Titre: Title:	Firm growth in industrial clusters of the United Kingdom
Auteurs: Authors:	Catherine Beaudry, & G. M. Peter Swann
Date:	2009
Type:	Article de revue / Article
Référence: Citation:	Beaudry, C., & Swann, G. M. P. (2009). Firm growth in industrial clusters of the United Kingdom. Small Business Economics, 32 (4), 409-424. https://doi.org/10.1007/s11187-007-9083-9

Titre de la revue: Journal Title:	Small Business Economics (vol. 32, no. 4)
Maison d'édition: Publisher:	Springer
URL officiel: Official URL:	https://doi.org/10.1007/s11187-007-9083-9
Mention légale: Legal notice:	This is a post-peer-review, pre-copyedit version of an article published in Small Business Economics (vol. 32, no. 4) . The final authenticated version is available online at: https://doi.org/10.1007/s11187-007-9083-9

Firm-Level Growth in Industrial Clusters: A Bird's Eye View of the United Kingdom [1]

CATHERINE BEAUDRY[i], *École Polytechnique de Montréal*

G. M. PETER SWANN, *Nottingham University Business School*

September 2004 – Revised: May 2007 – Final version August 2007

Abstract

This paper examines how firm growth is influenced by the strength of the industrial cluster in which the firm is located. The paper presents econometric estimates of firm-level growth models for 56 two-digit industries in the UK. In about half of these industries, there is a positive and statistically significant association between firm growth and own-sector employment. Significant associations between firm growth and other-sector employment are less common, but where these arise they are generally negative. We find that a weak rule of thumb applies in the great majority of industries: own-sector effects are positive or insignificant, while other-sector effects are negative or insignificant. Cluster effects are strongest in manufacturing, manufacturing-related or in key parts of the infrastructure, but weaker in services.

JEL classification: L10, O40, R12

Keywords: Industrial Clusters, Growth, Firm Performance

---

[i] **Corresponding author**: Département de mathématiques et de génie industriel, École Polytechnique de Montréal, C.P. 6079, succ. Centre-ville, Montréal (Québec) Canada, H3C 3A7, catherine.beaudry@polymtl.ca

I. Introduction

Research on clusters in some specific industrial sectors has found evidence that firms in strong industrial clusters often grow faster than average. The objective of this paper is to explore whether this conclusion applies across all industrial sectors, or if not, which sectors enjoy such cluster effects and which do not.

To this end, the paper presents econometric estimates of firm-level growth models for a range of 56 two-digit industries in the UK. In common with earlier work (Swann, 1998; Baptista and Swann, 1999; Swann and Prevezer, 1996; Beaudry, 2001; Cook *et al*., 2001; Pandit *et al*., 2001), the strength of the cluster is measured here by own-sector employment (employment in the firm's own 2-digit sector within its county), other-sector employment (employment in all other 2-digit sectors within its county), and several other variables including employment diversity. Rosenthal and Strange (2003) refer to these own-sector and other-sector effects as localisation effects and urbanisation effects respectively.

The earlier studies cited above – which relate to computing, biotechnology, aerospace, broadcasting and financial services – found that own-sector effects tend to be positive while other-sector effects are often negative. Or, to be more precise, companies located in clusters that are strong in their own sector (ceteris paribus) tended to grow faster than average, while companies located in clusters that are strong in other sectors, but not in their own (ceteris paribus) tended to grow slower than average. In what follows we shall focus on two questions. First, does this pattern apply more generally as a rule of thumb in a wide range of sectors across the economy? Second, in which sectors do we find the strongest clustering effects?

It is well recognised, of course, that co-location on its own does not necessarily yield economic advantage. Several qualitative studies and surveys have demonstrated that co-located companies do not necessarily expect to gain much from proximity to their neighbours.

But if companies from the same sector co-located in the same region tend to grow faster than average, then that is suggestive. Research on clustering is a classic example of the virtues of using a portfolio of research methodologies. Detailed case studies will tell us a lot about the detailed processes through which clustering brings economic benefits (e.g. Saxenian, 1994). By contrast, the broad-brush econometric approach of this paper captures a "bird's-eye" view of clustering in the UK, without the detail.

This paper uses a firm-level approach to study lifetime growth of firms located within industrial clusters. Most studies have used cluster-level approaches to examine employment growth or new firm formation (see for instance Henderson *et al.* (1995); Rosenthal and Strange, 2003; and Combes, 2000). This method used by Swann *et al.* (1998) and others has shown that firms grow faster individually when located with their peers while controlling for the firm characteristics. In a sense, this type of analysis lies mid-way between detailed case studies and cluster-level approaches.

The structure of the paper is as follows. Section II describes the logic behind the simple lifetime growth model used in the paper. Section III describes the data and data sources used in this study. Section IV presents and comments on the econometric results. Section V concludes.

II. Theoretical interpretation of models

A number of authors have examined the growth of industrial sectors within clusters. For instance, Glaeser *et al.* (1992), Combes (2000) and Henderson *et al.* (2003) have all tried to explain some form or another of employment growth within industry-cluster tandems with various specialisation, concentration, competition and diversity characteristics. Our goal in this paper is not to replicate these analyses for the UK, but to examine the growth of firms within clusters as opposed to the growth of clusters.

In earlier work we have looked at three broad categories of econometric models that analyse the performance of companies in clusters. One is the lifetime growth model analysed here. Another is a model of entry into clusters (Swann, 1998). Rosenthal and Strange (2003) have used similar methodologies to analyse the birth of new firms within clusters and the employment within these new firms. And the third is the analysis of innovation or patenting in clusters (Baptista and Swann, 1998; Beaudry and Breschi, 2003).

The econometric approach taken in this paper tries to identify whether firms located in strong clusters (with strong industry and/or a strong science base) grow faster than isolated firms. In its simplest form, a model of the lifetime growth of the firm can be estimated using employment as a measure of cluster strength. In the model, the trend growth of the firm is treated as a function of cluster employment – both in the firm's own sector, and in all other sectors (both in terms of scale and diversity measures). The model also takes account of any potential effects of a strong science base and other regional and sectorial "fixed effects" on the trend rate of growth of a firm. Accordingly, the model estimates a trend rate of growth, but also makes allowance for the possibility that growth may be influenced by clustering with similar firms, with dissimilar firms, or near the science base.

Figure 1 introduces the lifetime growth model in a very simple way. The graph shows (hypothetical) growth paths for employment in one stereotypical firm located in a cluster and another located outside a cluster. Because the vertical axis is on a log scale, the concave shape of the growth path implies that growth rates are high when the firm is young but tail off thereafter. This is to be expected, since small firms can achieve very high growth rates from a small base that could not be sustained when the firm gets larger. The growth path for the firm in a cluster lies above that for the isolated firm. The rate of growth at the start is greater and although growth rates are similar in maturity, the firm located in a cluster achieves a higher absolute size.

Suppose that all firms grow along these growth paths, and that we take a sample of firms at a particular date. Suppose also that the census we use only records firms when they reach a certain age or a certain size – this is true of the *FAME* database used for our work. More precisely, suppose that this census only records firms when they progress to the right of the line SS. Then the sample of firms from inside a cluster will lie along the upper growth path to the right of SS and the sample from outside a cluster will lie along the lower growth path to the right of SS.

Figure 1

Patterns of Employment Growth, for firms inside and outside clusters

In this case, we can see that if we fit two straight lines to these data - one for the cluster and the other for outside the cluster - then these are roughly parallel to each other. This is slightly unexpected, since the actual growth paths are certainly not parallel to each other. But it happens because the data represent an incomplete census of firms. If we estimate a lifetime growth model, where log size is regressed on age, then the cluster effect may show up as a *higher intercept* rather than a *steeper slope*. This was what we found in several previous studies (Swann, 1998).

What sort of sample selection bias will arise from this inevitable absence of small and new firms? It means that the fitted lines shown in Figure 1 have a flatter slope but a higher intercept than they would if we had complete data on all firms. And, as we measure the cluster effects by the intercept rather than the slope, this sample selection biases the cluster effects upwards.

