

THE PRODUCTIVE EFFICIENCY OF SCIENCE AND TECHNOLOGY WORLDWIDE: A FRONTIER ANALYSIS

ABSTRACT

We are interested in how codified knowledge is produced around the globe (which inputs are used to produce scientific articles and patented inventions) and the efficiency of the process (how do the best performers produce more with the same inputs or produce the same with less inputs). Using a Data Envelopment Analysis (DEA) efficiency frontier approach, we aim to determine which countries are more efficient at producing codified knowledge. We proxy knowledge production by publications and patents, obtained through human (researchers) and non-human (R&D expenditure) resources. We built a 15-year database with more than 800 observations of these and other variables. Our findings enable us to distinguish efficiency by country, geographical region, and income area. We run four different specifications and correlate the results with partial productivity indexes seeking consistency. Under constant returns to scale, the most traditional producers of knowledge are not fully efficient. Instead, small countries with limited resources appear to be efficient. When we add environmental conditions, both sets of countries are efficient producers of knowledge outputs. High-income regions, on the one hand, and East Asia, North America, and Europe and Central Asia, on the other, are the most efficient regions at producing knowledge.

KEYWORDS

Knowledge, science, productive efficiency; Data Envelopment Analysis, research

HOW TO CITE

Ferro G., Romero C. A. (2021) 'The Productive Efficiency of Science and Technology Worldwide: A Frontier Analysis', *Journal on Efficiency and Responsibility in Education and Science*, vol. 14, no. 4, pp. 217-230. <http://dx.doi.org/10.7160/eriesj.2021.140402>

Gustavo Ferro¹✉
Carlos A. Romero²

¹Universidad del CEMA (UCEMA) and CONICET, Argentina

²CONICET-Universidad de Buenos Aires. Instituto Interdisciplinario de Economía Política (IIEP-BAIRES), Argentina

✉ gferro05@yahoo.com.ar

Article history

Received

February 8, 2021

Received in revised form

May 17, 2021

Accepted

September 9, 2021

Available on-line

December 20, 2021

Highlights

- We build and examine a proxy of a production frontier for national innovation systems using DEA methodology.
- We can detect with this method which countries achieve more output of its resources.
- Our concern is to concentrate on the relationship between outputs and inputs. The method permits to distinguish between countries which are the best achievers.
- Policy design could use this to help respond "where we are and where we are intended to be", and as a catalyst for change (deepening on why? and how?).

INTRODUCTION

Public, private, and third sectors in every country devote resources through institutions (mainly universities, laboratories and research centers) to research activity. Produced knowledge can ultimately be applied to technology and yield developments. The knowledge production function is a multi-input and multi-output activity in which quality and environmental issues matter. The inputs consist of both human, non-human (scientific instruments, materials or financial resources), and intangible resources (accumulated knowledge, formal or informal networks of scientists and practitioners). The outputs can also be tangible (embedded in publications, patents, conference presentations, databases, etc.) and intangible (tacit knowledge, common

practices, etc.). Important activity areas for economic growth include the creation of private and public knowledge, human capital building, and knowledge infrastructure production (Abramo, D'Angelo and Di Costa, 2015).

The concept of National Innovation Systems holds that innovation results from complex interactions between actors who generate, diffuse, and apply knowledge. This concept was applied in several contexts: national, technological, and sectoral. Innovation Systems can be understood as a set of relationships between private firms, public authorities, research organizations, and other bodies, ideally structured and co-ordinated in some way so that linkages between actors stimulate collective learning, continuous innovation, and entrepreneurial activity (Njøs and

Jakobsen, 2018). Thus, a country's innovative achievements (research and development, R&D) depend on how those actors link up with each other as components of a collective creation system. Their contribution can be divided into knowledge production and application (Choi and Zo, 2019).

For R&D policy makers this kinds of study are key elements to allocate resources, to establish priorities, to set goals, to evaluate past initiatives, to compare with similar countries with best achievements, to extract lessons, to change course, to avoid misleading objectives or instruments, and to trigger delving deeper in details about the "why" and "how" of the observed performance. They are useful to project possible future paths, to identify commonalities and differences. The type of assessment we offer is an instrument to evaluate and monitoring the impact of national policies. The impacts can be measure according its absolute performance (production, inputs, evolution across time, etc.) and with respect to its relative performance (productivity and efficiency, across time, places, and productive units). The first approach is relatively simple and useful; the second one is superior since relates results to resources, compare best practices with standard (or even substandard) ones, identify costly ways of achieving the results, and challenge the researcher to define, conceptualize, and measure the phenomenon under study. The cost of more complete and deep assessments is some loss of simplicity, since simple ratios are easier to understand than more complex studies of efficiency and productivity. However, certain frontier techniques, as those here presented offer a reasonable trade-off between deepness of the analysis and loss of simplicity. An efficiency assessment helps identify typologies of knowledge generation in different countries and provides policy and managerial implications for each case, as well as detecting best-practices to identify benchmarks and discover weaknesses (Ferro and D'Elia, 2020). Thus, it is possible to evaluate whether some policy or line of incentives to research had some impact, such as budgetary funds allocated to certain goals or rewards and disincentives to certain practices. Most previous studies that examined the efficiency of National Innovation Systems comprise two stages: knowledge production and knowledge application (Lundvall, 2007; Marxt and Brunner 2013). The former is defined as generating knowledge outcomes by using research-related inputs. The latter stage consists of transforming the outcomes of the previous process into inputs for economic results. In developing countries, an additional component is knowledge absorption coming from developed countries (Choi and Zo, 2019). We focus our study on knowledge production.

The differences in efficiency (achievements given resources) between countries lying on the frontier and countries lying below the frontier, also show the possibilities of improvement. To catch-up the best performers at macro level, the following step is to explain those differences delving into the details of each national innovation system (organizations, institutions, incentives, policies to identify and retain talent, tools to induce certain research lines, etcetera). This deserves a deeper analysis at a micro level which is beyond our goals, nevertheless measuring a phenomenon can be a good first step for diagnosis, which in turn can be useful to develop policy guidelines as a second step. To apply policies aimed to generate knowledge, planning is essential, and planning mean a diagnosis which begins in data

collection, continues with transformation of data in information, follows with the analysis of information, and once some conclusions are drawn it is time for policy implementation, conducing to goals. Policy design could use this kind of analysis as a compass to help respond "where we are and where we are intended to be", and as a catalyst for change (if we are here, and we can be there, "why?" and "how?").

To guarantee the effectiveness of this instrument, it is necessary to analyze in detail certain features of the information and modelling. With respect to data, we build a homogeneous international database, paying attention to its coherence, quality, extension, comparability, and time span. And concerning models, pros and cons of different approaches are balanced. Simplicity is important for applying and for interpreting results within a interdisciplinary atmosphere such as national innovation systems. Data Envelopment Analysis (DEA), for instance, is particularly friendly to apply, and it is relatively straightforward to understand its results.

Based on data from R&D statistics on human and non-human resources as input variables, and publications and patents as output variables, we use an efficiency frontier method to determine what the countries obtain from them. It is not possible to draw a coherent conclusion from the patterns arising from the partial productivity indicators, such as papers per researcher or patents per unit of financial resources, since they yield, on occasion, contradictory rankings (Battese and Coelli, 1988). That is essentially why we have analyzed global indexes, such as frontier efficiency scores. Efficiency is understood as the relationship between inputs and outputs, paying due attention to environmental factors (contextual and mainly uncontrollable) and quality conditions (distinctive, and depending on volition and deliberate resource allocation decisions).

This study contributes to the literature by measuring the knowledge production of publications and patents, obtained from human and non-human resources at the country (national) level. This paper, builds and examines a proxy of a production frontier for national innovation systems using DEA methodology, contributing to a more complete evaluation of knowledge production than focusing either in comparing outputs, inputs, or simple ratios output/input (average productivity indexes). The method we apply allows us to determine with which combination of human and non-human resources (inputs), a certain set of outputs are produced (codified knowledge in the form of papers and patents over certain quality threshold). Moreover, we can detect with this method which countries achieve more output from its resources. Our concern is to concentrate on the observed relationship between outputs and inputs. The method permits to distinguish between units (countries in this case) which are the best achievers in comparison with other units which counting with the same resources achieve smaller results. Countries with the best results are in the efficiency frontier and countries with achievements below the frontier, by definition, produce only a fraction of the outputs the countries which lies in the frontier do. Quality and environmental issues should be addressed to differentiate the productivity of inputs and to compare the final outcomes. Quality results from deliberate actions to improve inputs and outputs, while environmental conditions are non-discretionary inputs.

