Università degli Studi del Molise
Facoltà di Economia
Dipartimento di Scienze Economiche, Gestionali e Sociali
Via De Sanctis, I-86100 Campobasso (Italy)

ECONOMICS & STATISTICS DISCUSSION PAPER
No. 42/08

Does Spatial Proximity Matter?
Micro-evidence from Italy

by

Giulio Cainelli
University of Bari and CERIS-CNR, Milan

and

Claudio Lupi
University of Molise, Dept. SEGeS, Campobasso

Does Spatial Proximity Matter? Micro-evidence from Italy *

Giulio Cainelli
University of Bari and CERIS-CNR, *Milan*

Claudio Lupi
University of Molise, Dept. SEGeS, Campobasso

Abstract

The effect of spatial agglomeration (localisation and urbanisation economies) on employment growth is explored over a balanced panel of 23,374 Italian firms, using a flexible Bayesian model. Contrary to previous research, the agglomeration economies measures are calculated using direct measures of physical distances between pairs of firms, rather than with respect to pre-specified geographical units. We find that localisation effects are positive but decreasing with distance, while the variety effects are negative for distances within 10 kilometers and become positive for distances in a range of 10–30 kilometers. Our results suggest that the use of geographic units such as standard metropolitan units, LLS, administrative regions or provinces can be misleading.

JEL classification: R11, O47
Key words: Proximity, agglomeration, knowledge spillover, employment.

1 Introduction

Endogenous growth theories (Romer, 1986, 1990; Lucas, 1988; Grossman and Helpman, 1991) emphasise the role of knowledge spillovers for enhancing technological change and long-term economic growth. A more recent strand of literature suggests that technological spillovers not only generate externalities, thereby fostering economic growth, but also tend to be spatially bounded (Jaffe et al., 1993; Audretsch and Feldman, 1996). Indeed, spatial proximity, stimulating face-to-face interactions between economic agents (firms and individuals), can facilitate the speed of spread

*This first version of this paper was prepared for inclusion in the book *Internationalization, Technological Change and the Theory of the Firm*, to be edited by Nicola De Liso and Riccardo Leoncini, and to be published by Routledge. We would like to thank, without implicating, Agostino Nobile for insightful comments.

of ideas, information (on new products, production processes, and markets), and different types of knowledge (codified and tacit, public and private) within a locality (Storper and Venables, 2004), thus causing differences not only in productivity at firm level, but also in growth rates at the local level. As Glaeser et al. (1992, p. 1127) correctly point out:

> the cramming of individuals, occupations, and industries into close quarters provide an environment in which ideas flow quickly from person to person.

In the early 1990s, these insights find empirical foundation in some seminal regional economics contributions, which extensively investigate the relationships between spatial agglomeration, knowledge spillovers, and economic growth at the urban level (Glaeser et al., 1992; Henderson et al., 1995). More specifically, Glaeser et al. (1992), using a cross-section of US cities, analyse the impact of three different forms of local knowledge spillovers — Marshall-Arrow-Romer (MAR), Porter and Jacobs externalities — on subsequent urban employment growth. In their paper, they show that localisation economies (also called MAR economies), arising from spatial concentration of firms belonging to the same industry, and captured by specialisation indicators, have a negative impact on urban economic growth, while urbanisation (or Jacobs) economies, spurred by the variety and diversity of geographically proximate industries, positively affect the subsequent growth of a metropolitan area. Using a similar empirical framework, Henderson et al. (1995) find that localisation has a positive role in mature capital-goods sectors, while differentiation of the productive structure (variety), which should generate cross-fertilisation of ideas between different industries, has a positive impact only in the case of high tech sectors. Finally, Forni and Paba (2002), using information on a cross section of 995 Italian Local Labour Systems (LLSs) for the period 1971-1991 find that in most cases specialisation and variety positively affect growth, but the variety is different for each industry. Moreover, they note that, consistent with Marshall (1920), in order to capture the spillover generating process a size effect needs to be added to the specialisation effect. The Glaeser et al. (1992) model has been replicated in the context of different countries in order to provide further evidence on these issues. Nonetheless, the various results obtained by the empirical research in this field are quite controversial such that currently there is no unique model explaining the link between labour growth and the structure of the local economy. In particular, some studies referring to the Italian case find that specialisation has a negative impact on local growth, while diversity plays a positive role (see, among others, Cainelli and Leoncini, 1999; Cainelli et al., 2001; Cunat and Peri, 2001; Usai and Paci, 2003; Paci and Usai, 2006; Mameli et al., 2007).

This empirical literature has been extended by some more recent studies (de Lucio et al., 2002; Henderson, 2003; Cingano and Schivardi, 2004). These papers analyse the impact of measures of agglomeration economies not only on employment growth (as in the original body of literature referred to above), but also on productivity growth or firms' total factor productivity growth (TFP). The findings within this new strand of empirical research are also rather puzzling. For example, de Lucio et al. (2002) investigating the relationship between labour productivity and spatial agglomeration at the level of the 50 Spanish provinces for the period 1978-1992, find that variety played a role in labour productivity growth, and find a U-shaped effect for specialisation. According to their results, low levels of specialisation reduce productivity growth and high levels foster it. In contrast, Cingano and Schivardi (2004), using firm-level based TFP indicators, show that specialisation, calculated at the level of the 784 Italian LLSs, has a positive impact on firm productivity growth, but that variety has no significant effect. Taking local employment growth as the dependent variable, Cingano and Schivardi (2004) show that the specialisation effect is reversed and becomes negative, while variety has a significant and positive impact on employment growth, thus confirming Glaeser et al.'s results. Finally, Henderson (2003), using the Longitudinal Research Database (LRD) of the US Census Bureau, find that localisation economies have strong positive effects on productivity at plant level in high tech industries, but not in machinery industries and found little evidence of urbanisation economies. The use of TFP measures is an obvious and notable improvement of these studies, which however must accept some of the drawbacks related to other measurement and empirical issues: for example, the use of sample of plant data (Henderson, 2003) or the problems of sample selection in the case of Cingano and Schivardi's paper. However, in our opinion, the most important shortcoming of all these studies is that they refer to exogenously defined geographic units such as standard metropolitan units, LLS, administrative regions or provinces.

A further consequence of this choice is that these contributions have difficulties in dealing with a rather relevant aspect of these phenomena: namely, the attenuation of agglomeration economies over space (Rosenthal and Strange, 2006). Recently, Desmet and Fafchamps (2005), using US county data for 1972 and 2000, try to overcome this problem by assuming that a county's employment growth is not only affected by the county under consideration, but also by all "near" counties. van Oort (2007) tries to tackle the same issue by considering spatial dependence, though he encounters some difficulties related to the robustness of his results. Our paper takes a similar, but alternative, view. Using a panel data set of 23,374 Italian manufacturing firms for the period 1998-2001 and estimating a flexible Bayesian model, the effect of spatial agglomeration economies — that is, localisation and urbanisation economies — on employment growth over discrete distances (not, as the previous literature, within pre-defined geographic units) is explored.

This paper thus makes two contributions to the empirical literature. First, by calculating precise actual distances between firms for each company in our large data set, we can compute agglomeration economies at different distances. This is accomplished by using available GIS location coordinates for each firm in our sample. Second, using these agglomeration variables we can study the impact of these measures on firms' employment growth. Thus, using measures of agglomeration economies over actual distances between firms, we can empirically identify the rate at which knowledge spillovers attenuate over space.

