# A spatial multilevel analysis of Italian SMEs Productivity

Giorgio Fazio*
University of Palermo
University of Glasgow

Davide Piacentino
University of Naples "Parthenope"

*June* 2009

**Abstract**

In this paper, we adapt multilevel analysis methods to investigate the spatial variability of SMEs productivity across the Italian territory, and account for differences in the socio-economic context. Our results suggest that to properly capture the variability of the data, it is important to allow for both spatial mean and slope effects. Social decay has the expected negative impact. However, while this effect is larger on firms with smaller capital intensity, firms with higher capital intensity seem to be less affected by geography. Greater territorial heterogeneity emerges among those firms with lower capital to labour ratios.

**JEL Codes:** C31, R11, R12, R30

**Keywords:** Firm heterogeneity, Spatial variability, Socio-economic Context, Multilevel Analysis

*Author for correspondence: DSEAF, Facoltà di Economia, Università di Palermo, ITA-90128, *Tel.+39 091 6626254, Fax. +39 091 422988, Email giorgio.fazio@unipa.it.* The authors would like to thank, without implicating, Luisa Corrado, Leonardo Grilli, Sergio Longobardi, Enza Maltese, Otto Raspe, Erasmo Vassallo and conference participants at the 2008 ERSA congress in Liverpool.

# 1 Introduction

Particular emphasis has recently been placed in spatial economics and regional science on the "territorial" determinants of economic activity (see, among the others, Camagni, 2009, Ottaviano, 2008, and Rodriguez-Pose, 1998, 2009).[1] However, the shortage and uneasy manageability of large microeconomic datasets has favoured empirical investigations at the aggregate rather than disaggregate level, where relations should apply in theory. While improvements in both data availability and data manageability have favoured microeconometric studies of firm performance, the role of spatial and socio-institutional differences is still overlooked at this level of analysis.

In this paper, we perform a micro-level analysis of productivity to estimate how much of the observed firm-level heterogeneity is due to firm-specific factors as opposed to the spatial economic and socio-institutional differences. In this respect, Italy's variegated economic and socio-institutional geography represents a particularly well suited field of analysis. Indeed, commentators have often abduced these factors to explain the lack of regional convergence in Italy (see, among the others, Byrne, Fazio and Piacentino, 2009).[2]

In order to extract the relative variability of firm specific versus spatial specific factors, we employ multilevel analysis methods. This methodology presents a number of benefits compared to alternative more traditional approaches. First, from the modelling point of view, it explicitly acknowledges the hierarchical nature of the problem: individuals operate within higher level environments that affect their decisions. In our example, we can consider firm-

[1] The interaction between firms and the economic space is at the centre of regional and geographical economics analysis at least since Marshall (1919) first introduced the idea of "industrial atmosphere".

[2] The few studies presenting microeconometric evidence on Italian firms (for example, Guiso and Parigi, 1999, Cingano and Schivardi, 2004; Bontempi, Golinelli and Parigi, 2007; Guiso and Schivardi 2007) have not focused explicitly on the role of the complex interactions between socio-institutional and economic contexts where firms take their decisions.

level production decisions as resulting from the interaction of individual behaviour and the socio-institutional setting. Ignoring the "hierarchical" structure of the data could seriously endanger the reliability of the empirical experiment. In this respect, multilevel analysis serves as a rather simple methodology to draw inference on complex data structures, such as spatially organised data. Second, it allows estimating the heterogeneity due to individual-specific components compared to the heterogeneity due to spatial factors, whose influence may operate both in terms of mean and slope effects. Third, multilevel analysis releases the assumption of zero intra-class correlation common to more conventional estimation procedures and so relevant when dealing with economic geography. Fourth, it allows to safely bypass the endogeneity and multicollinearity issues so critical in empirical studies using aggregate data to investigate the relevance of the socio-economic context for economic activity. Finally, it allows the inclusion of group level explanatory variables, which could not otherwise be modelled using fixed effects alternatives. Bearing these considerations in mind, the remainder of the paper is organised as follows: the next section sets out the empirical strategy and outlines the methodology. Section 3 describes the dataset and discusses the results. The last section concludes.

## 2 Empirical Strategy

We assume that individual decisions are taken under the influence of an economic space hierarchically organised, where firms occupy the first (lower) and geography the second (higher) level of the hierarchy. The role of external factors is then assessed by measuring production heterogeneity due to factors observed or unobserved at the firm-level, compared to factors observed or unobserved at the higher level in the spatial hierarchy, i.e. geography. In particular, in this study, we consider administrative units at the provincial level as the second level. Since we aim at separating longitudinal differences in firms' productivity into differences due to individual and spatial factors, we investigate spatial heterogeneity within a strictly cross-sectional framework.[3]

For reasons discussed above, Multilevel Analysis (MA) is a natural candidate to perform this exercise. In terms of empirical strategy, we first estimate the firm level relationship between labour productivity and capital intensity. This specification is subsequently used as a benchmark against multilevel alternatives, where the second level is modelled both in terms of random intercepts and random slopes of capital intensity. This allows the estimation of the firm level impact of relative inputs and the "spatial" variability of both productivity and capital intensity. Later, we construct a synthetic index of the level of "socio-economic territorial embeddedness" by the means of data reduction methods, in order to control more explicitly for geographic fixed effects at the second (provincial) level. The next section describes the employed methods in greater detail.

