Working Paper SIET 2014 - ISSN 1973-3208

Efficiency in regional investments in R&D: implications for territorial growth.*

Angela S. Bergantino† Claudia Capozza‡ Angela De Carlo§

Abstract

In this paper we measure technical efficiency in Research and Development (R&D) of Italian regions with the aim of understanding whether the variation in accessibility and transport infrastructure endowment across regions might be the cause of efficiency disparities. We use a semi-parametric method where in the first step we estimate bootstrapped efficiency scores through DEA while in the second step, efficiency scores are explained - using alternative estimation methods - in a regression model accounting for transport infrastructure related variables as non-discretionary inputs. We show that well-developed transport infrastructure networks and services significantly improve R&D efficiency by facilitating connections and, thus, knowledge transfer.

Key words: R&D, technical efficiency, transport infrastructures, accessibility.
JEL: C14; O32; H54

1 Introduction

It is widely recognised that Research and Development (R&D) activity is crucial for technological progress and, hence, for the long-run economic growth of a country. Starting from Griliches (1958) on the US agriculture, many scholars have devoted attention to the effects of R&D activity on growth. Among the others, Archibald and Pereira

*The authors would like to thank Stefano Usai and participants at the 54^{th} ERSA Congress as well as participant at the 34^{th} AISRe Conference in Palermo for very helpful discussion. All remaining errors are ours.

†University of Bari "Aldo Moro", email: angelastefania.bergantino@uniba.it
‡University of Bari "Aldo Moro", email: claudia.capozza@uniba.it
§University of Bari "Aldo Moro", email: angela.decarlo@uniba.it

(2003) investigate the long-run effects of public and private R&D, underlying the large return rate of publicly funded R&D projects on private-sector performance in US. Further, Goel et al. (2008) carry out an extensive study exploring the link between economic growth and R&D funding in US. The main result is that economic growth seems to have a stronger association with federal R&D than with non-federal R&D. Focussing on OECD countries, Guellec and van Pottelsberghe de la Potterie (2001) provide empirical evidence on the positive long-term effects of R&D on productivity. The effects become greater for R&D-intensive countries and for countries where the share of universities, rather than government labs, is higher. Using a panel of industries across OECD countries, Griffith et al. (2004) empirically prove that R&D stimulates growth either directly through innovation or indirectly through technology transfer. Indeed, Johnes (2002) shows in a theoretical model that the long-run growth in US is driven by the implementation of ideas discovered throughout the world.

The efficient use of R&D resources is, in effect, a fundamental issue for growth. It follows that the analysis of the determinants of R&D efficiency would be useful in order to identify appropriate policy measures to improve the resources' allocation and design appropriate policy measures. While the application of parametric, i.e., econometric, techniques to the study of regional economic and innovative performance has become standard, the implementation of non-parametric methods is still quite rare (Foddi and Usai, 2013). Some past papers on efficiency in R&D activity use Data Envelopment Analysis (DEA) to assess performance. For instance, Chen et al. (2004) look at R&D efficiency in the computer industry considering a sample of taiwanese firms. Sharma and Thomas (2008) evaluates relative efficiency of R&D process across developed and developing countries. As well, Wang and Huang (2007) conduct a more detailed analysis by using a three-step DEA to explore the relative efficiency of R&D activities across either OECD or non-OECD countries. After assessing inter-country performance, they find out that the main drivers of efficiency are the enrolment rate of tertiary education, the PC density, and, to a greater extent, the English proficiency. Employing the same method, Hsu and Hsueh (2009) measure the relative efficiency of government-sponsored R&D projects in Taiwan. Efficiency is significantly influenced by the firm size and by the ratio of public subsidy on R&D. In addition to Wang and Huang (2007), Wang (2007) evaluates efficiency by stochastic frontier analysis, showing that the higher the PC density and the economic freedom of a country, the lower R&D inefficiency. Instead, the government share in R&D expenditure is found to play no role in affecting efficiency. More recently, Thomas et al. (2011) calculate R&D efficiency across US states plus the District of Columbia as the ratio of R&D outputs over R&D inputs. For most of the states, R&D efficiency has decreased over time. Aguado et al. (2013) use DEA to assess the efficiency of Spain and Italy at regional level in two time periods (pre- and post-crisis). Then, using a cluster analysis, they point out differences in regions' performance between the two period of analysis. Their results show that northern Italian regions are the best performers not only among Italian regions but also considering the Spanish regions, either in the pre-crisis or in the post-crisis period. Further, southern Italian regions have improved their performance in the post-crisis period compared to the pre-crisis, though they are still far from the level of efficiency achieved by northern regions. On the contrary, Spanish regions show worse performance in the post-crisis period, mainly caused by an increase in the use of inputs which has not led to an expected increase in the level of output.

This paper contributes to the existing research by estimating R&D efficiency of the Italian regions with the aim of understanding whether the variation in transport infrastructure endowment across regions might be the cause of efficiency disparities. Our hypothesis is that transport infrastructures play a role in improving R&D efficiency by facilitating connections and, thus, knowledge transfer among firms and universities which are the main producers of R&D outputs. To test our hypothesis, we apply a semi-parametric method. In the first step, we estimate bootstrapped technical efficiency scores by the means of DEA. In the second step, we define a regression model, to explain efficiency scores, including transport infrastructure variables as non-discretionary inputs. Our results - which are robust to the alternative estimation methods we use - claim that transport infrastructures and territorial accessibility seriously improve technical efficiency in R&D by facilitating connections and, thus, information sharing and knowledge transfer among R&D producers. This is also consistent with Jakko and McCann (2008) whose findings show that knowledge exchanges play an important role in the innovation process.

