Explaining TFP at firm level in Italy. Does location matter?

Aiello, Francesco and Pupo, Valeria and Ricotta, Fernand

University of Calabria, Dept. of Economics and Statistics

30 December 2011

Explaining TFP at firm level in Italy. Does location matter?

Francesco Aiello, Valeria Pupo, Fernanda Ricotta
[f.aiello@unical.it; v.pupo@unical.it; f.ricotta@unical.it]
University of Calabria, Department of Economics and Statistics, Italy
87030 Arcavacata di Rende (CS)

Abstract This study considers how firms' internal variables and regional factors affect the total factor productivity of Italian manufacturing firms. Due to of the hierarchical structure of data in estimation, we employ a multilevel model. Results, which refer to 2006, show the importance of firm-specific determinants of TFP, but at the same time confirm the role of regional context in explaining the gap in TFP levels which exist between the South and the North of Italy. In this respect, we show that northern firms are localised in regions with adequate endowment of infrastructure, with efficient public administration and with high R&D intensity and, as a result of these factors, perform better than firms operating in less well endowed regions.

Key words: Manufacturing Firms; Total Factor Productivity, Italian Regional Divide; Multilevel Models.
JEL code: L60, O14, R11

1. Introduction

Italy is an interesting case-study in the field of economic development because of high and persistent disparity between the South and the rest of the country. The level of per capita income in southern regions was 17.324 euro in 2009, a much lower value than that (29.399 euro) observed in the Centre-North. This is a substantial gap which is also persistent given that it has not varied significantly over the last 30 years (ISTAT, 2005; 2010)[1].

As a result of these wide disparities, it has been necessary to adopt policies aimed at overcoming the dualist nature of the Italian economy. For instance, through the *Intervento Straordinario per il Mezzogiorno* (Special Funding Plan for the Development on the South Area) southern regions received a large amount of resources from 1951 to 1984. The effectiveness of this extraordinary intervention is debatable: while it may have been effective in fostering convergence up to the mid-'70s, it has not modified the structural conditions needed to ensure growth in the long term (Carey et al, 1955; Iuzzolino *et al.* 2011). Similar questions have arisen with regards more recent EU structural interventions which have only slightly contributed to tackling the Mezzogiorno issue (Aiello and Pupo, 2011). Among the several factors which have contributed to bringing about the failure of these policies, it is worth noting that they were mostly oriented towards boosting capital deepening, and in so doing they disregarded the fact that TFP, rather than capital intensity is the main determinant of income differences (Klenow and Rodriguez-Clare, 1997; Hall and Jones, 1999; Easterly

[1] See Iuzzolino *et al.* 2011 for a survey.

and Levine, 2001; Parente e Prescott, 2004; Prescott, 1998). This seems particularly critical for Italy, as differences in regional TFP explain a great deal regarding differences in regional per-capita GDP (Aiello and Scoppa, 2000; Ascari and Di Cosmo, 2005; Byrne et al 2009; Piacentino 2008).

This paper provides further evidence for the debate on the dualistic nature of Italian economy. To this end, it uses data at firm level retrieved from the survey carried out by UniCredit-Capitalia (2008). The focus is on the determinants of the differences existing in TFP between firms located in Italian southern regions and those operating in the northern area of the country. A distinguishing feature of the work is based on the belief that the environment in which firms operate matters (Krugman 1991) and thus, from our perspective, TFP is meant to depend not only on firms' internal factors - like size, type of economic activity and internal R&D - but also on external variables beyond firms control. In other words, we are interested in distinguishing between the impact on TFP brought about by internal variables and the role played by certain territorial factors (e.g. availability of infrastructure, local administration, propensity to innovate) which the related literature suggests might influence firms' performance. The key question in the paper is whether location in different regions matter in terms of firms' performance.

In order to provide an answer to this question we proceed as follows. While TFP is estimated at firm level by employing the Levinshon and Petrin's approach (2003), the empirical setting we propose is consistent with the type of analysis carries out in this paper. Indeed, in order to explain firms' TFP, we combine data at firm and regional levels and use the multilevel approach. This model - giving proper attention to nesting - allows the extent to which space matters in determining firms' performance to be explored. In other words, multilevel regressions combine different levels of data aggregation and relate them in ways that render the simultaneous existence of distinct level-one (*firms*) and level-two (*regions*) equations explicit. This represents a great methodological advantage with respect to single-level models which are too limited to handle hierarchical structures of data. Indeed, with nested data the single-level methodologies lead to correlated errors of firms belonging to the same region, so violating the assumption of independence among errors. Furthermore, in one-single models statistical inference is based on the entire sample size and this yields a high risk of type I errors because standards errors in level-two variables are underestimated.

With regards the Italian economic divide, no paper, to the best of our knowledge, has explored the influence of location on firms' performance, By using multilevel analysis, we can explain differences in firms' TFP and provide a clear distinction between firm and region-specific effects. Previous works generally use regions as the unit of analysis (Quartaro, 2006; Marrocu and Paci, 2010; Ascari-Di Cosmo, 2004; Destefanis and Sena, 2005). However, finding a correlation at the regional level does not necessarily mean that it also holds when using data at individual firm level. In the literature this is known as the ecological fallacy.

The empirical results obtained throughout the paper support the hypothesis that local environment conditions exert an influence upon firms' TFP. Since firms are clustered within regions, we find that operating in a high R&D-oriented region or in an area with good infrastructure and/or with efficient public services affect private performance. This is an important policy issue in Italy, as southern firms suffer from being located in regions which are poorly endowed with respect to such a pro-growth environment.

The work is organized as follows. Section 2 presents the micro-data used in the paper and points out how relevant the regional differences are in terms of labour productivity and TFP. Section 3 illustrates the empirical strategy followed in the estimations. Sections 4 discusses the results and section 5 concludes.

2. The economic divide in Italy and the role of TFP: what firm level data highlight

The firm level data used throughout this paper come from the Xth UniCredit-Capitalia survey (2008) of Italian manufacturing firms.[2] Table 1 presents some descriptive statistics of the sample of firms used in the empirical analysis.[3] In particular, it presents the distribution of firms by area and, for 2006, the labour productivity of firms grouped in terms of size and economic activity.