The traditional lifetime growth model first estimated for this paper is based on this simple construct. The format is as follows:

$$\ln e_{n\in\{I:c\}} = \alpha + \beta\, age_n + \gamma_1 \ln\left[\sum_{i\in\{I:c\}} e_i\right] + \gamma_2 \ln\left[\sum_{j\in\{\neg I:c\}} e_j\right] + \sum_v \delta_v \ln V_v + u \qquad \textbf{(1)}$$

$$\beta = 1 + \sum_{c=1}^{C-1} d_c D_c + \sum_{i=1}^{I} d_i D_i \qquad \textbf{(2)}$$

where:

$e_{n\in\{I:c\}}$ is employment in firm *n* from sector *i* at cluster *c*; [*CieEmp*]

age_n is the age of firm *n*; [*Age*]

$\sum e_{i\in\{I:c\}}$ is total employment in sector *I* in cluster *c*; [*OwnEmp*]

$\sum e_{j\in\{\neg I:c\}}$ is total employment in all sectors other than *I* at cluster *c*; [*OthEmp*]

V_v are other cluster strength variables;
D_c represents cluster dummy variables;
D_i corresponds to the sectorial dummy variables.
u is a disturbance term.

The variables used in this lifetime growth analysis are described in section III below. Probably the two most important shift variables in the model are what we shall call *own-sector* [*OwnEmp*] and *other-sector* [*OthEmp*] employment in that cluster. The first shift variable, own-sector, represents the extent of the localisation, or Marshall-Arrow-Romer (MAR), externalities, and is measured by the total number of employees in the same sector as firm *n* and located in the same cluster *c* as the firm. The second shift variable, other-sector, represents the scale of urbanisation, or Jacobs, externalities, and is measured by the total number of employees in all sectors except that of firm *n* and located in the same cluster *c* as the firm. The diversity aspect of urbanisation externalities is measured using a standard employment Herfindahl index[2] [*Emp2Herf*].

Thus the model estimates the trend rate of growth (β), the coefficient of the age of the firm [*Age*], but also makes allowances for the fact that growth may be influenced by the presence of similar firms (γ_1) i.e. from the same sector [*OwnEmp*], or of other firms (γ_2), i.e.

from other sectors [*OthEmp*]. In principle, $\gamma_2 \ln e_{jc}$ can be replaced by a sum of effects, one for each sector at cluster c, but given the likely collinearities, this was thought impractical. The coefficients of the growth rate dummy variables, d_c and d_i, allow us to measure whether companies in particular clusters and sectors grow faster than in others.

As indicated above, the dependent variable ($\ln e_n$) [ln *CieEmp*] is not limited in its range, so OLS can safely be applied to equation (1). Standard tests for the normality of the residuals, however, often reject OLS so we only report the negative binomial regression summarised results (the full results are provided in an unpublished working paper available upon request from the authors):

$$E\left[e_{n\in\{I:c\}} \middle| X_{nc}\right] = \lambda_n = \exp\left(\alpha + \beta\, age_n + \gamma_1 \ln\left[\sum_{i\in\{I:c\}} e_i\right] + \gamma_2 \ln\left[\sum_{j\in\{\neg I:c\}} e_j\right] + \sum_v \delta_v \ln V_v\right) \quad \textbf{(3)}$$

where the expected number of employees of a firm is an exponential function of its age, of the extent of the own sector and other sectors employment as well as other cluster strength and firm specific variables, and where β is defined as in equation (2).

Immediately, there are some problems with this approach. Three of the most important are as follows. First, it is a simple model of organic growth. But how does it cope with non-organic growth: with mergers, acquisitions and dispersals? If such companies are added to Figure 1 they will distort the picture. The data sources used indicate whether a company is a holding company, whether it has subsidiaries or whether it is a subsidiary. It also indicates whether accounts are consolidated. Accordingly, the approach taken here is simply to make a crude adjustment using dummy variables. As far as possible we have tried to avoid double counting where a subsidiary is both treated as a firm in its own right and as a consolidated part of a larger conglomerate. Wherever possible we work only with the smaller entity. Only 11% of firms are consolidated and a minority of their subsidiaries appear in the database because they do not file employment figures for their subsidiaries, so these companies are *de facto* eliminated from our database.[3, 4]

The second problem is that of heteroskedasticity. A quick glance at Figure 1 suggests that we would not necessarily expect the variance of actual (log) firm size around the growth path to be constant. We might expect this variance to increase with the age of the firm – and indeed the econometric work does find this. We have explored whether the heteroskedasticity can be modelled adequately by assuming that variance is proportional to the square of age, $\mathrm{var}(u) = \sigma^2 Age^2$, in which case we would work with a simple transformation of equation (1). However, we found that this approach does not capture the character of the heteroskedasticity. Various other simple functional forms were examined in an attempt to model the heteroskedasticity of the data without success. The form of the heteroskedasticity is more complex. Accordingly, we report robust standard errors (White, 1980) to allow for this heteroskedasticity.

The third problem is that of endogeneity. If own-sector employment is used as an explanatory variable in the lifetime growth model, we have a situation where the LHS variable $\ln\left(e_{n\in\{I:c\}}\right)$ is a part of an aggregate variable on the right hand side: $\ln\left(\sum e_{i\in\{I:c\}}\right)$ as shown below:

$$\ln\left[\sum_{i\in\{I:c\}} e_i\right] = \ln\left[e_{n\in\{I:c\}} + \sum_{\substack{i\in\{I:c\}\\ i\neq n}} e_i\right]$$

This endogeneity itself leads to two problems. First, if the focus of our interest is the effect of employment in *all other companies*, then the coefficient γ_I is an overestimate, because the right-hand side variable for aggregate own-sector employment includes the dependent variable for company employment. The second problem is a potential simultaneity bias. This arises because the fact that the dependent variable is included in the aggregate employment measure means that the disturbance term in equation (1) cannot be independent of the own-sector employment aggregate.

One strategy for dealing with these econometric issues would be to create an aggregate for own-sector employment *in all other companies.* However that is not as easy as it sounds. To subtract company employment from aggregate county employment only makes sense if companies operate from a single plant. That may be true for a large number of small companies, but these companies (accounting for only small employee figures) are unlikely to introduce significant problems. The real problems derive from companies with large employment figures, but these are most likely not to be single plant companies. To subtract these company employment figures (spread over many plants in several regions) from employment in one region makes no sense at all, and would introduce serious problems.

The endogeneity issue was examined using two-stage least squares and comparing it to the OLS results using a Hausman test as well as a Davidson and MacKinnon (1993) augmented regression test. The latter involves replacing the potentially endogenous right-hand-side variables with the predicted values of those right-hand-side variables, where the predicted values are a function of all exogenous variables. This test identified a problem of endogeneity in 9 of the 56 industries under consideration (SIC codes 21, 22, 32, 36, 45, 51, 52, 55 and 74). In all of these industries, there was also evidence of non-normality of the residuals. Results will thus be treated with the appropriate care.

The reader may wonder why we use a single census in this lifetime growth model rather than looking at year to year growth rates, which would avoid some of these problems. There are two reasons. First, as much of the literature has demonstrated, year-to-year company growth is more or less random – this is an old idea, dating back to the work of Hart and Prais (1956). Accordingly it would be difficult to make much headway in trying to model the factors that influence differences in year-to-year growth rates, because this is so volatile and so unpredictable. Second, while we can build up reasonable time series data on many companies, it is harder for some of the smaller companies in our sample. So while we could look at longer-period growth rates for some, we cannot do this for all - especially the smaller companies in our sample. And since small firms seem to be most dependent on clusters for superior performance, then it is especially important to include them in our sample. So while this lifetime growth model is a rough and ready method it is a natural approach to take in building up this bird's eye view.

