ this implies there may be some degree of endogeneity and IV is appropriate. This is known as the Wu-Hausman test[8].

Successful IV estimation often involves use of long time-lags of the instrumented variable (e.g. bus accessibility) as instruments in the (first stage) instrumenting regression. This is because it can be argued that these lagged values cannot have been influenced by the current level of the dependent variable (e.g. employment) and are thus exogenous. For our cross-sectional analysis we were able to make use of this, as the dataset did include lagged

[7] Sometimes this is mistakenly interpreted as meaning that the instrumenting variable(s) must be uncorrelated with the dependent variable in the instrumented (second stage) regression.

[8] An additional test known as the Sargan test uses over-identifying restrictions in a statistical model involving more than one instrument for each endogenous variable to test whether these instruments are truly exogenous. However, caution has to be exercised as weak instruments, i.e. instruments which are weakly correlated with the endogenous variable, can sometimes pass the Sargan test. If instruments are weak, the sampling distributions of the IV statistics are non-normal and the resulting IV estimates can be biased. There is no robust statistical test for weak instruments. Also, it is not appropriate to apply the Sargan test when just one instrument is used.

measures of accessibility for 2010 and 2009.[9] We could not adopt this approach the panel data analysis because we already used all the time series data for the panel element. The use of even a 1 year lag on accessibility would have meant losing 25% of our data and, in any case, as one year lag would not have served as much of an instrument. Other applications of this in panel datasets rely on much longer time periods which then facilitate semi-automated Generalised Methods of Moments (GMM) models first proposed by Arellano and Bond (1991).

IV estimation is inefficient, ie leads to lower t-statistics and less precise estimates than under OLS, so if OLS is appropriate it is preferable. If there is endogeneity and it is not controlled for then OLS estimates of the relationship may be biased, yielding parameters which do not accurately reflect the direction of causation.

4.3 WebTAG framework

For our estimation of the employment impacts, we have kept within the spirit of the WebTAG framework, but have taken a rather more parsimonious/pragmatic approach. This is tailored to the focus on public transport and the data available for the estimation of the sensitivity of employment to changes in generalised journey times through an ex-post analysis of impacts of differences in bus accessibility. Ideally we would have liked our measure of accessibility to include both time and fare but unfortunately fare data was not available at zonal level and we have worked on time only as our accessibility indicator.

We are also focusing on the level of employment rather than the number of hours worked.

There are several ways in which our approach differs in the detail to the current WebTAG framework:

- We focus on accessibility (bus travel times to local employment centres) measures rather than Generalised Cost. This allows us to avoid complex calculations of (modal weighted) generalised costs for zonal pairings.
- We are directly estimating changes in employment with respect to changes in accessibility rather than wage. This by passes issues to do with estimation of share of GC in Wages, and the appropriate value of time.
- We currently focus on estimation of employment impacts rather than GDP.
- We estimate a sensitivity parameter which captures the responsiveness of employment to changes in public and private transport accessibility rather than the labour supply elasticity. The use of the labour supply elasticity in this context implicitly assumes that the increase in labour supply will find its way into employment, i.e. that demand is perfectly elastic at the prevailing wage rate. Depending on the demand conditions, it seems reasonable to expect some extra workers might not actually be able to find work despite being willing. Our sensitivity parameter effectively conflates the demand and supply effects.

[9] Though the cross-section data did include 2007 accessibility measures we did find that, as for the panel data, these 2007 measures must have been created using a different methodology to subsequent years. Accessibility measures for 2008 were not available.

We think that one labour supply elasticity/employment sensitivity figure may be over simplistic. In lower-skilled occupations, labour supply is elastic because a pool of labour is employable at a fairly constant market wage rate, but where jobs require specific skills and training, the labour supply will be more inelastic because it is hard to expand the workforce in a short period of time when demand for workers has increased. As highlighted in the WebTAG guidance, a fixed value also assumes a fixed ratio of male to female workers, who have different labour supply elasticities. Our estimation approach has the flexibility to estimate different elasticity parameters for different area types, and to check whether a 'one size fits all' elasticity is appropriate. A more segmented approach may avoid issues of aggregation bias that could arise from using one value.

4.4 Panel Model

4.4.1 Basic Model

Spatial cross-sectional relationships between bus access, employment and earnings are difficult to establish, as many other factors influence these relationships, e.g. employment density is likely to be larger in towns/cities than rural areas but this is obviously not purely due to the better level of bus access. Some means to control for differences in employment arising from the different socio-economic make-up of areas needs to be considered.

This is typically addressed using panel data methods. Panel data involves repeated observations of accessibility of the spatial unit over time which allow the modelling to control for the unique and unobserved characteristics of each area. This approach is standard in analysis of this nature (see Gibbons et al 2012 for a recent application based on changes in road infrastructure).

We use panel data to look at temporal differences, ie changes over time in service levels and labour market outcomes. Using a fixed effects model applied to such data eliminates any time invariant unobserved heterogeneity, ie the impacts of area characteristics on the labour market discussed in Section 2, even though these variables are not directly observed. This allows us to isolate the impact of changes in service level on the labour market, given adequate controls for other key factors, which may include population and access to other modes.

This application requires marrying bus accessibility data with labour market data in a zonal model. These zones need to be small enough so that the accessibility measures are appropriate and sensitive to what would likely be small observed changes in fares/service levels.

Our model formulates the employment measure in area *i* at time *t* in the following way:

$$\text{Employment}_{it}=f(A_{it}, C_i, V_{it,} D_t)$$

where:

A_{it} is accessibility measure for area *i* in time *t*;

V_{it} are time variable factors such as population, skills level, car availabilty etc

C_i are area specific constants capturing impact of area characteristics

D_t are yearly *dummies* capturing the impact of changing macro-economic conditions.

Regression analysis using the natural log of our variables facilitates ease of interpretation. The estimated coefficients represent the sensitivities (i.e. elasticities) of employment to changes in the explanatory variables. This form is commonly used in the literature (see Gibbons et al 2012). These sensitivities can be used to evaluate the implied changes in employment resulting from, say, reductions in bus travel times. Other model forms were used, but these were discarded as they did not significantly improve model performance and were harder to interpret. However, as we did not have employment levels for 16-24 year olds, we used the log of the employment rate as our dependent variable.

A straightforward OLS approach to the analysis of these data would treat all areas as homogenous outside of observable differences. This would not account for differing unobserved spatial characteristics, as may have emerged through the existence of underlying geography, natural resources, established infrastructures, large employers, agglomeration density etc.