The sample comprises 3019 firms which are highly concentrated in traditional sectors (49% in the entire sample, and 62% when just considering the South) and in highly specialised sectors. The incidence of high-tech firms is residual (only 5% in the whole of Italy and 2% in the South). From a regional perspective, two third of the sample is comprised of firms located in the North of Italy, 16% in the Centre and 10% in the South. The proportion of small firms is high (about 56%) and uniform by area. This picture is representative of the Italian manufacturing industry, which is characterised by a predominance of firms located in the North and belonging to traditional sectors. Again, the share of small-sized firms is very high in Italy, whatever the area and the economic activity (Bank of Italy, 2009, Onida 2004).

Table 1 further confirms the dualist nature of Italian economy: we find that labour productivity of Southern firms is lower than that recorded in northern Italy. More importantly, this gap holds whatever the subgroup of firms we consider, except for science based firms operating in the South (table 1)..

What clearly emerges from table 1 is a sharp economic divide between firms operating in the South and the rest of the country, something which has been long debated in the literature. While there are numerous, complex reasons underlying this phenomenon, here we simply refer to the strand of literature explaining how regional differences in labour productivity may be mainly attributed to differences in TFP (amongst others Aiello *et al.*, 2009; Brandolini and Cipollone, 2001; Daveri and Jonia-Lasinio, 2005; ISTAT, 2007; OECD, 2007; Van Ark et al., 2007). In this sense, our analysis again points out the role of TFP. Indeed, after retrieving TFP from the Levinshon and Petrin's estimator (see Appendix A for the LP procedure), we find that the correlation between firms' labour productivity and TFP from 1998 to 2006 is, on average, 0.86% and that there is an even higher value (0.96%) over the 2001-2006

[2] The survey covers a sample of firms with 11 to 500 employees and all firms with over 500 employees. The Xth Capitalia-UniCredit survey questionnaire refers to 2004-2006 and contains information on firm structure, ownership structure, work force composition and investment activity in physical capital and innovation, as well as the degree of internationalization. The balance data refer, instead, to 1998-2006.

[3] Although the original data refer to 5,100 firms, a sample of 3000 firms obtained after carrying out a data cleaning procedure is used in the empirical analysis. The firms which presented negative values of value added have been eliminated from the original archive. Moreover, in order to eliminate *outliers*, firms with a growth rate of value added and of employees below the first or above the ninety-ninth percentile of the distribution have also been eliminated. Finally, firms for which at least 7 years data regarding employee numbers was not available were also excluded.

period. At regional level, this correlation ranges from 0.97 in the South of Italy and 0,82 in the North-West.[4]

Table 1. Main characteristics of the sample

	North west	North east	Centre	South	Italy
N. observations	1338	918	469	294	3019
	44.3%	*30.4%*	*15.5%*	*9.7%*	100%
Labour productivity*	58824	54775	55376	51305	56338
Pavitt Sectors					
Supplier dominated	54203	50683	52753	47220	51976
	42.9	49.9	53.7	62.2	48.6
Scale intensive	64377	60887	62846	59862	62740
	19.8	*16.0*	*23.0*	21.4	19.3
Specialised suppliers	59935	58379	52542	51003	58168
	31.2	*30.7*	*19.6*	13.9	27.6
Science based	66798	54048	61812	80604	64159
	6.1	*3.4*	*3.6*	2.4	4.5
By class of employees					
Small (11- 50)	58328	53484	54899	49878	55611
	57.9	*52.1*	*56.1*	55.1	55.6
Medium (50-250)	58605	55178	53704	52023	56059
	33.0	*38.0*	*35.4*	36.1	35.2
Large (>250)	63124	60001	65972	58675	62087
	9.0	*9.9*	*8.5*	8.8	9.2

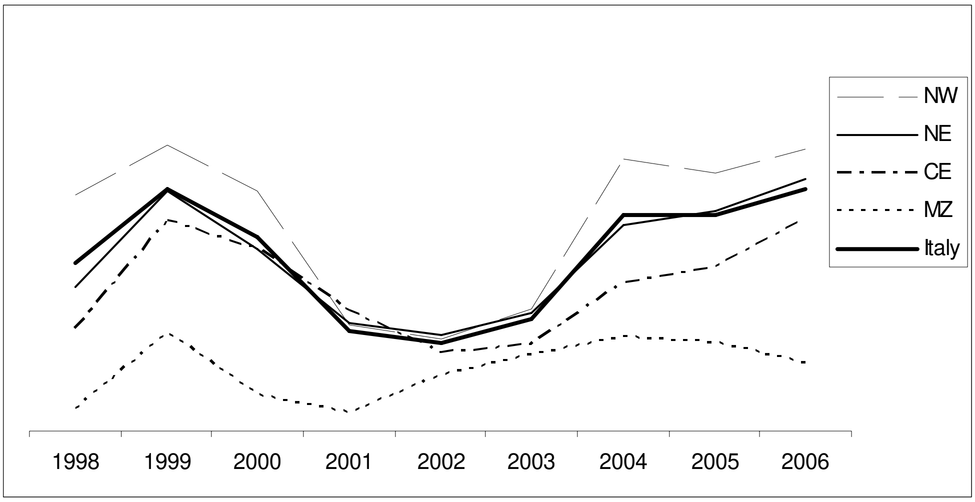
*In italic shares with respect to the total of the column.
All variables computed for 2006. Data in value deflated and expressed in euros.
The twenty Italian regions are grouped as follows. North-West: Piedmont, Lombardy, Aosta Valley and Liguria. North-East: Veneto, Emilia-Romagna, Friuli Venezia-Giulia,Trentino Alto-Adige. Centre: Tuscany, Marches, Lazio and Umbria. South: Campania, Apulia, Calabria, Basilicata, Abruzzo, Molise, Sicily and Sardinia.