## 2.1 Spatial Multilevel Analysis.

The first applications of Multilevel Analysis (see Hox, 2002; Goldstein, 2003; de Leeuw and Meijer, 2008) pertain to the study of pupils' performance, where higher "classes" or second levels are typically school or family effects. Only recently, MA has found application to regional economics and in particular to firm behaviour (see Raspe and van Oort, 2007). The features of MA make it a natural tool for spatial analysis, where the particular geography of a

[3]Adding a time dimension would in principle allow us to investigate variations due to the business cycle. These variations may or may not be relevant at the spatial level, depending for example on the extent of sectoral/spatial interdependence. However, data issues (described below) have prevented us to pursue this strategy at this stage. Also, given the more static nature of the socio-institutional environment, the time dimension would probably add very little our ability to investigate spatial heterogeneity, whilst subjecting our analysis to grater risk of endogeneity and serial correlation issues.

territory can be considered as a higher level effect on firm production decisions and performance. As mentioned above, MA presents a number of benefits compared to more traditional methods. It recognises the hierarchical nature of the data, it releases the over-binding assumption that observations within sub-units are zero-correlated, it allows the analysis of the level specific variability of output both through mean and the slope effects, and last but not least it avoids endogeneity issues between the observational unit (the firm) and the variables of interest.

In order to illustrate the methodology, we can develop from the familiar grounds of a plain-vanilla log-linearised Cobb-Douglas per worker production function:

$$y_{ij} = \beta_0 + \beta_1 k_{ij} + \epsilon_{ij}, \tag{1}$$

where the subscript $i$ refers to the individual unit or first level and $j$ refers to the second level in the hierarchy, $y = \log(Y/L)$ is the log of output per worker or *labour productivity*, $k = \log(K/L)$ is the log of the stock of capital per worker or *capital intensity*, and $\epsilon$ is a randomly distributed error term. Clearly, equation (1) makes no effort to accommodate for potential (and likely) heterogeneity which may arise at the $j$-level: all the geographical factors are assumed to have identical impact on the firm's per-worker production. Therefore, all firms are assumed to have identical intercepts, $\beta_{01} = \beta_{02} = \ldots = \beta_0$ and capital efficiency, $\beta_{11} = \beta_{12} = \ldots = \beta_1$. Multilevel analysis allows to explicitly model the potential hierarchical nature of the problem using a pair of *linked* models. Equation (1) can be easily extended to allow for second level mean-effects:

$$\begin{aligned} y_{ij} &= \beta_{0j} + \beta_1 k_{ij} + \epsilon_{ij}, & \epsilon_{ij} &\sim N(0, \sigma^2) & &\textit{Level}\ 1 \\ \beta_{0j} &= \gamma_{00} + u_{0j}, & u_{0j} &\sim N(0, \tau_{00}) & &\textit{Level}\ 2 \end{aligned}$$

At level 2, the spatial level intercept is specified as the sum of an overall mean ($\gamma_{00}$) and a series of random deviations from that mean ($u_{0j}$). Grouping the two levels, it is possible to obtain the following estimating equation:

$$y_{ij} = \underbrace{[\gamma_{00} + \beta_1 k_{ij}]}_{Deterministic} + \underbrace{[u_{0j} + \epsilon_{ij}]}_{Stochastic}, \tag{2}$$

where labour productivity is assumed to be the result of both a deterministic and a stochastic part (i.e. random effects), which in this case are the spatial level intercepts.

Equation (2) can also be easily extended to allow for (random) variations in the spatial slopes of capital intensity:

$$\begin{cases} y_{ij} = \gamma_{00} + \beta_{1j} k_{ij} + u_{0j} + \epsilon_{ij};\ \beta_{1j} = \gamma_{10} + u_{1j},\ u_{1j} \sim N(0, \tau_{10}) \\ y_{ij} = [\gamma_{00} + \gamma_{10} k_{ij}] + [u_{0j} + u_{1j} k_{ij} + \epsilon_{ij}] \end{cases}$$

In order to investigate the relevance of socio-economic factors for differences in firms' productivity, we add to the above specifications an indicator of socio-economic territorial embeddedness (SETE) calculated at the provincial level, as described in section 3. The specification then becomes:

$$y_{ij} = [\gamma_{00} + \gamma_{10} k_{ij} + \gamma_{20} SETE_j] + [u_{0j} + u_{1j} k_{ij} + \epsilon_{ij}] .$$

However, estimates of the above equations may turn inconsistent in presence of endogeneity between the level 1 explanatory variables and level 2 error terms. A simple endogeneity test (see discussion in the Appendix) can be performed by adding the level 2 mean of the level 1 explanatory variable, i.e. the provincial mean of capital intensity, and testing for its significance.