The remainder of the paper is organised as follows. In the next section we describe the data set and the empirical framework used in our analysis. In Section 3 the statistical model and the empirical results are presented and discussed. Section 4 concludes.

2 Data

The data source used in this paper is AIDA, a commercial database collected by Bureau Van Dijck. This large data set of Italian joint stock companies reports balance sheet data such as sales, number of employees, labour cost, etc., as well as the specific sector of activity.[1] In addition, it reports firms' street addresses, information that is very useful for this study. Using these data, we built a balanced panel data set composed of 24,089 Italian manufacturing firms for the period 1998-2001. The firms are located across 18 Italian regions. Firms in Sicily and Sardinia were excluded from our investigation because of the insular nature of these regions.

One novelty of this paper is that we are able to measure agglomeration economies over space, exploiting information on the actual distance between each pair of firms in the sample. The data on street addresses allow us to recover each firm's exact longitude and latitude coordinates. Then, using these coordinates — available for all the firms in our sample — we calculated, by means of a GIS programme, the actual distance (in meters) between each pair of firms (Wallsten, 2001). Finally, for each company, we computed the number of other firms located within different actual distances: i.e., from 0 to 2 kilometers, from 2 to 10 kilometers, from 10 to 30 kilometers. This allows us to perform a first qualitative analysis, whose results are listed in Table 1.

[Insert Table 1 about here]

Table 1 shows the distribution of sample firms by distance and industry. From this table it emerges the tendency for Italian manufacturing firms to locate close to one another. As a matter of fact, about 16% of total firms belonging to the same industry are located within 10 kilometers, and about 30% are located over a distance of 30 kilometers or less. It is also interesting that this tendency towards spatial agglomeration seems to vary according to the industry.

We use these data to calculate two different types of measures of agglomeration economies: (i) an indicator of localisation economies at different distances, and (ii) an indicator of urbanisation (or variety) at the same distances. The variable used to measure localisation (or MAR) economies is calculated as

$$L_{i,s}^{(d)} = n_s^{(d)}$$

where $n_s^{(d)}$ is the number of other firms belonging to the same industry s, located within distance d. As already noted, this indicator is calculated over different actual distances d: i.e., from 0 to 2 kilometers, from 2 to 10, and so on. The variable $V_{i,s}^{(d)}$ used to measure urbanisation (or Jacobs) economies — i.e., variety of the local productive structure — is calculated for a generic firm i belonging to industry s using Shannon's entropy index (Shannon, 1948) excluding sector s. Shannon's entropy index, also known as the Shannon-Wiener index has been widely used to measure biodiversity taking account of both the number and the evenness of species (see e.g.

[1] Each firm in the database is assigned to a sector according to the *Statistical Classification of Economic Activities in the European Community*, NACE Rev. 1.1 (2002). The correspondence between industry codes and their descriptions is provided in Table 7 in the Appendix.

Ricotta and Szeidl, 2006). We think that this can be very useful also in our context, where we have to measure the variety of "species" of firms in selected areas. Let S denote the number of sectors and N the total number of firms in the area. n_s is the number of firms in each sector $s \in [1, \ldots, S]$. Of course in each area the number of firms, N, is $N = \sum_{s=1}^{S} n_s$. Define also p_s as the proportion of the firms in sector s to the total number of firms in the area, $p_s = n_s/N$. Then the Shannon-Wiener index is defined as

$$H' = -\sum_{s=1}^{S} p_s \ln(p_s) .$$

It can be proved that H' is maximized when each sector is represented by an equal number of firms. It can also be shown that in this case

$$H_{\max} = \ln(S) .$$

Therefore it is also possible to compute a *relative* index by considering $H'/H_{\max}$.

In order to avoid some evident outlying observations, a very mild trimming (0.5% on both tails of the distributions) has been performed sequentially on employment growth (our dependent variable) and on production growth and labour cost per employee growth, respectively. This left us with 23,374 valid observations. In order to offer a visual summary of the main variables involved in the analysis, their estimated densities are plotted in Figure 1.

[Insert Figure 1 about here]

In this paper we want explicitly to study the properties of the sample at hand. We do not intend to draw inferences that are valid through the whole universe of Italian firms. Indeed, a potential problem with these kinds of samples is that firms are not randomly chosen (Cingano and Schivardi, 2004). However, comparisons with the whole population in terms of frequency distribution both by sector (Table 2), and by geographical areas (Table 3) show that the structure of our sample is generally well in accordance with that of Census data. The only (potential) selection problems are that average firm size in our sample is generally bigger than in the reference population, and that southern firms tend to be slightly under-represented.

[Insert Table 2 about here]

[Insert Table 3 about here]

The original 14 2-digit sectors have also been reduced to 11 by aggregating together sectors DB and DC (Manufacture of textiles and textile products and Manufacture of leather and leather products), DD and DN (Manufacture of wood and wood products and Manufacturing not elsewhere classified, including furnishings), and DF and DG (Manufacture of coke, refined petroleum products and nuclear fuel and Manufacture of chemicals, chemical products and man-made fibres).

3 The statistical model and the empirical results

We test the effects of agglomeration economies on firms' employment growth. More specifically, we adopt the following standard specification in long differences:[2]

$$\Delta_3 \log(e_i) = \beta_1 \Delta_3 \log(y_i) + \beta_2 \Delta_3 \log(w_i) + \sum_{d=1}^{D} \delta_{1,d} L_i^{(d)} + \sum_{d=1}^{D} \delta_{2,d} V_i^{(d)} + \xi_i$$

where Δ_3 is such that $\Delta_3 z_t := z_t - z_{t-3}$, e_i is employment, y_i is real output and w_i denotes real wage per employee. In addition, β_1 and β_2 are the elasticities with respect to output and wages, while $L_i^{(d)}$ and $V_i^{(d)}$ denote the localisation and urbanisation variables, respectively.

In a preliminary stage of our investigation we used a simple linear model relating employment growth to changes in wages and production, agglomeration variables, as well as size, geographical, and sector dummies. The model gave interesting results but it was open to three major objections. First, it is reasonable to think that sectoral effects cannot be fully controlled using intercept dummies. Second, residuals were approximately distributed according to a Student t random variable with about 3 degrees of freedom. Finally, we also observed that residuals variance was not constant across size classes. Therefore we decided to build a more sophisticated model that could explicitly address these points.

We finally decided to build a Bayesian model. The reason for this choice is threefold. First, it allows us to build an extremely flexible random coefficients model; second, it allows us to derive estimates of the full (posterior) distributions of the parameters, instead of simple point estimates; third, it allows us to explicitly address non-normality. Of course, this extra flexibility comes at the cost of a significant increase in the computational burden.

In many empirical applications observations are not (conditionally) normally distributed. Indeed, our preliminary results seemed to suggest that employment growth could be conditionally t-distributed. Modelling by using a Student-t distribution has also two practical advantages. In fact, using a t distribution allows us to obtain a model that is more *robust* to outlying observations and helps in coping with heteroscedasticity (see e.g. Gelman et al., 2004; Geweke, 1993). It should be emphasized that the degrees of freedom parameter of the t distribution is estimated within the model, so that our model encompasses a Gaussian one. In fact, the model is flexible enough to accommodate a Gaussian distribution, should this be the "true" conditional distribution of the data.