Bearing in mind the above-mentioned evidence, in what follows we focus on the dynamics of TFP over the period under scrutiny and highlight the differences from one area to another. In particular, it can be seen how TFP in the South was lower than in other areas for the whole period, underlining the technological gap in Italy which has already been discussed in literature (Byrne *et al.*, 2009; Ladu, 2010; Ascari-Di Cosmo, 2005). Results also show how this disparity was not uniform over time: although wide at both the beginning and at the end of the period, in 2002-2003 regional TFP converged (fig. 1) This, though, not so much due to the performance of Southern firms, but it mainly the result of what happened in the rest of Italy. As figure 1 shows, there was a decline in the TFP gap in Italy in 1999-2001. This though was mostly due to the dynamics of the North of the country, while an improvement in southern firms efficiency only took place subsequently. Again, it is important to emphasise that this recovery in the South was short-lived and much more limited than that registered elsewhere (figure 1).

While previous results bear out the dualistic nature of Italian economy, they leave the question about the reasons underlying the regional gap in TFP open. The next paragraph looks at this issue.

[4] Labour productivity is calculated as a weighted average of firms' productivity, using as a weight the firm's value added with respect to the group of reference (the whole sample or the value added of the area in the case of averages relative to the territory).

Figure 1. Average TFP by area from 1998 to 2006



Source: elaborations on data from UniCredit –Capitalia (2008)

3 Empirical setting

3.1 Methodology: the multilevel analysis

The understanding of how being located in different regions affects firms' performance is a typical issue with a hierarchical structure of data (Goldstein, 2003), in the sense that the units (firms) refer to different levels of aggregation (regions). The multilevel model addresses this issue because, combining different levels of data aggregation, allows a dependent variable to be related to variables at more than one level (Luke, 2004). In this model, variables at different levels are not simply add-ons to the same single-level equation, but are linked together in ways that make explicit the simultaneous existence of distinct level-one and level-two equations explicit. In such a way, a multilevel model permits to control for spatial dependence and correct the measurement of standard errors, so reducing the risk of type I error.[5] Moreover, in the multilevel approach level-two factors are used not just as independent variables to explain variability in a level-one dependent variable, but also because they explain variability in random intercept and random slopes (Bickel, 2007).

In what follows we briefly present the multilevel regression model and, in order to limit complexity, we consider a two-level model where firms are the first-level units and regions those at the second-level.

The dependent variable ω refers to firms and depend on a set *X* of variables measured at firm level and on a set *Z* of variables defined at regional level[6]. The variable ω may be predicted by just considering *X* as explanatory variables:

$$\omega_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij} \quad [1]$$

[5] In a single–level regression the number of level-one observations, not the number of groups, is used to test level-two coefficients.

[6] For the sake of the exposition, we assume that *X* and *Z* include only a single variable.

where β_{0j} is the intercept, β_{1j} is the slope coefficient and e_{ij} is the random error term with zero mean and σ_e^2 as variance, *j* is for regions (*j=1…r*) and *i* for firms (*i=1…N$_j$*). In eq. [1] the regression parameters β_j vary across level-2 units. This may be modelled as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \qquad [2]$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j} \qquad [3]$$

In so doing, β_{0j} and β_{1j} differ region-by-region and depend on *Zj*. u_{0j} e u_{1j} are random error terms defined at regional level with zero mean and assumed to be independent from e_{ij}. γ denote the fixed level-two parameters.

Combining the micro (eq. 1) and the macro models (eq. 2 and 3) produces a two level mixed model:

$$\omega_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + (u_{1j} X_{ij} + u_{0j} + e_{ij}) \qquad [4]$$

The deterministic part of the model, $\gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij}$, contains all the fixed coefficients, while the stochastic component is in brackets. The added complexity of the error term stems from the fact that it captures residual variance, in the same way as OLS regression does, as well as group-to-group variability in the random intercept relative to the overall intercept, and group-to-group variability in the random slope relative to the overall slope. It is clear that the error term displayed in eq. [4] is not independent. Indeed, as data are nested at different levels of analysis, firms operating in the same region tend to have correlated residuals, violating the assumption of independence.

Eq. [4] also allows identification of the errors due to differences across firms or regions. To this end, it is necessary to use to an "empty" model, i.e. a model without any explanatory variables:

$$\omega_{ij} = \gamma_{00} + u_{0j} + e_{ij} \qquad [5]$$

which allows decomposition of the variance of ω into two independent components, which are the variance of e_{ij} (σ_e^2), the so-called within-group variance, and the variance of u_{0j} (σ_{u0}^2), also known as between-group variance. Hence we can estimate the proportion of the total variance "explained" by the grouping structure, i.e. the intra-class correlation ICC:

$$ICC = \frac{\sigma_{\mu 0}^2}{\sigma_{\mu 0}^2 + \sigma_e^2} \qquad [6]$$

Furthermore, eq. [4] is general and contains different reduced specifications. For example, the so-called "random intercept model" with just the intercept as a function of level-two predictors, without considering the cross-level fixed effects, can be delineated. Since there are 20 Italian administrative regions, there is a relatively limited number of groups. Such a constrain limits the number of parameters to be estimated and, for this reason, we proceed by considering a two-level random intercept model with firms (lowest level) and regions (highest level)[7]. The specification used in this work is given by:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \quad [2']$$

$$\beta_{1j} = \gamma_{10} \quad [3']$$

$$\omega_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + u_{0j} + e_{ij} \quad [4']$$

The random component u_{0j} captures variability in the intercept across regions, while the fixed component γ_{00} is a weighted average for the intercept across all regions. Finally, equation [4'] allows us to account for the variability in the random component by introducing one or more level-two contextual variables[8].

3.2 Econometric specification and data

In line with eq. [4'], the model used in the empirical analysis is specified as follows

$$\omega_{ij} = \beta_0 + \sum_{s=1}^{k} \beta_s X_{sij} + \sum_{z=1}^{v} \eta_z Z_{zj} + \lambda South + \sum_{p=1}^{3} \lambda_p S_{pi} + \sum_{p=1}^{2} \lambda_p D_{pi} + \varepsilon_{ij} \quad [7]$$

where ω is the TFP of the i-th firm (in logarithm) operating in region j, *X* is a vector of firm-level variables which are meant to be important drivers of TFP and Z are variables at regional level. TFP determinants defined at firm level include the 2004 R&D intensity (R&D investments/sales), the white collar workers as share of total workers in 2006 and export intensity in 2006 (exports/sales)[9]. One of the basic empirical facts related to productivity is a strong positive association between productivity and exporting activity and, therefore, we include the share of exports in total sales among firms characteristics.[10] Similarly, it is widely argued that a firm's performance improves

[7] As we have already pointed out, in multilevel model when level-one coefficients are permitted to vary across groups, the number of groups, not the number of level-one observations, is used in tests of significance for level-one slopes. Consequently, cross-level interaction terms are likely to have unstable coefficients and uncertain inferential properties, unless there is a comparatively large number of cases at both level one and two (Bickel, 2007).

