SERC
Spatial Economics Research Centre

SERC DISCUSSION PAPER 94

Geographic Concentration and the Temporal Scope of Agglomeration Economics: An Index Decomposition

Colin Wren (SERC, Newcastle University Business School)

November 2011

This work is part of the research programme of the independent UK Spatial Economics Research Centre funded by the Economic and Social Research Council (ESRC), Department for Business, Innovation and Skills (BIS), the Department for Communities and Local Government (CLG), and the Welsh Assembly Government. The support of the funders is acknowledged. The views expressed are those of the authors and do not represent the views of the funders.

Geographic Concentration and the Temporal Scope of Agglomeration Economies: An Index Decomposition

Colin Wren*

November 2011

* SERC, Newcastle University Business School

Acknowledgements
This paper was presented at the 2010 European Regional Science Association conference, Jönköping, Sweden; the Spatial Economics Research Centre, London School of Economics, November 2010; and Department of Land Economy, Cambridge University, January 2011. The author is grateful to participants at these events, including Giordano Mion, Olmo Silva and Volodymyr Vakhitov. The comments of Steve Gibbons, Jonathan Jones, Francis Kiraly and Stephen McDonald are also gratefully acknowledged. This research was supported by the UK Economic and Social Research Council under the Spatial Economics Research Centre programme. The usual disclaimer applies.

Abstract

The paper decomposes a geographical concentration index to examine the temporal scope of a spillover, which is the period of time over which one firm's activity directly affects the location of other firms' activities. Natural advantages are fixed over reasonably long time periods, but if spillovers have a limited temporal scope then this can be used to identify these agglomeration economies. To operationalize the index decomposition the paper proposes an empirical methodology that is based on frequency estimator approach, which is applied across time periods. The approach is tested by numerical simulation and by application to a dataset on the location of new economic activity across British regions in the form of investment by foreign-owned plants. Overall, the results support the approach and indicate that the temporal scope of a spillover is on average about five years.

JEL classification codes: R12; R30; L10

Keywords: Industrial location, agglomeration forces, geographic concentration index, spillovers, temporal scope

1. Introduction

The geographic concentration of activity within industries is of great interest (e.g. Krugman, 1991; Rosenthal and Strange, 2003; Brülhart and Traeger, 2004), but while explanations rely on agglomeration economies in some form (see Döring and Schnellenbach, 2006), relatively little is known about the dynamics of these economies. This includes the temporal scope of an agglomeration economy, which is the period of time over which one agent's activity directly affects that of other time-separated agents (Rosenthal and Strange, 2004).[1] In the same way that agglomeration economies are found to diminish with the physical distance between firms (e.g. Rosenthal and Strange, 2003), the same may be true for firms that locate farther apart in time. This issue is little explored, although an exception is Henderson (1997), who finds that the concentration of economic activity within an industry affects employment in that industry, peaking in its effect at between two to five years and petering out after six years.

To examine the effect of agglomeration economies on economic activity, a difficulty is posed by the presence of 'natural advantages', which make some locations relatively more attractive to an industry. These are the availability of natural resources, a favourable climate, proximity to a coast and so on, while under a broader definition they also include the industry-wide advantages of a location that lower an industry's cost, such labour market conditions (e.g. skilled workers) and transportation networks (see Ellison and Glaeser, 1994, 1999). The difficulty posed by these natural advantages is that they are not directly related to the location of other activity in an industry, but they are likely to be correlated with it, giving rise to an identification issue. In the literature several approaches are taken to deal with this. The first is to include terms in a regression to control for these natural advantages, but a difficulty with this is that there are a large and perhaps unknowable number of these.[2] The other approach is to difference out the time invariant location attributes, as in Henderson (1997), but there may be issues of correlation between successively lagged agglomeration terms.

To distinguish agglomeration economies from natural advantages the approach of this paper is to decompose an index of geographic concentration. Whereas natural advantages are

---

[1] Glaeser *et al* (1992) and Henderson *et al (*1995) both find that the characteristics of a city impact on its growth over a 20-year period, but Rosenthal and Strange (2004) argue that this is unlikely to be a direct effect with a 20-year reach, but rather an accumulation of indirect effects that reflects the transitivity of these economies. It not only suggests that agglomeration economies are dynamic, but that their temporal scope is limited.

[2] Ellison and Glaeser (1999) include 16 terms for natural advantages, including transportation costs and labour inputs. They explain 20% of the mean geographic concentration, but reckon that a full set of such terms would explain at least 50%. More generally, when terms are included for natural advantages in regression work they are often treated as uninteresting controls, e.g. Rosenthal and Strange (2001) and Barrios *et al* (2005).

fixed over reasonably long time periods, if agglomeration economies have a limited temporal scope then this potentially provides a way of identifying these. The economies of interest are the agglomerative forces that lead to increased profits from locating close to other activity in the same industry, which is known as a spillover (Ellison and Glaeser, 1997).[3] These include the transfer of knowledge and technology, benefits from a shared labour market and inter-firm trade, but they do not include inter-industry relationships. There are good reasons to suppose that these spillovers have a limited temporal scope, since knowledge and technology are likely to lose their value in a location over time from obsolescence or spatial diffusion (see Howells, 2002; Karlsson and Johansson, 2005). Further, and more generally, competitive pressure and impatience will cause firms move sooner rather than later to exploit these spillovers.

Of course, ultimately, the temporal scope of a spillover is an empirical matter, but the index decomposition allows this to be explored in a general way that includes the possibility that spillovers are static or that they have the same temporal scope as a natural advantage. In this latter case, the decomposition is unable identify the spillovers from natural advantages, so that there is an 'observational equivalence', but this is like elsewhere.[4] Henderson (1997) defines the time invariant location attributes to include many of the natural advantages defined above, but there is the possibility that these will become exhausted or obsolete.[5] This will be reflected in the index of geographic concentration, and the performance of the decomposition is explored below in relation to this. To operationalize the index decomposition, an empirical methodology is proposed based on the frequency estimator approach of Maurel and Sédillot (1999). This treats the geographic concentration index as a combinatorial exercise, which is measured by the number of pairings of economic activities that occur within areas relative to the pairings that occur both within and between areas. This frequency estimator approach is applied in this paper to activity that locates across areas but in different time periods.

The paper explores the index decomposition and empirical methodology in two ways: by numerical simulation and by application to a dataset on the location of investment across the regions of Great Britain. This dataset is ideal for this purpose, as it records the location of

---

[3] These are the own-industry MAR economies, which over small areas are referred to as localisation economies.

[4] Ellison and Glaeser (1997) find that natural advantages and spillovers contribute to their index of geographic concentration index in an identical manner, so that there is an observational equivalence. It is argued here that differences in the timing with which activities locate in relation to one another can potentially reveal information about these processes, leading to an observational non-equivalence. Importantly, the approach does not restrict the temporal scope of the spillover, which is an empirical matter. Thus, should it turn out that this is the same as for the natural advantages then this means that the approach cannot address the identification issue.

[5] Specifically, the time invariant unmeasured location attributes include "regional resource endowments, notions of local culture affecting the local legal, business and institutional climate, and attributes of relatively immobile, specific skill portions of the local labour force" (Henderson, 1997, p. 450). Many of the other natural advantages described above are clearly fixed in time, such as a favourable climate or proximity to the coast.

new economic activity, whether through start-up or *in situ* activity, and on a consistent basis over a long time period. This relates to foreign-owned plants only, but this kind of investment is mobile in its location and amenable to spillovers or natural advantages (or possibly neither), while it is associated with the generation of spillovers (see Blomström and Kokko, 1998). The analysis of investment indicates that for many industries economic activity locates in relation to other activity in the same industry differently in the short run compared to how it locates in the long run. Since these effects are nearly always evident after a one-year and then decay, it not only suggests that spillovers are present, but it indicates that they have a limited temporal scope. On average, the one-year direct effect for agglomerative forces is about twice the long run effect, while spillovers have a temporal scope of about five years. Overall, it suggests that the approach developed in this paper can be employed to identify these spillovers.

In the next section the main indices of geographic concentration are briefly described. The index decomposition is given in section 3 and the empirical methodology is in section 4. The numerical simulation and results are presented in section 5 and section 6 concludes.

2. Geographic Concentration Indices

The geographic concentration index seeks to capture the extent to which economic activity in an industry locates unevenly across space, where this activity is measured by the number of plants or by the number of jobs in different plants. This section briefly presents the two main indices on which the index decomposition draws. To simplify the discussion this is in terms of plant location that is irrespective of the job scale. For ease, the spatial units are referred to as regions throughout and denoted by r (= 1 , …, R). Let n_r denote the number of plants in the industry in r, where $\sum_r n_r = n$, and s_r (= n_r / n) and x_r refer to the share of the number of plants in the industry and all industries respectively, where $\sum_r s_r = \sum_r x_r = 1$. The Herfindahl index is $H = (1 + c^2) / n$, where c is the coefficient of variation of the plant size distribution (Clarke, 1985), but as no account is taken of the job scale then this simplifies to $H = 1 / n$.

