GLAECONOMICS

Working Paper 17: Why distance doesn't die:
Agglomeration and its benefits

by Paul Ormerod, William Cook and Bridget Rosewell,
Volterra Consulting Ltd

June 2006

Transport for London

LONDON DEVELOPMENT AGENCY

MAYOR OF LONDON

copyright

Greater London Authority
June 2006

Published by
Greater London Authority
City Hall
The Queen's Walk
London SE1 2AA
www.london.gov.uk
enquiries **020 7983 4000**
minicom **020 7983 4458**

ISBN: 1 85261 881 7

This publication is printed on recycled paper.

For more information about this publication, please contact:
GLA Economics
telephone 020 7983 4922
email glaeconomics@london.gov.uk

GLA Economics provides expert advice and analysis on London's economy and the economic issues facing the capital. Data and analysis from GLA Economics form a basis for the policy and investment decisions facing the Mayor of London and the GLA group. The unit is funded by the Greater London Authority, Transport for London and the London Development Agency.

Contents

1. Introduction

Cities are core elements in the existence and development of civilisations. If we want to understand the origins and nature of any particular civilisation, whether ancient or modern, it is its cities that we need to unearth, investigate or visit.

Yet economic theory and analysis says little about these phenomena or to explain their existence, persistence and size. There is plenty of analysis of the firm, individual behaviour and countrywide aggregates. There is very little that recognises that all economic activity is situated in time and space and is also both constrained and energised by these limitations.

As a result, policy too often ignores the significance of the spatial location of activities; or has responded to spatial issues, such as regional policy, by focusing on sectors, on individual firms or on housing. This working paper explores the role of spatial concentration in generating economic benefits and making possible the range of activities that London exhibits. We begin by defining what agglomeration means, then explore the mechanisms of how agglomeration happens and finally look at the persistence of agglomeration over time.

2. Agglomeration

The term agglomeration refers to the way in which activities are stuck together. It gives a good indication as to what goes on in the situations where businesses, people and institutions come together. Even the word itself sounds gluey - and no parts of the economy act in isolation. Isolation is subsistence, where no trade takes place and everything has to be produced by an individual unit. It is the recipe for poverty at best and famine at worst.

Agglomeration gives the opportunity to trade and exchange as well as to build the institutions that regulate and empower such exchanges. It is important to realise that successive developments in technology, which has improved our ability to communicate, have not made cities redundant. Indeed a higher proportion of the world's population lives in cities than at any other time in history. In the USA, 75 per cent of people live in cities. Far from being spread evenly over the available geography, we concentrate in particular areas. The telegraph, trains, telephones, aeroplanes, the Internet and email don't spread us more evenly rather they concentrate us. We tend to travel more, and may engage in video conferencing but this does not necessarily replace the need for face-to-face interactions.

A recent description of the making of the Lord of the Rings films describes how the director, Peter Jackson, was needed simultaneously in London to record the music for the film and in Wellington, New Zealand, for the sound effects editing. Only bandwidth saved this situation, allowing him to view edits from across the world and videoconference with the Wellington team. But even so, the interview with the sound team ended with the observation that Jackson returned to Wellington for the final sound edits, and that this was essential. Moreover, each team, whether for the music or the sound effects, was large. Each needed a city to operate in, even if the scale of Wellington is very different from that of London.

3. The mechanisms of agglomeration

Cities exist as an agglomeration of economic activities in which the costs of being crowded must be outweighed by the benefits of the location to business and residents. Preferences appear to exist over a wide range of locations for co-location of many activities, including financial and business services.

Some of the benefits of this co-location may be direct cost reductions, for example access to concentrated markets and a straightforward reduction in costs of delivery. Others may not be directly costed, such as knowledge spillovers. Emergence of economies of scale will be a private benefit beyond the previous equilibrium outcome.

We might think of the issue first of all in a general equilibrium framework. By this we mean an analysis, which tries to incorporate all possible reactions to a changes across all elements of the economy. This allows us to abstract from the very important issues of the process and the time-scale over which a transition from one equilibrium Financial Business Services (FBS) cluster in the City to another FBS cluster in, for example, Leeds might take place. This issue will be returned to later.

Suppose the current location of industries in the UK corresponds to a general equilibrium. There is no role for increasing returns, whether internal or external to the firm. We now posit a change in relative prices. The City as a location becomes more expensive (e.g. through an increase in congestion). We abstract from the process of moving to the new equilibrium and imagine it is once again called into existence (possibly by the work of an auctioneer, who calls out all the possible prices until a new equilibrium is discovered). Can we say anything useful about the demand for locating in the City and whether firms might locate elsewhere in the UK?

The revealed preference of agents in this assumed general equilibrium is that firms in the FBS sector prefer to locate near to one another. We do not need to know why these preferences exist, but simply that revealed preference tells us that they do. Further, we must hold them fixed in order to compare general equilibrium situations in any meaningful way. It is important to remember that in general equilibrium, agents are maximising utility subject to their given tastes and preferences. These can differ across agents but the key point here is that they are assumed to be fixed.

Of course, if in some way during the process of auctioneering, firms could signal to one another that they would all be willing to locate to Leeds, this could emerge as a solution. We return to this point below. But this type of behaviour is certainly not envisaged in the process of establishing a general equilibrium solution. Given the tastes and preferences of agents, the auctioneer shouts out successive sets of relative prices until an equilibrium set can be found.

So long as the existing preference set is maintained, it seems unlikely that a migration of individual firms will take place – rather it would have to be a group. Alternatively, individual firms would have to either realise that their preference for co-location was misplaced or take the risk that it might be so. Either of these would require hypothetical information about what might happen in different circumstances, which is generally lacking.

If this is correct, the relocation to an alternative centre is the most likely option, or failure to expand in the way that might otherwise be possible with lower costs.

Even supposing a solution exists in which FBS in the City relocates to Leeds rather than to Frankfurt, there is the important question of the time-scale over which this process takes place. Little analysis has been done of the implied time scales in the transitions between equilibria in economic theory. But an important article by Atkinson[1], for example, suggests that in the neo-classical growth model, this is of the order of 100 years. Even with long-term investment projects this is well beyond normal timescales.

Thus far, we have considered the issue on the basis of a set of preferences. But these preferences are themselves based on real consideration.

Alfred Marshall asked long ago what sort of reasons might exist in practice for firms to favour agglomeration. We can list a number of important reasons why a firm might choose to locate in a place where agglomeration already exists. These do not need any form of increasing returns to production, either internal or external to the firm. And if we look more closely, we can see that these are easy to find in practice.

Marshall himself noted, for example:

- the availability of skilled workers
- the availability of intermediate goods (specialised inputs and services).

In other words, the costs of acquiring both information about these and then going on to acquire labour/services themselves are lower in an agglomeration.

Hotelling gave a famous example of the proximity of ice-cream sellers to the beach, illustrating the proximity to the output market. This assists cost minimisation on factors such as transport costs, marketing and promotional activity.

Marshall also noted a further feature, which is beginning to get close to the modern concept of increasing returns, which are external to the firm, namely, the relative ease of discussion and transmission of new ideas, which might be thought of as technology spillover.

So we may not need to posit increasing returns of any kind to order to explain *ex post* the existence of agglomerations of firms in FBS. Given the above features of an area, a profit-maximising firm producing under constant returns and with no benefits from increasing returns external to the firm, will choose to locate there. Of course, the

[1] AB Atkinson, *Review of Economic Studies*, 1969

existence of externalities of agglomeration external to the firm (analogous to those of post-neoclassical endogenous growth theory[2]) will reinforce any such decision.

The hypothesised absence of increasing returns in production does not detract from the fact that the agglomeration is still the most efficient place in which a firm of the relevant type locates.

Suppose the relative price of location in the City leads just one FBS firm to relocate. Where does the above approach suggest this marginal firm might move? It seems that it is likely to move to a place where the above cost-saving features might be best replicated. In other words, it would move to an existing FBS agglomeration outside the UK.

If a number of substantial firms decide to move at around the same time, there is no reason why the above logic should not hold for them also. The exception would be if in some way their decisions could be coordinated. If each agent knew that the others were likely to move to Leeds, say, rather than Frankfurt if every other relevant agent were going to move there as well, then it is possible that they all might do so.

In other words, in the absence of a Gosplan-like central planning agency to 'coordinate' firms' decisions, we might think of it in a game theoretic context. A firm wishes to move out of the City, and can choose, say, between Frankfurt, where an FBS agglomeration already exists, and Leeds, where one does not[3]. For simplicity, rather than envisaging a multi-agent game, just one of its rivals is also considering the question.

Essentially, we have a Prisoner's Dilemma-type game. Suppose that if both move to Leeds, substantial agglomeration benefits are created for both. Suppose further that these would be such as to outweigh the benefits of a Frankfurt location. Then the best solution for both is that they should both move to Leeds. But there is a temptation to defect. If one firm moves to Frankfurt whilst the other moves to Leeds, it gains over its rival, and vice versa.

It is well known that in the Prisoner's Dilemma with a fixed number of plays (in this case, just a single one), the best strategy for each player is to defect i.e. to move to Frankfurt.

In any event, we are making a rather important assumption on the pay-off matrix. Namely, that if both decide to move to Leeds, the benefits to both are greater than the benefits of them both moving to Frankfurt. This may or may not be justified empirically. Of course, the greater the number of agents we envisage moving, the more plausible this assumption becomes.

---

[2] for example, Mankiw, Romer and Weil, 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics,* 1992

[3] of course, there are FBS firms in Leeds, but the city does not immediately spring to mind as a world centre of FBS at present

Thinking about the process by which a Marshall/Hotelling-type agglomeration comes to exist in FBS is made even easier if we invoke increasing returns.

Krugman offers a simple regional model, which suggests many of the typical features of economic geography. A far from exclusive list would surely include the following three points:

- Self-organisation - atomistic interactions among individual agents can spontaneously generate large-scale order.
- Path dependence - small differences in initial conditions can have large effects on long-run outcomes.
- Discontinuous change - small quantitative changes in the underlying factors that drive location will sometimes produce large, qualitative changes in behaviour'[4].