[8] The possibility to employ contextual factors *(Z_j)* and, in the general specification (equ. [4]), cross-level interaction terms *($Z_j X_{jj}$)* to explain variability in random components is the main difference between multilevel model and random coefficient regression.

[9] It has been checked whether firm-level predictors are correlated and we find that the correlation coefficients are very low. This ensures that these variables capture different firms characteristics and thus that econometric results do not suffer from a serious problem of multicollinearity of firm predictors (the correlation matrix is available on request).

[10] In the literature two hypotheses about the positive correlation between export activity and productivity are investigated. The first hypothesis is that the most productive firms self select into foreign markets because they can overcome sunk costs associated with foreign sales (Melitz, 2003). The second hypothesis raises the possibility of "learning by exporting". Firms participating in international markets acquire knowledge and technology with positive feedback effects on firms' knowledge and technology accumulation. Furthermore, firms which are active in world markets are exposed to more intensive competition than firms which only sell their products domestically. In summarizing the results of the empirical research emerges that the more productive firms self-select into export market (ISGEP, 2008).

as a result of its innovative behaviour and in the presence of skilled workers (see, i.e., Krueger and Lindahl (2001) and Sveikauskas 2007).

As far as the regional level is concerned, we select three variables to be included in the analysis: the R&D intensity of the private sector, an index of infrastructure and the efficiency of public administration. This is done in line with the large amount of literature dealing with the reasons for the economic divide in Italy (an exhaustive survey can be found in Iuzzolino *et al.* 2011). In this paper, R&D intensity is measured as a share of private R&D expenditure in regional GDP in 2004 (data are from ISTAT). The index of total infrastructure (Italy = 100) is taken from the CNEL (National Council of the Economy and Labour) database Cnelstats[11] and summarises the availability of different kinds of productive infrastructure such as roads, railways, telecommunications, ports, airports and energy-environmental plants and networks as well as the presence of bank and miscellaneous services networks and cultural, education and health facilities. The index refers to the year 2004. The variable public efficiency variable refers to year 1997 and is measured by the difference between an evaluation of physical quantities of public goods and the corresponding expenditure made by the government, after controlling for regional differences in cost of public construction (Golden and Picci, 2004). All regional indicators are merged with firms' data on the basis of location of companies' headquarter. Table 2 displays the values of the regional variables used in the paper.

In estimating the multilevel equation we also control for location, sector and size effects by adding a set of dummy variables. South is a binary variable which takes the value one if the firm is located in the South of Italy and zero otherwise. The variable South variable is supposed to capture the non-observable differences between the Centre-North and in South of Italy. In order to control for sectoral heterogeneity in the production process, we include three sector dummies (*S*) according to the Pavitt taxonomy [S_2 is unity for firms belonging to scale intensive sectors, S_3 is unity for firms operating in specialised suppliers, while S_4 is unity when firms operate in science-based sectors. Firms in supplied-dominated sectors form the control group]. Finally, regressions also consider two dummy variables to control for size effect (D_M refers to medium-sized firms and D_L is for large-sized firms, where the control group is comprised of small firms).

TFP is estimated through the Levinshon and Petrin's approach and is expressed as the average of the three-year period 2004-2006. Although TFP at firm level is available for a longer period (see fig. 1), we restrict estimations to a cross-section analysis averaging TFP data from 2004 to 2006. This is done because of data constraints.[12] [13]

---

[11] The Cnelstats database built in cooperation with the Guglielmo Tagliacarne Institute provides both information and statistical indicators on economic trends, on productive network and social situation for Italy and the EU countries (http://www.cnel.it/cnelstats/index.asp).

[12] We average TFP over the three-year period 2004-2006 in order to control for the influence of shocks and measurement errors in specific year and to limit the extent of missing data.

[13] Equation [7] probably suffers from omitted variable problems since unit heterogeneity is not considered. One way to allow for unobserved heterogeneity is the fixed effects model. However, panel data analysis cannot be performed, due to the lack of time series in variables such as white collar share and export status. Moreover, endogeneity has been addressed by using lagged variables when available.

Table 2 Infrastructure, R&D Intensity and efficiency of public administration by region.

Regions	Infrastructure in 2004 (Italy=100)	Private R&D Intensity in 2004 (Italy =100)	Efficiency of Public Administration in 1997 (Italy =100)
Abruzzo	77.8	86.9	95.6
Basilicata	38.6	37.0	53.3
Calabria	74	3.7	40.8
Campania	95.7	77.7	36.2
Emilia-Romagna	109.8	125.7	161.1
Friuli-Venezia Giulia	123.9	98.0	107.7
Lazio	146.2	83.2	81.7
Liguria	191.2	120.2	66.8
Lombardia	123.9	153.5	116.1
Marche	88.6	49.9	131.2
Molise	50.6	11.1	58.2
Piemonte	88.3	244.1	163.8
Puglia	79	29.6	72.2
Sardinia	55.5	5.5	83.8
Sicily	84.2	42.5	60.7
Tuscany	111.4	64.7	161.3
Trentino-Alto Adige	60.2	46.2	123.5
Umbria	86.7	29.6	178.3
Valle d'Aosta	44.4	48.1	85.5
Veneto	117.3	51.8	122.0

Source: National Council of the Economy and Labour for infrastructure, National Institute of Statistics for R&D intensity and Golden and Picci (2004) for efficiency of public administration.

4 Econometric results

Results are displayed in tables 2 and 3. While table 2 refers to the estimates obtained when equation [7] is estimated through OLS, table 3 presents the evidence provided by the multilevel approach.

