XXXIV CONFERENZA ITALIANA DI SCIENZE REGIONALI

THE ROLE OF TRANSPORT INFRASTRUCTURES IN DETERMINING TECHNICAL EFFICIENCY IN R&D ACTIVITY OF ITALIAN REGIONS. A DOUBLE-BOOTSTRAPPED DEA PROCEDURE.

Angela S. BERGANTINO[1], Claudia CAPOZZA[2], Angela DE CARLO[3]

SOMMARIO

We measure technical efficiency in R&D of Italian regions with the aim of understanding whether the variation in 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. In the second step, efficiency scores are explained in a bootstrapped truncated regression using transport infrastructure variables as non-discretionary inputs. We find that well-developed transport infrastructures seriously improve R&D efficiency by facilitating connections and, thus, knowledge transfer.

[1] Università degli Studi di Bari, Largo Abazia Santa Scolastica 53, 70124,Bari, email: angelastefania.bergantino@uniba.it.

[2] Università degli Studi di Bari, Largo Abazia Santa Scolastica 53, 70124,Bari, email: claudia.capozza@uniba.it.

[3] Università degli Studi di Bari, Largo Abazia Santa Scolastica 53, 70124,Bari, email: angela.decarlo@uniba.it.

1 Introduzione

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, 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. Archibald and Pereira (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) provide 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. (2001) empirically prove that R&D stimulates growth either directly through innovation or indirectly through technology transfer.

The efficient use of R&D resources is, indeed, a fundamental issue for growth. It follows that the analysis of the determinants of R&D efficiency is needful in order to identify appropriate policy measures to improve the resources' allocation. This paper contributes to the existing research by estimating R&D efficiency of the Italian regions - there is a lack of evidence on Italy - 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, basically, the main producers of R&D outputs.

Some past papers on efficiency of R&D activity use Data Envelopment Analysis (DEA) to simply assess R&D performance. 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. A more detailed analysis is by Wang and Huang (2007). They use a three-step DEA to evaluate the relative efficiency of R&D activities across either OECD or non-OECD countries. After assessing inter-country performance, they find that the main drivers of efficiency are the enrolment rate of tertiary education, the PC density, and, to a greater extent, the English proficiency. Using the same method, Hsu and Hsueh (2009) measure 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) employs a stochastic frontier method to evaluate efficiency, showing that the higher the PC density and the economic freedom of a country, the lower R&D inefficiency. Instead, the governament share in R&D expenditure is found to have 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. They find out that, for most of the states, R&D efficiency has decreased over time.

To test our hypothesis, we apply the semi-parametric method by Simar and Wilson (2007). In the first step, we estimate bootstrapped technical efficiency scores by the means of DEA. In the second step, efficiency scores are explained in a bootstrapped truncated regression using transport infrastructure proxies as independent variables. Our results claim that transport infrastructures seriously improve technical efficiency in R&D by facilitating connections and, thus, information sharing and knowledge transfer among R&D producers. This highlight a multiplicative 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 the semi-parametric method by Simar and Wilson (2007) 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, or by using as little input as output production allows".[4] 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 is relevant for the present study as the innovation production function is certainly multidimensional. The drawback of DEA is that generates estimates biased upwards

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

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 truncated regression using non-discretionary inputs - transport infrastructure proxies - as independent variables.

2.1 First step

To estimate technical efficiency, a variable returns to scale envelopment problem is solved for each i^{th} DMU in the sample (Banker et al., 1984).[5] 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. Under the input-oriented approach, θ is the solution of the following linear program:

$$\min \theta \quad \textit{subject to}: \ \theta x_i - X\lambda \geq 0; \ \ Y\lambda - y \geq 0; \ \ e\lambda = 1; \ \ \lambda \geq 0 \qquad (1)$$

where:

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

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 another year.
Technical efficiency scores correspond to Debreau (1951) - Farrell (1957) measure of efficiency and are bounded between unity and infinity. A DMU is technically when θ=1, whereas a DMU is not technically efficient when θ>1. Scores have to be interpreted in terms of inefficiency with higher values indicating a greater inefficiency. As mentioned before, DEA tends to overestimate the true efficiency level, therefore scores are corrected by the bootstrap procedure developed by Simar and Wilson (2007).

[5] Formerly, Charnes, Cooper and Rhodes (1978) developed the DEA model assuming constant return to scale. Afterward, Banker et al (1984) relax this assumption and introduce variable return to scale.

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

2.2 Second step

In the second step, bootstrapped DEA scores are explained in truncated regression. We specify the following model:

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

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 macroeconomic factors equally affecting all regions. Finally, $\varepsilon_{i,t}$ is the idiosyncratic error term.

Coefficient are obtained using the maximum likelihood estimator.

3 Data

3.1 Inputs and outputs

Input-output data characterizing the R&D production function are chosen in line with the prevailing approach in the literature (see Pakes and Griliches, 1984) and are collected for the sample of 20 Italian regions.