Fixed effects regression is an established technique for dealing with panel data (see Greene 1997, chapter 14) and the appropriate starting point here given the very high degree of unobserved heterogeneity between the Local Authority Districts (LADs). This approach controls for the time-invariant local economy characteristics as far as is possible. In addition to these LAD fixed effects, year dummies were included to control for macroeconomic impacts on employment.

Though we focus primarily on the transport-employment relationship for the entire labour market, we also consider the 16-24 age group, who are more likely to be captive to public transport.

4.4.2 Serial Correlation in the Panel Model

The classical linear panel data model assumes that all effects are contemporaneous, ie that any effect of a change in one variable on another (dependent variable) will occur in the same time period. If this relationship holds then any errors in the model are serially uncorrelated, ie do not depend on past errors.

However, this assumption of absence of serial correlation is not valid if there is a dynamic relationship between variables such that effect on the dependent variable is often distributed over several time periods. For example, any change in bus accessibility may take time to manifest itself in changes in employment due to the time it takes commuters to react to such changes and the time it takes firms to adjust to a larger pool of labour. In such cases, serial correlation leads to lack of precision of parameter estimates, ie incorrectly inflated t-statistics. Nonetheless, it does not lead to biased estimates.

Modelling such dynamic processes involves the inclusion of lagged terms and thus reduces the already limited number of years we have in our panel model. Obviously, the less data we can estimate a model on, the less precise and robust are the estimates. However, we do estimate a dynamic version of the Fixed Effects model on the panel data (by including an autoregressive error structure). This allows us to test whether the inclusion of such an error structure is appropriate, ie whether we have significant serial correlation in our models.

4.5 Cross sectional Model

We estimate a cross sectional model of employment at the MSOA level, giving 6786 observations on social and labour market measures for England, matched again to public transport accessibility data as measured by bus travel times. The cross sectional approach allows us to investigate the relationship between spatial differences in public transport (and car) accessibility and differences in employment rates, controlling for other localised factors. In order to estimate this model we conduct fixed effects regression analysis by estimating a set of constants for each Local Authority District which capture area-wide unobserved characteristics influencing employment.

Our model formulates employment in MSOA *i* within LAD/UA *k* in the following way:

$$Employment_i = f(A_i, C_{LADk}, V_i)$$

where:

> A_i represents the accessibility measures for area *i*;
>
> V_i are variable factors such as population and labour force composition variables
>
> C_{LADk} are constants capturing the impact of unobserved variables within LAD area k

As in the panel model, we use a log-linear functional form to derive proportional responses (elasticities) for the impact of differences in travel time on employment directly from the parameters.

5 ANALYTICAL RESULTS

5.1 Panel Model

5.1.1 Basic Model Results

Table 5-1 shows the results of the estimation of the employment level regression. We use a log-linear functional form (by taking the natural log of the dependent and independent variables) so that resulting parameter estimates can be directly interpreted as elasticities.

We report four different regressions, all of which suggest evidence of statistically significant links between local employment and bus travel times. The first 2 columns estimate a single elasticity for employment with respect to changes in bus travel time with and without observations from London included.

The elasticity from these first two regressions are -0.018 and -0.016 and are significant at the 5% level and 10% level respectively. These coefficients suggest that for a 10% decrease in bus travel times, we expect a 0.16-0.18% increase in employment.

The final two columns segment the travel time variable to give different elasticities for the urban forms described in Section 3.2.3. Here we see that whilst the sensitivity in London is not significantly different from zero, the travel time elasticity is highest at -0.065 to -0.067 for Dense Urban areas outside London and significant at the 5% level. For Other Urban areas the elasticity is lower at -0.024 but only significant at the 10% level when London observations are included. In rural areas the employment to travel time elasticity is (statistically) close to zero.

These results are interesting as they show the higher sensitivity to employment for denser Urban areas, with the exception of London. This could be due to higher rates of non car availability in these areas and possibly because of differences in labour supply elasticities. The difference between the sensitivity of Dense Urban areas to Rural areas is significant at the 5% level, but insignificant for the other interactions[10].

The coefficient on the log of car travel time is positive but insignificant across all the regressions in Table 5-1 suggesting that car journey times do not impact on employment.

The other covariates were also logged in the regression estimation, although these were measured as percentages rather than levels. This requires some care in interpretation. The resultant coefficients on the other covariates represent elasticities of employment with respect to proportional changes in the percentage rate of the covariate. For example a parameter of -0.1 on a covariate indicates a 10% increase in the percentage value (for example from 50% to 55% and not from 50% to 60%) of the covariate would lead to a 1% reduction in employment.

[10] The necessary t-statistics for this observation were derived from the re-estimation of the models with a base level of Bus TT for the Rural areas, and additional interactions with Bus TT included for London, Dense Urban and Other Urban areas. The reported t-statistics from this form of the model were used to determine whether these areas had statistically significantly higher sensitivities than Rural areas. The results from this model are otherwise identical to those reported in **Table 5-1**.

Regarding these other socio-economic covariates, results are as expected, giving weight to the plausibility of our model. Employment is positively and significantly related to population size and population density. Clearly, we would expect higher levels of employment for higher populations. The positive relationship with population density indicates that there are employment effects for areas with higher concentrations of population, possibly due to the agglomeration impacts discussed in Section 3.2.4.

The impact of the proportion of those from ethnic minorities and elementary occupations is insignificant.

We find a higher proportion of males in the workforce also has a positive and significant impact on employment. This could be because females face relatively more barriers to employment, such as childcare duties and other family responsibilities.

We find higher educational attainment and training are associated with higher levels of employment – a higher skilled workforce, all else equal, will have a higher marginal product of labour and thus be more employable. This relationship is significant at the 5% level for both variables in all models in Table 5-1.

The proportion of those in public employment has a negative impact on overall employment suggesting public employment is higher in areas where the underlying level of employment is lower and is possibly being used to support local employment.

The macro-economic dummies show that, relative to 2007, 2008 and 2009 exhibited higher levels of employment and 2010 slightly lower, albeit insignificantly. This indicates a somewhat faltering economic recovery.

Table 5-2 shows the results of the estimation of the employment rate regressions for 16-24 year olds for the four different specifications as used in Table 5-1. Whilst the results are strong in terms of the other covariates, we do not find a statistically significant relationship between employment and public transport accessibility amongst this age group[11].

[11] We ran a regression of the employment rate for the 16-64 cohort to check that these results were not the artefact of the different form of the dependent variable, but results for the bus travel time related variables still were significant and of the expected sign.