OLS estimates are only used for reference and to verify the bias relating to the use of data at different levels of aggregation within a single-equation model. However, because OLS regressions are performed using micro and regional data, we control for the potential downward bias in the estimated errors by clustering firms at regional level.[14] In brief, two key results emerge from table 3. On one hand we find that location matters in determining firms' TFP. This can be seen from Model 2 which differs from Model 1 because of the South dummy, whose coefficient is negative and significant. This implies that, *ceteris paribus*, the level of TFP in firms located in southern regions is lower than that in firms located in the North of Italy. To some extent, similar evidence comes from Model 3, where the parameters associated with regional variables (R&S intensity, infrastructure index and the efficiency of public administration) are all

[14] Clustering data at regional level relaxes the assumption of independence and, therefore, increases the error term to accommodate the lack of independence of firms within regions. However, while clustered OLS leaves both the noise associated with difference between firms and noise associated with differences between regions in the error term, the multilevel model allows these two error component (see equ. [4] to be separated.

positive and highly significant. By referring to these results it is possible to argue that high regional R&D intensity, good infrastructure and high efficiency of public administration do help firms to improve performance. Again, in the case of Model 4 the result that the South dummy remains negative but not significant may be due to the fact that the variables which are observable at regional level capture all the differences in TFP and at the same time nothing else unobservable is left in the dummy South. On the other hand, by comparing data in tables 3 and 4 we obtain evidence that using a single-equation model when data are available in a hierarchical structure yields deflated standard errors. In this sense, interpretation of OLS results is bounded by the actual statistical significance of the estimated coefficients.

The results of the multilevel models are presented in table 4, where each column of data refers to different specifications of Eq. [7] according to the set of explanatory variables included in the model. Model 1 is the empty model, i.e. a model without regressors (eq. [5]), while Model [2] only includes level-1 predictors. In Models [3] and [4] different sets of level-2 regressors have been added.[15]

A key initial result comes from the likelihood-ratio test, which compares the empty model [eq. 5] with the standard linear regression. This test is highly significant, supports the use of multilevel methodology[16] and indicates that the intercept should be considered as a region-by-region variant coefficient. Moreover, the ICC value (*cfr* eq. [6]) indicates that 4.3% of firms' TFP can be assigned to the mere spatial location of firms (Model 1, table 4) while internal firm characteristics explain 95% of firms' TFP.[17]

A further interesting aspect of the approach refers to the possibility of using the variance at the relevant different levels of analysis to calculate the coefficient of determination and, in such a way, to obtain a proportional reduction in the estimated total residual variance. This is done by comparing the "empty model" with a given extended specification of the model. The coefficient of determination can be also calculated for each variance components (Rabe-Hesketh and Skrondal, 2008)[18].

For instance, we see in table 4 that variables at firms level as a whole are able to explain 28% of TFP firm variance (table 4, model 2). When including the region-level

---

[15] The multilevel analysis has been implemented in Stata using the "xtmixed" subroutine. All models were estimated using restricted maximum likelihood (REML) over maximum likelihood (ML) since the latter is more sensitive to loss of degrees of freedom when dealing with a small number of groups (Bickel, 2007).

[16] The null hypothesis is that $u_{0j} = 0$ or that there is no random intercept in the model. If the null hypothesis is true, an ordinary regression can be used instead of a variance-components model.

[17] Using data on firms in manufacturing and business services in the Netherlands, Raspe and Van Oort (2011) find that 2.3% of the firm productivity can be related to its location and that more than 97% to its internal characteristics.

[18] The coefficient of determination for two-level models is given by:

$$R^2 = \frac{(\sigma^2_{\mu 0N} + \sigma^2_{eN}) - (\sigma^2_{\mu 0M} + \sigma^2_{eM})}{\sigma^2_{\mu 0N} + \sigma^2_{eN}}$$

where N it states for the null model and M for the model of interest.

The proportional reduction in each of the variance components can be calculated separately. The proportion of the level-2 variance explained by the covariates is:

$$R^2_2 = \frac{(\sigma^2_{\mu 0N} - \sigma^2_{\mu 0M})}{\sigma^2_{\mu 0N}}$$

and the proportion of the level-1 variance explained is:

$$R^2_1 = \frac{(\sigma^2_{eN} - \sigma^2_{eM})}{\sigma^2_{eN}}.$$

predictors we find that the variance at regional intercepts decreases by 88%. In our case, this result is due to the fact that a large proportion of region-by-region variability in the intercepts has been accounted for the regional variables covered by the analysis (private R&D intensity, infrastructures and public administration efficiency). This evidence ensures that the selected regional factors of TFP capture a great deal of intercept variability, which we attribute to un-observed TFP heterogeneity when considering the empty model.

Moving on to discuss the results of estimated coefficients, table 3 shows that at firm level, the parameter associated with internal R&D has the expected positive sign and is highly significant. Firms investing more in R&D obtain higher TFP levels than firms with weak innovative activities. Again, an important role is played by the human capital employed by firms. The result is that TFP increases with human capital. These findings are in line with the literature showing that R&D and human capital induce higher firm TFP because they directly affect the possibility to introduce and use more productive processes and hence translate innovation efforts into profitable opportunities (Griliches 2000; Parisi et al 2006). Furthermore, and consistent with existing literature, we find that TFP tend to increase with exports. Many studies explain the positive relationship between export activity and productivity by self-selection of more efficient Italian firms into the export markets [see, for instance ISGEP, (2008), Benfratello and Razzolini (2008) Serti and Tomasi (2008) among many others], while few studies also find support for the "learning by exporting" hypothesis (ISGEP, 2008; Serti and Tomasi, 2008, and, then just for exporters with a high share of export intensity, Castellani, 2002).

In addition, the positive coefficients associated with D_M and D_L dummies highlight the role of size for TFP. Medium-sized firms perform better than small firms, but less well than large enterprises. In short, in the case of Italian manufacturing firms, TFP increases with firm size, indicating that economies of scale are at work. Another influential factor is the type of economic activity. It is widely accepted that TFP differs across sectors and it is found that firms in high-tech sectors perform better than others, followed by firms operating in scale intensive and specialised sectors. The lowest value of TFP is estimated for firms belonging to traditional sectors (our group of control). This result indicates that sectoral characteristics in producing innovative products allow high-tech firms to perform better than those operating in other sectors.