Overall, it emerges the additional effect of transport infrastructure investment on regional growth since well-developed transport infrastructures foster growth via two channel, one direct and the other indirect through R&D efficiency improvements.

The remainder of the paper unfolds as follows. In Section 2 we present the methodology, then in Section 3 we give a description of the data. In Section 4 we discuss the results and in Section 5 we draw conclusions. The robustness check is provided in the

Appendix.

2 Methodology

We apply a semi-parametric procedure to test the hypothesis that transport infrastructures affect technical efficiency in R&D activity of Italian regions.

In the first step, technical efficiency is estimated by the means of Data Envelopment Analysis (DEA), the non parametric approach introduced by Charnes et al. (1978). Technical efficiency refers to the "ability to avoid waste by producing as much output as input usage allows (output-oriented), or by using as little input as output production allows (input-oriented)".[1]

DEA has become the most popular technique for measuring efficiency. Actually, DEA is a very flexible tool. Firstly, it does not impose a functional form on the input-output relationship. Within the set of comparable Decision Making Units (DMUs), DEA identifies those that exhibit the best practice and constitute the efficient frontier. Deviations from the frontier are the result of inefficiency. Further, DEA manages multiple inputs and multiple outputs avoiding contrived output aggregation. This last point is relevant for the present study as the R&D production function is multidimensional, as we explain in the data section. The drawback of DEA is that generates estimates biased upwards since it overestimates the true efficiency level. To get rid of this downside, efficiency scores from the first step are corrected by the bootstrap procedure.

In the second step, technical efficiency scores are explained in a regression using non-discretionary inputs - transport infrastructure proxies - as independent variables.

We use different estimation methods, depending on the assumption we make on the distribution of the dependent variable.[2]

[1] See Lovel (1993) pg. 12.

[2] For greater details on the multi-step methodology for calculating efficiency determinants applied to different sectors see, among others, Bergantino et al., 2013; Bergantino and Musso, 2011; Bergantino and Porcelli, 2010 and 2011; Aubyn et al., 2009; Liu and Tone, 2008; Buzzo Margari, Erbetta, Petraglia and Piacenza, 2007; Worthington and Dollery, 2002.

2.1 First step

To estimate technical efficiency, the variable returns to scale envelopment problem is solved for each i^{th} DMU in the sample (Banker et al., 1984)[3]. We employ the standard input-oriented approach where technical efficiency is reached when inputs are minimized, keeping outputs fixed.

Consider the i^{th} DMU, with $i = 1, ..., N$, employing z inputs to produce q outputs. Then, θ is the solution of the following linear program:

$$\min_{\theta,\lambda} \theta \qquad subject\ to : \theta x_i \geq X\lambda; Y\lambda \geq y_i;\ e\lambda = 1;\ \lambda \geq 0 \tag{1}$$

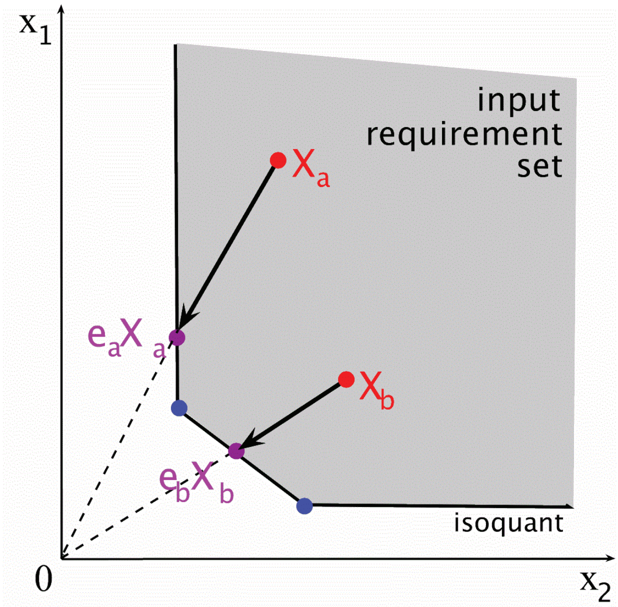
where:

- x_i is the $(z \times 1)$ input vector of the i^{th} DMU;
- y_i is the $(q \times 1)$ output vector of the i^{th} DMU;
- X is the $(z \times N)$ matrix of input vector in the comparison set;
- Y is the $(q \times N)$ matrix of output vector in the comparison set;
- λ is the $(N \times 1)$ intensity vector;
- e is the $(N \times 1)$ unity vector.

Technical efficiency scores correspond to Debreau (1951) - Farrell (1957) measure of efficiency and are bounded between unity and infinity. A DMU is technically when $\theta = 1$, whereas a DMU is not technically efficient when $\theta > 1$.

[3] Formerly, Charnes et al. (1978) developed the DEA model assuming constant return to scale. Afterward, Banker et al. (1984) relax this assumption by allowing for variable return to scale.

Figure 1. Graphical representation of the Debreu-Farrell index of technical efficiency.



Consider the graph in Figure 1 representing, for simplicity, the two-input/one-output production function. Both X_a and X_b are inefficient because they lie inside the input requirement set and their efficiency scores are the scalars $\frac{\overline{0e_aX_a}}{\overline{0X_a}}$ and $\frac{\overline{0e_bX_{jb}}}{\overline{0X_b}}$, which correspond to the minimum proportional reduction in both inputs necessary to hit the isoquant along the ray that connects the input combination to the origin. Efficiency scores have to be interpreted in terms of inefficiency with higher values indicating a lower efficiency.

