Determination of Total Factor Productivity in Italian Regions

Guido Ascari
(Università di Pavia)

Valeria Di Cosmo
(Università di Pavia)

170 (12-04)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Dicembre 2004

Determinants of Total Factor Productivity in the Italian Regions[*]

Guido Ascari[†]
University of Pavia

Valeria Di Cosmo
University of Pavia

October 2004

Abstract

This paper investigates the determinants of TFP for Italian regions. We find strong evidence in favour of the factors commonly suggested by the theoretical literature. In particular, R&D expenditures and the number of researchers are positively related to regional TFP. Moreover, human capital is an important determinant of TFP. Finally, we find a strong difference between Northern and Southern regions, particularly regarding the effect of research activity and social capital. Our results are robust across different estimation methods.

JEL Classification: O47, C23, R11

Keywords: Total factor productivity, Italian regions, panel data

[*] We particularly thank Carluccio Bianchi, Roberto Golinelli, Emanuela Marrocu and Eduardo Rossi for advices and useful discussions. We are grateful to Adriana Di Liberto for kindly providing us with the STATA procedure for implementing the KIVIET estimator, and to Giacomo Degli Antoni who kindly made available to us his measure of social capital for Italian regions.

[†] *Corresponding author*: Department of Economics and Quantitative Methods, Via San Felice 5, 27100 Pavia, Italy. *E-mail*: gascari@eco.unipv.it.

1. Introduction

The role of technological progress has been the subject of an increasing attention in the literature to understand the differences between developed and undeveloped countries. Prescott (1998), Prescott and Parente (2004) highlight the role of total factor productivity (TFP) in explaining international income differences (see also Easterly, 2001). As shown in Parente and Prescott (2004), TFP, together with human capital, can explain a large part of the income differences across countries. As Prescott (1998) writes, the standard economic growth theory first needs to analyse the TFP determinants to become also a theory of international income differences.

The same is true if one wants to explain the different degree of development within a country. This seems particularly critical for countries whose areas or regions are characterised by very different degree of development, as it is clearly the case of Italy. In this paper we look at the main determinants of TFP across Italian regions, using a panel data approach.

There are two obvious candidates to explain the different levels of TFP across countries or across regions within countries. The more important one is the amount of research carried out in that region/country. A vast literature investigating the national sources of economic growth (e.g., Cameron, 2003, Griffith et al., 2000, 2001) underlines the linkage between R&D expenditures TFP, and growth.

The second one is human capital. A sufficient level of knowledge in the workforce is necessary to acquire and exploit technology. Hence, the level of human capital is considered a natural necessary condition for high TFP level. There is however a potential problem: human capital can be thought as an input in the production process rather than a source of higher TFP (see, e.g., Mankiw et al., 1992 and Parente and Prescott, 2004). In this case, it is not clear if human capital should work also as a TFP determinant, through knowledge externality, as e.g., Lucas (1998). In the paper, we will tackle this issue constructing two different measures of TFP for Italian regions (one with and one without human capital as production input) and testing for human capital effects.

The literature analysed a third important channel that can affect TFP. Coe and Helpman (1995) find that international technological spillovers due to investments in R&D activity drive economic growth in underdeveloped countries. The authors stress the role of international trade in driving technological spillovers through the imitative process that determines the technological performance of countries that cannot sustain an endogenous technological growth process.[1] Cameron (2001) includes in TFP analysis R&D expenditures, human capital and the import/GDP ratio and finds that only R&D expenditures raise the rate of innovation, while international trade simply enhances the speed of technology transfer form the highly technological country.

The literature proposes other possible sources of TFP. Cameron (2003) includes in his analysis the proportion of manual male workers covered by collective agreements multiplied by the proportion of manual males in the workforce, to understand the role played by labour unions in the productivity growth. Senhadji (2000) and RiveraBatiz (2002), instead, investigate the social components of TFP, as external shocks, political stability and democracy. Many of these effects however have a country dimension so that they are common to regions in the same country. More generally, one may think that the so-called "social capital" differs across regions within a country. However, it is difficult to come up with a good proxy for social capital (see Temple, 2001).

Finally we have to consider the contribution of information and communications technology (ICT) to productivity. In the analysis proposed by Stiroh (2002) emerges that the relationship between ICT and TFP is weak: first, it is difficult to account for heterogeneity across productivity and industries when analysing ICT and TFP linkages and secondly, there is evidence of variation in the productive impact of different types of ICT capital. In other words, as found in Griliches (1994), the measurement errors are an overwhelming problem to account for the ICT role in TFP changing.

The above quoted analysis proposed to identify TFP determinants are principally direct to the US or UK economies. In this paper we would like to present an empirical approach for the 20 Italian regions. Using a panel data approach with a sample period from 1985 to 2000, we allow for heterogeneity across regions in estimating the TFP determinants. This is particularly important in analysing Italian economy since Italian regions present a great variability in income levels and almost opposite, if not divergent, development paths of two macro-regions: the North and the South.

Some recent works investigated TFP across Italian regions. Our paper is close in spirit to Aiello and Scoppa (2000). Aiello and Scoppa (2000) first make a growth accounting exercise for the year 1997 showing that a large part of regional disparities are due to differences in TFP level across regions; then estimate a model of TFP determinants in a cross section referred to the 20 Italian regions for the year 1997. Our paper expands this analysis to a panel data approach using a sample period ranging from 1985 to 2000.

[1] See also Keller (1998) and Edmund (2001) for a critique of Coe and Helpman (1995).

Marrocu et al. (2000) instead estimate the TFP across Italian regions and economic sectors, relaxing the assumption of constant return to scale, with a panel cointegration approach. The authors find the presence of constant return to scale at a national level, while the estimates of individual production functions for Italian regions show great differences in factor elasticities across regions and sectors.

Di Liberto et al. (2004) employ a panel data approach (based on Islam, 1995) to study if TFP is converging across Italian regions. This paper also highlights the heterogeneity of TFP levels across regions. While linked to ours, Di Liberto et al. (2000) deal with a different topic, focusing on TFP convergence and showing that only in the sub-period 1963-1978 they can detect convergence in TFP levels across regions.

Finally, La Ferrara and Marcellino (2000) analyse the link between public capital and TFP. They find an equivocal relationship between TFP and public infrastructure, that is, the growth rate of public capital is generally negative or not significant in determining the TFP growth rate, except for the South. Following these results, we abstract from public capital in determining TFP at the regional level.

The empirical literature therefore emphasizes the large differences in TFP levels across Italian regions, coherently with the well-known dualism between Northern and Southern Italy. None of the above paper however investigates the reasons of such heterogeneity and what are the key factors in determining such differences. This is instead clearly an important task also from a policy perspective. In this paper we therefore explore if the TFP determinants, as suggested by the theoretical literature, could explain the different regional TFP estimates.

The structure of the paper is as follows. Section 2 introduces the measure of TFP at national level and describes the behaviour of regional TFP across time and regions. Section 3 presents our benchmark econometric analysis and results. Section 4 modifies the TFP measure to take into account the possibility of human capital being an input to the production process. In Section 5 we break our dataset in two macro-regions, the North and the South and provide econometric estimates of our benchmark equation for these two macro-regions. Section 6 concludes.