Ellison and Glaeser (1997) derive their index, hereafter the EG Index, by modelling a sequence of profit-maximising location decisions by firms. Agglomerative forces arise either from natural advantages or spillovers, where these are represented by the parameters γ^{na} and γ^{s} respectively, which each lie in the unit interval. In addition, there are idiosyncratic plant-specific location effects. Let π_r denote the probability that a plant locates in region r.[6] Then, to model the natural advantages, two restrictions are placed on the moments of π_r:

$$\mathrm{E}(\pi_r) = x_r \quad \text{and} \quad \mathrm{var}(\pi_r) = \gamma^{na} x_r (1 - x_r). \qquad \text{(1a, 1b)}$$

The first means that on average the industry reproduces the location pattern of all industries, while the second captures the importance of natural advantages. When $\gamma^{na} = 0$ the natural advantages have no effect on location, so that each plant locates with a probability x_r and var (π_r) = 0. However, when $\gamma^{na} = 1$ the natural advantages overwhelm the plant-specific effects and the region with the best advantages attracts all plants, so that var (π_r) = x_r (1 - x_r).

Spillovers are modelled as a Bernoulli random event, which is equal to one with a probability of γ^s. Let u_{ir} be an indicator variable, such that $u_{ir} = 1$ if and only if plant i locates in region r. To reproduce the aggregate employment pattern it is again supposed that E(u_{ir}) = π_r. Since a plant locates in a single region only, var (u_{ir}) = π_r (1 - π_r), while corr (u_{ir}, u_{jr}) = γ^s ($i \neq j$), so that by the definition of covariance it follows that:

$$\mathrm{cov}(u_{ir}, u_{jr}) = \gamma^s \pi_r (1 - \pi_r). \qquad (2)$$

Based on these assumptions, the EG Index γ is derived as follows:

$$\gamma \equiv \gamma^{na} + \gamma^s - \gamma^{na}\gamma^s = \frac{\dfrac{G}{(1 - \sum_r x_r^2)} - \dfrac{1}{n}}{1 - \dfrac{1}{n}}, \qquad (3)$$

where $G = \sum_r (s_r - x_r)^2$ is raw geographic concentration. The index permits the geographic concentration of industries to be compared, being zero if plants are only as concentrated as would be expected based on the industrial concentration of the industry and the location of all industry across regions. Industrial concentration is measured by the Herfindahl index, $H = 1 / n$, which is independent of scale. The geography is captured by the location of all industry x_r, which is the benchmark against which the comparison is made. Observational equivalence

---

[6] In Ellison and Glaeser (1997) π_r is the average profitability from locating in region r relative to the profitability of all regions, but a simplified description of their model is given here, and likewise for spillovers.

arises as (3) shows that the EG Index γ is symmetric in γ^{na} and γ^{s}.

The index of Maurel and Sédillot (1999), hereafter the MS Index, does not distinguish between the agglomeration source, but focuses on the probability p that two plants i and j ($i \neq j$) locate in the same region r, i.e. $\mathrm{E}(u_{ir}, u_{jr})$. They obtain the following relationship:[7]

$$\mathrm{E}(u_{ir}, u_{jr}) = \mathrm{cov}(u_{ir}, u_{jr}) + \mathrm{E}(u_{ir})\,\mathrm{E}(u_{jr}) = \gamma\, x_r (1 - x_r) + x_r^2 . \tag{4}$$

Summing this across the R regions and using $\sum_r x_r = 1$ gives:

$$p \;=\; \gamma \left(1 - \sum\nolimits_r x_r^2\right) + \sum\nolimits_r x_r^2 . \tag{5}$$

Maurel and Sédillot (1999) propose a frequency estimator of p, which is discussed below. Substituting this for p in (5) gives the far right-hand side term in (3), but with raw geographic concentration now given by $G = \sum_r \left(s_r^2 - x_r^2\right)$. Maurel and Sédillot (1999) show that the expectation of the difference in the G terms between the EG and MS Indices is zero.

3. The Index Decomposition

The MS Index γ is now decomposed into components for natural advantages and spillovers by allowing for differences in the temporal scope. The decomposition is again for the location decision irrespective of its scale.[8] As a plant may add to its productive capacity at times other than entry, while spillovers may also be generated at times other than entry, then location now refers to plant entry and to other activities that are carried out by a plant *in situ* that add to the productive capacity of an industry, e.g. the introduction a new process or product. These are collectively referred to as activities. The index decomposition is in terms of these activities, where the same notation is used as above, except that it now refers to activities rather than to plants (e.g. n_r is the number of activities in region r).

It is assumed that each activity is mobile across across regions and time, and so in its

---

[7] This can be be derived from (1a) and (2), with γ^{s} replaced by γ. Given observational equivalence, Maurel and Sédillot (1999) note that it is only necessary to consider one of the processes, and they focus on the spillover model. However, since natural advantages are common to all plants in an industry, and location depends on π_r, then cov (u_{ir}, u_{ir}) = var (u_{ir}) = var (π_r), and this relationship also follows from (1b) with γ^{na} replaced by γ.

[8] This avoids the issue raised by Lafourcade and Mion (2007) that the index ignores correlations in location due to differences in the plant size. The conditions for the job size are given below in an appendix.

location potentially amenable to natural advantages or spillovers (possibly neither).[9] Further, where an activity generates a spillover, it is supposed that it is created at the time of location, whether this is plant entry or the introduction of an activity *in situ*. This is plausible, and it is supported by the empirical evidence below. As a further assumption the activities are fixed in scale, so that once located an activity neither expands nor contracts in scale, and neither does it exit or close. This means that a spillover is transmitted at a constant strength over time, but it does not mean that has a constant effect or an infinite temporal scope. This is because in a location it may decay in its usefulness to other firms due to obsolescence or diffusion, while firms may locate sooner rather than later due to competitive pressure or impatience.

In practice, an activity may expand or contract, but this is not relevant here as location is considered irrespective of scale. What is relevant is whether an activity exits or closes as this may truncate the transmission of a spillover. Empirically, it is difficult to observe exits, and particularly the closure of activities that are undertaken *in situ*, while allowing directly for these over-complicates the index decomposition. The approach is to adopt the assumption of fixed scales, so that a spillover is measured net of exits and closures, and to propose a method of correcting for this as part of the empirical methodology below.[10] This means that the index decomposition measures the temporal scope of a spillover net of exits and closures, which is likely to be an under-estimate. Further, the temporal scope is also considered net of negative spillovers, i.e. competition and congestion effects, which work in the same direction.

Consider a firm i that is choosing to locate its activity across regions r but in one of $\tau = 1, 2, \ldots, T$ time periods. Let $u_{ir\tau}$ be an indicator variable, such that $u_{irt} = 1$ if a firm locates its activity in region r at time t and $u_{ir\tau} = 0$ for all $\tau \neq t$. Defining u_{ir} as follows, then $u_{ir} = 1$ if $u_{irt} = 1$ for some t, in which case $u_{is} = 0$ for all $s \neq r$, which is as above:

$$u_{ir} \equiv u_{ir1} + u_{ir2} + u_{ir3} + \ldots + u_{irT}. \tag{6}$$

Like before, industry location reproduces the location pattern of all industries, but now in

---

[9] If a plant is observed to have more than one activity, then $u_{ir\tau}$ in (6) below is zero for some sub-periods, as an activity may not be mobile across all periods, e.g. an *in situ* activity cannot precede plant entry. It is handled by supposing that any non-mobile activities are randomly distributed over time periods, so that $\mathrm{E}(u_{ir\tau}) = x_{r\tau}$, in which case the same decomposition follows. This assumption is adopted in the empirical work below.

[10] The activities that are associated with exits or closures cannot be excluded, as they may transmit or be evidence of a spillover that is generated elsewhere. Data on exits and closures are not available in the empirical work. When account is taken of the scale of an activity a similar method of correction to that which is set out below may potentially be adopted to adjust for the expansions and contractions.

each time period, so that $\mathrm{E}(u_{ir\tau}) = x_{r\tau}$. Since $\sum_\tau x_{r\tau} = x_r$ it follows that $\mathrm{E}(u_{ir}) = x_r$, as in (1a). Further, since u_{ir} can be interpreted as a random variable, by the covariance addition of two sequences of random variables (see Mood *et al*, 1974) it follows that:

$$\mathrm{cov}(u_{ir}, u_{jr}) \equiv \mathrm{cov}\left(\sum_{\tau=1}^{\tau=T} u_{ir\tau}, \sum_{\upsilon=1}^{\upsilon=T} u_{jr\upsilon}\right) = \sum_{\tau=1}^{\tau=T}\sum_{\upsilon=1}^{\upsilon=T} \mathrm{cov}(u_{ir\tau}, u_{jr\upsilon}).$$