The above discussion allows us to formulate our research questions: What is a meaningful proxy for the production function of a national R&D system? What are the inputs, the outputs, and environmental conditions when considering a frontier where the unit of analysis is a country? What are the levels of productivity measured through conventional indicators, such as average productivity or average costs of such a system? What levels of productive efficiency can be estimated for country systems? What are the drivers that explain the differences in efficiency between knowledge producing countries? How can this study be enriched in the future?

After motivating the discussion, defining the objective, setting the perimeter, establishing the possibilities and limitations of the study and formulating research questions in this introduction, The second section, presents materials and method: it starts with the database and its analysis, describes the method to explain the efficiency differences, their drivers, presents the models to be run and the empirical results. The third section, of discussion, evaluates studies relevant for the specification of outputs, inputs, quality and contextual variables, as well as further explains the utility of the empirical results for policies at the stage of diagnosis and performance monitoring, and the fourth section concludes.

MATERIALS AND METHOD

Material: the Database

Our database covers the years 2003-2017. The original database is an unbalanced panel for 206 countries and territories across 16 years, which was shortened by balancing to 60 countries across 15 years. The outputs considered are published papers included in the Scimago (2020) database and patents (applications and grants) from the World Intellectual Property Organization (WIPO) database.

A publication, according to Scimago (2020), is a document published and indexed in a specific year, which satisfies certain scientific protocols (blinded refereeing and indexing for a database which admits publications over a certain quality threshold). According to our source (WIPO, 2020), a patent is 'An exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something, or offers a new technical solution to a problem. To get a patent, technical information about the invention must be disclosed to the public in a patent application [...] patent protection means that the invention cannot be commercially made, used, distributed, imported or sold by others without the patent owner's consent. Patents are territorial rights. In general, the exclusive rights are only applicable in the country or region in which a patent has been filed and granted, in accordance with the law of that country or region. The protection is granted for a limited period, generally 20 years from the filing date of the application.'

Knowledge production uses human resources, physical productive capital, research funds, knowledge embedded in human resources, machinery and equipment, public

involvement in R&D, and agglomeration effects. They are devoted to produce codified outputs (such as publications and patents), yielding non-codified outputs embedded in researchers, students and the community (extension or service activities). The inputs considered are non-human resources, proxied by R&D Expenditures (in constant 2005 Purchase Price Parity or PPP USD) and human resources proxied by Full-Time Equivalent (FTE) Researchers are both from UNESCO's compilation of statistics. In this study, each country is the decision-making unit of DEA, which uses human and non-human resources to produce scientific publications and patents. Quality is addressed through citations (for publications) and patent grants (for patents). The former to proxy the publication impact and the latter to proxy the patent's commercial application. According to WIPO (2020), 'Licensing a patent simply means that the patent owner grants permission to another individual/organization to make, use, sell etc. his/her patented invention. This takes place according to agreed terms and conditions, for a defined purpose, in a defined territory, and for an agreed period. Unlike selling or transferring a patent to another party, the licensor continues to have property rights over the patented invention.' Citation lag is not necessarily the same as publication lag, and patent grant lags are not necessarily the same as patent claim lags. Concerning publications, in an aggregated database, on the one hand, it is not possible to attribute each citation to its product. On the other hand, lags imply losing observations in a not very extensive series. Thus, we first correlate contemporaneous data of inputs to outputs and do the same for two-year lags. The correlations remain quite similar. Consequently, we opted to run the estimates in contemporaneous data for inputs and outputs.

Another decision concerns impact itself. Since we cannot trace citations to publications, and patent grants to patent claims, we run a Cobb-Douglas¹ cost function in logarithms with and without correction for quality. In the first version, we estimated costs with respect to outputs, the relative price of outputs and a time trend. Marginal costs of publications and patents proved to be highly significant and positive, as was expected. When we added two variables for quality consideration (citations for publications and patent grants for patents), the variable for citations was not significant and the inclusion of patent grants was highly significant, but invalidated the significance for patents. These results suggest that the marginal cost is positive for publications and for patents with commercial applications, but not necessarily so for citations (which depend on several factors) and for patent claims (not all patents have commercial potential). The Appendix presents the cost function estimates. Our outputs, corrected by quality (in the sense of a costly attribute added to the product to improve it) are, thus, publications and patent grants.

For environmental variables, we use data on GDP² as a proxy of the material resources each economy is able to produce, the population as a proxy of the potential dimension of the country's human resources, per capita GDP as the quotient of the former two, and the percentage of GDP³ devoted to R&D

1 We also ran a Translog version to address quadratic and crossed effects, but such terms were nonsignificant.

2 We use GDP at constant prices without correcting for PPP values, which is a better proxy of the size of each economy than the PPP value.

3 In this case, we use PPP values since it is a cost.

expenditure as a proxy of the relevance of the activity in the country under study.

We built two groupings for the countries based on World Bank criteria to classify countries into high-, low-, lower-middle, and upper-middle income, on the one hand, and employed geography, on the other. We also distinguish those countries

where English is the official language de jure or de facto. In each case, we studied correlations between these variables and outputs. We also present partial productivity ratios, which later compare with efficiency scores. Table 1 summarizes the variables: definition, source, units, and their role in the estimates.

Variables	Unit	Type	Concept
ydocum	Number of Documents published during a specific year	Output	Publications
qcitati	Total number of Citations contained in the documents published during the source year.	Quality	Quotes
ypateap	Number of Total Patent Applications (direct and Patent Cooperation Treaty PCT national phase entries). Yearly flow of new applications	Output	Patent applications
qpategr	Number of Total Patent Grants (direct and Patent Cooperation Treaty PCT national phase entries). Yearly flow of grants.	Quality	Patent grants
xfteres	Function-total Researchers Full Time Equivalent (FTE).	Input	Labor input
xgerdpp	Gross Domestic Expenditures in Research and Development (GERD) in Purchasing Power Parities (PPPs) at constant 2005 values.	Input	Non-labor input
GDPpc	GDP in 2010 constant million USD/Inhabitants	Environmental	GDP / Inhabitants
Partial productivity ratios			
doc_res	ydocum/xfteres (human resources partial productivity)		
doc_gerd	Ydocum/xgerdpp (non-human resources partial productivity)		
qpategr_res	Qpategr/xfteres (human resources partial productivity)		
qpategr_gerd	Qpategr/xgerdpp (non-human resources partial productivity)		

Table 1: Variable Definitions (sources: Scimago (2020) for publications, WIPO (2020) for patents, UNESCO (2020) for human and non-human resources and its elaborations, The World Bank (2020) for GDPpc and Population, World Bank Atlas for country classifications by income, 2019 GNI per capita, and geographical according WB. Wikipedia for “English as official language”)

Table 2 presents the descriptive statistics of the database. Our final database is comprised of 803 annual observations of 60 countries over 15 years (2003-2017) with an average of 13.4 observations per country (minimum 10, maximum 15).

Variable	Observations	Mean	Total	Std Dev	Minimum	Maximum
ydocum	803	41,732	33,511,125	94,378	92	669,588
qcitati	803	725,181	582,320,089	1,970,961	780	18,095,159
ypateap	803	38,871	31,213,618	124,776	1	1,306,080
qpategr	803	16,914	13,581,586	51,013	1	352,567
xfteres	803	113,256	90,944,836	262,111	142	1,740,442
xgerdpp	803	20,623,366	16,560,562,919	60,325,656	3,419	426,030,391
GDP pc	803	25,310	20,323,615	22,466	456	111,968
doc_res	803	0.533	428.335	0.414	0.028	3.437
doc_gerd	803	0.005	4.228	0.006	0.000	0.075
pategr_res	803	0.081	65.424	0.105	0.000	0.777
pategr_gerd	803	0.001	0.490	0.001	0.000	0.008

Table 2: Descriptive statistics of the variables (source: own elaboration on sources of Table 1)

The correlation between our quality variables and their output is high and positive: 0.8510 (qcitati and ydocum) and 0.9306 (qpategr and ypateap). Also, the correlations between inputs and outputs are high and positive: 0.9015 (ydocum and xfteres), 0.9630 (ydocum and xgerdpp), 0.8637 (ypateap and xfteres) and 0.8782 (ypateap and xgerdpp). GDPpc has a 0.1797 correlation with ydocum and 0.1151 with ypateap.