2. Measurement and data description

2.1. Measurement of inputs and output

We measure output as the value added at factor costs, in real Italian lire, at constant prices of 1995, source ISTAT. Labour input is simply the labour workforce, that is, we include both of autonomous and subordinate workers. The only measure of labour input available from ISTAT is the number of workforce, so there is no labour quality measure.

As a measure for capital input we use the gross capital stock data constructed from gross investment data, by Bonaglia and Picci (2000) and Picci (1999). Since this dataset ends in 1995, we calculate the capital stock for the sample 1996-2000 employing the perpetual inventory method and using a fixed depreciation rate calculated to be the average depreciation rate of the years 1990-1995.

2.2. TFP: measurement and description

We calculate TFP applying the standard Solovian growth accounting methodology. Denote regions by i =1,....N. GDP (*Y*) in each region at time *t* is produced with labour (*L*) and physical capital (*K*) according to a standard neoclassical production technology,

$$Y_{it} = A_{it} F(K_{it}, L_{it}) \qquad (1)$$

where A_{it} is the TFP. We consider a traditional Cobb-Douglas production function with constant return to scale, that is

$$Y_{it} = A_{it} K_{it}^{1-\alpha_i} L_{it}^{\alpha_i} \qquad (2)$$

Consequently, TFP is simply defined by: $A_{it} = \frac{Y_{it}}{K_{it}^{1-\alpha_i} L_{it}^{\alpha_i}}$. To calculate hence the TFP levels we just need a value for α_i. Under the (heroic) assumption of perfect factor markets, the exponent α_i equals the labour share of output, hence $\alpha_i = \frac{w_{id} L_i}{Y_i}$, where w_{id} is the nominal wage of subordinate workers at constant prices.[2] Note that we allow α_i to vary across regions, but not across time. This is because of the high variability across time of the labour share in the different regions, reflecting forces that are most probably not linked with technology. α_i instead is a technological parameter

[2] Since for Italian regions, there are no data available for the wages of autonomous workers, we consider the subordinate workers nominal wage w_{id} as a proxy for the autonomous one.

and should not vary too much in a sample of just 15 years. We therefore calculate α_i as the average of the labour share across the sample period for each region.

Before proceeding with the econometric analysis, it is instructive to describe the main features of the TFP levels of the Italian regions. As noted in the introduction, the literature suggests a large degree of heterogeneity among the Italian regions. Our results confirm such finding. Figure 1 plots the time series behaviour of TFP levels for Italian regions. Some features are evident:

1) Different regions exhibit very different TFP levels: the highest level (Lazio) is almost 3 times as large as the lowest one (Basilicata) along the sample period.

2) There is also a very high degree of persistence in the ranking of TFP levels across regions, which basically does not change along the sample.

3) There is mixed evidence on convergence. Figure 3 shows the standard deviation of TFP levels across regions. As regard TFP, while the '80's seem to be a period in which the gap between rich and poor regions widened, in the second part of the '90's some convergence took place.

4) As expected, TFP levels of Northern and Central regions (North) are substantially bigger than the ones of Southern regions (South), as evident from Figure 2 and Figure 4.[3] Figure 2 plots the average of TFP across the sample period for each region, showing that the Northern and Central regions are basically all above the average (roughly 0.15), while the Southern ones are below it. Figure 4 also visualises the same fact across time periods.

Finally, note that from equation (2) we can write: $\frac{Y_{it}}{L_{it}} = A_{it}\left(\frac{K_{it}}{L_{it}}\right)^{1-\alpha_i}$, which shows that the level of labour productivity is explained only by two factors: TFP and the capital-labour ratio. It follows that the regional differences in labour productivity can be explained through disparities in these two factors. Figure 5 shows that most of the heterogeneity in output-per-worker among different regions is explained by TFP rather than by the capital-labour ratio. Indeed, the correlation between TFP and output per worker is always above 0.9 in each year of the sample, while the correlation between capital-labour ratio and output per worker is rather low.

[3] We define North the macro-region composed of Northern and Central regions, that is, Piemonte, Valle d'Aosta, Lombardia, Trentino Alto Adige, Friuli, Liguria, Emilia Romagna, Toscana, Umbria, Marche and Lazio. The macro-region South is instead made up by Abruzzo, Campania, Puglia, Basilicata, Calabria, Sicilia and Sardegna.