Let the $(n \times k)$ matrix $\boldsymbol{X}$ denote the explanatory variables used in the model. The i-th row of $\boldsymbol{X}$, relative to the i-th firm, is

$$\boldsymbol{x}'_{i\cdot} := \left(\Delta_3 \log(w_i), \Delta_3 \log(y_i), \boldsymbol{d}'_i\right) ,$$

where $\boldsymbol{d}'_i$ is the row-vector containing the firm-specific agglomeration variables and geographical dummies.

[2]Long differences have been commonly used to eliminate region-specific effects and to capture medium- to long-run relationships between the variables of interest (see Holtz-Eakin and Schwartz, 1995; Boarnet, 1998; Picci, 1999; Brynjolfsson and Hitt, 2003, among others).

Specifically, for the i-th firm belonging to sector s and size class j we assume:

$$\begin{aligned}
\Delta_3 \log(e_i)|\boldsymbol{X} &\sim t\left(\mu_i, \tau_j, \nu_j\right) \\
\mu_i &= b_{1,s,j} + b_{2,s,j}\,\Delta_3 \log(w_i) + b_{3,s,j}\,\Delta_3 \log(y_i) + \\
&\quad + c_1 L1_i + c_2 L2_i + c_3 L3_i + d_1 V1_i + d_2 V2_i + d_3 V3_i + \\
&\quad + e_1 G1_i + e_2 G2_i + e_3 G3_i\,.
\end{aligned}$$

Here, $\Delta_3 \log(e_i)$, $\Delta_3 \log(w_i)$, and $\Delta_3 \log(y_i)$ indicate the 1998-2001 growth of employment, labour cost per employee, and production, respectively. $L1_i, \ldots, L3_i$ and $V1_i, \ldots, V3_i$ are the localisation and variety variables for various distances.[3] $G1_i, \ldots, G3_i$, indicate geographical dummies for the north-western, north-eastern, and southern regions, respectively.[4]

[Insert Table 4 about here]

The precision parameter,[5] τ_j, and the degrees of freedom parameter, ν_j, are assumed to vary with the size of the firm. Indeed, we found evidence in this sense in our initial linear model. Note that the degrees of freedom parameter is left unspecified and is estimated from data. The b parameters, that are related to technology and institutional factors, are instead assumed to vary with both sector and firm size.

Given our previous experience, we expect a low value for ν_j, so that we assign to it a prior such that $1/\nu_j \sim \mathrm{U}(0, 0.5)$. Of course this is a mildly informative prior, given that it assigns a value of ν_j between 2 and 4 with a probability $1/2$.

The other priors are fairly standard:

$$\begin{aligned}
\tau_j &\sim \text{Gamma}(0.001, 0.001) \\
b_{k,s,j} &\sim \mathrm{N}(0, 0.0001) \;\; k = 1, \ldots, 3;\; s = 1, \ldots, 11;\; j = 1, \ldots, 4 \\
c_k &\sim \mathrm{N}(0, 0.0001) \;\; k = 1, \ldots, 3 \\
d_k &\sim \mathrm{N}(0, 0.0001) \;\; k = 1, \ldots, 3 \\
e_k &\sim \mathrm{N}(0, 0.0001) \;\; k = 1, \ldots, 3
\end{aligned}$$

where $\mathrm{N}(\mu, \tau)$ denotes a normal distribution with mean μ and precision (the inverse of the variance) τ.

The model has been estimated by Monte Carlo Markov Chains (MCMC) using `R` and `WinBUGS` (R Development Core Team, 2006; Spiegelhalter et al., 2004). Three independent chains of length 1000 (excluding the burn-in replications) have been used to derive the posterior distributions of the parameters. The starting values of the chains were randomly selected from uniform distributions. Convergence was reached for all the parameters of the model.

Table 5 lists the descriptive statistics of the parameters of main interest in the model, together with the potential scale reduction factor $\hat{R}$ (Gelman and Rubin, 1992). If convergence is reached, $\hat{R}$ should be close to unity.

[3] The precise definition of the variables is offered in Table 4.

[4] With the exception of Sicily and Sardinia that have been excluded from the analysis, macro-areas are defined according to the official classification used by the National Institute of Statistics (Istat).

[5] In Bayesian analysis it is customary to use precision, the inverse of variance, to summarize distribution dispersion.

[Insert Table 5 about here]

Prior to commenting in detail the economic significance of our results, we want to check the plausibility of our model using posterior predictive analysis, along the lines exemplified, e.g., in Geweke and McCausland (2001).

Specifically, we compare the ability of our model to replicate some interesting features of the data, as opposed to a simple non-hierarchical Gaussian model. To derive the implications of the Gaussian model, we simply simulate $\boldsymbol{y}^{\dagger} \sim \mathrm{N}(\boldsymbol{X}\hat{\boldsymbol{\beta}}, \hat{\sigma}^2 \boldsymbol{I})$, with $\hat{\boldsymbol{\beta}}$ being the OLS estimates of the parameters. At each simulation step we compute the statistics of interest on $\boldsymbol{y}^{\dagger}$ and compare them with those computed on actual data. By repeating this procedure several times, it becomes possible to derive the quantiles of the distribution of the statistics of interest of the simulated data, as well as the p-values. Then, we use posterior predictive simulation to compare the implications of our model with respect to the same observed characteristics of the data. In the present study we focus on excess kurtosis, the skewness coefficient, and the quantile ratio defined as $(\max(\boldsymbol{y}) - \min(\boldsymbol{y}))/(y_{0.75} - y_{0.25})$ with $y_{0.75}$ and $y_{0.25}$ the third and first quartile of $\boldsymbol{y}$, respectively.

The results, reported in detail in Table 6, show that the Gaussian model is not able to reproduce the observed characteristics of the data. In no instance the observed values are included within the 5%–95% quantile interval. The p-values are always 0, up to the third decimal. On the contrary, our hierarchical model "fits" the data much better. The 5%–95% quantile intervals always cover the observed values and the p-values are always well above any conventional significance level. Of course, there might still be margins of improvement, but the results are strongly suggestive of a marked superiority of our model over the standard Gaussian alternative.

[Insert Table 6 about here]

The results confirm our intuition, based on preliminary results, that the data seem to be conditionally t-distributed with about 3 degrees of freedom, rather than being normally distributed (see the values of ν in Table 5). Indeed, given the importance of this parameter, it can be instructive to plot its posterior density. Figure 2 shows that the posterior median of ν_j is very close to 3 for any size class. Furthermore, all the upper bounds of the 90% highest posterior density (HPD) intervals are well below 4. We consider this result as a clear confirmation of non-normality.

[Insert Figure 2 about here]

Also, it is confirmed that there is substantial heteroscedasticity across size classes, as previously suggested (see the values of τ in Table 5).

An interesting side-result is that our estimates show that, controlling for the other factors, in the period 1998-2001 employment growth was higher in the southern and in the north-eastern regions (parameters e_1 and e_2 in Table 5) as compared to the remaining areas of the country. Employment growth in the central and north-western regions was instead of comparable size. Of course, the possibility that these growth differentials are extended over long horizons so to reduce the existing unemployment differentials is a different matter beyond the scope of the present work.[6]

A summary of the estimated intercepts and of the coefficients b_2 and b_3 is reported graphically in Figures 3–5. Estimates show that the parameters have the expected signs and that there is substantial variation across sectors and firms' size.