III. Data

Databases of regional data and company data were merged to construct the data used in this study. Regional data was obtained from the Office of National Statistics (ONS) Regional Trends database and from their data service. The *FAME* database from Bureau Van Dijk was used to extract the UK company data required for this study. From these data sources, four categories of variables were identified: employment, representing the size of the industry; firm type, categorising a company amongst its peers; economic data, illustrating cluster economic strength; and regional characteristics, showing general cluster strength. From the different databases available for this study, ten variables were extracted. Table 1 summarises the variables used in the lifetime growth model and Table 2 presents basic statistics for these. In the lifetime growth model, the left-hand side variable is the number of firm employees [*CieEmp*].

TABLE 1

Description of Variables Used in Model

TABLE 2

Elementary Statistics

The source of company data used in this study, FAME, gives data at the firm level rather than at the establishment level. The use of firm data raises some issues that need to be discussed here. How is a firm allocated to a region? Do any biases or inaccuracies arise from the use of firm data or the allocation of a multi-plant firm to one single region? Is it better to use data by establishment or by firm?

How is a firm allocated to a region? We allocate each firm to the region in which its headquarters are based. This means that the value of the cluster strength variable relating to a specific firm is measured as the level of employment in the region where that company's headquarters are based. In the case of single establishment firms, this allocation process seems unproblematic. But in the case of larger firms with plants in several locations (including some overseas) this may seem more questionable. In part, of course, the reason for our allocation process is a pragmatic one. FAME only indicates the address of company headquarters: it neither gives a list of other plants nor an indication of what proportion of activity is located in each place.

Do any biases or inaccuracies arise from the use of firm data or the allocation of multi-plant firm to one single region? In the case of a multi-plant firm, the cluster strength at its headquarters is not necessarily an accurate measure of the extent to which the firm as a whole benefits from a cluster. It might seem more appropriate to compute some sort of weighted average across all the regions in which the company is located, though for pragmatic reasons (described above) that is not possible. From this point of view, the use of a cluster strength variable relating to company headquarters rather than this weighted average could be seen as a problem of errors-in-variables. And as is well known, errors-in-variables does generally lead to biases in estimated parameters. In simple bivariate models this bias is towards zero, though in multivariate models the bias can go in either direction, depending on the detailed correlation structure of the explanatory variables.

We can, however, identify one assumption under which there is no such bias from allocating the firm to the region in which its headquarters are located. Suppose we can divide the firm's activities into those that are location-critical and those that are not. The first group contains activities that benefit from location in a cluster while the second group contains other activities that do not benefit from the cluster. If we assume that the firm conducts all location-critical activities at or near its headquarters, then the extent to which the company benefits from location in a cluster depends only on the strength of the cluster in which the company's headquarters are located. By contrast, the strength of the clusters in which other activities are located is not important. In this case, the appropriate measure of cluster strength to use in our analysis is the strength of the cluster in which the firm's headquarters are located – and not a weighted average. In this case, therefore, there would be no bias from our process of allocating the firm to one region.

Is the above assumption a reasonable one? In some case studies exploring the reasons why firms wish to locate their offices in strong clusters, Paton *et al*. (2007, Chapter 6) found that it is often essential to locate key R&D activities, or other activities involving high-level strategic interaction with customers and suppliers, at the company headquarters within a strong cluster. More generally, it is commonly found that a company's domestic R&D, at least, tends to be located at or near company headquarters (Howells, 1984, 1990). Moreover, interaction with customers and suppliers in a cluster is essential for competitive success in a wide range of industries (von Hippel, 1988). So there does seem to be some support for this assumption.

On the other hand, if the above simplifying assumption is not valid, and it is not the case that all location-critical activities take place at or near company headquarters, then a bias can arise. It is hard to generalise about the direction of this bias. If, as is the case for many large companies, their headquarters are located in London or the South-East of England while its plants are spread across the country, then the cluster strength variable that we use would exceed the ideal weighted average described above. This means that our measured explanatory variable sometimes over-estimates the relevant measure of cluster strength. Simple intuition might suggest this upward measurement error would bias the associated parameter towards zero but the realities of measurement error bias in a multivariate context are more complex and preclude such a generalisation.

In view of the above issues, would it be better to use establishment data rather than firm data in a model of this sort? In our view, the answer to that is "no". There is admittedly no problem of how to allocate an establishment to a region. However, to treat the establishment as the basic unit of analysis in our growth model would create even worse problems, because that would overlook the essential interdependencies within a firm. For example, suppose a firm has two establishments: the headquarters, located in a strong cluster, where all the location-critical activities go on; and a production plant, located in another region without a strong cluster. We cannot understand the growth of employment at the production plant by reference to the characteristics of its region alone. The growth of employment there will be highly dependent on the success of activities at headquarters, and to understand that we need to take account of the characteristics of the cluster in which the headquarters are located.

From the company database, FAME, employment for 136,304 firms was extracted. This database contains many more firms, but not all company records contain employment data for 1998. When employment in 1998 was missing, the figures for 1997 or 1999 were used as a proxy. Turnover and some other variables were considered as alternative measures of firm size, but these were rejected for a variety of reasons, including the problem that the number of usable observations available from the FAME database would drop considerably. Section II noted that small and very young companies are excluded from the database. This sample selection bias is likely to bias estimates of cluster effects upwards and that needs to be born in mind when interpreting the results. The region within which the firm operates as well as the main sector of operation were also added to the database as dummy variables, [D_c] and [D_i] at the one-digit and two-digit levels.

To provide a time frame in order to approximate firm growth, the age of the firm in 1999 was calculated from the year of incorporation [*Age*]. We are aware that firms that have changed their name or have been purchased by other firms might have a different age from the value calculated using the year of incorporation, but this represents a minority of companies in this database. So without a more accurate account of the age of the firm, the age from the year of incorporation to 1999 was used in this study. As mentioned before, the status of the company is also of interest and will be characterised by three dummy variables: whether the firm is consolidated or not, [*Dcons*], if it is a holding company, [*DHold*] and whether it is a subsidiary, [*Dsubs*].

As Table 2 shows, the average firm has been in business for more than 20 years and employs more than 180 people. Roughly 11% of firms in the database file consolidated accounts, 20% are holding companies, and 41% are subsidiaries. As can be deduced from the standard deviation, the type of firms included in the study, and certainly their size and age, are highly diverse.

Regional economic strength is measured by employment characteristics which are computed at NUTS[5] level 3 for each two-digit UK SIC code (rev. 1992).[6, 7] Regional employment figures for 1998 were obtained from the UK Annual Employment Survey. These employment figures were utilised to construct three regional variables at the NUTS level 3: own-sector employment [*OwnEmp*], other-sector employment [*OthEmp*], and a measure of employment diversity [*Emp2Herf*]. Own-sector employment corresponds to the number of employees in the county where firm *n* operates and are working in the same sector as firm *n*. Other-sector employment in that case represents all other employees in that county (the total working population of the county less employment within the same sector as firm *n*). Employment diversity for each county is measured by a Herfindahl index based on the variety of employees in different sectors. This variable will help determine whether a diverse environment is beneficial to firm growth.

In this paper, *OwnEmp* will represent the extent of the localisation externalities, while urbanisation externalities, which refer to the scale and diversity of the local environment will be split into two variables: *OthEmp*, for the scale of urbanisation effects, and *Emp2Herf*, for the diversity aspect. In essence, we use other-sector effects as measures of the scale of urbanization externalities, as in Rosenthal and Strange (2003), and a Herfindahl index of employment as the diversity measure of Jacobs externalities, as in Henderson *et al.* (1995) and Glaeser et al. (1992).