Hence, the first equality in (4) can now be written as:

$$\mathrm{E}(u_{ir}, u_{jr}) = \sum_{\tau=1}^{\tau=T}\sum_{\upsilon=1}^{\upsilon=T} \mathrm{cov}(u_{ir\tau}, u_{jr\upsilon}) + \mathrm{E}(u_{ir})\,\mathrm{E}(u_{jr}), \tag{7}$$

where by definition each covariance term is:

$$\mathrm{cov}(u_{ir\tau}, u_{jr\upsilon}) = \mathrm{corr}(u_{ir\tau}, u_{jr\upsilon})\,(\mathrm{var}u_{ir\tau})^{1/2}\,(\mathrm{var}u_{jr\upsilon})^{1/2}. \tag{8}$$

To arrive at the index decomposition each right-hand side term of (8) must be considered. In the case of the correlation term $\mathrm{corr}(u_{ir\tau}, u_{jr\upsilon})$ it is assumed that natural advantages γ^{na} are time invariant, but that spillovers γ^{s} have a limited temporal scope. The arguments for these were presented above. The temporal scope of a spillover is modelled by supposing that once created it has a use to other firms for q periods only, where $0 \le q \le T$. For simplicity, this is the same for all activities in the industry. Given this, the correlation term in (8) is:

$$\mathrm{corr}(u_{ir\tau}, u_{jr\upsilon}) = \begin{cases} \gamma^{na} & \text{if } |\tau - \upsilon| > q \\ \gamma^{na} + \gamma^{s} & \text{if } |\tau - \upsilon| \le q \end{cases}, \tag{9}$$

where τ, $\upsilon = 1, 2, \ldots, T$ refer to the time of location of activities i and j, and q is the temporal scope. Either activity i or j locates first, or they co-locate. If two activities locate more than q periods apart the correlation reflects natural advantages γ^{na} only, but if they locate within q periods of each other it reflects γ^{s} and γ^{na}, as natural advantages occur in each period.[11] Of course, for any given industry it may turn out that $\gamma^{na} = 0$ or $\gamma^{s} = 0$ (or both). Importantly,

---

[11] A simple sum of γ^{na} and γ^{s} is taken, as there is no reason to suppose otherwise, where $\gamma^{na} + \gamma^{s} \in [-1, 1]$.

(9) allows for the possibility that spillovers are static (i.e. $q = 0$), or that the temporal scope is the same as for natural advantages ($q = T$), so that there is observational equivalence.

Like Ellison and Glaeser (1997), the spillovers are of an "all or nothing" variety, so that they occur only within a region and are independent of the geographical distance between plants. However, unlike Ellison and Glaeser, in which spillovers are independent of the order in which firms make their location decisions, they are now symmetric for static externalities only, where activities co-locate. Otherwise, they are asymmetric, as dynamic spillovers (i.e. $q > 0$) only affect those location decisions that are later in time. Since spillovers are transitive, it is possible that a sequence of dynamic spillovers give rise to a longer-run indirect effect, but what is captured by (9) is the direct (dynamic) effect of a spillover over q periods.

The other right-hand side terms in (8) concern the variance. It is assumed that these vary across regions, but that they are constant over time, so that $\mathrm{var} u_{ir\tau} = \mathrm{var}(u_{ir} / T)$ for all i. This is plausible for activity that locates reasonably evenly over time, and it is consistent with the assumption about the expectation of $u_{ir\tau}$, while the results below are insensitive to this.[12] Since $\mathrm{var}(u_{ir}/T) = \mathrm{var}(u_{ir}) / T^2$ and $\mathrm{var}\, u_{ir} = x_r (1 - x_r)$ by (4) it follows that:

$$(\mathrm{var} u_{ir\tau})^{1/2} (\mathrm{var} u_{jrv})^{1/2} = x_r (1 - x_r) / T^2. \tag{10}$$

Substituting (9) and (10) into (8), and then substituting this into (7) and summing across the R regions, the probability p that a pair of activities locate in the same region is now given by:

$$p = \left[\gamma^{na} + \gamma^{s} \left\{ \frac{q\,(2T - q - 1) + T}{T^2} \right\} \right] \left(1 - \sum_r x_r^2\right) + \sum_r x_r^2 . \tag{11}$$

Of course, when no account is taken of the agglomeration source then p is given by (5), and so using (11) to substitute for p in (5), the index decomposition is:

$$\gamma = \gamma^{na} + \left\{ \frac{q\,(2T - q - 1) + T}{T^2} \right\} \gamma^{s} . \tag{12}$$

By construction, the left-hand side is the MS Index, so that (12) exactly decomposes this index

---

[12] The index decomposition is given by (12) below, where the term in curly brackets is the ratio of the number of time periods over which spillovers are observed relative to that in which natural advantages are observed. Since activities may arrive unevenly over time, the ratio could instead be measured according to the number of activity pairings, but this makes no qualitative difference to the results reported in tables 2 and 3 below.

into those parts that are due to natural advantages and spillovers. It has this form since there are T^2 covariance terms in (7), but whereas γ^{na} occurs in each of these, γ^{s} is in $q\ (2\ T - q - 1) + T\ (\leq T^2)$, which gives the term in curly brackets in (12), which weights these components.[13] If spillovers have the same temporal scope as natural advantages, i.e. $q = T$, they have the same weight and (12) reduces to $\gamma = \gamma^{na} + \gamma^{s}$. This differs from the index decomposition of Ellison and Glaeser in (3), but arising from the specification in (9).[14]

An implication of (12) is that if spillovers have a limited temporal scope, such that $q < T$, then the geographic concentration index γ will tend to reflect agglomeration effects due to natural advantages. This is because the term in curly brackets is less than unity. Indeed, since this term depends on $1\ /\ T$ and $q\ /\ T$ only, then γ is approximately equal to γ^{na} for large T.[15] It arises because natural advantages are present in every period, whereas the spillovers have a limited temporal scope, so that most correlations that make-up the index are due to the former. It does not mean natural advantages are more important than spillovers, as it could be that γ^{na} is small or zero, but it does mean that γ is likely to be small or zero in this case.

4. Empirical Methodology

Again, interest is in activities that add to capacity, irrespective of their scale. The empirical methodology for estimating γ^{na} and γ^{s} is based on the index decomposition, coupled with the frequency estimator approach of Maurel and Sédillot (1999). To understand the latter, for a distribution of activities across regions it expresses the number of pairings of activities that occur within regions relative to the number of pairings that occur within and across regions. If all activities locate in a single region it is equal to unity, but as activities become more spread out across regions it falls in value. It is written in a combinatorial form as follows,

---

[13] Given that activities i and j can each locate across the T periods, they co-locate in the same period in T of the T^2 possibilities, while they locate within q (> 0) periods of each other in a further $2\ \{(T - 1) + (T - 2) + (T - 3) + \ldots + (T - q)\}$ periods. These sum to $q\ (2\ T - q - 1) + T$. For example, if $q = 1$ there are $3\ T - 2$ observations on γ^{s}, of which T relate to the same period, $T - 1$ go from i to j over a single period and $T - 1$ go from j to i.

[14] If instead $\mathrm{corr}(u_{ir\tau}, u_{jrv}) = \gamma^{na} + \gamma^{s} - \gamma^{na}\gamma^{s}$ for $|\tau - v| \leq q$ in (9), then (12) reduces to (3) when $q = T$, but there seems no good reason to suppose this here, so that the simple sum is taken in (9).

[15] That is, γ tends to γ^{na} as T tends to infinity. For example, if $q = 1$ and $T = 20$, (12) gives $\gamma = \gamma^{na} + 0.145\gamma^{s}$.

where the numerator is the sum of the number of pairings between activities within each of the R regions and the denominator is for the R regions taken as a whole:[16]

$$\hat{p} \;=\; \frac{{}^{n_1}C_2 + {}^{n_2}C_2 + \ldots + {}^{n_R}C_2}{{}^{n}C_2} \;=\; \frac{\sum_{r=1}^{r=R} n_r\left(n_r - 1\right)}{n\left(n-1\right)} \;=\; \frac{\sum_r s_r^2 - 1/n}{1 - 1/n}. \qquad (13)$$

Expanding the first right-hand side term in (13) gives the second term, while dividing through by n^2 gives the final term. This is the frequency estimator of p in (5), and substituting this for p and rearranging gives the far right-hand side term in (3), which is the MS Index. Thus, the index can be interpreted as a counting exercise that is based on the number of pairings. As an activity pairing may be due to natural advantages or spillovers (or neither), then the frequency estimator approach may potentially be applied across time in order to quantify these.

Broadly, the empirical methodology is as follows. First, an estimate of γ^{na} is obtained, based on all the periods for which spillovers do not to occur, i.e. $|\tau - \upsilon| > q$ in (9). Second, using this, the MS Index and the index decomposition in (12), the estimate of γ^{s} is obtained. It requires q to be pre-specified, so that different values of q are taken. It is advantageous as it enables the time profile of the direct spillover effect to be explored, but to determine the (optimal) temporal scope of the spillover, denoted q^*, we are guided by Ellison and Glaeser (1997), although other approaches exist.[17] They regard a value of less than 0.02 as not very localized and a value of more than 0.05 as highly localized. As a specified value of q may be different from its optimal value q^* an important requirement of this approach is that the estimate of the natural advanatage term γ^{na} should be robust to q.

