Henry G. Overman, Stephen Gibbons, Sabine D'Costa, Giordano Mion, Panu Pelkonen, Guilherme Resende and Mike Thomas

Strengthening economic linkages between Leeds and Manchester: feasibility and implications: full report

Report

Strengthening Economic Linkages between Leeds and Manchester: Feasibility and Implications

Full Report

November 2009

Moving Forward:
The Northern Way

one NORTH EAST

Yorkshire Forward The Region's Development Agency

Northwest REGIONAL DEVELOPMENT AGENCY

SERC Spatial Economics Research Centre

This research programme was delivered by the **Spatial Economics Research Centre (SERC)** and was commissioned and sponsored by The Northern Way.

The SERC team, based at the **London School of Economics** comprised:

Henry Overman (LSE and SERC)
Steve Gibbons (LSE and SERC)
Sabine D'Costa (LSE and SERC)
Giordano Mion (LSE and SERC)
Panu Pelkonen (LSE and SERC)
Guilherme Resende (LSE and SERC)
Mike Thomas (LSE and SERC)

A Steering Group supported the implementation of the research programme, and policy implications were informed by discussions at a Policy Reference Group.

The following contributed to the work of these groups:

Department of Business, Innovation and Skills:
Adrien Amzallag, Andrew Cunningham-Hughes

Department of Communities and Local Government:
Daniel Thornton, Cathy Francis, Sarah James

Leeds City Region:
Matt Brunt, Rob Norrys

Manchester City Region:
Baron Frankel, Juan Gomez, Rupert Greenhalgh

North West Development Agency:
Damian Bourke, Nidi Etim

The Northern Way:
Andrew Lewis, John Jarvis, Richard Baker

Yorkshire Forward:
Nicky Denison, Simon Foy, Andrew Lowson

Independent Academic Advisor:
Professor Alan Harding, IPEG,
University of Manchester

The Northern Way
Stella House, Goldcrest Way, Newburn Riverside,
Newcastle upon Tyne NE15 8NY
Telephone: 0191 229 6200
Website: www.thenorthernway.co.uk

 We thank Lizze Diss at the Department for Transport and Dan Graham at Imperial College London for their help with the ward to ward GTC for driving. We thank We thank Peter Wiener at Steer Davies Gleeve and John Jarvis at Yorkshire Forward for advice on constructing the train counterfactuals.

Contents

1. Introduction

This report describes research by the Spatial Economics Research Centre that aims to understand economic integration and interaction between the Manchester and Leeds city-regions. As well as analyzing current patterns, the research assesses the possible economic impacts of increased integration.

The research was commissioned by The Northern Way, as a contribution to its Policy and Research programme, to provide robust evidence about the economic relationships between the two city-regions and to assess:

- economic opportunities which could accrue from closer links; to the two city-regions, other Northern territories and the wider UK
- risks, either in terms of adverse impacts on the economy of one of the two centres, or impacts on surrounding territories
- the potential and feasibility for public policy to stimulate and encourage such relationships.

The research involved a number of complementary projects, and it was undertaken between July 2008 and November 2009. Facilitated by The Northern Way, the project was supported by representatives from the two city-regions, Yorkshire Forward and the North West Development Agency, Government Departments and independent academic advisors.

This report, describes the detailed findings and methodology. Alongside a summary of findings and policy conclusions, it is available on the website of The Northern Way at www.thenorthernway.co.uk/leedsmanchester and SERC at www.spatialeconomics.ac.uk.

2. Background to the research

There is increasing interest in the role of cities in driving economic growth and development

An immediate focus is on the role cities may play in recovery from the current recession. However, beyond this the importance of cities to the economy and thus to economic policy is increasingly recognized at both national and international levels. In the UK, this increased interest reflects the fact that, after a long period of relative decline, a significant number of English cities have experienced improved economic performance (ODPM, 2006). At the same time, evidence about underlying structural changes, suggests there may be potential for continued long term growth in these cities.

In particular, if the UK economy continues its inexorable move from manufacturing to services, this will have important implications for continued growth in cities. There is a large body of evidence which suggests that producers of services benefit in a variety of ways when they locate in cities. Crucially, the benefits of this agglomeration appear to be larger for service producers than for manufacturers. A structural shift towards services, combined with the fact that services benefit more from cities, points towards a future in which more economic activity could be concentrated in a small number of larger cities.

Amongst policy makers in the UK, particularly those concerned with spatial disparities, this raises a number of important questions. Will this growth be concentrated mostly in London and the Greater South East? If so, is there anything that policy can, or should, do to counteract this? What role might future growth in Northern cities play in increasing growth in the wider northern economy? Which cities in the North might drive this growth and what, if anything, might be the appropriate role for policy? The research that we describe in this report is concerned with the last of these questions. In particular, we consider the implications and feasibility of developing stronger economic relationships between the Manchester and Leeds city-regions.

Recent reports for The Northern Way from IPEG/CUPS[1] and the Centre for Cities[2] have assembled extensive evidence describing the economic connections between Northern cities and between the Northern cities and London. This research has served to reinforce the longstanding sense within The Northern Way, and those working around it, that one of the key opportunities for the acceleration of growth available to the north of England as a whole may be the stimulation of higher levels of integration between the Manchester and Leeds economies. These cities are of particular interest because, while both cities have recently experienced strong growth, existing research finds little evidence of interaction in terms of business connections or commuting, despite their geographical proximity. Our research builds on this work to provide further evidence on the feasibility and implications of strengthening economic linkages between the Leeds and Manchester City Regions.

The fact that there is little evidence of interaction between Leeds and Manchester has led some commentators to conclude that the links between the two cities are somehow weaker than they should be and that increasing these links could play a part in improving economic performance of the Northern regions. In part, this conclusion is based on a comparison to the higher levels of interaction in other parts of the UK, in particular in London and the South East. In part it is based on international comparisons, where we observe stronger economic interactions

1. See ‘The Northern Connection’. IPEG/CUPS for The Northern Way, January 2008.

2. See ‘City Links’, Lucci & Hildreth, March 2008.

between similarly sized cities positioned close to each other. Commentators have sought to explain these weak links as arising from a number of factors including; topography (in particular the Pennines), cultural differences and poor transport connections.

In developing this research, we recognized that an analytical jump from the observation of low levels of interaction to the conclusion that integration *is weaker than it should be* is not warranted. Further evidence on the links between the two economies is needed to help assess the case for intervention and to understand whether increasing integration has any role to play in improving the economic performance of the two city regions or the wider northern economy.

To reach the conclusion that integration is weaker than expected, one needs to be able to make a comparison to an appropriate benchmark. Arguably, neither London and the Greater South East, nor a limited number of international cities provide particularly compelling comparators. Therefore, in the first stage of our analysis we revisit this issue and use regression analysis to construct more appropriate benchmarks based on observed behaviour across the whole of Great Britain. We start by considering commuting – the only "flow" between places in Great Britain for which we have reasonable data. We ask what determines commuting flows between places and, given this, whether flows between Manchester and Leeds are actually lower than expected. We then turn our attention to outcomes that are of greater policy interest, namely earnings, employment and output. Here, unlike with commuting, we are unable to directly observe the interactions between places. Instead, we consider the extent to which nearby places appear to experience similar levels of, and changes in, these outcomes. Again, we use observed behaviour across the whole of Great Britain to ask what determines these similarities and whether Manchester and Leeds are in any sense unusual.

Both these approaches essentially divide outcomes in to a part that can be explained by things that are observed about places and a residual, or unobserved, part. This distinction is of more than academic interest. If policy makers want to increase interaction between the Manchester and Leeds economies, then the appropriate policy response will depend crucially on what causes the degree of interaction to be low in the first place. Addressing cultural differences requires a different set of policies to those needed to address high travel costs. Knowing what helps explain the behaviour we observe is a good starting point for thinking about policy. It is in this sense – of trying to understand observed behaviour – that we consider questions of the "feasibility" of increasing integration between Manchester and Leeds. In the second part of the study we turn to the implications of doing so with a particular focus on the effect on economic performance in the north.

We approach this question of the implications for economic performance in two ways. The first is to view enhanced integration between Manchester and Leeds as a way of increasing the size of the local economy. A larger local economy may help firms be more productive. Such agglomeration economies – as economists refer to the beneficial effects of a larger local economy – may arise for a variety of reasons. In particular large local economies may facilitate sharing of resources (for example of large infrastructure such as airports), matching of capacity (for example of the right workers to the right firms) or learning (for example a transfer of knowledge from one firm to another). Can we say anything about the likely impact of these effects if we achieved increased integration between Manchester and Leeds? Existing work for The Northern Way has tackled this question by using estimates of the strength of agglomeration economies, coupled with assumptions on the extent to which integration would increase local economy size to work out the productivity impacts on different sectors of the economy. We use labour market data to try to understand whether this existing work captures all the likely impacts of increased integration.

There has been considerable speculation that the size of the Manchester and Leeds economies may have negative implications for labour market outcomes and that this may be an important factor in explaining their relative under performance. To examine this possibility we use data on individual wages to see how the level and growth of wages are affected by the size of the local labour market. We then assess the extent to which these benefits arise from changing composition (e.g. large cities have more educated workers) as opposed to higher wages for existing workers. We then use our estimates, coupled with realistic assumptions about policy induced changes in transport costs, to assess the impact of increased integration on labour market outcomes.

Our work on labour markets views enhanced integration between Manchester and Leeds as a way of increasing the size of the local economy and studies the impact on the structure of the economy and on wages. The method that we use, referred to as a “reduced form” approach, makes it hard to be specific about the economic channels through which these effects operate. This, in turn, means we cannot say anything precise about how these effects will impact on the wider Northern economy. In the final part of our research we examine these impacts using what economists refer to as a “structural model”. This model is very specific about the channels through which increased integration impacts productivity. We focus, in particular, on selection effects that are thought to generate a large part of the productivity increase that we observe as economies become more integrated. The strength of these selection effects depends on both the size of the local economy and the extent of integration with other local economies. This means that we can use the model, fitted to GB data, to examine how increased integration affects productivity across different economies and so get some idea of how closer Manchester-Leeds integration might affect other places in the North.

The rest of this report is structured as follows. Section 3 considers commuting flows. Section 4 considers linkages in output and employment. Section 5 considers the role of labour markets, while Section 6 outlines the findings from our structural model. Section 7 provides conclusions and considers policy issues.

3. Commuting flows

We are interested in understanding the determinants of commuting between Manchester and Leeds. We focus on commuting flows because (i) they are likely to be a very important driver of linkages between places; (ii) unlike, say, business input-output linkages, we have sufficiently detailed data to undertake analysis of the determinants. We have undertaken analysis to try to answer two related questions. First, given the overall level of commuting flows in and out of Manchester and Leeds, are the bilateral flows between Manchester and Leeds unusually high or low? Second, to what extent do characteristics of Manchester and Leeds (that is, characteristics that we can observe in the data – size, income, commuting costs) explain these patterns? To answer these questions we model data on commuting flows between Local Authority areas across Great Britain as a function of the characteristics (size, income, commuting costs) of those Local Authority areas. We then compare outcomes for Manchester and Leeds to those that we would predict based on their characteristics and the observed behaviour of commuters across Great Britain.

3.1 Gravity Model

We use a version of the 'gravity' model to explain commuting flows between places. In its most basic form, the model assumes that the degree of interaction between places depends positively on their "mass" – e.g. an area's population or employment – and negatively on the distance between places.

The gravity model has been widely used in the social sciences to study spatial interaction. In particular, it has been extensively used to examine trade between countries. See Overman, Redding and Venables (2003) and Anderson and van Wincoop (2004) for surveys.

Papers on commuting behaviour have also employed the gravity model, typically to study commutes within a single metropolitan area. Examples include studies of Washington D.C. by Levinson (1998) and the San Francisco Bay Area by Cervero and Wu (1997). Applications of the gravity model to commuting behaviour in the UK are scarce, though Coombes and Raybould (2001) investigate the regional characteristics associated with short commutes (less than 5km) in England and Wales.

In this project we use the gravity model to investigate commute patterns between Local Authority areas across the whole of Great Britain. The gravity model to explain the number of home-to-work commutes between any two areas can be expressed formally as:

$$T_{ij} = A_i B_j \exp[-\theta c_{ij}]$$

where T_{ij} denotes the number of commutes between i and j, A_i is an origin function, B_j a destination function and c_{ij} is a deterrence function[3]. The origin function captures anything specific to the origin (i.e. the point from which a commute starts) that might affect overall commuting flows to all destinations. The destination function does similarly for all destinations (i.e. the point at which a commute ends). The deterrence function captures the factors, such as distance, that might inhibit flows between locations i and j. We use an exponential deterrence function which is popular in the literature because it leads to a gravity model with a number of desirable properties. See Sen and Smith (1995) for further discussion.

3. Given that our focus is on commuting between Manchester and Leeds we ignore within Local Authority commuting flows.

Taking logs of both sides of the equation gives:

$$\ln (T_{ij}) = \ln (A_i) + \ln (B_j) - \theta c_{ij}$$

This is the version of the gravity equation that forms the basis for our empirical work.

In terms of the origin and destination characteristics, we model the number of commutes between two Local Authority areas i and j as driven by (i) the size of the Local Authority areas as measured by employment and (ii) the average wages in each Local Authority area. We also assume that, even allowing for these factors, some Local Authority areas will have high inward commutes (e.g. city centres) some high outward commutes (e.g. residential areas). Rather than trying to identify all the different factors that might cause areas to have high outflows or inflows in a given year, we just capture the effect in the model by including dummy (zero-one) variables that indicate a given origin or destination Local Authority area. These dummy variables allow the data to tell us when commutes are unusually high or low for a specific Local Authority area. We start by using straight line distances between areas as the factor that deters commutes. We then turn to more realistic measures of transport costs by road and train.

Technically, it is not possible to estimate the parameters on observable factors that are area specific (e.g. employment) -and at the same time control for area effects. Therefore, to demonstrate the effect of origin and destination characteristics we implement the gravity equation in three steps. We first estimate the model in terms of the influence of destination characteristics allowing for unexplained differences in the flows out of different origins using origin dummy variables. We then estimate the model in terms of the influence of origin characteristics allowing for unexplained differences in terms of the attractiveness of different destinations using destination dummy variables. Finally we combine the two estimates to calculate the residual or unexplained part of the commuting flows[4]. We base our final conclusions, however, on the more general model which allows origin and destination dummies to capture everything specific to origins and destinations allowing us to focus on the role of transport costs in deterring commuting.

We expect that distance and transport costs should have a negative effect on commutes. Wages and employment levels in the work area should have positive effects. In contrast, wages in the home area should have a negative effect. Home area employment may be positive (an overall size effect) or negative (people work locally). On average, we would expect the size effect to dominate and for the coefficient on home employment to be positive. Beyond these explanatory variables, as we just discussed, the fixed effects capture an area's overall tendency to be a home- or work-commute destination.

To reiterate, this model allows us to compare the number of commutes between Manchester and Leeds with commute patterns in the remainder of Great Britain and with other city pairs of interest. We can then use our understanding of how these determinants work on average across Great Britain to look at the specific factors that affect commuting between Manchester and Leeds.

3.2 Data Sources

To implement this methodology we need data on commuting, wages, employment, distances and transport costs between locations. An easily available source of

4. We estimate the residual as $r_{ij} = \ln (T_{ij}) - [\ln(\hat{T}^{o}_{ij}) + \ln(\hat{T}^{d}_{ij})]/2$ where a hat over the T indicates that it is predicted from the origin (superscript o) or destination (superscript d) regressions.

commuting data is the 2001 Census which is appealing as it is based on the entire population. However, this data is relatively old and so we instead use the Annual Survey of Hours and Earnings (ASHE) dataset. ASHE is constructed by the Office of National Statistics (ONS) based on a 1% sample of employees on the Inland Revenue PAYE register for February and April. It provides specific information on individuals including their home and work postcodes. The National Statistics Postcode Directory (NSPD) provides a mapping from every UK postcode to higher-level geographic units (e.g. output area, government office region, country, etc). Merging this data with each ASHE-individual's home and work postcode we are able to calculate the number of people commuting from one Local Authority area to another. We use these as our estimates of annual work-commute patterns across Great Britain for the years 2002-2005[5]. To increase the underlying sample size and to mitigate the problem that time series variation in this data can be driven by year-to-year variations in the sample we simply average across years and try to explain the average flows between 2002 and 2005 as a function of similarly time averaged area characteristics.

ASHE also includes information on occupation codes, industry code, private/public sector, age, gender and detailed information on earnings including base pay, overtime pay, basic and overtime hours worked. We use information on basic hourly earnings to calculate average wages by Local Authority area. This raises some concerns about local sample sizes from ASHE. Investigation suggests that this may be an issue for some rural LA, but not for primary urban areas. ASHE does not provide years of education so we construct these using cohort of birth-by-SOC matching on data from the LFS which contains information on both occupations and education. The way that we do that is described in Appendix 1 which also provides further details on the ASHE database.

To estimate employment in each area we use the Business Structure Database which provides an annual snapshot of the Inter-Departmental Business Register (IDBR). This dataset contains information on 2.1 million businesses, accounting for approximately 99% of economic activity in the UK and includes each business' name, postcode, industry code, number of employees, total employment (including owners), legal status and country of ownership[6]. From each firm's business address and total employment, we calculate the total employment in each Local Authority area.

We identify the centre of Local Authority areas using information on postcode locations and calculate distances as the straight line distance between these centroids. Coordinates (northing and easting) for all UK postcodes are provided by the NSPD. From these, we define an area's centroid as the average across all of its postcode coordinates. Since the number of postcodes increases roughly in proportion to population, this calculation of an area's centroid gives a rough estimate of the area's center-of-gravity by population. The distance between the centroids is then calculated using the Pythagorean Theorem. We construct Generalized Transport Costs (GTC) for train and driving as detailed in Appendix 2.

3.3 Results

As explained above we initially estimate two separate models. The first explains commuting as a function of destination characteristics allowing for unobserved characteristics of origins to affect commuting. The second explains commuting as a function of origin characteristics, allowing unobserved characteristics of destinations to affect commuting. For both origins and destinations we start with a very simple

5. Dan Graham (from Imperial College, London) has been working for DfT to assess whether the sample contained in ASHE is sufficiently representative to allow reasonable estimation of commuting flows (by comparing to the 2001 census). The results of this work are not yet published, but in private correspondence he has confirmed that ASHE based estimates of commuting flows should be sufficiently accurate for the kind of modeling exercise that we wish to undertake.

6. The 99% coverage was last verified in 2004/05, although there is no reason to think that this is not still the case.

model that only includes distance between Local Authority areas. Column 1 of Table 1 reports results. As expected, distance has a negative effect on commuting between Local Authority areas. The dependent variable (commuting) is in natural logarithms, but distances are in hundreds of kilometres (consistent with the exponential specification for the deterrence function as discussed above). The coefficient of 2.53 on origin-destination distance, implies that each 1km increase in distance reduces commuting by 2.5%. This means that commuting between Local Authority areas roughly halves every 30km. Column 2 shows what happens when we add in (log) employment as a measure of the size of Local Authority areas. Again, as expected, high employment destinations are associated with more commuting. The same is true of high employment origins suggesting that, as discussed above, the positive size effect dominates any negative local employment effect. Because origin and destination characteristics are entered in logs we can interpret the coefficients as telling us that commuting increases by 1.7% for a 10% increase in destination employment. The effect of origin employment is a little under half this. When introducing these measures of size, the coefficient on distance is essentially unchanged. Column 3 shows what happens when we include (log) wages. As with distance and employment, the coefficients are consistent with our prior expectations. High wage destinations are associated with more commuting, while high wage origins are associated with less. A 10% increase in destination wage increases commuting by 1%[7]. A similar increase in origin wage decreases commuting by 1.8%.

Table 1: Gravity of LA-LA commuting flows

Destination characteristics	1	2	3	4	5
log employment		0.1690***	0.1593***		
		0.0063	0.0066		
log wage			0.1025***		
			0.0256		
distance	-2.5346***	-2.7161***	-2.7109***	-2.9311***	7.9921***
	0.0401	0.0419	0.0421	0.0452	0.4062
GTC (driving)					-3.0857***
					0.1785
GTC (train)					-1.4001***
					0.0506
Observations	35584	35584	35584	35584	35584
R-squared	0.163	0.185	0.186	0.221	0.262
Origin characteristics	**1**	**2**	**3**	**4**	**5**
log employment		0.0659***	0.0818***		
		0.0056	0.0062		
log wage			-0.1830***		
			0.0309		
distance	-2.3539***	-2.3764***	-2.4153***	-2.9311***	7.9921***
	0.0368	0.0369	0.0382	0.0452	0.4062
GTC (driving)					-3.0857***
					0.1785
GTC (train)					-1.4001***
					0.0506
Observations	35584	35584	35584	35584	35584
R-squared	0.179	0.182	0.183	0.221	0.262

Notes: Table reports OLS regression coefficients and standard errors. Dependent variable is log commuting.
Source: ONS

7. This is the only coefficient that is sensitive to the inclusion of London. If we drop the one third of our sample for which at least one of the LAs is in the London City Region we no longer find a significant effect of destination wage on commuting flows. Our other results are otherwise unaffected.

Columns 1 to 3 separately model the effects of destination characteristics allowing for unobserved origin effects and origin characteristics allowing for unobserved destination effects. Column 4 shows the combined model where we allow for both unobserved origin and destination characteristics to drive commuting. As explained above, we can no longer separately identify the affect of observable origin and destination characteristics. We can, however, still identify the effect of distance on commuting flows (because this is origin-destination) specific. As is clear from column 4 the effect of distance is largely unchanged. Note also, that once we include both origin and destination fixed effects the origin and destination specifications are identical so we get the same results for both. Finally, column 5 shows what happens when we introduce measures of generalized transport costs (GTC) that capture both the monetary and time costs of travel. The effect of both train and driving GTC is negative while the impact of straight line distance is now positive. GTC are measured in £100's of pounds so the coefficients tell us that a £100 increase in GTC reduces commuting by 3.1% for driving, 1.4% for train. These effects are not particularly large (possibly because GTC's tend to respond positively to commuting flows). The positive coefficient on distance tells us that, once we condition on transport costs, distances are actually associated with more commuting (although the effect is small in magnitude). This is clearly not a causal linkage. A likely theoretical explanation for this finding is that longer commutes between cities with better (and therefore lower cost) transport links are more prevalent than other types of shorter LA-LA commute that involve similar transport costs. In practice, the distance and driving GTC variables are very highly correlated with each other, which makes it difficult to disentangle their separate effects on commuting[8]. One final thing to note, before turning to the specific question of the links between Manchester and Leeds is that the models are ranked in order of their ability to explain the overall variation in commuting (note that the R-square increases as we move across the columns).

As explained above, we can now use these gravity models to see whether commuting between Manchester and Leeds is higher or lower than expected. We do this by comparing predicted commuting between Local Authority areas in Manchester and Leeds with actual commuting. There are 15 Local Authorities in the Manchester City Region and 8 Local Authorities in the Leeds City-Region (Appendix 3 describes how we construct these city regions). Thus we need to compare the 120 bilateral flows between Manchester and Leeds to the 35,428 other Local Authority pairs that have positive commuting flows for which we are able to compare predicted to actual commuting. We use a simple regression analysis to make this comparison. Specifically, we regress the Local Authority area to Local Authority area residual log commuting flows on dummy variables that indicate whether these flows are to or from Manchester, to or from Leeds, or between Leeds and Manchester (in either direction). The top panel of Table 2 shows the results when running regressions on all Local Authority area to Local Authority area pairs. Each column 1 to 5 corresponds to the five specifications that we described above (the results of which are reported in columns 1 to 5 of Table 1).

Looking across the columns in the top panel it is evident that both Manchester and Leeds Local Authority areas have lower mean inflows and outflows than we would expect given their employment, wages and geographical position relative to other Local Authority areas. Even given these relatively low flows in and out of Manchester, commuting between Manchester and Leeds Local Authority areas is lower than expected, but the pattern of coefficients is not simple. If we just model commuting as driven by size (employment), wages and distance, commuting

8. We have tried alternative specifications using additional terms in the square of distance and transport costs, including the product of employment and the ratio of wages (to allow for more complex interactions). None of these changes alter the overall pattern of coefficients or our conclusions based on analysis of the residuals from these more complex specifications.

between Manchester and Leeds is about what you would expect. Specifically, it is about 96% of expected, but this effect is statistically insignificant (i.e. we have no confidence that the true figure isn't 100% and that the 96% hasn't just occurred by chance). If we take in to account all the origin and destination specific factors that might explain unusual flows, column 4, then is about 92% of predicted values. Once we take in to account the GTC of driving and trains this effect actually gets larger. The last column suggests that commuting between Manchester and Leeds is around 20% lower than between other Local Authority areas at similar distance and with similar commuting costs. Taken at face value, this suggests that it is not commuting costs that explain relatively low commuting between Manchester and Leeds but something else.

These comparisons, however, are rather misleading, because they compare inter-city region commutes between Manchester and Leeds to all Local Authority areas to Local Authority area commutes that may be both inter or intra-city region (or involve Local Authority areas not part of a city-region). To get round this, we pick a set of city region comparators (chosen before we conducted the analysis in consultation with the projects Steering Group) and compare commutes between Manchester and Leeds LAs to the inter-city-region commutes for these chosen comparators. After consultation we chose as comparators cities of a similar size and distance apart, specifically: Edinburgh-Glasgow, Bristol-Cardiff, Leeds-Sheffield, Manchester-Birmingham, Manchester-Liverpool, Nottingham-Derby, Leeds-Hull, Leeds-York. There are 566 inter-city region Local Authority area to Local Authority area commutes in this comparison sample. Results for the coefficient of interest (the Manchester-Leeds dummy) for this sample are reported in the second panel of Table 2. The point estimates in columns 1 to 3 suggest that commuting is lower than expected given the size (employment), wages and straight line distance. But these effects are very imprecisely estimated. Once we take all unobserved factors in to account, commuting is quite a lot lower given the straight line distance between Manchester and Leeds (column 4).This result tells us that commuting flows between Local Authority areas in Leeds and Local Authority areas in Manchester are around 37% lower than we would expect, when compared to inter-city-region commuting flows between other Local Authority areas at similar distance, Much, if not all of this gap is explained by relatively high commuting time and costs between Manchester and Leeds. Once we allow for the actual train and driving GTC between Local Authority areas the gap falls to about 22% but the estimate is very imprecise, and there is actually no statistically significant difference between Manchester-Leeds flows compared to the other city-region comparators (column 5).

Overall, compared to other city-region commutes between areas with similar economic activity, Manchester-Leeds commuting is low. However, most if not all of the difference is due to the time and monetary costs of commuting. Taking these costs into account, commuting between Manchester and Leeds does not appear to be much lower than is to be expected.

Table 2: Predicted versus actual commuting flows for Manchester-Leeds

All pairs	1	2	3	4	5
Manchester outflows	-0.0697***	-0.0813***	-0.0864***	-0.0466***	-0.0509***
	0.013	0.0129	0.0129	0.0128	0.0132
Leeds outflows	-0.0995***	-0.1233***	-0.1316***	-0.0721***	-0.0514***
	0.0157	0.0152	0.0152	0.0157	0.0158
Manchester inflows	-0.0852***	-0.1201***	-0.1167***	-0.1356***	-0.0995***
	0.0122	0.0122	0.0122	0.0116	0.0127
Leeds inflows	-0.0323**	-0.0937***	-0.0869***	-0.0697***	-0.0365**
	0.0164	0.0161	0.0161	0.0161	0.0178
Manchester´Leeds flows	-0.0362	-0.0375	-0.0394	-0.0771*	-0.2196***
	0.0448	0.043	0.043	0.0405	0.039
Observations	35584	35584	35584	35584	35584
R-squared	0.002	0.003	0.003	0.003	0.002
Comparator city-regions					
Manchester´Leeds flows	-0.1694	-0.1388	-0.1365	-0.3706**	-0.2213
	0.1554	0.1505	0.1505	0.1456	0.1439
Observations	566	566	566	566	566
R-squared	0.068	0.08	0.08	0.06	0.069

Notes: Table shows regression coefficients and standard errors. Dependent variable is residual log commuting flows, as described in text. ***, **, * denote significance at the 1%, 5% and 10% levels respectively. Source: ONS

In policy circles, some concern has been expressed about the ability of the two city regions to attract high skilled workers so, before leaving the issue of commuting, it is interesting to briefly consider the commuting patterns of higher skilled workers. We will have a lot more to say on other labour market issues in section 4. We proceed exactly as we did for overall commuting but now restricting our attention to commutes for the highest skilled workers only (see the data section for an explanation of the skills classification). We construct measures of skilled employment and average skilled wages using data from ASHE and, once again, average across time to increase the sample size. We present results for destination and origin regressions in columns 1 to 3 of Table 3. We start by including only distance (column 1) then add employment (column 2) and finally wages (column 3). As before destination employment and wage have a positive effect on commuting, origin employment also has a positive effect, while the effect of origin wage is negative. Distance continues to have a negative effect although the coefficient is slightly smaller in magnitude reflecting the higher commuting propensity for higher skilled workers. Columns 4 and 5 present results when including a full set of origin destination dummies (column 4) and then adding transport GTC (column 5). As before the impact of both driving GTC and train GTC is negative, while including them turns the coefficient on distance positive.