Two statistics are used for model evaluation. The first is the Intraclass Correlation Coefficient (ICC) which returns the amount of total variance accounted for by the variance between classes. Depending on the specification adopted, the ICC is:

$$\rho_1 = \frac{\tau_{00}}{\tau_{00}+\sigma^2} \quad \rho_2 = \frac{\tau_{00}}{\tau_{00}+\tau_{10}+\sigma^2}$$

The second statistic makes a simple Likelihood Ratio comparison between alternative models, i.e. given models A and B, $LR = -2(\log L_A - \log L_B)$, which under the null is distributed as a chi-square with degrees of freedom given by the difference in the number of parameters between the models. We now turn to the empirical implementation.

# 3 Empirical Implementation

## 3.1 Data.

In order to perform the empirical estimation, we have queried the Italian section of the Bureau Van Dijk Database (AIDA), which collects balance sheet data on almost 90 percent of the existing Italian firms with value of production beyond 100.000 Euros. Given our cross-sectional focus, we have collected data for the year 2005.[4] We have limited our sample to Small and Medium Enterprises (SMEs) using standard criteria (more than 10 and less than 250 employees and Total Assets between 2 and 43 million Euros). Arguably, SMEs are both more likely to be affected by the surrounding economic and socio-institutional environment and less able to influence it.[5] Moreover, the analysis has been restricted to manufacturing firms only by selecting the [15-37] sectoral range in the ATECO 2002 classification.[6] This further reduces the risk that sectoral effects might interfere with spatial effects, our main interest. This query together with further controls for data inconsistencies returned 7,097 observations distributed across the national territory.

[4]We have allowed a lag of two years to reduce the risk on data inconsistencies, which may be more present in more recent data.

[5]This also allows us to avoid issues of endogeneity running from level 1 to level 2.

[6]Sector 16, the Tobacco industry, has been excluded because it is an Italian State Monopoly.

Importantly, unlike more traditional approaches, such as fixed effects estimation, MA is robust the problem of irregular class frequencies originating from the uneven distribution of firms across the Peninsula.

The analysis has been performed on firm's output per worker, defined as the log of value added per employee. To avoid potential level 1 endogeneity, we have used as a firm level regressor a one year lag of capital intensity, defined as the log of capital stock per employee.[7]

## 3.2 Socio-Economic Territorial Embeddedness.

Defining the socio-economic territorial context is a complex and highly debated matter (see, for example, Rodriguez-Pose, 1998, 2009, and Camagni, 2009). For the purposes of this paper, we are interested in a synthetic indicator that may with reason represent the many socio-economic features of a particular territory. To construct such indicator, we have started off from the largest possible set of variables collected from the Italian National Statistical Office (ISTAT) database and have then applied two rounds of data reduction both on statistical and economic grounds. The 46 variables collected were first reduced to 11 by the means of a simple multicollinearity restriction exclusion, where only those correlated less than 80% were kept. Whenever two variables were correlated more than 80%, one was selected on the grounds of economic interpretation. The 11 indicators of socio-economic context are listed in table 1. Some of these refer more closely to the macroeconomic scenario of the manufacturing sector (e.g., labour productivity and employment), the economy as a whole (unemployment), or to the level of competitiveness (e.g. degree of openness and self-employment). Others are more relevant to the socio-institutional context (criminality, equality of opportunities (gender

[7] To increase the representativeness of the sample, we have only included firms active at least from 2003-2005.

employment), and infantile mortality). A final indicator captures the extent of human capital attraction (net brain flow).

Table 1 - Indicators of the Socio-Economic Context

| Label | Indicator | Years |
|---|---|---|
| Labour Productivity | Industry value added per labour unit | mean(1999-2003) |
| Openness | Export plus Imports over value added | mean(1999-2003) |
| Gender Employment | Male minus Female Employed (15-64 age range) | mean(1999-2003) |
| Employment | Employed in the Industry sector over total (percentage) | mean(1999-2003) |
| Self Employment | Self employed in the industry sector over total employed in the Industry sector | mean(1999-2003) |
| Unemployment | Unemployment rate | mean(1999-2003) |
| Crime1 | Number of voluntary manslaughter and attempted homicides per 100.000 inhabitants | mean(1999-2003) |
| Crime2 | Number of extorsions per 100.000 inhabitants | mean(1999-2003) |
| Crime3 | Number of bad cheques per 1.000 inhabitants | mean(1999-2003) |
| Net Brain Drain | Graduates born in other provinces or abroad per 100 graduates moved to other provinces | 1999 |
| Infantile Mortality | Infant mortality rate | mean(1999-2003) |