Formally, let $n_{r\tau}$ denote the number of activities locating in region r at time τ, where $n_\tau = \sum_r n_{r\tau}$, then the frequency estimator for the probability of a natural advantage p_q^{na} is:

$$\hat{p}_q^{na} = \frac{\sum_{r=1}^{r=R}\left\{n_r\left(n_r - 1\right) - \sum_{t=q+1}^{t=T}\left[2\, n_{rt} \sum_{\tau=1}^{\tau=q} n_{rt-\tau} + n_{rt}\left(n_{rt} - 1\right)\right]\right\}}{n\left(n-1\right) - \sum_{t=q+1}^{t=T}\left[2 n_t \sum_{\tau=1}^{\tau=q} n_{t-\tau} + n_t\left(n_t - 1\right)\right]}. \qquad (14)$$

---

[16] Spillovers are asymmetric over time, but for any given pair of activities either activity can locate first. This suggests the use of permutations, but combinations make no difference to the expressions that are obtained. If all activities locate in a single region, (13) gives $\hat{p} = 1$, and substituting this for p in (5) gives $\gamma = 1$. If activities locate according to all industry, (13) and (5) give $\gamma = 1 / (1 - n)$, which is (3) with $G = 0$, that is zero for large n.

[17] This follows the approach adopted elsewhere to interpret these indices. An alternative approach may involve constructing standard errors for these terms using bootstrapping techniques, but this is outside the paper's scope.

This expression is derived in Appendix A, where further explanation can be found. Basically, the numerator sums all the activity pairings that occur within regions, as in the second right-hand side term of (13). However, to exclude the activity pairings where spillovers are present it deducts the pairings for the activities that locate in the same region within q periods of one another, including in the same period. The denominator calculates the activity pairings on the same basis, as those pairings that both occur within and across regions minus those that occur across any pair of regions in the same q time periods. The probability estimate $\hat{p}_q^{na}$ lies in the unit interval, such that $\hat{p}_q^{na} = 1$ when all activities locate in a single region.

The probability of a natural advantage is given by (5) with $|\tau - \upsilon| > q$, for which (14) is an estimate. Making this substitution and rearranging, the estimate of the natural advantage term $\hat{\gamma}_q^{na}$ is as follows, where $\hat{p}_q^{na}$ is evaluated using (14):[18]

$$\hat{\gamma}_q^{na} = \frac{\hat{p}_q^{na} - \sum_r x_r^2}{1 - \sum_r x_r^2}. \tag{15}$$

To get the estimate for the spillover term, substituting (15) into the index decomposition in (12), where $\hat{\gamma}$ is the evaluation of the MS Index, gives:

$$\hat{\gamma}_q^s = \left\{\frac{T^2}{q\,(2T - q - 1) + T}\right\}\left(\hat{\gamma} - \hat{\gamma}_q^{na}\right). \tag{16}$$

By construction, the methodology exactly decomposes the Maurel and Sédillot index. If the activities locate across regions in the short run (i.e. within q periods of one another) the same as they locate in the long run, then $\hat{\gamma}_q^{na}$ is determined in an identical way to $\hat{\gamma}$, and $\hat{\gamma}_q^s$ is zero by (16). This could mean that spillovers are not present, or that they have the same temporal scope as natural advantages and that they cannot identified using this approach, so that there

---

[18] This necessitates a simplification, such that $x_{r t}$ is approximated by x_r. This is consistent with the assumption regarding the variance terms in (10), and it is reasonable as the purpose of this term is to capture the geography according to how all industry locates, although it potentially smooths out any trade cycle effect. More generally, $\mathrm{E}\left(u_{ir}\right)\mathrm{E}\left(u_{jr}\right)$ in (4) should be measured as $\sum_\tau \sum_\upsilon \mathrm{E}\left(u_{ir\tau}\right)\mathrm{E}\left(u_{jr\upsilon}\right)$, from which (5) is derived, and means that the x_r^2 term in the numerator of (15) should instead be $\sum_\tau \sum_\upsilon x_{r\tau} x_{r\upsilon}$. However, this greatly complicates the empirical work, as it means that the all industry share must be measured for the beginning (τ) and end (ν) of the time period over which *each* industry activity pairing is considered, so that a simplification is desirable.

is observational equivalence. However, if this is not the case, then $\hat{\gamma}$ - $\hat{\gamma}_q^{na}$ differs from zero, and so does $\hat{\gamma}_q^s$, which enables the spillover effect to be evaluated over q periods.

The index decomposition supposes that the activities are fixed in scale, and so do not exit or close. However, if each activity has a life of $\overline{q}$ periods, where $\overline{q} < q*$, then the index decomposition will under-estimate the optimal temporal scope of a spillover, as $\overline{q}$ rather than $q*$ is observed. The index decomposition in (16) continues to follow in this case, but with q replaced by $\overline{q}$. Hence, if the mean exit / closure rate over q periods is known, or it can be approximated, then the spillover estimate can be adjusted by multiplying it by $\{q\ (2\ T - q - 1) + T\}\ /\ \{\overline{q}\ (2\ T - \overline{q} - 1) + T\}$.[19] Of course, there is still the issue of how to calculate the MS Index in the presence of these exits / closures, which is considered below.

5. Numerical Explorations

The plausibility of the estimates from the index decomposition and empirical methodology is explored in two ways: by numerical simulation and by application to a dataset on location. In the latter case, this is for investment by foreign plants across British regions. These data have the major advantage that they not only identify investments in the form of new plant entry, but *in situ* activity in the form of re-investments. As these add to an industry's capacity, and may serve as substitutes for one another, it is important to observe both kinds of investment to capture the transmission and receipt of any spillover that is generated by a new activity. The data are available on a consistent basis over a long time period, which is important given the requirement that the natural advantage estimate should be robust to q.

5.1 Simulation

The purpose of the numerical simulation is to examine how the approach performs in relation to location patterns where the presence of spillovers or natural advantages can be reasonably easily observed. As such, to keep matters simple, an economy is considered with four regions ($R = 4$), four time periods ($T = 4$) and sixteen activities ($n = 16$), where four activities locate in each period in each case. It is supposed that the (optimal) temporal scope of a spillover is

---

[19] This supposes that $\overline{q}$ is constant across regions, so that the natural advantage estimate is not affected by this. If $\overline{q} \geq q*$ then (12) captures the spillover effect, as the exit or closure occurs after this effect is exhausted.

one period, i.e. $q = 1$. Figure 1 displays eight location patterns, which is sufficient to form a judgement, although the *a priori* characterisation of these in part depends on how all industry locates, reflecting the underlying geography, e.g. whether all industry locates all regions or not. There is no significant advantage to varying the dimensions of the economy (i.e. R or T), or indeed q, while the index and index decomposition are robust to n.[20]

Two regions are settled equally in (a) to (d) of figure 1, all four regions are settled equally in (g) and (h), and (e) and (f) are intermediate to these. The estimates associated with these are given in table 1 under two different scenarios about how all industry locates: either equally across all regions in part (i) of table 1 and equally across two regions only in part (ii). These represent a further sensitivity test of the estimates. As an *a priori* characterisation of the location patterns in figure 1, case (a) shows natural advantages but a complete absence of spillovers as no activity locates within a period of any other. Spillovers are present in (c) and (d), although more pronounced in (d) where eight activities locate immediately after other activities. Spillovers are also evident in (e) to (g), but in this case more pronounced in case (g) where the natural advantages are weak. Finally, cases (b) and (h) are equivalent under (i) and (ii) respectively, but in either case it is difficult to argue that spillovers are present.

Table 1 presents the frequency estimator $\hat{p}$ and MS Index $\hat{\gamma}$ for each location pattern in figure 1 under the two scenarios about how all industry locates. The estimator is based on (13) and is the same for (i) and (ii), while the MS Index is based on (5) and varies between these. The MS Index and its components can be large in magnitude, but reflecting the stark nature of the economies exhibited in figure 1. The index can also be negative, which is geographic deconcentration, e.g. the MS Index is negative for (e) to (h) in part (ii) of table 1, but this is because the industry locates in three or four regions, whereas all industry locates in just two regions only. To decompose the index the frequency estimator $\hat{p}_1^{na}$ is based on (14) with $q = 1$, $R = 4$ and $T = 4$, where the estimates are shown in table 1 (see note to this table for an explanation). From this, the index decomposition is made of $\hat{\gamma}_1^{na}$ and $\hat{\gamma}_1^{s}$ using (15) and (16). These sum to give the MS Index by (12) where the weight is equal to 0.625.