[Insert Figure 3 about here]

[Insert Figure 4 about here]

[Insert Figure 5 about here]

In terms of the main goal of the analysis, according to our estimates, localisation effects are positive, but decreasing with distance (see Figure 6). More precisely, the posterior distributions show that the model predicts that there is about a 99% probability that the localisation effect is positive within 2 kilometres. The probability decreases to about 89% and 82% for distances between 2–10 kilometres and 10–30 kilometres, respectively. The median effect decreases similarly. This evidence confirms the importance of taking into account attenuation phenomena when these type of agglomeration forces are considered. In addition, this evidence suggests that the use of geographic units such as standard metropolitan units, LLS, administrative regions or provinces (exogenously defined) can be misleading, since the impact of localisation economies on employment growth tends to change with distance.

[Insert Figure 6 about here]

It should be noted that this result, even if it is not counter-intuitive respect to the Italian experience (see, for example, the literature on Italian industrial districts (Signorini, 1994) which suggests a positive role for these kinds of agglomeration forces in explaining the success of these local production systems), it is not consistent with some previous contributions on the issue. In fact, these studies, using specialisation indicators, find a negative role for localisation economies (Glaeser et al., 1992; Henderson et al., 1995). At the same time, some studies (Forni and Paba, 2002) find that the impact of productive specialisation is positive when it is considered jointly with the size, in terms of employment, of the local industry under consideration. In other words, the main idea in these papers is that specialisation and size must be jointly considered in order to capture the knowledge spillover generating process. In our paper, however, localisation economies are not measured using specialisation indicators, as in previous studies (see, for example, Glaeser et al., 1992), but through direct measures of spatial agglomeration: that is, the number of firms actually located at different physical distances. In our view, these measures should partially capture the size, in this case in terms of number of firms, of the local industry under consideration. This may explain the finding that localisation economies matter for employment growth. A second explanation is that, as already noted, we do not use pre-defined geographic units such as LLSs. It is well known, that the size of these geographic units tends to

[6] For an assessment of unemployment geographical differences in Italy see Brunello et al. (2001), among others.

change considerably in terms of areas, population, and so on, thus imposing a predefined, non-empirically tested spatial boundary within which agglomeration forces act. In addition, this non-empirically tested hypothesis changes according to the LLS being considered.

At the same time, the variety effects are negative for distances within 2 kilometres and between 2–10 kilometres, while this form of agglomeration forces become positive for distances between 10–30 kilometres. The estimated probabilities of the negative effects are 100% and 97%, while the estimated probabilities of the positive effect is about 70%. In this case the importance of taking into account attenuation phenomena is also confirmed, when spatial agglomeration is being analysed. Moreover, our evidence suggests that variety of the production structure has a positive impact on firms' employment growth only for distances between 10–30 kilometres. In this sense, this latter finding supports those previous studies that identified a positive role of variety on employment growth (Glaeser et al., 1992), but also suggests that this positive role of variety needs space to become effective.

4 Conclusions

In this study we used a panel data set of 23,374 Italian manufacturing firms for the period 1998-2001 to estimate a flexible Bayesian model, to examine the impact of spatial agglomeration economies — that is, localisation and urbanisation economies — on employment growth over discrete distances.

Our main results can be summarised as follows. We find that localisation effects are positive, but decreasing with distance. More precisely, the posterior distributions show that the model predicts that there is approximately a 99% probability that the localisation effect is positive within 2 kilometres. This probability decreases to about 89% and 82% for distances between 2–10 kilometres and 10–30 kilometres, respectively. On the contrary, the variety effects are negative for distances within 2 kilometres and between 2–10 kilometres, while this form of agglomeration force becomes positive for distances between 10–30 kilometres.

To sum up, the main finding of this paper is that the use of geographic units such as standard metropolitan areas, LLSs, administrative regions, or provinces (exogenously defined) can be misleading, since the impact of spatial agglomeration economies on employment growth tends to change with distance, and local knowledge spillovers seem to attenuate over space.

Our study could be extended in a number of interesting ways. First, our dataset could be enlarged in order to include non-joint stock companies. This could be achieved by using, for example, microeconomic information drawn from Census data. Finally, in the presence of reliable data on firms' capital stock we could also investigate the impact of spatial agglomeration economies on firms' productivity growth or TFP.

Sectors	$\leq$ 2km	$\leq$ 10km	$\leq$ 30km
DA	3.92	9.23	20.10
DB	7.87	17.58	30.40
DC	7.89	15.37	27.58
DD	5.80	12.48	26.71
DE	9.07	23.63	35.93
DF	13.00	18.00	28.00
DG	9.68	24.25	39.14
DH	5.36	14.80	32.25
DI	4.96	11.70	22.18
DJ	4.90	14.25	29.97
DK	5.53	15.63	31.17
DL	6.98	20.74	35.88
DM	6.63	14.83	26.70
DN	5.61	13.09	26.59
All sectors	6.32	16.15	30.16

Table 1: Distribution of firms by distance and industry. The correspondence between industry codes and their descriptions is provided in Table 7 in the Appendix.

	Sample		Census 2001	
Sectors	Firms%	Employees%	Firms%	Employees%
DA	7.67	6.00	6.72	6.59
DB	13.47	9.71	12.33	10.87
DC	3.95	2.14	4.34	3.53
DD	2.37	1.23	2.78	1.68
DE	6.83	7.04	8.56	5.47
DF	0.42	0.50	0.27	0.71
DG	5.05	12.43	3.26	6.08
DH	5.55	5.05	5.18	5.36
DI	5.17	5.08	5.17	5.21
DJ	18.06	16.05	17.80	15.32
DK	14.48	14.95	13.94	15.37
DL	9.50	11.58	9.78	10.61
DM	2.34	5.56	2.26	7.83
DN	5.13	2.68	7.61	5.36

Table 2: Distribution of firms and employees by sector. Data refer to joint stock companies. The correspondence between industry codes and their descriptions is provided in Table 7 in the Appendix.

	Sample		Census 2001	
Area	Firms%	Employees%	Firms%	Employees%
North-West	46.40	52.05	38.56	46.63
North-East	32.51	29.37	26.11	27.78
Centre	14.80	14.37	19.24	15.31
South	6.28	4.21	16.10	10.28

Table 3: Distribution of firms and employees by geographical area. (The data refer to joint stock companies.)

Variable	Definition
$\Delta_3 \log e_i$	1998-2001 growth of employment (dependent variable)
$\Delta_3 \log y_i$	1998-2001 growth of production
$\Delta_3 \log w_i$	1998-2001 growth of labour cost per employee
$L1_i$	Centered log of number of firms of the same sector located within 2 kilometers
$L2_i$	Centered log of number of firms of the same sector located between 2 and 10 kilometers
$L3_i$	Centered log of number of firms of the same sector located between 10 and 30 kilometers
$V1_i$	Centered Shannon-Wiener index computed between 0 and 2 kilometers
$V2_i$	Centered Shannon-Wiener index computed between 2 and 10 kilometers
$V3_i$	Centered Shannon-Wiener index computed between 10 and 30 kilometers
$G1_i$	The firm is located in the North-West
$G2_i$	The firm is located in the North-East
$G3_i$	The firm is located in the South

Table 4: Definitions of the variables used in the empirical analysis.