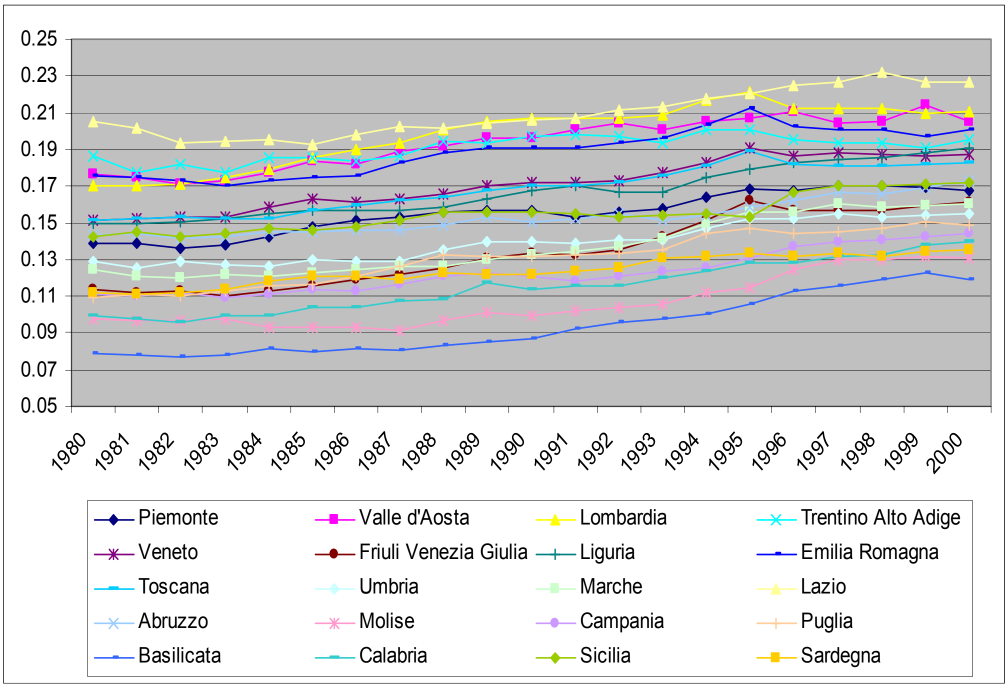

Figure 1. TFP levels of Italian regions

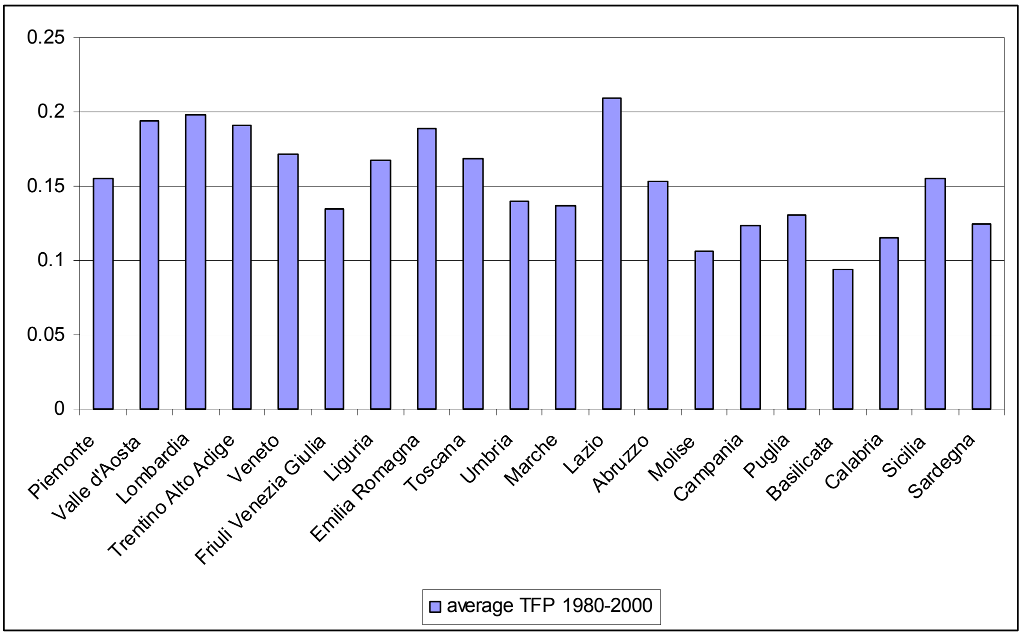

Figure 2. Average TFP levels for Italian regions 1980-200

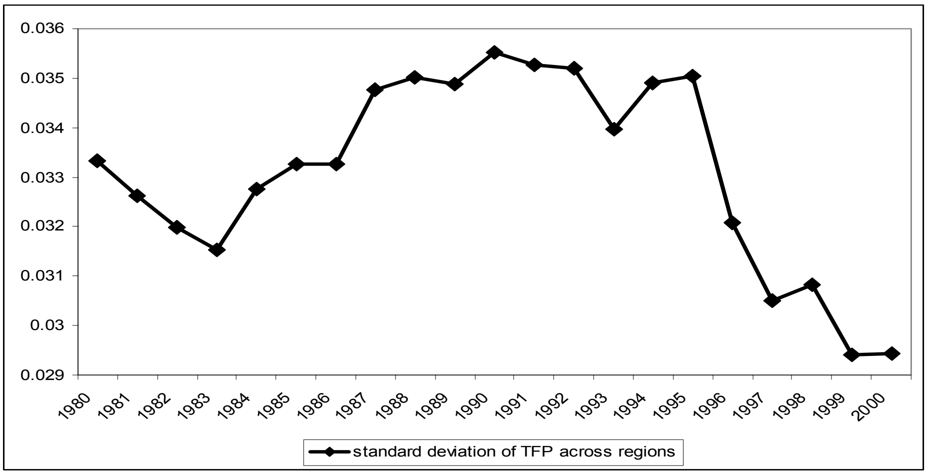

Figure 3. Standard deviation of TFP across Italian regions, 1980-2000

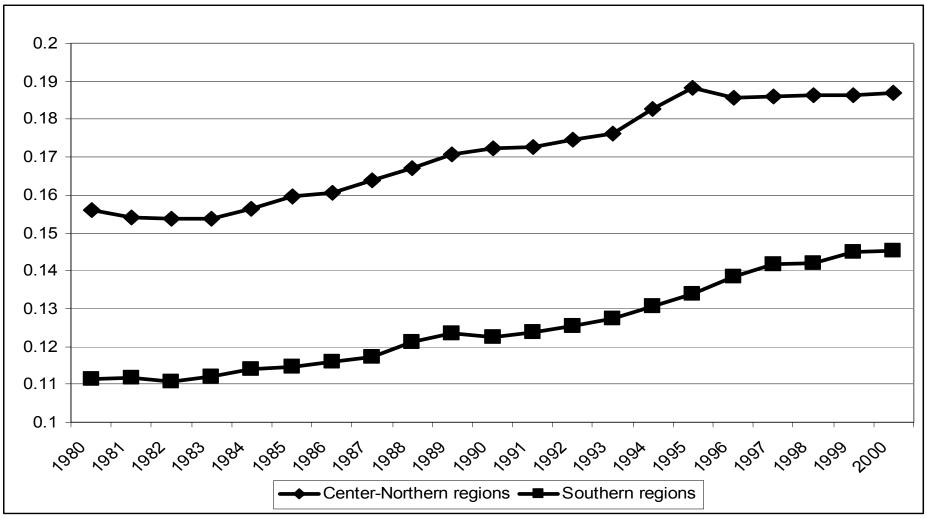

Figure 4. Average TFP levels for Central-Northern and Southern regions, 1980-2000

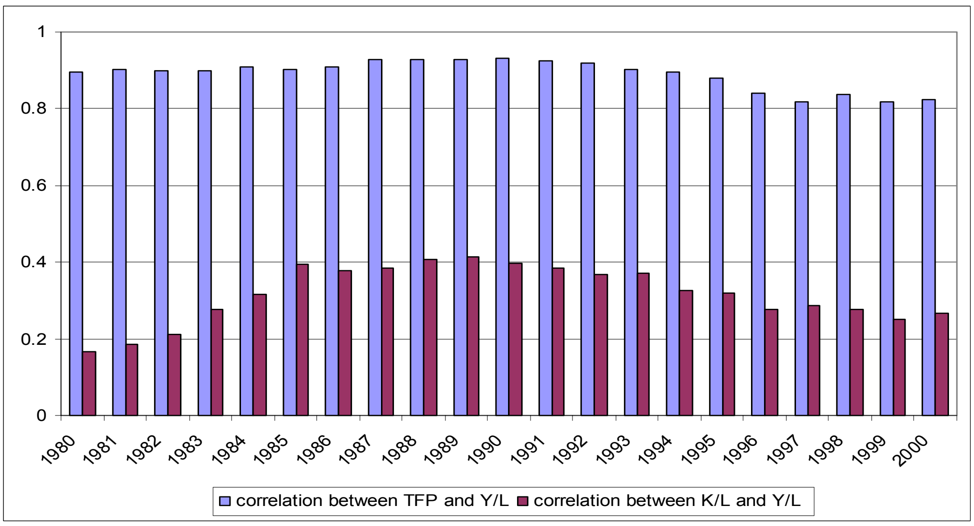

Figure 5. Correlation between TFP, output per worker and capital per worker, 1980-2000

3. Empirical results

As argued in the Introduction, the theoretical literature suggests the following variables as main candidates to affect the total factor productivity: research activity, human capital, trade and social capital. We therefore use the following variables in our benchmark regression equation:

- R&D, the stock of private enterprises spending in R&D divided by the total fixed investments;
- RICPOP, the number of researcher on total workforce;
- IMPORT, the ratio of total imports on GDP;
- EDU, the average years of schooling of the labour force;
- SOCIALK, used as a proxy for social capital. This variable is the main common factor of three indicators of crime and lawsuits. As stressed by Temple (2001), it is difficult to find a suitable variable for "social capital", but we propose to use SOCIALK as a proxy for trust and cohesion within the local community.[4]

Finally, we try to control the potential non-linear relation between the dependent variable and regressors taking the natural logarithm of TFP in our final regression. Our benchmark regression equation is then:

$$\begin{gathered}\log TFP_{it} = \alpha_{it} + \alpha_1 R\&D_{it} + \alpha_2 RICPOP_{it} + \alpha_3 IMPORT_{it} + \alpha_4 EDU_{it} + \alpha_5 SOCIALK_{it} + z_{it} \\ \text{where} \\ z_{it} = e_i + v_{it}\end{gathered} \tag{3}$$

3.1 Static Panel

In this section we present the results of estimation of (3) for the sample 1985-2000.[5] We use the within group estimator (LSDV), since the Hausman test rejects the null hypothesis of no correlation between regressors and error term v_{it}. In this case then GLS would be inconsistent and inefficient. Moreover, the Breusch-Pagan test suggests that the LSDV estimates better fit the true model than the pooled OLS estimates. The results are reported in Table 1.

[4] We gratefully thank Giacomo Degli Antoni, from the University of Parma, who kindly provided us with this variable. The way Degli Antoni builds this variable is described in the Appendix. Sources of all the data series are also presented in the Appendix.

[5] We have to restrict the sample with respect to the Figures in Section 2, because some of the regressors are available only from 1985.

Table 1. LSDV estimation results

logTFP		Coefficients
IMPORT		-.0940337
		(0.775)
R&D		.6543859 *
		(0.205)
RICPOP		2.48803 *
		(1.141)
EDU		.0987666 *
		(0.003)
SOCIALK		-.0042198
		(0.006)
	Piemonte	-2.818383
		(0.474)
	Valle D'Aosta	-2.530978
		(0.036)
	Lombardia	-2.566246
		(0.050)
	Trentino Alto Adige	-2.560995
		(0.039)
	Veneto	-2.646425
		(0.039)
	Friuli	-2.941153
		(0.040)
	Liguria	-2.783094
		(0.043)
	Emilia Romagna	-2.60142
		(0.039)
	Toscana	-2.686813
		(0.038)
	Umbria	-2.917172
		(0.037)
	Marche	-2.878422
		(0.036)
	Lazio	-2.613788
		(0.044)
	Abruzzo	-2.820766
		(0.039)
	Molise	-3.138294
		(0.035)
	Campania	-3.021776
		(0.037)
	Puglia	-2.90485
		(0.035)
	Basilicata	-3.229748
		(0.040)
	Calabria	-3.064465
		(0.037)
	Sicilia	-2.775011
		(0.035)
	Sardegna	-2.954355
		(0.036)
R^2		0.728
$corr(u_i, Xb)$		0.1542

Standard errors shown in parenthesis below each coefficient. Coefficient with * are significant at 5% level.

Note that both the variables relating to research activity (R&D and RICPOP) and human capital (EDU) are largely significant in the regression with the expected positive sign. Hence: (i) technological progress is an important determinant of Solow residuals; (ii) there are positive spillover effects from human capital and skilled labour workers on TFP (see also Cameron, 2001). As we noted, the theoretical literature suggests these two factors as the most important determinants of TFP, and our estimates do confirm this conjecture.

On the other hand, IMPORT and SOCIALK are not significant, even at 10% level. Our data hence seem to disagree with Coe and Helpman (1995) that point to trade as a possible important source of technology spillover. We are more inclined to think, however, that our data are not well suited to test for the effect of trade on TFP. Indeed, since there are no data for interregional trade, we cannot control whether the imported goods stay in the region or are just passed over to other regions in the country. Indeed, this seems to be the case: quite obviously there are some natural places (e.g., Milan, Rome) and borders (e.g., Brennero, Monte Bianco, Brindisi) that act just as a commercial entry to Italy for international goods that are then redistributed throughout the country. A second important reason why our data on import are not a particular good proxy for trade as a source of technology spillover lies in the fact that we cannot disentangle the type of imported goods to control for the amount of embedded technology and knowledge. Importing food is not the same as importing microelectronic equipments, as regards to possible technology spillover. Unfortunately, however, again our data do not allow us to do that. There are therefore good reasons for IMPORT to be not significant.

With respect to SOCIALK, the data show that social capital does not seem to be a determinant for regional TFP for Italy (but see Section5).

3.2 Dynamic Panel

The high t-values of intercepts in fixed effect estimates advise that in the static panel estimation above the regressors are unable to completely determine the dependent variable. To cope with the potential problem of omitted variables in the regression leads us to modify (3) to

$$\log TFP_{it} = \gamma \log TFP_{it-1} + \alpha_1 R\,\&\,D + \alpha_3 RICPOP + \alpha_4 IMPORT + \alpha_5 EDU + \alpha_5 SOCIALK + z_{it}$$

with (4)

$$z_{it} = e_i + v_{it}$$

e_i, v_{it} supposed normally and identically distributed.

It is well known that in the dynamic case, the LSDV estimator is inconsistent as the average of the lagged endogenous variable is always related with the vector of residuals.

Kiviet (1995) proposes a correction for panel data with a small time dimension, based on a two step procedure, where the residuals from a first step consistent estimator (typically obtained using Andersen and Hsiao instrumental variables) are used in a second stage calculation of the bias. We call the Kiviet corrected estimator LSDVc[6].

We can rewrite the (4) as

$$\log(TFP)_{it} = \gamma \log(TFP)_{it-1} + x'_{it}\alpha + e_i + \nu_{it} \tag{5}$$

We can also assume that TFP_{i0} may be correlated with e_i but not respect to ν_{it}, that is:

$$\begin{aligned} &E(\log(TFP_{i0})\nu_{it}) = 0 \\ &E(\log(TFP_{i0})e_i) = \text{unknown} \end{aligned} \tag{6}$$

and that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}x'_{it} \text{ converges to } A = (a_{ij})_{n\times n}, \text{where } \det(A) \neq 0 \tag{7}$$

Now (5) can be rewritten as

$$\log(TFP) = \gamma \log(TFP)^{(-1)} + X\alpha + (I_N \otimes \iota_T)e + \nu \tag{8}$$

where

$$\log(TFP) = \begin{bmatrix} \log(TFP)_{11} \\ . \\ . \\ . \\ \log(TFP)_{NT} \end{bmatrix}, \log(TFP)^{(-1)} = \begin{bmatrix} \log(TFP)_{10} \\ . \\ . \\ . \\ \log(TFP)_{N(T-1)} \end{bmatrix}, X = \begin{bmatrix} x'_{11} \\ . \\ . \\ . \\ x'_{NT} \end{bmatrix}, e = \begin{bmatrix} e_{11} \\ . \\ . \\ . \\ e_{NT} \end{bmatrix}, \nu = \begin{bmatrix} \nu_1 \\ . \\ . \\ . \\ \nu_N \end{bmatrix}, \iota_T = \begin{bmatrix} 1 \\ . \\ . \\ . \\ 1 \end{bmatrix}$$

Defining

$W = \left[\log(TFP)^{(-1)} \vdots X\right]$ and $\delta' = (\gamma, \alpha')$

[6] To perform the LSDVc estimator we use the procedure for STATA of Adam (1998).

then (8) can be written as

$$\log(TFP) = W\delta + (I_N \otimes \iota_T)e + \nu \tag{9}$$

the LSDV estimator of δ, can be expressed as

$$\hat{\delta} = (W'AW)^{-1}W'A\log(TFP) \tag{10}$$

in which $A = I_N \otimes A_T$ and $A_T = I_T - \frac{1}{T}\iota_T\iota'_T$; which is inconsistent for T finite and $N \to \infty$ as expressed above.