---

[20] Increasing R or T serves to replicate the kinds of scenario displayed in figure 1, while varying q makes it more difficult to assess *a priori* whether spillovers or natural advantages are present. If n increases, while the regional industry shares s_r are maintained, the index and decomposition are highly stable. For example, as n tends to infinity, for case (a) in part (i) of table 1 below, $\hat{\gamma}$ tends to 1 / 3, $\hat{\gamma}_1^{na}$ tends to 13 / 21 = 0.62 and $\hat{\gamma}_1^{s}$ tends to -16 / 35 = -0.46, which is like the estimates in this table. Likewise, for case (d) in part (i), where spillovers are prominent, $\hat{\gamma}$ again tends to 1 / 3, but now $\hat{\gamma}_1^{na}$ tends to -1 / 7 = -0.14 and $\hat{\gamma}_1^{s}$ tends to 16 / 21 = 0.76.

In general, the results in table 1 conform to prior expectations. The natural advantages come through in (a), while the negative spillover estimate indicates that activity locates away from regions in which it located in the previous period. Spillovers are stronger for (d) than (c), which is expected, as the number of activity pairings increases more than proportionately with the number of activities that locate in a region subsequent to other activities. Likewise for (e) to (g), the spillover estimate is greater for (f) than (e), and greater for (g). Overall, the estimates vary between (i) and (ii), which is expected, but the approach is able to pick-up the effect of spillovers under either scenario, even when the MS Index is negative.

As regards cases (b) and (h), the spillover estimate is (approximately) zero, and when the industry locates the same as all industry the natural advantage term is also zero, i.e. (b) in part (ii) and (h) in part (i), where identical estimates are obtained. However, if location across regions differs from that for all industry, so that the MS Index is non-zero, then it is captured as a natural advantage rather than as a spillover. Finally, the index decomposition supposes that the natural advantages are the fixed over time. Comparison of (b) and (g) shows that if there is a structural break in location (from regions r_1 and r_2 to r_3 and r_4), then this affects the MS Index, and which tends to be attributed to the spillovers. It suggests that those industries where there is a sharp shift in location pattern need to be identified *a priori*, and dealt with either by sub-period or by an adjustment to the empirical methodology to allow for this. In practice, such sharp shifts in location are likely to be the exceptional, while their effect will no doubt be lessened by the existence of many regions and time periods.

5.2 Application to a Dataset on Location

The approach is also explored by application to a dataset on the location of investment. This is a good test as investment represents the creation of new economic activity that adds to an industry's capacity. The investment data are available for foreign-owned plants only, so that what is examined is how foreign investment locates in relation to other foreign investment. A broad definition of entry is taken that includes start-ups and acquisitions (possibly of another foreign-owned plant), which may serve as substitutes for one another. The re-investments are major upgradings that add to a plant's capacity, e.g. a new production process or product (see Wren and Jones, 2009). Spillovers are likely to be important for foreign-owned plants, not least as investment by these plants may be associated with a 'specific advantage'.

The spillovers do not include linkages with domestic plants, but which are likely to be of far less significance, so that for practical purposes these are assumed to be part of the idiosyncratic effects.[21] Likewise, spillovers do not include the inter-industry effects that arise from linkages, externalities and co-agglomeration economies, but again this is like elsewhere, for which the same assumption is made.[22] Hence, the spillovers that are measured comprise the intra-industry agglomerative economies between foreign-owned plants.

The exit or closure of activities is not known, so that the (optimal) temporal scope that is measured may be an under-estimate, but possibly only by a year or so, while foreign-owned plants tend to be larger in scale (Jones and Wren, 2004), which lessens the significance of this. The MS Index is calculated for all the locations over the study period, including plants that exit, and it is this that is decomposed. It means that $\hat{\gamma}^{na}$ and $\hat{\gamma}^{s}$ are likely to be better determined as it includes all locations over the study period.[23] Rosenthal and Strange (2001) calculate a geographic concentration index for entrants only and find that it does not differ too much from the usual cross-section measure. Kim (1999) finds that the index is stable over long time periods, and Dumais *et al* (2002) find that births, exits, expansions and contractions act together to maintain the geographic concentration index over time.

The data give annual information on about 6,500 investments by foreign-owned plants across the regions of Great Britain over 1985-2005.[24] This is for 22 manufacturing industries, giving an average of 294 observations on location for each industry, which represents up to 43,000 pairings (i.e. ${}^{294}C_2$) to assign to natural advantages, spillovers or possibly neither. The industries are at the 2-digit level, but as FDI falls unevenly across these they are disaggregated

---

[21] The analysis is complementary to work on foreign direct investment (FDI) that measures MAR externalities by the number of locations by foreign plants in the same industry in the preceding period, e.g. Basile *et al* (2008) and Mariotti *et al* (2010). It is more general as it allows for *in situ* investment and for effects on location of more than a year. Numerous studies find that FDI has a greater effect on the location of other FDI than does domestic activity (e.g. Crozet *et al*, 2004; Head *et al*, 1995), while in net terms Mariotti *et al* (2010) find that spillovers flow from foreign to domestic plants, rather than in the converse direction.

[22] If two industries locate according to some common natural advantage then this co-agglomeration effect is not relevant as location will reflect the natural advantage. If they locate due to a spillover between them then the observed effect may depend on which of the industries locates 'first' and whether this is determined by natural advantages or spillovers. The assumption throughout is that these are idiosyncratic effects.

[23] The alternative is to calculate the geographic concentration index for a single cross-section of plants at the end of the sample period and to decompose this, i.e. base $\hat{\gamma}^{na}$ and $\hat{\gamma}^{s}$ on surviving plants only. However, any plant that exits prior to time T will not be taken into account, potentially biasing $\hat{\gamma}^{na}$ and $\hat{\gamma}^{s}$, while there is the issue of the closure of *in situ* activities. In any event, it is not an option as exits / closures are not observed.

[24] The data are supplied by the UK central government and used to report UK inward investment. These kind of data are used to examine location elsewhere (e.g. Dimitropoulou *et al*, 2006; Alegria, 2009; Jones and Wren, 2012), where further details can be found. The data are reckoned to be comprehensive of British inward FDI, and comparison with the published aggregate data (measured by net employment) in the UK production census shows no significant difference by region. The areas are the Government Office regions of Great Britain defined at *Eurostat* NUTS I level, where London is part of the South East, so that there are ten regions. Studies show that spillovers can extend over large areas (Döring and Schnellenbach, 2006; Jones and Wren, 2011). Start-ups, acquisitions and re-investments each account for about a third of the projects. The results will no doubt differ by entry mode, but given that these are substitute forms of entry then it is important to include them all.

or aggregated to form reasonably homogeneous groups, as in Appendix B. Some industries still have a small of observations, but it was shown above that the decomposition is robust to this. Agglomerative forces may be weaker at a higher level of industrial aggregation (Maurel and Sédillot, 1999), which may be reflected in the temporal scope, but the only other known evidence on this issue in Henderson (1997) is also for 2-digit industries. Natural advantages may include UK regional policy grants that have been offered to foreign investment in certain regions throughout the period (Wren and Jones, 2011), so that the benchmark regional share of investments x_r is calculated from the dataset for the manufacturing FDI as a whole.

5.3 Results

The results for $\hat{\gamma}$, $\hat{\gamma}_q^s$ and $\hat{\gamma}_q^{na}$ are given in table 2 based on the frequency estimates $\hat{p}$ and $\hat{p}^{na}$ reported in Appendix B. To aid clarity each index number is multiplied by 100, where to interpret these a value of less than 2.00 is not very localized and a value of more than 5.00 is regarded as highly localized. On average, the final row of table 2 shows that new economic activity is agglomerated by industry, i.e. $\hat{\gamma}$ = 3.27. The spillover estimates $\hat{\gamma}_q^s$ are generally positive, and on average they decline in the final row of table 2 as q increases from 4.00 over 1 year to 1.90 over 5 years. Further, the natural advantage estimates are robust to q, so that on average $\hat{\gamma}_q^{na}$ lies in the range 2.41 to 2.72 as q varies between 1 and 5 years. This reflects the robustness of the frequency estimate $\hat{p}_q^{na}$ for each industry shown in Appendix B.

At the industry level, about half the industries are geographically concentrated in table 2, while spillovers occur across a range of activities. These include labour-intensive industries (i.e. textiles, leather, publishing and furniture), high-tech industries (pharmaceuticals, TV and radio and office machinery) and capital-intensive industries (petroleum products, chemicals, basic metals and transport). These industries are found to be geographically concentrated at the 2-digit industry level elsewhere.[25] While the results suggest that spillovers are important to these industries, natural advantages also come through in many of these industries, although also in pulp, paper and metal products where a raw material source may be relevant. Natural advantages occur in electronic components and motor vehicles, which could be the presence

---

[25] These industries account for virtually all of the most geographically concentrated 2-digit industries in Maurel and Sédillot (1999), and which encompass the most localized 4-digit industries. The approach does not identify the source of a spillover, but it is reasonable that labour and knowledge are important in the labour-intensive and high-tech industries, and that intra-industry linkages are important in the capital-intensive industries.

of a skills base, which may characterize some regions. Potentially, there could be structural breaks in the importance of natural advantages, but in general the results indicate otherwise, as there is good evidence for the natural advantages across a range of industries.