	Mean	SD	Naive SE	MC Error	Batch SE	Batch ACF	Q(0.05)	Q(0.50)	Q(0.95)	$\hat{R}$
c_1	0.0040	0.0018	0.0000	0.0000	0.0000	0.0870	0.0011	0.0040	0.0069	1.0011
c_2	0.0018	0.0015	0.0000	0.0000	0.0000	0.0151	-0.0007	0.0018	0.0043	1.0005
c_3	0.0013	0.0014	0.0000	0.0000	0.0000	0.1296	-0.0010	0.0013	0.0036	1.0005
d_1	-0.0078	0.0023	0.0000	0.0001	0.0001	0.2876	-0.0116	-0.0078	-0.0039	0.9994
d_2	-0.0079	0.0041	0.0001	0.0001	0.0001	-0.2852	-0.0147	-0.0079	-0.0013	1.0013
d_3	0.0031	0.0062	0.0001	0.0002	0.0002	0.0887	-0.0070	0.0030	0.0132	1.0007
e_1	-0.0031	0.0042	0.0001	0.0001	0.0001	0.1000	-0.0101	-0.0032	0.0037	0.9991
e_2	0.0178	0.0042	0.0001	0.0001	0.0001	0.1245	0.0109	0.0179	0.0247	0.9990
e_3	0.0389	0.0068	0.0001	0.0002	0.0002	0.3481	0.0273	0.0390	0.0498	1.0003
τ_1	27.8026	0.8110	0.0148	0.0472	0.0475	0.1479	26.5495	27.7800	29.1805	1.0282
τ_2	50.0031	1.4835	0.0271	0.0900	0.0920	-0.0100	47.5600	50.0000	52.4500	1.0064
τ_3	65.5372	2.4488	0.0447	0.1448	0.1167	0.1613	61.5300	65.5600	69.5505	1.0174
τ_4	60.0378	5.7141	0.1043	0.3578	0.3005	0.0263	51.2190	59.6500	69.9810	1.0146
ν_1	2.5251	0.0817	0.0015	0.0054	0.0055	0.1880	2.3920	2.5240	2.6651	1.0296
ν_2	2.9253	0.1083	0.0020	0.0076	0.0083	-0.0871	2.7540	2.9220	3.1110	1.0131
ν_3	2.7469	0.1254	0.0023	0.0088	0.0065	0.1076	2.5580	2.7370	2.9670	1.0385
ν_4	3.0594	0.3820	0.0070	0.0243	0.0194	0.0566	2.5050	3.0250	3.7371	1.0366

Table 5: Estimates and convergence diagnostics of the main parameters of interest. Descriptive statistics are based on 3000 iterations (3 independent chains of 1000 iterations each). The Table reports the sample mean and the sample standard deviation of the posteriors in the first two columns. Three estimates of the standard error follow: "Naive SE" is a naïve estimate (the sample standard deviation divided by the square root of the sample size) which assumes the sampled values are independent, "MC Error" is a timeseries estimate (the square root of the spectral density variance estimate divided by the sample size) which gives the asymptotic standard error (Geweke, 1992), "Batch SE" is a batch estimate calculated as the sample standard deviation of the means from consecutive batches of size 50 divided by the square root of the number of batches. "Batch ACF" is the autocorrelation between batch means. Q(0.05), Q(0.50), and Q(0.95) are the 5% quantile, the median, and the 95% quantile of the posterior distribution. $\hat{R}$ is the potential scale reduction (Gelman and Rubin, 1992). $\hat{R} \approx 1$ denotes good convergence.

	Data	Gaussian model				Hierarchical model			
		Median	Q(0.05)	Q(0.95)	p-value	Median	Q(0.05)	Q(0.95)	p-value
Excess kurtosis	3.168	0.401	0.328	0.477	0.000	1.695	1.534	11.041	0.244
Skewness	-0.264	-0.093	-0.122	-0.066	0.000	-0.210	-0.406	-0.081	0.202
Quantile ratio	9.652	7.104	6.564	7.739	0.000	10.840	9.020	24.252	0.616

Table 6: Posterior predictive distributions and two-sided p-values. The first column reports the statistics of interest computed on actual data. "Median", Q(0.05) and Q(0.95) are the median, the 5% and 95% quantiles of the distributions of the statistics on simulated data, respectively. The results are based on 1000 simulations.

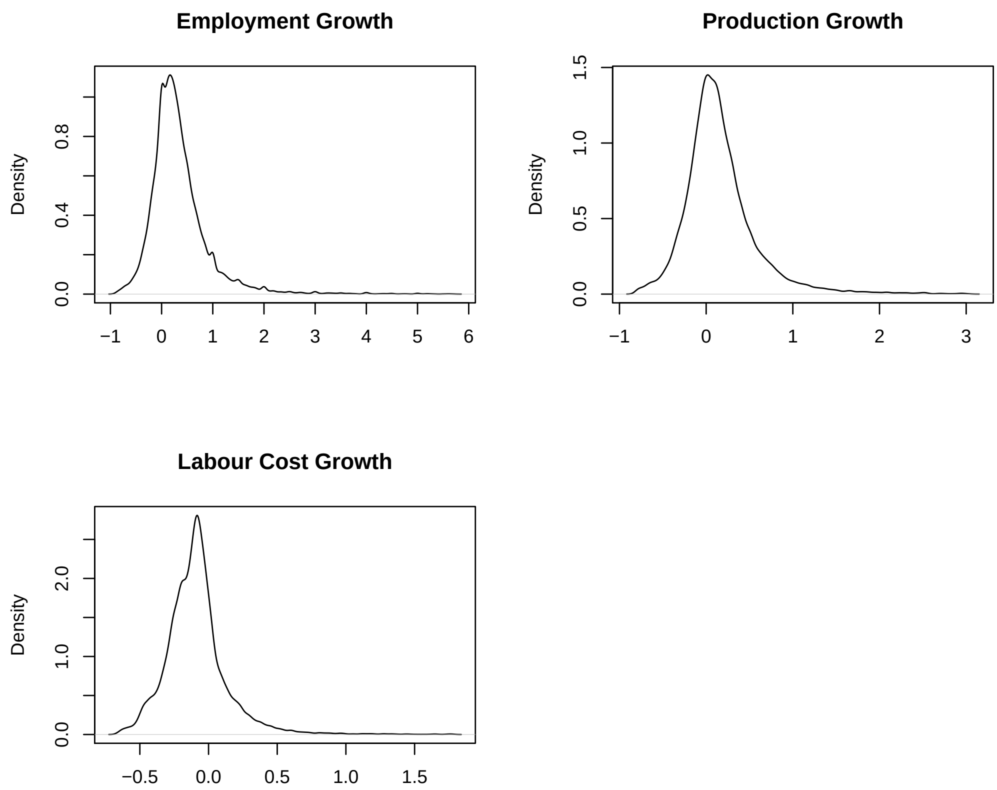

Figure 1: Estimated densities of the main variables.