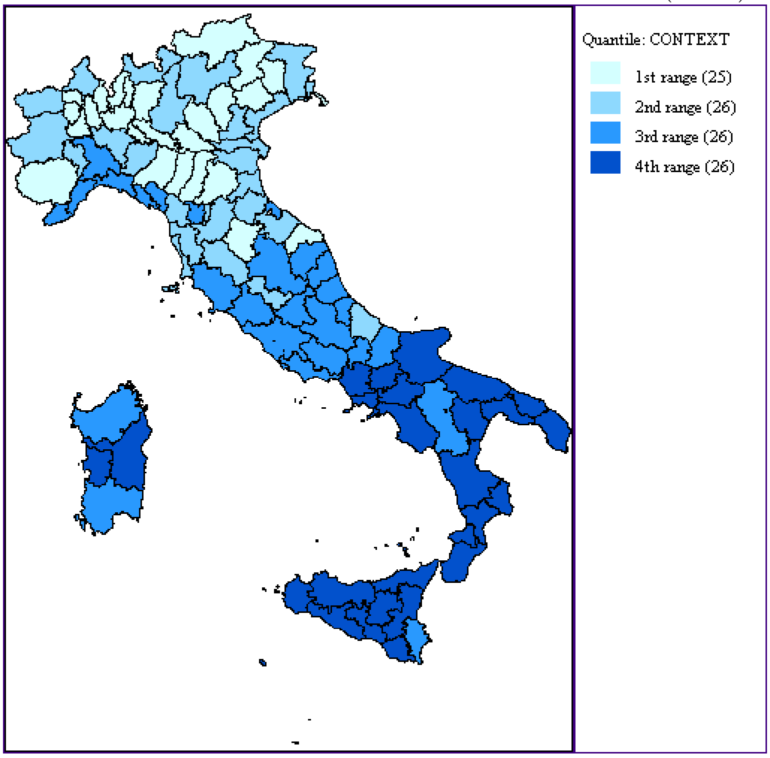
Secondly, a principal components analysis (PCA) was performed to collapse this set of variables into a synthetic indicator, which is a linear combination of the original variables, with weights derived to account for the largest part of data variability. Eigenvalues and eigenvectors of the PCA are reported in table 2. As it can be seen, the selected component explains just above 50% of the overall variance with an eigenvalue equal to 5.53. The eigenvector shows how the coefficients are similar, ranging (in absolute values) from 0.22 to 0.39; i.e. the variables enter the component with similar weights. The synthetic indicator obtained from the PCA is such that higher (positive) values denote worsening contexts, and viceversa. Hence, it can be considered as representative of the extent of the socio-economic "decay" of provinces. We refer to this indicator as the degree of Socio Economic Territorial Embeddedness (SETE).

Table 2 - Principal Components Analysis (PCA)

| | Component 1 (PCA) |
|---|---|
| Eigenvalue (% of variance) | 5.52948 (50.27) |
| Eigenvector | |
| Labour Productivity | -0.2368 |
| Openness | -0.2981 |
| Gender Employment | 0.3171 |
| Employment | -0.3157 |
| Self Employment | -0.2202 |
| Unemployment | 0.3918 |
| Crime1 | 0.2987 |
| Crime2 | 0.2394 |
| Crime3 | 0.3257 |
| Net Brain Drain | -0.3237 |
| Infantile Mortality | 0.3090 |

In figure 1, the provinces across the Italian territory have been coloured in relation to their quartile "score" in terms of the synthetic indicator, to highlight the spatial distribution of the virtuous (lighter colours) and the less virtuous (darker colours) provinces. Visually, a large degree of heterogeneity seems to emerge between macro-areas with levels of socio-economic decay increasing as one moves from North to South. In particular, according to our synthetic indicator, provinces in some parts of the North (mostly in Lombardia, Emilia Romagna and Veneto) come out as the ones with better socio-economic contexts. On the contrary, provinces located in the South of Italy are characterised by higher levels of socio-economic decay. In light of the existing evidence on the Italian territory, this is not too surprising.

Figure 1 - Indicator of Socio-Economic Territorial Embeddedness (SETE)




# 4 Results

As mentioned above, we begin from the benchmark single level specification and then start allowing for spatial (provincial) level 2 random intercepts and slopes of capital intensity, $k_{ij}$.[8] Regression results are presented in table 3, where together with coefficient estimates we report the residuals variance, $\sigma^2_\epsilon$, the variance of the second level intercepts, $\tau_{00}$, and slopes, $\tau_{10}$, and the Intra-Class Correlation (ICC). For model comparison, we have also included two sets of likelihood ratio tests, ($LR_1$ and $LR_2$), comparing respectively the

[8]Estimations have been carried out using Restricted Iterative Generalised Least Squares (RIGLS) in MlWin 2.10. As discussed in Aslam and Corrado (2007), the RIGLS estimator overcomes the potential downwards bias issue of the non restricted version. For robustness, we have also employed Restricted Maximum Likelihood Estimation (RMLE) in backstage regressions. Since no meaningful differences were detected, these results are not reported, but are available upon request from the authors.

estimated model to the constant only specification (not reported) and to the specification in the preceding column.