Krugman's model is of an economic region, but we can think of it as being a smaller entity such as the City. The area has two kinds of economic activities. First, the export base i.e. goods and services that it produces and which are sold outside the area, whether abroad or elsewhere in the same country. Second, the 'non-base' activities that provide goods and services to the local area.

In this context, we can think of the large number of firms which provide intermediate products to the financial firms in the City. The size of this 'non-base' demand is very important to them and to their decision as to where to locate.

In traditional regional economics, there is a proportional relationship between the export income of a region and its total income. If α is the fraction of income spent locally, X the income earned in the export sector, and Y the area's total income, it is easy to show that:

$$Y = X / (1 - \alpha)$$

But suppose now that the range of goods and service produced locally depends upon the size of the local market, because it is not worth producing some things locally unless the market is large enough. In other words, α is an increasing function of Y.

Krugman postulates a simple relationship in which the share of income spent locally in period t is proportional to the size of the local economy in the previous period, up to a maximum of 80 per cent:

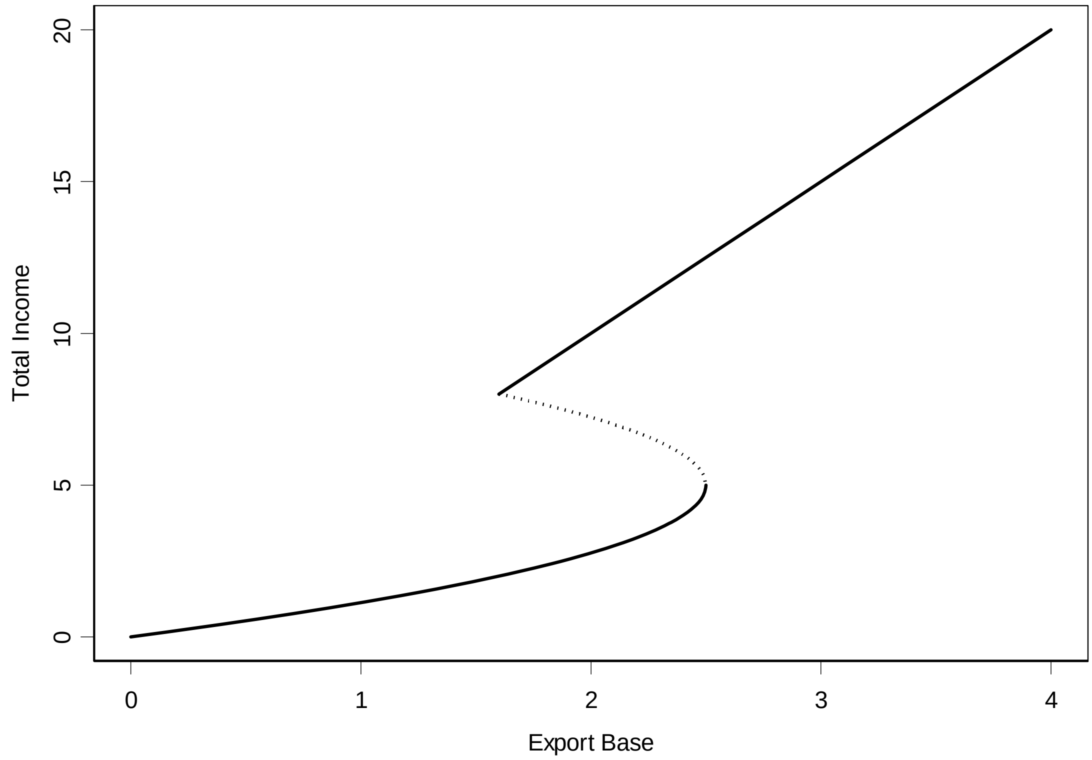
$$\alpha(t) = \min[\,0.8, 0.1Y(t-1)\,]$$

[4] P Krugman, 1997, 'How the economy organises itself in space', in Arthur, Durlauf and Lane, eds., *The Economy as an Evolving Complex System II*, vol. XXVII, Santa Fe Institute Studies in the Science of Complexity. Ormerod, 1998, provides a wide range of social and economic models with these properties (*Butterfly Economics*, Faber and Faber)

An equilibrium relationship can then be defined for Y, given X as a level of Y, such that Y(t+1) = Y(t).

The solution of this model is complex. There is a critical value of X above which there is a stable equilibrium, and there is a critical value below which another, separate equilibrium exists. But between these values, the equilibrium is unstable. Figure 3.1 plots the relationship between total income and the export base.

Figure 3.1: Equilibrium solution to Krugman base-multiplier model



The critical mass aspect of the model is apparent. If we start to expand the export base from a low level, the overall size of the local economy will expand gradually, on the lower of the two solid lines. But a critical point will be reached and the economy moves immediately to the upper solid line, to a much higher level. Alternatively, if the export base is shrinking down the upper line, at the critical point there will be a dramatic contraction.

We can readily think of the City as an area that at present sits well along the upper of the two solid lines. Other European financial centres such as Frankfurt or Paris also occupy this part of the graph, although presumably less stratospherically placed than the City. But one does not have to be a Lancastrian to imagine that Leeds, or indeed any other location for FBS in the UK, at present sits on the lower of the two solid lines.

This model is a helpful way of thinking about the problem of relocation. The export base firms in the City will have an understandably strong preference to be in a location that is on the upper part of the line. There, they reap the Marshall-Hotelling benefits.

If a location is chosen *de novo* as the potential basis for an FBS agglomeration, it is unlikely that we will find ourselves very far along the bottom line. If the alternative in the UK to the City is already near the point at which a critical mass forms, even one or two relocations could push it past the critical point. Eventually, this is an empirical issue.

Of course, to investigate the question in more depth, we require an agent-based model containing:

- workers (supply of skilled labour)
- intermediate firms (lawyers, merger and acquisition specialists)
- 'export' firms.

We can specify simple but reasonable behavioural rules for the agents, and explore the properties of the model with respect to location choice.

We need not necessarily specify externalities in production, as does the Krugman model. Instead, we can make the realistic assumption that the value of choosing a particular location for each set of these agents may be enhanced by the fact that others have previously chosen it. In other words, we can rely upon constant returns to scale, and the Marshall-Hotelling reasons for location.

In other words, the non-linearities arise in the process of establishing the agglomeration. Specifically, these are in the probabilities that agents will adopt a particular location, given that these probabilities will vary with the number that has already adopted it. Profit-maximising agents will be influenced in which location to choose, depending upon how many have already chosen it.

A number of powerful analytical results have been established which will certainly assist in helping to decide a number of assumptions in the model. For example, it appears to be the case that a key assumption is whether we assume that a fixed number of agents in the model make location decisions or whether there is an arbitrarily growing number of agents.

Kirman[5] provides a clear summary of this question. Suppose there are K locations, and x is the K vector of current shares of agents in each location. If p defines the probabilities for attracting a new agent, and denoting by x_i the proportion of agents already in the location, we have:

$$p = (p_1(x), p_2(x), \ldots, p_k(x))$$

where $p_i(x)$ is the probability that a new firm chooses location i given the proportions that have already adopted it. The mapping p takes the k simplex S^k into itself.

[5] A Kirman, 'The Economy as an Interactive System', in Arthur, Durlauf and Lane, 1997, op.cit.

Arthur[6] proves that if the number of agents grows, if p is continuous and denoting by B the set of fixed points in p, then x_n converges with probability one to a point z in B[7]. In other words, the system settles down eventually to fixed proportions. In a Föllmer-Kirman[8] type process, in which the number of agents is fixed, the proportions vary continuously.

More recently, the topology of the networks which connect agents have been shown to be important, with qualitatively different results arising with different classes of network[9].

But these are essentially questions to be considered in more detail when an agent-based model is constructed to examine the question of firm location and the possibility of a different UK location becoming a focus for FBS firms currently in the City.

In summary, the key points are:

- A general equilibrium framework suggests that fixed preferences will favour co-location.
- Agglomerations can exist under constant returns of production, both internal and external to the firm, provided that Marshall-Hotelling conditions hold. Profit maximising firms will then prefer to locate in an agglomeration.
- Thinking of the issue in a game-theoretic context implies that it is likely that firms re-locating from the City will go to other agglomerations in Europe rather than elsewhere in the UK.
- The process by which such a constant returns agglomeration comes into being implies the existence of externalities in the specific sense that the probability of a new agent choosing a particular location will depend endogenously on the number of agents which have already chosen it. This sets up a self-reinforcing process, which is difficult to break out of i.e., it makes it likely that firms would choose another European agglomeration rather than another UK location.
- Once increasing returns in production are admitted, moving from one agglomeration to one which is at a much lower level of development becomes even less likely – unless the alternative is already near a 'critical mass' point.

[6] W B Arthur, 'Competing Technologies, Increasing Returns and Lock-in by Historical Events', *Economic Journal*, 1989

[7] provided that the equivalent deterministic system of x possesses a Lyapunov function ν whose motion is negative outside $B = \{x|p(x) = x\}$

[8] H Follmer, 'Random Economies with Many Interacting Agents', *Journal of Mathematical Economics*, 1974 and A Kirman, 'Ants, Rationality and Recruitment', *Quarterly Journal of Economics*, 1993

[9] S.Mossa, M Barthelemy, HE Stanley, and LAN Amaral, 'Truncation of power law behaviour in scale free networks due to information filtering' arXiv:cond-mat 0201421, 2002, is a good example

4. The persistence of agglomeration

To look at the extent to which agglomeration both exists and persists in London we do two things:

- Establish the extent to which the employment mix within London differs from that in the rest of the UK.
- Establish the extent to which this has changed over the last 20 years and if agglomeration has intensified or reduced.

For the analysis we use employment data from the Annual Business Inquiry[10] which details employment at the 60 sector level within each of the 459 local area districts[11] (LADs) that make up England, Scotland and Wales. This data is recorded at two time points, 1983 and 2001. We immediately discard three sectors that contain zero employment for all areas in both years, reducing the data to 57 sectors[12]. We also remove one LAD from the analysis that has outlying observations, The Isles of Scilly. In 2001 this LAD had a recorded total employment of 860, over three times less than the next lowest LAD.

Table 4.1 shows the proportions of total employment for some of the larger UK sectors, along with the change between 1983 and 2001. We can see that the largest single sector in 2001 was retail, which grew by almost a third over the period. Even more impressive growth can be seen in the other business sector, which grew by 82 per cent, becoming the second largest employment sector in 2001.