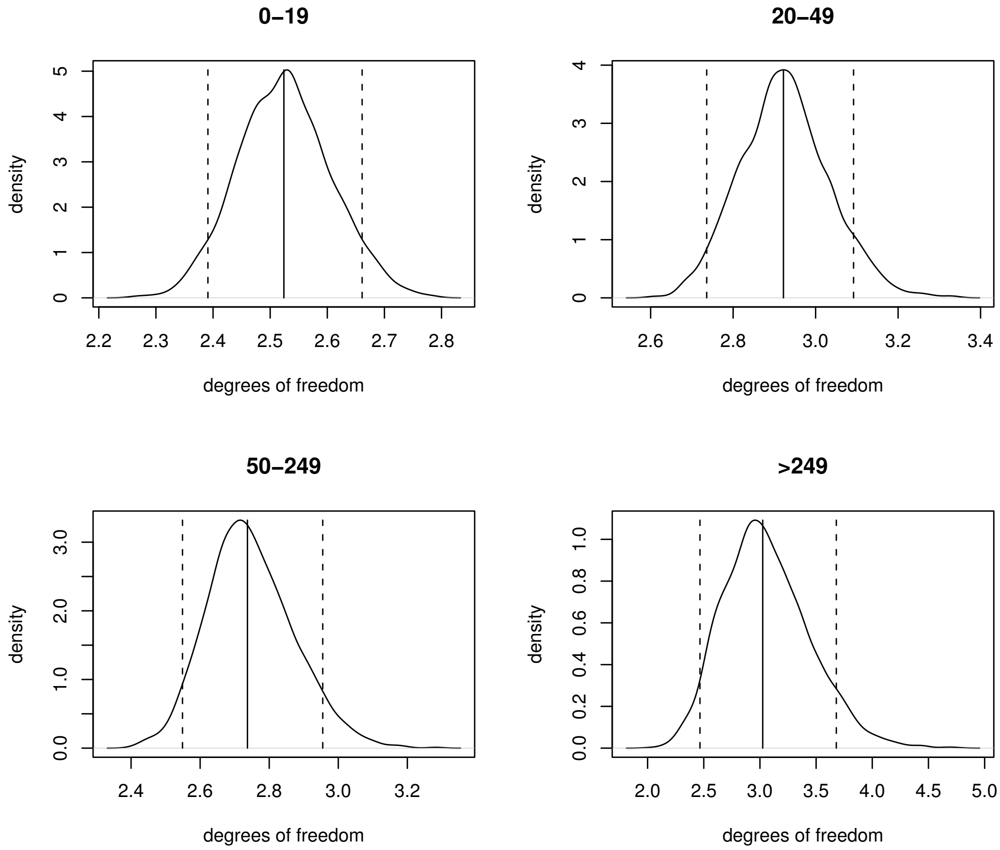

Figure 2: Estimated densities of the posterior distributions of the degrees of freedom parameter, ν_j. The title of the graph indicates the number of employees. The solid lines indicate the medians of the distributions. The dashed lines denote 90% highest posterior density (HPD) intervals.

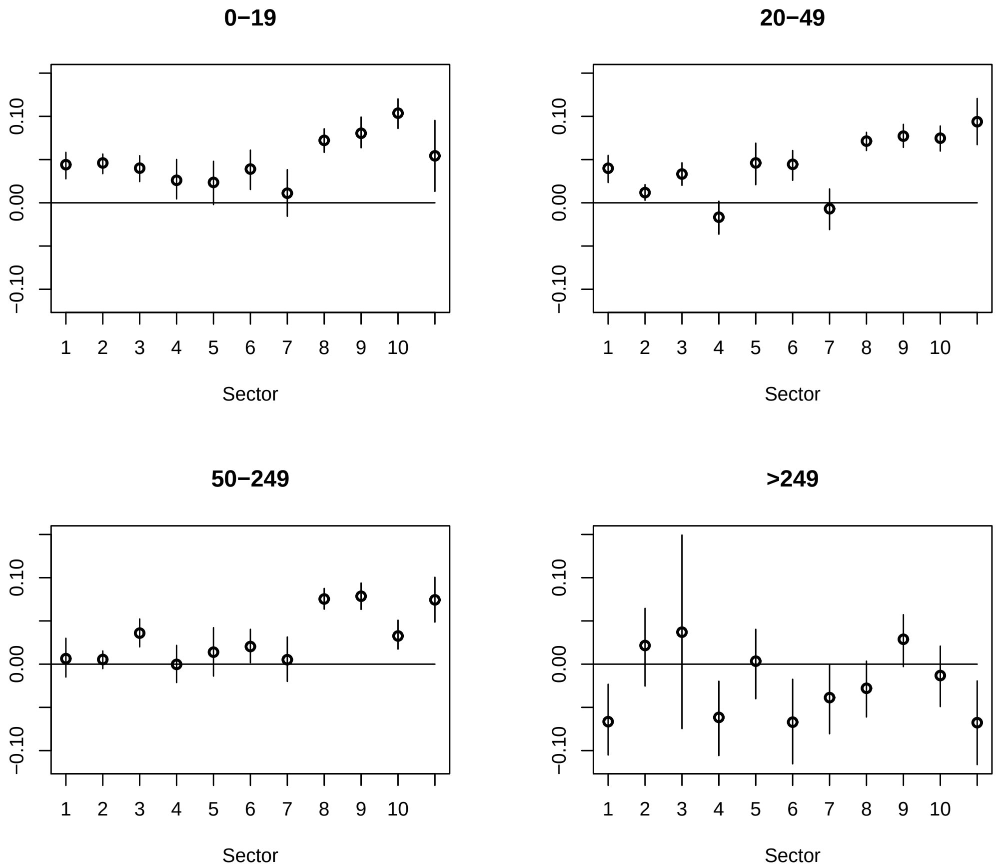

Figure 3: Estimated intercepts by sector and firm size. The title of the graph indicates the number of employees, the sectors are indicated from 1 to 11. The order of the sectors follows the official NACE Rev. 1.1 classification, with the aggregation described in the main text (DA, DB+DC, DD+DN, DE, DF+DG, DH, DI, DJ, DK, DL, DM). The points are the medians of the posterior distributions. Solid vertical lines represent 90% highest posterior density (HPD) intervals.

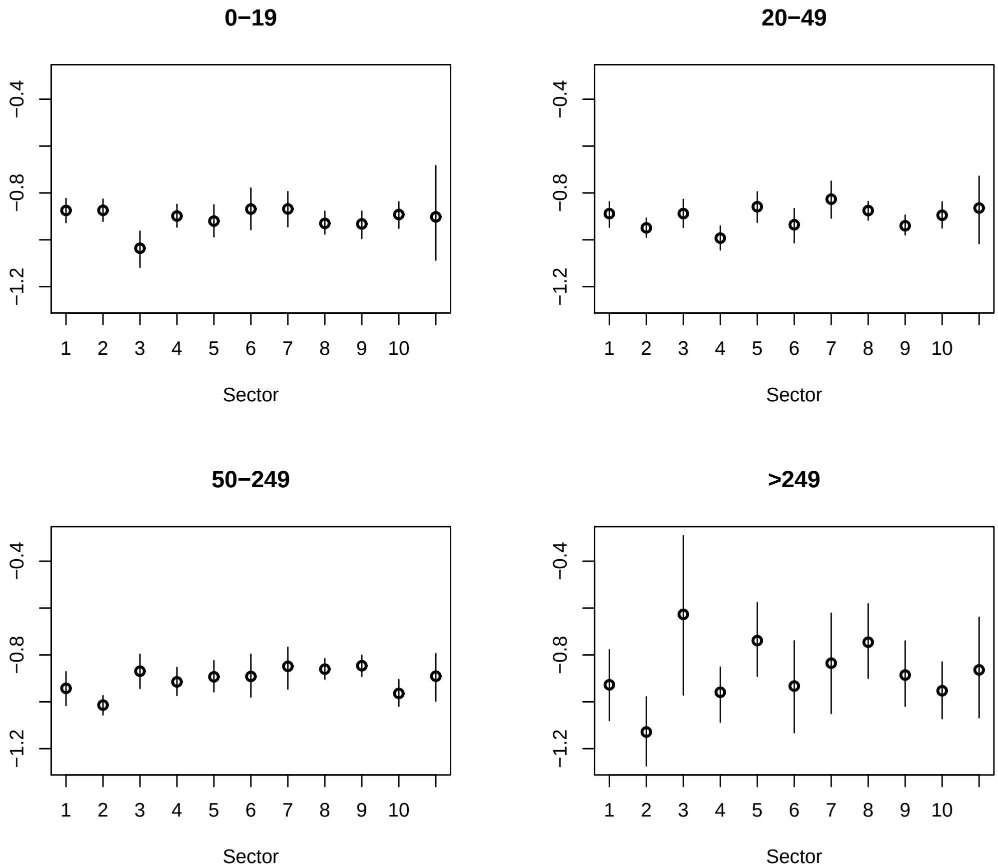

Figure 4: Estimated b_2 parameters by sector and firm size. The title of the graph indicates the number of employees, the sectors are indicated from 1 to 11. The order of the sectors follows the official NACE Rev. 1.1 classification, with the aggregation described in the main text (DA, DB+DC, DD+DN, DE, DF+DG, DH, DI, DJ, DK, DL, DM). The points are the medians of the posterior distributions. Solid vertical lines represent 90% highest posterior density (HPD) intervals.