Table 3: Gravity models of LA-LA commuting flows (skilled workers)

Destination characteristics	1	2	3	4	5
log employment		0.1574***	0.1468***		
		0.0066	0.0068		
log wage			0.1636***		
			0.0307		
distance	-1.9319***	-2.1622***	-2.1645***	-2.3001***	
	0.0425	0.0469	0.0473	0.0507	0.5315
GTC (driving)					0.2307
GTC (train)					0.0532
Observations	17411	17411	17411	17411	17411
R-squared	0.15	0.189	0.192	0.226	0.259
Origin characteristics	**1**	**2**	**3**	**4**	**5**
log employment		0.0955***	0.1042***		
		0.007	0.0075		
log wage			-0.1316***		
			0.0415		
distance	-1.9943***	-2.0193***	-2.0334***	-2.3001***	
	0.0432	0.0434	0.044	0.0507	0.5315
GTC (driving)					0.2307
GTC (train)					0.0532
Observations	17411	17411	17411	17411	17411
R-squared	0.19	0.198	0.198	0.226	0.259

Notes: Table reports OLS regression coefficients and standard errors. Dependent variable is log commuting flow. ***, **, * denote significance at the 1%, 5% and 10% levels respectively. Source: ONS

We use the same secondary regression to summarise whether commuting between Manchester and Leeds is higher or lower than expected. Results are shown in Table 4. As before, when comparing to all possible LA pairs Manchester-Leeds flows are significantly lower, by as much as 19% once we control for origin and destination effects and generalised transport costs. However, when we use the city-region comparators discussed above we find only a small (13%) gap between Manchester-Leeds and the rest when using straight line distance, but this is not a statistically significant difference. All of this gap is explained by commuting costs. Even if we ignore questions of significance, our specification that allows for unobserved origin and destination characteristics, distance and GTC has skilled commuting flows between Manchester-Leeds only 1.6% lower than expected. As we found for commuting for all workers, flows for skilled commuters between Manchester and Leeds may be lower than expected, but any gap is explained by commuting costs rather than any other more subtle unobserved factors about the Manchester-Leeds relationship.

Table 4: Predicted versus actual commuting flows for Manchester-Leeds (skilled workers)

All pairs	1	2	3	4	5
Manchester outflows	-0.1168***	-0.1311***	-0.1313***	-0.1047***	-0.1097***
	0.0128	0.0126	0.0125	0.013	0.0135
Leeds outflows	-0.0723***	-0.1065***	-0.1073***	-0.0630***	-0.0473***
	0.0155	0.0152	0.0151	0.0164	0.0166
Manchester inflows	-0.0950***	-0.1245***	-0.1176***	-0.1414***	-0.1110***
	0.0128	0.0126	0.0126	0.0131	0.0144
Leeds inflows	-0.0309*	-0.0946***	-0.0820***	-0.0673***	-0.0417**
	0.0185	0.0183	0.0183	0.0187	0.0205
Manchester´Leeds flows	-0.0786**	-0.0867**	-0.0898**	-0.1117***	-0.1885***
	0.0358	0.0349	0.0349	0.0369	0.037
Observations	0.0112**	0.0156***	0.0150***	0.0136***	0.0120***
R-squared	0.0046	0.0045	0.0045	0.0044	0.0044
Comparator city-regions					
Manchester´Leeds flows	-0.0338	-0.0148	-0.0172	-0.1308	-0.0161
	0.1209	0.1169	0.117	0.1179	0.1156
Observations	360	360	360	360	360
R-squared	0.09	0.105	0.104	0.066	0.092

Notes: Table shows regression coefficients and standard errors. Dependent variable is residual log commuting flows, as described in text. ***, **, * denote significance at the 1%, 5% and 10% levels respectively. Source: ONS

3.4 Conclusions

- The analysis of commuting between Local Authority areas in Britain suggests that, commuting is greater when Local Authority areas are larger in terms of employment. High wage Local Authority areas attract more commuters while low wage Local Authority areas generate more commuters. Transport costs reduce commuting. These findings are in line with previous research and theoretical predictions.

- Given the effects of straight line distance, size and employment, we find no difference in commuting between Manchester and Leeds and commuting between Local Authority areas for comparator city-region pairs. When we allow for all unobserved factors that may affect commuting our analysis does find some differences in commuting between Local Authority areas in Manchester-Leeds and between Local Authority areas in comparator city-region pairs. Specifically, commuting between the Manchester and Leeds City-Regions is about 40% lower than expected given the characteristics of the two cities and the physical distance between them. High overall commuting costs between the Manchester and Leeds City-Regions appear to be the main cause of this lower commuting. Once we include the overall costs of commuting between areas, both by car and train, in our analysis we can no longer be certain that there is any difference between the Manchester and Leeds city-regions and the other comparator city-region pairs.

- If we focus only on the highest skilled workers we reach essentially the same conclusions. Commuting between Manchester and Leeds is roughly what we would expect given the characteristics of LAs, and the distance and generalized transport costs of travelling between them.

- Economic factors, specifically the overall costs of commuting between the two cities, are the most important factor in explaining these relatively low commuting levels. This suggests that lowering these costs has an important role to play in increasing integration between the two city regions. This in turn may improve the economic performance of the two city-regions as we discuss further below.

- We do not examine the role of cultural or social factors directly. However, the fact that economic factors appear to explain low commuting levels leaves little room for cultural or social factors to play a large part in this story. This suggests that such factors are unlikely to act as a barrier to increased commuting between the two cities if transport investment lowers the overall costs of commuting, or if other economic factors lead to enhanced interactions.

4. Interactions in earnings, employment and output

While commuting is one of the most important ways in which interactions between cities occur, there are of course a number of others, including linkages between customers and suppliers. Unfortunately, there is very little, if any data, collected on these other linkages. There is certainly no systematic source of data collected for different places in different time periods. Therefore, for these other linkages, unlike with commuting, we are unable to directly observe the interactions between places. Instead, we have to turn our attention to the possible effects on outcomes, which are far harder to analyse than the interactions themselves. In this section we focus directly on outcomes by considering the extent to which outcomes (say increases in employment) of city-regions tend to move together. As for the work on commuting, the strategy is to identify general relationships for Great Britain and then ask whether the relationships between Manchester and Leeds are understandable in light of these general relationships using exploratory spatial data analysis and spatial econometrics. We outline the data first, before describing our approach and findings.

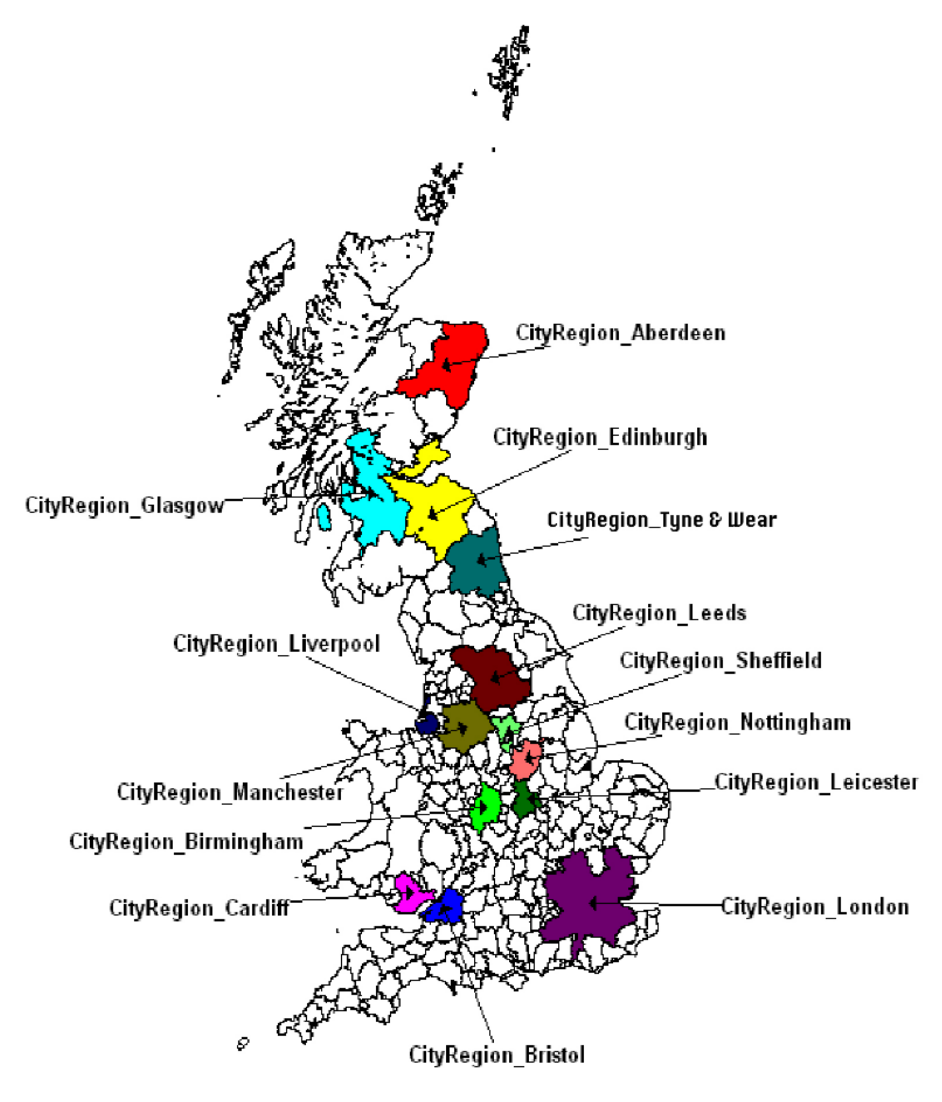
4.1 Data Description

We consider interactions in terms of earnings, employment and output per worker. This choice is governed by three considerations. First, the availability of data. Second, output per worker and wages are key outcomes that are explained by the structural economic model that we use later in this report to consider the implications of increased integration between Manchester and Leeds. Third, these are some of the most important outcomes from a regional economic development perspective.

Our units of analysis are a mix of Local Authorities and city regions in England, Wales and Scotland. We have 242 Local Authorities and 14 city regions. The city regions are aggregations of the remaining 161 Local Authorities into spatial units that better represent functional economic regions. In Appendix 3, we discuss how the city regions are constructed and list the districts which belong to each city region. Figure 1 shows the Local Authorities and 14 city regions with which we are working.

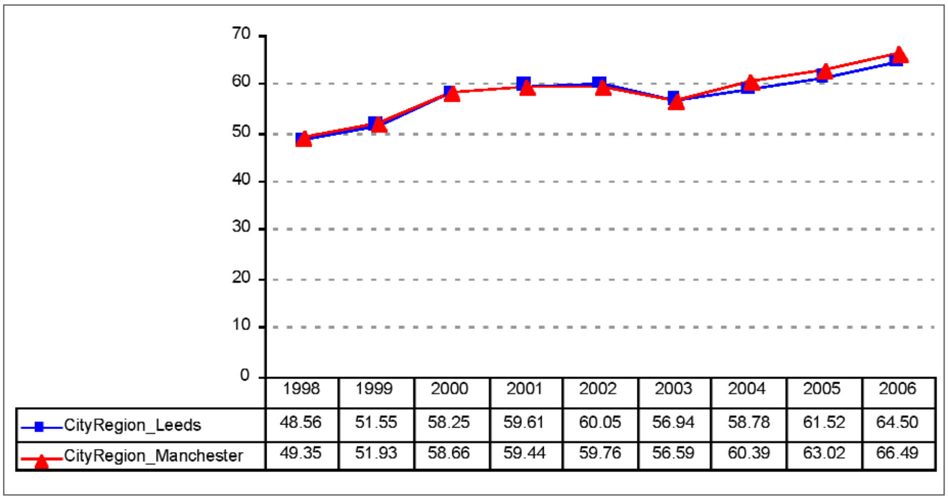
Local Authorities data for England, Wales and Scotland come from Nomis/ASHE (Annual Survey of Hours and Earnings) and Nomis/ABI (Annual Business Inquiry). Nomis/ASHE gives us the average hourly earnings of all full-time employees based on the location of workplace. Nomis/ABI gives us the number of employees based on the location of the workplace. Gross Domestic Product data (at current market prices in millions of euros) comes from Eurostat but is only provided at NUTS3 level. We estimate GDP at district level by distributing GDP according to employment shares calculated as the Local Authority share employment in total NUTS3 employment. Appendix 4 provides more detail on the data. The exploratory spatial data analysis will examine variables in levels in 2006 (the latest data available at Local Authority level), in differences between 1998 and 2006, and in annual growth rates between 1998 and 2006.

Figure 1 – City Regions and Local Authority Districts



Graphs 1, 2 and 3, and the associated tables, plot output per worker, earnings and employment for the Manchester and Leeds city regions from 1998-2006. They show the overall upward trend during this period in all three variables as well as the tendency for the two city regions to move together. The rest of this section is concerned with the extent to which these co-movements are stronger or weaker than might be expected in comparison to the rest of Great Britain.

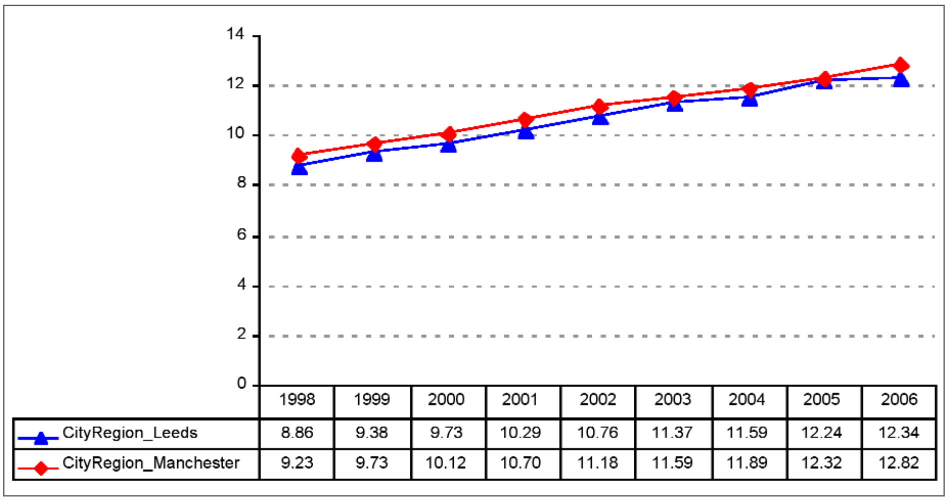
Graph 1 – GDP per worker between 1998 and 2006 (Manchester and Leeds)



	1998	1999	2000	2001	2002	2003	2004	2005	2006
CityRegion_Leeds	48.56	51.55	58.25	59.61	60.05	56.94	58.78	61.52	64.50
CityRegion_Manchester	49.35	51.93	58.66	59.44	59.76	56.59	60.39	63.02	66.49

Source: Eurostat and Nomis/ABI.

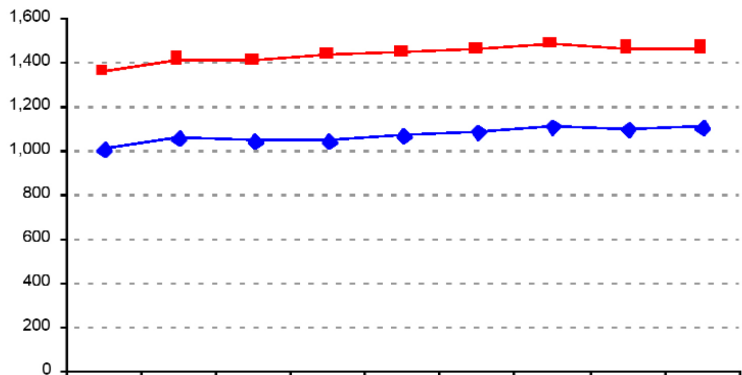
Graph 2 – Earnings between 1998 and 2006 (Manchester and Leeds)



	1998	1999	2000	2001	2002	2003	2004	2005	2006
CityRegion_Leeds	8.86	9.38	9.73	10.29	10.76	11.37	11.59	12.24	12.34
CityRegion_Manchester	9.23	9.73	10.12	10.70	11.18	11.59	11.89	12.32	12.82

Source: Nomis/ASHE.

Graph 3 – Employment between 1998 and 2006 (Manchester and Leeds)



	1998	1999	2000	2001	2002	2003	2004	2005	2006
CityRegion_Leeds	1,006.4	1,054.6	1,044.5	1,046.1	1,068.5	1,085.1	1,112.3	1,100.2	1,105.4
CityRegion_Manchester	1,356.8	1,413.6	1,406.7	1,430.6	1,446.6	1,455.8	1,480.0	1,460.7	1,465.3

Source: Nomis/ABI.

4.2 Exploratory spatial data analysis

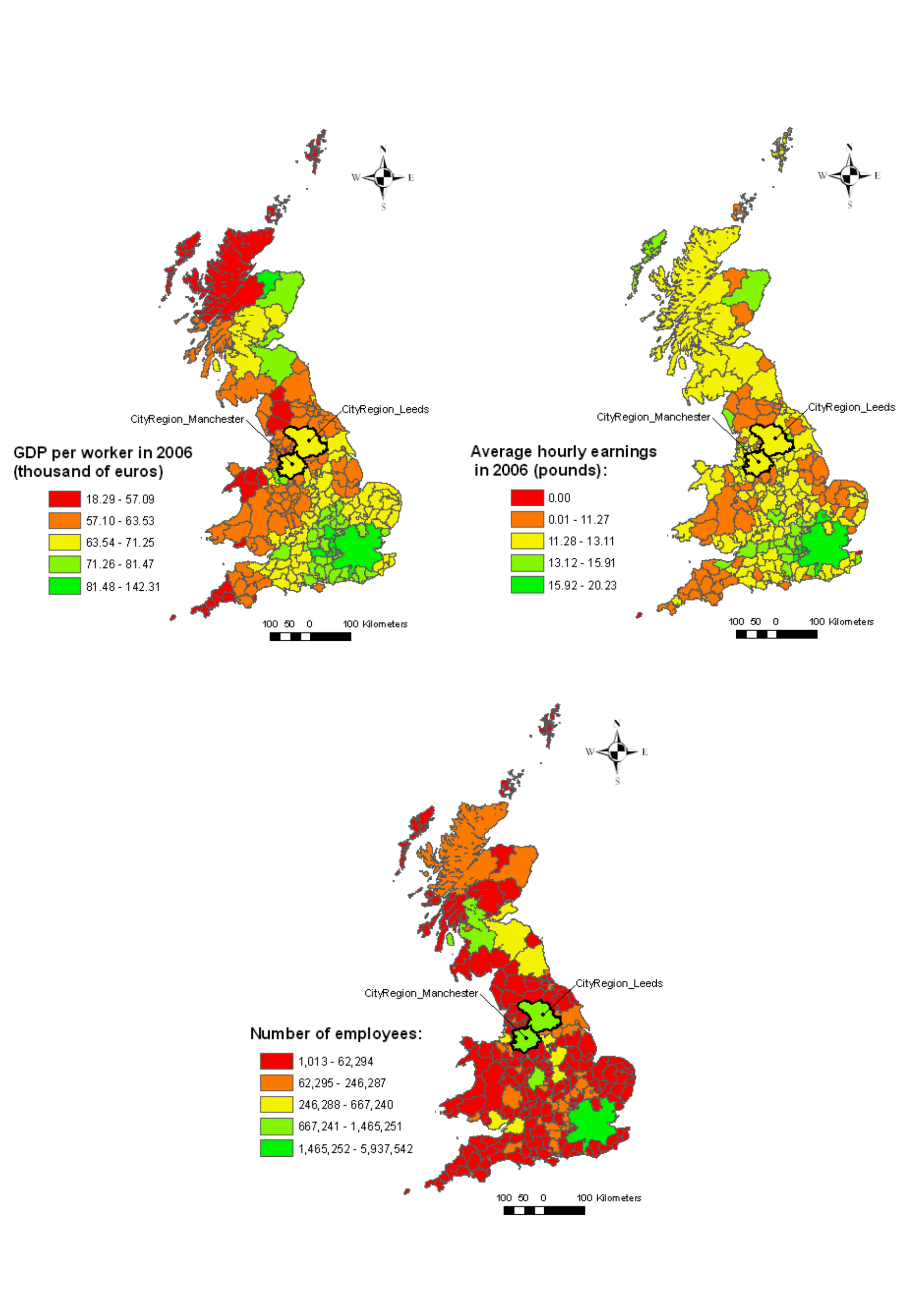
We examine the spatial interaction between areas using two related exploratory spatial data analysis (ESDA) techniques: Moran's I and Local Moran's I. ESDA is a set of techniques aimed at describing and visualizing spatial distributions and at detecting patterns of spatial association or clusters (Anselin, 1998a,b). Essentially, these methods measure global and local spatial autocorrelation. Indicators of spatial association measure the extent to which outcomes for "nearby" areas (in a sense to be made precise below) move in the same direction (positive spatial autocorrelation), move independently (zero spatial autocorrelation) or move in different directions (negative spatial autocorrelation). A global indicator of spatial association (e.g. Moran's I) captures the general pattern throughout Great Britain. Of course not all areas will follow the general pattern revealed by global indicators. Examining local spatial autocorrelation allows us to find neighbouring groups of areas which exhibit strong positive spatial autocorrelation, strong negative autocorrelation or show no spatial autocorrelation, within this general GB pattern. Comparing local patterns with global patterns shows us which groups of areas make the greatest contribution to the general pattern, and which groups of areas show a markedly different pattern of inter-relationships to those that hold more generally in Great Britain.

The exploratory spatial data analysis focuses on the nature of spatial interactions in levels, changes and growth rates for three variables:

1. Output per worker in 2006;
2. Difference in output per worker between 1998 and 2006;
3. Annual growth rates of output per worker between 1998 and 2006;
4. Earnings in 2006;
5. Difference in earnings between 1998 and 2006;
6. Annual growth rates of earnings between 1998 and 2006;
7. Number of employees in 2006;
8. Difference in employment between 1998 and 2006;
9. Annual growth rates of employment between 1998 and 2006.

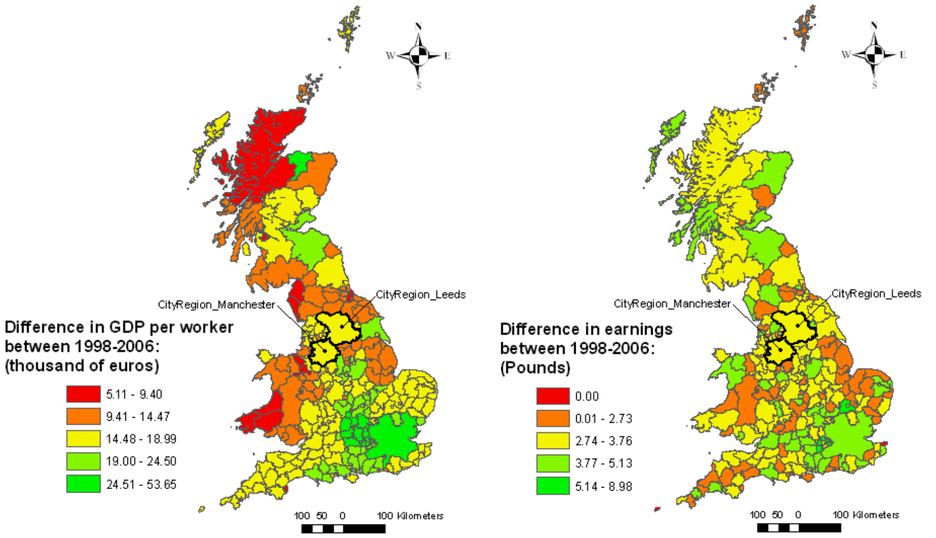
We start by mapping the data to give an impression of the spatial variations in outcomes for these variables. Figure 2 shows levels of output per worker, earning and employment at district/city-region level in 2006 for Great Britain. Figure 3 shows differences in output per worker, earnings and employment between 1998 and 2006. Figure 4 shows annual growth rates of output per worker, earnings and employment between 1998 and 2006.

Figure 2 - Levels of GDP per worker, earnings and employment in 2006

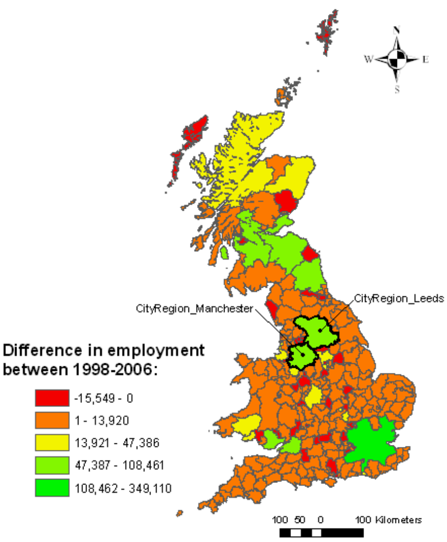


Souce: Eurostat, Nomis/ASHE, Nomis/ABI.

Figure 3 – Differences in GDP per worker, earnings and employment between 1998 and 2006

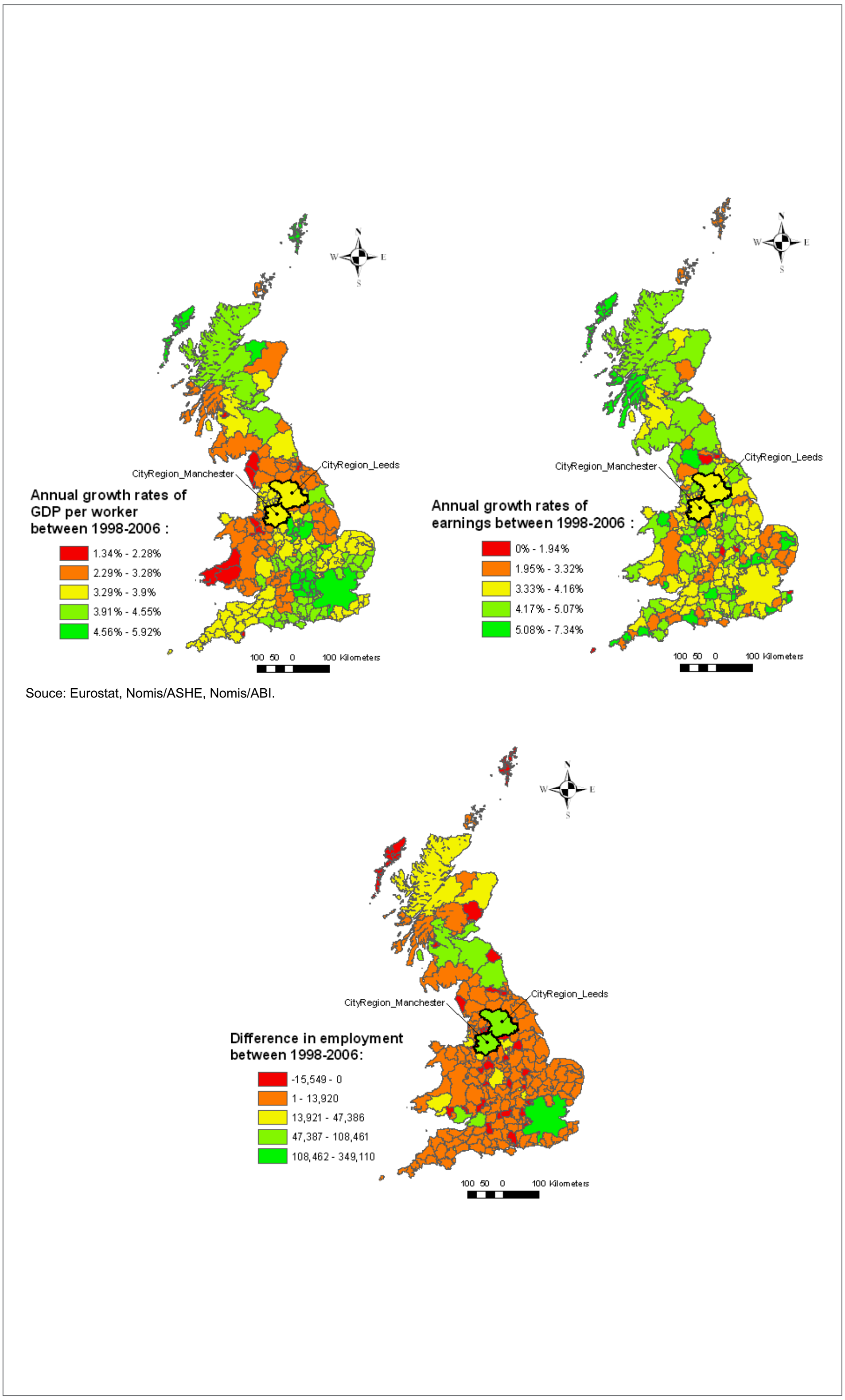


Souce: Eurostat, Nomis/ASHE, Nomis/ABI.



Souce: Eurostat, Nomis/ASHE, Nomis/ABI.

Figure 4 - Annual growth rates of GDP per worker, earnings and employment between 1998 and 2006



Souce: Eurostat, Nomis/ASHE, Nomis/ABI.

4.2.1 Moran's I statistics

Global spatial autocorrelation is based on Moran's I statistic (Cliff & Ord, 1981). For output per worker in 2006, this statistic is written in the following form:

$$I = \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} \qquad (1)$$

where w_{ij} are elements of a spatial weighting matrix (W) which is row-standardized such that the elements w_{ij} in each row sum to 1. y_i and y_j are the values of the outcome of interest (output per worker in 2006 for instance), $\bar{y}$ is the mean of the outcome and $\sum_i (y_i - \bar{y})^2$ is the variance normalization factor. We will discuss the spatial weighting matrix shortly. If $I \approx 0$, then there is no evidence of spatial autocorrelation, i.e., area outcomes tend to move independently. If Moran's I statistic is greater than zero, there is a positive spatial autocorrelation, i.e., areas with high output per worker in 2006 tend to be "near" to neighbouring areas with high output per worker in 2006 (and vice-versa). Finally, if Moran's I statistic is smaller than zero, there is a negative autocorrelation, i.e., districts with high output per worker tend to be close to neighbouring districts with low output per worker and vice versa. The statistical significance of Moran's I can be calculated using the permutation approach (Anselin, 1995)[9].