Other notable large increases over the period are in the computer sector, which almost quadrupled in size and real estate, which doubled. Proportionally, one of the larger falls here can be seen in agriculture, which dropped 88 per cent from 1.7 per cent of total employment in 1983 to 0.2 per cent in 2001. A similar table with all 57-employment sectors can be found in Appendix A. In this we can see that the largest proportion drop was in mining, which diminished in 18 years to just three per cent of its original 1983 proportional level.

[10] Annual Business Inquiry for 2001 data, Annual Employment Survey for 1983
[11] 1991 LAD definitions
[12] Mining of uranium and thorium ores, Private households with employed persons, Extra-territorial organisations/bodies

Table 4.1: Proportions of UK employment by selected sectors, 1983 and 2001

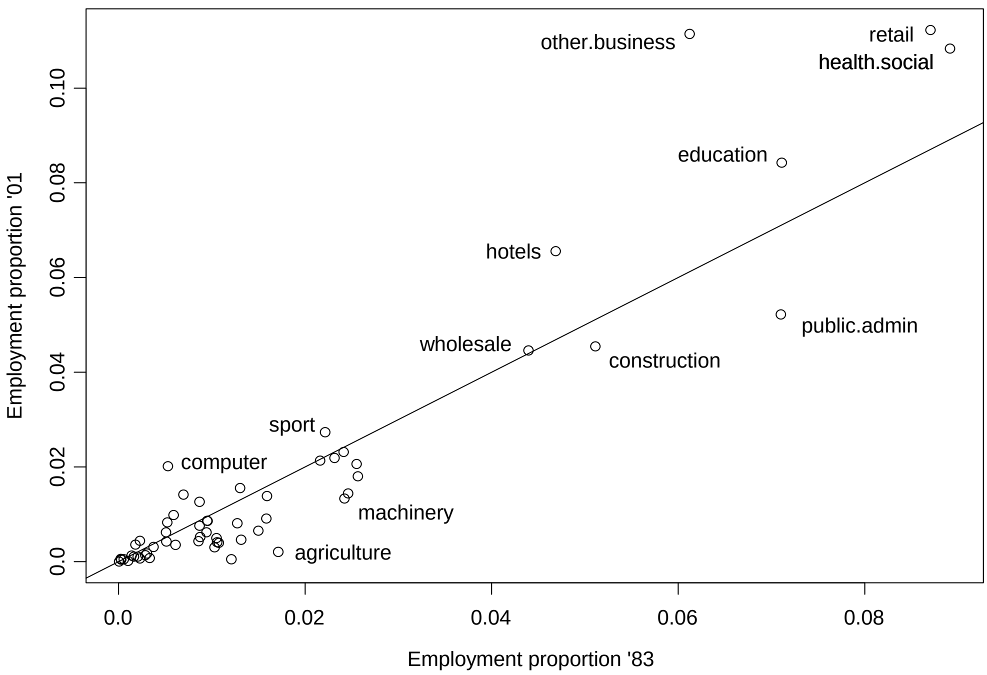
Sector	Percent of total employment		Percentage change
	1983	2001	
Retail	8.7	11.2	29
Health and social	8.9	10.8	22
Other business	6.1	11.1	82
Education	7.1	8.4	19
Public administration	7.1	5.2	-27
Hotels	4.7	6.6	40
Construction	5.1	4.5	-11
Wholesale	4.4	4.5	1
Sport	2.2	2.7	23
Land transport	2.6	2.1	-19
Financial intermediation	2.4	2.3	-4
Car sales etc	2.3	2.2	-6
Food	2.6	1.8	-30
Post	2.2	2.1	-1
Fabricated metals	2.5	1.4	-42
Machinery	2.4	1.3	-45
Printing	1.6	1.4	-13
Travel	1.3	1.6	19
Computer	0.5	2	280
Chemicals	1.6	0.9	-43
Other services	0.9	1.3	45
Other transport	1.5	0.7	-57
Real estate	0.7	1.4	103
Agriculture	1.7	0.2	-88

Note: Ordered by combined 1983 and 2001 proportions
Source: Annual Business Inquiry for 2001 data, Annual Employment Survey for 1983

Figure 4.1 plots the 1983 proportions against the 2001 proportions. Overlaid on the plot is the line of equality. Points above the line are sectors that have proportionally increased in size. Points below the line are sectors that have decreased proportionally in size. Other climbers that have been highlighted here include health and social work, education, hotels (including bars and restaurants) and sport. Large moves that join agriculture below the line include machinery, construction and public administration.

It is interesting to note at this point that the employment appears to have partially shifted from many small sectors to a few large sectors. Over the period 18 of the 57 sectors grew proportionally in size, while 39 of the 57 dropped. In part, this reflects the construction of the Standard Industrial Classification, which has far more detail in the manufacturing part of the classification than the services part. Since services has grown while manufacturing has declined this produces the statistical result. More detailed data on services would give us a more rewarding pattern.

Figure 4.1: Distribution of employment by sector, 1983 against 2001



Note: straight line indicates equality between 1983 and 2001
Source: Annual Business Inquiry for 2001 data, Annual Employment Survey for 1983

In order to find out how similar a specific LAD's employment is to the UK average we take two steps. Clearly, the total level of employment varies enormously between LADs, so the first step is to transform each of the recorded employment levels within each sector into a proportion of the total employment within the LAD. Using proportions instead of absolute levels allows us to compare different LADs on a like-for-like basis. Similarly, we also calculate the proportions of employment for the whole of the UK. This process is repeated for both time points being studied.

We measure the difference between the LADs' distributions of employment and the UK national average distributions of employment using two methods. The first is to perform what is known as a Chi-squared test. Using this test we can find the probability that an individual LAD's employment is similar to that of the national average. More precisely, we find the probability of observing the given employment levels within the LAD given that the employment has been drawn from the same distribution as the UK average.

The Chi-squared test was executed with all the data. The results of this however proved not to be useful for one simple reason. Nowhere in the UK has a distribution of employment that is the same as that of the UK, or in fact even close to it. In other words, there is no real place that approximates to the average. The p-values returned from the test were all identically equal to zero. Although interesting in itself, this does not help us to distinguish the degree to which different areas differ from the average. We therefore proceed to analyse the degree of similarity in a more classical way.

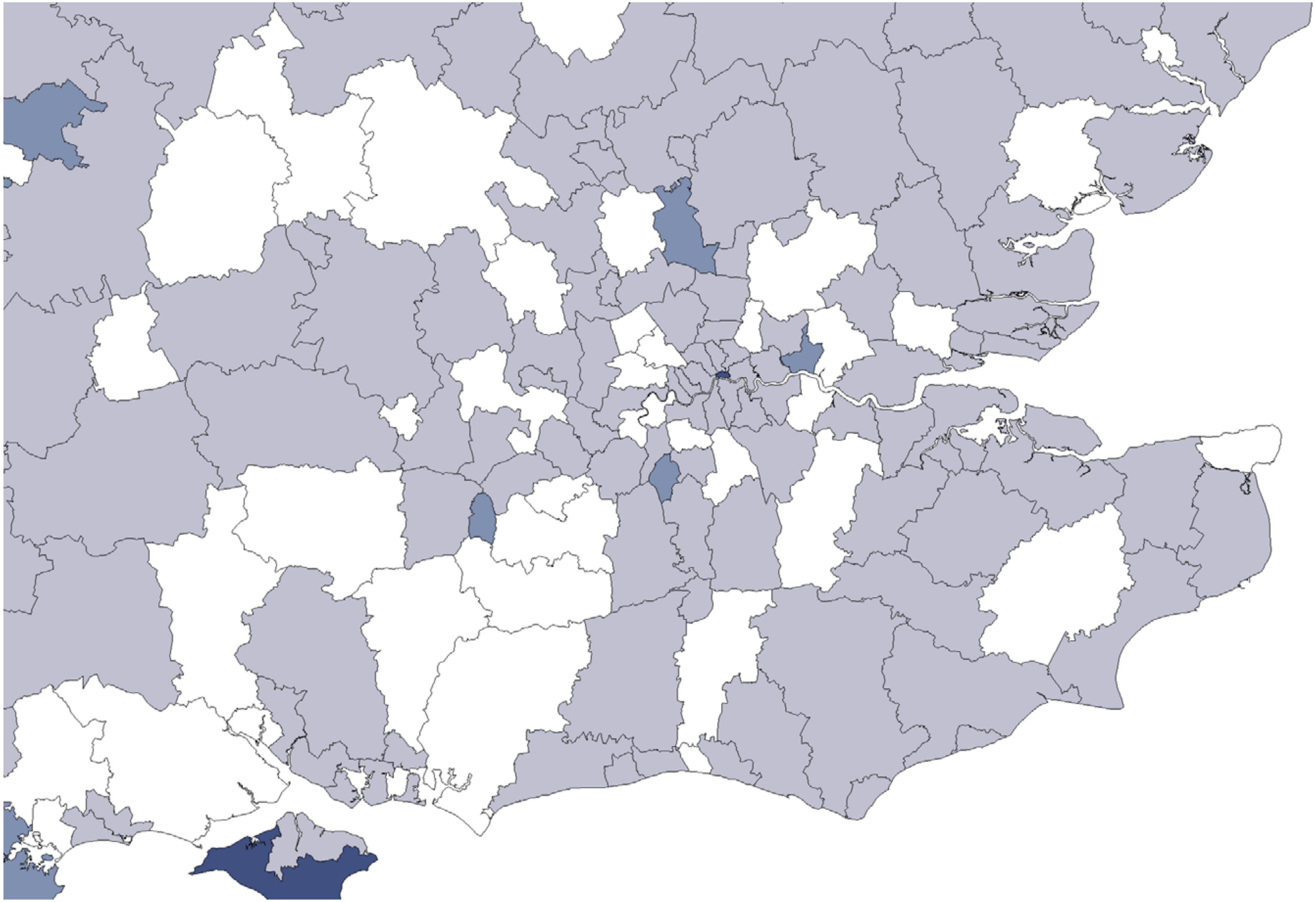
The Euclidean distance measure takes two vectors and calculates what is known as a distance between them. The closer the vectors are together, the lower the distance. If they are identically equal then the distance between them will be equal to zero. In a three dimensional space, where each element of the vector represents a coordinate in space, the distance between the vectors is actually equal to the physical distance between the two points that the vectors represent. Appendix B shows the calculations required.