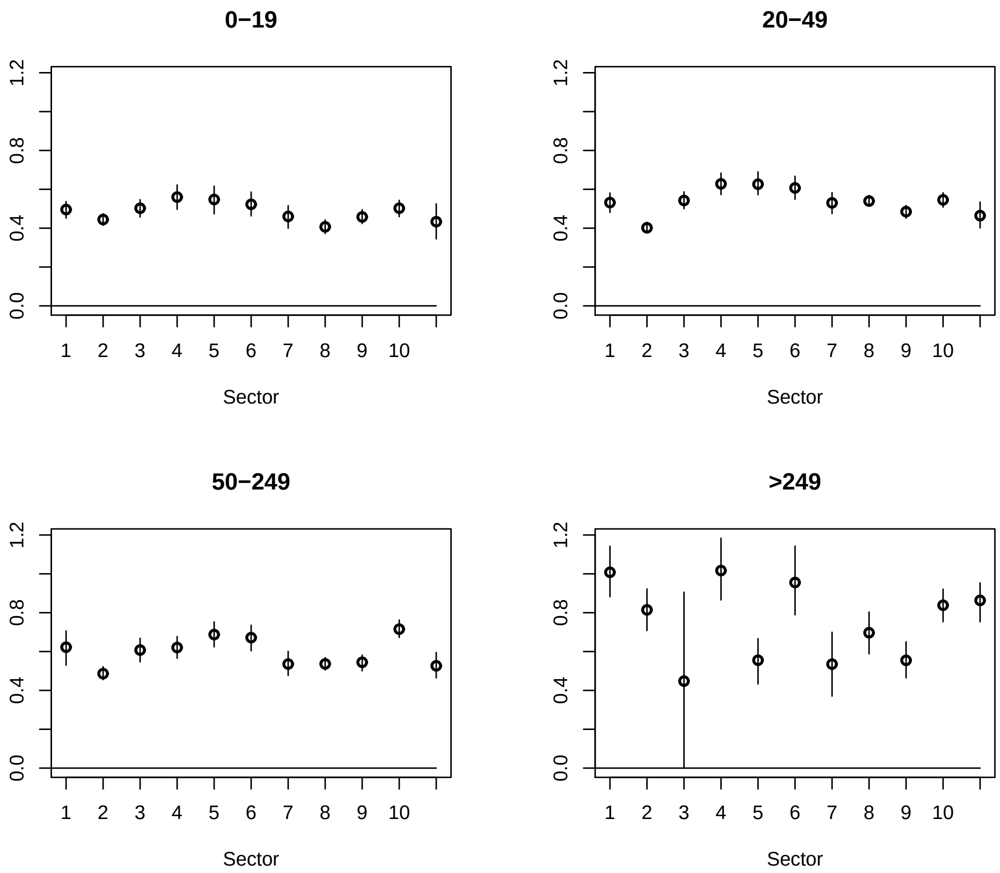

Figure 5: Estimated b_3 parameters by sector and firm size. The title of the graph indicates the number of employees, the sectors are indicated from 1 to 11. The order of the sectors follows the official NACE Rev. 1.1 classification, with the aggregation described in the main text (DA, DB+DC, DD+DN, DE, DF+DG, DH, DI, DJ, DK, DL, DM). The points are the medians of the posterior distributions. Solid vertical lines represent 90% highest posterior density (HPD) intervals.

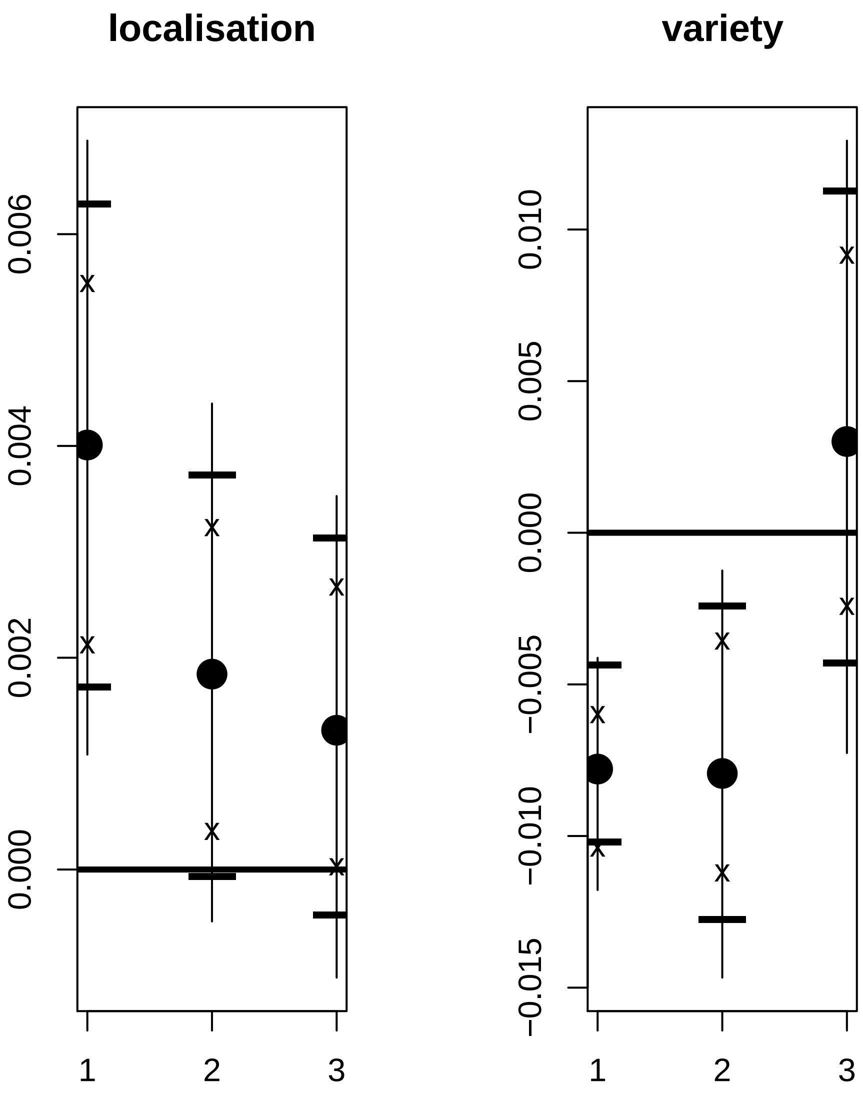

Figure 6: Estimated "localisation" (c_k, $k = 1, \ldots, 3$) and "variety" (d_k, $k = 1, \ldots, 3$) parameters. $k = 1$ indicates a distance $d \leq 2$km; $k = 2$ denotes 2km $< d \leq$ 10km; $k = 3$ stands for 10km $< d \leq$ 30km. Large points are the medians of the posterior distributions. Solid vertical lines represent 90% highest posterior density (HPD) intervals; 80% and 66% HPD intervals are represented by "–" and "x" respectively.