The inputs to innovation production activity are manpower and physical resources. We use data on the number of R&D personnel per 1,000 inhabitants and total R&D expenditures as percent of GDP, which are from the Italian National Statistical Institute (ISTAT).

The outputs of the innovation process are patents and publications. The number of patents is a widely recognized indicator of R&D output. 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). In addition, publishing articles is the way for delivering research outcomes.

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).

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).

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 granted patents (1996-2011)	320	571.11	1020.50	0	7564
Number of publications (1996-2011)	320	2107.83	2242.15	6	10090
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 granted patents (1997-2012)	320	560.99	1014.11	0	7564
Number of publications (1997-2012)	320	2211.01	2341.56	6	10547
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 granted patents (1998-2012)	300	545.98	952.175	0	6500
Number of publications (1998-2012)	300	2257.17	2373.83	6	10547

3.2 Environmental variables

In the second stage regression we include 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: Railway Network, Road Network and Air Transport Pax. Railway Network and Road Network are two indices for the extension of the railway and the road network, respectively. Either are measured in km per 100 km^2. Air Transport Pax is the total number of air passengers, capturing the intensity of airport activity.

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, Centre, South and Isles to control for geographical-specific effects (the omitted category is North). We also add a set of year dummies capturing the impact of common macroeconomic shocks.
Finally, descriptive statistics are reported in Table 2.

Table 2 - Descriptive statistics of Environmental variables

Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
Population over 65	320	19.581	4.062	12.218	68.594
Railway Network	260	7.198	30.295	1.773	492.934
Road Network	200	57.038	15.332	22.97	97.79
Air Transport Pax	260	5957.7	9220.8	0	40486

4 Results

4.1 Efficiency scors is 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 1 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 away following the introduction. After 2001, inefficiency in R&D activities steadily decreased till 2005. Henceforth, R&D inefficiency increased again.
Moreover, Figure 2 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, either are in the northern Italy. Among the regions belonging to the central Italy, Tuscany is the most efficient and Umbria the less efficient. Further, Apulia is the most efficient region in the south area, 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 in the appendix).

Figure 1 - Average bias corrected DEA scores across years

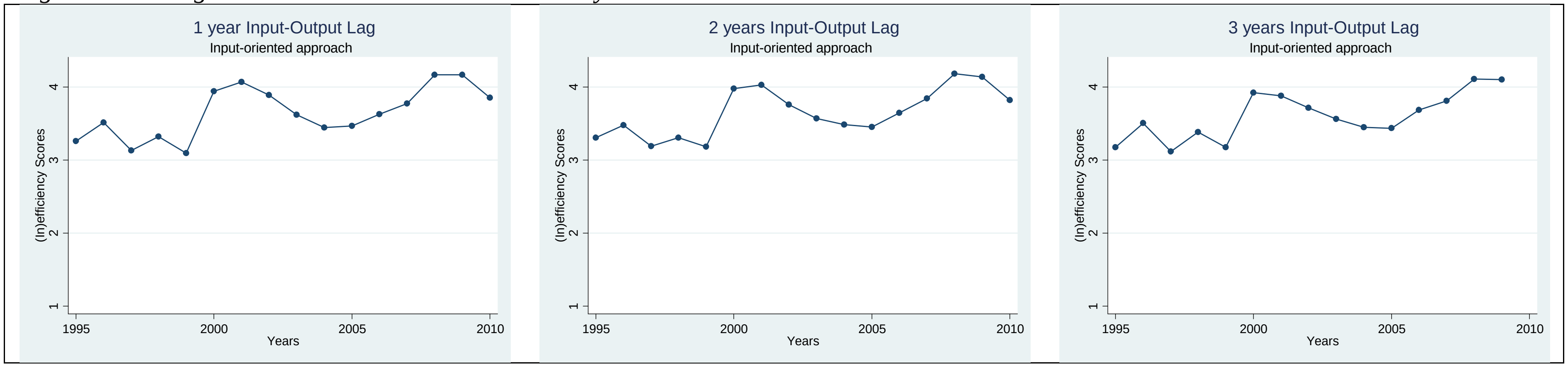
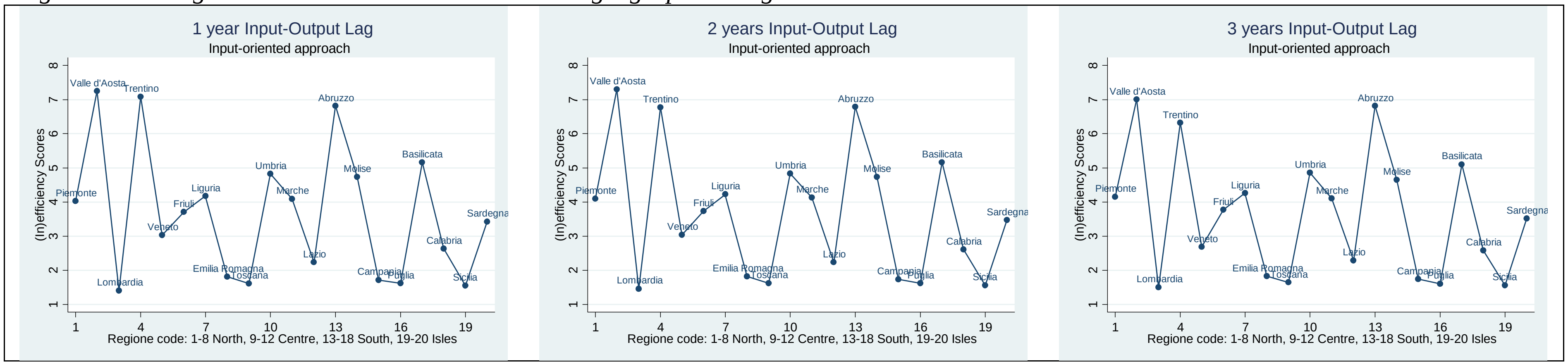