Table 5-1 Fixed effects regressions on ln(Employment Level), 2008-2011

	(1) With London	(2) Without London	(3) With London	(4) Without London
ln Bus T.T.	-0.0183	-0.0162		
	(-2.23)**	(-1.84)*		
ln Bus T.T. * London			-0.0538	
			(-1.42)	
ln Bus T.T. * Dense Urban			-0.0665	-0.0650
			(-2.78)**	(-2.67)**
ln Bus T.T. * Other Urban			-0.0241	-0.0222
			(-1.72)*	(-1.54)
ln Bus T.T. * Rural			-0.0063	-0.0054
			(-0.63)	(-0.51)
ln Car T.T.	0.0096	0.0088	0.0084	0.0070
	(1.36)	(1.09)	(1.18)	(0.86)
ln Population aged 16-64	0.5151	0.5080	0.5163	0.5096
	(28.16)**	(26.26)**	(28.24)**	(26.35)**
ln Population/SqKm	0.2926	0.2904	0.2911	0.2706
	(2.58)**	(1.98)**	(2.56)**	(1.85)*
ln % Ethnic min.	-0.0034	-0.0027	-0.0037	-0.0030
	(-1.11)	(-0.86)	(-1.20)	(-0.97)
ln % Male	0.2325	0.2327	0.2362	0.2362
	(2.84)**	(2.73)**	(2.89)**	(2.77)**
ln % NVQ3+ aged 16-64	0.0497	0.0447	0.0517	0.0471
	(3.25)**	(2.73)**	(3.38)**	(2.88)**
ln % GCSE+ aged 16-64	0.0166	0.0162	0.0168	0.0166
	(2.12)**	(1.92)*	(2.15)**	(1.97)**
ln % Public Employees	-0.0266	-0.0285	-0.0262	-0.0280
	(-3.99)**	(-4.07)**	(-3.93)**	(-4.00)**
ln % Elementary Occup.	0.0008	-0.0002	0.0006	-0.0006
	(0.19)	(-0.05)	(0.13)	(-0.13)
2008 dummy	0.0250	0.0241	0.0248	0.0235
	(6.06)**	(5.09)**	(6.02)**	(4.96)**
2009 dummy	0.0069	0.0070	0.0063	0.0060
	(1.87)*	(1.68)*	(1.71)*	(1.44)
2010 dummy	-0.0027	-0.0020	-0.0027	-0.0022
	(-1.04)	(-0.73)	(-1.04)	(-0.81)
Constant	2.2718	2.4306	2.2585	2.5193
	(2.84)**	(2.49)**	(2.82)**	(2.58)**
Observations	1290	1162	1290	1162
R^2	0.5064	0.4921	0.5101	0.4954

t statistics in parentheses
Notes: Fixed effects are for 324 Local Authority Districts.
Travel Times (T.T.) are to employment areas with more than 5000 workers.
* $p < 0.10$, ** $p < 0.05$

Table 5-2: Fixed effects regressions on ln(% Employment Rate 16-24 age group), 2008-2011

	(1) With London	(2) Without London	(3) With London	(4) Without London
ln Bus T.T.	0.0450	0.0329		
	(1.01)	(0.70)		
ln Bus T.T. * London			0.3181	
			(1.55)	
ln Bus T.T. * Dense Urban			-0.0616	-0.0645
			(-0.47)	(-0.49)
ln Bus T.T. * Other Urban			-0.0429	-0.0470
			(-0.56)	(-0.61)
ln Bus T.T. * Rural			0.0877	0.0859
			(1.62)	(1.52)
ln Car T.T.	0.0166	0.0170	0.0090	0.0095
	(0.43)	(0.39)	(0.24)	(0.22)
ln % Ethnic min.	-0.0492	-0.0471	-0.0512	-0.0487
	(-2.96)**	(-2.78)**	(-3.08)**	(-2.88)**
ln % Male	-0.1943	-0.2301	-0.2075	-0.2367
	(-0.45)	(-0.51)	(-0.48)	(-0.53)
ln % NVQ3+ aged 16-64	0.2255	0.2402	0.2281	0.2454
	(3.02)**	(3.05)**	(3.05)**	(3.11)**
ln % GCSE+ aged 16-24	-0.0151	-0.0226	-0.0191	-0.0244
	(-0.69)	(-0.97)	(-0.86)	(-1.05)
ln % Public Employees	-0.0799	-0.0960	-0.0777	-0.0942
	(-2.21)**	(-2.54)**	(-2.15)**	(-2.49)**
ln % Elementary Occup.	0.1210	0.1301	0.1199	0.1296
	(5.33)**	(5.49)**	(5.27)**	(5.47)**
2008 dummy	0.1321	0.1275	0.1299	0.1262
	(7.17)**	(6.57)**	(7.04)**	(6.50)**
2009 dummy	0.0648	0.0581	0.0614	0.0550
	(3.63)**	(3.09)**	(3.43)**	(2.91)**
2010 dummy	0.0094	0.0041	0.0077	0.0033
	(0.72)	(0.30)	(0.59)	(0.24)
Constant	3.7298	3.9250	3.7650	3.9660
	(2.15)**	(2.19)**	(2.17)**	(2.21)**
Observations	1290	1162	1290	1162
R^2	0.1533	0.1557	0.1574	0.1586

t statistics in parentheses
Notes: Fixed effects are for 324 Local Authority Districts.
Travel Times (T.T.) are to employment areas with more than 5000 workers.
* $p < 0.10$, ** $p < 0.05$
Population density was dropped as it was highly insignificant.

We experimented with the following variables as possible instruments for accessibility:

1. Concessionary fares: It can be argued that concessionary fares reimbursement positively affects bus service levels but does not directly affect employment levels
2. Population density: It can be argued that population density has an impact on transport service levels but little influence directly on employment (other than as a proxy for agglomeration)
3. BSOG (Bus Service Operators Grant) reimbursements were also considered but rejected as they are related to service levels so driven by bus accessibility rather than an explanatory variable of bus accessibility

We found Concessionary fares and population density were weak instruments, insofar as the fitted instrumented variable was a poor proxy for the original variable they were instrumenting.

5.1.2 Testing for Serial Correlation

Here we re-run the model estimated in Table 5-1 and Table 5-2 with an Autoregressive Error Structure of 1 lag, an AR(1) process, which enables us to test whether such a lag structure is appropriate, and what impact this has on the parameter estimates. Because of the limited number of time observations, we are gaining efficiency in our estimates by including a more appropriate dynamic error structure, but losing precision from 1 year less of data which is sacrificed in order to include a lagged term.

Based on the test statistics[12] reported in Table 5-3 and Table 5-4 all models demonstrate positive serial correlation at the 5% level of significance. This suggests that a dynamic model structure is appropriate.