As a result, the estimated bias is given by:

$$E(\hat{\delta} - \delta) = E\left\{\left[(W'AW)^{-1}W'A(W\delta + (I_N \otimes \iota_T)e + \nu)\right] - \delta\right\} = E(W'AW)^{-1}W'A\nu \tag{11}$$

Kiviet (1995) proposes a correction for the Nickell (1981) estimated bias for the γ coefficient, calculating the bias for $\hat{\delta}$ which includes both $\hat{\gamma}$ and $\hat{\alpha}$ coefficients. As expected, under the assumption of a normal and independent error terms distribution and the hypothesis (6) and (7), in the model without exogenous regressors, Nickell (1981) finds a bias with a residual term smaller than the one given by (11). Kiviet (1995) performs simulations to assess the sample bias of the LSDV estimator, and then corrects the LSDV estimator for this bias for dynamic models. In order to implement the unbiased estimator, one would need to know the variance of the residuals (namely, $\sigma_\nu^{\ 2}$). The correct residuals are obtained using an instrument variables dynamic estimator, as specified above. Monte Carlo evidence presented by the author and confirmed by Judson and Owen (1996) and by Blundell et al. (2001), finds that the LSDV corrected estimator perform better than the GMM estimator for a panel with a small temporary dimension (and large N).

Table 2 shows the coefficient estimates corrected for the bias due to the small T dimension using Kiviet's algorithm. The results confirm the strong persistence of the TFP series and indeed, the lagged TFP is strongly significant. Unfortunately, the other variables loose their explanatory power. We interpret this fact as due to the small cross-sectional dimension of our panel, since typically dynamic panel data estimator are advocated for panel with very high cross-sectional dimension.

Table 2. Kiviet correction bias for panel with small temporary dimension estimates

logTFP	Coefficient
TFP(-1)	0.977* *(0.352)*
IMPORT	0.079 *(0.096)*
R&D	0.119 *(0.1785)*
RICPOP	1.259** *(0.723)*
EDU	0.0040 *(0.0216)*
SOCIALK	-0.01* *(0.0037)*

Standard errors shown in parenthesis below each coefficient. Coefficient with * are significant at 5% level, with ** at 10%.

To investigate the robustness of our results, we also present the Arellano and Bond (1991) GMM estimation procedure, bearing in mind again the caveats of using dynamic panel estimator in macroeconomic studies where typically the N dimension is short.

The basic idea of Arellano and Bond (1991) GMM estimation procedure is that, given (5) and (6), assuming that both transient errors not serially correlated

$$E[\nu_{it}\nu_{is}] = 0 \text{ for } i = 1...19, t = 2...15$$

and

$$E[TFP_{i1}\nu_{i1}] = 0 \text{ and } E[e_i \Delta x_{it}] = 0 \text{ for } t = 2..15, i = 1..19$$

imply the following *m*=0.5(T-1)(T-2) moment restrictions

$$E[(TFP_{it-s}, x_{it-s})\Delta\nu_{it}] = 0 \text{ for } t = 3...15 \tag{12}$$

In other terms:

$$E[Z_i' \Delta\nu_i] = 0$$

with

$$Z_i^{'}=\begin{bmatrix} TFP_{i1}x_{it-1} & 0 & .. & 0 \\ 0 & TFP_{i1}x_{it-1}TFP_{i2}x_{it-2} & 0 & .. \\ .. & .. & TFP_{i1}x_{it-1}TFP_{i2}x_{it-2}TFP_{i3}x_{it-3} & .. \\ .. & .. & .. & .. \\ 0 & .. & .. & TFP_{i1}x_{it-1}TFP_{i2}x_{it-2}...TFP_{iT-2}x_{iT-1} \end{bmatrix}$$

$(T-2)\times m$ instruments matrix and Δv_i is the (T-2) vector $\left(\Delta v_{i3}....\Delta v_{iT}\right)'$.

The standard GMM estimator is then obtained simply differencing (5) and considering the residual condition[7]

$$E\left[\Delta v_i \Delta v_i^{'}\right]=\sigma_v^2(I_N \otimes G) \qquad (13)$$

in which G is a matrix given by

$$G=\begin{bmatrix} 2 & -1 & .. & .. & 0 \\ -1 & 2 & 0 & .. & .. \\ .. & .. & .. & .. & .. \\ .. & .. & -1 & 2 & -1 \\ 0 & .. & 0 & -1 & 2 \end{bmatrix}$$

as Δv_i is a MA(1) process.

We control for potential problems due to the presence of endogenous regressors, by using as instruments the lagged values of observed x_{it} and TFP_{it} series dated (T-2) and earlier which, given our assumptions, are valid instruments for (5) in differences. Otherwise, the GMM estimator could not be used with persistent series, since it is seriously biased as the autoregressive parameter approaches unity or as $\sigma_{e_i}^2$ increases with respect to $\sigma_{v_i}^2$, although the presence of other regressors, different from the lagged endogenous variable, may improve the behavior of GMM estimates. (See Blundell and Bond, 2001; Bond et al 2001). The GMM estimator implies the use of lagged level instruments and yields a consistent estimator for γ as $N\rightarrow\infty$ and T is fixed. Since consistency is for $N\rightarrow\infty$, again we stress that these estimators are well suited for micro studies where typically the N dimension is very large.

[7] The GMM estimator considers the residual condition given by (13) and use the orthogonality condition expressed in (12). Having these two conditions, the GMM is thus more efficient than the Andersen-Hsiao IV estimator.

The results of GMM estimation are shown in Table 3.

Table 3. Arellano-Bond dynamic panel data estimation

logTFP	**GMM1**	**GMM2**
logTFP(-1)	.793868 *	.8669724 *
	(0.055)	*(0.055)*
R&D	.2812823 **	.2190665 **
	(0.170)	*(0.121)*
IMPORT	.1339532	.0908795
	(0.090)	*(0.721)*
RICPOP	1.968069 *	1.441213 *
	(0.659)	*(0.531)*
EDU	.0078583	-.0115748
	(0.017)	*(0.010)*
SOCIALK	-.0013027	-.0034515
	(0.004)	*(0.003)*

GMM1 assumes regressors are exogenous, GMM2 assumes regressors are only weakly exogenous.
Robust standard errors are shown in parenthesis.
Coefficient with * are significant at 5% level, with ** at 10% level.