English as an official de jure or de facto language in our sample

exhibits a positive but low correlation with outputs: 0.3333 (d_eng and ydocum) and 0.1048 (d_eng and ypateap). As for impact, the English language is found to be more important: 0.3951 (d_eng and qcitati) but not so for patents: 0.1128 (d_eng and qpategr).

Thus, with respect to the correlation analysis and the correlation between geographical regions and outputs, two regions exhibit positive values: 0.5891 (d_nac and ydocum), 0.2900 (d_nac

and ypateap), 0.2250 (d_eas and ydocum), and 0.4951 (d_eas and ypateap). The last values also denote a relative pattern of specialization in publications in North America and patents in East Asia.

METHODOLOGY

The standard methods to estimate efficiency are the non-parametric Data Envelopment Analysis (DEA) and the parametric Stochastic Frontier Analysis (SFA). The respective advantages and disadvantages are well discussed in the literature. Parametric methods assume a specific functional form for the frontier, departing from some behavioral objectives (such as profit maximization or cost minimization); non-parametric methods do not given their greater flexibility to consider different decision-making unit behaviors. These methods can be deterministic or stochastic. Deterministic methods attribute the distance of a given decision-making unit from its frontier to inefficiency; stochastic methods assume that some of them can be attributed to randomness (“noise”) and try to separate both components from the error term.

We use DEA to determine which decision-making units (in this case, countries) form an envelope surface of the sample to which they belong. The efficient decision-making units are those yielding on the frontier, while those below it are deemed inefficient since they produce less than their “peers” in the frontier with the same inputs (or produce the same with fewer inputs). A score is attributed to each decision-making unit based on how much it differs from the most efficient “peers”. For each country, DEA solves an optimization problem seeking the optimal weights for the inputs and for the outputs, which maximize the ratio between the weighted sum of output divided by the weighted sum of inputs.

The efficiency measure (score) for any decision-making unit is obtained as the maximum ratio of weighted outputs to weighted inputs, subject to similar ratios for every decision-making unit being less or equal to unity. Following the Charnes, Cooper and Rhodes (1978) notation for n decision-making units ($j = 1, \dots, n$), s outputs and m inputs the problem is:⁴

$$\text{Max } \theta = \frac{\sum_{r=1}^s u_r y_{r0}}{\sum_{i=1}^m v_i x_{i0}} \quad (1)$$

Subject to:

$$\frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1; \quad j = 1, \dots, n$$

$$u_r, v_i \geq 0;$$

$$r = 1, \dots, s;$$

$$i = 1, \dots, m$$

Where θ is the maximum ratio for decision-making unit 0, y_r are the outputs (for $r = 1, \dots, s$), x_i are the inputs (for $i = 1, \dots, m$), outputs and inputs being positive. The $u_r, v_i \geq 0$ are the weights yielded by the solution of the problem, which relies on all decision-making units used as a reference set.

The efficiency of one decision-making unit of the sample is to be rated relative to the others, distinguishing it by “ θ ” in the functional (but preserving its original subscript in the constraints). This decision-making unit has the most favorable weighting allowed by the constraints (Charnes, Cooper and Rhodes, 1978). An optimal $\theta^* = \max \theta$ will always satisfy $0 \leq \theta^* \leq 1$ with optimal solution values $u_r^*, v_i^* > 0$ (Banker, Charnes and Cooper, 1984). Efficiency is defined as the score $E_r = y_r/Y_r$, where y_r is the actual output r produced by the decision-making unit under analysis, and Y_r is the maximum feasible output obtained by the same input set, where $0 \leq E_r \leq 1$. The weights are objectively determined to obtain a dimensionless E_r scalar measure of efficiency from observational data, subject only to the constraints established in (1). Therefore, no other set of common weights will give a more favorable rating relative to the reference set (Charnes, Cooper and Rhodes, 1978).

In the so-called CCR Model (Charnes, Cooper and Rhodes, 1978), the set of efficient decision-making units forms an envelope relative to observational data from all $j = 1, \dots, n$ decision-making units. The envelopment can differ because of the scale assumption with respect to the phenomena under study. They are customarily constant returns to scale (CRS) or variable returns to scale (VRS), encompassing both increasing and decreasing returns to scale. CRS implies that output will change by the same proportion as inputs do (at the same scale), while VRS assumption reflects that production will change in different proportion as input do (different scales, increasing, decreasing and as a special case, constant, that is, at the same scale). A priori can be reasons to assume certain returns to scale in each investigation, while the practice indicates that comparing results from different assumptions can be useful in another circumstances (Cooper, Seiford and Tone, 2007). Productivity and technical efficiency are equivalent only when the technology exhibits CRS and the model produces an “overall efficiency” rating. The BCC Model (Banker, Charnes and Cooper, 1984) applies to technologies with VRS, which helps compare the maximum average productivity attained at the most productive scale size with the average productivity at the actual scale of production to measure scale efficiency. Under VRS, it is possible to separate pure technical inefficiency from scale inefficiency. In this case, we only compare decision-making units of a similar scale. Units deemed inefficient under the CRS assumption can be efficient once we allow for VRS. As DEA is mainly a deterministic method, no accommodations have been made for bias resulting from environmental heterogeneity, external shocks, measurement errors, and omitted variables (Rhaïem, 2017).

⁴ This version of the problem is known as output-oriented. It can be also formulated as input-oriented, or not oriented at all. For the sake of brevity, we omit the last specifications. The way it is written is one of the three possibilities of presenting the problem. It intuitively presents the problem as maximizing the ratio between a weighted sum of outputs divided a weighted sum of inputs, where the key elements are the weights, different among efficient and non efficient decision-making units.

Models

We run two versions (CRS and VRS) for the two models, the first considering non environmental variables (CORE) and the second including the latter (ENV). We use GDPpc as a synthetic variable for development: GDP gives an indication of the economic size of the country, and the population proxies the country's potential in terms of human resources. GDPpc normalizes the first variable on the second. Moreover, GDP and population are partly exogenous (accounting for the endowments of resources) and partly endogenous (accounting for public policies, institutions and human

capital accumulation). From the point of view of knowledge generation, we consider it a non-discretionary input and have treated it as such in the estimates. It is reasonable to assume it thus over a short period, such as that of the sample. It is also plausible that the accumulation of physical and human capital plus technology change transforms the variable into an endogenous factor in knowledge production in the long run. The models we present in Table 3 are the result of several alternatives that we have tested and the reasoning underlying each one is more easily understood when examining the results below.

	Outputs + Inputs (CORE CRS Model)	Outputs + Inputs (CORE VRS Model)	Outputs + Inputs + Environmentals (ENV CRS Model)	Outputs + Inputs + Environmentals (ENV VRS Model)
Outputs				
Ydocum	X	X	X	X
Qpategr	X	X	X	X
Inputs				
Xfteres	X	X	X	X
Xgerdpp	X	X	X	X
Environmental				
GDPpc			X	X

Table 3: DEA Model Specifications (source: own elaboration)

RESULTS

Table 4 presents the results of the sample as a whole by year, while Table 5 shows the figures aggregating countries by region (according to the geographical and income classifications of

the World Bank). Table 4 shows the results for an unbalanced panel, which includes a minimum of 46 (2017) and a maximum of 59 countries (2011), and for four Models: CORE CRS, CORE VRS, ENV CRS, and ENV VRS.