The advantage of this measure for us is that it is simple to calculate and gives a fair weighting to all employment sectors. The disadvantage is that you cannot evaluate a probability on the basis of its value. This is not a problem however, as we have already discovered that all the LADs are highly significantly different from the UK average. We therefore proceed with this definition as our measure of similarity/dissimilarity.

Mapped out in figures 4.2 and 4.3 are the employment distances from the UK average in 1983 and 2001 respectively. These maps have been zoomed in to the South East of England. Looking firstly at the 1983 map we can see several things. Broadly speaking, the distribution of distances in this area of England is relatively random, with the possible exception of the central London boroughs, which are all within the 0.1 to 0.2 band. The City has a higher distance in 1983 of 0.33, the third highest in the country at the time.

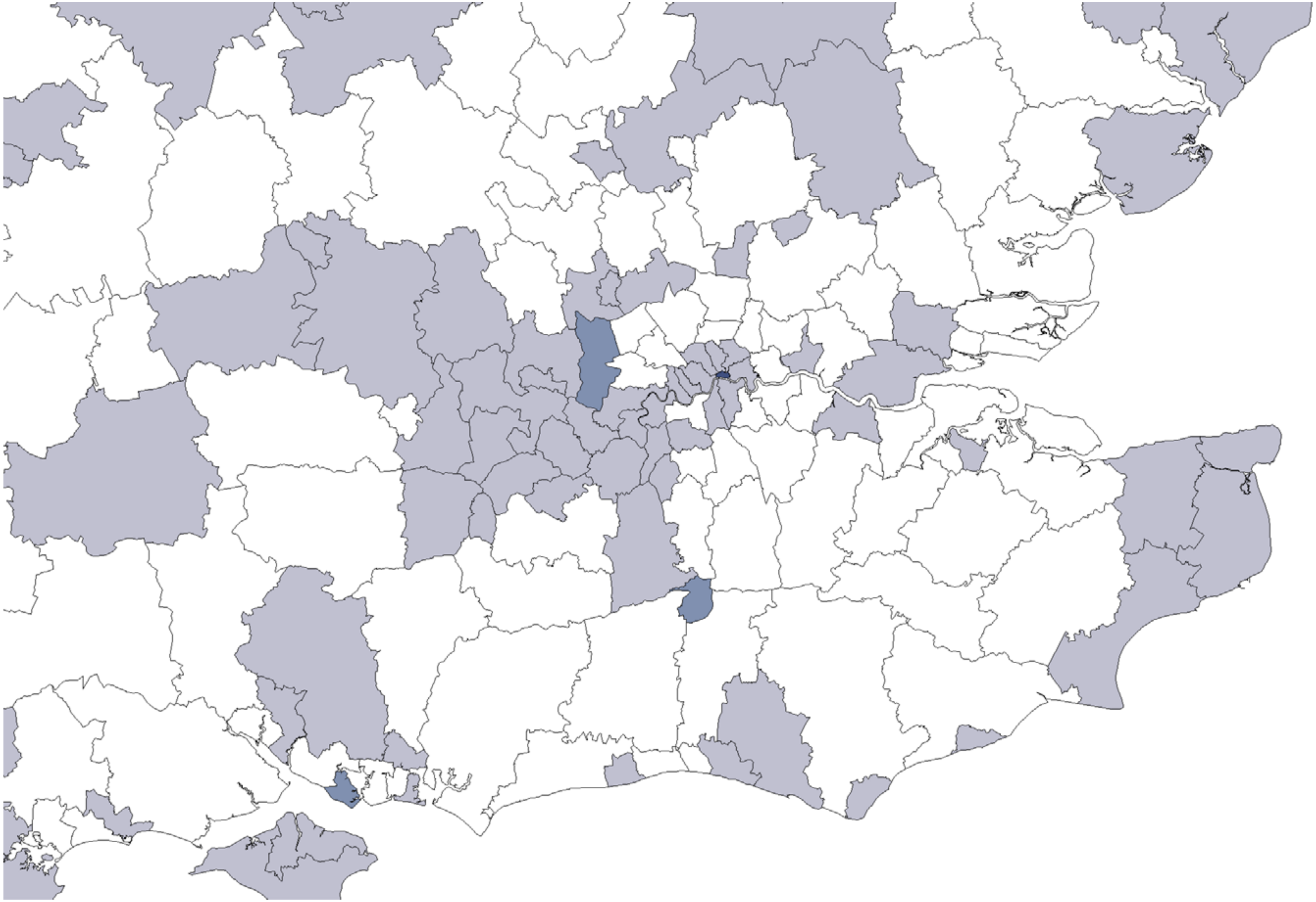
When we move forward in time to 2001 we see that many of the LADs in the South East have moved closer to the average, dropping out of the 0.1 to 0.2 band into the 0 to 0.1 band. Central London however does not drop into this band, along with a belt of LADs stretching West of London towards Oxford. The City distance increases from 0.33 to 0.38, making it the furthest from the UK average in 2001.

Figure 4.2: Distance from UK average distribution of employment, 1983



Source: Annual Employment Survey 1983

Figure 4.3: Distance from UK average distribution of employment, 2001



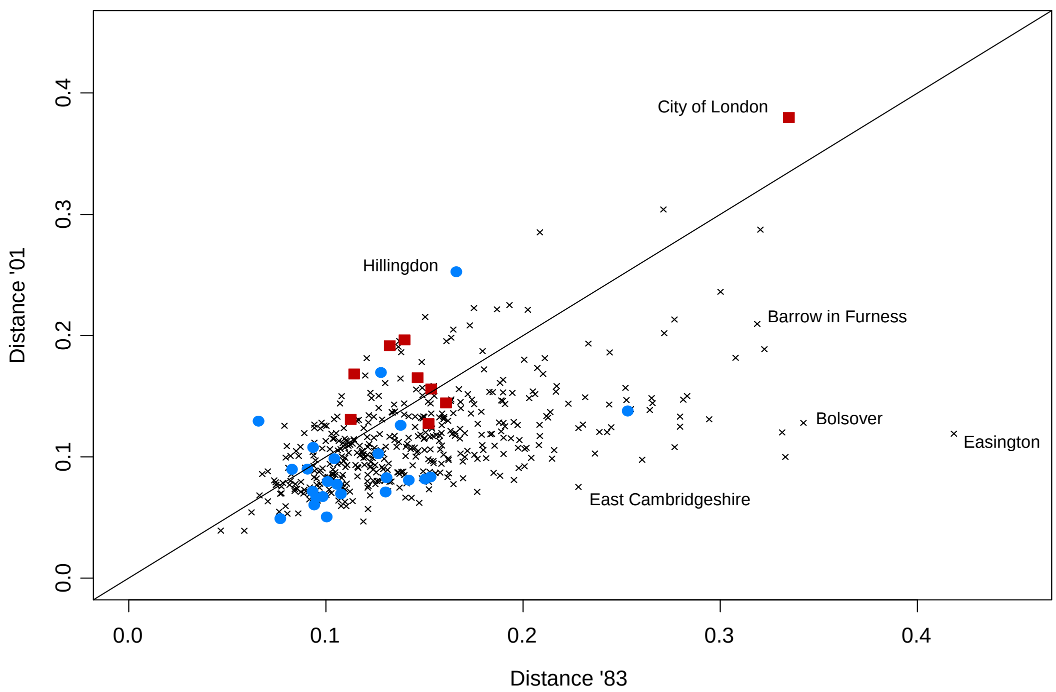
Source: Annual Business Enquiry 2001

We can examine the difference between 1983 and 2001 more closely by plotting the distances in these years against each other as shown in Figure 4.4. Overlaid on this plot is the line of equality again. This time points above the line represent LADs that have moved further away from the average, while points below the line represent LADs that have moved closer to the average.

We can see that in 2001 the City had the largest distance by a large margin, having increased significantly since 1983. When we examine the absolute levels of employment in the City we find that the majority of this change came about though a large proportional increase in the level employment in the other business sector. Similarly, Hillingdon, containing Heathrow airport, saw a shift away from the average, primarily due to the increase in air transport employment.

Points above the line, moving away from the average, are however in the minority, as many LADs move closer to the average distribution of employment in 2001. Good examples of such LADs are Bolsover and Easington, who saw a collapse in the coal mining industry, reducing their 'uniqueness'. Barrow in Furness lost much of its employment in the other transport sector, namely the shipping industry and East Cambridgeshire saw a large drop in agricultural employment.

Figure 4.4: Distance of employment distribution from UK average, 1983 against 2001



Note: straight line indicates equality between 1983 and 2001. Blue circles = outer London boroughs, red squares = central London boroughs

Source: Annual Business Inquiry for 2001 data, Annual Employment Survey for 1983

Counting the number of points above and below the line in Figure 4.4 we see that 100 of the 458 LADs (22 per cent) increased their distance from the average, while 358 (78 per cent) moved closer to the average. The coloured points of the plot represent the distances of the 33 London boroughs. The blue coloured circles correspond to the outer London boroughs and the red squares correspond to central London[13].