Appendix: Classification of manufacturing activities

Code	Numerical code	Description
DA	15, 16	Food products, beverages and tobacco
DB	17, 18	Textile and clothing
DC	19	Leather and leather products
DD	20	Wood and wood products
DE	21, 22	Pulp, paper and paper products
DF	23	Coke, refined petroleum products and nuclear fuel
DG	24	Chemicals, chemical products and man-made fibres
DH	25	Rubber and plastic products
DI	26	Non-metallic mineral products
DJ	27, 28	Basic metals and fabricated metal products
DK	29, 29	Machinery and equipment
DL	31, 32, 33	Electrical and optical equipment
DM	34, 35	Transport equipment
DN	36	Other manufacturing

Table 7: Classification of manufacturing activities

References

AUDRETSCH, D. AND M. P. FELDMAN (1996): "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86, 630–640.

BOARNET, M. G. (1998): "Spillovers and the Locational Effects of Public Infrastructures," *Journal of Regional Science*, 38, 381–400.

BRUNELLO, G., C. LUPI, AND P. ORDINE (2001): "Widening Differences in Italian Regional Unemployment," *Labour Economics*, 8, 103–129.

BRYNJOLFSSON, E. AND L. M. HITT (2003): "Computing Productivity: Firm-level Evidence," *Review of Economics and Statistics*, 85, 793–808.

CAINELLI, G. AND R. LEONCINI (1999): "Externalities and Long-term Local Industrial Development: Some Empirical Evidence for Italy," *Revue d'Economie Industrielle*, 90, 25–39.

CAINELLI, G., R. LEONCINI, AND A. MONTINI (2001): "Struttura produttiva e sviluppo regionale di lungo periodo in Italia," *Moneta e Credito*, 54, 461–485.

CINGANO, F. AND F. SCHIVARDI (2004): "Identifying the Sources of Local Productivity Growth," *Journal of the European Economic Association*, 2, 720–742.

CUNAT, A. AND G. PERI (2001): "Job Creation: Geography, Technology and Infrastructures," *Giornale degli Economisti e Annali di Economia*, 61, 23–54.

DE LUCIO, J. J., J. A. HERCE, AND A. GOILCOLEA (2002): "The Effects of Externalities on Productivity Growth in Spanish Industry," *Regional Science and Urban Economics*, 32, 241–258.

DESMET, K. AND M. FAFCHAMPS (2005): "Changes in the Spatial Concentration of Employment Across U.S. Counties: A Sectoral Analysis 1972–2000," *Journal of Economic Geography*, 5, 281–284.

FORNI, M. AND S. PABA (2002): "Spillover and the Growth of Local Industries," *Journal of Industrial Economics*, 50, 151–171.

GELMAN, A., J. B. CARLIN, H. S. STERN, AND D. B. RUBIN (2004): *Bayesian Data Analysis*, Texts in Statistical Science, Boca Raton, London, New York, Washington D.C.: Chapman & Hall / CRC, Second ed.

GELMAN, A. AND D. B. RUBIN (1992): "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–511.

GEWEKE, J. (1992): "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments," in *Bayesian Statistics*, ed. by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford: Oxford University Press, vol. 4.

——— (1993): "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics*, 8, S19–S40.

GEWEKE, J. AND W. MCCAUSLAND (2001): "Bayesian Specification Analysis in Econometrics," *American Journal of Agricultural Economics*, 83, 1181–1186.

GLAESER, E., H. D. KALLAL, J. A. SCHEINKMAN, AND A. SCHEIFER (1992): "Growth in Cities," *Journal of Political Economy*, 100, 1126–1152.

GROSSMAN, G. AND E. HELPMAN (1991): *Innovation and Growth in the Global Economy*, Cambridge, Mass.: MIT Press.

HENDERSON, V. J. (2003): "Marshalls Scale Economies," *Journal of Urban Economics*, 53, 1–28.

HENDERSON, V. J., A. KUNKORO, AND M. TURNER (1995): "Industrial Development in Cities," *Journal of Political Economy*, 103, 1067–1090.

HOLTZ-EAKIN, D. AND A. E. SCHWARTZ (1995): "Spatial Productivity Spillovers from Public Infrastructure: Evidence from State Highways," *International Tax and Public Finance*, 2, 459–468.

JAFFE, A. B., M. TRAJTENBERG, AND R. HENDERSON (1993): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 63, 483–499.

LUCAS, R. J. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3–42.

MAMELI, F., P. FAGGIAN, AND P. MCCANN (2007): "Agglomeration Economies and Employment Growth in Italian Local Labour Systems," in *Joint Congress of the European Regional Science Association and Association de Science Régionale de Langue Française*, Paris.

MARSHALL, A. (1920): *Principles of Economics*, London: MacMillan, 8th ed.

PACI, R. AND S. USAI (2006): "Agglomeration Economies and Growth. The Case of Italian Local Labour Systems, 1991-2001," Working Paper 2006/12, CRENOS.

PICCI, L. (1999): "Productivity and Infrastructure in Italian Regions," *Giornale degli Economisti e Annali di Economia*, 58, 329–353.

R DEVELOPMENT CORE TEAM (2006): "R 2.4.0 - A Language and Environment for Statistical Computing," Tech. rep., R Foundation for Statistical Computing, Vienna.

RICOTTA, C. AND L. SZEIDL (2006): "Towards a Unifying Approach to Diversity Measures: Bridging the Gap Between the Shannon Entropy and Rao's Quadratic Index," *Theoretical Population Biology*, 70, 237–243.

ROMER, P. (1986): "Increasing Returns and Long-run Growth," *Journal of Political Economy*, 94, 1002–1037.

——— (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98, S71–S102.

ROSENTHAL, S. AND W. STRANGE (2006): "The Attenuation of Human Capital Spillovers," mimeo, Syracuse University.

SHANNON, C. E. (1948): "A Mathematical Theory of Communication," *The Bell System Technical Journal*, 27, 379–423.

SIGNORINI, L. F. (1994): "The Price of Prato, or Measuring the Industrial District Effects," *Papers in Regional Science*, 73, 369–392.

SPIEGELHALTER, D. J., A. THOMAS, N. BEST, AND D. LUNN (2004): "WinBUGS User Manual," Tech. rep., MRC Biostatistics Unit, Institute of Public Health.

STORPER, M. AND A. J. VENABLES (2004): "Buzz: Face-to-face Contact and the Urban Economy," *Journal of Economic Geography*, 4, 351–370.

USAI, S. AND R. PACI (2003): "Externalities and Local Economic Growth in Manufacturing Industries," in *European Regional Growth*, ed. by B. Fingleton, Berlin: Springer-Verlag.

VAN OORT, F. G. (2007): "Spatial and Sectoral Composition Effects of Agglomeration Economies in the Netherlands," *Papers in Regional Science*, 86, 5–30.

WALLSTEN, S. J. (2001): "An Empirical Test of Geographic Knowledge Spillovers Using Geographic Information Systems and Firm-level Data," *Regional Science and Urban Economics*, 31, 571–599.