The coefficients on bus travel times and the interactions are very similar in magnitude, but the lower number of observations coupled with the modelling of the dynamic process means the accompanying t-statistics are lower than for the static model results in Table 5-1. Interestingly car travel times now become more (positively) significant, but still not so at the 5% level. Whilst the coefficient on population is very similar, the coefficient on population density is much higher and actually more significant in the dynamic model. Ethnicity becomes much less significant. The male coefficient falls slightly and has lower t-statistics. The education and training coefficients are of a similar magnitude to those in Table 5-1 but

[12] This involves testing the hypothesis of $\rho = 0$, (where ρ is the parameter on the lagged error) in a first-order autoregressive process produces test statistics with extremely complicated distributions. Bhargava, Franzini, and Narendranathan (1982) extended the Durbin–Watson statistic to the case of balanced, equally spaced panel datasets. Baltagi and Wu (1999) modify their statistic to account for unbalanced panels with unequally spaced data. In the same article, Baltagi and Wu (1999) derive the locally best invariant test statistic of $\rho = 0$. We report these two test statistics, Bhargava et al. Durbin-Watson statistic and the Baltagi-Wu test.
Generating these test statistics and the appropriate critical values is not straightforward. Bhargava et al (1982, Table I) gives us the closes measure of critical values where they have T=6, Obs=1000 and regressors (n)= 13 or 15, while we have T=4, Obs 966 to 870 and regressors (n)= 12 to 16. Their critical values are: $d_L = 1.952$ and $d_U = 1.963$, thus in every regression these test statistics indicate we have error autocorrelation.
For the Baltagi-Wu (1999) LBI (where LBI stands for Locally Best Invariant) test for unbalanced panels, our test statistics in the AR(1) version of table 2 (employment level) are interpreted as t-statistics, in which case all models again demonstrate positive serial correlation at the 5% level of significance.

estimated with lower t-statistics. The 2009 dummy is now positively significant in all specifications.

Table 5-3: Fixed effects regressions on ln(Employment Level), 2008-2011, with AR(1) errors.

	(1) With London	(2) Without London	(3) With London	(4) Without London
ln Bus T.T.	-0.0175	-0.0168		
	(-1.74)*	(-1.55)		
ln Bus T.T. * London			-0.0377	
			(-0.88)	
ln Bus T.T. * Dense Urban			-0.0562	-0.0555
			(-2.00)**	(-1.93)*
ln Bus T.T. * Other Urban			-0.0010	-0.0004
			(-0.06)	(-0.02)
ln Bus T.T. * Rural			-0.0174	-0.0177
			(-1.47)	(-1.43)
ln Car T.T.	0.0145	0.0149	0.0141	0.0145
	(1.74)*	(1.60)	(1.68)*	(1.55)
ln Population aged 16-64	0.5157	0.5115	0.5163	0.5120
	(24.18)**	(22.62)**	(24.19)**	(22.63)**
ln Population/SqKm	0.6524	0.6534	0.6562	0.6573
	(8.55)**	(8.08)**	(8.57)**	(8.10)**
ln % Ethnic min.	0.0000	0.0007	0.0007	0.0014
	(0.00)	(0.16)	(0.18)	(0.33)
ln % Male	0.1725	0.1905	0.1709	0.1892
	(1.80)*	(1.89)*	(1.78)*	(1.87)*
ln % NVQ3+ aged 16-64	0.0544	0.0470	0.0536	0.0462
	(2.90)**	(2.31)**	(2.85)**	(2.27)**
ln % GCSE+ aged 16-64	0.0168	0.0145	0.0169	0.0148
	(1.86)*	(1.48)	(1.87)*	(1.50)
ln % Public Employees	-0.0421	-0.0439	-0.0418	-0.0434
	(-5.18)**	(-5.10)**	(-5.13)**	(-5.04)**
ln % Elementary Occup.	0.0066	0.0072	0.0061	0.0067
	(1.34)	(1.38)	(1.23)	(1.27)
2009 dummy	0.0072	0.0073	0.0073	0.0074
	(2.39)**	(2.28)**	(2.40)**	(2.30)**
2010 dummy	-0.0001	0.0002	0.0001	0.0004
	(-0.04)	(0.08)	(0.06)	(0.15)
Constant	0.1407	0.2775	0.1225	0.2549
	(1.16)	(2.27)**	(0.98)	(2.02)**
Bhargava et al. Durbin-Watson	1.5543125	1.5631107	1.5640709	1.573511
Baltagi-Wu LBI	2.162092	2.1645034	2.1667122	2.1692339
Observations	966	870	966	870
R^2				

t statistics in parentheses
Notes: Fixed effects are for 324 Local Authority Districts.
Travel Times (T.T.) are to employment areas with more than 5000 workers.
* $p < 0.10$, ** $p < 0.05$

Table 5-4: Fixed effects regressions on ln(Official % Employment Rate 16-24 age group), 2008-2011, with AR(1) errors.

	(1) With London	(2) Without London	(3) With London	(4) Without London
ln Bus T.T.	0.0778	0.0617		
	(1.32)	(0.98)		
ln Bus T.T. * London			0.2742	
			(1.10)	
ln Bus T.T. * Dense Urban			-0.1196	-0.1260
			(-0.73)	(-0.75)
ln Bus T.T. * Other Urban			0.0883	0.0776
			(0.86)	(0.74)
ln Bus T.T. * Rural			0.0916	0.0836
			(1.31)	(1.15)
ln Car T.T.	0.0001	0.0075	-0.0043	0.0031
	(0.00)	(0.14)	(-0.09)	(0.06)
ln Population/SqKm	1.3620	1.3566	1.4097	1.3956
	(3.39)**	(3.23)**	(3.46)**	(3.29)**
ln % Ethnic min.	-0.0501	-0.0478	-0.0489	-0.0464
	(-2.15)**	(-2.01)**	(-2.09)**	(-1.95)*
ln % Male	-0.8645	-0.8388	-0.8537	-0.8348
	(-1.53)	(-1.42)	(-1.51)	(-1.41)
ln % NVQ3+ aged 16-64	0.1366	0.1651	0.1334	0.1649
	(1.35)	(1.53)	(1.32)	(1.53)
ln % GCSE+ aged 16-24	-0.0392	-0.0445	-0.0405	-0.0440
	(-1.38)	(-1.49)	(-1.42)	(-1.47)
ln % Public Employees	-0.1198	-0.1428	-0.1168	-0.1409
	(-2.47)**	(-2.80)**	(-2.40)**	(-2.76)**
ln % Elementary Occup.	0.1255	0.1369	0.1223	0.1335
	(4.28)**	(4.46)**	(4.15)**	(4.33)**
2009 dummy	0.0578	0.0531	0.0574	0.0523
	(2.97)**	(2.61)**	(2.92)**	(2.55)**
2010 dummy	0.0165	0.0109	0.0161	0.0110
	(1.23)	(0.77)	(1.19)	(0.78)
Constant	-2.0820	-1.8161	-2.4352	-2.0565
	(-1.70)*	(-1.50)	(-1.86)*	(-1.63)
Bhargava et al. Durbin-Watson	1.7834162	1.7864897	1.7989674	1.7938046
Baltagi-Wu LBI	2.3345735	2.3371169	2.3444712	2.3406586
Observations	966	870	966	870
R^2				