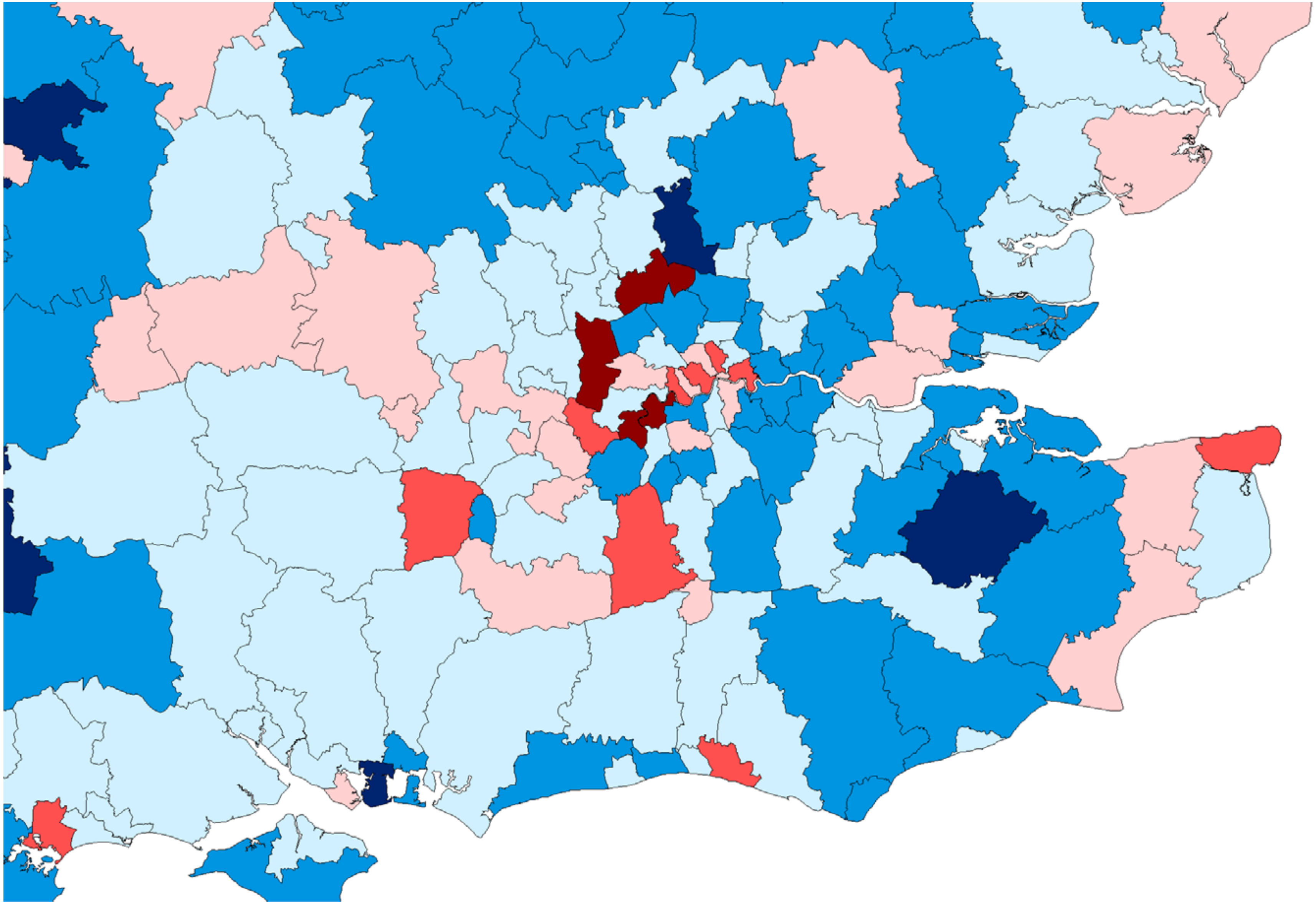
We can see here that only two of the nine central London boroughs saw a decrease in their distance, with seven moving further away. Conversely in outer London only five of the 24 boroughs moved away from the average, while 19 moved towards the average. The ratio of those increasing to those decreasing in the outer London boroughs matches the ratio of the whole dataset pretty closely. This means that outer London looks like the country as a whole – it is central London that is different.

In figure 4.5 we have mapped the percentage change in distance from the UK average between 1983 and 2001 for the South East. LADs that have moved away from the average have been coloured red, with those moving towards the average coloured blue. Darker colours indicate more extreme moves.

We can see that with the exception of Lambeth and Hackney, the distances from the average have increased for the central London boroughs. Outside central London the largest increases can be seen in Richmond, Hillingdon and Hertsmere. The majority of the remaining LADs in the South East have become more similar to the UK average distribution of employment.

[13] Nine central boroughs defined to be Camden, Hackney, Islington, Kensington and Chelsea, Lambeth, Southwark, Tower Hamlets, Westminster and the City of London

Figure 4.5: Percentage change in distance from UK average distribution of employment between 1983 and 2001



Source: Annual Business Enquiry 2001

Table 4.2: Change in distances from UK average of London boroughs

Borough	Distance from UK average		Percentage change
	1983	2001	
Barking and Dagenham	0.253	0.138	-46
Barnet	0.108	0.069	-36
Bexley	0.077	0.049	-36
Brent	0.091	0.090	-1
Bromley	0.106	0.077	-27
Camden	**0.147**	**0.165**	**13**
City of London	**0.335**	**0.380**	**13**
Croydon	0.093	0.072	-23
Ealing	0.083	0.090	8
Enfield	0.100	0.050	-50
Greenwich	0.142	0.081	-43
Hackney	**0.161**	**0.144**	**-10**
Hammersmith and Fulham	0.128	0.169	33
Haringey	0.101	0.079	-21
Harrow	0.098	0.067	-32
Havering	0.094	0.060	-36
Hillingdon	0.166	0.253	52
Hounslow	0.138	0.126	-9
Islington	**0.114**	**0.168**	**47**
Kensington and Chelsea	**0.153**	**0.156**	**2**
Kingston upon Thames	0.127	0.103	-19
Lambeth	**0.152**	**0.127**	**-16**
Lewisham	0.150	0.082	-46
Merton	0.094	0.108	15
Newham	0.130	0.071	-46
Redbridge	0.104	0.098	-6
Richmond upon Thames	0.066	0.129	96
Southwark	0.113	0.131	16
Sutton	0.131	0.083	-37
Tower Hamlets	**0.132**	**0.191**	**45**
Waltham Forest	0.095	0.067	-30
Wandsworth	0.153	0.083	-46
Westminster, City of	**0.140**	**0.196**	**40**
Greater London average	0.130	0.120	-8
Central London average	**0.161**	**0.184**	**15**
Outer London average	0.118	0.096	-19

Note: central London boroughs highlighted in bold
Source: Annual Business Inquiry for 2001 data, Annual Employment Survey for 1983

In order to concrete the inner/outer London agglomeration split seen in Figure 4.4 we present in Table 4.2 the complete list of distances and percentage changes for the London boroughs. Those boroughs making up central London are highlighted in bold.

At the bottom of the table we can see that London as a whole has on average moved towards the national average employment distribution over the period. When we split London into central and outer however a different picture emerges. Outer London has indeed seen a marked move towards the average, with the mean distance dropping by 19 per cent. Central London, on the other hand, has moved away from the average with the mean distance increasing by 15 per cent. Of the central boroughs Islington, Tower Hamlets and Westminster all saw large shifts away, increasing their distance by over 40 per cent.

The conclusions that can be drawn from this analysis are as follows:

- Between 1983 and 2001 employment has partially shifted away from many small industries (e.g. manufacturing) to a few large industries, such as other business, retail and hotel and catering.
- Over this period, the majority of local area districts in the UK have changed their distribution of employment to better represent the national average. Nationwide, uniqueness is becoming more rare.
- This shift of employment towards the average is also true in the South East of England and the outer boroughs of London.
- Central London, however, has seen a significant shift away from the national average distribution of employment, particularly in the boroughs of Islington, Tower Hamlets and Westminster.
- The City of London has also managed to move away from the average, despite already being the third furthest away in 1983. In 2001 it was the most unique LAD in the UK in terms of its employment characteristics.
- These consistently high and increasing distances in central London over this long period demonstrate a persistence to move away from the average distribution of employment.

5. Agglomeration and productivity

This section looks at estimating the extent to which agglomeration improves productivity. As earlier sections have shown, there are mechanisms which suggest that proximity confers advantages which outweigh the additional costs of commuting, land and so on. But it is also important to try to estimate how large these effects are. Productivity gains for businesses will be observed through two primary agglomeration sources. As more companies provide the same services in the area, competition increases, forcing increased productivity. In addition, as the distance and time between companies decreases, less time is wasted moving products and people around, again increasing productivity.

We quantify the benefits of employment agglomeration by breaking down Great Britain geographically and comparing the regional density of employment with average earnings in the same areas (where average earnings are taken as a proxy for output/productivity). We take data from two years, 2001 and 1989, along with the changes between these two years. We would expect to find that areas with different industrial employment structures would have different productivity, and therefore also different average earnings. We therefore condition on the regional differences in employment structure by calculating expected earnings for all areas.

We find strong evidence in this initial study for the existence of a positive, non-linear relationship between earnings differentials and employment density. There is evidence that the relationship between the two has strengthened over time at the higher levels of employment density.

Three types of data are required for the analysis:

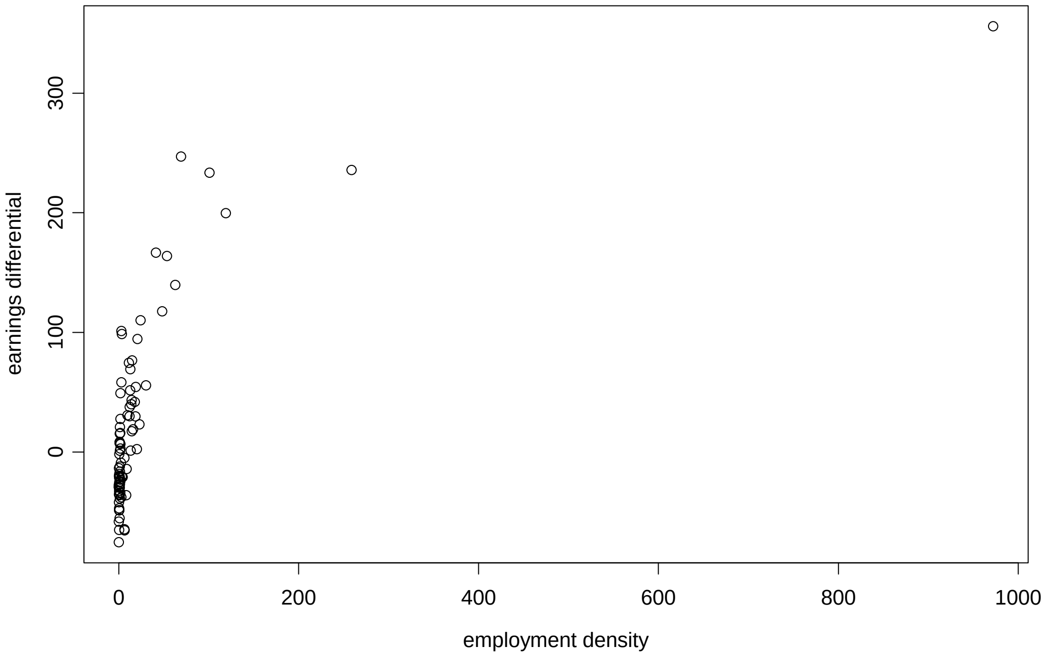
- average earnings data, at a regional level and national level broken down by sector
- employment data, broken down by region and sector
- physical land area data, at a regional level.