Year	#	Core Model							Env Model						
		CRS			VRS			Scale Effic.	CRS			VRS			Scale Effic.
		# Eff	AVG	Std Dev	# Eff	AVG	Std Dev		# Eff	AVG	Std Dev	# Eff	AVG	Std Dev	
2003	49	5	0.564	0.269	15	0.753	0.249	1.337	16	0.750	0.238	19	0.804	0.211	1.072
2004	48	5	0.540	0.266	15	0.771	0.232	1.427	14	0.736	0.241	18	0.814	0.204	1.106
2005	50	4	0.472	0.251	13	0.738	0.223	1.562	14	0.713	0.252	18	0.794	0.203	1.115
2006	52	4	0.505	0.228	14	0.713	0.233	1.411	12	0.690	0.235	17	0.758	0.214	1.098
2007	56	5	0.499	0.233	17	0.716	0.243	1.434	15	0.685	0.242	20	0.755	0.225	1.102
2008	57	7	0.521	0.237	19	0.743	0.234	1.426	14	0.714	0.233	20	0.777	0.208	1.088
2009	58	6	0.477	0.242	16	0.693	0.254	1.455	13	0.662	0.246	18	0.728	0.229	1.101
2010	54	4	0.496	0.241	13	0.692	0.250	1.394	12	0.676	0.238	15	0.727	0.225	1.076
2011	59	4	0.487	0.235	17	0.707	0.259	1.450	12	0.689	0.246	18	0.742	0.239	1.078
2012	57	3	0.440	0.236	14	0.699	0.260	1.589	11	0.691	0.250	20	0.767	0.249	1.111
2013	56	4	0.402	0.233	11	0.660	0.275	1.642	10	0.653	0.258	17	0.745	0.260	1.141
2014	56	2	0.413	0.213	12	0.688	0.278	1.667	12	0.685	0.264	20	0.766	0.260	1.119
2015	53	2	0.389	0.183	13	0.715	0.263	1.838	14	0.716	0.257	22	0.788	0.246	1.100
2016	52	4	0.404	0.221	14	0.726	0.268	1.794	17	0.727	0.267	20	0.775	0.255	1.065
2017	46	3	0.400	0.222	14	0.726	0.270	1.814	11	0.725	0.247	17	0.778	0.245	1.073
AVG	54	4	0.467	0.234	14	0.716	0.253	1.549	13	0.701	0.248	19	0.768	0.232	1.096

Table 4: Core and ENV Models, by year (CRS and VRS versions)(source: own elaboration)

All models show a slight decrease in the average efficiency scores, considering the extreme points of the series. The CORE CRS model, as was expected, registers few efficient countries: 4 on average⁵ (or 7.73 percent). The mean efficiency scores exhibit a decreasing pattern, averaging 0.467. The standard deviation, the dispersion and the minimum value show a quite stable behavior in the period. This model is influenced by the performance of a reduced set of small efficient countries. When we run the model omitting these countries to check for robustness, the average efficiency improves. Nonetheless, the objective is not to eliminate observations but to find a model that solves said aspect. Thus, we extend the models following two strategies: first, considering the potential influence of size (scale effect) in research and development activity (CORE VRS), and second considering the level of economic development as an environmental condition (ENV CRS and ENV VRS models).

The Core VRS Model exhibits 14 efficient countries on average⁶ (or 27.10 percent). The mean efficiency scores average 0.716, starting above, and the standard deviation is almost the same as in the CRS version. By using a VRS model, we assume that scale is important. The estimate seems to indicate that. In effect, the scale efficiency resulting from the quotient of the average efficiency scores under VRS on the same average under CRS is surprisingly high (1.550 on average and has an increasing pattern, reaching more than 1.80 in some years). These results should be taken with caution: technology enables us to share a significant degree of knowledge and experiences internationally, and a great deal of intellectual production comes from transboundary collaboration. Unfortunately, our database does not allow us to address that issue. Up to this stage, the results suggest strong economies of scale. Nonetheless, in that regard, some environmental conditions, correlated with the size of the country, can explain those results.

When accounting for environmental conditions, the scale efficiency presents moderate values. What hypothesis would explain the absence of economies of scale in knowledge generation? In a globalized world, researchers and resources for R&D can collaborate and even move at a relatively low cost. Mobility is a great incentive to concentrate human and non-human resources in developed countries. Yet, there are some reasons to limit the international mobility of researchers, which implies a sort of “Home Bias”⁷ in residence, research interests, and regional specialization. Differences in the cost of living, social status and growing communication possibilities are forces favoring low-scale knowledge production.

Thus, considering CRS ENV, the inclusion of GDPpc as an environmental variable yields results that are quantitatively similar to the Core VRS model. On average, 13 countries are efficient (24.72 percent of the sample), the mean efficiency score is 0.701, with similar dispersion and minimum to Core VRS model.

When analyzing VRS ENV, 19 out of 54 countries are efficient on average (or 34.88 percent), and standard deviation holds more or less the same average value as in the other three models. The minimum efficiency score is higher than in the rest of the models. However, the scale efficiency now shows a more stable and moderate pattern, averaging 1.096. We cannot confirm that variable returns to scale exist, moreover when considering aggregated information. The inclusion of a plausible environmental variable reduces the efficiency scores gap between CRS and VRS versions, in addition to the number of efficient countries. The difference in the average efficiency scores between both ENV model versions drops.

Table 5 shows that the production frontier’s average efficiency scores yield higher efficiency levels in regions and countries with a tradition in academic production. Thus, East Asia (EAS), Europe and Central Asia (ECS) and North America (NAC) exhibit comparatively high levels of productive efficiency, while Latin America (LCN), the Middle East and North Africa (MEA) and Sub-Saharan African Countries (SSA) show comparatively low levels of productive efficiency, the first case being more consistent across the models. In contrast, the figures improve in the second case if GDPpc is included as an environmental variable. However, the performance of the models differs.

The CORE CRS model is the most pressing one because it does not consider the size of the scientific community or the development level of the countries. In this model, the performance of the countries with the most noteworthy academic tradition (such as the United States, the United Kingdom and Germany) does not behave efficiently and the scores of these countries are not very far from the average performance of the sample. China, the rising star in publishing and patents claims and licensing – because of its impressive growth in the years of the sample – behaves poorly in efficiency terms. The reason for these results can be attributed to the efficient behavior of small countries with comparatively scarce resources but comparatively abundant production. In turn, this admits several possible explanations. For instance, the international financing of research and development has a great impact in countries with a low R&D budget, a critical weight of small high-productivity research groups, data errors, among others.

Thus, one interpretation is that the CORE CRS Model yields a higher inclusion error (i.e., considering a country efficient when it is not) and exclusion (considering a country inefficient when it is not). The adoption of the hypothesis of VRS or the consideration of GDPpc as an environmental variable reduces the standard deviation of the most traditional R&D producers, which implies a lower exclusion error. Nevertheless, it does not correct the inclusion error. The CORE VRS, ENV CRS and ENV VRS models (whose average level of efficiency and the number of efficient countries is quite similar, but not coincident in each model’s set of efficient countries, or in the same order)

5 The countries vary each year. Those which appear frequently are the Netherlands, Luxemburg, Japan, Republic of Korea.

6 The countries vary each year. Those which appear more frequently, besides those which appear as being efficiency under CRS are: United States, United Kingdom, Italy, Germany, China in the more recent years, as well of Chile, Russia, Poland, and Spain in some years.

7 It is a established hypothesis in finance. Portfolios in practice exhibit a lower internationalization in their composition than the levels suggested by theoretical portfolio models (Lewis, 1999).

show more foreseeable results for regions and countries with a greater tradition in R&D, particularly for these selected

countries (see Table 5). The same is valid for China. Efficiency scores also increase monotonically with income level areas.

Region*/Selected country	Mean 2003-2017 CORE CRS	Mean 2003-2017 CORE VRS	Mean 2003-2017 ENV CRS	Mean 2003-2017 ENV VRS
EAS	0.554	0.896	0.959	0.970
ECS	0.494	0.852	0.845	0.868
LCN	0.235	0.435	0.501	0.533
MEA	0.534	0.628	0.775	0.861
NAC	0.441	0.980	0.910	0.915
SSF	0.418	0.659	0.751	0.803
High Income	0.536	0.934	0.919	0.935
Low Income	0.031	0.051	0.040	0.067
Lower Middle Income	0.276	0.437	0.627	0.753
Upper Middle Income	0.260	0.675	0.801	0.813
Sample Average	0.467	0.716	0.701	0.768
China	0.235	0.809	1.000	1.000
Germany	0.598	0.945	0.921	0.945
USA	0.448	1.000	1.000	1.000
UK	0.437	1.000	1.000	1.000

* Each country score within the region is weighted by its GDP for summing up.

