Technovation 107 (2021) 102314

Contents lists available at ScienceDirect

Technovation

journal homepage: www.elsevier.com/locate/technovation

Is more always better? On the relevance of decreasing returns to scale on innovation

Javier Barbero [a], Jon Mikel Zabala-Iturriagagoitia [b,c,d], José L. Zofío [e,f,*]

[a] Oviedo Efficiency Group (OEG), Universidad de Oviedo, Asturias, Spain
[b] Deusto Business School, University of Deusto, Donostia-San Sebastián, Spain
[c] South Eastern University Norway, Kongsberg, Norway
[d] CIRCLE, Lund University, Lund, Sweden
[e] Department of Economics, Universidad Autónoma de Madrid, Spain
[f] Erasmus Research Institute of Management (ERIM), Erasmus University Rotterdam, the Netherlands

ARTICLE INFO

Keywords:
Innovation system
Innovation performance
European innovation scoreboard
Data envelopment analysis
TOPSIS
Innovation policy

ABSTRACT

We contribute to the literature on the assessment of innovation systems by relating the amount of inputs available to the system and its performance through the concept of returns to scale (increasing, constant or decreasing). We study to what extent the size or scale of innovation systems relates to their performance, which is estimated through frontier Data Envelopment Analysis-TOPSIS methods, which overcome several limitations of the standard DEA approach.

Using the same data provided by the European Innovation Scoreboard (*EIS)* for years 2010, 2013 and 2016, our results indicate that countries with a high innovation scale tend to overinvest in innovation inputs. This results into scale inefficiencies stemming from decreasing returns, leading to lower productivity levels. Thanks to DEA-TOPSIS we identify the best and worst performing innovation systems. This provides helpful information by setting suitable reference benchmarks for policy analysis and decision-making.

Our results question the current allocation of resources and call for a reconsideration of how innovation policies are designed in many European countries. We conclude that for the *EIS* to become a useful instrument for the definition of innovation policies, it should consider the nature of returns to scale. This would allow policymakers to identify problems and limitations related to the size of their respective innovation systems, and hence, design holistic innovation policies to act upon them.

1. Introduction

Innovation (the 'residual' in growth accounting) is the most important source of productivity growth and thereby of increased welfare. The European Commission has been one of the most active agents as to the measurement of innovation with the development of the European Innovation Scoreboard (between 2010 and 2015, Innovation Union Scoreboard) and the implementation of the Community Innovation Surveys (*CIS*). Other scoreboards also include the UK Competitiveness Index, the index of the Massachusetts Innovation Economy, the Global Innovation Index, the Nordic Innovation Monitor or the Bloomberg Innovation Index to mention a few. What these approaches share is that they all are based on the use of a synthetic scalar measure that, through composite indicators, provides a ranking of the countries under study, with consequent political implications.

However, the simplistic use of synthetic composite indicators may be dangerous because the rankings derived from them are often taken for granted, without any deliberation of their validity (Grupp and Mogee, 2004; Grupp and Schubert, 2010: 69). They can have a communication function to raise awareness about innovation policy, but they should not be instrumentally used to make policy decisions without relevant qualifications (see Edquist et al., 2018). However, if innovation scoreboards are expected to have a real impact on the definition of innovation policies, it is essential that they set the ground for an exhaustive characterization of innovation systems. We believe that such characterization can only be achieved when it is based on sound scientific concepts and methodologies, which are able to identify the strengths and weaknesses of every innovation system, so the policy design can take this

* Corresponding author. Department of Economics, Universidad Autónoma de Madrid. Francisco Tomás y Valiente 5, 28049 Madrid, Spain.
E-mail addresses: jose.zofio@uam.es, jzofio@rsm.nl (J.L. Zofío).

https://doi.org/10.1016/j.technovation.2021.102314
Received 9 January 2020; Received in revised form 19 March 2021; Accepted 22 May 2021

diagnosis as a point of departure.