In our sample we observe input-output data on Italian regions over the years. The linear program is solved by using a pooled approach where only one production frontier is estimated. In this way, each region is compared with all other regions and also with itself in different years (inter-regional and inter-temporal comparison).

As mentioned before, DEA tends to overestimate the true efficiency level, thus scores are corrected by the bootstrap procedure (2000 reps) developed by Simar and Wilson (1998, 2000).

In order to check the robustness of results, we also estimate efficiency using the output-oriented approach, where technical efficiency is reached when outputs are maximized, keeping inputs fixed (see the Appendix).

2.2 Second step

In the second step, bootstrapped DEA scores are explained in a regression model using non-discretionary inputs as independent variables. We specify the following equation:

$$\theta_{i,t} = \beta_0 + \beta_1 X_{i,t} + \delta_t + \varepsilon_{i,t} \tag{2}$$

where i identifies the region and t the time.

The dependent variable $\theta_{i,t}$ is the vector of efficiency scores. Further, $X_{i,t}$ is the set of environmental variables that might influence the efficiency and δ_t is the set of year dummies which capture the impact of macroeconomic factors equally affecting all regions. Finally, $\varepsilon_{i,t}$ is the idiosyncratic error term.

To estimate regression parameters we use three alternative strategies:

1. Following Simar and Wilson (2007), in the second step we implement a bootstrapped (2000 reps) left-truncated regression whose coefficients are obtained using the maximum likelihood estimator (MLE). Formerly, in the second step, the Tobit model was largely used. However, Simar and Wilson (2007), using Monte-Carlo experiments, show that Tobit provides poor results for either estimation or inference. They also prove that the double bootstrapped procedure performs very well, in terms of estimated confidence intervals and root mean square error.

2. In order to account for the potential non-linearity, we exploit the fact that the DEA efficiency scores have an ordinal meaning. Firstly, we compute the inverse ratio of efficiency scores, which are bounded between 0 and 1. In this case, the higher the score, the greater the efficiency. We define the dependent variable as a binary indicator taking value 1 if the DMU is above the mean of the inverse ratio of the efficiency scores, 0 otherwise. Then, we use the Probit estimator to obtain coefficients.

3. For robustness check and to further tackle the issue of non-linearity, we use the Bernoulli quasi-MLE as proposed by Papke and Woolridge (2008), where the dependent variable is defined as at point 2.

The output is presented and analysed in the following sections.

3 Data

3.1 Inputs and outputs

Different kinds of inputs have been adopted in the literature to characterised the R&D production function. Personnel and financial resources are, indeed, fundamental for developing a research project. In fact, for a successful R&D activity, the effort of manpower is a key factor. In addition, financial resources for purchasing technology and equipments are necessary to carry out R&D projects (see, among others, Pakes and Griliches, 1984; Guan and Wang, 2004). The outputs of the R&D activity are, mainly, patents and publications. The number of patents is a widely accepted indicator of R&D output. Basically, a patent indicates the presence of a non-negligible expectation on the product or the idea as to its ultimate utility and marketability (see Griliches, 1990); moreover, Acs et al. (2002) empirically prove that patents are a fairly reliable measure of innovative activity. Publications represent an indicator of academic productivity and are the way for sharing research results (Wang and Huang, 2007 and Sharma and Thomas, 2008).

In line with the prevailing approach, we define a two-inputs/two-outputs production function. We use, as inputs, scientific manpower and financial resources devoted to R&D. The required data on the number of *R&D personnel per 1,000 inhabitants* and total *R&D expenditures as percent of GDP* are collected from the Italian National Statistical Institute (ISTAT). As mentioned, outputs are *patents* and *publications*. Data on the number of patents granted by agencies in each region are collected from the Ufficio Italiano Brevetti e Marchi (UIBM). Data on the number of articles published are retrieved from Web of Knowledge, Science Citation Index (SCI), a well accepted source of data on publications.[4]

It is worth noting that the R&D production process requires time to be completed and to realize outputs. Therefore, we account of a time lag between inputs and outputs, defining three production functions according to the three different time lags (1-year, 2-years and 3-years, see Table 1). Data are collected for the sample of 20 Italian regions over the time span 1995 to 2012.

[4]For greater details on the definitions and characteristics of the inputs and outputs the reader is referred to: Bergantino, Capozza and De Carlo (2012).

Table 1. Descriptive statistics of R&D inputs-outputs

	Obs	*Mean*	*St. Dev*	*Min*	*Max*
INPUT					
R&D personnel per 1,000 inh (1995-2010)	320	2.557	1.385	0.08	6.19
R&D expenditures %GDP (1995-2010)	320	0.892	0.423	0.056	1.955
OUTPUT					
Number of patents granted (1996-2011)	320	571.11	1,020.50	0	7,564
Number of publications (1996-2011)	320	2,107.83	2,242.15	6	10,090
INPUT					
R&D personnel per 1,000 inh (1995-2010)	320	2.557	1.385	0.08	6.19
R&D expenditures %GDP (1995-2010)	320	0.892	0.423	0.056	1.955
OUTPUT					
Number of patents granted (1997-2012)	320	560.99	1,014.11	0	7,564
Number of publications (1997-2012)	320	2,211.01	2,341.56	6	10,547
INPUT					
R&D personnel per 1,000 inh (1995-2009)	300	2.510	1.363	0.08	6.19
R&D expenditures %GDP (1995-2009)	300	0.881	0.423	0.056	1.955
OUTPUT					
Number of patents granted (1998-2012)	300	545.98	952.175	0	6,500
Number of publications (1998-2012)	300	2,257.17	2,373.83	6	10,547

3.2 Environmental variables

We include in the second stage regression model a set of environmental variables, which might affect R&D efficiency, mainly related to the transport infrastructure endowment. Specifically, environmental variables related to transportation are *Transport infrastructure endowment* and *Air accessibility.* The former is the sum of extension of the railway and the road network[5] measured in km per 100 km^2, capturing the level of infrastructure endowment. The latter is logarithm of the total number of air pas-

[5] For robustness check, we also carry out regressions using the extension of the railway network and the extension of the road network as two separate variables.

sengers, capturing the intensity of airport activity and, thus, the accessibility of the territory.