Table 3 - Multilevel Analysis (RIGLS)

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $k_{ij}$ | 0.05 | 0.061 | 0.074 | 0.075 | 0.075 |
| | (0.005) | (0.005) | (0.007) | (0.007) | (0.007) |
| $SETE_j$ | | | | -0.01 | -0.01 |
| | | | | (0.003) | (0.004) |
| $\bar{k}_j$ | | | | | -0.017 |
| | | | | | (0.040) |
| Constant | 4.557 | 4.487 | 4.425 | 4.414 | 4.493 |
| | (0.021) | (0.022) | (0.033) | (0.034) | (0.183) |
| $\sigma_\epsilon^2$ | 0.041 | 0.040 | 0.039 | 0.039 | 0.039 |
| $\tau_{00}$ | | 0.002 | 0.032 | 0.034 | 0.034 |
| $\tau_{10}$ | | | 0.001 | 0.002 | 0.002 |
| ICC | | 0.05 | 0.44 | 0.45 | 0.45 |
| $-2\log L$ | -2442.89 | -2669.858 | -2692.155 | -2706.053 | -2706.2 |
| $LR_1$ | 122.968 | 349.936 | 372.233 | 386.131 | 386.278 |
| $LR_2$ | 122.968 | 226.968 | 22.297 | 13.898 | 0.147 |

Standard Errors are reported in parentheses

Column one presents the estimates of the benchmark model , where as expected capital intensity shows up as a positive and highly significant determinant of firm level productivity. The variance of the residuals is a mere 4 percent. In column two, we have allowed intercepts to vary across provinces. This second model should allow us to capture the provincial effects in the variation of firm level productivity. The likelihood ratio test indicates that this specification yields greater information than the benchmark. Surprisingly, the variance of intercepts is extremely small at just 0.2 percent and the ICC concludes that only 5 percent of the variability of firm-level productivity is due to provincial spatial variations. Given the known dispersion of economic activity across Italy, we would have expected spatial variations to be far more relevant. The next specification, however, shows that this result may be due to unaccounted provincial variability in the slopes of capital intensity, as its effect on productivity may differ across space. This is tested in column three, where slopes are also allowed to vary across provinces. The variance of the slopes is extremely small at only 0.1 percent. However, allowing for such small variance in the spatial slopes of capital intensity allows us to capture the variation due to spatial intercepts that we were expecting. The intercept variance is now around 3 percent and the ICC shows that 44 percent of the firm level variance is due to the variation across provinces. It is important to stress that if we had not allowed for slopes to vary, we would have been misled to under represent the spatial differences between firms across Italian provinces. Allowing for the slope of capital intensity to vary across space has also increased slightly the "deterministic" coefficient on capital intensity. The likelihood ratio tests concludes that this specification is superior to the one where varying slopes are not allowed.

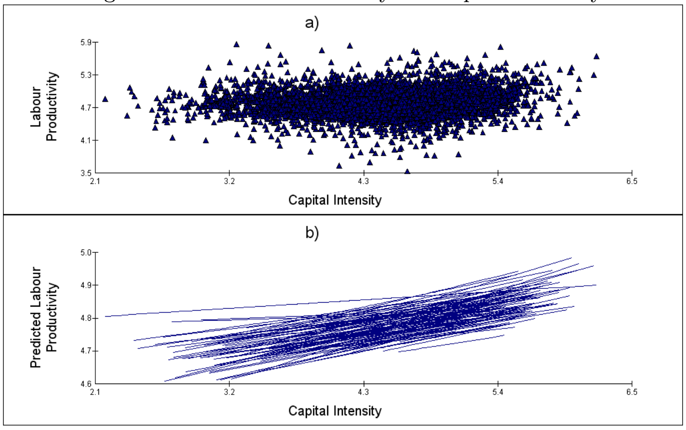
We can now introduce (see column four) the index of Socio-Economic Territorial Embeddedness (SETE) in order to assess how socio economic decay affects the productivity of firms. Our constructed indicator, SETE, enters the regression with the expected negative sign and is statistically significant. Worse socio economic scenarios lead to lower firm-level productivity, as predicated by the recent regional and geographical economics literature. The likelihood ratio also shows that this variable adds significant information to the previous specification.[9]

Figure 2 presents the scatter plot of firm level labour productivity against capital intensity (panel a) and the estimated provincial slopes from the model in column four of table 3 (panel b). Comparison of the two panels allows us

[9]Clearly, we can imagine the negative effects of socio-economic decay to impact also on firm natality and, therefore, on firms not present in the sample, because never born. Accounting for the role of the territory on such unborn, and hence unobserved, firms is a challenging and interesting line for future research.

to assess how well our spatial multilevel model is able to capture the data dispersion in the scatter plot. Indeed, a single level model would have missed the great variation across provinces evident both in the intercepts and in the slopes of the relationship between productivity and capital intensity.