A second source of regional data consists of a measure of expenditure on research and development [*R&D*] in each region of the UK in 1997. These data were provided by the Office of National Statistics data service. Unfortunately in this case, data is only available at the NUTS level 2 (36 groups of counties) because of the way it is collected.[8] The database provides separate estimates of research and development expenditures emanating from business [*R&DBus*], public [*R&DGov*] and higher education [*R&DHE*] institutions. For the majority of counties, business research and development is the main constituent of total R&D expenditure. Indeed, for the entire country, business R&D expenditure represents a share of 68.5%, Government, 11% and Higher Education, 20.5% (1997 data). Population density was also introduced in the analysis to control for a scale effect. Data for the UK population in 1997 at the NUTS level 3 was obtained from the Office of National Statistics [9, 10].

IV. Results

We estimated equations (1), OLS, and (3), negative binomial, for 56 two-digit industries. In the interests of brevity, we shall only report here a summary of the main parameters of interest. The complete regression results (both OLS and negative binomial) are available in an unpublished working paper available upon request from the authors.

Our principal interest in this paper is to examine the effects of localisation (own-sector employment) and urbanisation (other-sector employment and diversity) on firm growth. These effects are summarised in Figures 2 and 3. Figure 2 shows a scatter plot of the estimated regression parameters for own-sector employment and other-sector employment, while Figure 3 shows a scatter plot of the estimated regression parameters for own-sector employment and the Herfindahl index of employment diversity.

In Figure 2, the majority of the sub-sectors lie in the bottom right quadrant where the coefficient of own-sector employment is *positive* and that of other-sector employment is *negative.* When the effect of own-sector employment on growth is positive, this implies that a firm located in a cluster that is strong in its own industry has a tendency to grow faster than a firm that is not surrounded by its peers. Conversely, when the effect of other-sector employment on growth is negative, this means that a firm located in a district that is dominated by employment in other sectors will tend to grow slower than average. This negative effect may be due to congestion and competition in overcrowded clusters; in other words, the urbanisation externalities have a detrimental effect on firm growth.

Figure 2

Coefficients on Own-Sector and Other-Sector Employment

Two-digit industry regressions

Figure 3

Coefficients on Own-Sector Employment and Employment Herfindahl

Two-digit industry regressions

Turning to Figure 3, the coefficient on the Herfindahl index is negative in over 64% of cases (36 out of 56). This means that a more diverse employment environment (low Herfindahl) has a positive effect on firm growth. This is in accord with Audretsch and Feldman's (1999) finding that diversity matters more than specialisation in industrial clusters. Figure 3 also shows a large number of industries in the bottom right quadrant where own-sector employment has a *positive effect* and diversity has a *negative effect*. In this quadrant, firms that are co-located with others from the same industry but in a generally diverse employment environment will grow faster than the rest.

Figures 2 and 3 show the absolute magnitudes of estimated parameter coefficients. But to draw reliable conclusions from our regression results, we need to consider not just the absolute magnitude of coefficients but also their *statistical significance*. This is done in Tables 3 and 4. Both contain 9 cells, in three rows and three columns. The columns refer to the sign and significance of the own-sector coefficients, while the rows refer to the sign and significance of the other-sector coefficients (Table 3) or of the employment Herfindahl coefficients (Table 4).

Turning first to Table 3, the left-hand column shows sectors for which the own-sector regression coefficient is negative and significantly different from zero at the 10% level. The middle column shows all the sectors for which the own-sector regression coefficient is not significantly different from zero at the 10% level. The right-hand column shows sectors for which the own-sector regression coefficient is positive and significantly different from zero at the 10% level. Turning to the rows, the top row shows sectors for which the other-sector regression coefficient is positive and significantly different from zero at the 10% level. The middle row shows all the sectors for which the other-sector regression coefficient is not significantly different from zero at the 10% level. The bottom row shows sectors for which the other-sector regression coefficient is negative and significantly different from zero at the 10% level.

What does Table 3 tell us? Starting with the own-sector effects, we find that in slightly under a half of the industries (23 out of 56, or 41%), there is a positive and statistically significant association between firm growth and own-sector employment. We find that the greater number of these are manufacturing sectors. In most of the other industries (a further 29 out of 56, or about 52%), there is no significant association. In a few (4 out of 56, or about 7%) there is a significant but negative association. Turning next to the other-sector effects, we find that in almost two thirds of industries (36 out of 56, or 64%), there is no significant association between firm growth and other-sector employment. But in a majority of sectors where there is a significant association, it is a negative one (13 out of 56, or 23%) rather than a positive one (7 out of 56, or 12%).

TABLE 3

Summary of Own Sector (Localization) and Other Sector (Urbanization) Effects, by Two-Digits Industrial Sector

TABLE 4

Summary of Own Sector (Localization) and Diversity (Urbanization) Effects, by Two-Digits Industrial Sector

Turning now to Table 4, we find a slightly smaller proportion of non-significant associations between firm growth and diversity of employment (35 out of 56, or 62%). In nine sectors (16%), there is a positive and significant association between firm growth and the Herfindahl index of employment diversity – meaning that firms grow faster when employment is *less* diverse in a cluster. On the other hand, in 12 sectors (21%) there is a negative and significant association between firm growth and the Herfindahl index - meaning that firms grow faster when there is *more* employment diversity.

Two Questions

We said in the introduction that we would focus on two main questions in this study. First, does the pattern of positive own-sector effects and negative other-sector effects found in some earlier studies apply more generally as a *rule of thumb* over a wide range of sectors? Second, which are the sectors with the strongest clustering effects? Let us address these two questions in turn.

We referred to a rule of thumb from earlier studies. We can state this in two ways. The *strong* rule of thumb is that own-sector effects are significant and positive, while other-sector effects are significant and negative. The *weak* rule of thumb is that own-sector effects are significant and positive *or insignificant*, while other-sector effects are significant and negative *or insignificant*.

Table 3 shows that the *strong rule of thumb* only applies in a minority of industries. In only 11 out of 56 (or 20% of) sectors are own-sector effects positive and significant while other-sector effects are negative and significant But the *weak rule of thumb* applies in the great majority of industries. In 47 out of 56 industries (about 84%) we find that *own-sector* effects are positive or insignificant, while *other-sector* effects are negative or insignificant.

In answer to this first question, therefore, we can say that the results of earlier studies do indeed apply over a wide range of sectors. There are 8 exceptions, and these are essentially service industries. This result is consistent with that of Combes (2000) which identified five service sectors that benefited from a larger local economy (density measured by the ratio of total local employment to the local area, a variable closely related to our *other-sector* variable).

Next, we turn to the second question: where are the strongest cluster effects? From Table 3 and 4, we can identify those sectors where the own-sector clustering effect is positive and significant at the 10 % level. Many of these are *manufacturing* sectors (e.g. textiles, clothing; leather; wood products; rubber and plastic; non-metallic mineral products; office machinery and computers; radio, television and communication equipment; motor vehicles, trailers and semi-trailers; other transport equipment). Many of these industries are also identified in Henderson *et al.* (1995) as benefiting from localisation externalities. Others are *manufacturing-related* (agriculture; mining; construction; extraction of crude petroleum and natural gas, and related services), or relate to key parts of the *infrastructure* (electricity, gas, steam and hot water supply; air transport, water transport; education). Rather fewer are in *services* (retail trade; insurance and pension funding; activities auxiliary to financial intermediation; other business activities).

Negative own-sector effects are found in a few service industries: recycling; auxiliary transport activities; hotels and restaurants; and real estate activities. In two of these (hotels and restaurants, real estate), the other-sector effects are positive and significant, suggesting that companies in such sectors benefit from location in a general purpose cluster with companies from other industries, but not from co-location with others in their own industry.

In contrast to the results of Henderson *at al.* (1995) for the high-tech sectors, we find that other-sector effects (a measure of Jacobs or urbanisation externalities) do not play an important role in the development of high-tech sectors, while the own-sector effects (localisation externalities) are generally positive.

In general, when we compare the two aspects (size and diversity) of urbanisation externalities, we find that in most industries one effect is significant when the other is not. In short, while it is quite common to find that one measure of Jacobs externalities or another is important, it is rare to find that both are significant at the same time (only 5 out of 56, or 9%).