We include some control variables. *Population over 65* is defined as the percentage of population over 65 years. It measures the population aging and, indirectly, the attitude to innovate of a region. We introduce four macro-area dummies, *North-West*, *North-East, Centre* and *South and Isles* to control for geographical-specific effects (the omitted cathegory is *North-West*). We also add a set of *year dummies* to account for the impact of common macroeconomic shocks. Descriptive statistics are reported in Table 2.

Table 2. Environmental variables

Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
Transport infrastructure endowment	200	62.328	16.634	25.510	107.013
Air accessibility	260	5,957.7	9,220.8	0	40,486
% Population over 65	320	19.581	4.062	12.218	68.594

4 Results

4.1 Efficiency scores in R&D

We compute bootstrapped efficiency scores which have to be interpreted in term of inefficiency (i.e. the higher the score, the lower the efficiency).

Figure 2 shows the pattern followed by technical efficiency in R&D across years. One might expect that the level of inefficiency in R&D decreases over time thanks to a learning-by-doing process. However, we find a dramatic increase of inefficiency between the 1999 and 2000. Actually, the euro currency official introduction in 1999 may have caused an imbalance affecting R&D performance in the years right after the introduction. From 2001 to 2005, the level of inefficiency steadily decreased while, henceforth, R&D inefficiency increased again.

Figure 3 shows the pattern followed by technical efficiency in R&D across regions. On the x-axis, regions are ordered from north to south. Within each macro-area, regions appear to be heterogeneous in the level of efficiency achieved (central regions are relative less heterogeneous). According to our estimates, the most efficient region is Lombardy whereas the less efficient is Aosta Valley, both are in the Northern Italy.

Among the regions belonging to Central Italy, Tuscany is the most efficient and Umbria the less efficient. Further, Apulia is the most efficient region in the South, whereas Abruzzi is the less efficient. Finally, Sicily appears to be more efficient than Sardinia.

Efficiency scores look very similar across production functions (i.e. whatever the time lag considered between inputs and outputs). This would suggests that the ability of regions to perform efficiently does not depend on the time required to complete the R&D production process. Indeed, the Spearman correlation among rankings is at least equal to 0.98 (see the top-left part of Table 4 reported in the Appendix).

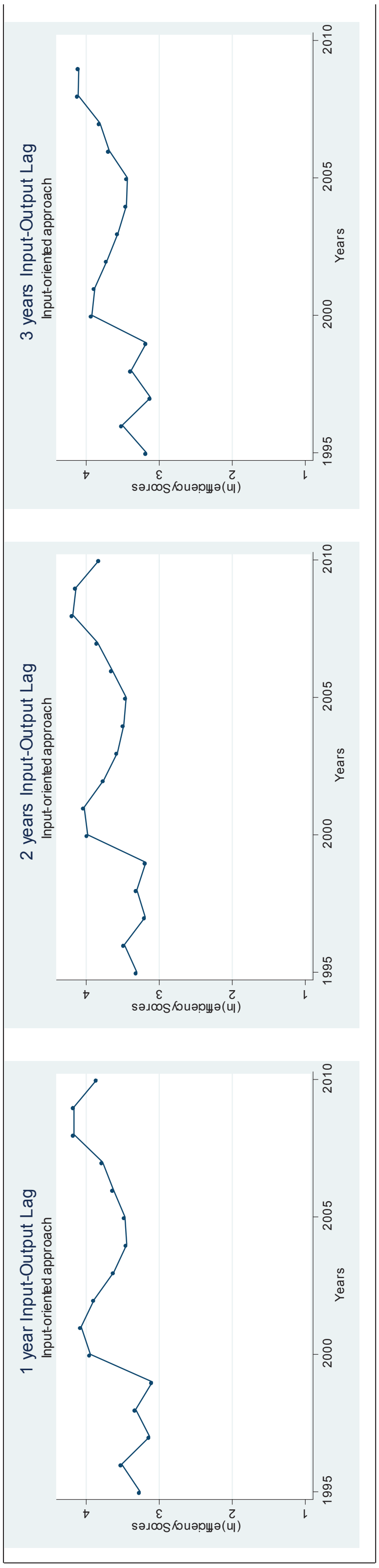


Figure 2. Bootstrapped DEA scores averaged across years.