In the GMM case, the results are quite similar to the static panel one since both R&D and RICPOP are significant (even if now R&D is significant only at the 10% level). Moreover, both IMPORT and SOCIALK are again not significant. There is an important difference however with respect to the variable EDU, that becomes not significant at the conventional levels. This fact is surprising given the very strong significance level of this variable in the LSDV estimation. A first interpretation is that EDU was just picking up some of the strong autocorrelation of the TFP series, working then as a proxy for TFP_{-1}, which is indeed very strongly significant in the GMM Arellano-Bond and Kiviet estimates. Another possibility, after investigating the properties of the series, is that the strong significance of the EDU variable in the LSDV estimate may be due to a spurious regression problem. Both the TFP levels (see Figure 1) and the variable EDU in fact are characterised by a positive trend, which is instead not common to the other variables in the regression.

3. TFP and human capital as an input to production

After the seminal works of Lucas (1988) and Mankiw et al. (1992), it is common in the literature to consider human capital as an input of the production function. In this case, we can recalculate the TFP as

$$A_{it} = \frac{Y_{it}}{K_{it}^{1-\alpha_i}(L_{it}H_{it})^{\alpha_i}} \qquad (14)$$

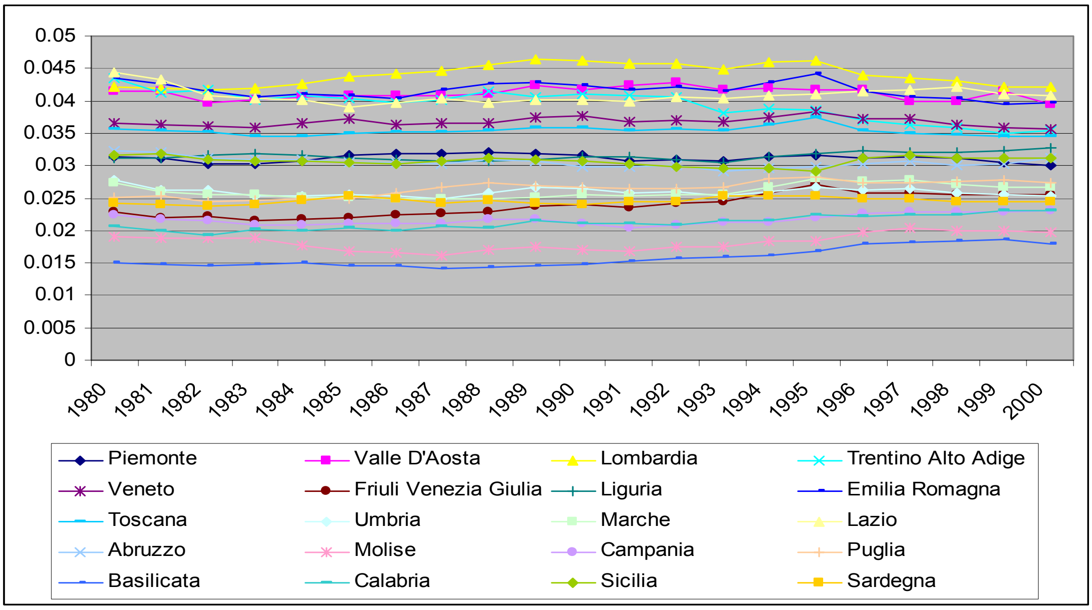

Figure 1. TFP levels of Italian regions, considering human capital as an input to production

Figure 6 plots the time series behaviour of the TFP for the sample 1980-2000. The main regional TFP features described in Section 2 are qualitatively still valid, especially regarding the North vs. South difference. The main, and important, change is that now the TFP does not show any evident trend.[8]

We then perform the same analysis of section 3. We would like, however, to test for possible external effects generated by human capital, as put forward first by Lucas (1988). In particular, we want to catch the idea of possible spillover effects when the labour force is composed of a large fraction of specialised workers. Following Temple (2001) and Cameron (2003), we therefore define the variable SKILL, that is the proportion of administrative workers on the total labour force. We can estimate the following equation

$$\log TFP_{it} = \alpha_{it} + \alpha_1 R\&D_{it} + \alpha_2 RICPOP_{it} + \alpha_3 IMPORT_{it} + \alpha_4 SKILL_{it} + \alpha_5 SOCIALK_{it} + z_{it}$$

where (15)

$$z_{it} = e_i + v_{it}$$

As in the previous section, we employ different estimation methods to check the robustness of our results. The results are summarised in Table 4. The main finding is that the variables related to

[8] This may suggest that TFP and EDU were cointegrated in the regressions of the previous section. We chose to ignore the non-stationarity issue in the econometric treatment of the previous section for two reasons. (i) 15 time observations do not seem to be enough to allow any sensible panel cointegration analysis; (ii) the analysis in this section does not seem to suffer of this problem, while it anyway confirms our result. Cointegration analysis is then left to further research.

research activity (R&D and RICPOP) are always significant across estimation methods.[9] This is an important result since it indicates that research activity is the main determinant of TFP and the main cause of TFP (and income) differences across Italian regions. This finding is robust across estimation methods and TFP definitions.

Table 4. Estimation results for the TFP calculated considering human capital as an input to production

		LSDV	LSDVc	GMM1	GMM2
logTFP(-1)			0.9527645*	.7014857 *	.7886282 *
			(0.385)	*(0.057)*	*(0.054)*
IMPORT		.0874451	0.09392705	.1926244 **	.145223
		(0.076)	(0.107)	*(0.104)*	*(0.088)*
R&D		.4272384 *	0.1286362	.2254378**	.2285486 *
		(0.020)	(0.181)	*(0.147)*	*(0.116)*
RICPOP		2.349573 *	1.2317725**	2.174854 *	1.813295 *
		(1.155)	*(0.708)*	*(0.642)*	*(0.534)*
SOCIALK		-.0051879	-0.03461425*	-.0067923	-.0063622
		(0.006)	*(0.005)*	*(0.004)*	*(0.004)*
SKILL		.1897384 *	-0.06345312	-.1452536	-.1313839
		(0.072)	*(0.008)*	*(0.142)*	*(0.111)*
Fixed effects	Piemonte	-3.599025			
		(0.040)			
	Valle D'Aosta	-3.27663			
		(0.029)			
	Lombardia	-3.238216			
		(0.045)			
	Trentino Alto Adige	-3.328551			
		(0.030)			
	Veneto	-3.374345			
		(0.031)			
	Friuli	-3.822226			
		(0.034)			
	Liguria	-3.590201			
		(0.037)			
	Emilia Romagna	-3.276014			
		(0.030)			
	Toscana	-3.433604			
		(0.031)			
	Umbria	-3.739752			
		(0.028)			
	Marche	-3.724728			
		(0.026)			
	Lazio	-3.358764			
		(0.043)			
	Abruzzo	-3.600242			
		(0.030)			
	Molise	-4.095354			
		(0.026)			
	Campania	-3.942216			
		(0.034)			
	Puglia	-3.22703			
		(0.030)			
	Basilicata	-4.218584			
		(0.026)			
	Calabria	-3.942645			
		(0.032)			
	Sicilia	-3.594486			
		(0.033)			
	Sardegna	-3.788179			
		(0.029)			

[9] With the exception of R&D in Kiviet's LSDVc.