For some industries in table 2 the spillovers do not to decay very quickly (e.g. textiles and chemicals), while for other industries the estimates actually increase (i.e. pharmaceuticals, minerals products and electronic components). To examine this, the spillovers were examined over longer time horizons of 7 and 10 years, and the results are given in table 3. In so doing, the 10-year spillover estimate is sometimes perverse (i.e. negative and large), but this indicates that the data are being stretched too far.[26] Table 3 suggests that the spillovers can extend over long time horizons, but that they eventually decay and are nearly always exhausted over a 10-year time span. In pharmaceuticals and electronic components the spillovers build-up slowly, but where present they are nearly always evident over a one-year time horizon.

Overall, the results indicate that spillovers decay with time, and that on average they have an (optimal) temporal scope of about five years. This is similar to Henderson (1997), but what is measured here is a mean effect over q years, which may produce a longer temporal scope. Contrary to this, the spillover is measured net of exits, which means that it is an under-estimate, although under reasonable assumptions this may only be by a year or so.[27] There are variations across industries, but on average the combined one-year effect of agglomerative forces (i.e. spillovers and natural advantages) is twice the long run direct effect. Finally, given that spillovers decay and have a limited temporal scope, the results point to the importance of natural advantages to the overall value of the geographic concentration index.[28]

6. Conclusions

This paper examines the temporal scope of a spillover, which is the period of time over which one agent's activity directly affects that of other agents. In the same way that agglomeration

---

[26] It arises as when q is large the number of observations on the pairings that are due solely to natural advantages (e.g. locations more than 10 years apart) is much smaller, so that $\hat{\gamma}_q^{na}$ can be poorly determined, and likewise for $\hat{\gamma}_q^{s}$. This may also be the case for mineral products when measured over the 5-year time horizon.

[27] If the mean exit / closure rate is 20% over this period say, then in gross terms the (optimal) temporal scope is about 6 years. That is, given $q = 5$, $\overline{q} = 4$, and $T = 20$ then according to the method of correction outlined above the adjustment involves multiplying $\hat{\gamma}^{s}$ by $\{q\,(2\,T - q - 1) + T\} / \{\overline{q}\,(2\,T - \overline{q} - 1) + T\} = 1.1875$.

[28] Across the 22 industries in table 2, the correlation coefficient between $\hat{\gamma}$ and $\hat{\gamma}_q^{na}$ is 0.96, 0.94, 0.96 and 0.96 for $q = 1, 2, 3$ and 5 respectively, but between $\hat{\gamma}$ and $\hat{\gamma}_q^{s}$ it is much lower, lying between 0.34 and 0.41.

economies diminish with the physical distance between firms, these spillovers may be smaller for firms that locate farther apart in time. The issue is explored in this paper by decomposing an index of geographic concentration in which spillovers have a limited temporal scope. The index decomposition is general, as it allows for the possibilities that the spillovers are static or are long-lived, so that they affect location over the same period as natural advantages. While the natural advantages are fixed, the approach may be adapted to handle breaks in these that are identified *a priori*, but in general the results suggest that this is not an issue.

To evaluate the spillovers, the decomposition is coupled with the frequency estimator approach, and tested by numerical simulation and by application to a dataset on foreign direct investment, representing the location of new economic activities. This suggests that activity locates in relation to other activity in the same industry differently in the short run compared to how it locates in the long run, such that the one-year direct effect of agglomerative forces (spillovers and natural advantages) is about twice the long run effect. Given that the natural advantages are fixed, then it suggests that where they exist the spillovers generally decay and have a temporal scope that is on average about five years. The results accord with the limited evidence elsewhere. Overall, the paper offers a new approach for determining the relative importance of spillovers and natural advantages, which may be applied elsewhere.

Figure 1: Simulation of Activity Location Patterns

(a)

	r_1	r_2	r_3	r_4
t_1	4	0	0	0
t_2	0	4	0	0
t_3	4	0	0	0
t_4	0	4	0	0

(b)

r_1	r_2	r_3	r_4
2	2	0	0
2	2	0	0
2	2	0	0
2	2	0	0

(c)

r_1	r_2	r_3	r_4
4	0	0	0
0	4	0	0
0	4	0	0
4	0	0	0

(d)

r_1	r_2	r_3	r_4
4	0	0	0
4	0	0	0
0	4	0	0
0	4	0	0

(e)

	r_1	r_2	r_3	r_4
t_1	2	2	0	0
t_2	2	2	0	0
t_3	0	0	2	2
t_4	2	2	0	0

(f)

r_1	r_2	r_3	r_4
2	2	0	0
2	2	0	0
2	2	0	0
0	0	2	2

(g)

r_1	r_2	r_3	r_4
2	2	0	0
2	2	0	0
0	0	2	2
0	0	2	2

(h)

r_1	r_2	r_3	r_4
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

Note: Each cell gives number of 16 activities locating across $r = 4$ regions and $t = 4$ time periods.

Table 1: Numerical Explorations

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
$\hat{p}$	0.47	0.47	0.47	0.47	0.27	0.27	0.20	0.20
$\hat{p}_1^{na}$	0.70	0.48	0.41	0.11	0.21	0.14	0.03	0.22
(i) All industry locates equally across all regions ($\sum_r x_r^2$ =0.25):								
$\hat{\gamma}$	0.29	0.29	0.29	0.29	0.02	0.02	-0.07	-0.07
$\hat{\gamma}_1^{na}$	0.60	0.31	0.21	-0.19	-0.05	-0.14	-0.30	-0.04
$\hat{\gamma}_1^{s}$	-0.50	-0.03	0.13	0.77	0.11	0.26	0.37	-0.05
(ii) All industry locates equally across two regions ($\sum_r x_r^2$ = 0.50):								
$\hat{\gamma}$	-0.07	-0.07	-0.07	-0.07	-0.47	-0.47	-0.60	-0.60
$\hat{\gamma}_1^{na}$	0.41	-0.04	-0.19	-0.78	-0.58	-0.71	-0.94	-0.56
$\hat{\gamma}_1^{s}$	-0.77	-0.05	0.19	1.00	0.18	0.38	0.54	-0.06

Notes: Estimates based on location patterns in (a) to (h) of figure 1 with q = 1. $\hat{p}$ and $\hat{p}_1^{na}$ given by (13) and (14) and $\hat{\gamma}_1^{na}$ and $\hat{\gamma}_1^{s}$ by (15) and (16). To explain the calculation of (14), consider case (a) in figure 1, where n_r (n_r - 1) = 8 x 7 = 56 and n (n - 1) = 16 x 15 = 240. The first term in square brackets in the numerator of (14) is zero, since positive activity is followed by zero activity, and conversely, while for the second term three regions have non-zero activity in time periods 2, 3 and 4, so that this is 3 x (4 x 3) = 36. In the denominator the first term in square brackets is 2 x {(4 x 4) + (4 x 4) + (4 x 4)} = 96, while the second term is the same as in the numerator. Hence, $\hat{p}_1^{na}$ = (2 x (8 x 7) - 0 - 36) / (240 - 96 - 36) = 0.70.

Table 2: Results for the Index Decomposition

Industry	$\hat{\gamma}$	$\hat{\gamma}_q^s$				$\hat{\gamma}_q^{na}$			
		1 yr	2 yr	3 yr	5 yr	1 yr	2 yr	3 yr	5 yr
Food, Beverages and Tobacco	**-0.67**	0.39	0.24	0.32	-0.16	-0.72	-0.72	-0.77	-0.60
Textiles and Textile Products	**2.72**	6.90	6.20	6.43	5.87	1.77	1.33	0.75	0.05
Leather and Leather Products	**8.29**	31.76	22.42	12.77	5.71	3.89	3.25	4.38	5.69
Wood and Wood Products	**1.56**	1.05	1.90	1.58	0.38	1.42	1.14	1.08	1.39
Pulp, Paper and Paper Products	**5.44**	0.85	1.32	1.48	0.22	5.32	5.14	4.98	5.34
Publishing and Printing	**10.43**	4.37	1.58	-0.31	0.99	9.83	10.07	10.52	9.97
Coke, Refined Petroleum Products	**2.16**	2.83	3.06	0.17	0.91	1.77	1.48	2.11	1.75
Chemicals	**1.36**	2.31	2.30	2.10	1.96	1.04	0.84	0.72	0.47
Pharmaceuticals	**4.66**	3.74	3.70	3.63	4.29	4.14	3.83	3.55	2.70
Rubber and Plastic Products	**0.28**	0.78	0.84	0.62	-0.02	0.17	0.09	0.09	0.29
Mineral Products	**1.54**	0.07	0.22	0.55	1.99	1.53	1.49	1.37	0.64
Basic Metals	**2.93**	4.15	3.74	2.81	2.37	2.36	2.09	2.07	1.85
Metal Products	**2.73**	0.98	0.68	0.08	0.07	2.59	2.58	2.71	2.70
Machinery	**0.92**	0.70	0.54	0.58	0.04	0.82	0.80	0.74	0.90
Office Machinery	**4.23**	2.26	1.77	2.24	1.82	3.92	3.83	3.54	3.40
Electrical Machinery	**0.62**	1.76	0.97	0.50	-0.19	0.37	0.41	0.47	0.71
Electronic Components	**10.84**	1.70	1.86	3.47	4.56	10.60	10.42	9.77	8.76
TV and Radio	**1.36**	2.16	1.74	1.15	1.56	1.06	0.97	1.01	0.65
Medical and Optical Instruments	**0.59**	0.94	0.54	0.81	0.93	0.46	0.47	0.35	0.17
Motor Vehicles	**6.23**	0.32	0.32	0.54	0.69	6.18	6.16	6.06	5.91
Other Transport	**0.61**	8.98	5.55	3.02	3.02	-0.63	-0.64	-0.32	-0.77
Furniture and Leisure Goods	**3.15**	8.99	8.14	6.89	4.72	1.90	1.32	1.03	1.00
Mean	**3.27**	**4.00**	**3.17**	**2.34**	**1.90**	**2.72**	**2.56**	**2.56**	**2.41**

Note: Decomposition of MS Index in (12), based on (15), (16) and frequency estimates in Appendix B. Each index number multiplied by 100, where q = 1, 2, 3 and 5 years. x_r is regional share of FDI across all manufacturing industries.