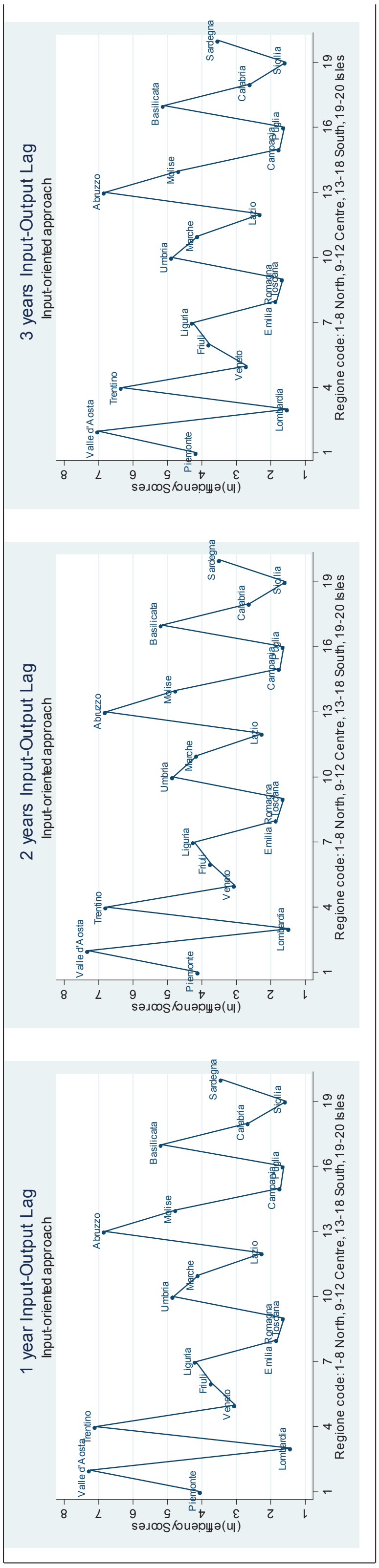


Figure 3. Bootstrapped DEA scores averaged across regions.

4.2 The impact of transportation infrastructures and accessibility

Environmental aspects cannot be directly included in the production function when using DEA, still they affect regional efficiency in R&D. In our study we make the hypothesis that the ability to perform R&D activity in an efficient way might be influenced by the variation in transport infrastructure endowment and territorial accessibility across regions. In the second stage regression we include some proxies for the railway, the road and the air transport infrastructures in order to verify whether they explain the efficiency in R&D. We further control for geographic and demographic characteristics.

As mentioned in Section 2.1, we implement three alternative estimation methods. We also consider three different time lag between inputs and outputs (1-year, 2-years, 3-years). In this way we also verify the sensitiveness of estimates to different specification of the production function.

In Table 3 we report results obtained using the left-truncated regression model. The dependent variable is the bootstrapped Debreau-Farrell efficiency scores, with higher values indicating lower efficiency.

Table 3

Estimation results of bootstrapped left-truncated regression using MLE.

Dependent variable: Bootstrapped Debreau-Farrell efficiency scores.

	Transport Infrastructures			*Accessibility*		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	1-y lag	2-y lag	3-y lag	1-y lag	2-y lag	3-y lag
Transport infrastr. endowment	-0.126***	-0.118***	-0.119***			
	(0.024)	(0.022)	(0.023)			
Air accessibility				-1.008***	-0.973***	-0.924***
				(0.087)	(0.086)	(0.093)
Pop over 65	0.865***	0.844***	0.869***	0.053	0.054	0.228***
	(0.155)	(0.148)	(0.153)	(0.079)	(0.083)	(0.063)
North-East	-0.654	-0.957	-1.181	-0.427	-0.594	-0.574
	(0.920)	(0.852)	(0.861)	(0.367)	(0.369)	(0.383)
Centre	-1.988**	-1.920**	-1.889**	-1.673***	-1.690***	-1.610***
	(0.856)	(0.783)	(0.770)	(0.379)	(0.373)	(0.421)
South and Isles	1.978**	1.890**	1.915***	0.273	0.161	0.700
	(0.825)	(0.750)	(0.743)	(0.456)	(0.471)	(0.447)
Observations	200	200	200	258	258	258

Time dummies are included but not reported. Robust standard errors in parentheses.

*** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Results show that the transport infrastructures and accessibility exert a negative effect on inefficiency. Specifically, the coefficienct of *Transport infrastructure endowment* and *Air accessibility* are always negative and highly significant across regressions. This underlyines the strong impact of the rail and road network extension and of the volume of air passenger traffic, which are found to reduce the inefficiency of R&D activity.

The variable *Population over 65*, a proxy for the innovation attitude, has the expected sign. It is positive and highly significant, meaning that the higher the proportion of population aged 65 and over, the higher the inefficiency of R&D activity.

In Table 4 we show coefficients obtained from the probit estimator and in Table 5 we illustrate results using the Bernoulli quasi-MLE. In either cases we report the marginal effects. It is important for the reader to bear in mind that the dependent variable is defined in a different way compared to the left-truncated regression. It is, indeed, a binary indicator which takes value equal to 1 if the efficiency score of a given DMU is above the mean of the inverse ratio of efficiency scores, 0 otherwise.

Table 4

Estimation results of Probit estimator.

Dependent variable: binary indicator taking value 1 if the efficiency score

is above the mean of the inverse ratio of efficiency scores, 0 otherwise.

	Transport Infrastructures			*Accessibility*		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	1-y lag	2-y lag	3-y lag	1-y lag	2-y lag	3-y lag
Transport infrastr. endowment	0.165**	0.084	0.084			
	(0.075)	(0.084)	(0.051)			
Air accessibility				0.774*	1.033***	0.599
				(0.400)	(0.351)	(0.572)
Pop over 65	-2.953***	-3.188***	-1.012***	-0.231	-0.354	-0.421
	(0.641)	(0.478)	(0.350)	(0.364)	(0.324)	(0.429)
North-East	-2.923	-3.126	-0.370	1.442	1.735	0.833
	(2.579)	(3.819)	(2.226)	(2.159)	(2.523)	(3.272)
Centre	4.773	7.667*	0.219	2.017	2.332	1.274
	(2.979)	(4.153)	(2.120)	(2.181)	(2.371)	(2.946)
South and Isles	-8.662***	-9.127**	-3.066	1.333	1.857	1.142
	(3.333)	(4.019)	(2.173)	(2.293)	(2.490)	(4.466)
Observations	200	200	200	258	258	258

Time dummies are included but not reported. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.10.