Figure 2 - Average bias corrected DEA scores across geographical regions



4.2 The impact of transportation infrastructures

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 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.

Coefficient estimates are reported in Table 3. The dependent variable is the set of efficiency scores obtained from the production functions specified. It is important to bear in mind that they have to be interpreted in term on inefficiency. As stated before, we consider three different time lag between inputs and outputs (1-year, 2-years, 3-years). In this way we also check the robustness of estimates.

Our results show that the transport infrastructures exert a negative effect on inefficiency. Specifically, the coefficient of Railway Network is negative although significant only in regression with 3-years lag between inputs and outputs. This would indicate that the railway network improves R&D performance when assuming that inputs in the production process requires three years to produce outputs. In other words, railway network positively affects medium-run performance in R&D rather than short-run performance. Moreover, coefficients of the variables Road Network and Air Transport Pax are always negative and highly significant across regressions, thus underlying a strong impact of the road network extension and the volume of air passenger traffic which reduce the inefficiency of R&D activity.

Finally, 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.

Table 3 - The impact of transportation infrastructures on efficiency in R&D

Variable	*1-y lag*	*2-y lag*	*3-y lag*	*1-y lag*	*2-y lag*	*3-y lag*	*1-y lag*	*2-y lag*	*3-y lag*
Railway Network	-0.072	-0.076	-1.520***						
	(0.747)	(0.725)	(0.230)						
Road Network				-0.140***	-0.131***	-0.133***			
				(0.027)	(0.026)	(0.028)			
log(Air Trans Pax)							-0.931***	-0.899***	-0.865***
							(0.083)	(0.080)	(0.090)
Pop over 65	0.554***	0.565***	0.837***	0.736***	0.738***	0.769***	0.108***	0.112***	0.198***
	(0.202)	(0.195)	(0.125)	(0.136)	(0.132)	(0.132)	(0.040)	(0.040)	(0.053)
Centre	-3.644***	-3.390***	-3.007***	-1.323*	-1.171*	-1.056	-1.393***	-1.330***	-1.231***
	(1.165)	(1.066)	(0.602)	(0.737)	(0.687)	(0.660)	(0.298)	(0.294)	(0.318)
South	0.303	0.394	0.708	3.178***	3.168***	3.293***	1.014***	0.995***	1.235***
	(0.888)	(0.821)	(0.437)	(0.737)	(0.730)	(0.714)	(0.277)	(0.274)	(0.286)
Isles	-5.085**	-4.425**	-3.601***	-1.310	-0.859	-0.668	-4.092***	-4.079***	-3.245***
	(2.032)	(1.861)	(0.800)	(1.002)	(0.852)	(0.846)	(0.652)	(0.661)	(0.578)
Observations	260	260	240	200	200	200	258	258	240

Time dummies are always included but not reported. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

5 Summary and conclusions

In this paper we measured the efficiency of R&D activities across Italian regions with the aim of shedding light on the role of transport infrastructures in promoting R&D efficiency. We employed the double-bootstrapped semi-parametric method (Simar and Wilson, 2007).

Results confirm our initial hypothesis on the influence of well-developed transport infrastructures on regional efficiency in R&D. In particular, a greater extension of road network and a greater volume of air passengers seem to reduce the inefficiency, while a more widespread extension of railway network appears to reduce the medium-run production inefficiency of R&D activity. This might happen because transportation facilitates the information sharing and knowledge transfer, allowing producers to learn from the best practice and, thus, to improve the production process.

This finding might be of interest for policymakers. Investing for developing transport infrastructures is, indeed, a well known way to stimulate economic growth. However, a multiplicative effect comes out 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, these arguments take on a greater importance since transport infrastructures are also the key means to reduce regional disparities.

References

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.

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.

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.

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.

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

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

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.

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.

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

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

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.

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.

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

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.

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.

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

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.

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.
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.
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.

Appendix - Robustness check

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

Table 4 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 4 - 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