The local version of Moran's I statistic is an example of a Local Indicator of Spatial Association (LISA). Anselin (1995) defines a LISA as any statistic satisfying two rules: the LISA for each spatial unit should give an indication of significant spatial clustering of similar values around that unit and the sum of the LISA for all spatial units should be proportional to a global indicator of spatial association. Thus the local-Moran statistic indicates to what extent a specific area is surrounded by areas with high or low values of the outcome analyzed. A LISA measure of spatial association (Anselin, 1995) can be defined as:

$$I_i = \frac{(y_i - \bar{y}) \sum_j w_{ij} (y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2 / n} \qquad (2)$$

where n is the number of observations, w_{ij} are the elements of the spatial weight matrix and as above y_i and y_j are the values of the outcome of interest, $\bar{y}$ is the mean of the outcome and $\sum_i (y_i - \bar{y})^2$ is the variance normalization factor. In this case a significant positive result indicates the existence of a cluster of similar values surrounding area i, that we would be unlikely to see if the values were randomly distributed over space. Using local spatial association tests we can detect hot spots or areas showing values of output per worker far above the average, as well as clusters of areas with significantly low values.

Comparing these clusters across the whole of Great Britain we can evaluate the strength of the correlation between nearby areas and assess, for example, whether Manchester-Leeds represent a cluster of areas that have unusual spatial autocorrelation on particular outcomes.

4.2.2 Spatial Weight Matrix (W)

So far, we have been deliberately vague about what is meant by "nearby" areas. We now consider this issue. The NxN spatial weight matrix (W) provides the 'structure' of spatial relationships by defining what we mean by "nearby" areas. The most common weight matrix is a standardized first-order contiguity matrix (also called Queen contiguity matrix): that is, the element w_{ij} in the matrix is 1 if areas i

9. All computations were carried out using Geoda and ArcGIS9.

and j share borders or vertices and 0 otherwise. Another common choice is "Rook contiguity" which uses only common borders. For real geographies like Great Britain the choice between Rook or Queen contiguity will not make any difference, but it would if our areas were defined as squares on a regular grid like a chessboard. In this report given the rather uneven size of the spatial units we prefer to use a distance based matrix. Specifically, we use a distance band of 70 kilometres of one another. In other words, the element w_{ij} in the matrix is 1 if areas i and j are within 70 kilometres and 0 otherwise. We have chosen a 70 kilometres cut-off because it is the travel distance (by car) between Manchester and Leeds. Therefore, this spatial weight matrix captures the spatial interdependence between areas in a way that allows us to say more about the relationship between Manchester and Leeds. It is important to note that analysis will be conditional on the choice of this spatial weight matrix[10].

4.2.3 ESDA Results

Figure 5, Figure 6 and Figure 7 show Moran's I statistic and the Moran scatterplot for output per worker, earnings and employment (respectively, levels in 2006, differences between 1998 and 2006, and annual growth rates between 1998 and 2006). The Moran scatterplot is a useful way of visualising the spatial interactions captured by the global and local I. The scatterplot displays the "spatial lag" of an outcome for each area plotted against the outcome for each area. The spatial lag is constructed as W times the variable of interest. For example, for the (row-standardised) 70 kilometres cut-off matrix we are using here the spatial lag for any area is just the average of the outcomes for neighbouring areas which are within 70 kilometres of that area.

Figure 5 gives the Moran scatterplot for 2006 output per worker, earnings and employment. For the first two variables the Moran's I statistic is positive (0.4501 and 0.3329 respectively) and highly significant. This can be easily seen in the scatterplots by noticing that most of the points lie either in the Low-Low (south west part of the diagram) or the High-High (north east part of the diagram). These parts of the diagram capture places which exhibit positive spatial autocorrelation (high values with high values, low values with low values). The off diagonal areas Low-High (in the north-west part of the diagram) and High-Low (in the south east part) represent negative spatial autocorrelation, indicating spatial clustering of dissimilar values. It's clear that positive spatial autocorrelation is far more common than negative.

These results suggest that the null hypothesis of no spatial autocorrelation is rejected and that the distributions of output and earnings variables are by nature clustered. In other words, areas (districts/city regions) with relatively high values are near other areas with relatively high values and vice-versa. This is, of course, not particularly surprising, but Moran's I provides us with a way of quantify the extent and significance of these positive correlations. On the other hand, Moran's I statistics for the level of employment in 2006 shows a non-significant value close to zero (-0.0136) with a p-value equal to 0.1854. This finding presumably reflects the interplay of two offsetting effects – employment outcomes for large employment centres do show a broad spatial pattern, but these employment centres are often surrounded by commuting areas of low employment (see Figure 2).

For differences (Figure 6) and growth rates (Figure 7) of output per worker and earnings the values of the Moran's I statistics decrease compared to those for the variables in levels. That is, changes and growth of output per worker and earnings

10. The results we report are robust to using alternative distances as the cut-off to define neighbouring areas.

show more of a random spatial pattern than levels. Interestingly, however, the difference and growth rates of employment become positively statistically significantly correlated at the 5% and 1% level, respectively. For difference in output per worker between 1998 and 2006 (Figure 6), the Moran's I statistic is positive and significant (0.4010) with a p-value equal to 0.0001. For differences in earnings the Moran's I statistic is 0.0808 with a p-value equal to 0.0015. Finally, in Figure 7 we observe that the null hypothesis of no spatial autocorrelation for growth rates of employment is rejected (p-value=0.0001) showing that the distributions of this variable is by nature clustered over the period 1998-2006.

In summary, large significant and positive values of Moran's I reveal the presence of spatial association of similar values among neighbouring areas in output per worker and earnings in 2006. However, when their differences and growth rates are analyzed the values of the statistic decrease. The main finding to emerge at this point is that Moran's I values in levels are higher than those for differences and growth rates. We now turn to the question of local associations and the specific question of the relative strength of the spatial correlation between Manchester and Leeds.

Figure 5 – Moran's I (scatter plots) of levels of GDP per worker, earning and employment in 2006

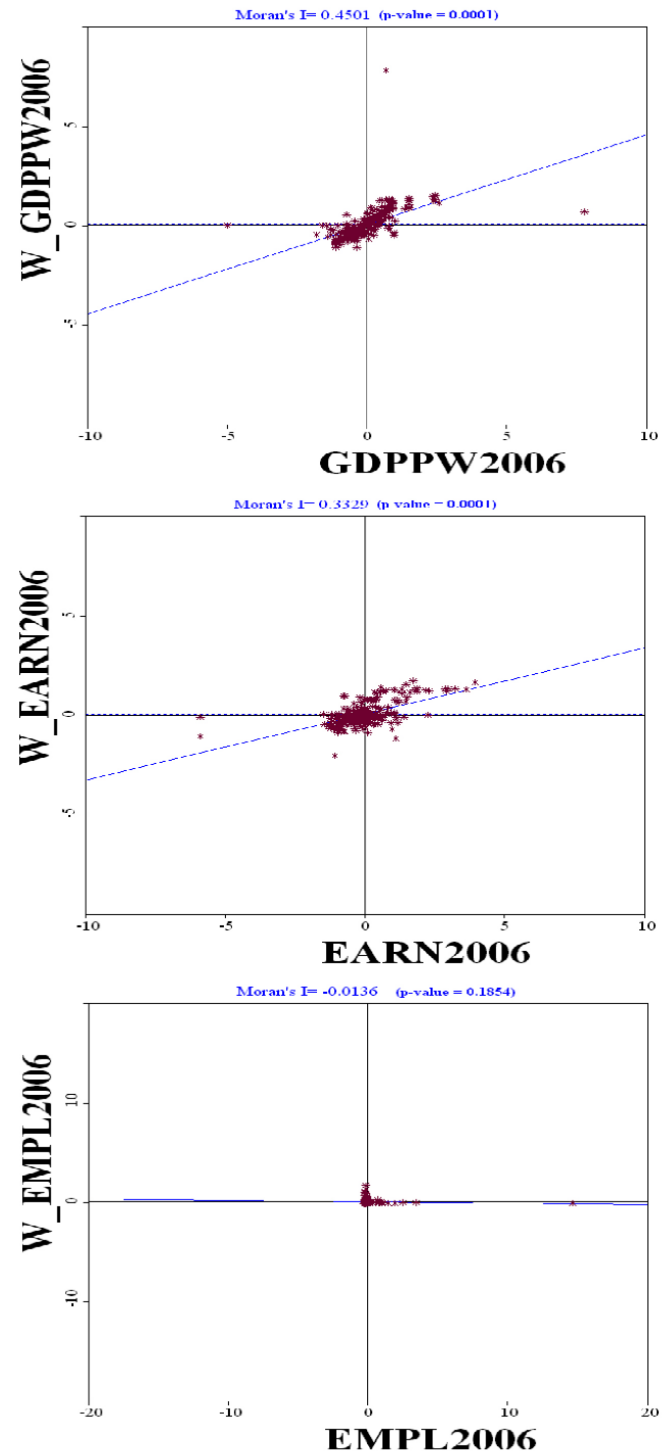


Figure 6 – Moran's I (scatter plots) of Differences in GDP per worker, earning and employment between 1998 and 2006

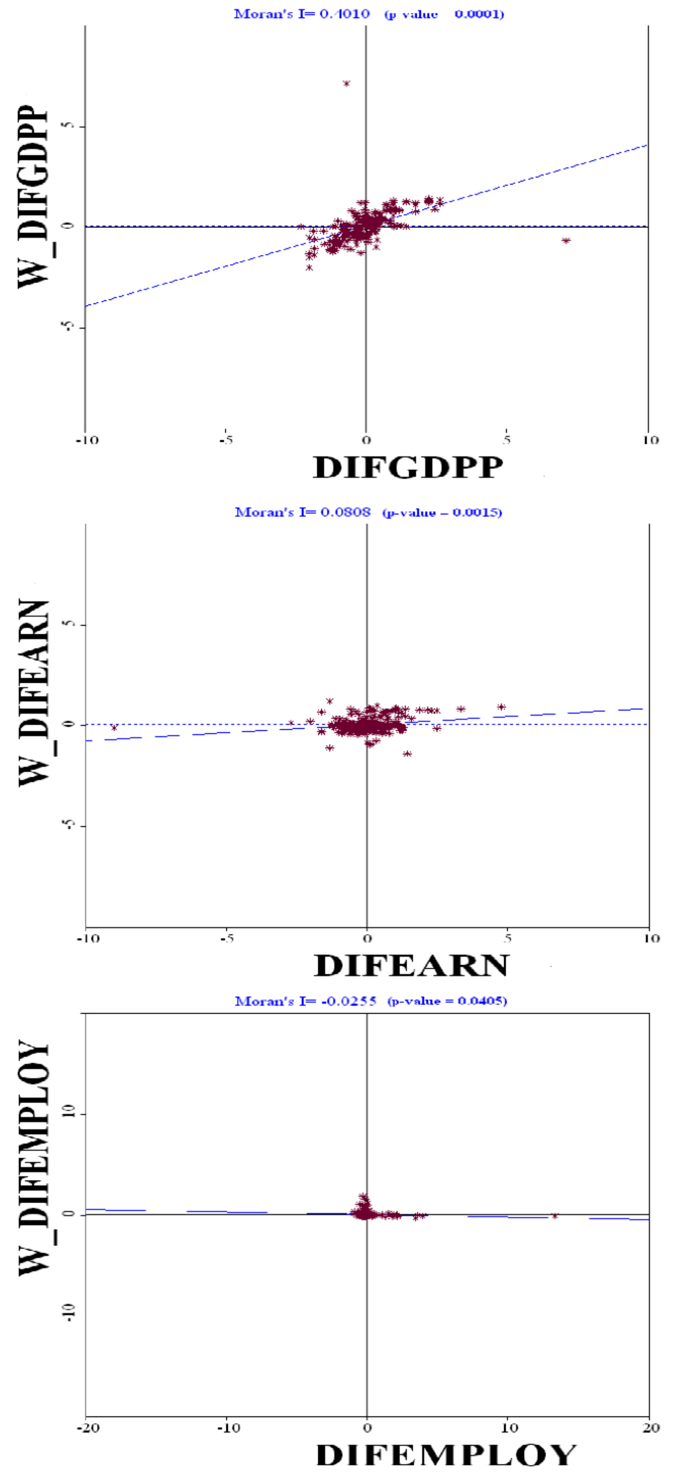


Figure 7 – Moran's I (scatter plots) of annual growth rates of GDP per worker, earnings and employment between 1998 and 2006

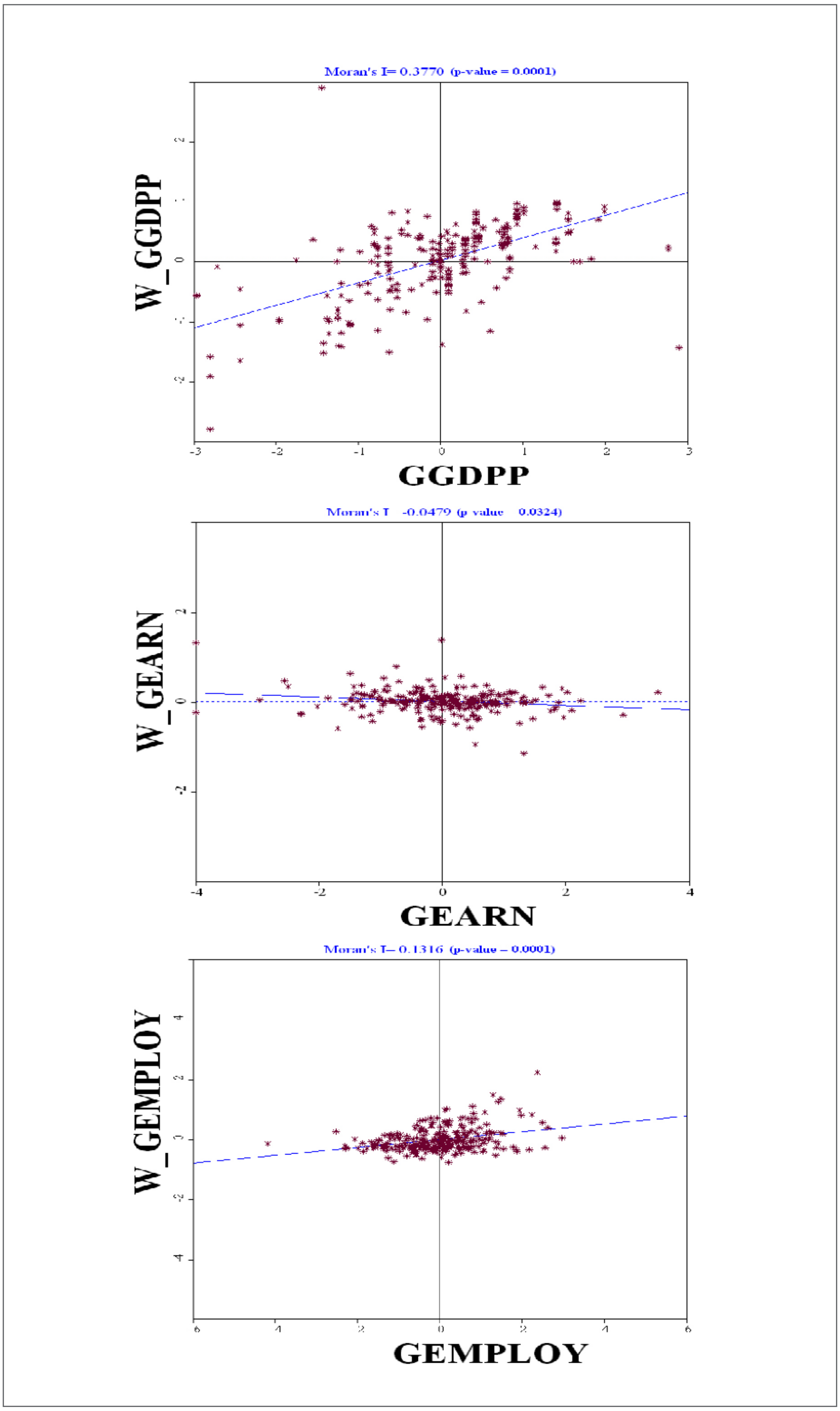


Figure 8, Figure 9 and Figure 10 show cluster maps for the LISA analysis for levels, differences and growth rates, respectively. A positive value for LISA indicates spatial clustering of similar values (either high or low) whereas a negative value points to spatial clustering of dissimilar values between an area and its neighbours.

Figure 8 clearly identifies high-high (HH) clustering (in red) in Southeast England for output per worker and earnings. On the other hand, clusters of low values in output per worker and earnings are located in the North, South-West and Wales. It turns out that much of the clustering of low values in Wales and the North is an artifact of the way we have allocated output per worker from the NUTS3 level to Local-Authority/City-region level. Appendix 5 shows the picture for output per worker at the original NUTS3 level, and shows a cluster of low output per worker areas around and to the south west of North Manchester, but less evidence of this in Wales, further north or around Leeds. Whether we look at Local-Authority/City region level or NUTS3, the results show a low output per worker cluster (in blue) around Manchester and its neighbours (which includes Leeds), suggesting that output per worker in this neighbourhood is unusually low given the output per worker in Manchester. Output per worker in Local-Authority areas to the south west of Manchester (Cheshire, and South Manchester – see the NUTS3 analysis in Appendix 5) tends to be high relative to the surrounding areas.

It is interesting to note that a cluster around Manchester is not revealed in the pattern for earnings. In this case, there is no statistical evidence that areas around Leeds or Manchester tend to be similar in their levels of earnings. However, as is clear from the map, the South-East has a cluster of relatively high-earning areas (unsurprising) and is the outlier when compared to the rest of Britain. Manchester and Leeds are not unusual relative to other larger urban areas located outside the South East. For employment, Manchester and Leeds do not show a particularly strong pattern but the contrast here is no longer with the South East but with a few geographically small areas that have particularly low employment.

Figure 8 – LISA Cluster Map (levels of GDP per worker, earning and employment in 2006)

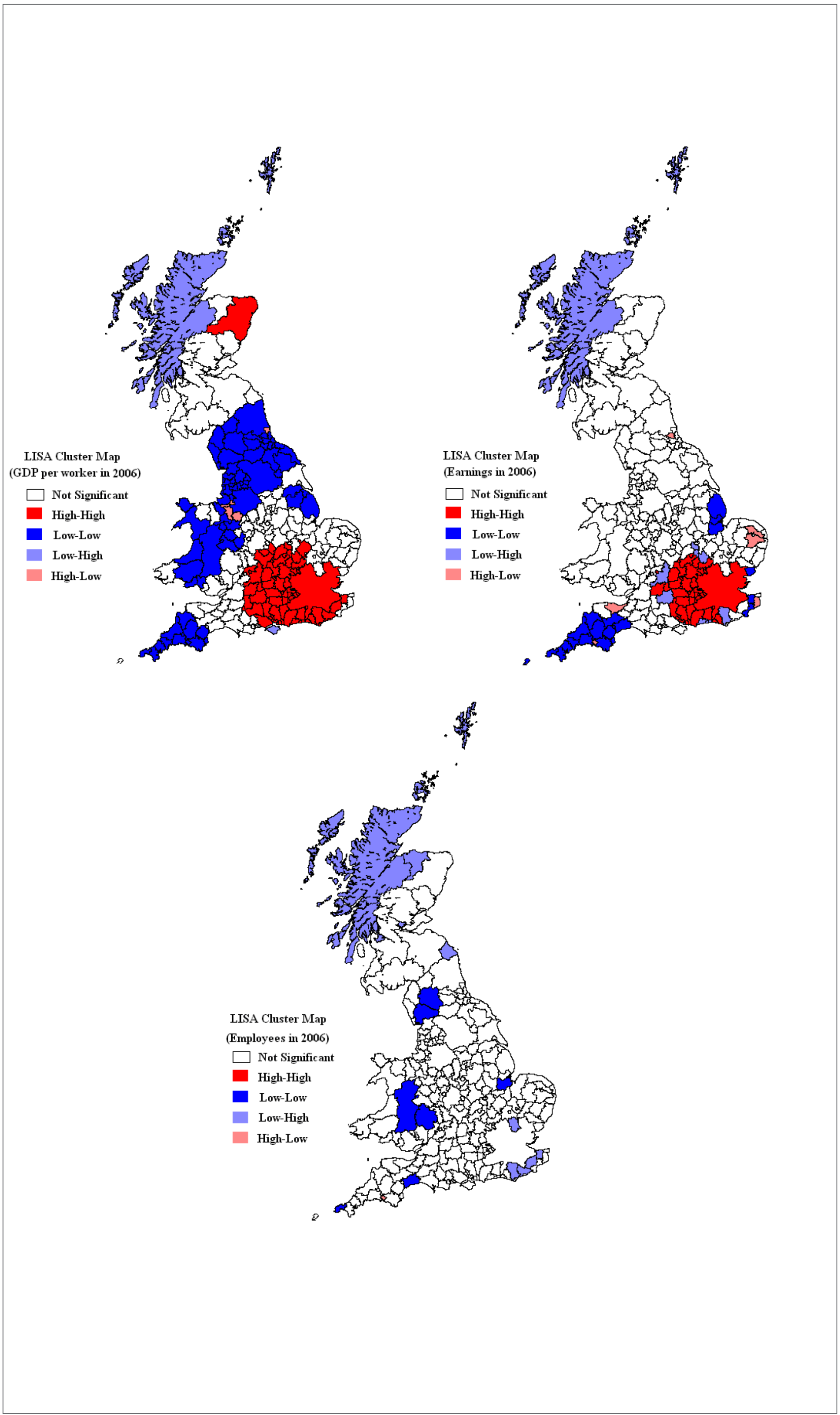


Figure 9 – LISA Cluster Map (Differences in GDP per worker, earnings and employment between 1998 and 2006)

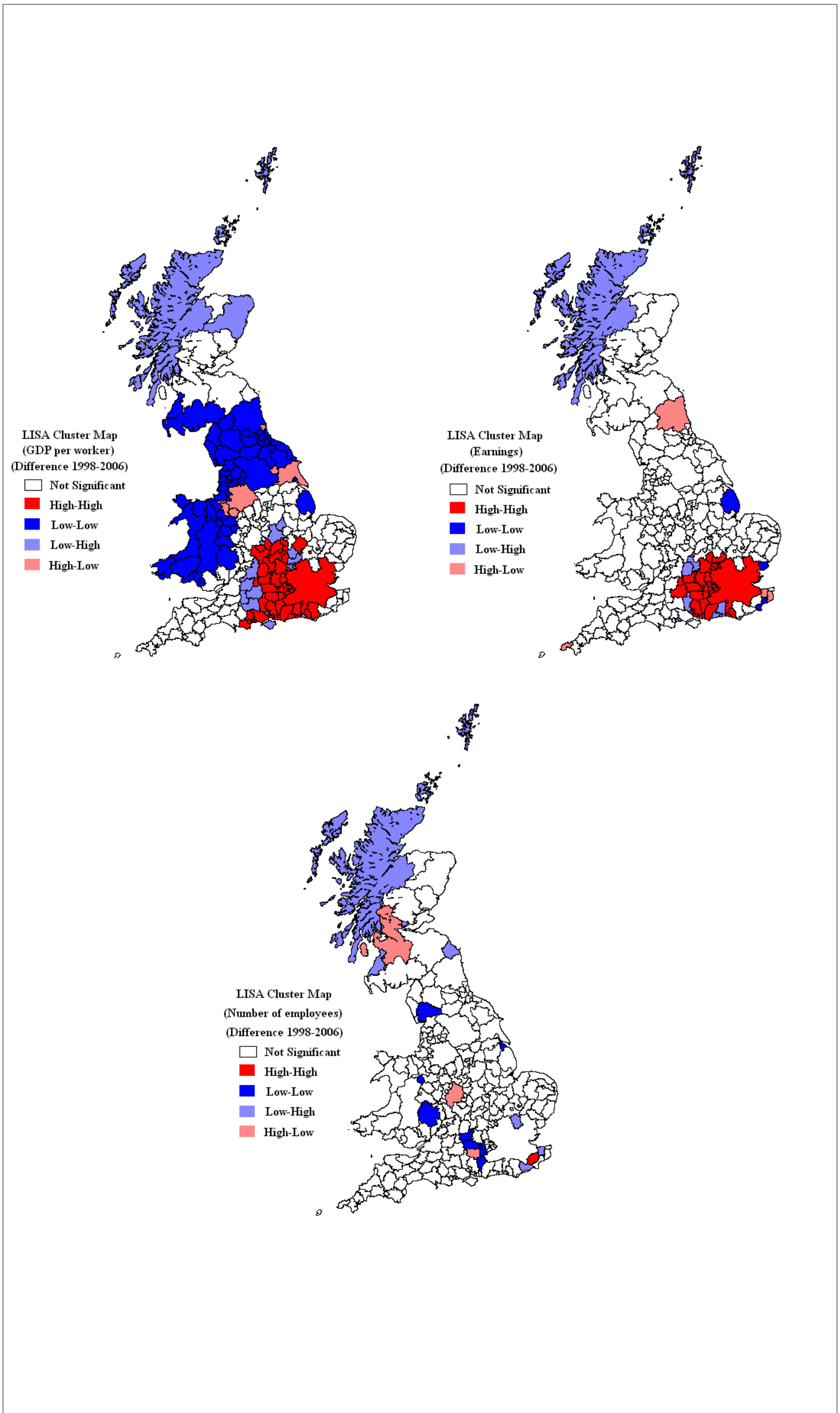
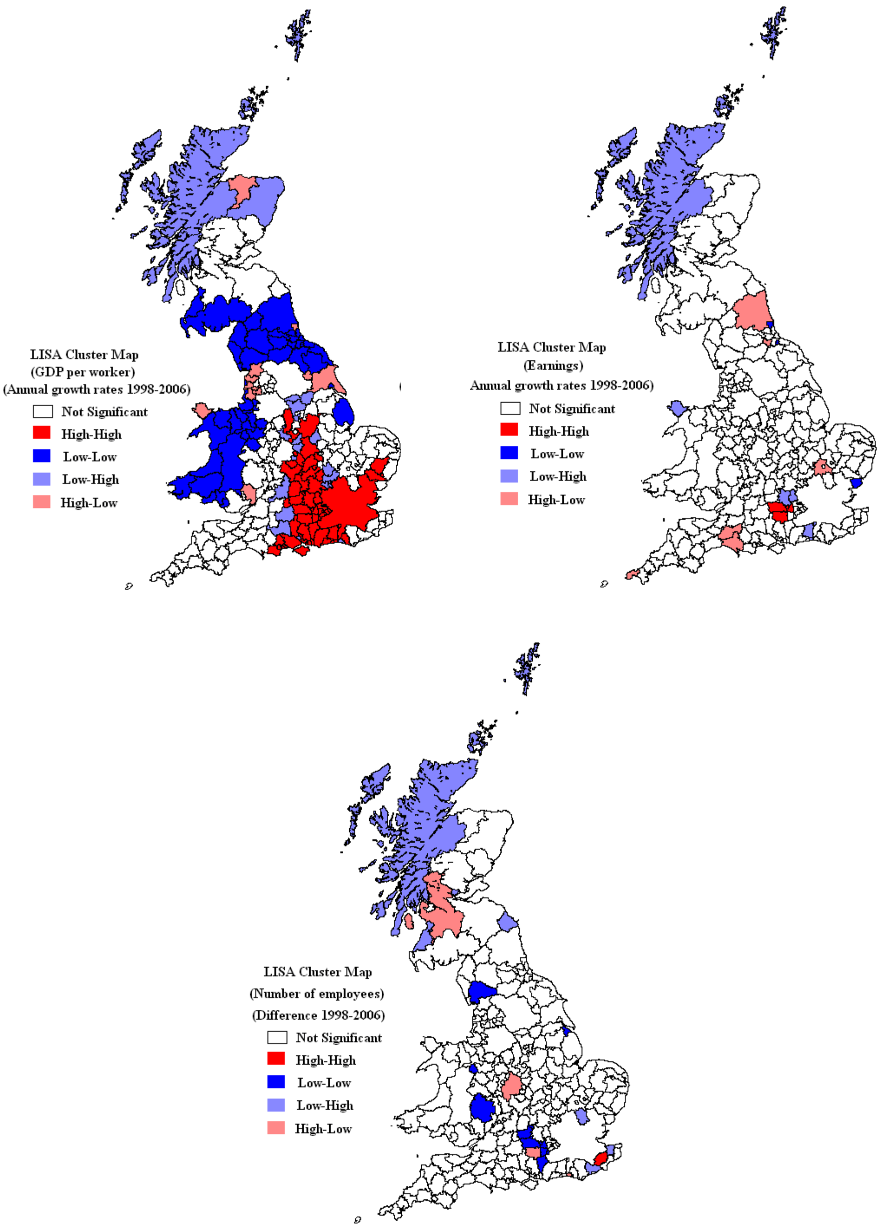


Figure 10 – LISA Cluster Map (Annual growth rates of GDP per worker, earnings and employment between 1998 and 2006)



The patterns in Figure 8 show us where there are spatial clusters of high and low values for the levels of output per worker, earnings and employment. But the more interesting question when considering economic linkages is whether there is a tendency for these values to move together over time in neighbouring places. To answer this question, Figure 9 and Figure 10 show the cluster maps of differences and growth rates of output per worker, earning and employment. Again, there are issues with allocating output per worker to Local-Authority/city-region level, so Appendix 5 provides a NUTS3 level analysis for comparison.