Figure 2 - Labour Productivity vs. Capital Intensity




To see whether the variability of intercepts and slopes across provinces follows the canonical North-South dichotomy, we compare the distributions of the intercepts and the slopes of capital intensity of Southern provinces with that of provinces in the Centre-North. In figure 3 we show how the kernel density of the intercepts (left panel) of Southern provinces lies to the left of that of Northern provinces and the kernel density of the slopes (right panel) lies to the right of that of Northern provinces (and to the right of the overall Italian distribution).

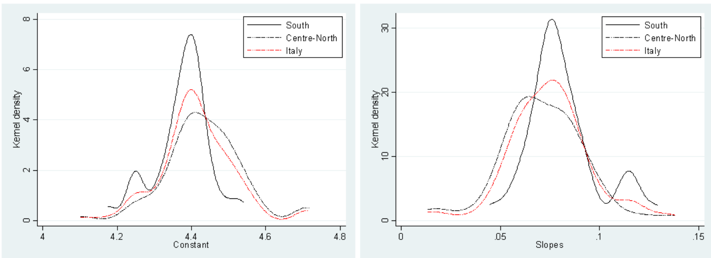
In table 4 we present a Kolmogorov-Smirnov test and a t-test of sample mean equality for the intercepts and the slopes. The first suggests that the distributions of intercepts and slopes for the two subgroups are statistically

Table 4 - Tests of mean equality for constants and slopes

| | | | |
|---|---|---|---|
| | Kolmogorov-Smirnov Test | | |
| | D | p-value | corrected |
| Intercepts | 0.3659 | 0.004 | 0.002 |
| Slopes | 0.3062 | 0.027 | 0.016 |
| | | | |
| | T-test | | |
| | $H_0 : diff. = mean(South)–mean(Centre - North) = 0$ | | |
| | $H_a : diff. < 0$ | $H_a : diff. \neq 0$ | $H_a : diff. > 0$ |
| Intercepts | Pr(T < t) = 0.0010 | Pr(\|T\| > \|t\|) = 0.0019 | Pr(T > t) = 0.9990 |
| Slopes | Pr(T < t) = 0.9933 | Pr(\|T\| > \|t\|) = 0.0135 | Pr(T > t) = 0.0067 |

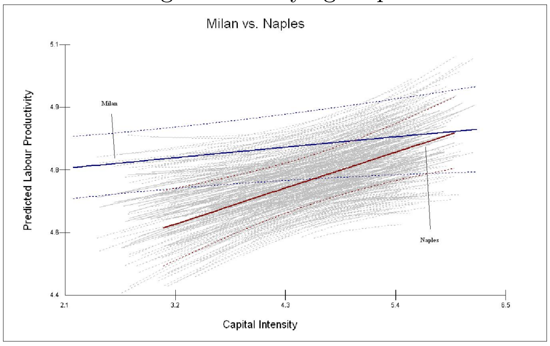
different. The two sample t-test shows how the mean of the intercepts of provinces in the Centre-North is significantly greater that the mean of the intercepts of provinces in the South. However, the mean slope of capital intensity for provinces in the South is significantly greater than the mean slope for provinces in the Centre-North.

Figure 3 - Kernel Density of Intercepts (left panel) and Capital Intensity Slopes (right panel)




To provide a graphical example, we have highlighted in figure 4 the capital intensity slopes (and 95% confidence bands) of two provinces, Milan and Naples, representative respectively of the North and the South. Clearly, the relationship between labour productivity and capital intensity is flatter in Milan than in Naples. The confidence bands allow us to conclude that when capital intensity is higher productivity is not statistically different in the two provinces. On the other hand, statistically significant spatial differences in productivity emerge when capital intensity is lower.

Figure 4 - Varying Slopes

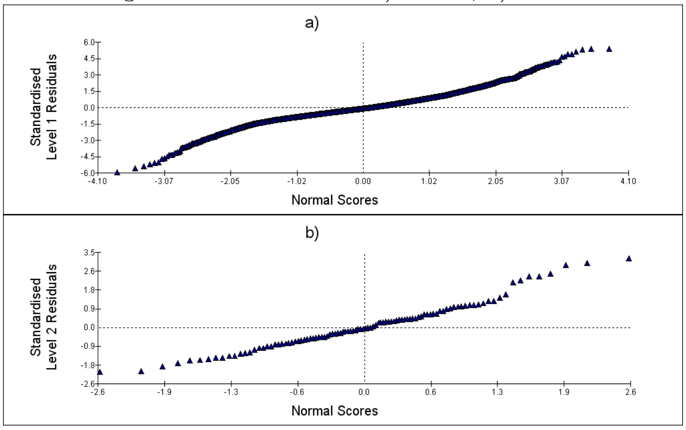



*Robustness*

Two checks of model validity and robustness are important. Firstly, we want to be reassured that the residuals at each level follow normal distributions. In order to verify this assumption, we have represented in figure 5 the Normal probability plots (i.e. the ranked residuals vs. the corresponding points on a normal distribution curve) of the level 1 (panel a) and level 2 (panel b). Both plots look fairly linear, reassuring us on the Normality assumption.