Effects of Other Variables

While the effects of the other variables in equation (1) are not our main concern here, a brief summary is useful. The coefficient on age in the regression should be interpreted as the trend rate of growth for the company - see Figure 1. For all but 4 of the regressions, this trend rate of growth is positive. The median rate of growth is one per cent per annum. This may seem rather low, but remember that because of the sample selection characteristics mentioned above (Figure 1), this is the trend rate of growth amongst older companies.

Population density was introduced in the model to take account of a scale effect - the hypothesis that companies in more *densely* populated areas grow faster. This variable has a positive coefficient in about 64% of regressions (36 out of 56).

The sign of coefficients for the R&D variables depend on what type of R&D is involved. For Business R&D, only 36% are positive. For Government R&D, 57% are positive. But for Higher Education R&D, 54% are positive. These differences are in line with observations by Dasgupta and David (1994), who point out that public research is organised to disseminate and spill over, while private R&D is not. The positive and significant coefficients in sector 0 (agriculture) are not surprising since rates of return to public-based agricultural research are known to be high (e.g. Griliches, 1992).

For all three dummy variables, the vast majority of estimated parameters are positive. As expected, consolidated firms appear to grow faster than average. Being part of a greater organisation, either as a subsidiary or as a holding company, has a lesser - but still positive - effect on firm growth.

V. Conclusions

The main objective of this paper was to take a bird's eye view of clustering in the UK, and examine whether companies located in strong clusters (or districts) performed better than average. In particular our focus is on the two questions set out in the introduction. First, do the results from earlier studies of computing, biotechnology, aerospace and financial services generalise to a broad range of industries. Specifically, do we continue to find that own-sector effects are positive while other-sector effects are negative? Second, which sectors show the strongest clustering effects?

Starting with first question, we find that in slightly under half of the industries, there is a positive and statistically significant association between firm growth and *own-sector employment*. In most of the other industries, there is no significant association, though in a few there is a significant but negative association. Turning to the other-sector effects, we find that in almost two thirds of industries, there is no significant association between firm growth and other-sector employment. But in those remaining sectors where there is a significant association, it is generally a negative one.

We showed that a *weak rule of thumb* applies in the great majority of industries. In most industries we find that *own-sector* effects are positive or insignificant, while *other-sector* effects are negative or insignificant. In short, if there are any significant clustering effects, this is the form they generally take.

A positive own-sector employment effect can be interpreted as evidence of localisation (or Marshall-Arrow-Romer) externalities. In contrast, a positive other-sector employment effect can be interpreted as evidence of urbanisation scale (or Jacobs) externalities. But in addition, it is common to use diversity measures to test for urbanisation effects. In this paper, we have therefore taken two different approaches to measuring urbanisation externalities: a scale measure (the other-sector effect) and a diversity measure (the Herfindahl index). In general, the scale measure of urbanisation appears to be detrimental to firm growth, while the diversity of employment measure appears to be beneficial for firm growth.

Turning to our second question, where are the strongest cluster effects? We saw that many of these are to be found in *manufacturing* or *manufacturing-related* industries, or in key parts of the *infrastructure*, but fewer are in *services.* By contrast, negative own-sector cluster effects are found in a few *service* industries.

In conclusion, we can draw out one implication of these findings for policy. In the UK, the development of successful industry clusters is seen as an important part of industrial policy (DTI 1999, 2001; DETR 2000). Our results suggest that in slightly under half of the industries considered, companies co-located with others from their own sector tend to grow faster than average. We could describe these as industries that enjoy beneficial cluster effects. As noted already, they tend to be concentrated in manufacturing or key parts of the infrastructure. Fewer service industries enjoy these beneficial cluster effects, and they are concentrated in utilities, transportation, retail and financial services. The rationale for cluster policy would appear strongest when it is focussed on those industries that enjoy such cluster effects.

References

Audretsch, D. and Feldman, M., 1999, 'Innovation in Cities: Science-based Diversity, Specialisation, and Localised Competition', *European Economic Review* **43**, 409-29.

Baptista, R. and Swann, G. M. P., 1998, 'Do Firms in Clusters Innovate More?', *Research Policy* **27**, 525-40.

Baptista, R. and Swann, G. M. P., 1999, 'The Dynamics of Firm Growth and Entry in Industrial Clusters: A Comparison of the US and UK Computer Industries', *Journal of Evolutionary Economics* **9** (3), 373-99.

Beaudry, C. and Breschi, S., 2003, 'On clustering of innovative firms: empirical evidence from Italy and the UK'. *Economics of Innovation and New Technology* **12** (4), 325-42.

Beaudry, C., 2001, 'Entry, Growth and Patenting in Industrial Clusters: A Study of the Aerospace Industry in the UK', *International Journal of the Economics of Business* **8** (3), 405-436.

Combes, P.-P., 2000, 'Economic Structure and Local Growth: France, 1984-1993', *Journal of Urban Economics* **47**, 329-355.

Cook G., Pandit, N. and Swann, G. M. P., 2001, 'The Dynamics of Industrial Clustering in UK Broadcasting', *Information Economics and Policy* **13**, 351-375.

Dasgupta, P. and David, P. A., 1994, 'Towards a New Economics of Science', *Research Policy* **23**, 487-521.

DETR, 2000, *Planning for Clusters: A Research Report,* DETR (June). www.communities.gov.uk/pub/674/PlanningforClustersAResearchReport_id1144674.pdf

DTI, 1999, *Biotechnology Clusters: Report of a Team led by Lord Sainsbury, Minister for Science*, DTI (August). www.dti.gov.uk/sectors/biotech/page28453.html

DTI, 2001, *Business Clusters in the UK - A First Assessment*, prepared by Trends Business Research for the DTI (February). www.dti.gov.uk/regional/clusters/clusters-assessment/page17380.html

Ellison, G. and Glaeser, E. L., 1997, 'Geographic Concentration in the U.S. Manufacturing Industries: A Dartboard Approach', *Journal of Political Economy* **105** (5), 889-927.

Glaeser, E. L., Kallal, H. D., Scheinkman, J., Shleifer, J. and A., 1992, 'Growth in Cities', *Journal of Political Economy* **100** (6), 1126-1152.

Griliches, Z., 1992, 'The Search for R&D Spillovers', *Scandinavian Journal of Economics* **94** (S), 29-47.

Hart, P. E. and Prais, S. J., 1956, 'The Analysis of Business Concentration: A Statistical Approach', *Journal of the Royal Statistical Society (series A)*, 119-150.

Henderson, V., Kuncoro, A. and Turner, M., 1995, 'Industrial Development of Cities', *Journal of Political Economy* **103** (5), 1067-1090.

Howells J. (1990) "The Location and Organisation or Research and Development: New Horizons", *Research Policy* **19**, 133-146.

Howells, J., 1984, 'The Location of Research and Development: Some Observations and Evidence from Britain', *Regional Studies* **18**, 13-29.

Howells, J., 1990, 'The Location and Organisation of Research and Development: New Horizons', *Research Policy* **19**, 133-46.

Jaffe, A. B., Trajtenberg, M. and Henderson, R., 1993, 'Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations', *Quarterly Journal of Economics* **108**, 577-598.

MacKinnon, J. G. and White, H., 1985, 'Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties', *Journal of Econometrics* **29**, 305-325.

Pandit N., Cook, G. and Swann, G. M. P., 2001, 'The Dynamics of Industrial Clustering in UK Financial Services', *The Services Industry Journal* **21** (4), 33-60.

Paton, D., Swann, G. M. P., Thompson, S., Girma, S., Asmild, M., Hanley, A. (2007) *Competitiveness in the UK Electronics Sector*, Report for the UK Department of Trade and Industry: Industry Economics and Statistics Division, May 2007.

Rosenthal, S. S. and Strange, W. C., 2003, 'Geography, Industrial Organization, and Agglomeration', *Review of Economics and Statistics* **85** (2), 377-393.