Here, the stand out point relating to output per worker around Manchester and Leeds can be observed in the first panel of Figure 9 (and second panel of Appendix 5). Manchester and Local Authority areas to its south west (specifically South Manchester) have moved in the opposite direction to their neighbours in terms of changes in output per worker. This is an unusual pattern relative to the rest of Britain, and overall there is no evidence of a tendency for Manchester and Leeds to move together in terms of change in output per worker. Again, the picture is different for earnings and employment change, with no indication of any spatial linkages in Local Authority areas around Manchester-Leeds, and little sign of linkages anywhere else outside hotspots in the South East. Switching to growth (% change) as the metric in Figure 10, shows no spatial linkages between Manchester-Leeds and their neighbours in output per worker or earnings, although Manchester appears as a low employment growth cool spot in the third panel. The most interesting point to take out of this analysis is that recent changes in output per worker in the Manchester City Region has been unusual positive relative to the rest of Great Britain, but it has not been associated with similar changes in the surrounding areas. More generally however, there are few signs that Manchester-Leeds are exceptional in their strength or weakness in spatial linkages. London and the South East appear predominantly as the outlier in respect of strong positive spatial linkages relative to the rest of Great Britain. We now turn to try to understand what might have caused these patterns.

4.3 Spatial Econometric Analysis

So far, we have examined the tendency for area outcomes to move with their neighbours. We defined the spatial weight matrix (a 70 kilometres band) so that Manchester and Leeds count as “neighbours” and then we used exploratory spatial techniques to examine whether their interaction is different from what we see in Great Britain as a whole.

Given the results we obtained, it is interesting to attempt to further explore the extent to which measurable characteristics of areas drive these interactions. Since we have found that Manchester and Leeds are particularly unusual in terms of their strong spatial autocorrelation in output, weak correlation in terms of earnings and employment and particularly weak in terms of some aspects of growth, it would be interesting to know what characteristics of Manchester and Leeds might explain this. As with our work on commuting, the strategy is to look at the nature of these relationships across Great Britain to help identify the factors that cause the patterns we observe.

We start our analysis with a basic equation:

$$Y = X\beta + \varepsilon$$

where Y is the dependent variable for each area. Eight dependent variables are used:

(i) Output per worker in 2006;
(ii) Earnings in 2006;
(iii) Difference in output per worker between 1998 and 2006
(iv) Difference in earnings between 1998 and 2006;
(v) Difference in employment between 1998 and 2006;
(vi) Annual growth rates of output per worker between 1998 and 2006;
(vii) Annual growth rates of earnings between 1998 and 2006;
(viii) Annual growth rates of employment between 1998 and 2006.

X is a matrix of area characteristics that may be important in explaining the behavior of the dependent variables and is the error (or unexplained part of the dependent variable). We include a group of variables which represent local sectoral composition, local occupation composition and education levels (all these variables are in percentage terms) as well as the average age of the population. See appendix 4 for exact definitions.

Once we have estimated these equations we take the residual, or unexplained part of the dependent variable ε and examine whether it is spatially autocorrelated using the same approach as we did above. If we find it is not, then the spatial interaction that we see between Manchester and Leeds is explained by them sharing similar characteristics. If we continue to observe spatial interaction in these residuals, then we would conclude that there is something specific about the interaction between Manchester and Leeds that cannot be explained by observed characteristics.

Table 5: OLS Results and Diagnostics for Spatial Dependence (Levels)

	Dependent Variable	Level of GDPpw 2006		Level of Earnings 2006	
	Estimation method: OLS	coefficient	P-value	coefficient	P-value
	Constant	27.038	0.209	9.703	0.006
Local sectoral composition	Agriculture and fishing in 2006(%)	-55.885	0.596	1.812	0.916
	Energy and water in 2006(%)	8.376	0.887	22.488	0.020
	Manufacturing in 2006(%)	29.336	0.037	0.862	0.706
	Construction in 2006(%)	-22.959	0.488	-0.679	0.899
	Distribution, hotels and restaurants in 2006(%)	13.377	0.446	-4.051	0.157
	Transport and communications in 2006(%)	20.008	0.241	0.606	0.827
	Banking, finance and insurance, etc in 2006(%)	59.745	0.000	13.400	0.000
	Other services in 2006(%)	26.806	0.547	-3.879	0.592
Education	Level 1 in 2001(%)	96.550	0.000	4.921	0.234
	Level 2 in 2001(%)	53.851	0.071	3.267	0.500
	Level 3 in 2001(%)	44.250	0.300	-6.648	0.339
	Level 4 and 5 in 2001(%)	50.564	0.004	11.120	0.000
Local occupation composition	Managers and Senior Officials in 2006(%)	5.730	0.787	4.311	0.212
	Professional Occupations in 2006 (%)	-10.327	0.642	4.072	0.260
	Associate Prof & Tech Occupations in 2006(%)	-24.420	0.262	0.120	0.973
	Admin and Secretarial Occupations in 2006(%)	-27.756	0.243	5.368	0.165
	Skilled Trades Occupations in 2006(%)	-30.616	0.138	-2.001	0.550
	Personal Service Occupations in 2006(%)	21.559	0.380	2.640	0.509
	Sales & Customer Service Occupations in 2006(%)	-0.633	0.981	4.272	0.326
	Process, Plant and Machine Operatives in 2006(%)	-32.594	0.192	2.182	0.591
	Average age of the population in 2006	-0.185	0.587	-0.099	0.076
	Adjusted R-squared	0.33464		0.548677	
	Diagnostics For Spatial Dependence	Test	P-Value	Test	P-Value
	Moran's I (error)	3.763	0.000	2.209	0.027
	Lagrange Multiplier (lag)	80.326	0.000	0.829	0.363
	Robust LM (lag)	72.482	0.000	0.166	0.684
	Lagrange Multiplier (error)	8.422	0.004	2.247	0.134
	Robust LM (error)	0.577	0.447	1.584	0.208

Notes: To avoid perfect multicollinearity, some variables are excluded from the regressions. Local sectoral composition, excluded variable: Public administration, education & health. Education level, excluded variable: No qualification and other qualifications/level unknown. Local occupation composition, excluded variable: Elementary occupations.

Table 6: OLS Estimation Results and Diagnostics for Spatial Dependence (in Differences

	Dependent variable	Diff in GDPpw		Diff in Earnings	
	Estimation method: OLS	**coeff.**	**P-value**	**coeff.**	**P-value**
	Constant	-21.155	0.066	6.434	0.012
Local sectoral composition	Agriculture and fishing in 1998 (%)	116.165	0.014	2.042	0.844
	Energy and water in 1998 (%)	-17.838	0.441	0.377	0.943
	Manufacturing in 1998 (%)	10.525	0.072	-2.032	0.121
	Construction in 1998 (%)	-37.176	0.019	0.126	0.972
	Distribution, hotels and restaurants in 1998 (%)	12.134	0.149	-2.646	0.158
	Transport and communications in 1998 (%)	6.152	0.414	-0.570	0.740
	Banking, finance and insurance, etc in 1998 (%)	33.945	0.000	2.259	0.159
	Other services in 1998 (%)	-6.962	0.744	-2.326	0.625
Education	Level 1 in 2001 (%)	36.683	0.008	-1.208	0.690
	Level 2 in 2001 (%)	1.703	0.904	2.195	0.480
	Level 3 in 2001 (%)	9.330	0.656	-7.243	0.120
	Level 4 and 5 in 2001 (%)	6.335	0.444	4.436	0.017
Local occupation composition	Managers and Senior Officials in 1998 (%)	0.987	0.901	-0.767	0.663
	Professional Occupations in 1998 (%)	10.165	0.300	-1.421	0.517
	Associate Prof & Tech Occupations in 1998 (%)	11.715	0.277	2.275	0.343
	Administrative and Secretarial Occupations in 1998 (%)	-6.643	0.568	-1.168	0.652
	Skilled Trades Occupations in 1998 (%)	1.419	0.855	-1.757	0.315
	Personal Service Occupations in 1998 (%)	14.917	0.180	-2.994	0.225
	Sales and Customer Service Occupations in 1998 (%)	-4.642	0.686	-1.550	0.546
	Process, Plant and Machine Operatives in 1998 (%)	-5.215	0.643	-2.681	0.285
	Average age of the population in 1998	0.097	0.543	-0.058	0.105
	GDP per worker in 1998	0.294	0.000		
	Earnings in 1998			0.059	0.349
	Employment in 1998				
	Adjusted R-squared	0.415		0.220	
	Diagnostics For Spatial Dependence	**Test**	**P-value**	**Test**	**P-value**
	Moran's I (error)	3.785	0.000	-0.163	0.871
	Lagrange Multiplier (lag)	22.744	0.000	0.401	0.526
	Robust LM (lag)	14.767	0.000	0.055	0.815
	Lagrange Multiplier (error)	7.988	0.005	0.458	0.499
	Robust LM (error)	0.011	0.918	0.111	0.739

Notes: To avoid perfect multicollinearity, some variables are excluded from the regressions. Local sectoral composition, excluded variable: Public administration, education & health. Education level, excluded variable: No qualification and other qualifications/level unknown. Local occupation composition, excluded variable: Elementary occupations.

and Growth rates between 1998 and 2006)

Diff in Employment		Growth of GDPpw		Growth of Earnings		Growth of Employment	
coeff.	P-value	coeff.	P-value	coeff.	P-value	coeff.	P-value
-7610.629	0.730	-0.001	0.942	0.100	0.000	-0.025	0.437
-103242.200	0.254	0.206	0.013	0.026	0.808	-0.047	0.728
-22825.960	0.608	-0.013	0.748	-0.013	0.812	-0.074	0.266
-23143.050	0.040	0.022	0.030	-0.024	0.079	-0.041	0.016
41996.150	0.169	-0.065	0.019	-0.006	0.862	0.095	0.037
-189.927	0.991	0.019	0.196	-0.020	0.307	0.042	0.081
-16493.580	0.256	0.013	0.331	-0.014	0.436	-0.032	0.137
-4830.735	0.708	0.067	0.000	0.003	0.846	-0.011	0.577
46784.880	0.256	-0.022	0.548	-0.026	0.603	0.026	0.670
19527.530	0.454	0.070	0.004	-0.012	0.700	-0.021	0.581
3057.103	0.910	0.018	0.469	0.026	0.431	0.103	0.011
560.568	0.989	0.027	0.461	-0.059	0.227	0.075	0.214
12448.660	0.431	0.011	0.428	0.044	0.023	0.020	0.385
-2387.840	0.875	0.001	0.929	-0.018	0.343	-0.041	0.070
-5315.851	0.778	0.020	0.231	-0.035	0.134	-0.003	0.910
2274.411	0.913	0.022	0.236	0.012	0.644	-0.021	0.491
18820.650	0.400	-0.012	0.555	-0.016	0.546	0.000	0.996
3957.034	0.792	0.002	0.896	-0.027	0.142	0.005	0.834
-5339.903	0.802	0.030	0.127	-0.045	0.082	-0.011	0.725
-308.669	0.989	-0.010	0.600	-0.023	0.395	-0.075	0.022
22620.400	0.297	-0.008	0.675	-0.024	0.370	0.061	0.058
56.118	0.854	0.0003	0.332	-0.001	0.125	0.0004	0.388
		-0.0003	0.002				
				-0.002	0.014		
0.065	0.000					0.0000	0.843
0.912		0.226		0.056		0.178	
Test	P-value	Test	P-value	Test	P-value	Test	P-value
0.293	0.770	6.097	0.000	-0.280	0.780	2.485	0.013
1.348	0.246	51.485	0.000	1.229	0.268	6.806	0.009
1.388	0.239	28.122	0.000	0.616	0.433	5.524	0.019
0.064	0.800	23.939	0.000	0.613	0.434	3.023	0.082
0.104	0.747	0.577	0.448	0.000	0.984	1.740	0.187

Table 5 and Table 6 report the Ordinary Least Square (OLS) results and diagnostics for spatial dependence. The way to read the table is to look for p-values less than 0.05 or 0.10. These identify variables that have a statistically significant effect on the dependent variable at the 5% or 10% level, respectively. So, for example, for the level of output per worker, places with high shares of manufacturing and banking sector have higher output per worker. Places with lower levels of education (level 1: e.g., Foundation GNVQ) and higher levels of education (levels 4 and 5: e.g., Higher Degree) also have higher output per worker. With the relatively broad occupational classifications that we use we do not observe any significant effect on output per worker or earnings (and similarly for age composition). There are certainly some puzzles in these results, but our key interest is whether the spatial dependence between Manchester and Leeds remains or disappear after conditioning on the exploratory variables that, themselves, have very strong spatial or geographic patterns.

To consider this, as explained earlier, we take the residuals from the OLS estimation of equation 3 (Table 5 and Table 6) and produce the same LISA maps as in Figure 8, Figure 9 and Figure 10. The results are reported in Figure 11, Figure 12 and Figure 13, respectively. From the maps we can see that Manchester-Leeds spatial dependence is no longer unusual once we control for observable characteristics of Manchester and Leeds. In other words, this result tells us that the characteristics we included in the models (Table 9 and Table 10) explained the unusualness identified in the exploratory spatial data analysis (ESDA).

Figure 11 – LISA Cluster Map (OLS Residuals of levels of GDP per worker and earning in 2006)

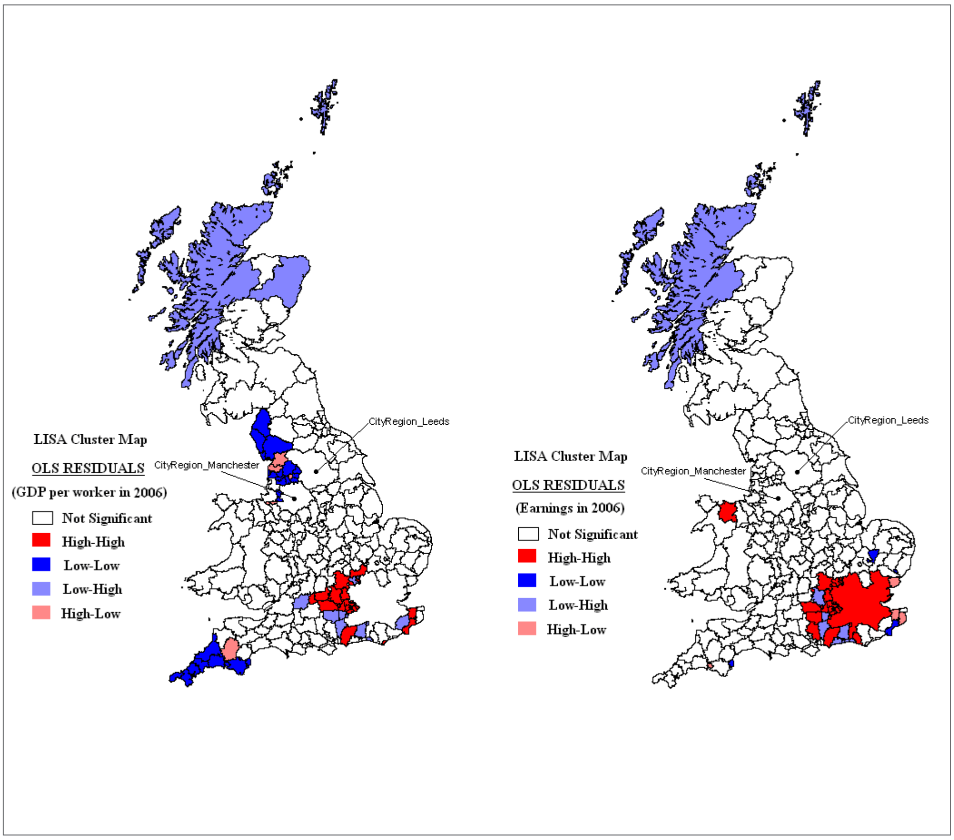


Figure 12 – LISA Cluster Map (OLS Residuals of Differences in GDP per worker, earnings and employment 1998-2006)

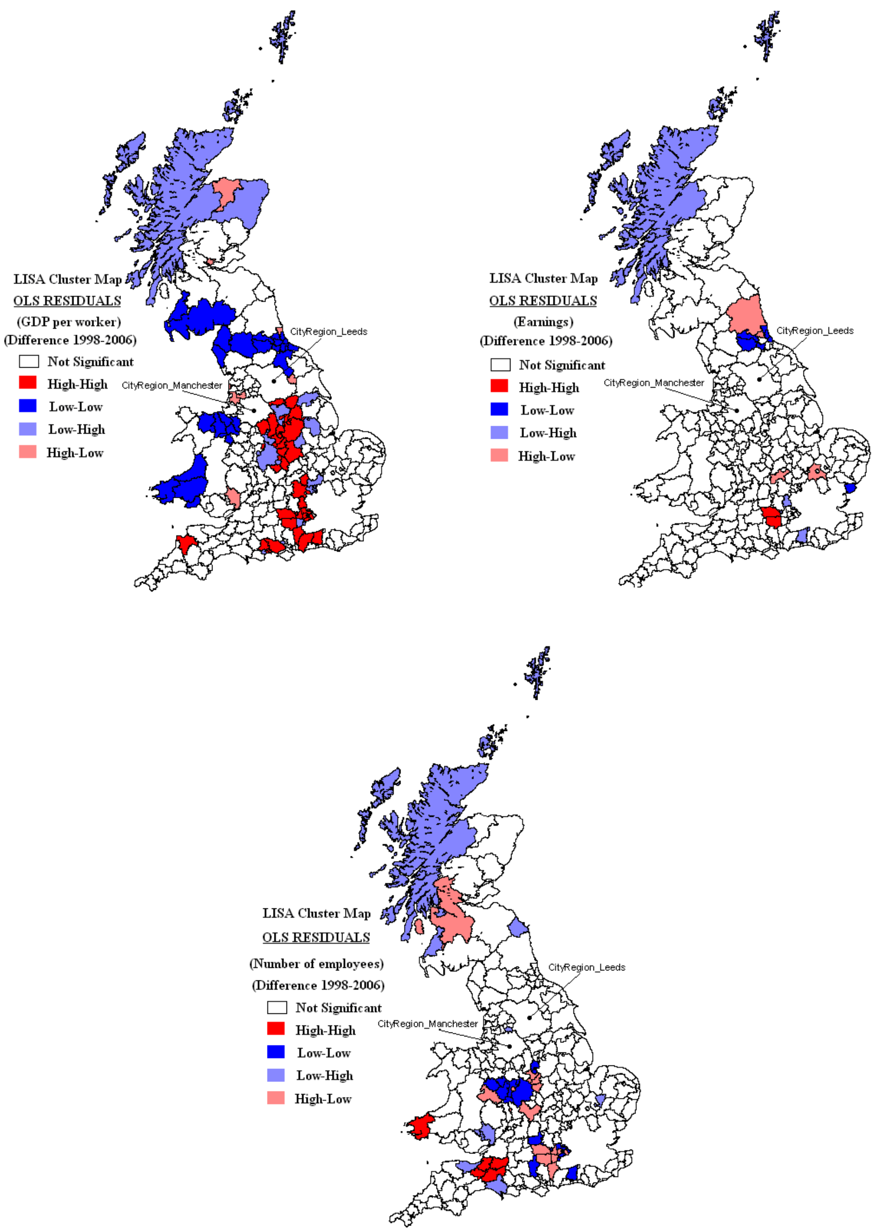
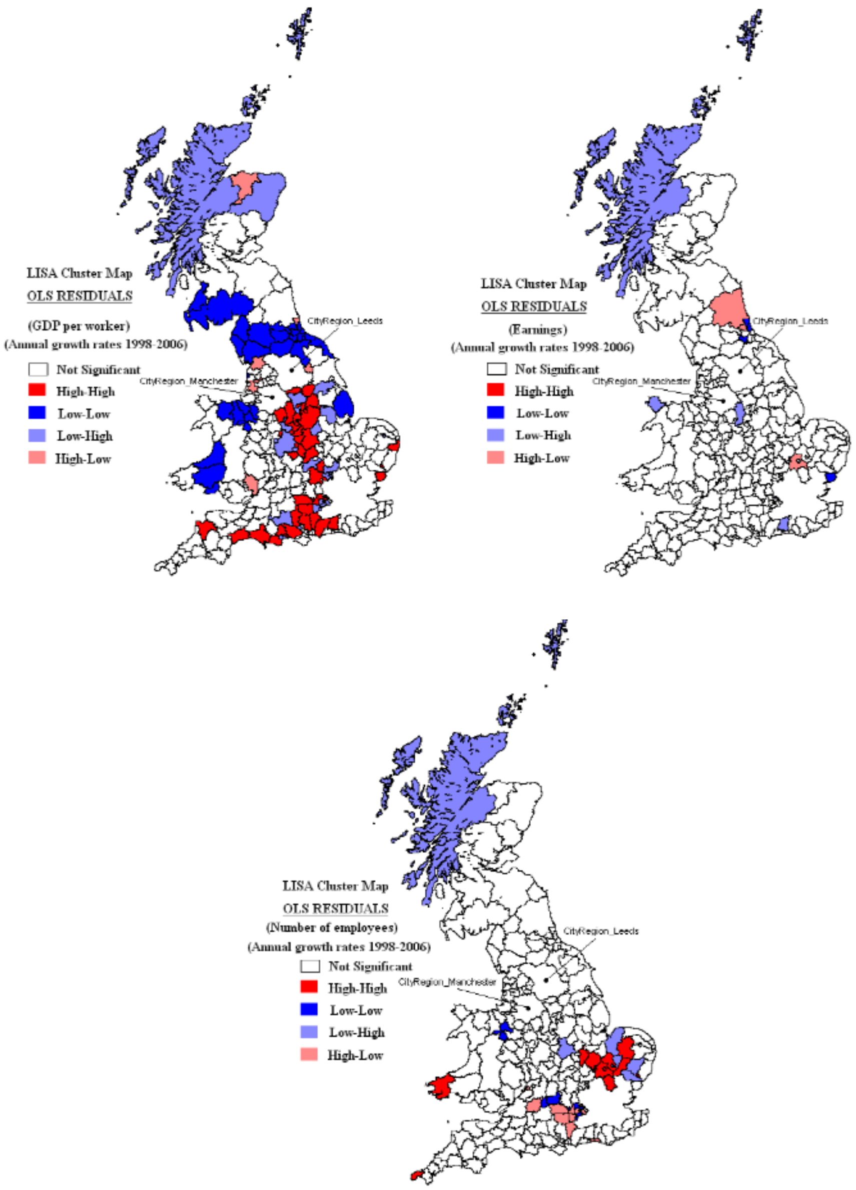


Figure 13 – LISA Cluster Map (OLS Residuals of Annual growth rates of GDP per worker, earnings and employment 1998-2006)



4.4 Conclusions – Spatial Econometric Analysis

- Exploratory spatial econometric analysis indicates that there are some distinct spatial patterns in output per worker, and changes in output per worker around Manchester and Leeds. In particular, South Manchester and areas to its south west are unusual in the extent to which recent positive changes in output per worker have not been linked to positive changes in surrounding areas.

- More generally, for earnings, employment and growth in output per worker there is little evidence of co-movement between Manchester-Leeds and their neighbours. Manchester-Leeds is not unusual in this respect when judged against other urban areas in Britain. Manchester-Leeds appears unusual when compared to London and the South East, because this area exhibits unusually strong positive links in terms of levels and changes in economic indicators.

- Any differences from general GB patterns are explained by a few structural, economic characteristics of the two areas. As with commuting, this finding points away from social, cultural or similar factors as drivers of weak linkages between the cities (although we do not study these factors directly). It suggests that other unexplained factors are unlikely to constrain Manchester and Leeds from following the general GB pattern.

- This analysis reminds us that the interactions between places are as much outcomes of the underlying structural characteristics of those places as they are drivers of changes in those structural characteristics. Given the current industrial and skills structures of the Manchester and Leeds city regions the correlations in terms of outcomes are about what we would expect.

- Overall, this suggests that structural change would be likely to play an integral part in increasing the extent of observed interaction between the two city-region economies.

5. Agglomeration and labour markets

The work discussed so far describes and analyses existing interactions in terms of a direct measure of linkages (commuting) and outcomes (earnings, employment and output per worker). As we explained in the introduction we view this as both telling us something about current behaviour and the *feasibility* of increasing interaction. We now turn to the possible impacts of increasing integration. In this section we consider the agglomeration benefits of increased productivity focusing, in particular, on the functioning of labour markets.

Our starting point is the observation that, all else equal, larger places tend to have higher productivity and wages. Economists refer to the productivity benefits associated with increased levels of economic activity as agglomeration economies (or benefits). At their broadest level, *agglomeration economies* occur when individuals and firms benefit from being near to others. We will refer to this as the effect of better access to economic mass. This report focuses on agglomeration economies that arise in production. It is important to remember, however, that there may be other benefits of agglomeration, for example in terms of consumption.

With this focus, agglomeration economies arise because of the production benefits of physical proximity. Physical proximity to other firms, workers and consumers, may help firms in the day-to-day business of producing goods and services. This implies that the productivity, of individual firms will rise with the overall amount of activity in other nearby firms, or with the number of nearby workers or consumers. Physical proximity may also facilitate the flow of ideas and knowledge leading firms to be more creative and innovative. Higher productivity, in turn, tends to lead to higher wages for workers.

The literature traditionally emphasises three sources of agglomeration economies: linkages between intermediate and final goods suppliers, labour market interactions, and knowledge spillovers. Input-output linkages occur because savings on transaction costs means firms benefit from locating close to their suppliers and customers. Larger labour markets may, for example, allow for a finer division of labour or provide greater incentives for workers to invest in skills. Finally, knowledge or human capital spillovers arise when spatially concentrated firms or workers are more easily able to learn from one another than if they were spread out over space. In this report, we are only concerned with the overall effect of increased access to economic mass. MIER (2008) includes a much more detailed discussion of the sources of different agglomeration economies and provides a review of the existing literature.

Existing work for The Northern Way[11] has followed Department for Transport guidance on evaluating the wider economic impacts of transport schemes to address this question. This approach uses estimates of the strength of agglomeration economies, coupled with assumptions on the extent to which integration would increase local economy size to work out the productivity impacts on different sectors of the economy. We use labour market data to try to understand whether this existing work captures all the likely impacts of increased integration.

It has been suggested that the size of the Manchester and Leeds economies may have negative implications for labour market outcomes, particularly for more highly educated workers, and that this may be an important factor in explaining their relative under performance[12]. To examine this possibility we use data on individual wages to see how the level and growth of wages are affected by the size of the local labour

11. See Agglomeration Simulation Exercise, Steer Davies Gleave (November 2006) for The Northern Way; and Model Development and Results for The Northern Way using the South and West Yorkshire Dynamic Model, Steer Davies Gleave (December 2006).

12. See, for example, the Manchester Independent Economic Review's work on skills which considers this issue.

market. This is just another way of trying to identify the overall agglomeration benefits that have been studied in the previous research referred to above.

In the academic literature it is increasingly recognized, however that the changing composition of the labour market may account for a large part of the overall positive relationship between wages, productivity and city size. For example, it may be that large cities tend to attract more educated workers. Because more educated workers also tend to earn more this leads to a positive relationship between city size and wages. In this scenario when we measure agglomeration economies by looking at how wages change with city size we are capturing the fact that the *composition* of the labour force changes with city size. Alternatively, it could be that larger cities actually make workers more productive whatever their education level. That is, there is a *place-based* effect whereby larger cities pay higher wages. Existing work on agglomeration economies (including that for the MIER, DfT and The Northern Way) has focused on assessing the overall impact of city size. In contrast, our research assesses the extent to which these overall benefits arise from changing composition as opposed to higher wages for existing workers. We then use our estimates, coupled with realistic assumptions about policy induced changes in transport costs, to assess the impact of increased integration on labour market outcomes. This allows us to paint a much richer picture of the potential sources of gains, the distribution of benefits and the kinds of structural changes that might be needed to achieve them.

5.1 Methodology and Data

To assess the magnitude of overall agglomeration benefits we need to see how wages differ with city size. We then want to break these overall benefits down in to those that come from changing composition versus those that come from place-based effects. To do this, we need to be able to look at the wage levels for individuals who are otherwise identical but who happen to live in different cities. Ideally, we would do this by randomly allocating people across cities. In reality, fortunately, the UK government does not decide in which cities people live. This creates a problem for us, however, because people are able to sort across cities in non-random ways. Imagine, for a moment, that we are able to observe everything about an individual (age, sex, education) that might affect their wage. Then, even in the absence of random allocation across cities, we can still identify place effects on wages by comparing wages for people with identical observable characteristics who happen to live in different places.

Unfortunately, even with quite detailed data, we cannot usually be certain that we are observing everything about an individual that might affect wages. For example, in the data that we use, we have no information on cognitive abilities or motivation. So when we compare two people with identical observed characteristics it may be that the one with higher ability lives in the larger city and thus earns a higher wage. It is the unobserved individual characteristics (ability) that explains the higher wage of the individual in the larger city but we mistakenly attribute it to the effect of city size. If, on average, higher ability individuals live in bigger cities, then we will find a statistical relationship between city size and wages even though city size has no direct causal effect on wages. In this example, the relationship between city size and wages comes about instead because labour markets in bigger cities differ in their ability composition from labour markets in smaller cities.

One way to get round this is to follow the same individual as they move across cities. Providing that ability is fixed over time if we see the same individual earning more in larger cities we may be more confident in attributing this to a place-based effect rather than a composition effect. Even then, we cannot rule out the possibility that something changed for the individual that both affected their earnings potential and their choice of place to live. Still, in the absence of random allocation of people (or a policy change that as good as randomly assigns people) tracking individuals over time is the best that we can do to identify true place-based effects of changing city size.

As is clear from this discussion we need data on individuals that provides information on where they work, on their wages and on the individual characteristics that might affect wages. We would also like to be able to follow individuals over time, particularly as they move across cities. Such data is available from the New Earnings Survey and the Annual Survey of Hours and Earnings (NES/ASHE).