Due to the impact that innovation is having in most economies, and also as a consequence of increasing interests from policy-makers concerning public accountability (Lovell, 2002; Batterbury, 2006), there has been an increasing development, use and exploitation of indicators to improve the measurement of innovation systems (Castro-Martínez et al., 2009; Dziallas and Blind, 2019). Several concepts have been introduced in the literature to assess and characterize innovation systems, such as innovation capacity, innovation potential, propensity to innovate, innovativeness or innovation performance to mention a few (Carayannis et al., 2015; Furman et al., 2002; Hagedoorn and Cloodt, 2003; Jordan, 2010; Mairesse and Mohnen, 2002; Prajogo and Ahmed, 2006; Zabala-Iturriagagoitia et al., 2007a). However, few have tackled the actual meaning behind these concepts, neither offering robust definitions that allow distinguishing them, nor discussing their potential complementary effects (Carayannis and Grigoroudis, 2014; Lee, 2015).

In this paper we relate the volume of innovation inputs available to an innovation system and its performance. *Innovation inputs* is here used as a measure of the amount of resources that are invested in the innovation system. In turn, *innovation performance* is defined as the relationship between these resources (i.e. innovation inputs) and the results (i.e. innovation outputs) that emanate from the system. Hence, innovation performance is defined as a measure of the efficiency levels achieved by a particular innovation system, or, equivalently, its relative productivity (see Edquist et al., 2018). Both concepts are rendered operational in the paper through corresponding scalar measures of input size and efficiency. These two concepts are also related through the analysis of the returns to scale (increasing, constant or decreasing) to clarify if an increase in the amount of innovation inputs available to the innovation system also leads to higher innovation performance. If countries increase innovation inputs and the resulting increase in outputs is proportionally larger, equal or smaller than the increase in the inputs, then they face increasing, constant or decreasing returns to scale, respectively.[1] In the latter case of decreasing returns, continuous efforts to increase the amount of innovation inputs will result in successive reductions in the productivity of the system (as outputs grow at a lower rate than inputs, and therefore their ratio decreases), which will eventually result in lower efficiency levels, when compared to those of other countries. This justifies the importance of determining the nature of returns of scale of innovation systems and testing the previous hypothesis.

A key feature of the concept of innovation performance used in this paper is that it is defined in relative terms.[2] Production theory studies the relation between the amount of inputs used within a system (i.e., the scale of the system) and the amount of outputs that such system is capable of producing (Shephard, 1970). From this relation a natural measure of relative performance emerges by comparing the two through the concept of efficiency or productivity (i.e. the ratio of an aggregate output index to an aggregate input index). Hence, to assess the innovation performance of national innovation systems we rely on the literature on efficiency and productivity (Fried et al., 2008). This is interactively determined by multi-lateral comparisons of multiple input-output combinations (Guan and Chen, 2012). To measure innovation performance in a robust manner we introduce to the field of innovation the advanced Data Envelopment Analysis techniques related to the Technique for Order of Preference by Similarity to Ideal Solutions (DEA-TOPSIS).[3] The rationale for using the DEA-TOPSIS methods lies in that it helps overcome the severe limitations of the standard DEA approach, which has already been applied to assess innovation performance in previous research efforts (e.g., Zabala-Iturriagagoitia et al., 2007b; Edquist et al., 2018).

The previous methodology is applied to the data provided by the European Innovation Scoreboard (*EIS*) for years 2010, 2013 and 2016.[4] The *EIS* aims to "provide a comparative assessment of the research and innovation performance of the EU Member States and the relative strengths and weaknesses of their research and innovation systems" (European Union, 2017: 8). The *EIS* is the main instrument used by the European Commission to monitor the results achieved by the Innovation Union, which is one of the flagship initiatives defined by the European Union within its Europe 2020 Strategy to create an innovation-friendly environment that supports the generation, emergence and diffusion of innovations. The *EIS* calculates a Summary Innovation Index (*SII*) that synthetizes all the indicators included in the *EIS*, regardless of their character (i.e. inputs, outputs, determinants, outcomes, impacts), by calculating their arithmetic mean (i.e. the basic aggregating function one may rely on to obtain a single scalar measure). The *SII* ranks all EU countries according to what is explicitly called "EU Member States' Innovation Performance" (European Union, 2017), so the underlying logic is that the bigger the *SII* ('size' in terms of all indicators), the better the innovation performance.[5]