A further important robustness check pertains to the possibility of cross-level endogeneity (between the level 1 explanatory variables and level 2 error terms), which would make results inconsistent. To test for endogeneity, we follow the approach suggested by Grilli and Rampichini (2006) and include the level 2 mean of the level 1 explanatory variable, $\bar{k}_{ij}$, as an additional regressor

Figure 5 - Model Check: a) level 1; b) level 2




(see Appendix A). As it can be seen in column five of table 3, this variable is not statistically significant. According to Grilli and Rampichini, this can be taken as an indication of no endogeneity problems. This strenghtens the robustness of our previous estimates.

# 5 Conclusions.

Recent contributions in spatial economics and regional science have emphasised the role of "territorial" factors for differences in economic performance. Empirical tests of this relationship, however, have often been at the aggregate rather than the disaggregate level. In this paper, we have adapted multilevel methods to the analysis of spatial differences in the firm level productivity of small and medium enterprises across the Italian peninsula. Compared to standard approaches, this approach yields a number of important benefits. It allows to simply and explicitly model the hierarchical structure of the data, which may arise both at the mean and the slope spatial levels, allowing to capture data heterogeneity to a greater extent. It lets a simple estimation of the firm-level heterogeneity compared to the spatial level heterogeneity. Unlike traditional regression methods, it acknowledges the non-zero intra-class correlation, and allows a solution to the endogeneity and multicollinearity issues which hamper empirical studies on aggregate data. Finally, it allows a mix of fixed and random effects, allowing direct tests of the relevance of spatial factors. In particular, we introduce and test explicitly an indicator of socio-economic decay across Italian provinces.

A number of results have emerged. A first notable result is that allowing for the variability of both intercepts and slopes leads to a significant improvement in our ability to capture firm-level heterogeneity in productivity. Second, our result suggest that in order to avoid underestimating spatial differences, it is important to allow for spatial slope effects. Third, we are able to conclude that worse territorial socio-economic conditions do lower firm-level productivity. A final interesting result is that while Northern provinces have on average greater productivity of Southern ones, the latter have on average greater slopes of capital intensity.

Since capital intensity is not a measure of firm size, but an indication of firm technology, this implies that less capital intensive (more labour intensive) firms are more affected by location and geography than firms with high capital intensity. To further stretch the argument, we can conjecture that firms with higher capital intensity in Southern provinces seem able to overcome the negative effects of social decay. We can probably think of this as a form of "internalisation" of the negative externality of social decay, where individual efforts and abilities are able to counteract the negative influence of worse geography. On the other hand, firms with lower capital intensity are the ones who benefit more greatly from the spillovers of a better location or suffer more from a negative socio-economic background. Firms in the Centre-North are then able to enjoy greater productivity levels even in presence of lower levels of capital investment. Firms in the South need to invest more if they want to compete at the levels of firms in the rest of the country.

## References


Aslam, A. and L.Corrado (2007) No man is an island: The inter-personal determinants of regional well-being and life satisfaction in europe, Cambridge Working Paper in Economics 0717.

Bontempi, E., R. Golinelli, and G. Parigi (2007) Why demand uncertainty curbs investment: Evidence from a panel of italian manufacturing firms, Temi di discussione Banca d'Italia(Economic Working Papers) 621.

Byrne, J.P., G. Fazio, and D. Piacentino (2009) Total factor productivity convergence among italian regions: Some evidence from panel unit root tests, Regional Studies 43, 63-76.

Camagni, R. (2009) Territorial capital and regional development, in R.Capello, and P. Nijkamp, ed.: Handbook of Regional Growth and Development Theories, Cheltenham: Elsevier, 118-132.

Cingano, F., and F.Schivardi (2004) Identifying the sources of local productivity growth, Journal of the European Economic Association, 2, 720-742.

de Leeuw, J., and E.Meijer (2008) Handbook of multilevel analysis, Springer.

Goldstein, H. (2003) Multilevel Statistical Models, Arnold.

Grilli, L., and C. Rampichini (2006) Model building issues in multilevel linear models with endogenous covariates, Proceeding of IASC-INTERFACE-IFCS Workshop, http://www.iasc-isi.org/Proceedings/2006/COMPSTAT_Satellites/KNEMO

Guiso, L., and G.Parigi (1999) Investment and demand uncertainty, The Quarterly Journal of Economics, 114, 185-227.