With the data available to us, our study area consists of England and Scotland, broken down to the geographical level consisting of counties, metropolitan counties and the boroughs of London. We break down employment to the 2-digit SIC92 level for the 2001 data, and the 2-digit SIC80 level for the 1989 data.

The full analysis data-set consists of 87 different geographic areas with 56 employment industries in 2001 and 60 employment industries in 1989. Further details on the construction of this consistent data-set can be found in Appendix C.

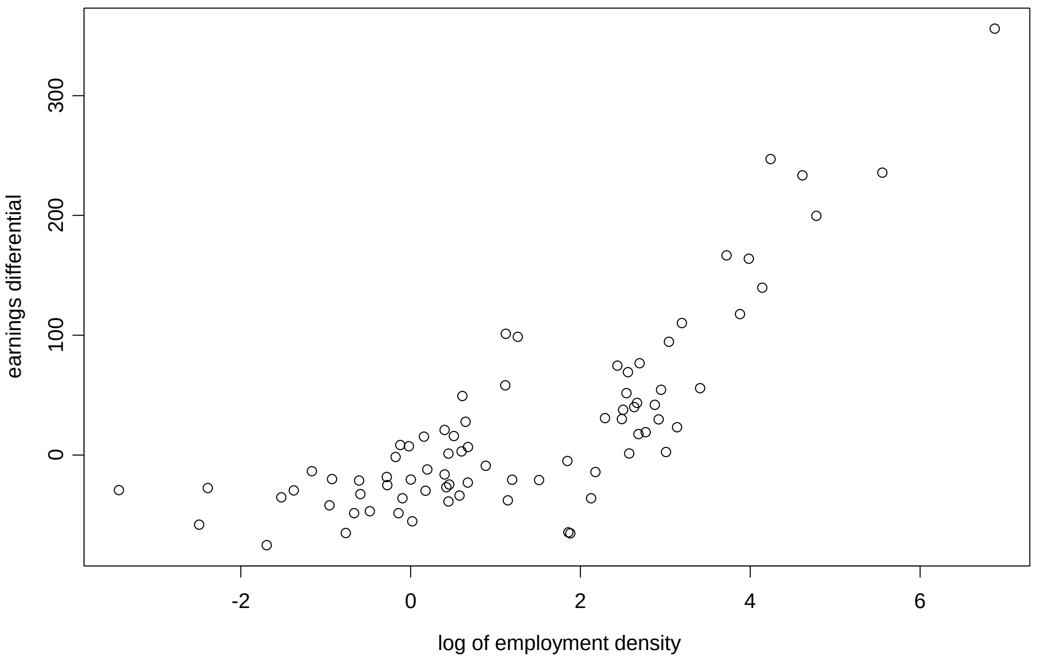
Initial inspection of the data suggests that there is a strong relationship between the earnings differential and employment density. Further, it is highly non-linear. Figure 5.1 plots the basic data for 2001.

Figure 5.1: Earnings differential and employment density, 87 GB areas, 2001



The observation at the far right hand top of the chart is the City of London. It appears to be an outlier. However, a simple transformation of the employment density variable into log form produces a more tractable graph.

Figure 5.2: Earnings differential and log of employment density, 87 GB areas, 2001



The simple correlation between the raw variables is 0.745, falling to 0.646 when the City is omitted. With the employment density variable in log form, however, the City is much less obviously an outlier, and the two correlations are 0.793 and 0.766 respectively.

A standard least squares fit between the earnings differential and the log of employment density sets a benchmark for the regression work:

Table 5.1: Linear regression of the earnings differential on the log of employment density, 2001 data

```
Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept) -18.1611   6.6493    -2.7313   0.0078
 log(emp01)  33.0982   2.8787    11.4975   0.0000

Residual standard error: 49.46 on 78 degrees of freedom
Multiple R-Squared: 0.6289
```

However, not surprisingly from Figure 5.2, a Ramsey test for specification shows a distinct degree of non-linearity. The Ramsey test augments the specification in Table 5.1 by adding powers of the fitted values of this regression. A standard procedure is to include the square, cube and fourth power, and these are very significantly different from zero.

We searched for an appropriate non-linear specification using the general technique of local regression[14]. A standard simple linear regression model takes the form:

$$y = \alpha + \beta X + \varepsilon$$

In local regression, we build the smooth function s(X) as follows. Take a point, say X_0. Find the k nearest neighbours of X_0, which constitutes a neighbourhood $N(X_0)$. The number of neighbours, k, is specified as a percentage of the total number of points, called the span.

Calculate the largest distance between X_0 and another point in $N(X_0)$:

$$\Delta(X_0) = \max_{N(X_0)} |X_0 - X_i|$$

Assign weights to each point in $N(X_0)$ using the tri-cube function:

$$W\left(\frac{|X_0 - X_i|}{\Delta(X_0)}\right)$$

where

$$W(u) = \left\{(1 - u^3)^3\right\} \text{ for } 0 \le u \le 1 \text{ and } W(u) = 0 \text{ otherwise}$$

[14] S-Plus 2000, *Guide to Statistics, vol.1*, Data Analysis Products Division, MathSoft Inc., Seattle

Calculate the weighted least squares fit of y on the neighbourhood $N(X_0)$. Take the fitted value $\hat{y}_0 = s(X_0)$. This can then be repeated for each of the predictor values. With a simple linear regression we have two parameters in the model, the coefficient of the explanatory variable and the constant term. More generally in a polynomial regression we have (n+1) parameters, where n is degree of the polynomial being fitted. For example, with quadratic regression (of degree 2) we have three parameters: the same two coefficients as the linear model with a further coefficient for the squared term. When we move to a local regression, the number of parameters is not fixed, as the coefficients vary for each area of the curve.

A distinct advantage of local regression is that, as the span is increased, the approach converges on the classical linear regression. We can therefore use analysis of variance to investigate the appropriate degree of non-linearity.

The formal approach to testing for non-linearity is as follows. We start by constructing a local regression with a high span value, which can be taken to be approximately linear. We then construct a second local regression with a marginal increase to the non-linearity, by decreasing the span value. Using an Analysis of Variance test, commonly known as an ANOVA test, we can then see if there has been any measurable gain from the decreased span. We continue to decrease the span and perform ANOVA tests until we cannot significantly improve the model.

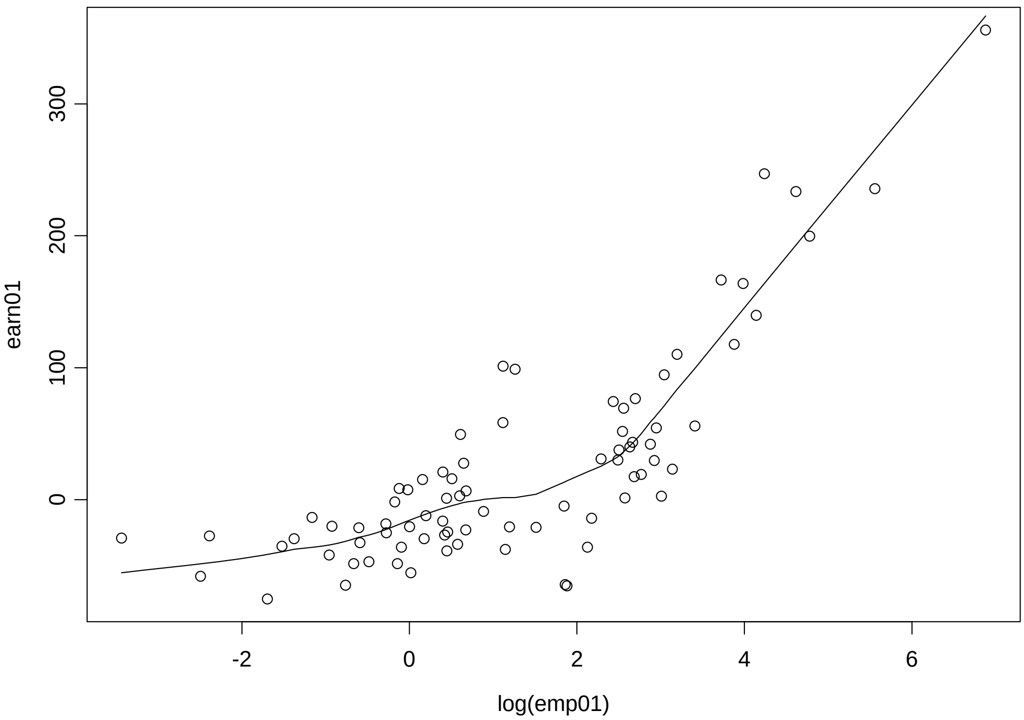
One of the output characteristics associated with local-regressions is known as the effective number of parameters. This value allows us to gauge the degree of fitting that is being allowed by the process. With a high span value, and an approximately linear fit we find the effective number of parameters approaching 1. As we decrease the span, the effective number of parameters rises.

Considering now the earnings differential data, we look for non-linearity and reduce the span by 0.1 at a time, in each case carrying out the ANOVA test. The effective number of parameters with a span of 0.9 is 2.8, with 0.8 it is 3.3, 0.7 is 3.6, 0.6 is 4.0 and 0.5 is 4.8.

The null hypothesis that the residual sum of squares are identical was only rejected at p = 0.11 for spans of 0.6 and 0.5, so we choose a span of 0.6 as our preferred model. The effective number of parameters being 4.0, this suggests that a relationship of the same order as a cubic should give a good approximation to the relationship between the earnings differential and the log of employment density.

Figure 5.3 plots the fitted value of the local regression with span = 0.6.