We use data on individual wages calculated as the basic hourly wage from NES/ASHE. Data on individual characteristics – age, age squared, gender and occupation come from the same source. ASHE does not provide years of education so we construct these using cohort of birth-by-SOC matching on data from the Labour Force Survey (LFS) which contains information on both occupations and education. The way that we do that is described in Appendix 1. Individual occupation levels comes from NES/ASHE and are recorded using SOC1990 for 1998-2001 and SOC2000 for 2002 onwards. Data on the characteristics of an individual's job (public sector, part time, collective agreement) also come from NES/ASHE. NES/ASHE also includes the industry of occupation recorded using SIC2003. The information on aggregate employment and the industrial structure of an area comes from the Business Structure Database (BSD) which records the postcode address, employment, and turnover of all VAT or PAYE registered businesses in the UK. For the occupation structures of areas we need to aggregate up from the individual data in ASHE. Other area level variables – a Herfindahl index of industrial diversity, shares of industry sectors at 1-digit level – are based on BSD data. Finally, area proportions of workers belonging to high-skill and intermediate-skill groups are based on LFS data. More detail on the NES/ASHE dataset is provided in the commuting section of this report.

We follow existing research for The Northern Way, Dft and the MIER by focusing on the relationship between wages and 'access to economic mass' rather than between wages and city size. The problem with the latter is that it requires us to impose city boundaries on a map and talk about workers being located in a particular city. In a sub-national context these boundaries are essentially arbitrary (at least when it comes to the working of agglomeration economies). Measures of access to economic mass treat space as continuous by taking into account access to all other areas discounted by distance or transport cost to these areas, and avoid the need to impose such arbitrary boundaries. We construct two measures of access to economic mass as follows. The first measure is based on Generalised Transport Costs (GTC) when driving. We first use a mapping from postcodes to wards to calculate total employment in each ward from the BSD. The access to economic mass for ward j in a given year is calculated by adding up contemporaneous employment in all other wards using inverse-GTC (driving) weighting. This inverse-GTC weighting applies a weight of GTC^{-1}_{ij} to the employment in ward j, where GTC_{ij} is the ward-to-ward GTC for driving as described in the commuting section. Therefore, a ward is assigned an aggregate of employment in other wards, with employment in more distant places contributing less than employment close by. The equation for the access to economic mass measure in ward i is thus: $A_{it} = \Sigma emp_{jt} \, x \, GTC^{-1}_{ij}$. To allow employment in ward i to contribute to its own access to economic mass, we set $GTC_{ii} = 0.5 \, x \, GTC_{i1}$ where GTC_{i1} is the minimum ward-to-ward GTC for ward i (i.e. the "closest" ward). Each worker is assigned the access to economic mass value equal to the ward in which their employer is located[13]. Note that this index of access to economic mass is identical to the effective density index used by Graham (2006) (although we prefer to refer to it as a measure of *access to economic mass* or *employment accessibility*). Our second access to economic mass measure is calculated using train GTC. We only have LA-to-LA GTC so LAs, rather than wards, are the underlying unit used to construct the index. The index is otherwise constructed in an identical manner to the index based on driving GTC. For simplicity, we will refer to the employment accessibility measure based on driving GTC as "Car Accessibility" and that based on train GTC as "Train Accessibility".

5.2 Results

We start by considering a simple model that captures overall agglomeration economies, ignoring the distinction between composition and place-based effects. To do this, we run regressions that explain the wages of an individual as a function of the access to economic mass at the individual's work place:

$$\ln(w_{it}) = \alpha_t + \theta \ln(A_{it}) + \varepsilon_{it}$$

where w_{it} is individual i's wage at time t, A_{it} is one, or both, of the access to economic mass variables described above, ε_{it} is a residual, α_t is a time varying parameter and θ a time invariant parameter (both to be estimated). The alphas capture the fact that wages tend to rise over time, while the theta captures the impact of increases in access to economic mass (the impact assumed constant over our relatively short time period). Results are reported in Table 7. We report the estimated coefficient, the associated standard error, the R-squared (which tells us the percentage of variation in wages that are explained by access to economic mass) and the number of observations.

In the first panel we report results using only Car Accessibility (column 1), then only Train Accessibility (column 2) then both together (column 3). We see that when entered separately, unsurprisingly, both are positively and statistically significantly associated with wages. When we include both together we find the effect of Car Accessibility is positive but insignificant while that of Train Accessibility is both positive and significant. At this stage, one shouldn't read too much in to the difference between the coefficients on the two measures. As we will see below this difference depends crucially on what other characteristics of individuals are being controlled for in the regressions. The effect of access to economic mass remains essentially unchanged if we drop all individuals that work in the London Travel to Work Area (column 4). Finally, for comparison we present results based on TTWA employment rather than access to economic mass (column 5).

In terms of magnitudes, the coefficient on TTWA employment in column 5 is the easiest to interpret. As it is in logs, the coefficient is an elasticity and tells us that a 10% increase in TTWA employment is associated with a 0.7% increase in wages. This is consistent with the existing literature of the effect of city size on productivity which reports the effect of a 10% increase in city size varying from around 0.2% to 2% with most under 1%. The coefficients on the access to economic mass measures are harder to interpret because they are calculated using GTC weighting of employment across all wards (driving) and LAs (train). Taken at face value, the

13. While ASHE contains information on both home and work postcode, NES only provides the latter so we need to base our measure of access to economic mass on work rather than home location.

coefficient of 0.344 for Train Accessibility implies that a 10% increase in employment in all wards in Britain, or a 10% reduction in the GTC between all Local Authority areas in Britain, would increase wages by around 3.4%. For the moment, it is simplest to focus on how these coefficients change as we introduce individual characteristics. Later, however, we calculate changes in economic mass consistent with proposed transport interventions which we then use to give a feeling for the magnitude of the effects on wages.

Table 7: Regressions of wages on access to economic mass

	1	2	3	4	5
	Car only	Train only	Both	Without London	Using TTWA Employment
ln Car Accessibility	0.230*		0.084	-0.040	
	(0.092)		(0.122)	(0.035)	
ln Train Accessibility		0.344***	0.258**	0.217***	
		(0.093)	(0.093)	(0.036)	
ln Employment					0.069***
					(0.008)
R2	0.085	0.086	0.09	0.06	0.085
Observations	1102527	1119582	1102527	884953	1119582

Notes: All models have log hourly earnings as a dependent variable and the explanatory variables of interest are logarithms of car and train accessibility, or log TTWA employment. All estimations are base on panel data for 1998-2007, and include year effects. Errors are clustered at the TTWA level. ***, **, * denote significance at the 1%, 5% and 10% levels respectively.

As we explained above, the problem with using these figures (and the others available in existing studies on agglomeration effects in Manchester and Leeds) is that our results so far do not distinguish between two different explanations of the positive correlation between access to economic mass and wage. To recap, what we have called a *people*-based or composition-based explanation relies on individuals who would be better paid *where-ever they live* choosing to work in places with higher access to economic mass. For example, high ability people may live disproportionately in larger cities. The alternative, *place*-based, explanation would attribute the effect directly to the place so that otherwise identical individuals earn higher wages in places with greater access to economic mass.

To separate out these effects, we need to control for the fact that individual characteristics that affect wages may be correlated with access to economic mass. To do this, we include these individual characteristics in our wage regressions to give:

$$\ln (w_{it}) = \alpha_t + \beta X_{it} + \theta \ln (A_{it}) + \varepsilon_{it}$$

where X_{it} are individual characteristics, beta is a parameter to be estimated and all other notation is as before. Beta captures the effect of individual characteristics on wages leaving theta to capture the effect of access to economic mass after controlling for the composition effects that we described above.

We have a large number of individual level variables that could be included in the regression. In order to separate people/composition-based effects from place-based effects we want to control for predetermined productive characteristics of individuals that are correlated with the economic mass of the cities in which they choose to live. These characteristics – e.g. gender, age – arise independently of city size, but can become correlated with city size because individuals with different productivities sort into cities of different sizes. Clearly the sex of a worker is determined independently of access to economic mass even if males and females then choose to live in different places so that sex is correlated with access to economic mass. A similar argument applies to age.

However, there are some characteristics of individual that may at least partly be *determined* by economic mass. For example, people in different places may have access to different opportunities in terms of occupations. If good access to economic mass causes a person to choose a more productive (and higher paid) occupation (which is possible if agglomeration economies cause some occupations to be more prevalent in big cities) then we may want to attribute the resulting effect on wages to access to economic mass *not* to occupation. Controlling for occupation in our wage regressions will yield estimates of the effect of economic mass that net-out any effects arising from the occupational choice of individuals. Similarly, if big cities encourage development of human capital (education), controlling for individual education in our regressions nets-out any wage effects on individuals arising via the educational choices of individuals. Similar arguments apply to industry choices.

An additional challenge is that an association between composition and economic mass could arise because, historically, more productive workers tend to live closer together. It could also arise because better transport connections have evolved between labour markets with more productive workers. The reason we need be very cautious here is that the direction of causality may not run from economic mass to labour market composition, but in the opposite direction: Productive labour markets encourage better transport linkages. If this is the case then improving transport linkages will not be effective in changing the composition of the labour market or raising productivity. Therefore the estimates in Table 7 are upward biased and the economic benefits that they imply will never be fully realised by improving transport connections or otherwise increasing economic mass.

These issues complicate our analysis. On the one hand, we want to purge estimates of compositional biases arising from the sorting of more productive people into places of higher economic mass. For example, individuals who choose to work in finance may be highly paid in finance in whatever size city they worked. But finance jobs tend to be located in places of high economic mass, so there is a correlation between economic mass and individual wages arising through the sorting of finance workers into dense city centres. It would be wrong in this first case to attribute the *individual's* higher wage to the fact that they work in a dense place rather than that they work in finance. To avoid this type of bias, we need to control for the industry in which an individual works.

We also want to control for industry, occupation and other characteristics to try to purge our estimates from bias arising from the potential reverse causality running from productive labour markets to economic mass discussed above. For example, London's productive finance sector probably started life in response to London's position as a port and trade centre. London's economic mass and transport infrastructure grew as a consequence. Therefore, it would be wrong to attribute all London's productivity and its higher proportion of more educated workers to the existence of economic mass and transport infrastructure, since it is the productivity which has caused infrastructure and mass to grow.

On the other hand, a non-financial worker who lives in a big city with a large financial sector may be encouraged to work in finance because of the prevalence of well paid financial jobs. If we are interested in the overall effect of economic mass on individual wages, we would like to include this effect that works through industry choice. In this second case, we do not necessarily want to control for industry in our wage regressions, because controlling for industry will eliminate one of the channels through which economic mass acts on wages.

In short, in our wage regressions we want to control for individual variables that can be regarded as predetermined, in the sense that they are not determined by access to economic mass in the city in which a person currently lives and works. But there are some characteristics like occupation, education and industry which are partly predetermined, but partly determined by the economic mass of the city in which a person lives and works. If we control for all these factors, we control for all compositional effects arising both through sorting (which we want to eliminate), and through changes in individual characteristics induced by economic mass (which we do not necessarily wish to eliminate).

One modelling approach would be to control for all characteristics to obtain a fully specified model of wage determination, and then estimate separate models which deal with the mechanisms by which economic mass determined occupation, industry and education choice probabilities. This challenging undertaking is beyond the scope of this report.

The more feasible approach which we employ here is simply to estimate wage equations using various individual control variable sets, whilst recognising that controlling for variables that are partly determined by economic mass is likely to yield lower bounds to the overall impact of economic mass on individual wages, whereas failing to control for predetermined characteristics is likely to upward bias our estimates. In reality, of course, we do not usually know which characteristics are pre-determined and which are channels through which access to economic mass operates. We proceed by introducing variables starting with those that are most likely to be pre-determined and then adding in variables where we are less certain.

We start by introducing sex, age and age squared which, as argued above, are certainly predetermined. Results for the coefficient on the two access to economic mass measures are reported in Table 8, while we report the full results in Table 1 in Appendix 6. Column 1 just replicates the results from Table 7 where we only consider the effect of access to economic mass and do not control for any individual characteristics. Column 2 shows what happens when we control for sex, age and age squared (sometimes thought of as capturing the effect of experience). The coefficient on Car Accessibility drops while that on Train Accessibility increases. Taken literally this tells us that people who work in places with good Car Accessibility tend to be male and middle aged. We know this because the fall in the coefficient on Car Accessibility means that we must have included characteristics that must be positively correlated with both wage and Car Accessibility. Middle age and being male are the individual characteristics positively associated with wages (see Table 1 in Appendix 5) so it must be these characteristics that are associated with higher Car Accessibility. In contrast, the increase in the coefficient on Train Accessibility suggests the opposite. In fact, neither of the changes in coefficient are statistically significant so, although these patterns arguably make sense one should not read too much in to the changes.

Table 8: Regressions of wages on access to economic mass and other variables

	1	2	3	4	5	6	7
ln Car Accessibility	0.084	0.074	0.071	0.054	0.046	0.069***	0.070***
	(0.122)	(0.118)	(0.080)	(0.066)	(0.058)	(0.016)	(0.021)
ln Train Accessibility	0.258**	0.277***	0.173**	0.165***	0.170***	0.049***	0.030***
	(0.093)	(0.090)	(0.059)	(0.049)	(0.044)	(0.014)	(0.010)
R2	0.090	0.218	0.513	0.622	0.638	0.918	0.918
Observations	1102527	1091551	1091551	1091551	1090528	1090528	1090528

Notes: All models have log hourly earnings as a dependent variable and the explanatory variables of interest are logarithms of car and train accessibility variables. All estimations are based on panel data covering years 1998-2007, and include year effects. Column [1] has no controls; [2] adds age, age squared and gender; [3] adds years of education; [4] adds occupational characteristics (1-digit level) and dummies for part-time, public sector and collective wage agreement); [5] adds 1-digit industry controls; [6] adds individual fixed effects; [7] adds area level characteristics as described in the text. Standard errors are clustered at the travel-to-work area level. ***, **, * denote significance at the 1%, 5% and 10% levels respectively.

The next individual characteristic that we include is education. Although there is some evidence linking educational outcomes to access to economic mass the causal effect is not large (if the effect is causal at all). Given our rather aggregated skills classification we would argue it makes sense to see education as predetermined. However, as column 3 makes clear, sorting means that education is quite strongly correlated with access to economic mass, at least for Train Accessibility. Again, the interpretation of this is that higher educated workers get paid higher wages and tend to live in areas with higher access to economic mass by train. Once we control for this the association between wages and Train Accessibility is significantly weakened.

Next we introduce variables to control for occupation, whether the individual works in the public sector, works part time and is subject to a collective pay agreement. We could think of these characteristics as associated with either the individual or the job. If the latter, it is a little harder to be certain that these characteristics are predetermined. Fortunately this issue is moot as introducing these controls has little affect on the coefficients on the two access to economic mass variables (although the effect is slightly larger for Car Accessibility) as can be seen by looking at the results in column 4. A similar story applies with the introduction of industry controls as can be seen from results reported in Column 5.

To summarise the results so far, when we control for composition based on the observable characteristics of individuals (and jobs) the effect of access to economic mass is reduced by somewhere between a quarter and a third.

We can go one step further in our attempt to control for confounding factors that might be driving the relationship between access to economic mass and wages. So far, we have only controlled for the observable characteristics of individuals (things for which we have data in ASHE). Given that we observe individuals over time, however, we can use panel data techniques to control for unobservable characteristics of individuals that might be positively associated with both wages and access to economic mass. For example, as explained above, higher ability individuals will likely get higher wages and may also tend to live in larger cities. As ability is not recorded in our data, we would attribute the positive correlation between city size and wages to city size, when it was actually due to the sorting of higher ability individuals in to larger cities. To allow for this possibility, we estimate the model as above, but we now include individual fixed effects to control for time invariant unobservables. This implies that the effects of access to economics mass are estimated from individuals that move over time (otherwise, for individuals that do not move, we cannot be sure whether the higher wages are something to do with that individual or something to do with the place in which they live). The specification is thus:

$$\ln(w_{it}) = \alpha_t + \beta X_{it} + \theta \ln(A_{it}) + \lambda_i + \varepsilon_{it}$$

where everything is as defined before, except for the inclusion of individual unobserved fixed effects λ_i.

As can be seen from Column 6 in Table 8, the effect on the coefficients on the measures of access to economic mass are considerable. The change is most obvious for Train Accessibility where the coefficient is decreased by a factor of 3. In fact, the coefficient on Train Accessibility reduces sufficiently that the coefficient on Car Accessibility is now larger (although not significantly so). There is a further, more subtle, impact on the coefficient on Car Accessibility. Looking at the standard errors we see that the increase in the coefficient combined with a decrease in the standard error makes Car Accessibility significant for the first time. Economists would generally consider these results that control for both the observed and unobserved characteristics of individuals as providing the best estimate of the relationship between wages and access to economic mass controlling for composition. It is sensible to view these coefficients as capturing the upper bound of the likely effect on individuals who do not change sex, age, education etc as a result of increasing access to economic mass. In short, once we allow for composition, both Car and Train Accessibility are positively related to wages although the relationship is much weaker than suggested by our initial results that did not adjust for composition.

In the results reported so far, we only allow for place-based effects to be explained by access to economic mass. It is, of course, possible that other area characteristics that are positively correlated with both access to economic mass and wages might actually be the source of place-based effects. To consider this we calculate a number of additional area based characteristics that might potentially be associated with wages. Following Wheeler's (2008) work on wages growth for the US these include a measure of TTWA industrial diversity and occupational diversity to allow for the fact that diversity might be more important for wages than size per se. Industrial diversity of a TTWA j is calculated using a Herfindahl index: $\Sigma_j(E_{ijt}/E_{it})^2$ where j is two digit industry, i is TTWA and t is year. Occupational diversity is an analogous measure using 2-digit occupational code-level employment instead of SIC. Although we cannot include a measure of access to economic mass disaggregated by skills (see the discussion above) we can include the share of high and intermediate skills in TTWA j's working age population (with low skills the omitted category). Finally, we include measures of average TTWA industry shares (two digit) to see if the TTWA-wide industrial structure makes any difference.

Column 7 shows what happens to the coefficients on the two access variables when we include these additional area characteristics. The effect of Car Accessibility is essentially unchanged, while that of Train Accessibility falls somewhat further. Results reported in Appendix 6 shows that TTWA percentage high skills and the share of activity in Other Services are the only two area characteristics that have a significant effect on wages. These appear, perhaps unsurprisingly, to be positively correlated with Train Accessibility which leads to overall reduction in the coefficient on that measure of accessibility. Of course, these results may partly reflect the fact that large places attract lots of skilled workers. There is an active debate in the econometric literature about whether it is appropriate to control for the channels through which an explanatory variable might impact outcomes at the same time as trying to identify the overall effect of the explanatory variable (see Angrist and Pishke 2009). Without more evidence on the channels, and given that the coefficients on the access variables do not change too markedly, we prefer to use the results in Column 6 (ignoring other area characteristics) to assess the likely impact of the counterfactuals described below. We note that our results in column 7 provide some preliminary evidence that part of these effects may work through increasing the proportion of high skill.

Results reported in Table 9 show what happens when we exclude all individuals working in London. The overall story is much as before. Car Accessibility has an insignificant effect on wages, while Train Accessibility is significant (column 1). Adding in observable characteristics of individuals leaves the coefficient on Car Accessibility essentially unchanged while decreasing that on Train Accessibility (columns 2-5). As before, introducing individual fixed effects to control for unobservable characteristics more than halves the effect of Train Accessibility, while making the effect of Car Accessibility positive and significant (column 6). Additional area controls make Car Accessibility insignificant and slightly reduce the effect of Train Accessibility. As before, our preferred specification is column 6 which shows that excluding London does not make that much difference to the Train Accessibility coefficients that are the main focus of our counterfactual analysis below.

Table 9: Regressions of wages on access to economic mass and other variables excluding London

	1	2	3	4	5	6	7
ln Car Accessibility	-0.040	-0.046	-0.011	-0.014	-0.015	0.017*	0.007
	(0.035)	(0.035)	(0.021)	(0.017)	(0.016)	(0.007)	(0.006)
ln Train Accessibility	0.217***	0.229***	0.131***	0.121***	0.117***	0.055***	0.044***
	(0.036)	(0.037)	(0.023)	(0.019)	(0.018)	(0.007)	(0.006)
R2	0.055	0.190	0.505	0.616	0.630	0.917	0.917
Observations	884953	876198	876198	876198	875416	875416	875416

Notes: Regressions as in Table 8. but excluding observations for London TTWA.

Finally, we can consider whether these effects differ depending on the skill level of workers. We are limited in what we can do here, because our measures of economic mass are based on employment from the IDBR which only provides employment classified by industry (not skill or occupation). This prevents us from recalculating our measures of access to economic mass based on employment split by skill. Instead, we simply run our regressions separately for each skill group continuing to use access to economic mass based on overall employment. Table 10 shows the results for the three skill groups for our preferred specification (including individual fixed effects) and when we introduce area characteristics. It is interesting to note that the lower skilled (group 1) benefit from Car Accessibility, but not Train Accessibility. The highest skilled (group 4) benefit from both, although the effect of Car Accessibility is still slightly larger. The proportion of high skilled in the area has such a strong impact on the highest skilled that it essentially explains all of the effect of increased access to economic mass so that both access variables are insignificant once we introduce additional area controls. We use the average effects in what follows, ignoring the individual channels through which the effects might operate and ignoring the fact that the differences might differ somewhat across individuals. The results in Table10 do suggest, however, that the effects of improving both Car and Train Accessibility may actually be slightly stronger for those with intermediate level skills while higher skill gain less and lower skill may not benefit from increased Train Accessibility at all.

Table 10: Regressions of wages on access to economic mass and other control variables split by skill group.

	Skill group 1		Skill groups 2-3		Skill group 4	
	FE	FE+Area	FE	FE+Area	FE	FE+Area
ln Car Accessibility	0.054***	0.049***	0.074***	0.076***	0.049***	0.036
	(0.011)	(0.010)	(0.017)	(0.022)	(0.010)	(0.018)
ln Train Accessibility	0.003	-0.015	0.054***	0.036***	0.019*	-0.003
	(0.015)	(0.017)	(0.016)	(0.011)	(0.009)	(0.012)
R2	0.82563961	0.82587542	0.89440883	0.89447582	0.8629299	0.86306324
Observations	46057	46057	894873	894873	149598	149598

Notes: Regressions dividing by skill groups. FE reports coefficients from a regression including a full set of individual controls and is equivalent to specification [6] in Table 8. FE+Area adds a full set of area characteristics and is equivalent to specification [7] in Table 8. ***, **, * denote significance at the 1%, 5% and 10% levels respectively.

5.3 The labour market impacts of closer integration

We can now use our results to assess the labour market impact of improving access to economic mass. To do this we construct a number of counterfactual access to economic mass measures based on several different scenarios: (1) a 40 minute reduction in train travel time between Leeds and London; (2) a 40 minute reduction in train travel time between Leeds and London; (3) a 20 minute reduction in train travel time between Manchester and Leeds; (4) a 1% reduction in train travel times *within* Leeds and *within* Manchester; (5) a 1% reduction in train travel times *between* Manchester and Leeds; (6) and (7) as for (4) and (5) but for driving times; (8) and (9) a 10% increase in employment in all LAs in Manchester and Leeds (with the effects on Car Accessibility and Train Accessibility reported separately). In the first three scenarios we also allow for the second round (or knock on) effects on journeys between LAs not directly affected (e.g. Liverpool to Hull) that may see improved journey times as a result of the improved network. For the final four transport scenarios we just assess the impact of the first round effects. We do not view any of this as a serious transport modeling exercise, but instead use it to help quantify the effects and identify key messages emerging from our analysis. Appendix 2 gives more details on the construction of counterfactuals, while Table 11 reports the percentage changes in access to economic mass for each of the counterfactuals. To translate these percentage changes in access to economic mass in to changes in wages we multiply by the relevant coefficients on Car and Train Accessibility that we have estimated and reported in Table 8. Because the change in economic mass for each LA will be multiplied by the same coefficient, the impact on wages is proportional to the change in access to economic mass. This means that the figures in Table 11 allow us to assess the relative magnitudes and the distribution of the impacts of different types of changes.

Starting with the first two columns we see, unsurprisingly, that the impact of a Leeds-London reduction in train times is felt disproportionately in Leeds. Manchester LAs do benefit somewhat because the journey time matrix that we have at our disposal suggests that for all Manchester LAs some journeys to LAs elsewhere in Great Britain become faster if they use the new faster Leeds-London segment (see the section on counterfactuals in the appendix for more discussion of this). Similarly all Leeds LAs benefit from improved Manchester-London journey times but the effects are relatively small compared to the benefits to Manchester LAs. In line with existing work for The Northern Way, the proportionate benefits are, overall, larger for Leeds (in column 1) than for Manchester (in column 2) because external accessibility is more important for the smaller economy. The differential impact of a 20 minute Manchester-Leeds travel time reduction is consistent with this overall pattern, although it is clear that the most directly affected journeys benefit most (e.g. Manchester and Leeds). The impact on accessibility of within city GTC reductions are always larger than for between city GTC reductions. This is true for both Train and Car Accessibility. It is beyond the scope of this report to assess whether between or within GTC reductions are equally achievable but we provide the coefficients so that they can be used in future work to assess the likely impact if such figures become available for specific schemes. Finally, we see that 10% increases in population for all LAs have a fairly uniform effect on accessibility with, unsurprisingly, central LAs tending to see larger increases in access to economic mass in response to these changes.

Table 11: Percentage change in access to economic mass for different counterfactuals

LAD NAME	CR	L-Lon -40m	M-Lon -40m	L-M -20m	WT -1%	BT -1%	WC -1%	BC -1%	TP +10%	CP +10%
Bradford	L	2.58	0.78	6.59	0.23	0.09	0.24	0.11	3.18	3.44
Calderdale	L	2.72	1.21	6.05	0.17	0.11	0.21	0.14	2.80	3.42
Craven	L	2.57	0.75	6.30	0.12	0.08	0.14	0.12	2.03	2.58
Harrogate	L	2.17	0.85	6.98	0.14	0.08	0.17	0.09	2.18	2.58
Kirklees	L	2.18	1.26	6.00	0.17	0.12	0.20	0.13	2.92	3.30
Leeds	L	2.43	0.86	9.75	0.27	0.09	0.25	0.09	3.53	3.42
Selby	L	2.45	0.93	6.51	0.13	0.09	0.16	0.08	2.14	2.38
Wakefield	L	1.38	0.68	10.26	0.19	0.08	0.20	0.10	2.71	2.91
Bolton	M	0.72	2.59	6.17	0.21	0.07	0.28	0.08	2.79	3.50
Bury	M	0.73	2.45	6.24	0.20	0.08	0.28	0.09	2.76	3.59
Congleton	M	0.27	1.01	6.29	0.12	0.05	0.17	0.05	1.70	2.22
High Peak	M	0.67	1.70	5.22	0.13	0.07	0.19	0.08	2.03	2.71
Macclesfield	M	0.44	1.21	7.84	0.19	0.07	0.23	0.06	2.52	2.86
Manchester	M	0.69	2.36	10.07	0.27	0.07	0.32	0.07	3.45	3.90
Oldham	M	0.77	1.98	4.56	0.19	0.08	0.27	0.10	2.65	3.66
Rochdale	M	1.36	1.90	4.34	0.19	0.09	0.25	0.11	2.79	3.54
Salford	M	0.73	1.90	4.42	0.20	0.08	0.32	0.07	2.74	3.86
Stockport	M	0.74	1.38	7.62	0.22	0.07	0.29	0.07	2.90	3.59
Tameside	M	1.21	1.68	4.12	0.18	0.09	0.27	0.09	2.70	3.60
Trafford	M	0.71	1.58	6.40	0.22	0.07	0.31	0.06	2.87	3.71
Vale Royal	M	0.42	1.52	6.21	0.17	0.06	0.19	0.06	2.24	2.45
Warrington	M	0.38	1.35	6.86	0.19	0.06	0.25	0.06	2.51	3.05
Wigan	M	0.44	1.25	6.47	0.18	0.06	0.24	0.07	2.35	3.03

Notes: Table shows percentage change in access to economic mass from a 40 minute reduction in train journey time Leeds-London (L-Lon -40m); 40 minute reduction Manchester-London (M-Lon-40m); 20 minute reduction Leeds-Manchester (L-M-20m); Within and Between train GTC reductions of 1% (WT, BT -1%); Within and Between driving GTC reductions of 1% (WC, BC, -1%) and population plus 10% using train and car (TP, CP +10%). WC, BC and CP use LA-to-LA driving GTC, all others based on train GTC. CR indicates City-Region (L=Leeds, M=Manchester)

We now turn to the impact on wages obtained by multiplying the percentage changes in accessibility by the relevant coefficients on Car and Train Accessibility that we have estimated and reported in Table 8. Table 12 works through the example of a 20 minute reduction in train journey time between Manchester and Leeds (which happens to deliver the largest impact on wages of any of the exercises we consider). The column marked L-M -20m gives the percentage change in Train Accessibility and just replicates column 3 of Table 11. The first column reports the total effects of this change (including any compositional changes). These range from a 2.7% increase in wages in Wakefield to a 1.06% increase in Tamefield. Column 2 shows what happens as we control for age and sex. Consistent with the discussion above, the estimate of the percentage wage effect increases slightly because the coefficient on Train Accessibility is slightly higher (for convenience we repeat these coefficients in the last row of the table). Column 3 controls for education which leads to the first big reduction in the estimated size of the effect. Columns 4 and 5 show smaller changes as we first introduce occupation and then industrial controls. Finally column 6 shows the large reduction when we allow for unobservable individual characteristics. As a reminder column 6 is our preferred estimate of the impact of increased accessibility controlling for the effects of composition. We see the results range from a high of 0.50 of a percent for Wakefield to a low of 0.20 of a percent for Tameside. As is clear, compositional changes account for the vast majority of the estimated overall impact on wages.

We view this as the fundamental policy message to emerge from our work on labour markets: most of the overall agglomeration gains come from the changing composition of labour markets not from improved wages for those that do not change education, occupation, industry or ability in response to increased accessibility. As the composition of the Manchester-Leeds economies shifts towards higher educated, higher ability workers average wages will rise by somewhere between 1.06% (Tameside) and 2.65% (Wakefield). But the gains to existing workers who do not change their characteristics in response to increased integration are considerably smaller. We return to the implications of this in our conclusions.