Table 5: Core and ENV models by region (CRS and VRS versions) (source: own Elaboration)

DISCUSSION

Conceptual discussion

This study contributes to the literature by measuring the knowledge production of publications and patents, obtained from human and non-human resources at the country (national) level. As stated, a national innovation system can produce knowledge, apply knowledge, and in certain context absorb knowledge coming from developed countries. Choi and Zo (2019) concentrates in the three stages and with cross sectional information estimates efficiency in each of the stages. Their contribution is important for its reach; however, they are considering only one year of information, and a subset of countries. Cross sectional databases have its merits however they do not allow to examine evolution over time. More mundane, outliers or error in data are hardly detected. We develop from Choi and Zo (2019). Our purpose is first to expand its time frame, second to extend geographically the reach (both things imply taking several instead of one picture, to detect evolution patterns), and third to focus the reach of the analysis to knowledge generation for deepening on this issue. In Choi and Zo (2019), papers and patents are the knowledge outputs. In our contribution we add quality dimensions (papers over certain internal quality assessment, and patents not only registered but licensed, that is, with current application potential) and environmental or contextual issues to address for differences in standard of living among countries. Also, we delve into the role of scale in knowledge production, that is, whether size of national innovation system is conducive (or not) to high efficiency.

As research activity is a production process, we analyze it from the perspective of the microeconomic theory of production. Performance should be evaluated with respect to aims and stated in measurable terms that represent the desired outcome (Abramo and D'Angelo, 2014). Aksnes et al. (2017) investigate methodological problems in measuring research productivity at the national level. Problems arise with the comparability of input and output statistics, as well as with the different National Innovation Systems themselves. Reports on resources and outcomes are often presented separately by different reporters, instead of combining them in measurements of productivity or efficiency.

Note that the productive units have different aims and institutional contexts. Fair comparisons of output-input efficiency must be managed by means of environmental variables, considering the uncontrollable inputs of knowledge production. The aims and rewards of private sector producers are supposedly different to public sector institutions' objectives and incentives. Nevertheless, public research organizations and universities face increasing demands to extend their teaching and research activities through the licensing of inventions, spin-off creation, research collaborations, and partnerships with private companies, etc. (Abramo, D'Angelo and Di Costa, 2015).

A time lag exists between the start of research and when the results ultimately appear as published articles. A standard in scientific production is that It can take about two years, on average, to achieve publication (Leydesdorff and Wagner, 2009). Lags also exist between the claim and award of patents, between the time patents are awarded and licensed with commercial aims, and from the time a publication is

edited and cited. Nevertheless, the precise attribution of lags in empirical work is feasible if the data make it possible to attribute the citations to the publications, or the award to the patent, which is difficult when the level of aggregation of the data increases.

We base our selection of the outcomes that synthesize the countries' knowledge codified production on the character of public and private new knowledge. It is unlikely that the output of basic research is patentable since it is difficult or almost impossible to exclude third parties from using it (free access). Thus, private incentives to produce that knowledge are low, while public or third sector bodies interested in the production of public goods and positive externalities will be biased to produce them. This kind of knowledge can also surge and disseminate through accumulation and casual events (serendipity). New knowledge, a rival of established knowledge, and excludable, is deemed private or exclusive knowledge. Indeed, a patent confers exclusive rights on its owner. A publication is free knowledge (you pay for its supporting platform) (Elías, Ferro and García, 2019).

The results of research in the business sector are rarely published in scientific journals. If they are, it is often because the research was conducted together with actors from other sectors. Thus, we can expect publications to proxy well (mainly) publicly aimed and financed research, and that patents proxy well (mainly) privately aimed and financed research. Specific empirical evidence supports this assertion (Schmid and Fajebe, 2019). However, the boundary between knowledge outcomes is blurring. The recent worldwide moves towards incentivizing "impact" within the research funding system pose a growing challenge to academic research practices that produce both scientific and social impact, which differ in terms of their epistemic qualities and value as academic work. A growing number of countries are adopting incentives for research "relevance" to fund research. In this "culture of accountability", universities are becoming increasingly more "entrepreneurial" at raising funds (Bandola-Gill, 2019).

Rhaim (2017) distinguishes the following outputs for knowledge production: refereed articles, books, book chapters, refereed conferences, professional publications, dissertations, other deliverables, and quality indicators to build a hierarchy among them. Refereed articles ensure sufficient homogeneity, which is a necessary condition to estimate a production function. To further homogenize outcomes, they can be corrected by quality, proxied by its impact, and measured by its citations. In addition to publications, several authors have included industry grants, third-party contracts, technology transfer to businesses, the number of patents, the number of expert reports, and teaching activities in the case of universities. It seems, thus, that publications and patents are good candidates for proxy knowledge production.

Research production is usually understood as the number of publications produced by a given unit and analyzed in such a way as to consider impact, efficiency, or quality components in research production to determine its drivers (Thelwall and Fairclough, 2017). Publications are external, objective means to measure outcomes, unlike internal and more subjective

metrics. In effect, in assessing research group output, two approaches are generally employed: peer reviews and bibliometric methods. Each method has its pros and cons. However, publications appear to summarize the consensus of committees. The peer-review method is based on perceptions of well-informed experts about different quality dimensions of research production. It is subjective and depends on the committee's composition. Groot and García-Valderrama (2006) conclude that publications well proxy the scientific output as scientists understand it. A scientific paper has become the reference unit of bibliometric research because of the rules of the game of such outputs (blind refereeing, the transparency of the process, the need to present evidence and more recently, anti-plagiarism software), and the possibility of standardizing and auditing those rules through indexing in different databases (Abramo and D'Angelo, 2014). While the core bibliometric indicator is the number of published papers by an institution or country within a period, nevertheless, it should take the resources that are correlated with intellectual production into consideration (Bornmann et al., 2020). Partial productivity indexes allocates the papers to full-time equivalent researchers or non-human resources used in the productive process.

Bibliometric methods also have their limitations since they are restricted to written outputs. The evaluation results are influenced by the measurement methods applied, and the assessment of which publications are acceptable depends on publication and citation practices in different disciplines (Groot and García-Valderrama, 2006). Uncodified new knowledge is hard to measure, and when codified, some conventions should be included to identify and measure its various forms. A publication is an established measure of codified knowledge in most sciences, while it is not the standard output for the arts, humanities, and part of the social sciences. Compilations of science indicators measuring national research performance in the international context describe the development of a field of science or a production unit with the help of bibliometric means. Nevertheless, not all researchers produce bibliometric outputs, and these are heavily concentrated in a minority of researchers (Glänzel, 2003).

In specific environments, some scholars challenge publication in international journals (see Sahoo et al. (2015) for a discussion). They consider them biased against developing countries' problems, issues, data, and researchers. The prejudice extends to language since the international scientific community works almost exclusively in English, with little room for other native languages. The discussion concerns the legitimacy of using indexed international publishing as the main output of the scientific community. Nevertheless, the growing tendency to collaborate through international co-authorship is an important counterargument against the isolation argument ("a native writes for natives") (Sahoo et al., 2015).

The use of publications as an output measure of research has the disadvantage of being retrospective. Likewise, the use of research grants in efficiency studies has its proponents and opponents. Research grants are defined as additional resources in an institution's budget to promote research and the work of young scientists. Research grants proxy the market value of financed research and can also be considered a proxy of its

quality. Nevertheless, the funds are spent not only on research assistance but also on other facilities, which are inputs for production. Thus, some authors consider research grants to be a measure of research output (because they are assigned to “virtuous” researchers or valuable proposals), but they mostly consider them an input because they support research projects (Gralka, Wohlrabe and Bornmann, 2019).

Given the diversity of research activities, several experts have proposed including indicators, such as patenting and spin-offs, in performance measurements (Gralka, Wohlrabe and Bornmann, 2019). Patents are often published to be presented to scientists (Abramo, D’Angelo and Di Costa, 2015). However, knowledge used for practical applications can differ from academic knowledge in terms of the type of problems being dealt with, incentives, timelines, accountability standards, procedures, and institutions (Bandola-Hill, 2019).