t statistics in parentheses
Notes: Fixed effects are for 324 Local Authority Districts.
Travel Times (T.T.) are to employment areas with more than 5000 workers.
Variable PopPerSqKm omitted because of statistical insignificance.
* $p < 0.10$, ** $p < 0.05$

In the regression results for 16-24 year olds, the impact of bus and car travel time remains insignificant in each specification and interaction. Given this, further comment on the other covariates is not necessary.

5.2 Cross sectional Model

5.2.1 Basic OLS Model Results

Table 5-5 reports the results of the OLS regression (with Fixed Effects for LADs) on the natural log of employment levels. The first two columns are for employment of 16-74 year olds[13] and the final two columns for 16-24 year olds. Whilst we have chosen similar explanatory variables to the panel data analysis, the census allowed us to also include a measure of those with No Car Availability.

The elasticity of employment with respect to bus travel times from the first two regressions in Table 5-5 are -0.022 and -0.015 with and without London MSOAs respectively and are both the expected sign and significant at the 5% level. The similarity with the results from the Panel data analysis is striking and it suggests that the results are robust to the choice of data source and modelling methodology.

These coefficients indicate that a 10% reduction in bus travel times leads to, all else equal, a 0.15-0.22% increase in employment.

The second two regressions show that for 16-24 year olds elasticities are slightly larger and more significant, at around -0.029. This implies that differences in public transport accessibility matters more to the employment prospects of younger people.

The proportion of those with No Car Availability has a significant and negative impact on employment. This variable is highly correlated with the socio economic makeup of an area, with higher social class more likely to have access to a car. This variable also varies systematically with urban density – people in denser urban areas are less likely to need a car, or more likely to be put off owning a car due to parking issues and congestion.

The results for the other socio-economic covariates are broadly as expected. Employment is positively and significantly related to population. The impact of the proportion of those from ethnic minorities is insignificant for the 16-74 year olds but significant and negative for the younger cohort. The proportion of those with English as a first language has a significant and positive impact on employment. As well as being important influences in themselves, these two variables may also be correlated with other socio-economic variables that might influence employment outcomes.[14]

Other variables that captured various socio-economic aspects such as qualifications, social class and measures of deprivation, as reported in *Table 3-4*, were dropped as they were found these to have very high correlation coefficients in excess of 0.9 with other variables in the regression. These dropped variables would otherwise give rise to the problem of ‘collinearity’ and lead to counter-intuitive signs on estimated coefficients.

[13] The 16-74 measure was reported in the Census rather than the 16-64 value we used in the panel data

Table 5-5: OLS estimation on ln(Employed) in 2011

	(1) Aged 16-74, incl. London	(2) Aged 16-74, excl. London	(3) Aged 16-24, incl. London	(4) Aged 16-24, excl. London
ln Bus T.T.	-0.0223	-0.0150	-0.0288	-0.0290
	(-5.04)**	(-3.09)**	(-5.32)**	(-4.98)**
ln Car T.T.	0.0005	-0.0039	0.0024	-0.0002
	(0.11)	(-0.69)	(0.39)	(-0.02)
ln Popul. aged 16-74	0.9915	0.9730		
	(141.36)**	(128.30)**		
ln Popul. aged 16-24			0.8256	0.8198
			(148.22)**	(140.58)**
ln % No Car Availiabilty	-0.0920	-0.0944	-0.0400	-0.0392
	(-27.43)**	(-26.82)**	(-9.46)**	(-9.02)**
ln % Ethnic minority	-0.0043	0.0061	-0.0424	-0.0300
	(-1.34)	(1.75)*	(-10.68)**	(-7.12)**
ln % Econ active males	-0.0864	-0.1478	-0.0147	0.0021
	(-1.82)*	(-2.66)**	(-0.25)	(0.03)
ln % English 1st lang.	0.5153	0.5881	0.4005	0.5388
	(17.61)**	(15.66)**	(11.11)**	(11.95)**
Constant	-2.0300	-1.9826	-0.8815	-1.5561
	(-6.82)**	(-5.48)**	(-2.45)**	(-3.65)**
Observations	6635	5706	6635	5706
R^2	0.8210	0.8126	0.8414	0.8459

t statistics in parentheses
Notes:
326 LAD dummies included in regressions with London.
294 LAD dummies included in regressions without London.
Travel times (T.T.) are for work areas of 5000 plus.
* $p < 0.10$, ** $p < 0.05$

5.2.2 IV Model Results

Table 5-6 shows the results from our IV estimation using two-stage least squares (2SLS). In these second stage regressions, we include estimated measures of (ln) Bus Travel times based on first stage OLS regressions with these as a function of the explanatory variables and instruments.

We experimented with various choices of instruments, including population density and various lags of bus and car travel times (and squared values of these lags). The literature suggests longer lags of the endogenous variables are more likely to be suitable instruments as any direct effect of the current explanatory variable would be less likely. We had measures of travel times for 2007, 2009 and 2010. Although we found 2007 travel times to be a good instrument of 2011 travel time we also discovered that it had originally been constructed using a different methodology to subsequent years and we were therefore unable to use it This left us with 2009 and 2010 travel times. It was felt that 2010 was too recent. We found that using lagged car travel times and population density led to poor results as these were jointly weak instruments probably because they are highly correlated with one another. The IV results report in Table 5-6 are based on the use of a single instrument, bus travel times in 2009 which is then subjected to exogeneity tests.