Table 12: Percentage change in wages for a 20 minute reduction in Manchester-Leeds train time

LAD NAME	CR	L-M 20m	1	2	3	4	5	6
Bradford	L	6.59	1.70	1.83	1.14	1.09	1.12	0.32
Calderdale	L	6.05	1.56	1.68	1.05	1.00	1.03	0.30
Craven	L	6.3	1.63	1.75	1.09	1.04	1.07	0.31
Harrogate	L	6.98	1.80	1.93	1.21	1.15	1.19	0.34
Kirklees	L	6	1.55	1.66	1.04	0.99	1.02	0.29
Leeds	L	9.75	2.52	2.70	1.69	1.61	1.66	0.48
Selby	L	6.51	1.68	1.80	1.13	1.07	1.11	0.32
Wakefield	L	10.26	2.65	2.84	1.77	1.69	1.74	0.50
Bolton	M	6.17	1.59	1.71	1.07	1.02	1.05	0.30
Bury	M	6.24	1.61	1.73	1.08	1.03	1.06	0.31
Congleton	M	6.29	1.62	1.74	1.09	1.04	1.07	0.31
High Peak	M	5.22	1.35	1.45	0.90	0.86	0.89	0.26
Macclesfield	M	7.84	2.02	2.17	1.36	1.29	1.33	0.38
Manchester	M	10.07	2.60	2.79	1.74	1.66	1.71	0.49
Oldham	M	4.56	1.18	1.26	0.79	0.75	0.78	0.22
Rochdale	M	4.34	1.12	1.20	0.75	0.72	0.74	0.21
Salford	M	4.42	1.14	1.22	0.76	0.73	0.75	0.22
Stockport	M	7.62	1.97	2.11	1.32	1.26	1.30	0.37
Tameside	M	4.12	1.06	1.14	0.71	0.68	0.70	0.20
Trafford	M	6.4	1.65	1.77	1.11	1.06	1.09	0.31
Vale Royal	M	6.21	1.60	1.72	1.07	1.02	1.06	0.30
Warrington	M	6.86	1.77	1.90	1.19	1.13	1.17	0.34
Wigan	M	6.47	1.67	1.79	1.12	1.07	1.10	0.32
Multiply percentage change by			0.25800	0.27700	0.17300	0.16500	0.17000	0.04900

Notes: Table shows percentage change in accessibility for a 20 minute reduction in train journey times between Manchester and Leeds (L-M-20m). Column [1] shows total effects including any compositional changes; [2] controls for age, age squared and gender; [3] controls for years of education; [4] controls for occupational characteristics (1-digit level) and dummies for part-time, public sector and collective wage agreement); [5] controls for 1-digit industry; [6] controls for individual fixed effects. The final row corresponds to the coefficients in columns [1] to [6] reported in Table 8.

5.4 Results: wage growth

To reiterate, our results so far suggest that any significant impact on wage levels from greater integration of Manchester and Leeds labour markets come mostly from changing the composition of individuals and partly from changing the composition of work via effects on industrial structure and occupation. The effects on workers who are unable to change individual characteristics (education, ability) are quite small. In this sub-section we focus on the related question of whether access to economic mass plays a role in driving individual wage growth rather than levels. That is, we consider the possibility that increased access to economic mass is more important for understanding the dynamics of the labour market. That is, we address the possibility raised by some commentators that the problem for Manchester and Leeds is that there smaller access to economic mass means that labour markets are “thin” preventing workers from moving around between jobs as a way of achieving faster wage growth.

The sample of individuals that we use to study wage growth is essentially the same as the one that we use for the results reported above (we apply some additional trimming to eliminate very large growth rates). The dependent variable is annualised percentage wage growth over the period of observation of the individual: $\ln(w_T - w_{t0})/(T - t_0)$ where w_{t0} is the individuals' annual wage in the first year they are observed and w_T is the wage in the final year. The percentage wage growth is normalized by the number of years $T - t_0$ over which the individual is observed to allow for the fact that we observe different individuals for different lengths of time.

We work through exactly the same set of specifications as we did for wages. That is, we start by introducing controls for sex, age and age squared. We then control for education followed by "job" characteristics (occupation, part time, public sector, collective agreement) and industry. Finally, we control for other area characteristics. The access to economic mass and area variables are constructed as above. Note that, as we only have one observation of wage growth for each individual we cannot include individual fixed effects to control for unobserved ability (as we did in our preferred specification above). We discuss the likely implications of this further below.

Because we are looking at wage growth over a period of years we need to decide which variables we measure at the start of the period and which we allow to vary over time. Sex is obviously fixed and, without loss of generality, we can also measure age and experience at the start of the period (because age increases linearly across time). For the remaining individual and job characteristics we simply take the average over the period for which we observe the individual. We also time average accessibility and area characteristics for each individual (thus allowing for the fact that individuals may move across TTWAs over time).

We start, as with the level of wages, by regressing growth in wages on the access to economic mass variables separately and together. Results are reported in Table 13 where, once again, for comparison we also include the coefficients from a regression of wage growth on (log) TTWA employment. We can see straight away that the effects are an order of magnitude smaller than those on levels. This is reassuring as large differences in growth rates across places quickly translate into very large differences in the levels of wages across places (because of the "compound interest" nature of wage growth). The meaning of the coefficient of 0.067 on Train Accessibility is that a 10% improvement in train accessibility increases annual wage growth by roughly 0.7 percentage points. Similarly to the wage levels case, removing individuals with at least one year's work experience in London leaves the results unchanged (column 4).

Table 13: Regressions of wage growth on access to economic mass

	1	2	3	4	5
ln Car Accessibility	0.020**		-0.018**	-0.016**	
	(0.007)		(0.006)	(0.004)	
ln Train Accessibility		0.043**	0.067**	0.067**	
		(0.013)	(0.008)	(0.006)	
ln TTWA employment					0.010*
					(0.005)
Observations	248118	251915	248068	195212	252128
R-squared	0.00	0.00	0.00	0.00	0.00

Notes: All models have annualised percentage wage growth over the period of observation of the individual as dependent variables and the explanatory variables of interest are logarithms of car and train accessibility variables, or log TTWA employment. ***, **, * denote significance at the 1%, 5% and 10% levels respectively

We now start to introduce individual characteristics in the same order as for the wage regressions. As before, we report the coefficients on access to economic mass in the text and the full results in Table 2 of Appendix 5. Column 1 of Table 14 just repeats results when we enter the two access to economic mass variables with no controls. Adding sex, age and age squared (column 2) makes the negative effect on Car Accessibility insignificant and substantially reduces the coefficient on Train Accessibility. Adding education (column 3) has a similar effect. Adding occupational controls (1 digit occupation dummies plus part time, public sector and collective agreement) turns Car Accessibility positive and Train Accessibility negative (column 4). Once we include industry dummies (column 5) we are left with a very small effect of Car Accessibility on wage growth, but no effect from Train Accessibility. When we add in area controls such as high and intermediate skill shares, diversity measures and industry shares (column 6), even the effect of Car Accessibility disappears (mostly, as with wage levels because the share of high skilled workers in the TTWA is now significant). Overall, we do not find particularly strong evidence of an impact from access to economic mass on wage growth.

Table 14: Regressions of wage growth on access to economic mass and other variables

	1	2	3	4	5	6
ln Car Accessibility	-0.018**	-0.003	-0.003	0.006*	0.007**	0.005
	(0.006)	(0.005)	(0.006)	(0.003)	(0.003)	(0.004)
ln Train Accessibility	0.067**	0.0152**	0.010*	-0.012*	-0.011	-0.001
	(0.008)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Observations	248068	246125	246125	246125	246125	246125
R-squared	0.00	0.08	0.08	0.10	0.10	0.10

Notes: All models have annualised percentage wage growth over the period of observation of the individual as dependent variables and the explanatory variables of interest are logarithms of car and train accessibility variables. Column [1] has no controls; [2] adds age, age squared and gender; [3] adds years of education; [4] adds occupational characteristics (1-digit level) and dummies for part-time, public sector and collective wage agreement); [5] adds 1-digit industry controls; [6] adds area level characteristics as described in the text. ***, **, * denote significance at the 1%, 5% and 10% levels respectively.

Of course, this does not directly tackle the question of whether "thin" labour markets prevent workers from moving around between jobs as a way of achieving faster wage growth. To consider this question we decompose wage growth in to wage growth on the job and wage growth that occurs because workers are moving between jobs. We call these components "within" and "between" wage growth.

Table 15 shows what happens when we take each of these components and regress them on exactly the same explanatory variables as we did for overall wage growth (so columns 1 to 6 in Table 15 correspond exactly to columns 1 to 6 in Table 14).

Table 15: Regressions for "within" and "between" job wage growth

Within	1	2	3	4	5	6
ln Car Accessibility	-0.008*	-0.004	-0.004	0.001	0.003	0.005**
	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
ln Train Accessibility	0.027**	0.014**	0.013**	0.001	-0.004	-0.010**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Observations	248068	246125	246125	246125	246125	246125
R-squared	0.00	0.02	0.02	0.03	0.03	0.03
Between	1	2	3	4	5	6
ln Car Accessibility	-0.012	0.002	0.001	0.012**	0.011**	0.009*
	(0.006)	(0.005)	(0.006)	(0.003)	(0.003)	(0.004)
ln Train Accessibility	0.056**	0.007	0.003	-0.023**	-0.017**	-0.009
	(0.006)	(0.004)	(0.004)	(0.005)	(0.004)	(0.006)
Observations	197056	195866	195866	195866	195866	195866
R-squared	0.00	0.09	0.09	0.12	0.12	0.12

Notes: Dependent variables are within and between job wage growth as described in the text and the explanatory variables of interest are logarithms of car and train accessibility variables. Column [1] has no controls; [2] adds age, age squared and gender; [3] adds years of education; [4] adds occupational characteristics (1-digit level) and dummies for part-time, public sector and collective wage agreement); [5] adds 1-digit industry controls; [6] adds area level characteristics as described in the text. ***, **, * denote significance at the 1%, 5% and 10% levels respectively.

Starting with column 1 we see that the effect of Train Accessibility is about twice as large for between wage growth as for within wage growth if we ignore the role played by individual characteristics. As we add in more individual characteristics we see that the positive effect of Train Accessibility just captures the fact that it tends to be positively correlated with other characteristics of individuals associated with both higher within and between job wage growth. Once we allow for these factors we see that access to economic mass plays essentially no role in determining the within component of wage growth. There are contrasting effects on the between component. Overall, better Car Accessibility is associated with higher between wage growth, but train Accessibility is actually associated with lower between wage growth. Adding in area controls (column 6) does not change these conclusions substantially although the significance of the effect of Train Accessibility disappears. Overall, we do not find particularly strong evidence for the idea that larger labour markets have a strong effect on the amount of wage growth that occurs because of between job moves. The effects of larger labour markets on within job wage growth are statistically significant but very small in magnitude.

Table 16 shows that these conclusions are not substantially changed if we remove observations for the London TTWA. The table shows results from two specifications, one with no individual controls and one with the full set of controls for the overall, within and between specifications. The pattern of coefficients is essentially unchanged as can be seen by comparing results to those reported in columns 1 and 5 in Table 14 and Table 15.

Table 16: Wage growth regressions without London

	Overall	Overall	Within	Within	Between	Between
ln Car Accessibility	-0.016**	0.005*	-0.005**	0.004**	-0.009**	0.010**
	(0.004)	(0.003)	(0.002)	(0.001)	(0.003)	(0.003)
ln Train Accessibility	0.067**	-0.012*	0.031**	-0.001	0.051**	-0.022**
	(0.006)	(0.005)	(0.003)	(0.003)	(0.006)	(0.005)
Observations	195212	193611	195212	193611	152520	151546
R-squared	0.00	0.08	0.00	0.03	0.00	0.11

Notes: Dependent variables are overall, within and between job wage growth as described in the text and the explanatory variables of interest are logarithms of car and train accessibility variables. For each component of wage growth, the first column reports a specification with no individual controls, the second reports a specification with a full set of controls. ***, **, * denote significance at the 1%, 5% and 10% levels respectively.

Another possibility is that the effects might differ between high and low skilled individuals.

Table 17 suggests that this is indeed the case with both the overall and between effects tending to be driven by high skilled individuals. The table reports results for access to economic mass variables in specifications including a full set of individual level controls (so results should be compared to those in column 5 of Table 14 and Table 15). Again, however, we see that while Car Accessibility is positively associated with wage growth from job moves, Train Accessibility is negatively associated.

Table 17: Wage growth regressions for lowest and highest skilled

	Overall	Overall	Within	Within	Between	Between
	Low	High	Low	High	Low	High
ln Car Accessibility	-0.002	0.015**	0.008	0.001	0.003	0.009*
	(0.007)	(0.005)	(0.005)	(0.003)	(0.010)	(0.004)
ln Train Accessibility	-0.005	-0.030**	-0.022	0.001	-0.008	-0.028**
	(0.016)	(0.010)	(0.012)	(0.005)	(0.015)	(0.008)
Observations	14717	32096	14717	32096	9224	24675
R-squared	0.04	0.03	0.03	0.02	0.05	0.04

Notes: Dependent variables are overall, within and between job wage growth as described in the text and the explanatory variables of interest are logarithms of car and train accessibility variables. All specifications include a full set of controls. For each component of wage growth, the first column reports results when restricting the sample to low skilled workers, the second when restricting the sample to high skilled workers. ***, **, * denote significance at the 1%, 5% and 10% levels respectively.

In short, overall, we do not find particularly strong evidence for the idea that larger labour markets have a strong affect on overall wage growth or on the amount of wage growth that occurs because of between job moves.

5.5 Labour markets and agglomeration: conclusions

- Closer integration between Manchester and Leeds may deliver additional benefits in terms of increased wages for workers. Our largest estimate (for a 20 minute reduction in train journey times between Manchester and Leeds) has wages increasing by between 1.06% and 2.7%. However, nearly all of these wage effects come through the changing the *composition* of the workforce (arising through sorting, and/or because people change their characteristics in response to changes in economic mass). The effects for any given individual who does not increased their education or skill levels (the place-based effects) are small at somewhere between 0.20 and 0.50 of a percent.

- Consistent with this, individual wage growth is faster in places with better access to economic mass, but this effect appears to be driven by the fact that these cities tend to have more educated workers. Once we control for this there is essentially no relationship between city size and wage growth.

- We do not find particularly strong evidence for the idea that larger labour markets have a strong affect on overall wage growth or on the amount of wage growth that occurs because of between job moves.

- Overall, we fine that the aggregate effects of closer integration may be larger than the individual benefits. This relies on structural changes moving the composition of the Leeds and Manchester workforces towards higher skilled jobs. From a traditional cost-benefit perspective, these effects would *not* be counted as additional for individual investment projects if, as is likely, they come about because of greater attraction or retention of existing skilled workers. If they occur because existing workers increase their education or skills in response to changing economic opportunities some part of these higher gains may be additional (to the extent that the individual benefits of increasing, say, education, outweigh the costs).

- Regardless of the mechanism, if increased integration does lead to structural change these compositional changes will increase aggregate output in Manchester and Leeds, and this will be of interest to policy-makers interested in the performance of these places and of the wider North.

6. A structural model to examine the impacts of policy

The commuting and spatial econometric work we reported in Sections 3 and 4. help us understand the nature of current linkages between Manchester and Leeds and what might explain these. That does not help us answer the "what if" question of the impact of various possible policy interventions. Section 5 discussed one approach to this question based on (what economists call) reduced form models of the impact of increased access to economic mass on labour market outcomes. In this section we use a different approach based on a simple structural model. We have developed this model to focus more explicitly on the way that changes in Manchester and Leeds might affect other areas in the rest of the North (and more widely). The research in the previous section made assumptions about the distance decay effects of agglomeration and estimated the strength of agglomeration economies. This allowed us to focus on the extent to which the resulting changes in wages were driven by compositional changes versus size effects on productivity (which should underlie the "pure size" effects on wages). In this section we are interested in understanding how these productivity effects might spill out over space. To examine this we use a structural model where we indirectly assume the size of agglomeration economies (by using existing parameter estimates of the firm level spread of productivity) but where we estimate the strength of linkages across space. This allows us to ignore questions about composition and instead to focus clearly on the spatial distribution of the impacts of counterfactual changes similar to those discussed in Section 5.

This approach delivers two key policy messages. First, it reminds us the more conservative estimates of the impact on wages produced in section 5 should be seen as an upper bound for the *additional* benefits of increased integration net of compositional changes. The structural model we present here carefully considers one of the mechanisms that could lead to such effects and finds that they are smaller than the effects that we identified in section 5. Second, the structural model clearly highlights the fact that the spatial distribution of changes in response to the counterfactuals is complicated. This is an important finding, because the selection mechanism that we study is one of a very limited number of situations where we are able to articulate a model which allows us to capture the spatial distribution of changes. While it remains popular in policy circles to expand on the likely impacts of policy changes in one place on outcomes in other places our understanding and modeling of these impacts remains in its infancy.

The model that we use to do this draws on insights from the heterogenous firm trade literature which considers the impact of trade integration on aggregate productivity. Before outlining the model, it is useful to briefly consider the pros and cons relative to a fully specified regional model (e.g. a computable general equilibrium model). A fully calibrated model would have the advantage of being 'more general' in the sense that it would account for a larger number of real world features (like the housing market, capital accumulation, input-output relationships across industries) that are not really dealt with in the model we have developed. However, current CGE models rest on the strong assumption that productivity and/or technological progress are exogenous. A point which is related to this assumption is that these models usually provide implausibly low figures on the effect of regional policies. The urban literature clearly points out that productivity and innovation are endogenous, that they are related to the spatial distribution of firms and workers, and that there are large differences across space. The model we will use provides a micro-founded mechanism for these differences (the selection of the most competitive firms) that is consistent with empirical evidence. As long as the primary concerns of a regional policy are competitiveness, productivity and wages, then the model we use provides a robust and simple tool to explore more accurately these issues. We now explain the model and mechanisms at work before explaining how we estimate and use it. The model is explained in detail in Behrens et al (2008) and in an appendix to the this report available from The Northern Way website. In the text, we focus on explaining the economic mechanisms and how the model is used.

6.1 An Introduction to Heterogeneous Firm Models

Recent models featuring heterogeneous firms have pointed out that trade integration across regions, at both international and intra-national level, has a positive impact on aggregate productivity through the selection of the best firms (Bernard et al., 2003; Melitz, 2003; Melitz and Ottaviano, 2008). The reason is a combination of increasing competition from importing firms (that forces small and unproductive indigenous firms to shut down) and increased access to export markets (that induces a reallocation of economic resources to those large and productive firms that export the most). There is a clear parallel here with the economic forces that should affect firms if we use transport policy to achieve greater integration between Manchester and Leeds. As transport policy is one of the few direct policy levers that would facilitate greater interaction we have focused our modelling efforts on these economic mechanisms.

The detailed impacts on firms of increasing integration work through as follows. The model features spatial competition among firms with heterogeneous productivities and endogenous wages. The productivity of a region is endogenous in the model due to a Darwinian mechanism of selection of the best firms. Trade costs (broadly defined as all impediments in doing business in different locations) “protect” some local unproductive firms from the competition of firms located in other regions. At the same time, such costs limit the potential of local productive firms that cannot expand their production due to the difficulty of reaching consumers in other regions. Transportation policies that successfully reduce trade costs will induce both the exit of low productive firms and a reallocation of market shares towards the most productive and competitive firms. This will in turn increase the aggregate productivity of the integrating regions. At the same time, pressure on the labor market due to increased aggregate production will push wages, and so production costs, up especially in the regions experiencing the highest productivity gains. Endogenous wages thus “mitigate” differences in competitiveness by counterbalancing productivity changes in such a way that, in a long term situation where trade across regions is approximately balanced, average production costs do not display too great an imbalance.

For international trade, this mechanism finds strong empirical support in firm-level analyses that have tried to identify the direction of causation hidden in the positive correlation between the export status of a firm and its productivity (called ‘exceptional exporter performance’ by Bernard and Jensen, 1999). This is a crucial issue for trade policy. Causation going from export status to firm performance would reveal the existence of ‘learning by exporting’ and therefore call for export promotion. However, apart from peculiar cases, most of the evidence supports reverse causation in the form of ‘selection into export status’: firms that already perform better have a stronger propensity to export than other firms (Tybout, 2002). Once again there is a clear parallel with integration at the sub-national level. Often policy markers place emphasis on encouraging supplier-customer links across space because they are positively correlated with productivity. But in reality we do not know whether this correlation reflects productivity increasing as a result of supplier-customer links or instead, whether the most productive firms are those that end up supplying customers in the other region. In the absence of convincing evidence at the sub-regional level, our model will focus on the mechanism (selection rather than learning) which receives most support from the international trade literature.

The model usually predicts that trade integration is particularly beneficial to small regions as they gain access to large markets and can disproportionately benefit from increased production and competition. However, this result rests on the hypothesis that consumers and workers do not change their location. Unfortunately, we cannot yet endogenise consumer and worker location responses, but we can consider policies that attract workers from other locations by treating the changes as exogenous shifts in population. The effect of such shifts is certainly beneficial in terms of productivity and wages in the model. By enlarging local demand, an inflow of workers/consumers will promote the creation of new local firms thus increasing both competition and consumption variety within the region.

6.2 Data

We estimate the model using the same data and spatial units as in the section on spatial econometrics. The reader is referred to that section for an explanation of both the data and geography. We have used additional data on the Generalised Transport Costs between areas (as described in the section on commuting). We provide figures for changes in both output per worker and average wage induced by the counterfactuals discussed below. It is important to bear in mind that the wage changes we provide are nominal wage changes with respect to a numeraire region (we choose Aberdeen). The underlying theoretical model is in fact invariant to a change in the unit of measurement for wages and a numeraire is needed. By contrast, given the underlying model, absolute real wage changes in a region, which are probably the ones that are more interesting for policy analysis, are equivalent to changes in output per worker.

6.3 Counterfactuals

We report results from four illustrative counterfactuals to demonstrate the way the model works and to provide evidence on the implications of different policy interventions. The counterfactuals are as follow:

1) Counterfactual 1: A decrease of 20 minutes in the rail journey time between Manchester and Leeds that takes into account network effects on other regions.
2) Counterfactual 2: A decrease of 40 minutes in the rail journey time between Leeds and London that takes into account network effects on other regions.
3) Counterfactual 3: A decrease of 40 minutes in the rail journey time between Manchester and London that takes into account network effects on other regions.
4) Counterfactual 4: An improvement in the internal transportation network of both Manchester and Leeds equivalent to an X% decrease of the generalized transport costs within the two regions. Two scenarios will be considered: A) A reduction of 1%; B) A reduction of 5%.
5) Counterfactual 5: An increase in the housing stock of both Manchester and Leeds that is able to attract families and workers from other regions in such a way that the population of both cities increases by 10%. Two scenarios will be considered: A) Migrants coming from the North region only (defined as North-East, plus North West, plus Yorkshire and the Humber); B) Migrants coming from all over Great Britain.

6.3.1 The impact of transport policies

Figure 14 shows the impact on GDP per worker of counterfactual 1. Both Manchester and Leeds gain from the reduction in travel time between the two cities (leading to a 0.101 % increase in output per worker for Leeds and a 0.0390% increase for Manchester). Wages increase by 0.020% in Leeds and 0.016% in Manchester. These look like small numbers, although one has to keep in mind that they represent a permanent increase in productivity that will be experienced for many years to come. For example, considering total 2006 GDP in the two regions and a discount rate of 3%, the policy would be worth £2.7 billion (assuming benefits persist indefinitely). In terms of the spatial distribution, as highlighted by Figure 14, many other areas in the North would gain from speeding up the connection between Manchester and Leeds. Gains follow a clear geographical pattern around the Pennines with some areas (Doncaster, North Lincolnshire, Stoke-on-Trent, York, Stafford, Crewe, Nantwich, and West Lindsey) actually gaining more than Leeds in percentage terms. For the entire North area the present discounted value of this policy equals £6.7 billion. Many regions in Scotland would also experience small gains due to increased accessibility while the South of Great Britain suffers small loses. These loses occur because falling transport costs increase the productivity of firms in Manchester, Leeds and other Northern locations as a result of stronger selection effects. In turn, firms in less affected regions find it more difficult to enter the Manchester and Leeds market reducing their sales and profitability. It is important to note that the network effects of the change are crucial for the spread of these benefits across the North. The smaller these network effects the more concentrated are the benefits on Manchester and Leeds. At the extreme, as we shall see later, changes that only benefit Manchester and Leeds tend to lead to loses elsewhere in the North.

Figure 14: Percentage Change in GDP per worker in response to a 20 minute reduction in train journey time between Manchester and Leeds

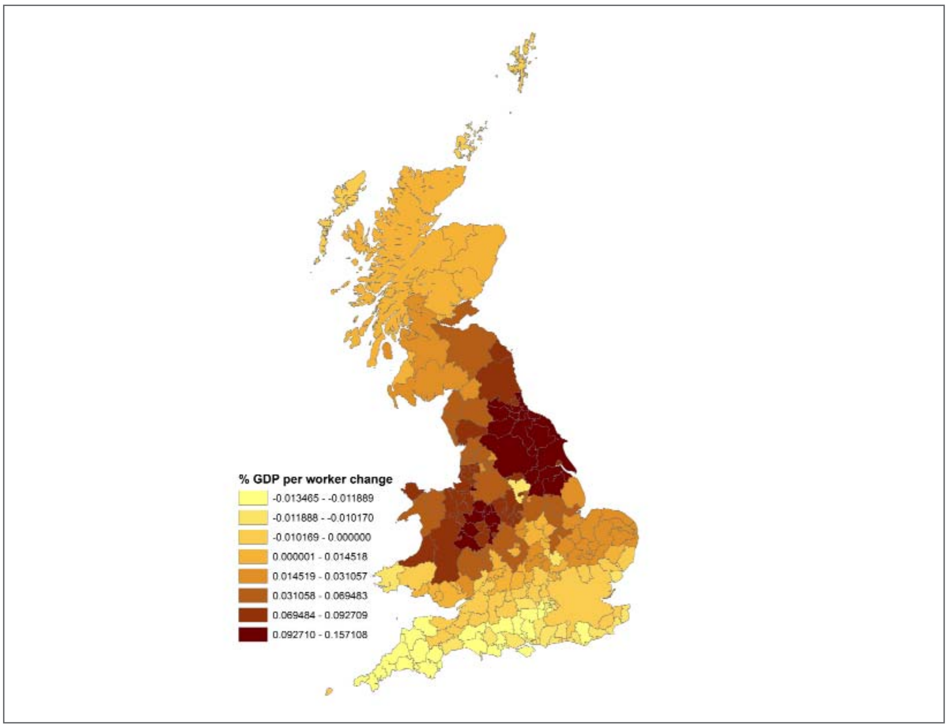
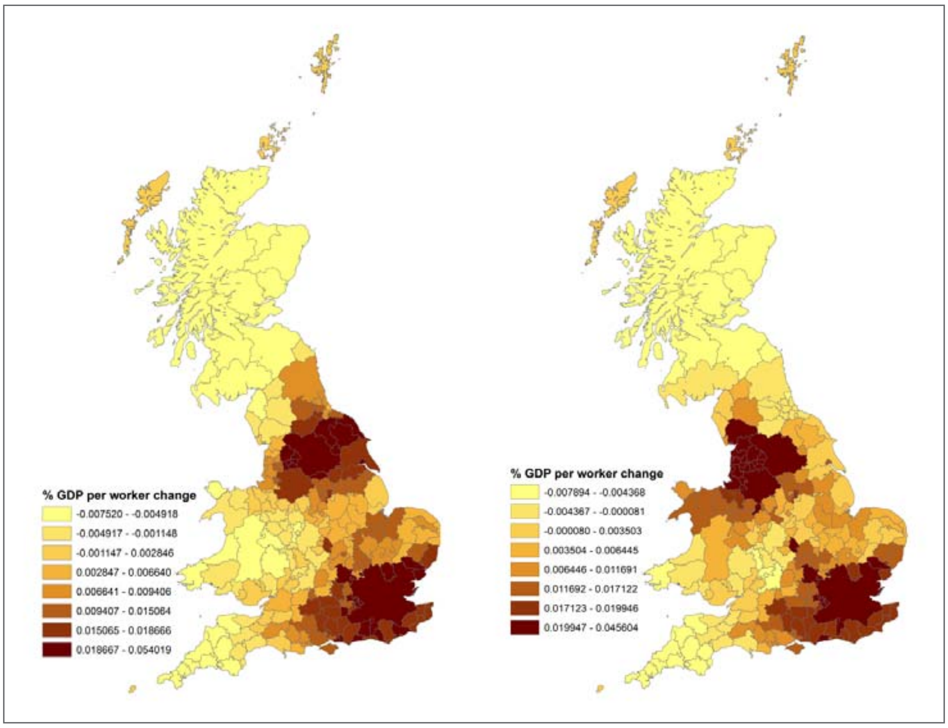


Figure 15 shows the impact on GDP per worker of counterfactuals 2 and 3 involving faster rail links between Leeds-London and Manchester-London respectively. Notice the significant differences in the overall pattern of benefits. In counterfactual 2 (left panel), decreasing train journey times between Leeds and London sees output per worker gains of 0.054% in Leeds and 0.020% in London. Manchester, which benefits indirectly from a better accessibility to London, gains 0.017%. Wage increases would be 0.012% for Leeds, and 0.005% for London and Manchester. The expected return of this policy for Leeds and London with a discount rate of 3% is £3.4 billion. If we consider both the North and London then the present discounted value of the policy would be £4.4 billion. Note that, in this scenario, most of Scotland loses as a result of increased local competition in the North and London that makes it more difficult for Scottish firms to profitably sell there. But the bigger percentage loses are concentrated in the areas sandwiched between the Northern and Southern gainers.