With respect to inputs of knowledge production, Rhaïem (2017) distinguishes human resources (academic staff and non-academic staff), physical productive capital (building spaces, laboratories, equipment, libraries, computers, etc.), research funds (capital and operating), knowledge embedded in human resources, machinery and equipment, public involvement in R&D, and agglomeration effects, which refer to a regional concentration of research effects and the links that could appear in an entrepreneurial environment. The identification of production factors other than labor, and the assessment of their value and share by field is often hard to quantify. Sometimes, unobserved effort could yield no results, and conversely, given that ‘serendipity and luck may yield huge returns at little cost’ (De Fraja and Valbonesi, 2012: 322). At the same time, some factors can be independent of the capacities of the staff for the units under examination, because of returns to scale, returns to scope or available capital resources (Abramo, D’Angelo and Di Costa, 2015).

Quality can and ideally should be assessed either in outputs or inputs for fair and meaningful comparisons through different coefficients or dummy variables (See Abramo, D’Angelo and Murgia, 2016, for seniority and qualification of researchers and its impact, for example). In bibliometric studies, research productivity distinguishes a researcher’s publications from their impact (Abramo and D’Angelo, 2014). Although the databases differ in scope, the volume of data and coverage policies, the countries’ outputs (papers) and impacts (citations) are extremely correlated (Archambault et al., 2009). Co-authorship, references, and citations are qualitative elements that denote the impact of the contribution. Processes are important when considering whether an output is deemed a scientific product: sources, procedures, and techniques should be reliable and documented and the reproducibility of results should be guaranteed. The screening procedures for publication ensures that scientific character in particular. Scientific knowledge production has historically developed within an international community of scholars for whom values such as objectivity and de-contextualization are epistemic virtues, and prerequisites for communication (Bandola-Gill, 2019).

There is a growing demand to expand the societal impacts of academia although the logic behind the achievement of scientific impact and societal relevance is different: aims

(reputation or “scientific quality” against the practical application or “societal relevance”) and processes diverge, as well as reward mechanisms. The reshaping of incentives (funding in particular) eventually reconciles the divergence of skills and interests. It also facilitates specialization among those that continue to produce knowledge (seeking peer recognition) and those who redirect their efforts towards communication and fundraising (D’Este et al., 2018).

The value of research is measured by its impact on scientific advancements. The impact is proxied by citations, which reveal knowledge dissemination. Not all given citations indicate quality. Heavy criticism can reflect their true impact. Irrelevance is a major reason not to be cited. Citation impact is mainly influenced by the subject matter, the paper’s age, its “social status”, the document type, and the observation period (Glänzel, 2003). Citation behavior differs across fields. At times, citation scores can be inflated, favoring popular authors, topics, fields, and established journals (Groot and García-Valderrama, 2006). An otherwise important paper that is casually dismissed as common knowledge may not be cited at all. Authors working on niche areas are cited less. Citation-based analyses can also be biased due to selective citations or self and mutual citations (Sahoo et al., 2015).

When measuring research productivity, as Abramo, Costa and D’Angelo (2015) recommends the specifications for the exercise must also include the publication period and the “citation window” (to address the already mentioned time lag of publications and the so called shelf life of papers, see Ferro and D’Elia (2020)). The “publication window”, again according Abramo, Costa and D’Angelo (2015) refers both to the date of a paper’s original submission to a journal to its date of acceptance, and then from acceptance to its actual publication. These vary greatly within the same discipline. Publication delays differ across fields. Publication intensity is linked to the type of research, as well as to the entire research life cycle: a scientist, say, could appear to be completely unproductive if evaluated during the launch of a new research program. The reliability of citations to approximate the publication impact is higher when the length of the “citation window” is longer (Abramo, Costa and D’Angelo, 2015).

The most important indicators of co-authorship are the number and share of co-authored papers of a unit, joint publications of different units, the strength of co-authorship links, and the profile and citation impact of co-publications. Co-authorship weight can be assigned following fractional counting, first address count, or full or integer counting for each contributor. The first and the second are problematic when the practice of the discipline is simply alphabetical order. In some fields, instead, the tendency is to put the most relevant contributor first (Glänzel, 2003).

Many publications are internationally co-authored and result from collaborative efforts involving more than one country. While different principles and counting methods can be applied in bibliometric studies, the most common method is whole counting. Thus, every country receives full credit. An alternative is fractionalized counting, where the credit is divided proportionally between the participating countries. Whole counting reflects the number of papers in which the country

has “participated”. The choice of counting method influences the output variable because the proportion of internationally co-authored publications varies across countries. According to Aksnes et al. (2017), small countries tend to obtain higher productivity results from whole counting than fractional counting.

The importance of patents refers to the subsequent technological change they induce, and it can be measured by citations received from new patents. Forward citations are positively correlated with the market value of a patent. The generality of patents concerns the technological scope of a technology’s impact on subsequent innovation and is frequently measured by a concentration measure of a patent’s forward citations. Patents can be issued for trivial or incremental inventions, while other patented innovations can yield subsequent technological progress for decades. During the application process, patent applicants and the patent examiner are required to cite antecedents for the proposed innovation. Experts consider highly cited patents important. Patents cited by patents from vast technology fields are deemed more general than those cited by patents with applications from few subfields. Patent quality is also expected to be positively correlated with the number of jurisdictions in which it has been filed (Schmid and Fajebi, 2019). An alternative measure of patents is less academic and more commercial, attributing impact on patents that reveal, through grants, economic impact.

Once inputs and outputs have been specified, it is important to identify possible environmental variables that may explain the differences in the efficiency scores of decision-making

units. Environmental variables make it possible to address observable heterogeneity owing to uncontrollable factors. The main difference between environmental and production drivers is that the latter influence technology structure, while the former influence the efficiency with which the drivers are converted into outputs.

Empirical results discussion

One value added of the paper is to study the evolution of efficiency across the time: take a subset of best achievers in terms of gross production (USA, the UK, Germany and China), the evolution of efficiency scores across the time provides an impressive tendency to convergence in efficiency (clear in CRS CORE Model, and less clear but also present in VRS CORE Model) (See Figure 1). It is explained by the growth of China: in the period, this country increased its ydocum production by 553 percent, its qpategr in 2,842 percent, its full-time equivalent researchers in 102 percent and its R&D expenditures (at PPP values in 568 percent). In the same period, its GDP increased 226 percent. The ratio ydocum / researcher increased 223 percent in the period; and the ratio qpategr / researcher increased in 1357 percent. In 2003 the comparative productivity in terms of papers produced by researcher of China was slightly less than 20 percent the average researcher in the USA. In the last observation the value is almost 60 percent. The average researcher produced in China 1/13 patents with respect to the USA. In the last observation, the proportion converged. The comparison with USA, UK and Germany is presented in Table 6.

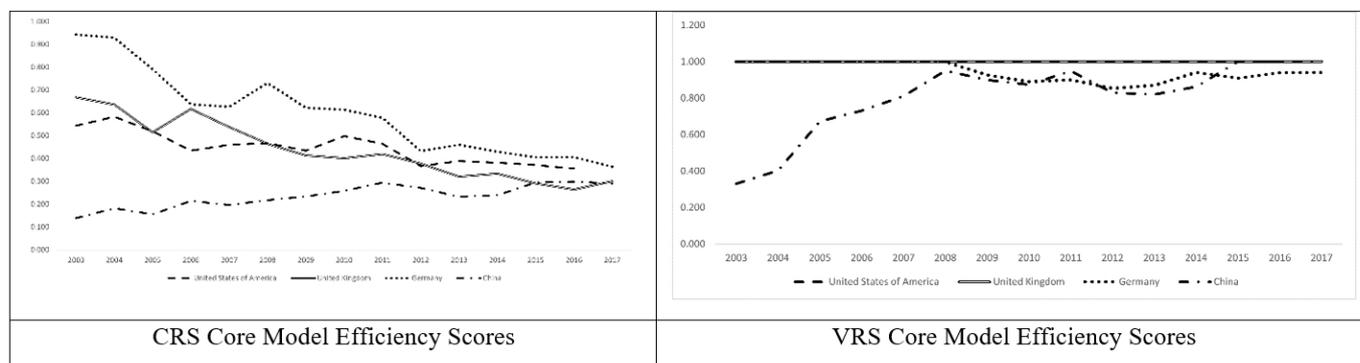


Figure 1: The evolution of efficiency in selected countries

GROWTH RATE		ydocum	qpategr	xfteres	xgerdpp	GDP pc
China	2017 / 2003	553%	2,842%	102%	568%	226%
Germany	2017 / 2003	76%	46%	54%	51%	23%
UK	2017 / 2003	81%	37%	34%	30%	12%
USA	2016 / 2003	55%	82%	21%	37%	14%
MEAN PRODUCTIVITY GROWTH		Ydocum / xfteres	Ydocum / xgerdpp	Qpategr / xfteres	Qpategr / xgerdpp	
China	2017 / 2003	223%	-2%	1,357%	340%	
Germany	2017 / 2003	15%	16%	-5%	-3%	
UK	2017 / 2003	35%	39%	3%	6%	
USA	2016 / 2003	28%	13%	50%	33%	