Swann, G. M. P. and Prevezer, M., 1996, 'A Comparison of the Dynamics of Industrial Clustering in Computing and Biotechnology', *Research Policy* **25**, 1139-1157.

Swann, G. M. P., Prevezer, M. and Stout, D., eds., 1998, *The Dynamics of Industrial Clusters: International Comparisons in Computing and Biotechnology*, Oxford University Press: Oxford.

von Hippel E. (1988) *The Sources of Innovation*, Oxford: Oxford University Press

TABLE 1

Description of Variables Used in Model

Variable	***Units***	***Description***		***Source***	***Year***
CieEmp	Units	Number of employees of each firm	[e_n]	Bureau van Dijk FAME database	1998
Age	Units	Number of years since the incorporation of the firm	[age_n]	Bureau van Dijk FAME database	1998
DCons	Discrete	Dummy variable taking the value 1 if the company files consolidated accounts, and 0 otherwise		Bureau van Dijk FAME database	1998
DHold	Discrete	Dummy variable taking the value 1 if the firm is a holding company, and 0 otherwise		Bureau van Dijk FAME database	1998
DSubs	Discrete	Dummy variable taking the value 1 if the company is a subsidiary, and 0 otherwise		Bureau van Dijk FAME database	1998
OwnEmp	Units	Total employment in own two-digit UK SIC (rev. 1992) industrial sector per NUTS 3 region	[e_{ic}]	ONS, UK Annual Employment Survey	1998
OthEmp	Units	Total employment in other two-digit UK SIC (rev. 1992) industrial sectors per NUTS 3 region, i.e. without *own employment*	[e_{jc}]	ONS, UK Annual Employment Survey	1998
Emp2Herf	Index	Employment diversity measured by a Herfindahl index of two-digit UK SIC (rev. 1992) industrial sectors per NUTS 3 region		ONS, UK Annual Employment Survey	1998
R&D	Millions	Total expenditure in research and development per NUTS 2 region		ONS	1997
R&DBus	Millions	Expenditure in *private* research and development per NUTS 2 region	[V_v]	ONS	1997
R&DGov	Millions	Expenditure in *public* research and development per NUTS 2 region		ONS	1997
R&DHE	Millions	Expenditure in *higher education* research and development per NUTS 2 region		ONS	1997
PopDens	Inhabitants/km^2	Population density per NUTS 3 region		ONS	1997
D_i	Discrete	Sector dummy variables, for each 60 two-digit UK SIC (rev. 1992) codes	[D_i]	ONS, UK Annual Employment Survey	1998
D_c	Discrete	Cluster dummy variables, for each NUTS 3 region	[D_c]	Bureau van Dijk FAME database	1998

Note : The means of the cluster variables such as *OwnEmp*, *OthEmp*, *Emp2Herf*, *PopDens*, *R&D*, *R&DBus*, *R&DGov* and *R&DHE* were weighted using the number of firms in each cluster.

TABLE 2

Elementary Statistics

Variable	***n***	***Mean***	***Standard Deviation***	***Minimum***	***Maximum***
CieEmp	137816	186	2020	1	185580
Age	137816	20.25	17.83	0	99
DCons	137816	0.1148	0.3188	0	1
Dhold	137816	0.2046	0.4034	0	1
Dsubs	137816	0.4018	0.4903	0	1
OwnEmp	137816	73161	145620	1	640326
OthEmp	137816	1259081	1327576	24406	3614792
Emp2Herf	137816	0.0615	0.0076	0.0470	0.0785
R&D	137816	811.85	531.26	6.17	1514.08
R&DBus	137816	470.42	273.87	2.59	978.41
R&DGov	137816	81.28	66.40	0.01	260.80
R&DHE	137816	260.15	293.40	0	729
PopDens	137816	1730.78	1812.08	7.83	4513.42

TABLE 3

Summary of Own Sector (Localization) and Other Sector (Urbanization) Effects

by Two-Digit Industrial Sector

		Coefficient on Own Sector Employment (localisation effects)		
		negative	***non significant***	***positive***
Other-Sector Employment (urbanisation effects)	***positive***	2 S55 (**, *) S70 (*** , **)	2 M23 (, **) M27 (, *)	3 M17 (** , **) S62 (* , ***) M30 (*** , **)
	non significant	2 S37 (*** ,) S63 (** ,)	25 M02 (,) S41 (,) M05 (,) S50 (,) M15 (,) S51 (,) M20 (,) S60 (,) M21 (,) S64 (,) M24 (,) S71 (,) M28 (,) S72 (,) M29 (,) S73 (,) M31 (,) S75 (,) M33 (,) S85 (,) S90 (,) S91 (,) S92 (,) S93 (,) S95 (,)	9 M01 (*** ,) S61 (*** ,) M14 (** ,) S66 (** ,) M19 (** ,) S80 (* ,) M22 (** ,) M25 (** ,) M34 (*** ,)
	negative	0	2 M36 (, **) S65 (, **)	11 M10 (*** , **) S40 (*** , ***) M11 (*** , ***) S45 (*** , **) M18 (*** , *) S52 (* , *) M26 (*** , ***) S67 (* , *) M32 (*** , ***) S74 (* , **) M35 (*** , *)

Note: *** shows significance at the 1 % level, ** shows significance at the 5 % level and * shows significance at the 10 % level. *M* represent sectors related to manufacturing (secondary sectors) as well as primary sectors (agriculture, mining, etc.), *S* identify tertiary sectors, mainly services.

TABLE 4

Summary of Own Sector (Localization) and Diversity (Urbanization) Effects

by Two-Digit Industrial Sector

		Coefficient on Own Sector Employment (localisation effects)		
		negative	***non significant***	***positive***
Coefficient on Employment Herfindahl (urbanisation effects)	***positive***	0	6 M02 (, ***) S60 (, ***) M23 (, **) S73 (, ***) M29 (, **) S91 (, *)	3 M17 (** , **) M26 (*** , ***) M34 (*** , *)
	non significant	3 S37 (*** ,) S55 (** ,) S70 (*** ,)	18 M05 (,) S41 (,) M15 (,) S50 (,) M20 (,) S64 (,) M21 (,) S65 (,) M24 (,) S72 (,) M27 (,) S75 (,) M28 (,) S90 (,) M31 (,) S95 (,) M33 (,) M36 (,)	14 M10 (*** ,) S40 (*** ,) M11 (*** ,) S45 (*** ,) M14 (** ,) S52 (* ,) M18 (*** ,) S66 (** ,) M25 (** ,) S67 (* ,) M30 (*** ,) S80 (* ,) M32 (*** ,) M35 (*** ,)
	negative	1 S63 (** , ***)	5 S51 (, *) S71 (, *) S85 (, **) S92 (, *) S93 (, *)	6 M01 (*** , ***) S61 (*** , *) M19 (** , *) S62 (* , **) M22 (** , ***) S74 (* , **)

Note: *** shows significance at the 1 % level, ** shows significance at the 5 % level and * shows significance at the 10 % level. *M* represent sectors related to manufacturing (secondary sectors) as well as primary sectors (agriculture, mining, etc.), *S* identify tertiary sectors, mainly services.