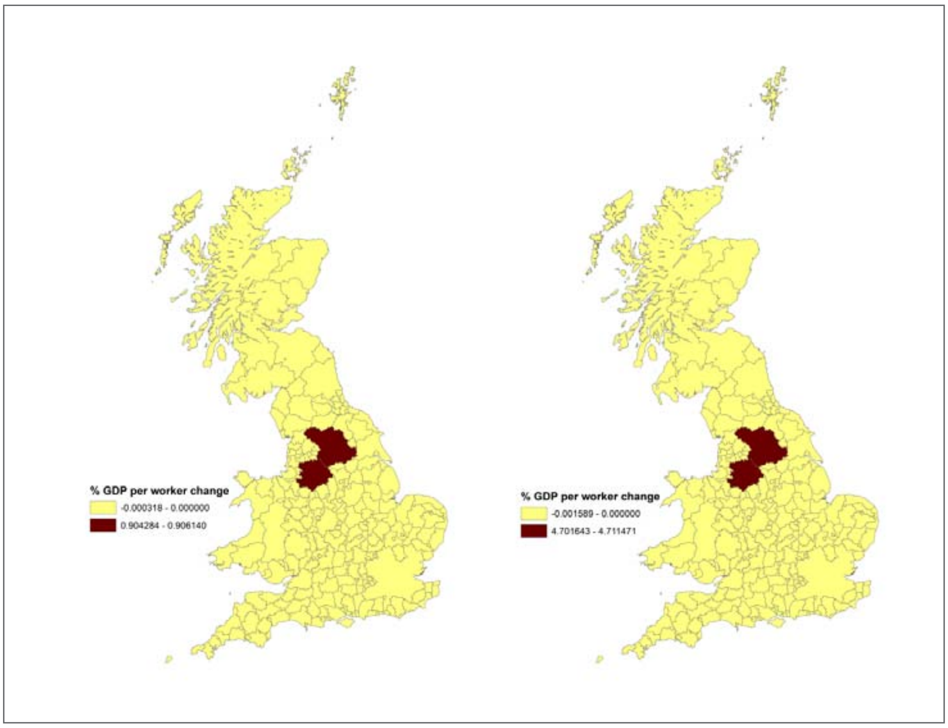
Figure 15: Percentage Change in GDP per worker in response to a 40 minute reduction in train journey time between Leeds and London (left) and Manchester and London (right)



The right panel in Figure 15 shows the pattern of changes if the link is between Manchester and London rather than Leeds and London. The overall picture of gains and loses is similar to that mapped in the left panel except that now Manchester (and its near neighbours) are those enjoying the highest gains due to increased accessibility. Manchester now sees a 0.043% gain in output per worker, London gains 0.021% while Leeds, which benefits indirectly from better accessibility to London, gains 0.023%. Wages would rise by 0.006% in Leeds and London, and by 0.010% in Manchester. The expected return of this policy for Manchester and London is £3.6 billion. If we consider both the North and London then the present discounted value of the policy would be £4.6 billions. A similar set of regions lose for this Manchester-London improvement as do from the Leeds-London improvement discussed above. It is interesting to note that, consistent with our findings in section 5, the percentage gains of increased access to London are higher for the smaller (Leeds) than for the larger economy.

Figure 16 displays the outcome of counterfactuals 4A and 4B. They have an identical spatial pattern and the percentage change in output per worker stemming from a 1% and 5% reduction of internal GTC costs in Manchester and Leeds are essentially proportional to GTC changes. For the case of a 5% internal GTC costs reduction, Leeds would experience an increase of 4.70% in output per worker while Manchester would gain 4.71%. Other regions experience negligible losses. Wages only increase by 0.006% in Leeds and 0.008% in Manchester because the increase in productivity would almost entirely be translated in to a reduction of selling prices. Compared to previous scenarios, an internal GTC reduction works as a simple cost cut for firms without any gain in accessibility to other markets. This cost reduction is entirely passed through to lower consumer prices affecting nominal wages only slightly. By contrast real wages in Manchester and Leeds will, as explained above, increase by the same magnitude as output per worker. The present discounted value of this policy for Manchester and Leeds is £185.4 billion reminding us that the beneficial price effects may far outweigh any nominal wage effects.

Figure 16: Percentage Change in GDP per worker in response to a 1% (left) and 5% (right) GTC reduction within Leeds and Manchester



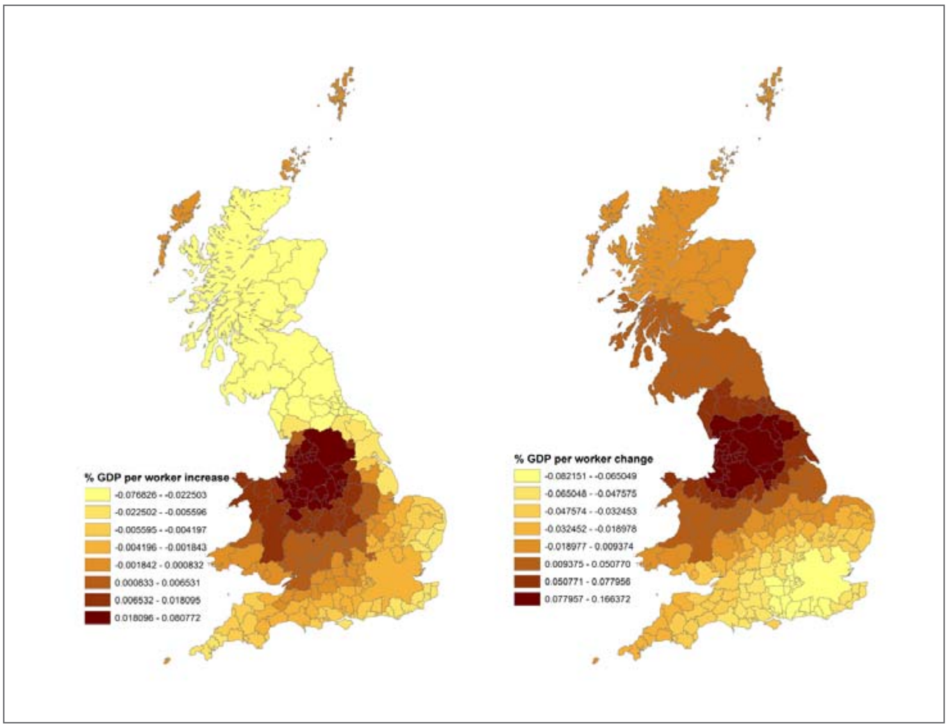
Putting aside the relative costs of these two policies, it is clear that policies targeting within city transport costs deliver larger benefits for the cities themselves but have relatively little impact on surrounding regions. The key insight to understand these results is that the two types of policies affect different numbers and types of firms. A between city transport improvement makes it easier for all firms to better access new markets. However, as some trade impediments will still remain in place, only the few firms that sell in other markets (the most productive ones) will actually take advantage of such a policy with the vast majority of firms who still only sell locally suffering from tougher external competition. There would certainly be an increase in the profitability of the whole population of firms, which would then materialize in to increased productivity, but the impact of such a positive change is rather low and has a strong spatial pattern. By contrast, a policy targeting within city transport costs would be beneficial for the whole population of domestic firms because each and every firm will be able to reach a large part of of its total demand (i.e. local demand) at a lower total cost (production + delivery). On the other hand, other regions are only slightly affected because, from a production costs point of view, firms in the region where the policy takes place have not become any “better” and between-city transportation costs have not changed.

The impact of population changes
Figure 17 displays gains in output per worker resulting from counterfactuals 5A and 5B. Starting with the left panel one can see that Manchester and Leeds would both gain by expanding their population. Some areas in the south of Manchester and Leeds (especially in the Midlands) gain from this southern-shift of the population but the rest of the Northern region lose. In this case, as before, the larger market leads to a stronger selection effect and more competitive firms in Manchester in Leeds. This is bad news for firms in areas that have lost population to Manchester and Leeds who get hit by the double whammy of lower own market size as well as stronger competition from firms in Manchester and Leeds. For areas south of Manchester and Leeds the stronger competition effect is actually offset by the gains that come from improved access to the newly expanded Manchester-Leeds market. These areas gain overall. Scotland is quite negatively affected because of the loss of market in areas to the south of it, while Wales and the South experience only marginal losses. Leeds would see its ouptut per worker increase by 0.062% while Manchester would gain 0.081%. The expected return of the policy for Manchester and Leeds in terms of productivity gains is around £2.9 billion. Taking into account gains and losses of other areas, such a change in the spatial distribution of population would deliver a present discounted value of £2.3 billion for the North as a whole. Clearly, these figures do not consider the overall increase in *total* output in Manchester and Leeds (that would be roughly 10%) due to the influx of new labor force. As for wages, the model predicts an increase of 0.018% for Leeds and 0.022% for Manchester.

Turning to the right hand side of Figure 17 one notices that the spatial distribution of gains is shifted North due to the fact that most new workers would come from the South and the Wales. Indeed, as we have assumed that all regions would loose the same percentage share of population in favor of Manchester and Leeds, the highly populated regions in the South will be those experiencing the largest drops in local market size (as they start off with higher populations). On the other hand, while Scotland is now losing some population, it can gain from a closer proximity to the large consumer mass that has shifted northwards into Manchester and Leeds and is ultimately better off under this second scenario. Leeds would see its output per worker increase by 0.16% while Manchester would gain 0.17%. The expected return of the policy for Manchester and Leeds in terms of productivity gains is estimated to be around £6.4 billion. Taking into account gains and losses of other regions, such a change in the spatial distribution of population would deliver a present discounted value of £10.9 billion for the North as a whole. As for wages, the model predicts an increase of 0.031% for Leeds and 0.032% for Manchester.

It is important to note that when population comes from the rest of the North, these changes in wages will be swamped by the gains that come from the fact that Manchester and Leeds are more productive (and pay higher wages) than the regions from which population is being drawn. When population comes from across Great Britain this effect is reversed and the gains are substantially offset by the fact that population is moving out of the higher productivity south towards Manchester and Leeds.

Figure 17: Percentage Change in GDP per worker in response to a 10% increase in population in Leeds and Manchester with a corresponding reduction in northern regions (left) and the whole GB (right)



6.4 Conclusions – structural model

- Both intra and inter city-region transport schemes will deliver productivity benefits as a result of the selection effects generated by greater competition. These effects are positive, and occur in addition to the user benefits identified in a traditional cost benefit analysis.

- Inter-city schemes favour Leeds, while intra-city schemes favour Manchester. Better connecting Manchester and Leeds to London delivers larger gains than linking Manchester-Leeds. Given total 2006 GDP in the two city-regions and a discount rate of 3%, a 20 minute reduction in train journey times between Manchester and Leeds would be worth £2.7 billion (assuming benefits persist indefinitely). Reducing train journey times between Leeds-London by 40 minutes is worth £3.4 billion to those two cities, while the same time reduction between Manchester and London is worth £3.6 billion to those two cities. For both the improvements involving London, percentage increases are greater in the Northern cities but the larger size of the London economy means that it accounts for a larger share of the total gains[15].

15. This is in line with evidence presented in Steer-Davies-Gleave North-South Connections report for The Northern Way.

- Increasing Manchester and Leeds population leads to small wage and productivity gains but quite large total GDP gains. However, if this population increase came by drawing in workers from the rest of the North the aggregate gains coming from increased productivity are swamped by the gains coming from moving workers from lower productivity regions to the higher productivity regions of Manchester and Leeds. If population moves from all over Great Britain the fact that some higher productivity places (in the South) are losing population offsets this effect.

- The broad spatial distribution of gains and losses is usually quite intuitive although the details can be quite complicated. The spatial patterns vary markedly by counterfactual.

7. Overall conclusions

- Against the benchmark of other comparable city pairs within Great Britain, we find evidence that commuting between Manchester and Leeds is around 40 per cent lower than expected, given the characteristics of the two cities and the 40 miles distance from centre to centre.

- This is explained partly by overall transport costs between the two cities and partly by their current industrial and occupational composition. This suggests that lowering these costs has an important role to play in increasing integration between the two city regions. This in turn may improve the economic performance of the two city-regions.

- Although we do not examine their role directly, the fact that economic factors explain these low commuting levels appears to leave little room for cultural or social factors to play a large part in explaining overall commuting patterns. This suggests that such factors are unlikely to act as a barrier to increased commuting between the two cities if transport investment lowers the overall costs of commuting, of if other economic factors lead to enhanced interactions.

- Differences in the correlation between the city-regions' growth and levels of earnings, employment and GDP (relative to GB benchmarks) are explained by patterns of industrial and skill structure.

- Overall, this suggests that structural change would be likely to play an integral part in increasing the extent of observed interaction between the two city-region economies.

- We draw two key policy messages from these findings. First, unusually poor links between the two city-regions are not a convincing explanation of current economic performance. Second, we find little evidence that interactions between the two city-regions depend on unobserved characteristics. This suggests that these factors should not be a barrier to increasing integration between the two-city regions if this was considered desirable.

- Closer integration between Manchester and Leeds (from a 20 minute reduction in journey time) could increase wages by 1.06%-2.7%. This impact is dependent on induced changes in the industrial structure, composition and skill levels of the population. It represents an upper bound of the possible effects as we cannot rule out the possibility that some of this effect runs from the composition of the labour market to lower transport costs (rather than vice versa). We find evidence that the effect on wages for individuals who do not change their personal or job characteristics are small (between 0.2% - 0.5%). This modest impact on the wages of workers whose characteristics remain unchanged is likely to be offset or even reversed by induced increases in the cost of living.

- This finding suggests that the effects on Manchester and Leeds will be bigger if policy interventions, such as improved transport links, induce structural change, particularly by changing the composition and skills of the workforce. In the analysis of specific transport projects, whether these wider economic impacts should be seen as additional to traditional user benefits depends crucially on the policy objective. From a national cost-benefit perspective, these effects would not be counted as additional if, as is likely, they come about because of attraction or retention of skilled workers at the expense of other places. In a policy context which aims to address the underperformance of the North, or address spatial disparities, these effects would be of more importance.

- Inter-city transport schemes appear to favour Leeds, while intra-city schemes favour Manchester. Better connecting Manchester and Leeds to London (through a 40 minute reduction in journey times) generates larger overall gains, with larger aggregate increases in GDP in London, but larger percentage increases in the North. Taken individually, links from Manchester and Leeds to London generate some wage reductions in parts of the east and west sides of the North respectively. In contrast, Leeds-Manchester links concentrate more of the benefit in the North, and generate a rather greater impact on the north-south economic differential, although with some negative impacts possibly experienced in more peripheral areas within the North.

8. References

Anderson, James E. and van Wincoop, Eric. “Trade costs”. Journal of Economic Literature”, Sep. 2004, 42 (3), pp. 691-751.

Angrist, J and S. Pischke (2009) Mostly Harmless Econometrics, Princeton University Press

Anselin, L. (1988). Spatial econometrics: methods and models. Dorddrecht: Kluwer Academic.

Anselin, L. (1995). Local Indicators of Spatial Association — LISA. *Geographical Analysis*, 27: 93–115.

Anselin, L (1998a). Interactive techniques and exploratory spatial data analysis. In: Longley PA, Goodchild MF, Maguire DJ,Wind DW(eds) Geographical information systems: principles, techniques, management and applications. Wiley, NewYork

Anselin, L (1998b). Exploratory spatial data analysis in a geocomputational environment. In: Longley PA, Brooks SM, McDonnell R, Macmillan B (eds) Geocomputation, a primer. Wiley, NewYork.

Aw B., S. Chung and M. Roberts (2000) Productivity and Turnover in the Export Market: Micro-level Evidence from the Republic of Korea and Taiwan (China), World Bank Economic Review 14, 65-90.

Behrens K., G. Mion, Y. Murata and J. Südekum (2008) Trade, wags and productivity, IZA DP.

Bernard A., J. Eaton, J. Jensen and S. Kortum (2003) Plants and Productivity in International Trade, American Economic Review 93, 1268-1290.

Bernard A. and B. Jensen (1999) Exceptional Exporter Performance: Cause, Effect, or Both?, Journal of International Economics 47, 1-25.

Bernard A, J. Jensen and P. Schott (2003) Falling Trade Costs, Heterogeneous Firms, and Industry Dynamics, NBER Working Paper No.9639.

Cervero R, Wu K-L, 1997, "Polycentrism, commuting, and residential location in the San Francisco Bay area" *Environment and Planning A* 29(5) 865 – 886

Clerides S., S. Lach and J. Tybout (1998) Is Learning by Exporting Important? Micro-dynamic Evidence from Colombia, Mexico, and Morocco, Quarterly Journal of Economics 113, 903-947.

Cliff A. D. & Ord, J.K. (1981). Spatial processes: models and applications. Pion, London

Coombes, M. and S. Raybould. 2001. Commuting in England and Wales: “People” and “place” factors. In: Pitfield, D., ed. *Transport Planning, Logistics and Spatial Mismatch: A Regional Science Perspective*, pp. 111-134. London: Pion Limited.

Del Gatto G., G. Mion, and G.I.P Ottaviano (2006) Trade Integration, Firm Selection and the Costa of non-Europe, CEPR Discussion paper 5730.

DETR (2001) Transport Economics Note, DETR, London.

Eaton J. and S. Kortum (2002) Technology, Geography, and Trade, Econometrica 70, 1741-1779.

Eurostat, Regional Statistics, REGIO database, EUROSTAT.

Head, K. and T. Mayer (2002) "Market Potential and the Location of Japanese Investment in the European Union," Centre for Economic Policy Research discussion paper no, 3455

Levinson, D. Accessibility and the journey to work, Journal of Transport Geography, Volume 6, Issue 1, March 1998, Pages 11-21, ISSN.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. Review of Economic Studies, 60: 531-42.

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

Melitz M. and G. Ottaviano (2008) Market size, trade and productivity, Review of Economic Studies 75, 295-316.

MIER (2008) The Case for Agglomeration Economies, Manchester Independent Economic Review.

Overman, Henry G., Redding, Stephen J. and Venables, Anthony J. “The economic geography of trade, production and income: A survey of empirics”. In Handbook of International Trade, edited by E. Kwan-Choi and J. Harrigan. Basil Blackwell, 2003.

Pavcnik N. (2002) Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants, Review of Economic Studies 69, 245-276.

Sen, A. and T.E. Smith (1995) Gravity Models of Spatial Interaction Behavior. NY: Springer.

Tybout J. (2002) Plant and Firm-Level Evidence on New Trade Theories. In: Harrigan J. (ed.) Handbook of International Economics, Vol. 38, Basil-Blackwell.

World Economic Forum (2005), Global Competitiveness Report 2005-2006, Palgrave Macmillan.

Appendices

1. The ASHE and NES databases

We have checked the ASHE and NES database for consistency. A few observations with inconsistencies such as miscodings in age or gender have been either corrected (e.g. by using the annual nature of the survey to correct age and by using modal gender to correct year-on-year changes in classification) or dropped. To reduce the impact of outliers, we drop 0.5% of observations from both the top and the bottom of the wage distribution for each year 1998-2007. If an individual has multiple jobs, only the main job is used in the analysis.

Section 5 uses the following SIC 1-digit coding:
1 = Agriculture, Fishing, Mining, Leather products
2 = Manuf. of wood, chemicals, metals
3 = Manuf. of electrical equipment, vehicles
4 = Electricity, Gas, Water, Construction
5 = Wholesale, retail, Hotels, Restaurants
6 = Transport, Communication, Finance
7 = Real estate, Computing, R&D, Public Admin
8 = Education, Health, Social Work
9 = Other services

Section 5 uses the following SOC 1-digit coding:
1 = Managers and Senior Officials
2 = Professional Occupations
3 = Associate Prof & Tech Occupations
4 = Administrative and Secretarial Occupations
5 = Skilled Trades Occupations
6 = Personal Service Occupations
7 = Sales and Customer Service Occupations
8 = Process, Plant and Machine Operatives
9 = Elementary occupations

We classify workers in to four skill groups based on a mapping of two digit SOC codes provided by the SOC2000 documentation. Two digit SOC1990 codes were then mapped in to this classification using information contained in the SOC1990 documentation.

Table A1.1: The definition of skill levels based on 2-digit occupational coding

Skill level	SOC1990	SOC2000
4 = highest	10-15, 19-27, 29	11, 21-24
3	16, 17, 30-39, 50-61, 70, 71, 90	12, 31-35, 51-54
2	40-46, 49, 62-67, 69, 72, 73, 79, 80-89	41, 42, 61, 62, 71, 72, 81, 82
1 = lowest	91-95, 99	91, 92

Notes: In ASHE data (1998-2008), 51% of the individuals move between occupations of different skill levels. 13% remain always in skill level 4, while 4% remain always at level 1

Proxy for years of education is based on the occupational category SOC2000 for years 2002-2008 and SOC90 for years 1998-2001. January-March sweeps of LFS of years 1997-2000 and 2002-2008 have been used to predict years of education as a linear function $Educ = \alpha + \beta_1 C + \beta_2 C^2$ for each occupation group at the lowest possible level (3 digit for SOC1990; 4 digit for SOC2000), where C is the birth cohort. Coefficients α, β_1 and β_2 are then applied to the NES/ASHE data, and individual-specific medians are taken as fixed “years of education”.

2. Generalised Transport Costs and Counterfactuals

2.1 Generalised Transport Costs (GTCs)

Ward to ward generalized costs (driving) were provided by the DfT. These costs comprise fuel and on-fuel vehicle operating costs and the value of time multiplied by the travel time. The data have been averaged for peak and off peak. The exact formulae for these calculations can be found in the DfT's Transport Economics Note (DETR 2001).

Ward to ward driving GTC were provided in 1998 prices. To be consistent with the rail GTC we follow standard practice and update to 2004 values using the Retail Price Index (RPI) available from:
http://www.statistics.gov.uk/CCI/nugget.asp?ID=21.

Local Authority to Local Authority driving GTC (used in the commuting section and to construct one of the counterfactuals) are constructed by averaging the ward-to-ward GTCs, across Local Authorities. That is, for a given pair of LAs (a and b) we average the GTCs for all ward-ward pairs i and j, where i is in Local Authority a and j is in Local Authority b. This averaging is carried out using ward weights based on the number of postal delivery points recorded in each ward in the National Statistics Postcode Directory (which therefore approximates residential population weighting). To be more precise, to get an estimate of the GTCs for Local Authority a to Local Authority b, we proceed in two steps: 1) average the GTCs for all wards in a, to all destination wards j in Local Authority b, using the number of postal delivery points (addresses) in wards j as weights; 2) average the GTCs to Local Authority b from each ward i in Local Authority a created in step 1, using the number of delivery points in wards i as weights. This procedure is equivalent to that described on p.21. of Head and Mayer (2002).

The information for time travel by train stems from Base Year (2004) Rail 'Level of Service' skims based on the Midman rail data. The original data provided to us by DfT are split by travel purpose: Employer Business (EB), Commute (or Work, W) and Other (O). This train time data was provided in the form of Local Authority to Local Authority origin-destination matrices.

The final GTC matrix by train is a weighted sum[16] of in-vehicle, wait and walk times (multiplied by the respective time value) and fare matrices. Specifically, for each travel purpose [Employer Business (EB), Commute (or Work, W) and Other (O)], four skims are used to construct the GTC by train:

- fare in British pounds;
- in vehicle time in hours
- average total wait in hours; and
- access time in hours.

Information about values of travel time were taken from DfT "Values of Time and Operating Costs" available at:

//www.webtag.org.uk/webdocuments/3_Expert/5_Economy_Objective/3.5.6.htm. These are the latest values of time recommended by the DfT for use in most routine economic appraisals of transport projects. The prices in the document are in 2002, but we inflated the values to 2004 prices (since Fares are in 2004 values) using the index suggested in the DfT documentation: 4.244% for working time (EB), and 3.3881% for non-working time (commuting and other).

16. The weights are defined below.

The time values used to compute the GTC by train are thus as follow:

In-vehicle: Values of Working (EB) and Non- Working (C and O) Time per person (£ per hour, 2002 prices and values)

In-vehicle: Values of Working (EB) and Non- Working (C and O) Time per person (£ per hour, 2002 prices and values)

Vehicle Occupant	Perceived Cost
Rail passenger (Employer's Business)	30.57
Commuting (C)	5.04
Other (O)	4.46

Notes: Time spent travelling during the working day is a cost to the employer's business. 'Commuting' is travelling to and from the normal place of work. 'Other' is travel for other non-work purposes, for example leisure trips. There is no differentiation of 'commuting' and 'other' values of time by mode.

Waiting: The values for non-working time ('commuting' and 'other') spent waiting for public transport is two and a half times the 'commuting' and 'other' values. In the appraisal process, changes in travel time on employer's business are valued the same whatever stage of the journey is involved, i.e. there is no weighting applied to take account of the reluctance of passengers to walk to/from or wait for transport services.

Access (walking): Where walking and cycling is used as a means of inter-changing between modes of transport, the non-working values ('commuting' and 'other') of walking and cycling is twice the 'commuting' and 'other' values. In the appraisal process, changes in travel time on employer's business are valued the same whatever stage of the journey is involved, i.e. there is no weighting applied to take account of the reluctance of passengers to walk to/from or wait for transport services.

Each matrix (404x404) should have 163,216 observations[17]. There are, however, 90 missing observations in the wait times, in-vehicle times, and fare matrices. There are no missing observations for the walk time matrices (EB, Work and Other). Rather than drop these observations, we predict missing values as follows:

In the Wait time Matrices (EB, Work and Other) we use the average for Local Authority i of all non-missing Local Authority i to Local Authority j wait times.

To predict the missing values in the In-vehicle Time and Fare Matrices (EB, Work and Other) we regress In-vehicle time or fare on driving GTC (see above) and straight line distances calculated using Local Authority centroids as described in the text. The estimated coefficients for each travel purpose are as follow[18]:

17. Great Britain has 408 Local Authorities (LA)/Districts. Four LAs were not considered in the analysis since there are no links by train: Eilean Siar, Isles of Scilly, Orkney Islands, and Shetland Islands.

18. Note that the travel times in the Employer Business (EB) and Other (O) matrices are the same.

	IN-VEHICLE TIME results			FARE results		
	EB	Work	Other	EB	Work	Other
Driving GTC	4.39	4.46	4.39	0.114	0.067	0.039
s.e	0.02	0.02	0.02	0.008	0.005	0.003
Driving (kms)	-0.46	-0.48	-0.46	0.221	0.129	0.076
s.e	0.005	0.005	0.005	0.002	0.001	0.00069
Intercept	35.66	37.54	35.66	20.68	12.072	7.138
s.e	0.19	0.20	0.19	0.074	0.043	0.026
Adjusted R-squared	0.862	0.855	0.862	0.855	0.855	0.855
Observations	163,126	163,126	163,126	163,126	163,126	163,126

Finally, we weight each of the different GTC using the following proportions of trips by journey purpose [Employer Business (EB), Commute (or Work, W) and Other (O)] provided in the DfT documentation "Values of Time and Operating Costs"[19]:

Proportion of Trips Made in Work and Non-Work Time

Mode / Vehicle Type & Journey Purpose	All Week Average
Heavy Rail	
Employer Business (EB)	7.6%
Commuting (or Work, W)	52.2%
Other(O)	40.3%
Total	100%

Given the above numbers, the formulas to construct the final GTC by train are (note that time travel in the matrices are in minutes, so we divide by 60 to get time per hour):

Cost_EB_final = ((TIME_IVT_EB_final/60)*(30.57*1.04244)) + ((TIME_WAIT_EB/60)*(30.57*1.04244)) + ((TIME_WALK_EB/60)*(30.57*1.04244));

Cost_Work_final = ((TIME_IVT_Work_final/60)*(5.04*1.033881)) + ((TIME_WAIT_Work/60)*(5.04*1.033881)*2.5) + ((TIME_WALK_Work/60)*(5.04*1.033881)*2.0);

Cost_Other_final = ((TIME_IVT_Other_final/60)*(4.46*1.033881)) + ((TIME_WAIT_Other/60)*(4.46*1.033881)*2.5) + ((TIME_WALK_Other/60)*(4.46*1.033881)*2.0);

Weighting matrices using proportions of trips by journey purpose:

WEIGHTED_COST_TRAIN = (Cost_EB_final*0.076) + (Cost_Other_final*0.403) + (Cost_Work_final*0.522);

FARES_WEIGHTED = (FARES_EB_pounds_final*0.076) + FARES_OTHER_pounds_final*0.403) + (FARES_WORK_pounds_final*0.522);

Finally, we sum the weighted matrix of in-vehicle, wait, walk times (multiplied by the respective time value) and the weighted fare matrix to get the Generalized Transportation Costs (GTC) by train between Local Authorities:

GTC_TRAIN = WEIGHTED_COST_TRAIN + FARES_WEIGHTED;

19. Available at: http//www.webtag.org.uk/webdocuments/3_Expert/5_Economy_Objective/3.5.6.htm.

Within Local Authority times are not provided in any of the data sets. We estimate within Local Authority GTC by taking the minimum GTC to neighbouring Local Authorities and assuming employment is uniform within the circle of radius given by that distance (this amounts to using the value of 35.4% of the minimum GTC to neighbouring Local Authorities).

2.2 Conterfactuals

We simulate the effects of various transport policies (the counterfactuals) by modifying the train and road GTC origin-destination matrices. Counterfactuals involving a k% reduction can be computed by simply reducing the costs in the matrices by k% in the appropriate areas e.g. reducing costs by 1% on all links between areas within the Manchester and Leeds city regions to simulate a 1% reduction in driving costs within these city regions.