Table 5

Estimation results of Bernouli quasi-MLE.

Dependent variable: binary indicator taking value 1 if the efficiency score is above the mean of the inverse ratio of efficiency scores, 0 otherwise.

	Transport Infrastructures			*Accessibility*		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	1-y lag	2-y lag	3-y lag	1-y lag	2-y lag	3-y lag
Transport infrastr. endowment	0.012***	0.012***	0.013***			
	(0.003)	(0.003)	(0.003)			
Air accessibility				0.216***	0.224***	0.194***
				(0.032)	(0.034)	(0.031)
Pop over 65	-0.134***	-0.137***	-0.140***	-0.061***	-0.065***	-0.059***
	(0.018)	(0.019)	(0.020)	(0.021)	(0.022)	(0.015)
North-East	0.030	0.055	0.146	0.273**	0.308***	0.286***
	(0.138)	(0.138)	(0.138)	(0.112)	(0.113)	(0.067)
Centre	0.294***	0.296***	0.260**	0.512***	0.527***	0.292***
	(0.112)	(0.112)	(0.122)	(0.132)	(0.135)	(0.077)
South and Isles	-0.314***	-0.323***	-0.335***	0.147	0.159*	0.145*
	(0.096)	(0.095)	(0.098)	(0.095)	(0.094)	(0.084)
Observations	200	200	200	258	258	258

Time dummies are included but not reported. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.10.

Results are consistent with those obtained with the left-truncated regression. In fact, a higher level of transport infrastructure endowment and a greater territorial accessibility increase the probability of being deemed as efficient. Focusing on the probit estimation, the coefficient of *Transport infrastructure endowment* is positive, although significant only in regression with 1-years lag between inputs and outputs. Instead, it appears to be positive and always significant under the Bernoulli quasi-MLE. Moreover, coefficients of the variable *Air accessibility* are, in general, positive and significant across regressions. Finally, the variable *Population over 65* is negative and always significant with the Bernoulli quasi-MLE, while it is negative and significant with the probit estimator only for the set of regression reported in columns 1 to 3, concerning transportation infrastructure endowment.

5 Summary and conclusions

R&D activities are an essential element of economic growth. Understanding the sources of (in)efficiency in the regional production processes of R&D is, thus, crucial for developing targeted economic policies for innovation activities and R&D promotion. We provide evidence on Italian regions' performance over a fifteen years period and identify contextual factors which might play a role in determining it. In particular, in this paper we measure the efficiency of R&D activities across Italian regions with the aim of shedding light on the role of transport infrastructures and accessibility in promoting R&D efficiency. We carry out a two step analysis. (In)efficiency scores calculated in the first step are regressed in the second step of the analysis, through various estimation methods, on measures of transport infrastructure endowment and accessibility in order to identify the effects of transport related variables in explaining regional performances, after controlling for other contextual factors.

The analysis follows two steps. Firstly, we implement DEA, which provides evidence of a dualistic pattern in the regional R&D activities, with the most efficient territories located in the center-north areas of the country. Conversely, the lowest efficiency scores are shown by regions located in peripheral, southern areas, especially in the insular regions. Furthermore, the estimations carried out in the second part of the analysis, shows that R&D performance is explained also by a number of contextual factors, among which transport related indicators play a relevant and significant role. Results obtained appear to be consistent to the alternative estimation methods used. The results obtained with the three different estimation models used (bootstrapped left-truncated regression using maximum likelihood estimator, probit estimator and, finally, the Bernoulli quasi-MLE as proposed by Papke and Woolridge (2008)) are, in fact, coherent.

Our findings confirm our initial hypothesis on the positive and significant influence of the availability of an extensive transport infrastructure network and favourable accessibility conditions on regional efficiency in R&D. In particular, a greater extension of road and railway network and a greater volume of air passengers seem to stimulate the efficiency of R&D activity: transportation facilitates information sharing and knowledge transfer, allowing producers to learn from the best practice and, thus, to improve the production processes.

Policies disregarding the role of physical accessibility in promoting R&D and in favouring a convergence process among macroareas within the Italian territory, thus, might leave out an important element of evaluation. This might be of interest for policy-makers. Investing in developing transport infrastructures is, indeed, a well known way to stimulate economic growth. However, an additional effect emerges from our work. Well-developed transport infrastructures improve efficiency in R&D activity that, in turn, stimulates growth. In other words, transport infrastructures foster growth via two channel, one direct and the other indirect through R&D efficiency improvements. Recalling the historical North-South gap (Bergantino, 2013a and 2013b), these arguments take on an even greater importance since transport infrastructures are also the key means to reduce regional disparities and to promote convergence in R&D performance.

References

[1] Aubyn, M. S., Pina, A., Garcia, F., & Pais, J. (2009). Study on the efficiency and effectiveness of public spending on tertiary education. *European Economy Economic Papers,* n. 390, November 2009.

[2] Acs, Z. J., Anselin, L., Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research policy*, 31(7): 1069-1085.

[3] Aguado, R., Jabier, M., Cermelli, M.(2013). Patterns of efficiency in R&D expenditure. A compared analysis of spanish and italian regions. AISRe, *Proceedings of the XXXIV AISRE Conference*, Palermo.