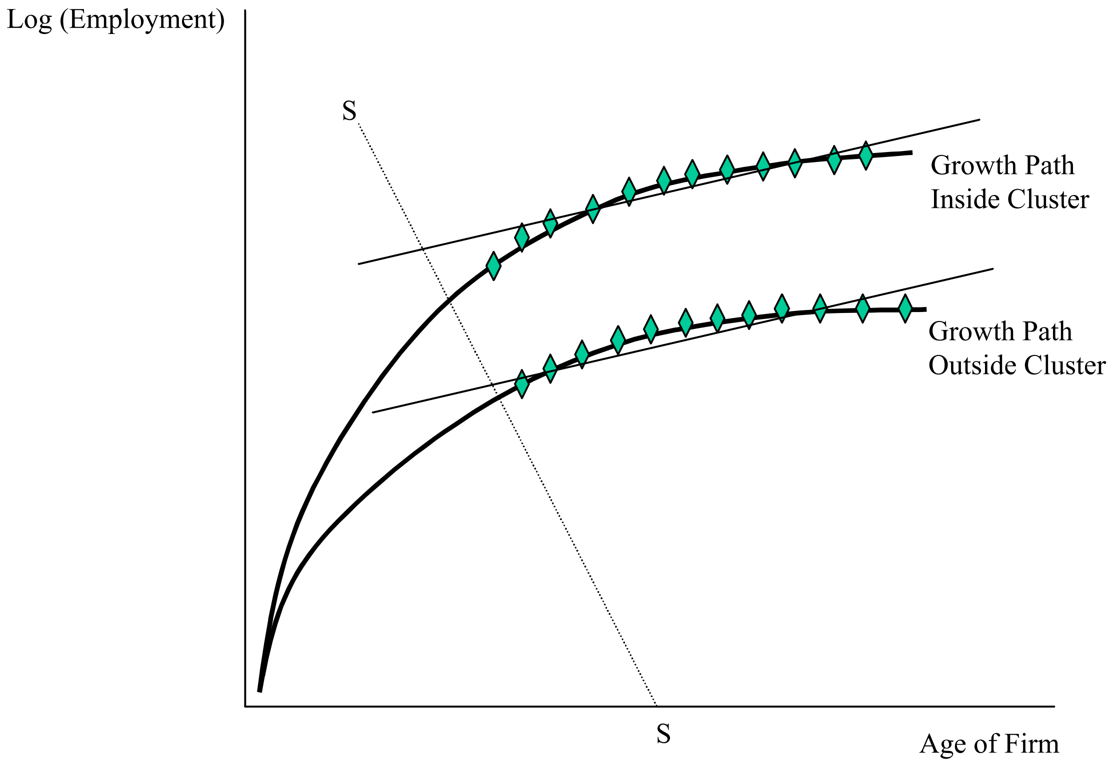

FIGURE 1 Patterns of Employment Growth, for firms inside and outside clusters

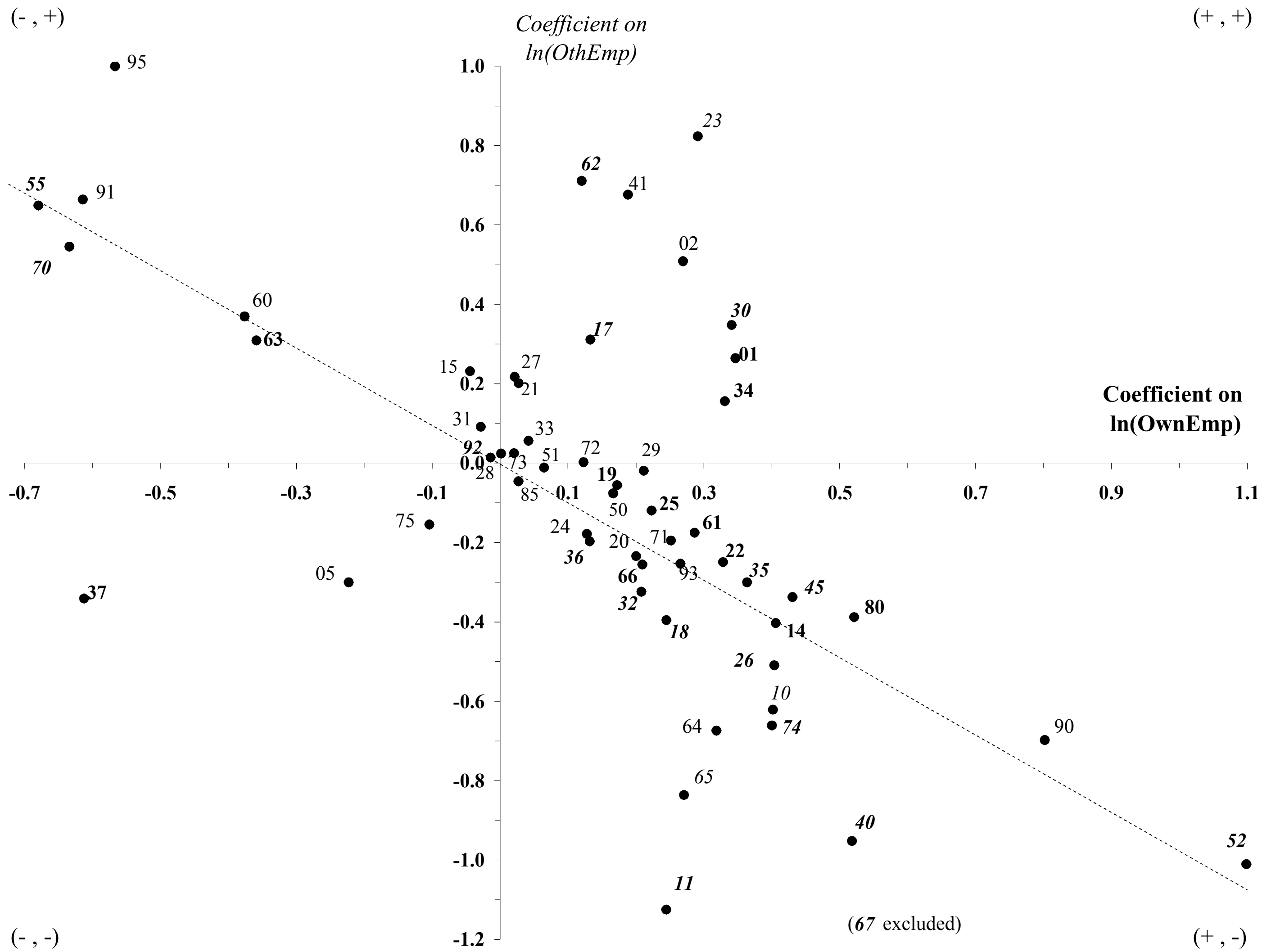

FIGURE 2 Coefficients on Own-Sector and Other-Sector Employment, Two-digit Industry Regressions