For the GMM1 case the significance is at the 10% level, but we present this case just for completeness, since it is clear that the variable SKILL cannot be considered exogenous, and therefore the GMM2 estimations are more reliable for our model.

With respect to the spillover effects of a qualified workforce, the evidence is mixed. While the static LSDV estimator finds a strongly significant and positive external effect, SKILL is not significant using the dynamic GMM Arellano-Bond estimator. As in the previous section the human capital variable is strongly significant in the static estimates, but its significance vanishes in the dynamic ones, as if the human capital variable was able to catch some of the persistence of the TFP.

5. TFP: North and South

In this section, we analyse the well-known dichotomy between the North and the South among Italian regions. Our data allow us to separately investigate the TFP determinants for the Northern regions (i.e., Piemonte, Valle d'Aosta, Lombardia, Trentino Alto Adige, Friuli, Liguria, Emilia Romagna, Toscana, Umbria, Marche and Lazio) and the Southern ones (i.e., Abruzzo, Campania, Puglia, Basilicata, Calabria, Sicilia and Sardegna). For this exercise we use the TFP as calculated in our benchmark model in section 3.

Given the previous result and the discussion in section 3, we exclude from our analysis the variable IMPORT, basically because both this type of data cannot pick up the correct external trade of the different regions and it is never significant in the regressions presented above. So, the benchmark equation given by (4) becomes:

$$\log TFP_{it} = \alpha_{it} + \alpha_1 R\&D_{it} + \alpha_2 RICPOP_{it} + \alpha_3 EDU_{it} + \alpha_4 SOCIALK_{it} + z_{it}$$

in which (16)

$$z_{it} = e_i + v_{it}$$

5.1 North

The Hausman test suggests estimating the model with random effects[10]. Table 5 shows that in this case all the variables have the expected positive effect on the endogenous variable, and they are all strongly significant, except for RICPOP , that is significant at 10% level.

Table 5. Estimates for the Northern regions

[10] The test accepts the null hypothesis with a p value of 0.9999

LogTFP	GLS	LSDVc	GMM1	GMM2
logTFP(-1)		0.921165*	.8735172 *	.9027004*
		(0.393)	*(0.007)*	*(0.0798)*
R&D	.3950904 *	0.053582	.0915161	.0944263
	(0.175)	*(0.169)*	*(0.106)*	*(0.1048)*
RICPOP	1.425288 **	0.259528	.7402415	.485252
	(1.025)	*(0.811)*	*(0.446)*	*(0.4904)*
EDU	.0800056 *	-0.00202	.0082371	.0110651
	(0.004)	*(0.022)*	*(0.015)*	*(0.0149)*
SOCIALK	.0213811 *	0.006935	.0033585	.0053312
	(0.010)	*(0.011)*	*(0.006)*	*(0.0072)*
R^2	0.747			

Standard errors are reported in parenthesis, coefficients wit * are significant at 5% level, with ** at 10% level.

Note that the estimates for the Northern regions exhibit a significant effect of the social capital variable. This is noteworthy since in the previous sections SOCIALK was not significant when all the Italian regions were included in the panel. The positive sign is the expected one, meaning that TFP increases with this social capital variable, considered as a proxy for unity and trust in the social community. This would signal that a more friendly and serene relationship within the community creates the right environment for the development of economic activity.

When we perform dynamic estimation, all coefficients became non significant at standard levels, except the lagged endogenous variable. We interpret this result as evidence that the undersized cross-sectional dimension do not justify the use of either the Kiviet and the GMM estimators.

5.2 South

We estimate the static panel with random effects, because Hausmann test accepts the null hypothesis with a p-value of 1.000. Moreover, we again stress that the use of dynamic estimators with such a low numbers of regions may be not justified, but we choose to present the estimates mainly for completeness.

Table 6. Estimates for the Southern regions

logTFP	GLS	LSDVc	GMM1	GMM2
logTFP(-1)		0.915626	.76341 *	.7794907 *
		(0.6971)	*(0.107)*	*(0.112)*
R&D	-.3633251	0.307919	.1565446	.2162585
	(1.115)	*(1.134)*	*(0.497)*	*(0.431)*
RICPOP	3.159734	4.443149*	4.390318 *	4.440764 *
	(2.944)	*(1.9721)*	*(1.412)*	*(1.469)*
EDU	.1477072 *	0.00667	.0829374 **	.0772602 **
	(0.007)	*(0.667)*	*(0.044)*	*(0.044)*
SOCIALK	-.0143419 **	-0.00549	-.0082565 *	-.0088776 *
	(0.008)	*(0.008)*	*(.0024)*	*(0.002)*
R^2	0.8417			

Standard errors are reported in parenthesis, coefficients with * are significant at 5% level, with ** at 10% level.

Strangely enough, however, the variable R&D, which so far has been always significant across estimation techniques, is now not significant in all our performed estimations for the South. RICPOP is also not significant in the GLS estimates, while it is significant in the dynamic ones (i.e., LSDVc and GMM). These findings suggest that the amount of research activity is a less important determinant of TFP in the Southern regions with respect to the Northern ones.

Moreover, the SOCIALK variable is significant but with a negative sign. Together with the opposite finding for the North, this reconciles the evidence of a not significant SOCIALK variable for Italy as a whole in the regressions of section 3 and 4. This result however stresses even further, and possibly in a more worrying way, the Italian North-South dichotomy. There are in fact two possible interpretations for a negative coefficient on the SOCIALK variable. The first one is simply that this variable is not a good proxy for the social capital in the Southern regions, but then it is not clear why it should on the contrary be so for the Northern ones. The second one, instead, points to the strong peculiarity of the Italian Mezzogiorno: a widespread and powerful organised crime. A quiet environment can then basically correspond to a “social peace” secured by organised crime with a detrimental effect of economic activity and TFP: Whatever one may think, surely, the estimated regression model works much better for the North than for the South, pointing to a peculiarity of the South which deserves further research.

6. Conclusions

A large literature by now points to TFP as the main determinant of income differences across countries and regions within countries. In this paper we try to empirically assess the determinants of TFP levels for Italian regions. This seems a quite important task given the great disparity in the level of development between the different Italian regions.

Our results do confirm the prediction of the theoretical literature, pointing to research activity and human capital as the main determinants of TFP differences across Italian regions. This I particularly true for the static panel estimations. Moreover, we do use also dynamic panel data estimation techniques, namely Kiviet’s (1995) LSDVc and Arellano and Bond’s (1991) GMM, even if we acknowledge from the outset that the limited cross sectional dimension of our panel put some doubts on the reliability of these dynamic estimates. Overall we con conclude that our results are quite robust across different estimation procedure, and particularly the GMM2 dynamic estimates tend to confirm the results from the static panel. Finally, when we divide the Italian regions in two macro-regions (North and South), the notorious dualism of Italian development clearly emerges, with respect both to TFP levels and to estimation result, notably with respect to the social capital variable.