[4] Archibald, R.B., Pereira, A.M. (2003). Effects of Public and Private R&D on Private-Sector Performance in the United States. *Public Finance Review*, 31(4): 429-451.

[5] Banker, R.D., Charnes, A., Cooper, W.W. (1984). Some Models for Estimating Technical and scale inefficiencies in DEA. *Management Science,* 32(9): 1613-1627.

[6] Bergantino, A.S., Musso E., Porcelli, F. (2013). Port management performance and contextual variables: Which relationship? Methodological and empirical issues. *Research in Transportation Business & Management*, 9: 81-91.

[7] Bergantino, A.S., Capozza, C. De Carlo, A. (2012). Efficiency in regional investments in R&D: implications for territorial growth. *Proceedings of the XIV SIET Conference*, Bari.

[8] Bergantino, A. S., Musso, E. (2011). The role of external factors versus managerial ability in determining seaports' relative efficiency: An input-by-input analysis through a multi-step approach on a panel of Southern European ports. *Maritime Economics & Logistics*, 13(2), 121–141.

[9] Bergantino, A.S., F. Porcelli (2011). A measure of Italian local government spending efficiency: The case of transport related expenditure. A preliminary analysis. In: E. Mercucci and E. Musso (eds.) *Sostenibilità, qualità e sicurezza nei sistemi di trasporto e logistica*. Milano, Italy: FrancoAngeli: 24–35.

[10] Bergantino, A.S., F. Porcelli (2010). Is efficiency in urban transport related expenditures capitalised in housing market quotations? The case of Italian municipalities. Identità, Qualità e Competitività Territoriale. *Proceedings of the XXXI AISRE Conference*, Aosta.

[11] Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6): 429-444.

[12] Chen, C., Chien, C., Lin, M., Wang, J. (2004). Using DEA to Evaluate R&D Performance of the Computers and Peripherals Firms in Taiwan. *International Journal of Business*, 9(4): 347-360.

[13] Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19(3): 273-292.

[14] Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120(3): 253-281.

[15] Goel, R.K., Payne, J.E., Ram, R. (2008). R&D expenditures and U.S. economic growth: A disaggregated approach. *Journal of Policy Modeling*, 30(2): 237-250.

[16] Foddi, M., Usai, S. (2013). Regional Knowledge Performance in Europe, *Growth and Change*, 44(2): 258–286.

[17] Griffith, R., Redding, S., Reenen, J.V. (2004). Mapping the two faces of R&D: Productivity growth in a panel of OECD industries. *Review of Economics and Statistics*, 86 (4): 883-895.

[18] Griliches, Z. (1958). Research costs and social returns: Hybrid corn and related innovations. *Journal of Political Economy,* 66(5): 419-437.

[19] Griliches, Z. (1990). Patent statistics as economic indicators: a survey? *Journal of Economic Literature*, 28(4): 1661-1707.

[20] Guan, J., Wang, J. (2004). Evaluation and interpretation of knowledge production efficiency. *Scientometrics*, 59(1) :131-155.

[21] Guellec, D., van Pottelsberghe de la Potterie, B. (2001). R&D and Productivity Growth: Panel Data Analysis of 16 OECD Countries. OECD Science, Technology and Industry Working Papers 2001/3, OECD Publishing.

[22] Hsu, F.M., Hsueh, C.C. (2009). Measuring relative efficiency of government-sponsored R&D projects: a three-stage approach. *Evaluation and program planning*, 32(2): 178-186.

[23] Jones, C. I. (2002). Sources of economic growth in a world of ideas. *American Economic Review*, 92(1): 220-239.

[24] Liu, J., & Tone, K. (2008). A multistage method to measure efficiency and its application to Japanese banking industry. *Socio-Economic Planning Sciences*, 42(2), 75–91.

[25] Lovell, C. K. (1993). Production frontiers and productive efficiency. In Fried, A. O., Lovell, A. K., and Schmidt, S. S., editors, *The Measurement of Productive Efficiency*, chapter 1, pages 3-67. Oxford University Press.

[26] Margari, B., Erbetta, F., Petraglia, C., Piacenza, M. (2007). Regulatory and environmental effects on public transit efficiency: A mixed DEA–SFA approach. *Journal of Regulatory Economics*, 32: 131–151.

[27] Pakes, A., Griliches, Z. (1984). Patents and R&D at the firm level: a first look. In: Griliches, Z. (Ed.), *R&D Patents and Productivity*. University of Chicago Press, Chicago.

[28] Papke, L.E., Wooldridge, J.M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145(1-2): 121-133.

[29] Sharma, S., Thomas, V.J. (2008). Inter-country R&D efficiency analysis: An application of data envelopment analysis. *Scientometrics*, 76(3): 483-501.

[30] Simar, L., Wilson, P.W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44 (1): 49-61.

[31] Simar, L., Wilson, P.W. (2000). A general methodology for bootstrapping in non-parametric frontier. *Journal of Applied Econometrics*, 27 (6): 779-802.

[32] Simar, L., Wilson, P.W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1): 31-64.

[33] Thomas, V.J., Sharma, S., Jain, S.K. (2011). Using patents and publications to assess R&D efficiency in the states of the USA. *World Patent Information*, 33(1): 4-10.

[34] Wang, E.C., Huang, W. (2007). Relative efficiency of R&D activities: A cross-country study accounting for environmental factors in the DEA approach. *Research Policy*, 36(2): 260–273.