When considering the first level of the analysis results indicate how firms internal factors are relevant in determining the level of TPP. However, there is also interest in the role of variables defined at regional level. A first finding regards the role of infrastructure. We find that TFP at firm level is positively affected by the endowment of regional infrastructure, in the sense that firms' localised in regions with an adequate provision of infrastructure benefit more than firms operating in under-endowed regions. Due to the sharp differences in regional endowment of infrastructure (see table 2), this result indicates that, other things being equal, TFP of southern firms will be lower (fig. 1) because they operate in areas suffering from a lack of public capital. This is in line with the conclusions drawn, for instance, by Aiello *et al.* (2010), De Stefanis and Sena (2005) and Marrucu and Paci (2010) according to which investment in core infrastructure has a greater impact in the South than in the North of Italy also because of the low level of initial infrastructure stock in lagging regions. With regards regional private R&D activity, we find a positive impact on firms' TFP (the coefficient ranges from 0.0775 to 0.0851). This corresponds with the literature (see. e.g, Camagni 1991; Ciccone and Hall, 1993). It is an indication for the spillovers effect as a product of innovations, in the sense that being located in regions with high innovation-creating potential makes individual firms perform better. In this respect, it appears clear that the TFP divide between northern and southern firms that we have recorded is also due to the differences in regional innovativeness: the data used in this paper regarding 2004 2004 R&D intensity indicate that the level of innovative efforts made by the private sector in southern regions is very low and far less than that recorded in the rest of the country (table 2). Therefore, stimulating and increasing R&D investments in the South of Italy has to be a priority in policy agenda because so that might help to build a R&D environment from which firms may acquire innovative opportunities tht can be translated into internal efficiency. Finally, we find a positive and significant coefficient associated with the efficiency of public administration: efficient provision of public services is an efficient way is an important factor for firms' productivity. In Italy, public administration is most successful in providing services in the northern regions (Bank of Italy, 2009; Tabellini 2010). From our perspective this fact contributes to explaining why TFP of firms operating in that area of the country is higher than the TFP levels observed in the South of Italy. Firms operating in regions with efficient public institutions benefit from a reduction in the transaction costs they face when introducing more productive activities and to creating an environment which is conducive to growth.

Table 3
Explaining TFP of Italian manufacturing firms in 2004-2006: OLS Results

Explanatory Variables	Model 1	Model 2	Model 3	Model 4
Firm level covariates				
R&D Investments/Sales	0.00002***	0.00002***	0.00002***	0.00002***
	(7.85)	(7.52)	(9.54)	(9.50)
White Collar Share	0.26421***	0.25228***	0.21599***	0.21597***
	(6.72)	(6.89)	(6.789)	(6.80)
Exports/Sales	0.00087***	0.00066***	0.00059**	0.00059**
	(3.87)	(3.24)	(2.67)	(2.71)
Medium firms (DM)	0.28628***	0.28794***	0.29112***	0.29107***
	(32.25)	(30.62)	(23.27)	(23.62)
Large firms (DL)	0.70147***	0.70332***	0.69024***	0.69020***
	(23.95)	(25.09)	(24.50)	(24.47)
Scale intensive (S2)			0.16957***	0.16959***
			(8.64)	(8.64)
Specialised suppliers S3)			0.13288***	0.13288***
			(7.01)	(7.02)
Science based (S4)			0.23032***	0.23034***
			(4.12)	(4.12)
Regional level covariates				
Private R&D over Regional GDP			0.07898***	0.07882***
			(5.18)	(4.85)
Index of Infrastructure			0.00181***	0.00177**
			(4.86)	(2.43)
Efficiency of Public Administration			0.00102***	0.00099*
			(3.33)	(1.89)
South		-0.17450***		-0.00421
		(-5.03)		(-0.07)
Constant	6.35030***	6.37730***	5.92164***	5.92993***
	(235.46)	(289.67)	(89.12)	(41.12)
R-squared	0,24	0,26	0,29	0,29
Observations	2941	2941	2941	2941

Dependent variable: log of TFP (average values for 2004-2006 period). In parentheses, t-values based on standard errors clustered at firm level. Level of significance: *** 1%, ** 5%, * 10%.

Table 4 Explaining TFP of Italian manufacturing firms in 2004-06: multilevel regressions

Explanatory Variables	Model 1	Model 2	Model 3	Model 4
Fixed effects				
Level 1: Firms				
R&D Investments/Sales		0.00002***	0.00002***	0.00002***
		(4.79)	(4.70)	(4.73)
White Collar Share		0.216***	0.215***	0.214***
		(7.71)	(7.66)	(7.64)
Exports/Sales		0.00064***	0.00060**	0.00060**
		(2.54)	(2.40)	(2.38)
Medium firms (DM)		0.291***	0.291***	0.291***
		(18.28)	(18.29)	(18.26)
Large firms (DL)		0.690***	0.690***	0.690***
		(24.45)	(24.41)	(24.40)
Scale intensive (S2)		0.164***	0.166***	0.165***
		(8.32)	(8.42)	(8.35)
Specialised suppliers S3)		0.130***	0.130***	0.129***
		(7.33)	(7.32)	(7.26)
Science based (S4)		0.227***	0.227***	0.225***
		(6.18)	(6.15)	(6.12)
Level 2: Regions				
Private R&D over Regional GDP			0.0775**	0.0797
			(1.92)	(1.57)
Index of Infrastructure			0.00150***	0.000783
			(2.98)	(0.97)
Efficiency of Public Administration			0.000836**	0.000111
			(2.32)	(0.16)
South				-0.0759
				(-1.07)
Constant	6.589***	6.257***	5.975***	6.151***
	(246.21)	(232.47)	(88.14)	(39.06)
Random-Effects				
Variance				
Region	0,010	0,006	0,001	0,002
Firms	0.213	0.154	0.155	0.155
Interclass correlation	4,3%			
R		0,28	0,30	0,29
R^2 *level 2*		0,33	0,88	0,76
R^2 *level 1*		0,28	0,27	0,27
LR test	61.00***			
Number of observations	3006	2941	2941	2941

Dependent variable: log of TFP (average values for 2004-2006 period). In parentheses, z-values. Level of significance: *** 1%, ** 5%, * 10%.