Figure 5.3: **Earnings differential and log of employment density, 87 GB areas, 2001.**



Note: The solid line is the fitted values from the general non-linear local regression technique, with span = 0.6

An important practical feature of the result is the way in which the slope of the fitted line increases beyond a critical point. There is a positive relationship between earnings differentials and employment density even at relatively low levels of density. But the relationship intensifies beyond a critical point. The observations to the right of this point in Figure 5.3 are Greater Manchester, Tyne and Wear, West Midlands, and 27 London boroughs – in other words, dense urban areas.

We carried out additional specification tests on the residuals of this preferred regression. The null hypothesis that the residuals are normally distributed is only rejected on the general Kolmogorov-Smirnov test at $p = 0.437$. The simple correlation between the residuals and fitted values is –0.003, and between the square root of the absolute value of the residuals and the fitted values is 0.110. Neither of these are significantly different from zero at the conventional level of significance of $p = 0.05$, so there is no evidence of heteroskedasticity in the residuals.

The results for the 1989 data are qualitatively very similar. The simple regression between earnings differential and the log of employment density is set out in Table 5.2.

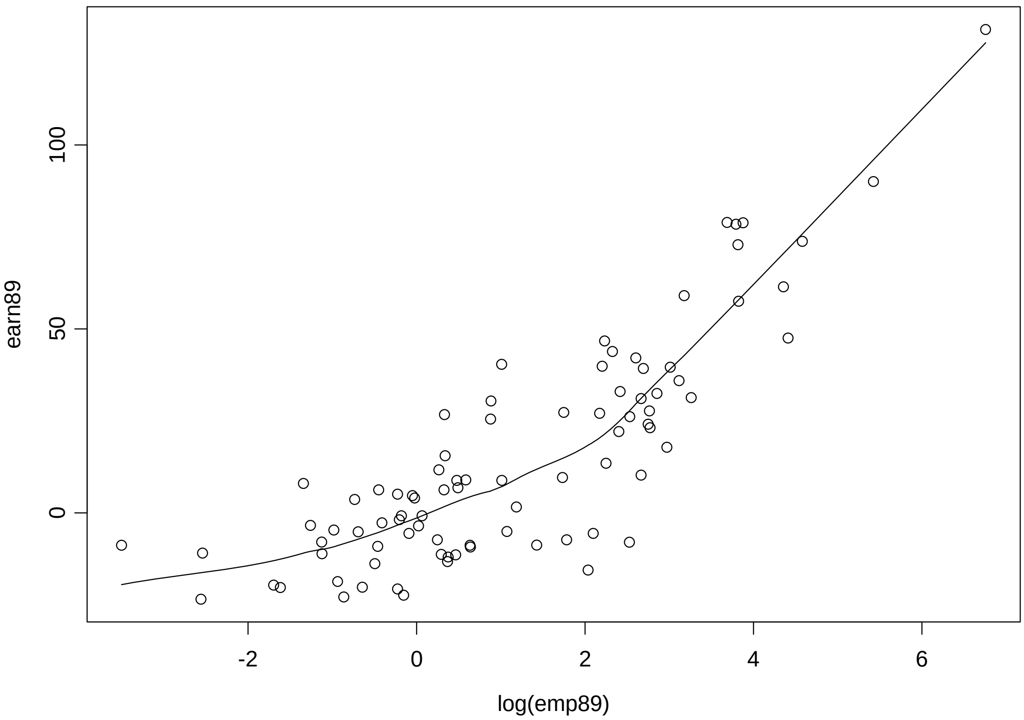
Table 5.2: Linear regression of the earnings differential on the log of employment density, 1989 data

```
Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  -0.2558   2.1138    -0.1210   0.9040
 log(emp89)  13.2030   0.9461    13.9547   0.0000

Residual standard error: 16.84 on 85 degrees of freedom
Multiple R-Squared: 0.6961
```

Using the local regression approach, again a span of 0.6 is indicated, and the fitted values from this are plotted in Figure 5.4.

Figure 5.4: Earnings differential and log of employment density, 87 GB areas, 1989.



Note: The solid line is the fitted values from the general non-linear local regression technique, with span = 0.6

We again observe the strengthening of the relationship beyond a critical point, and the observations to the right of it are Greater Manchester, Tyne and Wear, West Midlands, and almost all the London boroughs.

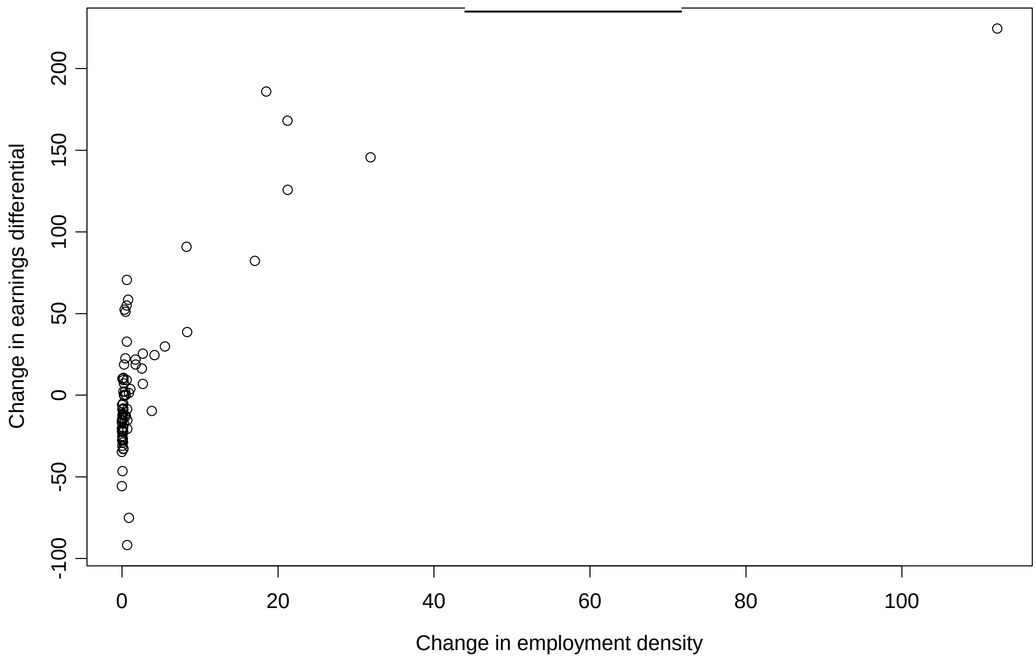
The equation is again well specified. The null hypothesis of normality of the residuals is only rejected at p = 0.50. The correlation between the residuals and the fitted values and the square root of the absolute values of the residuals and the fitted values is 0.002 and 0.060 respectively.

The data for both 1989 and 2001 compare areas at a point in time. Given that we are working out of necessity at the 2-digit SIC level, it is possible that the strong relationships we have identified might arise in part from the fact that industries at this level of disaggregation might not be strictly comparable across areas. In other words, each 2-digit industry is made up of different types of economic activity. These obviously have a lot in common, given that they share the same 2-digit code. But two authorities could have the same proportion of their total employment in a given 2-digit industry, but each could have a different mix of the component parts.

We therefore examined the relationship between the changes in the earnings differential and employment density between 1989 and 2001. Within each area, the mix of component industries within each 2-digit industry might plausibly be thought to be reasonably stable over time.

The results are in fact very similar to those of the individual years. Figure 5.5 plots the basic data.

Figure 5.5: Changes in earnings differential and employment density, 87 GB areas, 1989 – 2001



In terms of estimating the relationship between the two, four areas in the data sample experienced a fall in employment density between 1989 and 2001,so they were omitted in order to take the log of employment density.

The simple regression is given in Table 5.3.

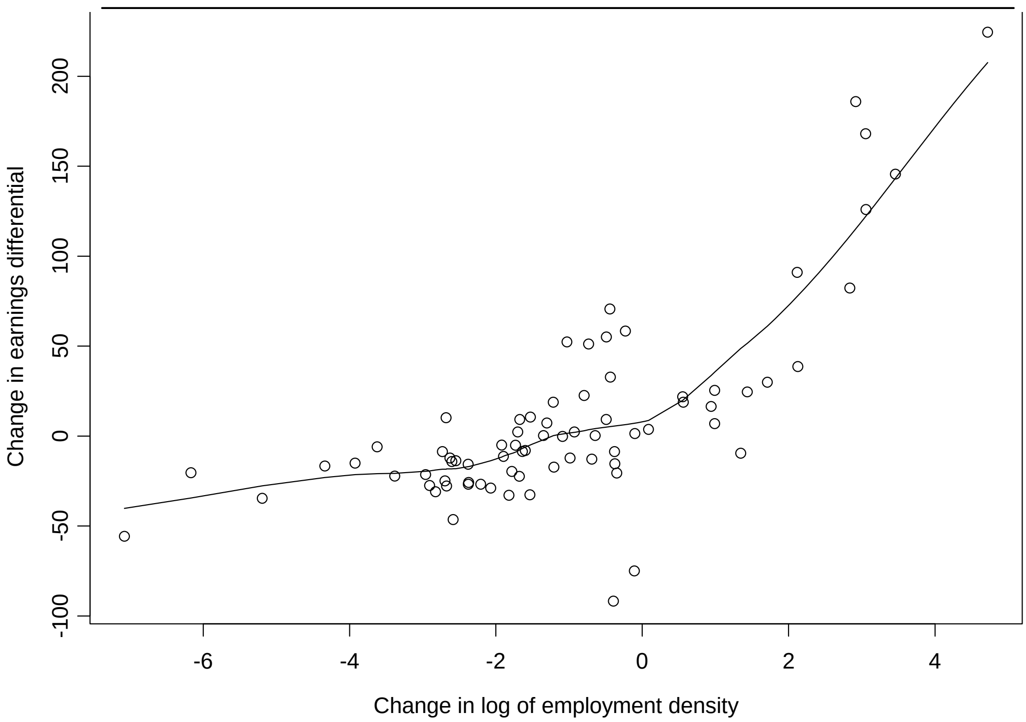
Table 5.3: Linear regression of the change in earnings differential on the log of the change in employment density, 2001 data

```
Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) 28.5849  4.7809     5.9789  0.0000
  log(empc) 18.5692  2.0231     9.1788  0.0000

Residual standard error: 37.15 on 73 degrees of freedom
Multiple R-Squared: 0.5358
```

Application of the local regression procedure suggested a specification again with a span of 0.6 and an effective number of parameters of 4.3.