Table 5-6: IV on ln(Employed) in 2011

	(1) Aged 16-74, incl. London	(2) Aged 16-74, excl. London	(3) Aged 16-24, incl. London	(4) Aged 16-24, excl. London
fit(ln Bus T.T.)	-0.0336	-0.0238	-0.0395	-0.0414
	(-5.25)**	(-3.33)**	(-5.04)**	(-4.84)**
ln Car T.T.	0.0091	0.0029	0.0105	0.0094
	(1.48)	(0.42)	(1.39)	(1.14)
ln Popul. aged 16-74	0.9901	0.9721		
	(140.61)**	(127.81)**		
ln Popul. aged 16-24			0.8249	0.8190
			(147.64)**	(140.08)**
ln % No Car Availability	-0.0936	-0.0955	-0.0413	-0.0406
	(-27.40)**	(-26.67)**	(-9.63)**	(-9.22)**
ln % Ethnic minority	-0.0051	0.0054	-0.0431	-0.0308
	(-1.59)	(1.55)	(-10.81)**	(-7.28)**
ln % Econ active males	-0.0897	-0.1464	-0.0181	0.0040
	(-1.89)*	(-2.63)**	(-0.31)	(0.06)
ln % English 1st lang.	0.5167	0.5909	0.4016	0.5425
	(17.65)**	(15.71)**	(11.14)**	(12.02)**
Constant	-1.9878	-1.9755	-0.8470	-1.5517
	(-6.67)**	(-5.45)**	(-2.35)**	(-3.63)**
H0: Xs exogenous	.	.	.	.
Durbin Chi^2	6.3234	2.9646	3.7741	4.1346
Durbin prob.	0.0119	0.0851	0.0521	0.0420
Wu-Hausman Chi^2	6.0108	2.8091	3.5862	3.9186
Wu-Hausman prob.	0.0142	0.0938	0.0583	0.0478

t statistics in parentheses
Notes:
326 LAD dummies included in regressions with London.
294 LAD dummies included in regressions without London.
Travel times (T.T.) are for work areas of 5000 plus.
ln Bus T.T. instrumented using: lnBusTT5k_2009
* $p < 0.10$, ** $p < 0.05$

The use of the 2009 lagged bus travel times as an instrument yielded a good estimate of 2011 bus travel times. The resulting parameters for (fitted) bus travel times have the right sign, are significant and are approximately 50% higher than those from the Fixed Effects panel regressions. Interestingly this suggests that the OLS elasticities may be under-estimates of the true relationship between travel times and employment. However, when testing for the exogeneity of the other explanatory variables, we find that for models 2 and 3 the Durbin and Wu-Hausman tests do not reject the null hypothesis that these explanatory variables are exogenous[15].

Parameter estimates for the remaining co-variates are similar to those in the OLS model.

[15] The use of 1 instrument precludes us from estimating the Sargan test.

5.2.3 Segmented models

We looked at alternative versions of the model with urban form segmentations in Table 5-1. Experimentations on these segmentations of the travel times by urban form were not successful. Given the large number of observations in the data we decided to estimate the elasticities through separate models for each urban area. The results are reported in Table 5-7. They show the highest elasticity for London at -0.031 with dense urban areas at -0.025. Elasticities for other urban and rural areas are lower, but just insignificant at the 10% level.

Table 5-7: OLS estimation on ln(Employed) in 2011 - models for each area (16-74 year olds)

	(1) Aged 16-74, London	(2) Aged 16-74, Dense Urban	(3) Aged 16-74, Other Urban	(4) Aged 16-74, Rural
ln Bus T.T.	-0.0305	-0.0253	-0.0150	-0.0078
	(-3.31)**	(-2.08)**	(-1.56)	(-1.53)
ln Car T.T.	0.0260	0.0293	-0.0107	-0.0131
	(2.76)**	(1.83)*	(-0.84)	(-2.46)**
ln Popul. aged 16-74	1.1233	0.9297	0.9189	1.0253
	(70.20)**	(50.17)**	(59.15)**	(133.47)**
ln % No Car Access	-0.0316	-0.1437	-0.1275	-0.0576
	(-2.81)**	(-16.26)**	(-17.93)**	(-15.34)**
ln % Ethnic minority	-0.1214	0.0162	-0.0181	0.0044
	(-10.77)**	(2.26)**	(-2.58)**	(1.05)
ln % Econ active males	0.1272	0.4418	-0.0853	-0.0584
	(1.44)	(3.43)**	(-0.76)	(-0.95)
ln % English 1st lang.	0.2579	1.0011	0.2926	0.1755
	(5.31)**	(12.20)**	(4.45)**	(2.66)**
Constant	-2.6070	-5.7953	-0.2644	-1.0645
	(-4.66)**	(-6.74)**	(-0.39)	(-2.39)**
Observations	951	1268	1876	2540
R^2	0.8875	0.7572	0.7540	0.9072

5.2.4 Interactions Model

As another approach, we derived variable elasticities for different areas by interacting car availiability and bus travel time. The results are reported in Table 5-8. The reasoning is that the impact of changes in travel times is likely to be higher in areas with low levels of car availability. We then derive elasticities across the range of levels of non-car access. We also calculate elasticities for the different urban forms based on the average level of non-car access in each of these areas. These interactions are found to be very significant across the 4 reported models in Table 5-8.

Table 5-8: OLS estimation on ln(Employed) in 2011, with car non availability interactions

	(1) Aged 16-74, incl. London	(2) Aged 16-74, excl. London	(3) Aged 16-24, incl. London	(4) Aged 16-24, excl. London
ln Bus T.T.	0.0500	0.0567	-0.0027	-0.0039
	(10.08)**	(10.54)**	(-0.45)	(-0.61)
ln Bus T.T. * ln N.C.A.	-0.0220	-0.0224	-0.0075	-0.0074
	(-22.61)**	(-21.97)**	(-6.24)**	(-6.05)**
ln Car T.T.	-0.0010	-0.0054	0.0013	-0.0012
	(-0.20)	(-0.94)	(0.21)	(-0.18)
ln Popul. aged 16-74	0.9895	0.9722		
	(138.61)**	(125.66)**		
ln Popul. aged 16-24			0.8202	0.8144
			(147.82)**	(140.19)**
ln % Ethnic minority	-0.0061	0.0046	-0.0438	-0.0310
	(-1.86)*	(1.29)	(-10.96)**	(-7.30)**
ln % Econ active males	-0.1116	-0.1729	-0.0132	0.0023
	(-2.29)**	(-3.04)**	(-0.22)	(0.03)
ln % English 1st lang.	0.5557	0.6434	0.4212	0.5643
	(18.75)**	(16.88)**	(11.68)**	(12.51)**
Constant	-2.3888	-2.4219	-1.0741	-1.7611
	(-7.92)**	(-6.58)**	(-2.98)**	(-4.12)**
Observations	6635	5706	6635	5706
R^2	0.8146	0.8051	0.8402	0.8446

t statistics in parentheses
Notes:
326 LAD dummies included in regressions with London.
294 LAD dummies included in regressions without London.
Travel times (T.T.) are for work areas of 5000 plus.
* $p < 0.10$, ** $p < 0.05$

Because of the interaction term, the resulting elasticities are not easy to infer from the parameter estimates in Table 5-8. Figure 5-1 thus illustrates how the elasticity of employment with respect to changes in bus travel times varies depending on the level of non-car availability in each MSOA. We have plotted the estimated values from model 1 based on the interquartile range of non-car availability values, 13% to 34%. This shows that the employment elasticity with respect to bus travel time works in the expected way and is higher (in absolute terms) for areas with higher levels of non-car availability, i.e. where more people are captive to public transport modes. This function allows us to derive elasticities for the different urban forms outlined in section 3.2 based on the average level of non-car availability in each of these areas, as shown in Table 5-9.