[35] Wang, E.C. (2007). R&D efficiency and economic performance: A cross-country analysis using the stochastic frontier approach. *Journal of Policy Modeling*, 29 (2): 345-360.

[36] Worthington, A.C., and Dollery, B. (2002). Incorporating contextual information in public sector efficiency analysis: A comparative study of NSW local government. *Applied Economics*, 34, 453–464.

Appendix A

In order to check the robustness of our results, we compute R&D technical efficiency using the production functions specified in Section 3 under the output-oriented approach, where technical efficiency is reached when outputs are maximized, keeping inputs fixed.

Table 6 reports the matrix of Spearman correlation. The Spearman correlation between rankings of technical efficiency scores obtained using the input-oriented and the output-oriented approach is at least equal to 0.91. This would suggest that results are robust and are not influenced by the approach used to solve the linear program.

Table 6. Spearman correlation between rankings.

		Input-oriented			*Output-oriented*		
		1-y lag	*2-y lag*	*3-y lag*	*1-y lag*	*2-y lag*	*3-y lag*
Input-oriented	*1-y lag*	1					
	2-y lag	0.9933	1				
	3-y lag	0.9838	0.9903	1			
Output-oriented	*1-y lag*	0.9204	0.9145	0.9019	1		
	2-y lag	0.9152	0.9195	0.9080	0.9925	1	
	3-y lag	0.9053	0.9099	0.9164	0.9825	0.9893	1

A further robustness check consists of running the same set of regressions introducing in the model the variables *Road Network* and *Railways Network* separately, in place of the variable *Transport infrastructure endowment*, which is given by the sum of *Road Network* and *Railways Network*. Overall, results appear to be consistent with the main estimates (see Table 7 to 9).

Table 7

Estimation results of bootstrapped left-truncated regression using MLE.

Dependent variable: Bootstrapped Debreau-Farrell efficiency scores.

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	1-y lag	2-y lag	3-y lag	1-y lag	2-y lag	3-y lag
Railway Network	-0.091	-0.092	-1.491***			
	(0.734)	(0.692)	(0.204)			
Road Network				-0.132***	-0.124***	-0.125***
				(0.028)	(0.025)	(0.026)
Pop over 65	0.645***	0.635***	0.925***	0.849***	0.829***	0.855***
	(0.222)	(0.211)	(0.135)	(0.165)	(0.155)	(0.159)
North-East	-0.311	-0.635	-1.918**	-0.607	-0.916	-1.145
	(1.345)	(1.294)	(0.803)	(0.964)	(0.894)	(0.891)
Centre	-3.979***	-3.823***	-4.142***	-1.930**	-1.853**	-1.817**
	(1.480)	(1.442)	(0.763)	(0.891)	(0.824)	(0.821)
South and Isles	-0.536	-0.550	-1.019	2.110**	2.023**	2.056***
	(1.369)	(1.266)	(0.706)	(0.901)	(0.796)	(0.763)
Observations	260	260	240	200	200	200

Time dummies are included but not reported. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.10.

Table 8

Estimation results of Probit estimator.

Dependent variable: binary indicator taking value 1 if the efficiency score is above the mean of the inverse ratio of efficiency scores, 0 otherwise.

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	1-y lag	2-y lag	3-y lag	1-y lag	2-y lag	3-y lag
Railway Network	0.390***	0.270***	1.702***			
	(0.060)	(0.065)	(0.519)			
Road Network				0.165**	0.037	0.072
				(0.080)	(0.087)	(0.047)
Pop over 65	-3.553***	-2.492***	-1.685***	-2.880***	-3.312***	-1.019***
	(0.457)	(0.587)	(0.337)	(0.607)	(0.635)	(0.295)
North-East	0.803	1.635	0.702	-3.243	-2.622	-0.671
	(4.244)	(3.675)	(2.289)	(2.817)	(3.213)	(1.916)
Centre	5.747	5.746	2.425	1.781	8.959**	1.207
	(4.879)	(3.789)	(2.175)	(3.013)	(4.382)	(2.054)
South and Isles	-1.204	-4.635	-1.419	-8.726***	-7.705**	-2.683
	(4.353)	(4.181)	(1.993)	(3.187)	(3.253)	(1.651)
Observations	260	260	260	200	200	200

Time dummies are included but not reported. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.10.

Table 9

Estimation results of Bernouli quasi-MLE

Dependent variable: binary indicator taking value 1 if the efficiency score is above the mean of the inverse ratio of efficiency scores, 0 otherwise.

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	1-y lag	2-y lag	3-y lag	1-y lag	2-y lag	3-y lag
Railway Network	0.011***	0.011***	0.102***			
	(0.002)	(0.002)	(0.010)			
Road Network				0.012***	0.012***	0.013***
				(0.003)	(0.003)	(0.003)
Pop over 65	-0.117***	-0.120***	-0.083***	-0.131***	-0.134***	-0.137***
	(0.016)	(0.016)	(0.014)	(0.018)	(0.018)	(0.020)
North-East	-0.029	-0.011	0.117***	0.013	0.037	0.125
	(0.108)	(0.110)	(0.043)	(0.134)	(0.134)	(0.135)
Centre	0.343***	0.347***	0.184***	0.278**	0.280**	0.243**
	(0.090)	(0.090)	(0.030)	(0.112)	(0.112)	(0.120)
South and Isles	-0.182**	-0.176**	-0.006	-0.322***	-0.330***	-0.345***
	(0.084)	(0.084)	(0.071)	(0.095)	(0.095)	(0.097)
Observations	260	260	240	200	200	200

Time dummies are included but not reported. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.10.