Figure 5.6: **Changes in earnings differential and log of employment density, 87 GB areas, 1989 - 2001.**



Note: The solid line is the fitted values from the general non-linear local regression technique, with span = 0.6

The residuals are once again homoskedastic, the correlation between the residuals and the fitted values and the square root of the absolute value of the residuals and the fitted values being 0.038 and 0.169 respectively. The latter is not statistically significant from zero at $p = 0.05$. There is, however, evidence of non-normality of the residuals, with the null hypothesis of normality being rejected at $p = 0.012$

The existence of a positive, non-linear relationship between earnings differentials and employment density therefore appears to well established.

There is evidence that the relationship between the two has strengthened over time at the higher levels of employment density. In 1989, the mean value of employment density for the areas in the sample is 22.80 with a maximum of 859.9 in the City of London. There are small increases in these in the 2001 data, to 26.97 and 972.1 respectively. However, beyond the critical value of employment density identified in Figures 5.3 and 5.4, the slope of the regression between the two variables has increased between the two years. Taking a linear approximation between earnings differentials and the log of employment density for all observations where the log of density is greater than 2, we obtain an R^2 of 0.714 in the 1989 sample and 0.861 in the 2001 sample. But the coefficient on the log of employment density rises from 24.6 to 78.7, both of which are significantly different from zero at $p = 0.0000$. In other words, the strength of the agglomeration effect seems to have intensified between 1989 and 2001.

The included work establishes strong initial evidence for the existence of an agglomeration effect in the UK. Further, this effect is non-linear. Output per head initially rises linearly with the log of employment density, but beyond a critical value this relationship breaks down and the gradient of this relationship increases.

At any rate, there is so far no evidence that allows an estimate of the density at which agglomeration effects would tail off and crowding costs would outweigh such benefits.

Other work has also recently looked at these issues and estimated agglomeration effects. Different data and attempting to control for different variables has added to the robustness of the conclusion that these effects are important. One study[15] used a different definition of density and individual firm data, while another[16] has controlled more effectively for education differences.

Further work would still be useful to use more disaggregated data and to condition more effectively for spatial effects of various kinds. Effective density, where all areas contribute somewhat to the density is one such possibility but accounting more directly for local spatial effects would also be useful.

However, all the work conducted agrees that these effects are both significant and persistent and need to be even better understood.

[15] Dan Graham
[16] Venables and Rice

Appendix A

Figure A.1: Proportions of UK employment by sector, 1983 and 2001

Sector	Percentage of total employment		Percentage change
	1983	2001	
Agriculture	1.7	0.2	-88
Forestry	0.1	0	-15
Fishing	0	0	34
Coal	1.2	0	-96
Petroleum	0.1	0.1	-11
Mining	0	0	-97
Quarrying	0.2	0.1	-49
Food	2.6	1.8	-30
Tobacco	0.1	0	-87
Textiles	1.3	0.5	-65
Apparel	1	0.3	-71
Tanning	0.3	0.1	-78
Wood	0.4	0.3	-17
Paper	0.6	0.4	-43
Printing	1.6	1.4	-13
Coke	0.2	0.1	-38
Chemicals	1.6	0.9	-43
Rubber	1	0.9	-10
Non-metallic	1.1	0.5	-53
Metals	1.1	0.4	-63
Fabricated metals	2.5	1.4	-42
Machinery	2.4	1.3	-45
Office machinery	0.3	0.2	-44
Electrical machinery	0.9	0.6	-34
Radio	0.9	0.4	-50
Medical	0.9	0.5	-41
Motor vehicles	1.3	0.8	-36
Other transport	1.5	0.7	-57
Furniture	0.9	0.8	-13
Recycling	0	0.1	109
Electricity	1.1	0.4	-61
Water	0.3	0.1	-53
Construction	5.1	4.5	-11
Car sales.etc	2.3	2.2	-6
Wholesale	4.4	4.5	1
Retail	8.7	11.2	29
Hotels	4.7	6.6	40

Land transport	2.6	2.1	-19
Water transport	0.2	0.1	-72
Air transport	0.2	0.4	97
Travel	1.3	1.6	19
Post	2.2	2.1	-1
Financial intermediation	2.4	2.3	-4
Insurance	1	0.9	-11
Financial auxilliary	0.6	1	66
Real estate	0.7	1.4	103
Machinery renting	0.5	0.6	21
Computer	0.5	2	280
Research	0.5	0.4	-18
Other business	6.1	11.1	82
Public admin	7.1	5.2	-27
Education	7.1	8.4	19
Health social	8.9	10.8	22
Refuse	0.2	0.4	91
Miscellaneous	0.5	0.8	58
Sport	2.2	2.7	23
Other service	0.9	1.3	45

Appendix B: Distance definition

Starting with the employment data we construct here the definition of Euclidean distance.

Let e_{ij} = employment of LAD i in sector j i = 1, …, 458
j = 1, …, 57

The proportion employment in each LAD is defined as:

$$p_{ij} = \frac{e_{ij}}{\sum_{j=1}^{57} e_{ij}}$$ i = 1, …, 458
j = 1, …, 57

The national proportions of employment are:

$$P_j = \frac{\sum_{i=1}^{458} e_{ij}}{\sum_{i=1}^{458}\sum_{j=1}^{57} e_{ij}}$$ j = 1, …, 57

Define the Euclidean distance between the employment proportions of LAD i and the national average as being:

$$d_i = \sqrt{\sum_{j=1}^{57}\left(p_{ij} - P_j\right)^2}$$ i = 1, …, 458

Appendix C: Analysis data description

In order to construct a suitable data-set to be analysed for this project, original data had to be manipulated in order to maintain consistency across both study years, 1989 and 2001. More specifically, the breakdown of geographic areas had to be an exact match. Given the changes in commonly used area definitions between the two periods, along with the inconsistencies in the area definitions used by different institutions to record data, this was not a simple exercise.

The area definition used is the lowest common denominator in terms of the level of disaggregation, breaking down England and Scotland into 87 areas. This is based upon the earnings data provided by the ONS for 1989, which uses a combination of counties, metropolitan counties and London boroughs. No consistent tally between the area definitions used for Wales in the two years could be found. Consequently, Wales was excluded from the analysis. The original data used was as follows.

Average earnings data
From the New Earnings Survey, 1989 and 2001, provided by the Office for National Statistics. Four data-sets used:

1989 Regional gross average weekly earnings. GB broken down into 97 counties, metropolitan counties and London boroughs

1989 National gross average weekly earning, broken down into 60, 2-digit SIC80 employment sectors

2001 Regional gross average weekly earnings. GB broken down into 173 unitary authorities and metropolitan counties

2001 National gross average weekly earnings, broken down into 56 2-digit SIC92 employment sectors

Employment data
From the Census of Employment (1989) and the Annual Business Inquiry (2001), provided by NOMIS online. Four data-sets used:

1989 GB employment broken down by county and 2-digit SIC80 employment sectors, providing data-set of size 66 x 60

1989 London employment broken down by local area district definitions (boroughs) and 2-digit SIC80 employment sectors, providing data-set of size 33 x 60

2001 GB employment broken down by county and 2-digit SIC92 employment sectors, providing data-set of size 66 x 56

2001 London employment broken down by local area district definitions (boroughs) and 2-digit SIC92 employment sectors, providing data-set of size 33 x 56

Supplementary employment data[17]

2001 GB employment broken down by unitary authority and 2-digit SIC92 employment sectors, providing data-set of size 203 x 56

Land area data

Land areas provided by the Office of National Statistics, breaking down the GB to the unitary authority and county level, along with ONS London borough land areas.

We start with the 97 GB areas in the 1989 regional earnings data. Unable to tie-up the Welsh data between 1989 and 2001 we exclude Wales, which consists of nine areas. We also exclude the Scottish islands, which are recorded as a single observation. This leaves 87 areas for the area definition to be used.

Excluding the 22 Welsh observations and three Scottish islands' observations[18] from the 2001 earnings data leaves 148 observations. These can be directly tied up with the 87 areas from the 1989 earnings data.

The four employment data-sets are combined to form two data-sets covering the 87 areas. Of the 66 areas in the county employment data-sets we exclude eight Welsh records, the three Scottish island records and one for Greater London. This leaves 54 records, which when combined with the 33 recordings from the London data-sets brings us up to the predefined 87 areas.

Combining area and employment data is simply a matter or adding the observations. In order to combine the 2001 average earnings data we need to scale by the total employment within areas beings combined. To do this we use one further data-set, detailing 2001 employment at the unitary authority level. This consists of 203 observations, reducing to 178 when we exclude the 22 Welsh areas and three Scottish islands. By combining suitable unitary authorities we form the corresponding metropolitan counties of the earnings data-set and reduce to 148 recordings. This data-set then allows the reduction of the 2001 earnings data-set onto the 1989 area definitions.

The final data-sets we use for the analysis are as follows:

[17] Data-set used as weightings for combining 2001 average earnings data
[18] Western Isles, Orkney Islands and Shetland Islands

Earnings data

1989 Average earnings for 87 areas of England and Scotland

1989 Average earnings by the 60 2-digit SIC80 employment sectors

2001 Average earnings for 87 areas of England and Scotland

2001 Average earnings by the 56 2-digit SIC92 employment sectors

Employment data

1989 Employment split by 87 areas and 60 2-digit SIC80 sectors

2001 Employment split by 87 areas and 56 2-digit SIC92 sectors

Land area data

Land areas of the 87 areas of England and Scotland

Mathematical specification of definitions used in report

We define the following variables:

W_j = Average national gross weekly wages for industry j

w_i = Average gross weekly wages for area i in the country

e_{ij} = Employment for industry j in area i

A_i = Physical land area of area i

$w_i^* = \Sigma_j (e_{ij} W_j / \Sigma_j e_{ij})$ Expected average wages in area i

$d_i = w_i - w_i^*$ Earnings differential

$\lambda_i = \Sigma_j e_{ij} / A_i$ Employment density