Table 3: Index Decomposition over Longer Time Horizons

Industry	$\hat{\gamma}$	$\hat{\gamma}_q^s$				$\hat{\gamma}_q^{na}$			
		3 yr	5 yr	7 yr	10 yr	3 yr	5 yr	7 yr	10 yr
Food, Beverages and Tobacco	**-0.67**	0.32	-0.16	-1.00	-1.41	-0.77	-0.60	-0.08	0.40
Textiles and Textile Products	**2.72**	6.43	5.87	5.68	0.76	0.75	0.05	-0.61	2.15
Leather and Leather Products	**8.29**	12.77	5.71	4.11	6.52	4.38	5.69	5.87	3.39
Wood and Wood Products	**1.56**	1.58	0.38	-0.39	-1.32	1.08	1.39	1.79	2.56
Pulp, Paper and Paper Products	**5.44**	1.48	0.22	-0.77	-3.15	4.98	5.34	5.89	7.80
Publishing and Printing	**10.43**	-0.31	0.99	-0.66	-0.50	10.52	9.97	10.82	10.80
Coke, Refined Petroleum Products	**3.75**	0.17	0.91	-2.94	-12.46	2.11	1.75	3.89	11.52
Chemicals	**1.36**	2.10	1.96	0.58	-0.29	0.72	0.47	1.01	1.58
Pharmaceuticals	**4.66**	3.63	4.29	4.46	2.00	3.55	2.70	2.04	3.15
Rubber and Plastic Products	**0.28**	0.62	-0.02	-0.53	-0.54	0.09	0.29	0.59	0.68
Mineral Products	**1.54**	0.55	1.99	1.65	0.69	1.37	0.64	0.57	1.02
Basic Metals	**2.93**	2.81	2.37	1.39	1.27	2.07	1.85	2.11	1.98
Metal Products	**2.73**	0.08	0.07	0.11	-1.55	2.71	2.70	2.67	3.89
Machinery	**0.92**	0.58	0.04	-0.25	-0.68	0.74	0.90	1.06	1.43
Office Machinery	**4.23**	2.24	1.82	1.52	1.40	3.54	3.40	3.34	3.18
Electrical Machinery	**0.62**	0.50	-0.19	-0.43	-0.51	0.47	0.71	0.87	1.00
Electronic Components	**10.84**	3.47	4.56	5.24	2.80	9.77	8.76	7.76	8.73
TV and Radio	**1.36**	1.15	1.56	0.07	-0.38	1.01	0.65	1.32	1.64
Medical and Optical Instruments	**0.59**	0.81	0.93	0.91	0.47	0.35	0.17	0.06	0.24
Motor Vehicles	**6.23**	0.54	0.69	0.45	0.26	6.06	5.91	5.96	6.03
Other Transport	**0.61**	3.02	3.02	4.15	-4.42	-0.32	-0.77	-1.83	3.93
Furniture and Leisure Goods	**3.15**	6.89	4.72	3.13	1.57	1.03	1.00	1.31	1.97
Mean	**3.27**	**2.34**	**1.90**	**1.20**	**-0.43**	**2.56**	**2.41**	**2.56**	**3.59**

Notes: Decomposition of geographic concentration index for $q = 7$ and 10 years, based on (14), (15) and (16). Each index number multiplied by 100, with those for 3 and 5 years reproduced from table 2 (see note to table 2).

Appendix A: Derivation of the Frequency Estimators for p^{na}

To derive the estimator of p^{na} for the location of an activity, we first note that the total number of activity pairings within regions is $\sum_{r=1}^{r=R} n_r(n_r - 1)/2$, and that between and within regions it is $n(n - 1)/2$, so that the ratio of these gives (13).

Initially, suppose spillovers extend over a single period only, i.e. $q = 1$, so that they occur between $\tau = t - 1$ and t. Then to get p_1^{na}, from each of the above we deduct the relevant number of activity pairings occurring over a single time period and within the same period of location at time t. The number of pairings between and within times $t - 1$ and t for region r is ${}^{n_{rt}+n_{rt-1}}\mathrm{C}_2$, but excluding those at $t - 1$ it is ${}^{n_{rt}+n_{rt-1}}\mathrm{C}_2 - {}^{n_{rt-1}}\mathrm{C}_2$, so the deduction for region r is:

$$\frac{(n_{rt}+n_{rt-1})(n_{rt}+n_{rt-1}-1)}{2} - \frac{n_{rt-1}(n_{rt-1}-1)}{2} = n_{rt}n_{rt-1} + \frac{n_{rt}(n_{rt}-1)}{2}.$$

Hence, for all regions the deduction is $\sum_{r=1}^{r=R}\left[n_{rt}n_{rt-1}+n_{rt}(n_{rt}-1)/2\right]$, and over the T periods it is $\sum_{t=2}^{t=T}\sum_{r=1}^{r=R}\left[n_{rt}n_{rt-1}+n_{rt}(n_{rt}-1)/2\right]$. By the same reasoning the deduction from the denominator is $\sum_{t=2}^{t=T}\left[n_t n_{t-1}+n_t(n_t-1)/2\right]$, and hence:

$$\hat{p}_1^{na} = \frac{\sum_{r=1}^{r=R}\left\{n_r(n_r-1) - \sum_{t=2}^{t=T}\left[2n_{rt}n_{rt-1} + n_{rt}(n_{rt}-1)\right]\right\}}{n(n-1) - \sum_{t=2}^{t=T}\left[2n_t n_{t-1} + n_t(n_t-1)\right]}. \tag{A1}$$

More generally, if spillovers extend over q periods then it amounts to substituting $\sum_{\tau=1}^{\tau=q} n_{rt-\tau}$ for n_{rt-1} in (A1), and this gives $\hat{p}_q^{na}$ in (14) of the text.

The approach can be extended to consider the scale of different activities, as measured by the number of jobs, so that the more general form of the Herfindahl index H is relevant. The number of job pairings between different activities within regions is $\left(\sum_{r=1}^{r=R} s_r^2 - H\right)/2$, and between and within regions it is $(1 - H)/2$, where s_r is the share of the toal jobs in region r. The ratio of these gives the frequency estimator in (13) with $1 / n$ replaced by H. To get $\hat{p}_1^{na}$, then like above, it can be shown that the number of job pairings between activities over a single time period and within the same period at time t is $\sum_{t=2}^{t=T}\left[s_t s_{t-1} + s_t\left(s_t - 1\right)/2\right]$ and that over T periods it is $\sum_{t=2}^{t=T}\sum_{r=1}^{r=R}\left[s_{rt} s_{rt-1} + s_{rt}\left(s_{rt} - 1\right)/2\right]$, so that the expression is now:

$$\hat{p}_1^{na} = \frac{\sum_{r=1}^{r=R}\left\{s_r^2 - \sum_{t=2}^{t=T}\left[2s_{rt}s_{rt-1} + s_{rt}\left(s_{rt} - 1\right)\right]\right\} - H}{1 - \sum_{t=2}^{t=T}\left[2s_t s_{t-1} + s_t\left(s_t - 1\right)\right] - H}. \tag{A2}$$

As a check, this reduces to (A1) if activities are of equal size. This can be seen by writing $s_r = n_r / n$, noting that $H = \sum_r n_r / n^2 = 1 / n$ and likewise for the t-subscripted terms. More generally, if the spillovers extend over q periods then the expression is:

$$\hat{p}_q^{na} = \frac{\sum_{r=1}^{r=R}\left\{s_r^2 - \sum_{t=q+1}^{t=T}\left[2s_{rt}\sum_{\tau=1}^{\tau=q}s_{rt-\tau} + s_{rt}\left(s_{rt} - 1\right)\right]\right\} - H}{1 - \sum_{t=q+1}^{t=T}\left[2s_t\sum_{\tau=1}^{\tau=q}s_{t-\tau} + s_t\left(s_t - 1\right)\right] - H}. \tag{A3}$$

Following the same reasoning as above, then as a check this reduces to (14). The methodology is otherwise basically the same as in the text, but where x_r now refers to the regional share of jobs.