Data Appendix

Real Output: Data on real output are taken from the Office of National Statistics (ISTAT). As in 1995 a new accounting methodology (Sec95) was imposed by the new European classification, we recalculate the old series (Sec79) as follows. We assume that the proportion existing between the new 1995 series and the old one also exist between the old series (1980-1994) and the new ones. That is, we impose the condition

$$1994_{Sec95} = \frac{1995_{Sec95} * 1994_{Sec79}}{1995_{Sec79}}$$

Value added :Data are taken from the CRENoS database NewRegio02 available on the web site www.crenos.it

Labour input : Total employment is from the Office of National Statistics (ISTAT).

Capital input: Data for individual capital are supplied by Bonaglia and Picci (2000) and Picci (1999).

Total Factor Productivity: This is calculated from data on real output, labour input and capital input as above. As explained in the main text, the measure of TFP is based on growth accounting methodology.

R&D: data on enterprise R&D spending are provided by ISTAT, as enterprise total fixed investments.

RICPOP: the source is again ISTAT and it is just the number of researchers over the total workforce.

SKILL: This measure of human capital, as suggested by Cameron (2003), is the ratio of non-manual (i.e. administrative, clerical and technical staff to total workers) to total workers. This is taken from ISTAT.

EDU: We define the total stock of human capital of the labour force as Di Liberto and Symons (1998), that is:

$$EDU_i = \sum_J AS_{Ji} \cdot HK_{Ji}$$

where J is the schooling level, AS_{Ji} is the number of years of schooling represented by level J, and HK_J is the fraction of the labour force for which the Jth level of education represents the highest level attained. Within the Italian system, primary level includes eight years of schooling, secondary level is attained after 5 years, and we consider that university courses include four years of attendance. As explained by Di Liberto and Symons (1998), this indicator represents a measure of the average years of schooling of the labour force. Data are again taken from ISTAT.

SOCIALK : This is a proxy for social capital, obtained by Giacomo Degli Antoni, University of Parma, through the statistical analysis of principal component of three variable (all provided by ISTAT): (i) the number of bill protests, (ii) reports to the police, (iii) labour lawsuit proceedings. SOCIALK is then the first principal component. Degli Antoni shows that this common factor explains around 75% of the comovement of the variables.

References

Adam, C.S. 1998. Implementing small-sample bias corrections in dynamic panel data estimators using STATA, mimeo University of Oxford.

Aiello, F; Scoppa, V. 2000, Uneven regional development in italy: explaining differences in productivity levels, *Giornale degli Economisti e Annali di Economia*, **59**, 270-98.

Anderson, T and Hsiao, C. 1981. Estimation of dynamic models with error components, *Journal of American Statistical Association*,**76**, 598-606.

Arellano, M.; Bond, S. 1991. Some tests of specification for panel data, *Journal of Econometrics* **59**, 87-97.

Blundell, R; Bond, S and Windmeijer, F. 2000. Estimation in dynamic panel data models : improving on the performance of the standard GMM estimators, *IFS Working Papers*, **00/12**.

Bonaglia, F; Picci, L .2000. Il capitale nelle regioni italiane, *DSE,Università di Bologna,***374.**

Bond, S; Hoeffler, A and Temple, J. 2001. GMM estimation of empirical growth models, *CEPR Discussion Paper*, **3048**.

Cameron, G. 2003. Why did UK manufacturing productivity growth slow down in the 1970s and speed up in the 1980s?, *Economica*, **70** 121-41.

Cameron, G; Proudman, J and Redding, S. 2001. Technological convergence, R&D, trade and productivity growth, *Bank of England Working Paper*

Coe, D; Helpman, E. 1995. International R&D spillovers, *European Economic Review*, **39** 859-87.

Di Liberto, A; Symons, J.1998. Human capital stocks and the development of italian regions: a panel approach, *Quaderni di ricerca CRENoS*

Di Liberto, A; Mura, R and Pigliaru, F. 2004. How to measure the unobservable: a panel technique for the analysis of TFP convergence, *Quaderni di ricerca CRENoS*

Easterly, W; Levine, R. 2001. It is not factor accumulation: stylized facts and growth models, *The World Bank Economic Review,***15**, 177-219.

Edmund, C. 2001. Some panel cointegration models of international R&D spillovers, *Journal of Macroeconomics*, **23**, 241-60.

Griffith, R; Redding, S and Van Reen, J. 2000, Mapping the two faces of R&D: productivity growth in a panel of OECD industries, *CEPR discussion paper*, **2457**.

Griffith, R; Redding, S and Van Reen, J. 2003. R&D and absorptive capacity: theory and empirical evidence, *Scandinavian Journal of Economics*, **105**, 99-118

Griliches, Z. 1994. Productivity, R&D and the data constraint, *The American Economic Review*, **84**, 1-23.

Judson, R; Owen, A. 1999. Estimating dynamic panel data models: a practical guide for macroeconomists, *Economics Letters*, **65**, 9-15.

Keller, W. 1998. Are international R&D spillovers trade-related? Analyzing spillovers among matched trade partners, *European Economic Review*, **42**, 1469-81.

Kiviet, J. 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models, *Journal of Econometrics,***68,** 53-78.

Islam, N .1995.Growth empirics : a panel data approach, *Quarterly Journal of Economcis,* **110** , 1127-70.

La Ferrara, E; Marcellino, M. 2000. TFP,costs and public infrastructure: an equivocal relationship, *IGIER working papers* **176**.

Mankiw, N; Romer, D and Weil, N. 1992. A contribution to empirics of economic growth, *The Quarterly Journal of Economics*, **107**, 407-37

Marrocu, E; Paci, R and Pala,R. 2000. Estimation of total factor productivity for regions and sectors in Italy. A panel cointegration approach *Quaderni di ricerca CRENoS*

Nickell, S. 1981. Biases in dynamic models with fixed effects, *Econometrica*, **49**, 359-86

Lucas, R. 1998. On the mechanics of economic development, *Journal of Monetary Economics*, **22**, 3-42.

Parente, S. and Prescott, E. 2004.A unified theory of the evolution of international income levels, *Federal Reserve Bank of Minneapolis Staff Reports*, **333**.

Picci, L., 1999, Productivity and infrastructure in the Italian regions, *Giornale degli Economisti e Annali di Economia*, **58**, 329-353.

Prescott, E. 1998. Needed: a theory of total factor productivity, *International Economic Review*, **39**, 525-551.

Rivera-Batiz, F; Rivera-Batiz, L.2002.Democracy, partecipation and economic development: an introduction, *Review of Development Economics*, **6**, 135-50

Senhadji, A. 2000. Sources of economic growth: an extensive growth accounting exercise, *IMF Staff Papers,* **47**.

Stiroh, K. 2002. Are ICT spillovers driving the new economy? *Review of Income and Wealth,***48**, 33-57

Temple, J. 2001. Growth effects of education and social capital in the OECD countries, *CEPR Discussion Paper* **2875**.