Guiso, L., and F. Schivardi (2007) Spillovers in industrial districts, Economic Journal, 117, 68-93.

Hox, J. (2002) Multilevel Analysis: Techniques and Applications, Lawrence Erlbaum Associates.

Marshall, A. (1919) Industry and Trade, London: Macmillan.

Mundlak, Y., (1978) On the pooling of time series and cross section data, Econometrica, 46, 69-85.

Ottaviano, G. (2008) Infrastructure and economic geography: An overview of theory and evidence, EIB Papers, 13, 9-35.

Raspe, O., and F. van Oort (2007) What when space matters little for firm productivity? a multilevel analysis of localised knowledge externalities, Papers in Evolutionary Economic Geography, 0706.

Rodriguez-Pose, A. (1998) Dynamics of regional growth in Europe: Social and political factors, Oxford University Press.

Rodriguez-Pose, A. (2009) Do institutions matter for regional development?, RSA Annual Conference: Understanding and Shaping Regions: Spatial, Social and Economic Futures.

# A Appendix: Endogeneity in Multilevel Models

In general, endogeneity arises when unobserved (omitted) covariates are correlated with the observed (included) covariates and with the response variable. In multilevel models, however, a further type of endogeneity can arise if the level 2 random effects are correlated with a level 1 covariate (level 2 endogeneity). Since in presence of endogeneity standard estimators become inconsistent, this issue is receiving increasing attention in the literature, and a new detection method has been recently proposed by Grilli and Rampichini (2006). To illustrate this method, let us consider a random intercept model with a single level 1 covariate $X_{ij}$:

$$Y_{ij} = \alpha + \beta X_{ij} + v_j + \epsilon_{ij} \tag{A.1}$$

Level 2 endogeneity occurs when $Cov\left[v_j, X_{ij}\right] \neq 0$, so that $E\left[v_j \mid X_{ij}\right] \neq 0$ and thus the standard estimators produce inconsistent estimates of $\beta$. In order to investigate the presence of level 2 endogeneity, Grilli and Rampichini (2006) treat the covariate $X_{ij}$ as a random variable with between and within clusters variations:

$$X_{ij} = X_j^B + X_{ij}^W \tag{A.2}$$

where it assumed that $X_j^B$ are *i.i.d.* with mean $\mu_X$ and variance $\tau_X^2 > 0$, $X_{ij}^W$ are *i.i.d.* with zero mean and variance $\sigma_X^2 > 0$, and $X_j^B \perp X_{ij}^W \quad \forall\, i, j$.

Taking into account the decomposition in A.2, A.1 can be written as follows:

$$Y_{ij} = \alpha + \beta^W X_{ij}^W + \beta^B X_j^B + v_j + \epsilon_{ij} \tag{A.3}$$

From an alternative parameterisation of A.3, we can obtain the following specification:

$$Y_{ij} = \alpha + \beta^W X_{ij} + \delta X_j^B + \upsilon_j + \epsilon_{ij} \tag{A.4}$$

where $\delta = \beta^B - \beta^W$.

If $X_j^B$ is omitted from equation A.4, and consequently included in the level 2 error term, the model becomes as follows:

$$Y_{ij} = \eta + \beta^W X_{ij} + \nu_j + \epsilon_{ij} \tag{A.5}$$

where $\eta = (\alpha + \delta\mu_X)$ and $\nu_j = \delta\left(X_j^B - \mu_X\right)$, with $E(\nu_j) = 0$ and $Var(\nu_j) = \tau^2_{Y|X} = \delta^2 \cdot \tau^2_X + \tau^2_{Y|X^B X^W}$.

Hence, $Cov(\nu_j, X_{ij}) = Cov\left(\nu_j, X_j^B\right) = \delta\tau^2_X$, which is different from zero if and only if $\delta \neq 0$. In other words, when the between and within slopes differ, the correlation between $\nu_j$ and $X_{ij}$ implies $E(\nu_j \mid X_{ij}) \neq 0$, i.e. a certain degree of level 2 endogeneity.

Unfortunately, $X_j^B$ and $X_{ij}^W$ are unobservable and consequently model A.4 cannot be fitted. However, Grilli and Rampichini (2006) suggest using as observable analogues the cluster mean $\overline{X} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}$ for $X_j^B$ and the deviation from the cluster mean $\widetilde{X}_{ij} = X_{ij} - \overline{X}_j$ for $X_{ij}^W$. Then, equation A.4 becomes:

$$Y_{ij} = \alpha + \beta^W X_{ij} + \delta\bar{X}_j + z_j + \epsilon_{ij} \tag{A.6}$$

where $z_j = \upsilon_j - \delta\overline{X}_j^W$ with $E(z_j) = 0$. Under $H_0 : \delta = 0$, there is no level 2 endogeneity. As suggested by Mundlak (1978), this test on the slope of $\overline{X}_j$ in model A.6 is equivalent to an Hausman test.