5. Conclusions

Many recent works investigated the economic divide in Italy focussing on the role played by TFP. This literature comes to the conclusion TFP is a key variable in explaining the regional differences in Italian growth,. Following this line of reasoning, we provide empirical evidence of the likely causes of different levels of efficiency in regional economic systems. With respect to the most literature, this article makes two main contributes. Starting from an estimation of TFP based on firm data, the work provides further evidence regarding the TFP divide in Italy, with particual interest in understanding whether location matters in determining firms' TFP. This is done by quantifying the role of local environment conditions and in this respect the analysis allows the impact of firms and regional factors on productivity to be disentangled. Estimations have been used to quantify the relative importance of different territorial factors in the two main areas of Italy (Mezzogiorno and Centre-North).

An initial result of the paper confirms that Italian firms' TFP differs according to their localisation across regions. By confirming that firms located in the South of Italy are less efficient than those operating in the rest of the country, the analysis supports the hypothesis that southern regions are technologically lagging behind northern regions. Throughout the paper we find that location matters in determining the level of firms' TFP. This emerges, for instance, from the basic multilevel model which points out the incidence of variant regional intercepts in explaining the variance of firms' TFP. Similar conclusions may be drawn from multilevel regressions which include the regional determinants of TFP, i.e., R&D, the efficiency of public administration and infrastructure. These variables explain a large proportion (more than 88%) of the average level of regional TFP. This outcome supports our choice of environmental factors of firms' TFP. Therefore, it appears clear that, given the high share of TFP variability explained by these environmental variables, any excluded regional factor of TFP can only be of marginal importance.

Interesting insights come from the signs of regional variables used in estimations. The relevant regional factors - R&D, the efficiency of public administration and infrastructure – always register the expected positive sign. The parameters associated with regional determinants of TFP are highly significant in OLS regressions, but this is not surprising because in one-single equation standard errors are deflated by using the entire sample of firms. More importantly, the impact of regional factors remains highly significant in multilevel regressions and this is not a foregone conclusion. Indeed, this approach uses the numbers of regions instead of the entire sample of firms and therefore statistical significance may be lost. The fact that they are still significant implies that they are important causes of TFP differences across Italian regions. In a nutshell, we find that operating in R&D oriented regions, which guarantee good quality public services and high endowment of infrastructure ensures that firms will achieve a high level of economic performance. Data also reveal that southern regions are poorly endowed in terms of infrastructure, dedicate very little effort to innovative activities and suffer as a result of the high level of local administration inefficiency. In terms of policy implications these findings suggest that, on one hand, there is still a need for state intervention aimed at reducing the obstacles to growth in the South of Italy. It is sure that firms' performance in this area of the country is negatively conditioned by marked deficiencies of public goods. On the other, while the demand for greater state intervention is highly justified, it should be underlined that, since the 1950s enormous financial efforts have been made in national and European policies which have been aimed specifically at overcoming the Italy's North-South dualism. However, the fact that this divide persists indicates that these policies have been uneffective and hence that more efforts must be made to understand why some development policies work well everywhere except in Italy.

Appendix A– A measure of TFP

TFP at firm level is estimated by using Levinshon and Petrin's approach (2003). Productivity was estimated using the following log-linear specification of a production function:

$$y_{it} = \beta_0 + \beta_K^{MAT} k_{it}^{MAT} + \beta_l l_{it} + u_{it} \qquad \text{(A1)}$$

with $i = 1, \ldots\ldots N$ firms, t = 1998, ……2006 and where *y* represents the value added, *l* the number of employees, k^{MAT} the stock of physical capital, β_0 measures the average efficiency and u_{it} represents the deviation of firm *i* from this average at time *t*. The error term can be decomposed into two parts:

$$u_{it} = \omega_{it} + \eta_{it} \qquad \text{(A2)}$$

where the term ω_{it} represents the productivity of firm *i* at time *t* and η_{it} is a stochastic term which includes not only the measurement error, but also the shocks which are unobservable to firms, and, therefore, do not correlate with inputs.

Productivity ω_{it} is known to the firm which, therefore, in the case of positive shocks to productivity, can decide to increase production by raising the level of inputs. This determines a problem of simultaneity which Levinshon and Petrin (2003) resolved by identifying in the demand for intermediate goods as a proxy for the variations in TFP known to firms.

Equation (A1) was estimated by utilizing the tangible fixed assets as a proxy for the stock of physical capital and the demand for intermediate goods was measured by the operating costs. The value added has been deflated by using the ISTAT production price index available for each ATECO sector. As regards the tangible fixed assets, data have been deflated by using the average production price indices of the following sectors: machines and mechanical appliances, electrical machines and electrical equipment, electronics and optics and means of transport. For the operating costs, we adopt the intermediate consumption deflator calculated by using data from ISTAT.

REFERENCES

Aiello F., Scoppa V. (2000), Uneven regional development in Italy: explaining differences in productivity levels, Giornale degli Economisti 60, 270–298.

Aiello F., Iona A., Leone L. (2010), Regional Infrastructure and Firm Investment. Theory and Empirical Evidence for Italy, Empirical Economics, DOI 10.1007/s00181-010-0445-9.

Aiello F., Pupo V. (2011), Structural Funds and the Economic Divide in Italy", Journal of Policy Modeling (2011), doi:10.1016/j.jpolmod.2011.10.006.

Aiello F., Pupo V., Ricotta F. (2009), Sulla dinamica della Produttività Totale dei Fattori in Italia. Un'analisi settoriale", l'industria – Rivista di economia e politica industriale, n. 3, pp. 405-426.

Ascari G., Di Cosmo V. (2005), Determinants of total factor productivity in the Italian regions, Scienze Regionali – Italian Journal of Regional Science 4, 27–49.

Bank of Italy (2009), Rapporto sulle tendenze nel sistema produttivo italiano, Questioni di Economia e Finanza, n. 45.