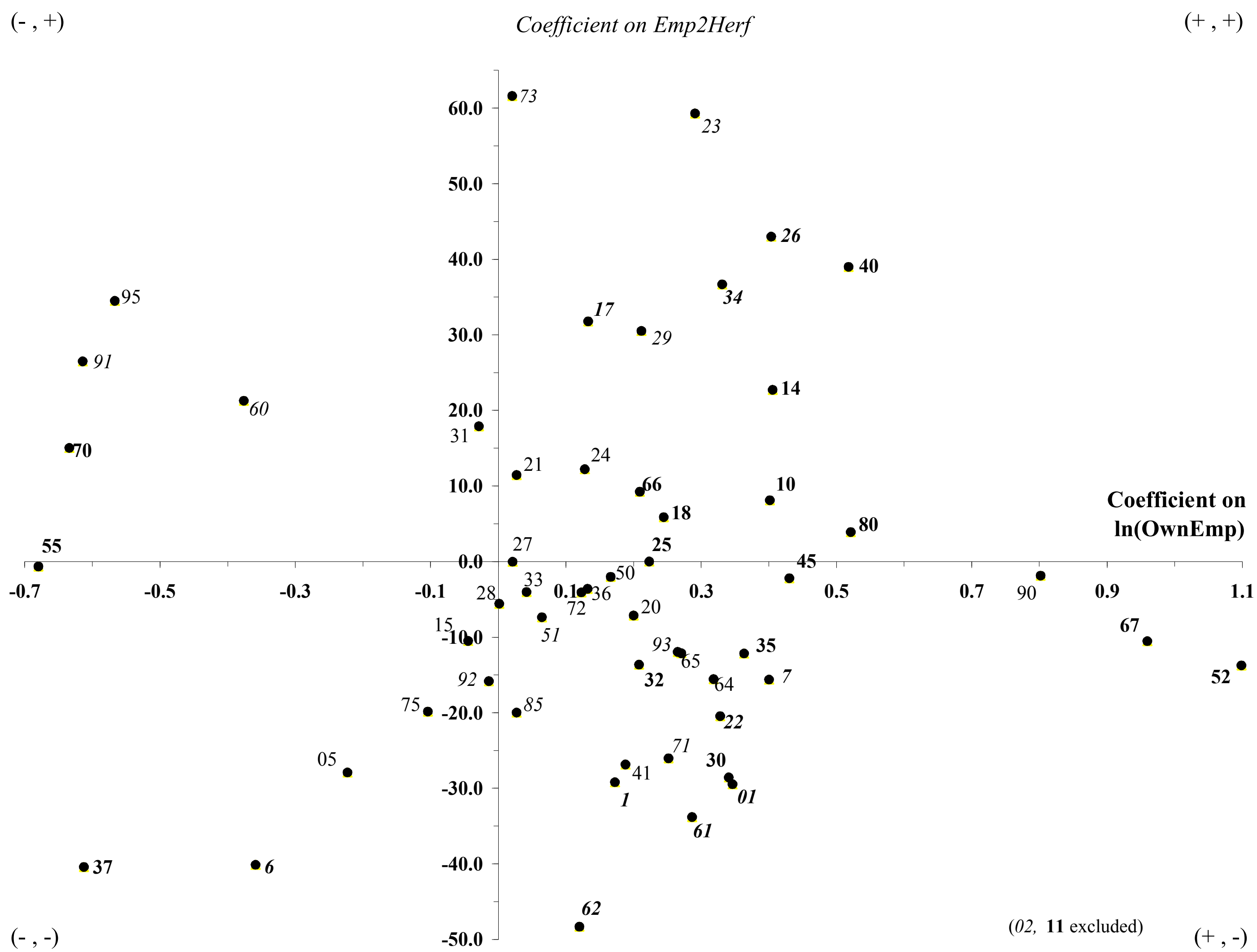

FIGURE 3 Coefficients on Own-Sector Employment and Employment Herfindahl, Two-digit Industry Regressions

APPENDIX A

Two-Digit UK Standard Industrial Classification (rev. 1992)

	Description		
01	Agriculture, hunted and related service activities	40	Electricity, gas, steam and hot water supply
02	Forestry, logging and related service activities	41	Collection, purification and distribution of water
05	Fishing, operation of fish hatcheries and fish farms; Service activities incidental to fishing	45	Construction
10	Mining of coal and lignite; Extraction of peat	50	Sale, maintenance and repair of motor vehicles and motorcycles; Retail sale of automotive fuel
11	Extraction of crude petroleum and natural gas; Service activities incident to oil and gas extraction excluding surveying	51	Wholesale trade and commission trade, except motor vehicles and motorcycles
14	Other mines and quarrying	52	Retail trade, except motor vehicles & motorcycles; Repair of personal goods
15	Manufacture of food products and beverages	55	Hotels and restaurants
17	Manufacture of textiles	60	Land transport; transport via pipelines
18	Manufacture of wearing apparel; Dressing and dying of fur	61	Water transport
19	Tanning and dressing of leather; Manufacture of luggage, handbags, etc.	62	Air transport
20	Manufacture of wood and of products of wood and cork, except furniture; Manufacture of articles of straw and plaiting materials	63	Supporting and auxiliary transport activities; Activities of transport agencies
21	Manufacture of pulp, paper and paper products	64	Post and telecommunications
22	Publishing, printing and reproduction of recorded media	65	Financial intermediation, except insurance and pension funding
23	Manufacture of coke, refined petroleum products and nuclear fuel	66	Insurance and pension funding, except compulsory social security
24	Manufacture of chemicals and chemical products	67	Activities auxiliary to financial intermediation
25	Manufacture of rubber and plastic products	70	Real estate activities
26	Manufacture of other non-metallic mineral products	71	Renting of machinery, equipment without operator, personal & household goods
27	Manufacture of basic metals	72	Computer and related activities
28	Manufacture of fabricated metal products, except machinery & equipment	73	Research and development
29	Manufacture of machinery and equipment nec	74	Other business activities
30	Manufacture of office machinery and computers	75	Public administration and defence; Compulsory social security
31	Manufacture of electrical machinery and apparatus nec	80	Education
32	Manufacture of radio, television and communication equipment	85	Health and social work
33	Manufacture of medical, precision & optical instruments, watches & clocks	90	Sewage and refuse disposal, sanitation and similar activities
34	Manufacture of motor vehicles, trailers and semi-trailers	91	Activities of membership organisations nec
35	Manufacture of other transport equipment	92	Recreational, cultural and sporting activities
36	Manufacture of furniture, manufacturing nec	93	Other services activities
37	Recycling	95	Private households with employed persons

NOTES

---

[1] This research was funded by Stanford University as part of the SIEPR project on *Silicon Valley and its Imitators*. We are grateful to the Office of National Statistics for providing data. We acknowledge helpful comments from the associate editor, Simon Parker, and two anonymous referees as well as from our partners in this research programme, notably Ashish Arora, Alfonso Gambardella, Paul Romer, Annalee Saxenian, and also discussions with Gary Cook and Naresh Pandit, and the earlier work of Rui Baptista developing econometric models of entry, growth and innovation. None of these, however, are responsible for any remaining errors.

[2] More complex measures of diversity exist in the literature (see Henderson *et al*., 1995; Glaeser *et al.*, 1992; or Combes, 2000). In our regressions, we tried these different measures, but the simplest Herfindahl worked best and will be used in this paper.

[3] There are two ways to treat the consolidated account problem. First, holding companies can be treated separately from their subsidiaries. A problem then arises when we encounter subsidiaries of subsidiaries. Second, dummy variables for holding companies and consolidated accounts can be introduced. This second method is preferred because it does not reduce the sample size more than is necessary. Another issue is how we should treat entry by and growth of multi-sector firms. In this study, such companies were only counted in the principal sector where the firm was active in the sample year. That may be an unsatisfactory assumption when the firm has diversified from one original sector into others, but in the absence of detailed information on such diversification in each firm, this seems a reasonable working assumption.

[4] We have checked for the robustness of the regressions with respect to outliers that may arise from firms filing consolidated accounts, and found no such problem: coefficients and standard errors showed only slight variations when such observations were removed.

[5] NUTS: Nomenclature des Unités Territoriales Spatiales. In the UK, NUTS level 1 corresponds to 11 regions, while NUTS level 3 represents the 65 counties.

[6] The use of NUTS level 3 data is an improvement on some previous studies (which used NUTS levels 1 and 2), but because of this regional disaggregation, we had to compromise on the industrial disaggregation. In an ideal world, we would have NUTS level 3 data using 3- or

4-digit industrial codes. However, for the present study the 56 two-digit industrial codes (1992 rev.) give a sufficient disaggregation for our immediate interest.

[7] A description of these codes is provided in appendix A.

[8] This is somewhat unsatisfactory, since the cluster is generally smaller than a NUTS 2 area - really more in line with a NUTS 3 area. However, studies such as Glaeser *et al.* (1992) and Jaffe *et al.* (1993) have shown that external effects of the kind that are explored here seem to grow stronger as the regional unit becomes smaller. Any bias introduced here should be to underestimate the strength of clustering effects. Nevertheless, the use of regions as a spatial unit has some administrative sense. In some countries, for example, government policies and incentives towards new industries are to some extent defined at a regional level.

[9] For England and Wales, this data was obtained from the British office, while for Scotland, population estimates were provided by the Scottish office.

[10] One of the problems encountered with data published by the Office of National Statistics, especially within Regional Trends, is the change in the classification of the regions. For a number of years now, and probably since the introduction of the Scottish Parliament and the Welsh Assembly, the boundaries of various 'official' regions in Scotland and Wales have changed and no longer correspond strictly to the NUTS level 3 regions. It was however possible to obtain proxy measures for the NUTS level 3 regions from the various government offices responsible to provide this information to Eurostat, the European office of statistics.