Figure 5-1: Employment elasticity and Bus travel time from Table 5-8

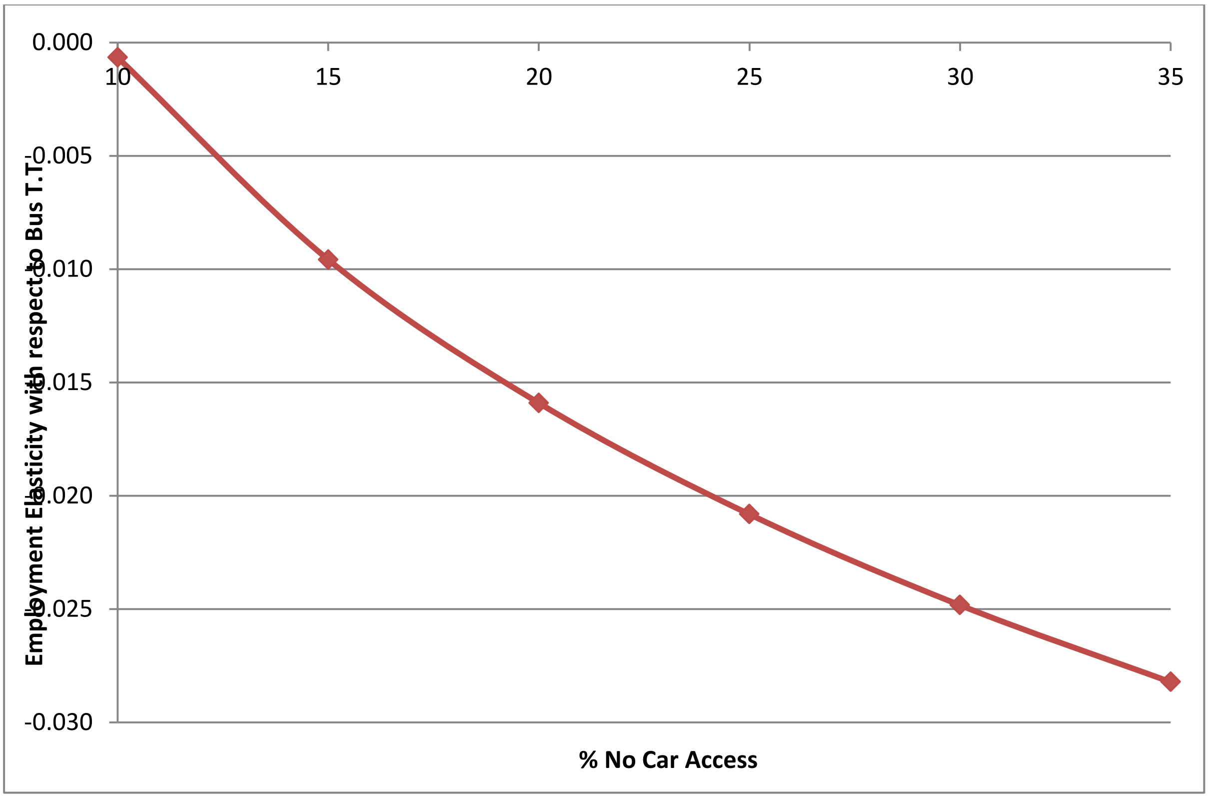


Table 5-9: Derived travel time elasticities

Urban Type	**Average NCA (%)**	**Derived employment elasticity from Table 5-8)**	**Estimated elasticity from** Table 5-7
London	40.36	-0.031##	-0.031**
Dense Urban	29.82	-0.025##	-0.025**
Other Urban	22.83	-0.019##	-0.015
Rural	11.70	-0.004##	-0.008

Notes:
LAD dummies included in all regressions.
$p < 0.05$
- elasticities estimated based on significant parameters for travel time and travel time and car availability interaction

The similarity between these two sets of elasticities is striking, especially for London and Dense Urban areas where they are significant.

5.3 Employment elasticities for Car Travel Time

The estimated models indicate that car travel time is seldom important in determining employment levels. In other words, the estimated parameters are small, sometimes positive or negative, and often statistically close to zero. The explanation for this might be twofold, it might be capturing a true feature of car travel time and/or it might be a consequence of the available data.

At the margin, extra car travel time might not have a large impact on employment levels. In other words, extra minutes of car travel might not significantly influence employment decisions. In part, this might be because one's own vehicle can be seen as a comfortable environment to spend time in and because of the sunk costs of car ownership. Another possibility is that car ownership is lower amongst the unemployed, thus car travel times are less of a factor for those at the margins of employment. The unimportance of car travel time on employment could also be an artefact of the data. Car travel times have been calculated from traffic speed data, with an assumption of a minimum of five minutes from the door to the start of the journey, no time for parking and no time at junctions. Bus travel times include walk times to and from bus stops, plus waiting times between busses. As a result, car travel times may be artificially low and show very little variation. A model requires variability in order to explain any underlying relationships.

The bus and car travel times are measured on a range from 5 minutes to 120 minutes. Table 5-10 shows that in the cross-section dataset 58.9% of car journeys take less than 10 minutes while for bus travel this proportion is only 6.6%. These values are similar in the panel dataset, 61.5% of car journeys and 0.6% of bus journeys take less than 10 minutes. This lack of variation in measured car journey times is also reflected in the low standard deviations for these variables. The standard deviations (measure of dispersion) for car travel times are 9.0 and 10.4 for the cross-section and panel datasets respectively. The corresponding standard deviations for bus travel times are much higher, at 24.9 and 23.5 for the cross-section and panel datasets respectively.