The procedure for constructing counterfactuals for rail link improvements is more involved, because we want to permit a change in costs between two Local Authorities containing terminus stations to affect the costs on a wider range of Local Authority to Local Authority pairs. For example, a 20 minute reduction in journey times between Leeds and Manchester has three first order effects: 1) it reduces the journey time between Leeds and Manchester Local Authorities directly; 2) it reduces the journey times between Local Authorities that are already linked along the quickest route via Leeds and Manchester, and 3) it reduces the journey times between Local Authorities that were not connected via Leeds and Manchester, but are now connected more quickly via the Leeds-Manchester link.

There are also second order effects arising because any Local Authority to Local Authority link journey time that is reduced via the first order effects in 1) and 2) above, may provide a new quicker route for other Local Authority to Local Authority journeys.

It is not feasible to model these second order effects without analysing a fully specified rail network model. So, we simply estimate the changes in transport costs arising from first order effects, using the Local Authority to Local Authority origin-destination GTC matrix (constructed as explained above). The procedure is as follows, using a 20 minute reduction in the Leeds (l) – Manchester (m) journey time as an example. Define the cost of a journey between Local Authority a and Local Authority b as cost_a_b.

1) reduce cost_l_m and cost_m_l by the £ value corresponding to a 20 minute reduction of in-vehicle time
2) for an origin-destination Local Authority pair a,b, compute the new alternative route costs via Leeds and Manchester i.e. newcost1_a_b = cost_a_l + cost_l_m + cost_m_b and newcost2_a_b = cost_a_m + cost_m_l+ cost_l_b
3) replace the existing cost_a_b with the minimum new cost via Leeds and Manchester if (i) the minimum new cost is lower than cost_a_b, and (ii) the minimum cost via Leeds and Manchester was not already lower than cost_a_b. This second condition is based on the assumption that there may be unobserved factors (other than GTCs) that rule out travel via Leeds and Manchester.
4) repeat steps 2-3 for all a-b pairs.

3. Definitions of city regions

The geographical definitions of the city regions used in this report follow the criteria established by the Manchester Independent Economic Review (See the Regeneris report for the MIER):

- The travel to work patterns of people in higher managerial and professional occupations are used as the basis for defining the city regions.

- For each city, a core employment area has been identified. This is the local authority district or combination of districts which can reasonably be regarded as the central employment area.

- Census data showing the travel to work movements of higher managerial residents of districts to the core employment area determine the boundaries of the city region. A 15% threshold is set for inclusion, so any local authority district which sends 15% or more of its residents to the core employment area is defined as being within the city region. This is one method among several which are commonly used to define city regions, and is one which reasonably reflects the pull of a major employment area on an HMP population.

Table A4.1: List of City-Regions and Local Authorities

City Region	Local Authorities
Manchester	Bury; High Peak; Macclesfield; Manchester; Oldham; Rochdale; Salford; Stockport; Tameside; Trafford; Bolton; Warrington; Wigan; Vale Royal; Congleton
Sheffield	Rotherham; Sheffield; North Derbyshire; East Derbyshire
Nottingham	Nottinghamshire; Broxtowe; Erewash; Gedling; Rushcliffe; Ashfield; Newark and Sherwood
Newcastle	Newcastle; Gateshead; North Tyneside; South Tyneside; Blyth Valley; Wansbeck; Castle Morpeth; Tynedale; Derwentside; Chester-le-Street; Alnwick
Liverpool	Liverpool; Knowsley; Sefton; Wirral
Edinburgh	Edinburgh; Mid, East and West Lothian; Scottish Borders; Fife; Falkirk
Cardiff	Cardiff; Caerphilly; Merthyr Tydfil; Rhondda; Vale of Glamorgan; Newport
Birmingham	Birmingham; Solihull; Bromsgrove; Sandwell; Tamworth; Walsall; North Warwick; Lichfield; Redditch; Dudley
Leeds-Bradford	Bradford; Leeds; Craven; Harrogate; Selby; Wakefield; Kirklees; Calderdale
Glasgow	Glasgow; West, East Dunbartonshire; East Renfrenshire; North, South Lanarkshire; Inverclyde; East, North Aryshire; Stirling
Bristol	Bristol; North Somerset; South Gloucester; Bath; North East Somerset;
Aberdeen	Aberdeen; Aberdeenshire;
Leicester	Leicester; Oadby & Wigston; Biaby; Harborough; Charnwood;
London	Barking and Dagenham; Barnet; Basildon; Bexley; Braintree; Brent; Brentwood; Bromley; Broxbourne; Camden; Castle Point; Chelmsford; Chiltern; City of London; Colchester; Croydon; Dacorum; Ealing; East Hertfordshire; Elmbridge; Epping Forest; Epson and Ewell; Gravesham; Greenwich; Guildford; Hackney; Hammersmith and Fulham; Haringey; Harlow; Harrow; Havering; Hertsmere; Hillingdon; Hounslow; Islington; Kensington and Chelsea; Kingston upon Thames; Lambeth; Lewisham; Maidstone; Maldon; Medway; Merton; Mid Sussex; Mole Valley; Newham; North Hertfordshire; Redbridge; Reigate and Bansted; Richmond upon Thames; Rochford; Sevenoaks; South Bucks; Southend on Sea; Southwark; St Albans; Sutton; Swale; Tanbridge; Three Rivers; Thurrock; Tonbridge and Malling; Tower Hamlets; Tunbridge Wells Uttlesford; Waltham Forest; Wandsworth; Waverley; Welwyn Hatfield; Westminster; Woking ; Stevenage; Enfield; Dartford; and Watford.

4. Data for spatial econometrics & structural model

This appendix describes the data used in sections 4 and 6. The data is at the level of the Local Authorities Districts (LAD) and City regions of England, Wales and Scotland. A consistent data set has been set up from different databases for the period 1998-2006.

Employment: Annual Business Inquiry (ABI/Nomis) gives the number of employees based on the location of the workplace (employment in thousand of persons) at Local Authority District level. The employment level for the years 1998-2008 is based on the ABI workplace employee analysis available on the Nomis database (the Office for National Statistics' on-line labour market statistics database). The Annual Employment Survey (AES) was replaced by the Annual Business Inquiry (ABI), from 1998.

Average hourly earnings: Annual Survey of Hours and Earnings (ASHE) gives estimates of the average hourly earnings of all full-time employees at the level of Unitary Authority and Local Authority (UALAD) districts in England based on the location of workplace from 1998 to 2007. ASHE is based on the 1% random sample of employee jobs taken from HM Revenue & Customs PAYE (Pay-As-You-Earn) records.

GDP (per employee): The regional Gross Domestic Product (GDP) series are taken from the most recent version of the EUROSTAT-REGIO. GDP estimates are available annually from 1995 to 2006 at NUTS 3 level. The values are at current market prices in millions of euros from 1.1.1999 and in millions of ECU up to 31.12.1998. To calculate GDP per employee we employ the ABI (Annual Business Inquiry) workplace employee information described above.

Population: Mid-year estimates of the total number of persons resident in British districts are available from Office for National Statistics between 1981 and 2007.

Local sectoral composition: ABI/Nomis provides aggregated data in broad industrial groups between 1998 and 2007:

1 Agriculture and fishing
2 Energy and water
3 Manufacturing
4 Construction
5 Distribution, hotels and restaurants
6 Transport and communications
7 Banking, finance and insurance, etc
8 Public administration, education & health
9 Other services

Local occupation composition: Annual Population Survey – APS (workplace analysis) provides occupation at district level between 2004 and 2007. The data set is split into nine categories:

1 Managers and Senior Officials
2 Professional Occupations
3 Associate Prof & Tech Occupations
4 Administrative and Secretarial Occupations
5 Skilled Trades Occupations

6 Personal Service Occupations
7 Sales and Customer Service Occupations
8 Process, Plant and Machine Operatives
9 Elementary occupations

Local average age of the population: Mid-year population estimates from ONS provide total population by age (Aged under 1 year, Aged 1 - 4 years, Aged 5 - 9 years, etc). WE use these to caculate average age of the population at the district level between 1998 and 2006.

Local educational level: This data set is found in Census 2001/Nomis for England and Wales and in Scotland's Census (2001) Results OnLine (SCROL) for Scotland. The categories are (at district level):

Local Education level	
No qualifications	No academic, vocational or professional qualifications.
Level 1	1+ 'O' levels/CSE/GCSE (any grade), NVQ level 1, Foundation GNVQ.
Level 2	5+ 'O' levels, 5+ CSEs (grade 1), 5+ GCSEs (grade A - C), School Certificate, 1+ 'A' levels/'AS' levels, NVQ level 2, Intermediate GNVQ or equivalents.
Level 3	2+ 'A' levels, 4+ 'AS' levels, Higher School Certificate, NVQ level 3, Advanced GNVQ or equivalents.
Level 4/5	First degree, Higher Degree, NVQ levels 4 - 5, HNC, HND, Qualified Teacher Status, Qualified Medical Doctor, Qualified Dentist, Qualified Nurse, Midwife, Health Visitor or equivalents.
Other qualifications/ level unknown	Other qualifications (e.g. City and Guilds, RSA/OCR, BTEC/Edexcel), Other Professional Qualifications.

5. Additional results for spatial econometrics section

Figure 18 – Moran's I (scatter plots) of GDP per worker in 2006, Differences in GDP per worker and annual growth rates of GDP per worker between 1998 and 2006 (128 NUTS3)

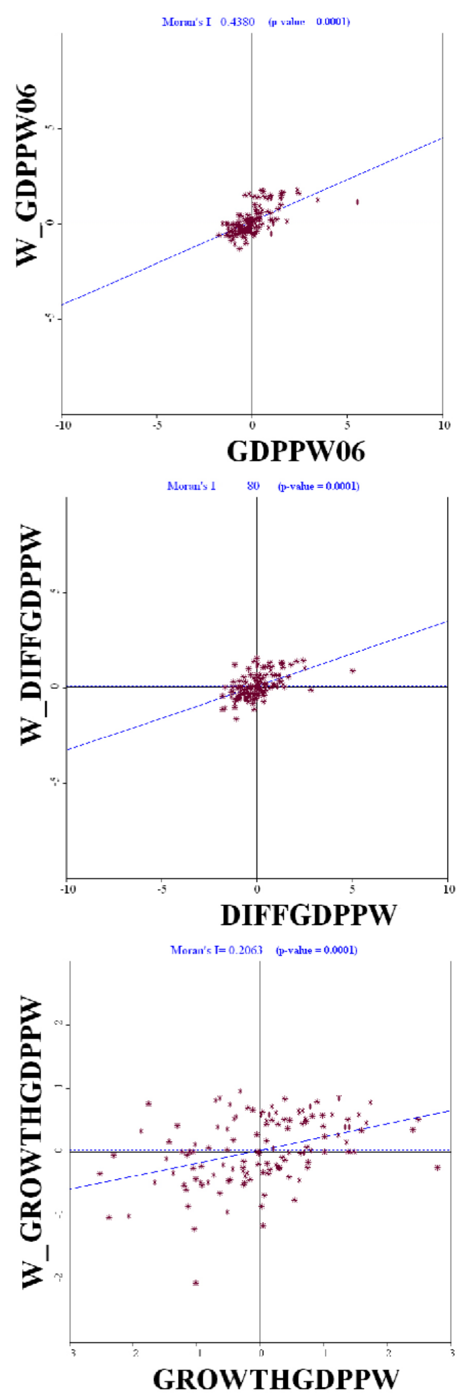
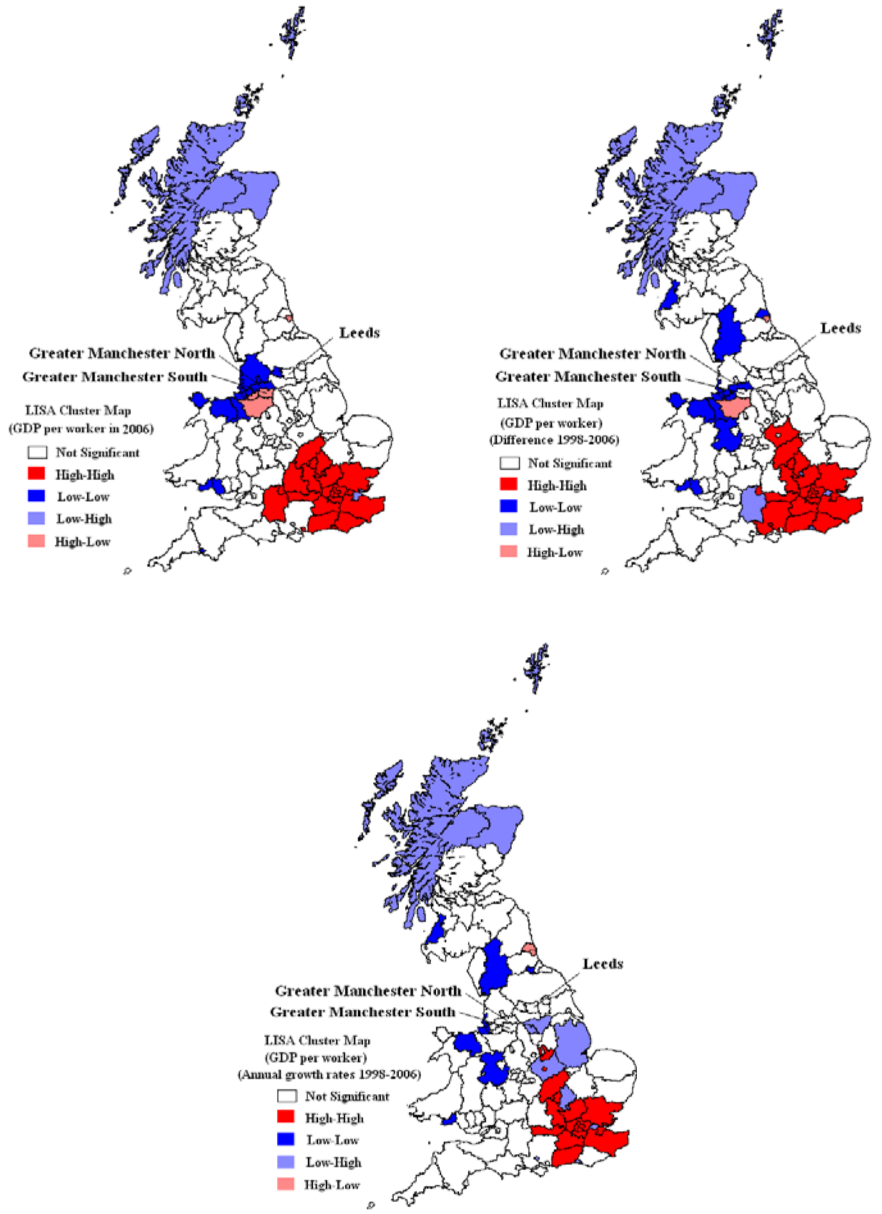


Figure 19 – LISA Cluster Map (GDP per worker in 2006, Differences in GDP per worker and annual growth rates of GDP per worker between 1998 and 2006, 128 NUTS3)



6. Full results for labour market regressions

The following table provides the results for all coefficients for the wage levels specifications described in the text.

Results for all coefficients for wage level specifications

	1			2			3		
Variable	Coef.	S.E.		Coef.	S.E.		Coef.	S.E.	
ln Density (Car)	0.084	0.122		0.074	0.118		0.071	0.080	
ln Density (Train)	0.258	0.093	***	0.277	0.090	***	0.173	0.059	***
Female				-0.197	0.007	***	-0.221	0.010	***
Age				0.083	0.003	***	0.071	0.003	***
Age2				-0.001	0.000	***	-0.001	0.000	***
Years of Education							0.199	0.001	***
Public Sector									
Part-time									
Collective agreement									
SOC90 = 2									
SOC90 = 3									
SOC90 = 4									
SOC90 = 5									
SOC90 = 6									
SOC90 = 7									
SOC90 = 8									
SOC90 = 9									
SOC2000 = 2									
SOC2000 = 3									
SOC2000 = 4									
SOC2000 = 5									
SOC2000 = 6									
SOC2000 = 7									
SOC2000 = 8									
SOC2000 = 9									
Industry SIC 2									
Industry SIC 3									
Industry SIC 4									
Industry SIC 5									
Industry SIC 6									
Industry SIC 7									
Industry SIC 8									
Industry SIC 9									
Industrial diversity									
Occupational diversity									
Proportion highs skills									
Proportion intermed. skills									
Proportion in SIC = 2									
Proportion in SIC = 3									
Proportion in SIC = 4									
Proportion in SIC = 5									
Proportion in SIC = 6									
Proportion in SIC = 7									
Proportion in SIC = 8									
Proportion in SIC = 9									
ln Distance to London									
Year = 1999	0.055	0.003	***	0.054	0.002	***	0.041	0.002	***
Year = 2000	0.095	0.004	***	0.091	0.003	***	0.069	0.003	***
Year = 2001	0.147	0.004	***	0.145	0.004	***	0.115	0.003	***
Year = 2002	0.180	0.005	***	0.179	0.005	***	0.140	0.003	***
Year = 2003	0.250	0.003	***	0.250	0.002	***	0.193	0.002	***
Year = 2004	0.242	0.004	***	0.245	0.005	***	0.191	0.003	***
Year = 2005	0.267	0.004	***	0.273	0.005	***	0.213	0.004	***
Year = 2006	0.300	0.005	***	0.305	0.007	***	0.236	0.006	***
Year = 2007	0.329	0.005	***	0.340	0.007	***	0.266	0.006	***
Constant	-2.515	0.745	***	-4.192	0.814	***	-5.213	0.608	***
Individual Fixed Effects	No			No			No		
Observations	1102527			1091551			1091551		
R2	0.0902			0.2178			0.513		

4			5			6			7		
Coef.	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
0.054	0.066		0.046	0.058		0.069	0.016	***	0.070	0.021	***
0.165	0.049	***	0.170	0.044	***	0.049	0.014	***	0.030	0.010	***
-0.134	0.006	***	-0.117	0.007	***						
0.049	0.002	***	0.046	0.001	***	0.051	0.002	***	0.051	0.002	**
0.000	0.000	***	0.000	0.000	***	-0.001	0.000	***	-0.001	0.000	***
0.103	0.002	***	0.103	0.001	***						
0.042	0.004	***	0.058	0.004	***	0.047	0.005	***	0.047	0.005	***
-0.118	0.005	***	-0.096	0.004	***	-0.007	0.002	***	-0.007	0.002	***
0.011	0.007		0.007	0.007		0.004	0.002	*	0.004	0.002	*
-0.126	0.015	***	-0.119	0.011	***	0.024	0.004	***	0.024	0.004	***
-0.221	0.004	***	-0.223	0.004	***	-0.082	0.002	***	-0.082	0.002	***
-0.409	0.004	***	-0.429	0.004	***	-0.161	0.002	***	-0.161	0.002	***
-0.388	0.009	***	-0.396	0.005	***	-0.146	0.003	***	-0.146	0.003	***
-0.541	0.006	***	-0.531	0.005	***	-0.200	0.005	***	-0.200	0.005	***
-0.596	0.015	***	-0.547	0.009	***	-0.234	0.005	***	-0.233	0.005	***
-0.517	0.009	***	-0.541	0.008	***	-0.194	0.004	***	-0.194	0.004	***
-0.631	0.015	***	-0.623	0.013	***	-0.237	0.005	***	-0.237	0.005	***
-0.085	0.009	***	-0.073	0.006	***	0.005	0.006		0.005	0.006	
-0.175	0.014	***	-0.177	0.014	***	-0.060	0.003	***	-0.060	0.003	***
-0.438	0.010	***	-0.456	0.009	***	-0.144	0.005	***	-0.144	0.005	***
-0.390	0.008	***	-0.400	0.010	***	-0.107	0.005	***	-0.107	0.004	***
-0.521	0.011	***	-0.500	0.012	***	-0.167	0.006	***	-0.167	0.006	***
-0.565	0.008	***	-0.509	0.005	***	-0.172	0.003	***	-0.172	0.003	***
-0.510	0.007	***	-0.534	0.007	***	-0.144	0.004	***	-0.144	0.004	***
-0.648	0.005	***	-0.654	0.004	***	-0.217	0.004	***	-0.217	0.004	***
			0.062	0.010	***	0.028	0.004	***	0.028	0.004	***
			0.099	0.016	***	0.034	0.006	***	0.035	0.006	***
			0.085	0.010	***	0.019	0.005	***	0.019	0.005	***
			-0.088	0.011	***	-0.066	0.003	***	-0.066	0.003	***
			0.128	0.018	***	0.055	0.009	***	0.056	0.009	***
			0.035	0.007	***	-0.013	0.004	***	-0.013	0.004	***
			-0.028	0.009	***	-0.025	0.005	***	-0.025	0.005	***
			-0.061	0.009	***	-0.042	0.005	***	-0.042	0.005	***
									0.066	0.065	
									0.075	0.380	
									0.149	0.029	***
								-0.020	0.020		
									-0.038	0.089	
									-0.065	0.094	
									-0.043	0.142	
									0.463	0.366	
									0.202	0.140	
									-0.610	0.334	
									1.118	1.858	
									-1.440	0.322	***
									0.000	0.007	
0.044	0.001	***	0.045	0.001	***	0.067	0.001	***	0.065	0.002	***
0.074	0.002	***	0.074	0.002	***	0.125	0.002	***	0.121	0.002	***
0.121	0.002	***	0.122	0.002	***	0.198	0.004	***	0.194	0.003	***
0.165	0.009	***	0.168	0.008	***	0.276	0.008	***	0.272	0.007	***
0.221	0.011	***	0.223	0.011	***	0.345	0.008	***	0.338	0.008	***
0.227	0.008	***	0.232	0.007	***	0.388	0.008	***	0.382	0.008	***
0.256	0.006	***	0.260	0.005	***	0.443	0.008	***	0.435	0.008	***
0.280	0.004	***	0.284	0.004	***	0.498	0.009	***	0.489	0.009	***
0.314	0.004	***	0.315	0.003	***	0.555	0.010	***	0.545	0.010	***
-2.853	0.544	***	-2.785	0.538	***	-0.355	0.079	***	-0.068	0.304	
No			No			Yes			Yes		
1091551			1090528			1090528			1090528		
0.6223			0.6375			0.9182			0.9183		

The following table provides the results for all coefficients for the overall wage growth specifications described in the text.

Annualized % between job wage growth	(1)	(2)	(3)	(5)	(6)	(8)	(9)	(10)	(11)	(12)	(13)
lmed_driving_GTC	0.02**		-0.02**			-0.00	-0.00	0.01*	0.01**		0.00
	(2.88)		(3.20)			(0.53)	(0.49)	(2.17)	(2.60)		(1.21)
lmed_train_GTC		0.04**	0.07**			0.02**	0.01*	-0.01*	-0.01		-0.00
		(3.47)	(8.17)			(2.79)	(1.98)	(2.04)	(1.95)		(0.12)
lmtotempl				0.01*	0.01*					0.00*	
				(2.26)	(2.06)					(2.54)	
lmarea					0.01*					-0.00	
					(2.02)					(1.23)	
female						0.03**	0.03**	-0.02**	-0.02**	-0.02**	-0.02**
						(8.16)	(8.55)	(9.56)	(9.86)	(9.67)	(10.27)
age						-0.04**	-0.04**	-0.03**	-0.03**	-0.03**	-0.03**
						(31.24)	(31.65)	(32.33)	(31.83)	(31.54)	(31.27)
exp_start						0.00**	0.00**	0.00**	0.00**	0.00**	0.00**
						(27.77)	(29.90)	(31.45)	(31.14)	(31.05)	(30.59)
eduy							0.01**	0.03**	0.04**	0.03**	0.04**
							(10.65)	(22.31)	(22.57)	(23.21)	(23.21)
av_socds2_w								-0.03**	-0.03**	-0.03**	-0.03**
Av 1 digit (indiv level)											
occ dummies)								(3.57)	(3.83)	(3.83)	(3.15)
av_socds3_w								0.03**	0.03**	0.03**	0.03**
								(4.50)	(4.61)	(4.52)	(5.14)
av_socds4_w								0.05**	0.05**	0.05**	0.06**
								(9.06)	(9.63)	(9.81)	(11.28)
av_socds5_w								0.04**	0.04**	0.04**	0.05**
								(7.08)	(6.85)	(6.95)	(7.55)
av_socds6_w								0.04**	0.04**	0.04**	0.05**
								(6.23)	(5.49)	(5.61)	(6.80)
av_socds7_w								0.08**	0.07**	0.07**	0.08**
								(7.94)	(7.82)	(7.63)	(7.88)
av_socds8_w								0.04**	0.04**	0.04**	0.06**
								(6.83)	(7.71)	(8.23)	(10.35)
av_socds9_w								0.04**	0.04**	0.04**	0.05**
								(7.65)	(7.34)	(7.46)	(9.22)
av_socd90s2_w								-0.12**	-0.13**	-0.12**	-0.11**
								(10.18)	(10.76)	(10.94)	(9.78)
av_socd90s3_w								-0.07**	-0.07**	-0.07**	-0.06**
								(15.17)	(15.67)	(15.43)	(11.30)
av_socd90s4_w								-0.07**	-0.06**	-0.06**	-0.05**
								(14.15)	(12.93)	(13.48)	(9.78)
av_socd90s5_w								-0.01	-0.00	-0.00	0.02**
								(1.21)	(0.76)	(0.62)	(2.62)
av_socd90s6_w								0.05**	0.05**	0.05**	0.06**
								(8.22)	(7.30)	(7.55)	(8.95)
av_socd90s7_w								0.13**	0.12**	0.12**	0.14**
								(10.45)	(9.09)	(8.94)	(10.36)
av_socd90s8_w								-0.01	0.00	0.01	0.03**
								(1.08)	(0.99)	(1.34)	(4.90)
av_socd90s9_w								0.04**	0.04**	0.04**	0.06**
								(7.22)	(7.16)	(7.29)	(9.79)
av_pubsec								-0.01	-0.01**	-0.01*	-0.01**
								(1.96)	(2.80)	(2.51)	(3.02)

Annualized % between job wage growth	(1)	(2)	(3)	(5)	(6)	(8)	(9)	(10)	(11)	(12)	(13)
av_parttime								0.15**	0.14**	0.14**	0.14**
								(40.89)	(38.34)	(37.74)	(38.88)
av_colag								-0.01*	-0.01*	-0.01*	-0.00
								(2.00)	(2.09)	(2.13)	(0.85)
av_s_lsic1d_2_w									-0.03**	-0.03**	0.00
Av 1 digit (indiv level)											
industry dummies									(4.08)	(3.96)	(1.04)
av_s_lsic1d_3_w									-0.02**	-0.03**	0.01
									(4.80)	(4.94)	(1.36)
av_s_lsic1d_4_w									0.01*	0.01	0.04**
									(2.54)	(1.91)	(8.37)
av_s_lsic1d_5_w									0.01	0.01	0.04**
									(1.29)	(0.67)	(7.96)
av_s_lsic1d_6_w									-0.02**	-0.03**	0.00
									(4.52)	(5.21)	(0.50)
av_s_lsic1d_7_w									-0.01	-0.01*	0.02**
									(1.44)	(2.23)	(3.91)
av_s_lsic1d_8_w									0.01	0.00	0.03**
									(1.21)	(0.42)	(6.62)
av_s_lsic1d_9_w									0.00	-0.00	0.03**
									(0.19)	(0.70)	(4.96)
mhiskills											0.15**
											(5.53)
mintskills											0.10
											(1.20)
moccdiversity2											0.21
											(0.67)
mdiversity2											0.26
											(1.01)
ms1d2											-0.32*
Av 1 digit (ttwa level)											
industry shares											(2.40)
ms1d3											0.08
											(0.56)
ms1d4											-1.46**
											(2.73)
ms1d5											-0.82
											(1.32)
ms1d6											-0.13
											(0.57)
ms1d7											-0.17
											(0.50)
ms1d8											2.84
											(1.40)
ms1d9											-0.66
											(1.16)
Constant	-0.11	-0.40*	-0.46**	0.03	0.01	0.83**	0.73**	0.46**	0.44**	0.37**	0.00
	(1.18)	(2.51)	(3.07)	(0.57)	(0.18)	(6.51)	(4.96)	(4.28)	(4.20)	(5.20)	(1.21)
Observations	248118	251915	248068	252128	252128	246125	246125	246125	246125	250154	-0.00
R-squared	0.00	0.00	0.00	0.00	0.00	0.08	0.08	0.10	0.10	0.10	(0.12)

Robust t statistics in parentheses* significant at 5%; ** significant at 1%

If desired, one can use the following employment shares to weight for the aggregate city-region impact of the counterfactuals reported in section 4.

LA	City	Pop	share of City	share of Total
Bury	Manchester	62402	0.042	0.024
Wigan	Manchester	103058	0.069	0.039
Bolton	Manchester	109127	0.073	0.041
Oldham	Manchester	80895	0.054	0.031
Salford	Manchester	116384	0.078	0.044
Rochdale	Manchester	77359	0.052	0.029
Tameside	Manchester	75148	0.050	0.028
Trafford	Manchester	130452	0.087	0.049
Congleton	Manchester	34275	0.023	0.013
High Peak	Manchester	31005	0.021	0.012
Stockport	Manchester	124014	0.083	0.047
Manchester	Manchester	297346	0.199	0.112
Vale Royal	Manchester	50299	0.034	0.019
Warrington	Manchester	119333	0.080	0.045
Macclesfield	Manchester	81883	0.055	0.031
Leeds	Leeds	425978	0.369	0.161
Selby	Leeds	31791	0.028	0.012
Craven	Leeds	28688	0.025	0.011
Bradford	Leeds	202521	0.176	0.077
Kirklees	Leeds	161203	0.140	0.061
Harrogate	Leeds	76253	0.066	0.029
Wakefield	Leeds	140948	0.122	0.053
Calderdale	Leeds	86363	0.075	0.033
Manchester total		1492980		
Leeds-Bradford total		1153745		
Total		**2646725**		

Source: ONS