Benfratello L., Razzolini T. (2008), Firms' productivity and internationalisation choices: evidence for a large sample of manufacturing firms in Multinationals and local competitiveness, Lucia Piscitello and Grazia Santangelo (Eds.), Franco Angeli

Bickel R. (2007), Multilevel Analysis for Applied Research, It's Just Regression!, The Guilford Press, New York.

Brandolini A., Cipollone P. (2001), Multifactor productivity and Labour Quality in Italy, 1981-2000, in Temi di Discussione, n. 422, Bank of Italy.

Byrne J. P., Fazio G., Piacentino D. (2009), Total Factor Productivity convergence among Italian Regions: Some evidence from panel unit root tests, Regional Studies, 43:1, 63-76.

Camagni R. (1991), Innovation Networks: Spatial Pespectives (eds), Belhaven Press, London.

Carey J.P.C , Galbraith C. A. (1955), The South of Italy and the Cassa per il Mezzogiorno, *Political Research Quarterly*, December 1955 8: 569-588.

Castellani, D. (2002), Export behaviour and productivity growth: evidence from Italian manufacturing firms, Weltwirtschaftliches Archiv, vol. 138, pp. 605–28.

Ciccone A., Hall R.E. (1993), Productivity and the Density of Economic Activity, American Economic Review, 86: 54-70.

Cingano F., Schivardi F. (2004), Identifying the sources of local productivity growth, Journal of the European Economic Association, 2(4): 720-742.

Daveri F., Jona Lasinio C. (2005), Italy's decline: getting the facts right, in "Giornale degli Economisti e Annali di Economia", vol. 64, n. 4, pp. 365-410.

Destefanis S., Sena V. (2005), Public capital and total factor productivity: New evidence from the Italian regions, 1970-98, Regional Studies, 39, 5, pp. 603-617.

Easterly W., Levine R. (2001), It's not factor accumulation: stylized facts and growth models, World Bank Economic Review, 15, 177–219.

Eckaus R. (1961), The North-South differential in Italian economic development. Journal of Economic History 20: 285–317.

Golden M.A, Picci L. (2004), Proposal for a New Measure of Corruption, Illustrated Italian Data, mimeo, UCLA, downloadable at http://golden.polisci.ucla.edu.

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

Griliches, Z. (2000), R&D, Education and Productivity A Retrospective. Harvard University Press, Cambridge, Massachusetts.

Hall R., Jones C. (1999), Why do some countries produce so much more output per workers than others?, Quarterly Journal of Economics 114, 83–116.
Hox, J.J. (1995), Applied Multilevel Analysis, TT-Publikaties, Amsterdam
ISGEP, International Study Group on Exports and Productivity (2008), Understanding Cross-Country Differences in Exporter Premia: Comparable Evidence for 14 Countries, Review of World Economics, Volume 144, Issue: 4, December 2008, pp. 596 – 635.
ISTAT (2005), Conti Economici Regionali, 1980-2004, Rome.
ISTAT (2007), Misure di produttività, in Statistiche in breve, 5 ottobre, Roma.
ISTAT (2010), Conti Economici Regionali, 1995-2009, Rome.
Iuzzolino G., Pellegrini G. Viesti G. (2011), Convergence among Italian Regions, 1861-2011, Bank of Italy, Economic History Working Papers, n. 22.
Klenow P. J., Rodriguez-Clare A. (1997), The neoclassical revival in growth economics: has it gone too far?, in BERNANKE B. and ROTEMBERG J. (Eds) NBER Macroeconomics Annual, pp. 73–103. MIT Press, Cambridge, MA.
Krueger A.B., Lindahl M. (2001), Education for Growth; Why and for Whom? Journal of Economic Literature, 39, 4, pp. 1101-1136.
Krugman P. (1991), Geography and Trade, Cambridge University Press.
Ladu M. G. (2010), Total factor productivity estimates: some evidence from European regions, WIFO Working Papers, 380.
Luke A. (2004), Multilevel Modelling. London: Sage Publications
Marrocu E., Paci R. (2010), The effects of public capital on the productivity of Italian regions, Applied Economics, 42, 989-1002.
Melitz M. J. (2003), The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. Econometrica 71 (6): 1695–1725.
OECD (2007), Economic Survey: Italy, OECD, Parigi.
Onida F. (2004), Se il piccolo non cresce. Piccole e medie imprese in affanno, Il Mulino, Bologna.
Parente, S., Prescott E. (2004), A unified theory of the evolution of international income levels, Federal Reserve Bank of Minneapolis Staff Reports, 333.
Parisi M. L., Schiantarelli F., Sembenelli A. (2006), Productivity, Innovation and R&D: Micro Evidence for Italy, European Economic Review, 50(8), pp. 2037-2061.
Piacentino D. (2008), Productivity, Infrastructures and Convergence: Panel Data Evidence on Italian Regions, Scienze-Regionali/Italian-Journal-of-Regional-Science, 7(2): 5-26.
Prescott E. (1998), Needed: a theory of total factor productivity, International Economic Review, 39, 525-551.
Quatraro F. (2006), Technological change and productivity growth in Italian regions, 1982-2001, Région ed Développement, n.24.
Rabe-Hesketh S., Skrondal A. (2008), Multilevel and Longitudinal Modeling Using Stata. Stata Press, 2nd edition.
Raspe O., Van Oort F. (2011), Firm heterogeneity, productivity and spatially bounded knowledge externalities, Socio-Economic Planning Sciences, 45, 38-47;
Serti F., Tomasi C. (2008), Self-Selection and Post-Entry Effects of Exports: Evidence from Italian Manufacturing Firms, Review of World Economics, Volume 144, Issue: 4, December, pp. 660 – 694.
Sveikauskas L. (2007), R&D and Productivity Growth: a Review of the Literature, US Bureau of Labour Statistics, BLS Working Papers n. 208.
Tabellini G. (2010), Culture and Institutions: Economic Development in the Regions of Europe, Journal of the European Economic Association, 8, 4, 677-716.

UniCredit-Capitalia (2008), Decima indagine sulle imprese manifatturiere italiane. Rapporto Corporate, n.1.

Van Ark B., O'Mahony M., Ypma G. (2007), Productivity in the European Union: A comparative Industry Approach, in "The EU KLEMS Productivity Report", n. 1.