Table 5-10: Distribution of Travel times across the two datasets

	% Less than 6 minutes	***%Less than 10 minutes***	***%Less than 15 minutes***
Cross Sectional Bus Travel time	0.3	6.6	23.4
Cross Sectional Car Travel time	32.2	58.9	79.4
Panel Data Bus Travel time	0.3	0.6	6.4
Panel Data Car Travel time	29.8	61.5	80.4

6 SUMMARY AND RECOMMENDED VALUES

6.1 Summary

The work reported on here uses econometric models to analyse the impact of changes and differences in bus accessibility on the labour market.

We analyse a panel dataset and a cross-sectional dataset of bus accessibility indicators, labour market indicators and socio demographic information to examine effect of differences in public transport (primarily bus) journey times on Local Authority District (Panel) and MSOA (Cross-section) areas' labour market outcomes). We examine different model specifications and discuss the use of methods to establish the direction of causation in the relationship between bus accessibility and employment.

Our results add to the existing literature on labour supply elasticity/employment sensitivity within the spirit of the current WebTAG framework.

The outcomes of our model estimates using the panel data and the cross section census data lead us to the following conclusions:

- Across both our datasets, we found a statistically significant and negative relationship between public transport travel time accessibility and employment, which varies in magnitude by urban type and level of car availability. In most cases, the proportional responses (elasticities) in the cross section data were similar to those found for the panel data. Our models appear plausible in terms of signs and magnitudes for all estimated coefficients. We take this consistency between datasets to be an indication of robustness in the results.

- For the panel data, we were able also to estimate models, which were segmented by urban form. These indicate that the denser an urban area is, the more susceptible employment is to changes in bus travel times.

- For the panel data, the small number of years limited the ability to investigate the issue of causation, as we could not apply a suitable instrument for public transport travel times. However, the Fixed Effects approach will have removed any time-invariant causes of endogeneity.

- We found some evidence of serial correlation in the panel model suggesting a dynamic approach using lagged terms is appropriate. However, the four year panel is very short and a quarter of these observations were lost through using a dynamic model.

- In the cross section data, we were able to investigate the use of instrumental variable (IV) estimation to control for possible endogeneity between employment and public transport travel times. The results actually indicate OLS estimates may be under-reporting the effect of differences in travel times on employment levels. However, the tests for rejection of OLS were equivocal suggesting that instrumenting might not be necessary. Given this and the inherent dangers in using IV approaches, we prefer the OLS estimates.

- With the cross section data, we estimated separate employment elasticities for urban forms directly from regression models for each urban area.. However, these were slightly insignificant for Other Urban and Rural areas. Using bus travel times and an interaction term between non-car access and bus travel times (which were both highly significant) were also able to impute different elasticities, which were highest for London and lowest for Rural Areas. They were also very similar to the directly estimated parameters.

- We cannot impute labour supply elasticities. Whilst we could impute the effect on net wages of those in or entering the labour market who use bus services, we do not observe the number of people using bus services following an improvement in travel times. This would require more consideration and utilisation on data on modal market shares. Our parameters thus conflate the impact of different urban types (and/or levels of non-car availability) and any differences in elasticities of the overall sensitivity of employment to changes in bus travel times.

6.2 Recommended Values

Table 6-1 summarises the derived elasticities from our different models. The table highlights the consistency in the results for England as a whole, with or without London. The AR(1) results which control for Serial Correlation are very similar to the FE results, but with lower levels of significance. The Cross Section OLS results are smaller than the IV results. The reasoning behind both these findings was outlined in Section 5.

Table 6-1: Estimated Bus travel time elasticities across the models

	Panel Data FE	*Panel Data AR(1)*	*Cross Section OLS*	*Cross Section OLS Urban Form Models*	*Cross Section IV*
Source	Table 5-1	Table 5-3	Table 5-5 & Table 5-8	Table 5-7	Table 5-6
England inc. London	-0.0183**	-0.0175*	-0.0223**		-0.0336**
England exc. London	-0.0162*	-0.0168	-0.0150**		-0.0238**
London	-0.0538	-0.0377	-0.0314**#	-0.0305**	
Dense Urban/ Conurbation	-0.0650**	-0.0555*	-0.0247**#	-0.0253**	
Other Urban	-0.0222	-0.0004	-0.0188**#	-0.0150	
Rural	-0.0054	-0.0177	-0.0041**#	-0.0078	

#These values are derived from the coefficients on bus travel time and the interaction with levels of car availability so do not have separate significance levels.
**Indicates significance at the 5% level; *Indicates significance at the 10% level

These values can be interpreted as follows: as an example, the elasticity from Cross Sectional regression including London was -0.0223. This suggests that for a 10% decrease in bus travel times across England, we would expect a .22% increase in employment, amounting to over 50,000 extra jobs (based on an employment level in England of around 25 million). For an average Dense Urban conurbation (with a level of non-car availability of 30%), we would estimate a higher elasticity from which we would expect a .25% increase in employment.

Overall, we prefer the results derived from the Cross Section OLS models in Table 5-5 and Table 5-8. These results are based on more observations and are more robustly estimated parameters than the Panel Data results and the separate Cross Section OLS models of Table 5-7. Although the models in Table 5-8 did not directly estimate separate elasticities for the urban types, they control for the level of car availability. Apart from the insignificant (and small) values in Rural areas, the results from Table 5-8 and Table 5-7 are remarkably similar.

Having investigated IV estimation, we found endogeneity is not an issue in all models and indeed it leads to higher estimates so the OLS results can be seen as conservative values.

7 REFERENCES

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

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

Bhargava, A., Franzini, L. and Narendranathan, W.: 1982, Serial correlation and the fixed effects model, *Review of Economic Studies* **49**(4), 533–49. 1, 2, 3, 5, 10

Department for Transport (2012), Accessibility Statistics 2012, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/66795/accessibility-statistics-guidance.pdf

Department for Transport (2014), Transport Analysis Guidance Unit A2.1, https://www.gov.uk/government/publications/webtag-tag-unit-a2-1-wider-impacts

DEFRA (2009), Defra Classification of Local Authority Districts and Unitary Authorities in England, http://archive.defra.gov.uk/evidence/statistics/rural/documents/rural-defn/LAClassifications_technicalguide.pdf

Greene, W.G., Econometric Analysis, Third edition, Prentice Hall International, ISBN 0-13-724659-5

Mackie, P., Laird, J. and Johnson D., (2012), *Buses and Economic Growth Final Report*, Greener Journeys, www.greenerjourneys.com/wp-content/uploads/2012/06/BusesEconomicGrowth_final-report.pdf

Standing Advisory Committee On Trunk Road Assessment (SACTRA), (1999) *Transport and the Economy*. Norwich:TSO.

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