Table 6: The evolution of outputs, inputs, environmental variable and productivities (selected countries) (source: own elaboration)

The method we apply does not compare performance among incomparable units, instead it does compare “peers”. Among “peers” (in our case, countries with similar resources) you can find best performers, middle performers, and low performers. If we say that (for example in CRS ENV Model) Chile is fully efficient and Japan is also fully efficient, we are not saying that Chile and Japan produce the same quantity or quality of research products, what we are saying is that Chile performs better than countries with similar resources (its peers) and the same is true for Japan. Chile and Japan are in different points of the frontier. On average for the whole period, Chile ydocum figure is 8,770 papers and qpatagr is 258 patents, while Japan ydocum figure is 127,816 papers and qpatagr is 262,009 patents. To focus on policy, one national authority can watch macro-tendencies. The first question after seen these evidence is which micro determinants explain the results?

The scale consideration, in turn, tries to address whether there is an advantage of being big, which can have some rationale at laboratory level, and to certain extent, but it is not for sure at country levels. Finally, the environmental condition we test is another way to be fair in the comparison. Argentina and Brazil, for instance, are in a similar 0.19 low efficiency level in the CRS CORE model; if corrected by per capita GDP (the environmental condition), a bit lower in Brazil than in Argentina, Brazil goes to 0.55 and Argentina to 0.34. The interpretation is that Argentina wastes part of its slight per capita GDP advantage against Brazil, at least producing knowledge, all the other things equal. On the other hand, if compared the CRS CORE with the VRS CORE, Argentina’s score is 0.33 and Brazil 0.40, the difference being attributable to the scale advantage due to the greater size of Brazil.

The correlation between partial productivity ratios and efficiency scores is comparatively higher in CRS models, and the highest corresponds to the CORE CRS model. On the other hand, the efficiency scores correlate 0.66 (both CRS) and 0.88 (both VRS), while CORE CRS and CORE VRS correlate 0.69, and ENV CRS and ENV VRS correlate 0.85. The low positive correlations between researchers’ partial productivity ratios indicate some degree of complementarity between both outputs, while the low but negative productivity of financial resources devoted to both outputs suggests that they compete for the R&D funds.

CONCLUSION

Our study builds and examines a proxy of a production frontier for national science systems using DEA methodology. To that end, we concentrate on two outputs: publications and patents, and on two inputs – researchers (human resources) and R&D expenditure (non-human resources). We aggregate our measures to country levels as the unit of analysis.

Conventional indicators of our sample show that the average researcher in the sample produces 0.53 publications per year (3.43 maximum) and 0.17 patents (1.28 maximum). We go beyond partial productivity analysis and estimate production frontiers. We do not make behavioral assumptions about the mechanisms of national innovation systems. As the frontier is a non-parametric estimate, the orientation (to inputs or to outputs) is only a criterion to determine which variable is discretionary. A non-oriented method is appropriate when both

inputs and outputs can be modified discretionally. We estimate a “CORE” model following the economic theory (production requires human and non-human resources to produce outputs) and explore a comprehensive environmental variable adjusting that core model in an “ENVIRONMENTAL” (ENV) one. We do not conjecture a priori in favor of or against the presence of economies of scale. Thus, we estimate CRS and VRS versions of CORE and ENV models.

The results appear to be broadly consistent (when considering average values) with partial productivity indexes: efficiency and productivity have relatively high and positive correlations. The CORE CRS model reveals that the expected regions are the most efficient and that efficient areas are coincident with their affluence. Nevertheless, there are some surprises when analyzing individual countries. Small countries are positioned as efficient in the CORE CRS model while, contrary to what we expected, the most traditional producers of knowledge do not have outstanding efficiency results.

When VRS is incorporated, for instance, the United States, the United Kingdom become efficient, and Germany is quite close to being fully efficient. We follow another road, incorporating an encompassing environmental variable: GDPpc. When the ENV model is run under CRS, Germany, the USA, and the United Kingdom improve, the latter two becoming fully efficient. On the other hand, when GDPpc is incorporated as an environmental variable, they become efficient. The same is true for China, which is by no means efficient in the CORE model.

The role of variable returns to scale is unclear, especially when the level of aggregation is as high as it is in this analysis. The CRS assumption is only fitting when all of the decision-making units are operating at an optimal scale. Banker et al. (1984) suggest extending the CRS DEA model to account for variable returns to scale (VRS) situations. The use of the CRS specification when not all decision-making units are operating at the optimal scale will result in measures of total efficiency. One shortcoming of this measure of efficiency is that the value does not indicate whether the producer is operating in an area of increasing or decreasing returns to scale. This can be determined by running an additional DEA problem with a non-increasing return to scale (NIRS) assumption.

Is it meaningful to speak of constant or variable returns to scale in this context? It is probably useful, since a CORE CRS model yields a higher inclusion error (i.e., considering a country efficient when it is not) and exclusion (considering a country inefficient when it is not). The adoption of the hypothesis of VRS, or the consideration of GDPpc as an environmental variable, reduces the standard deviation of the most traditional R&D producers, which implies a lower exclusion error. However, it does not correct the inclusion error. More levels of disaggregation of the information would improve the results.

For R&D policy makers this kinds of study are key elements to allocate resources, to establish priorities, to set goals, to evaluate past initiatives, to compare with similar countries with best achievements, to extract lessons, to change course, to avoid misleading objectives or instruments, and to trigger delving deeper in details about the “why” of the observed performance. They are useful to project possible future paths, to identify commonalities and differences.