Appendix B: Industry Classification and Frequency Estimates

Industry	**Number of investments**	$\hat{p}$	$\hat{p}_1^{na}$	$\hat{p}_2^{na}$	$\hat{p}_3^{na}$	$\hat{p}_5^{na}$
Food, Beverages and Tobacco (15 and 16)	360	0.115	0.115	0.115	0.115	0.116
Textiles and Textile Products (17 and 18)	170	0.145	0.143	0.141	0.140	0.138
Leather and Leather Products (19)	26	0.194	0.184	0.183	0.189	0.195
Wood and Wood Products (20)	94	0.135	0.134	0.133	0.133	0.135
Pulp, Paper and Paper Products (21)	201	0.169	0.169	0.168	0.167	0.169
Publishing and Printing (22)	131	0.213	0.211	0.212	0.214	0.213
Coke, Refined Petroleum Products (23)	25	0.154	0.150	0.149	0.152	0.153
Chemicals (24, excl. 24.4)	473	0.133	0.132	0.131	0.131	0.130
Pharmaceuticals (24.4)	231	0.162	0.161	0.160	0.159	0.158
Rubber and Plastic Products (25)	418	0.123	0.123	0.123	0.123	0.123
Mineral Products (26)	155	0.134	0.134	0.134	0.134	0.132
Basic Metals (27)	193	0.147	0.146	0.145	0.145	0.144
Metal Products (28)	380	0.145	0.145	0.145	0.145	0.145
Machinery (29)	797	0.129	0.129	0.129	0.129	0.129
Office Machinery (30)	131	0.158	0.158	0.158	0.158	0.157
Electrical Machinery (31)	399	0.126	0.126	0.126	0.126	0.127
Electronic Components (32.1)	493	0.216	0.216	0.215	0.213	0.210
TV and Radio (32.2 and 32.3)	271	0.133	0.132	0.131	0.132	0.131
Medical and Optical Instruments (33)	339	0.126	0.126	0.126	0.126	0.125
Motor Vehicles (34)	779	0.176	0.175	0.175	0.175	0.175
Other Transport (35)	189	0.126	0.124	0.125	0.125	0.124
Furniture and Leisure Goods (36)	217	0.148	0.146	0.144	0.144	0.145
Mean	**294**	**0.150**	**0.149**	**0.148**	**0.149**	**0.149**

Notes: Industries defined by NACE (rev. 1), with the 2 or 3-digit code given in parentheses. Number of investments is for period 1985-05. $\hat{p}$ is calculated from (13) and $\hat{p}_q^{na}$ from (14) for q = 1, 2, 3 and 5 years.

References

Alegria, R. (2009), 'The Location of Multinational Firms in the UK: Sectoral and Functional Agglomeration', *Spatial Economics Research Centre Annual Conference*, London School of Economics, May.

Barrios, S., Bertinelli, L., Strobl, E. and Teixeira, A-C., (2005), 'The Dynamics of Agglomeration: Evidence from Ireland and Portugal', *Journal of Urban Economics*, 57.1, 170-88.

Basile, R., Castellani, D. and Zanfei, A. (2008), 'Location Choices of Multinational Firms in Europe: The Role of EU Cohesion Policy', *Journal of International Economics*, 74.2, 328-40.

Blomström, M. and Kokko, A. (1998), 'Multinational Corporations and Spillovers', *Journal of Economic Surveys*, 12, 247-77.

Brülhart, M. and Traeger, R. (2004), 'An Account of Geographic Concentration Patterns in Europe', *Regional Science and Urban Economics*, 35, 597-624.

Clarke, R. (1985), *Industrial Economics*, Blackwell, Oxford.

Crozet, M., Mayer, T. and Mucchielli, J-L. (2004), 'How Do Firms Agglomerate? A Study of FDI in France', *Regional Science and Urban Economics*, 34, 27-54.

Dimitropoulou, D., Burke, S. and McCann, P. (2006), 'The Determinants of the Location of Foreign Direct Investment in UK Regions', *Regional Science Association International: British and Irish Section Conference*, Jersey, August.

Döring, T. and Schnellenbach, J. (2006), 'What Do We Know about Geographical Knowledge Spillovers and Regional Growth? A Survey of the Literature', *Regional Studies*, 40.3, 375-95.

Dumais, G., Ellison, G. and Glaeser, E. L. (2002), 'Geographic Concentration as a Dynamic Process', *Review of Economics and Statistics*, 84.2, 193-204.

Ellison, G. and Glaeser, E. L. (1994), 'Geographic Concentration in US Manufacturing Industries: A Dartboard Approach', *NBER Working Paper no. 4840*, Cambridge MA.

Ellison, G. and Glaeser, E. L. (1997), 'Geographic Concentration in US Manufacturing Industries: A Dartboard Approach', *Journal of Political Economy*, 105.5, 889-927.

Ellison, G. and Glaeser, E. L. (1999), 'Evolution of the Geographic Concentration of Industry', *American Economic Review*, 89.2, 311-16.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A. and Shleifer, A. (1992), 'Growth in Cities', *Journal of Political Economy*, 100, 1126-52.

Head, K., Ries, J. and Swenson, D. (1995), Agglomeration Benefits and Location Choice: Evidence from Japanese Manufacturing Investments in the United States, *Journal of International Economics*, 38, 223-47.

Henderson, J. V., Kuncoro, A. and Turner, A. (1995), 'Industrial Development in Cities', *Journal of Political Economy*, 103, 1067-85.

Henderson, J. V. (1997), 'Externalities and Industrial Development', *Journal of Urban Economics*, 42, 449-70.

Howells, J. R. L. (2002), 'Tacit Knowledge, Innovation and Economic Geography', *Urban Studies*, 39, 871-84.

Jones, J. and Wren, C. (2004), 'Inward Foreign Direct Investment and Employment: A Project-Based Analysis in North East England", *Journal of Economic Geography*, 4.5, 517-44.

Jones, J. and Wren, C. (2011), 'On the Relative Importance of Agglomeration Economies in the Location of FDI across British Regions', *SERC Discussion Paper 89*, Spatial Economics Research Centre, London School of Economics, London.

Jones, J. and Wren, C. (2012), 'FDI Location across British Regions and Agglomerative Forces: A Markov Analysis', *Spatial Economic Analysis*, forthcoming.

Karlsson, C. and Johansson, B. (2005), 'Regional Development and Knowledge', *CESIS Working paper no. 76*, Stockholm.

Kim, S. (1999), 'Regions, Resources, and Economic Geography: Sources of US Regional Comparative Advantage, 1880-1987', *Regional Science and Urban Economics*, 29, 1-32

Krugman, P. (1991), *Geography and Trade*, MIT Press, Cambridge, MA.

Lafourcade, M. and Mion, G. (2007), 'Concentration, Agglomeration and the Size of Plants', *Regional Science and Urban Economics*, 37, 46-68.

Mariotti, S., Piscitello, L. and Elia, S. (2010), 'Spatial Agglomeration of Multinational Enterprises: The Role of Information Externalities and Knowledge Spillovers', *Journal of Economic Geography*, 10.4, 519-38.

Maurel, F. and Sédillot, B. (1999), 'A Measure of Geographic Concentration in French Manufacturing Industries', *Regional Science and Urban Economics*, 29, 575-604.

Mood, A. M., Graybill, F. A. and Boes, D. C. (1974), *Introduction to the Theory of Statistics*, McGraw-Hill, Singapore, Third Edition.

Rosenthal, S. and Strange, W. (2001), 'The Determinants of Agglomeration', *Journal of Urban Economics*, 50, 191-229.

Rosenthal, S. and Strange, W. (2003), ‘Geography, Industrial Organization, and Agglomeration’, *Review of Economics and Statistics*, 85, 377-93.

Rosenthal, S. and Strange, W. (2004), ‘Evidence on the Nature and Sources of Agglomeration Economies’, in J. V. Henderson and J-F. Thisse (eds), *Handbook of Regional and Urban Economics*, Volume 4, Elsevier, North Holland, Amsterdam.

Wren, C. and Jones, J. (2009), ‘Re-investment and the Survival of Foreign-Owned Plants’, *Regional Science and Urban Economics*, 39, 214-23.

Wren, C. and Jones, J. (2011), ‘Assessing the Regional Impact of Grants on FDI Location: Evidence from UK Regional Policy, 1985-05’, *Journal of Regional Science,* 51.3, 497-517.

BIS Department for Business Innovation & Skills

Communities and Local Government

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Spatial Economics Research Centre (SERC)
London School of Economics
Houghton Street
London WC2A 2AE

Tel: 020 7852 3565
Fax: 020 7955 6848
Web: www.spatialeconomics.ac.uk

SERC is an independent research centre funded by the Economic and Social Research Council (ESRC), Department for Business Innovation and Skills (BIS), the Department for Communities and Local Government (CLG) and the Welsh Assembly Government.