REFERENCES

- Abramo, G. and D'Angelo, C. A. (2014) 'How do you define and measure research productivity?', *Scientometrics*, Vol. 101, No. 2, pp. 1129–1144. <https://doi.org/10.1007/s11192-014-1269-8>
- Abramo, G., D'Angelo, C. A. and Di Costa, F. (2015) 'A New Approach to Measure the Scientific Strengths of Territories', *Journal of the Association for Information Science and Technology*, Vol. 66, No. 6, pp. 1167–1177. <https://doi.org/10.1002/asi.23257>
- Abramo, G., D'Angelo, C. A. and Murgia, G. (2016) 'The combined effects of age and seniority on research performance of full professors', *Science and Public Policy*, Vol. 43, No. 3, pp. 301–319. <https://doi.org/10.1093/scipol/scv037>
- Abramo, G., Costa, C. and D'Angelo, C. A. (2015) 'A multivariate stochastic model to assess research performance', *Scientometrics*, Vol. 102, pp. 1755–1772. <https://doi.org/10.1007/s11192-014-1474-5>
- Aksnes, D. W., Sivertsen, G., van Leeuwen, T. N. and Wendt, K. K. (2017) 'Measuring the productivity of national R&D systems: Challenges in cross-national comparisons of R&D input and publication output indicators', *Science and Public Policy*, Vol. 44, No. 2, pp. 246–258. <https://doi.org/10.1093/scipol/scw058>
- Archambault, É., Campbell, D., Gingras, Y., and Larivière, V. (2009) 'Comparing bibliometric statistics obtained from the Web of Science and Scopus', *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 7, pp. 1320–1326. <https://doi.org/10.1002/asi.21062>
- Bandola-Gill, J. (2019) 'Between relevance and excellence? Research impact agenda and the production of policy knowledge', *Science and Public Policy*, Vol. 46, No. 6, pp. 895–905. <https://doi.org/10.1093/scipol/scz037>
- Banker, R., Charnes, A. and Cooper, W. W. (1984) 'Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis', *Management Science*, Vol. 30, No. 9, pp. 1078–1192. <https://doi.org/10.1287/mnsc.30.9.1078>
- Battese, G. and Coelli, T. (1988) 'Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data', *Journal of Econometrics*, Vol. 38, No. 3, pp. 387–399. [https://doi.org/10.1016/0304-4076\(88\)90053-X](https://doi.org/10.1016/0304-4076(88)90053-X)
- Bornmann, L., Gralka, S., de Moya Anegón, F., and Wohlrabe, K. (2020) 'Efficiency of Universities and Research-Focused Institutions Worldwide: An Empirical DEA Investigation Based on Institutional Publication Numbers and Estimated Academic Staff Numbers', *CESifo Working Paper*, No. 8157.
- Charnes, A., Cooper, W. W. and Rhodes, E. (1978) 'Measuring the Efficiency of Decision-Making Units', *European Journal of Operational Research*, Vol. 2, No. 6, pp. 429–444. [https://doi.org/10.1016/0377-2217\(78\)90138-8](https://doi.org/10.1016/0377-2217(78)90138-8)
- Choi, H. and Zo, H. (2019) 'Assessing the efficiency of national innovation systems in developing countries', *Science and Public Policy*, Vol. 46, No. 4, pp. 530–540. <https://doi.org/10.1093/scipol/scz005>
- Cooper, W. W., Seiford, L. M. and Tone, K. (2007) *Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, 2nd Edition, Boston, MA: Springer. <https://doi.org/10.1007/978-0-387-45283-8>
- De Fraja, G. and Valbonesi P. (2012) 'The Design of the University System', *Journal of Public Economics*, Vol. 96, No. 3–4, pp. 317–330. <https://doi.org/10.1016/j.jpubeco.2011.10.005>
- D'Este, P., Ramos-Vielba, I., Woolley, R. and Amara, N. (2018) 'How do researchers generate scientific and societal impacts? Toward an analytical and operational framework', *Science and Public Policy*, Vol. 45, No. 6, pp. 752–763. <https://doi.org/10.1093/scipol/scy023>
- Elías, J., Ferro, G. and García, A. (2019) 'A Quest for Quality: Creativity and Innovation in the Wine Industry of Argentina', *AAWE Working Paper* 242, 09/2019.
- Ferro, G. and D'Elia, V. (2020) 'Higher Education Efficiency Frontier Analysis: A Review of Variables to Consider', *Journal of Efficiency and Responsibility in Education and Science*, Vol. 13, No. 3, pp. 140–153. <https://doi.org/10.7160/eriesj.2020.130304>
- Glänzel, W. (2003) *Bibliometrics As A Research Field. A Course on Theory and Application of Bibliometric Indicators*, Course Handouts.
- Gralka, S., Wohlrabe, K. and Bornmann, L. (2019) 'How to measure research efficiency in higher education? Research grants vs. publication output', *Journal of Higher Education Policy and Management*, Vol. 41, No. 3, pp. 322–341. <https://doi.org/10.1080/1360080x.2019.1588492>
- Groot, T. and García-Valderrama, T. (2006) 'Research Quality and Efficiency. An Analysis of Assessments and Management Issues in Dutch Economics and Business Research Programs', *Research Policy*, Vol. 35, No. 9, pp. 1362–1376. <https://doi.org/10.1016/j.respol.2006.07.002>
- Leydesdorff, L. and Wagner, C. (2009) 'Macro-level indicators of the relations between research funding and research output', *Journal of Informetrics*, Vol. 3, No. 4, pp. 353–362. <https://doi.org/10.1016/j.joi.2009.05.005>
- Lewis, K. K. (1999) 'Trying to Explain Home Bias in Equities and Consumption', *Journal of Economic Literature*, Vol. 37, No. 2, pp. 571–608. <https://doi.org/10.1257/jel.37.2.571>
- Lundvall, B.-Å. (2007) 'National Innovation Systems—Analytical Concept and Development Tool', *Industry and Innovation*, Vol. 14, No. 1, pp. 95–119. <https://doi.org/10.1080/13662710601130863>
- Marxt, C. and Brunner, C. (2013) 'Analyzing and Improving the National Innovation System of Highly Developed Countries—The Case of Switzerland', *Technological Forecasting and Social Change*, Vol. 80, No. 6, pp. 1035–1049. <https://doi.org/10.1016/j.techfore.2012.07.008>
- Njøes, R. and Jakobsen, S.-E. (2018) 'Policy for Evolution of Regional Innovation Systems: The Role of Social Capital and Regional Particularities', *Science and Public Policy*, Vol. 45, No. 2, pp. 257–268. <https://doi.org/10.1093/scipol/scx064>
- Rhaim, M. (2017) 'Measurement and Determinants of Academic Research Efficiency: A Systematic Review of the Evidence', *Scientometrics*, Vol. 110, No. 2, pp. 581–615. <https://doi.org/10.1007/s11192-016-2173-1>
- Sahoo, B., Singh, R., Mishra, B. and Sankaran, K. (2015) 'Research Productivity in Management Schools of India: A Directional Benefit-of-Doubt Model Analysis', *MPRA Paper*, No. 67046. <https://mpra.ub.uni-muenchen.de/67046/>
- Schmid, J. and Fajébe, A. (2019) 'Variation in patent impact by organization type: An investigation of government, university, and corporate patents', *Science and Public Policy*, Vol. 46, No. 4, pp. 589–598. <https://doi.org/10.1093/scipol/scz010>
- Scimago (2020) *Scimago Journal & Country Rank*, [Online], Available: <https://www.scimagojr.com/countryrank.php> [20 Mar 2020].

Thelwall, M. and Fairclough, R. (2017) 'The research production of nations and departments: A statistical model for the share of publications', *Journal of Informetrics*, Vol. 11, No. 4, pp. 1142–1157. <https://doi.org/10.1016/j.joi.2017.10.001>

UNESCO (2020) *UNESCO Institute for Statistics (UIS)*, [Online], Available: <http://data.uis.unesco.org/> [20 Mar 2020].

WIPO (2020) *Information Resources on Patents*, [Online], Available: <https://www.wipo.int/patents/en/> [20 Mar 2020].

World Bank (2020) *World Bank Open Data*, [Online], Available: <https://data.worldbank.org/> [20 Mar 2020].

APPENDIX

A LOG-LOG COBB-DOUGLAS COST FUNCTION

The model estimates the financial resources devoted to R&D (in logarithms) as a function of the logs of both outputs (ydocum and ypateap), the relative price of publications in terms of patents (estimated as the quotient of zgerdpp on xfteres) and a time trend (model 1). After reaching a satisfactory core model, we try to address quality issues, approximating the logs of qcitati for publications and qpategr for patents. The first variable proves to be non-significant and the second is significant, but the log of ypateap loses its statistical significance (Model 2). We hold that the marginal cost of publishing is positive (in human and non-human resources), while it is also positive for licensing patents. Thus, we run a third model, using the logs of publications and licensing of patents and do not include the logs of qcitati and ypateap (Model 3). We use this as a guide to design the CORE and ENV DEA models in Section 3.

Independent variables (t):	Dependent Variable: lxgerdpp	Model 1	Model 2	Model 3
lydocum		0.8080*** (25.29)	0.7916*** (16.04)	0.7945*** (26.61)
lypateap		0.0749*** (4.03)	0.0230 (0.98)	
zgerd_fteres		0.0026*** (4.03)	0.0026*** (10.23)	0.0026*** (10.27)
Trend		-0.2069*** (-6.62)	-0.0212*** (-4.24)	-0.0210*** (-6.89)
lqpategr			0.0682*** (3.62)	0.0807*** (5.28)
lqcitati			-0.0231 (-0.07)	
Constant		6.4566*** (31.84)	6.5837*** (29.20)	6.6190*** (31.86)
Observations		930	930	930
Groups		102	102	102
R-sq: within		0.3857	0.3935	0.3956
R-sq: between		0.9367	0.9366	0.9356
R-sq: overall		0.9387	0.9391	0.9381
Wald chi2(4)		1,832.9500	1,864.4800	1,783.7700
Prob > chi2		0.0000	0.0000	0.0000
Sigma_u		0.5596	0.5579	0.5800
Sigma_e		0.2633	0.2610	0.2621
Rho (fraction of variance due to u_i)		0.8187	0.8124	0.8304

*** 0.99 significant, ** 0.95 significance, * 0.90 significance

Table A1: Cobb-Douglas in logs cost function. Random-effects GLS regression, unbalanced panel. (102 groups, minimum observations per group 1, maximum observations per group 15) (ource: own elaboration)