The rest of the paper is organized as follows. Section 2 presents the indicators that constitute the 'standard' or base model followed in the paper to characterize national innovation systems based on the data provided by the *EIS*. In section 3 we present the applied methodology and discuss the advantages of using the DEA adaptation of the multi-criteria TOPSIS method compared to the standard DEA approach that has been applied in the literature for the purpose of measuring the efficiency of innovation systems. Section 4 compares the results achieved when DEA-TOPSIS and standard DEA methods are applied to the same models for years 2010, 2013 and 2016. Section 5 shows the conclusions that can be deduced from the previous results, while providing a discussion of the main findings and its relevance for the practice of innovation policymaking.

2. Characterizing innovation systems

Innovation systems are composed by a complex network of interacting organizations, policies and institutions whose main purpose is to improve the conditions under which the emergence, generation, diffusion and uptake of innovations take place (Metcalfe, 1995; Palmberg, 2006). Since its advent, the characterization of innovation systems has always constituted a challenge, not only for policy-makers but also for researchers engaged in the innovation studies community. The original works by Nelson (1993), Lundvall (1992) and Edquist (1997), among many others, started studying the main characteristics of a set of national innovation systems, analyzing the organizations embedded in them and the institutions affecting those. These studies set the ground for the emergence of a wider set of quantitative studies. In Europe in particular, the data provided by the *CIS* and the *EIS* facilitated the rapid

[1] Numerically, the concept of decreasing returns to scale takes place when a proportional increase in inputs (e.g., doubling the amount of inputs, 2x) results in a lower proportional growth in outputs (i.e., less than double, <2x). If outputs and inputs increase in the same proportion, then constant returns to scale are observed (i.e., double, = 2x). Finally, increasing returns to scale are observed if the outputs grow at a higher rate than the inputs (greater than double, >2x).

[2] The main purpose of developing comparative studies "is to assist policy by summarizing a range of innovation indicators at the national, regional or sector level, by permitting a comparison of the relative success or failure of the innovation system, or through the identification of specific aspects of the innovation system which perform well or poorly" (Arundel and Hollanders, 2008: 30).

[3] See Section 3 for a detailed description of the DEA-TOPSIS methodology.

[4] The data are retrieved from the 2017 edition of the European Innovation Scoreboard.

[5] The *EIS* does not provide any specific definition of innovation performance, beyond the Summary Innovation Index (*SII*). It can thus be said that for the *EIS*, innovation performance is understood as the arithmetic mean of all the indicators in the *EIS* (i.e. the *SII*).

development of quantitative works. However, despite the fact that the number of indicators available for measuring innovation has increased over the last three decades, the characterization problem still remains (Dziallas and Blind, 2019).

As already introduced, the aim of this paper is to relate the amount of innovation inputs available to the innovation system and its innovation performance through the concept of returns to scale. The idea of addressing the performance of innovation systems has been already discussed in the literature (Zabala-Iturriagagoitia et al., 2007b; Cherchye et al., 2008; Nasierowski and Arcelus, 2012; Carayannis et al., 2015; Kou et al., 2016; Edquist et al., 2018). A good example in this regard is the Global Innovation Index (Dutta et al., 2017), which calculates the so-called "*Innovation Efficiency Ratio*", defined as the productivity ratio between an aggregate innovation output sub-index over an aggregate innovation input sub-index (p. 12).[6] Indeed, this approach is, in its basic characteristics, very similar to the "*Productivity Innovation Index*" proposed by Edquist et al. (2018). However, the former is based on composite indicators while the latter resorts to optimizing Data Envelopment Analysis (DEA) techniques for aggregation. As regards the concept of innovation scale, the Global Innovation Index provides a synthetic input score measuring the size of innovation systems, depending on the availability of inputs that are put into the system.

Nonetheless, there are still several methodological limitations that previous studies have not managed to elucidate, and which justifies the novel approach adopted here (see Section 3.1). Despite the extensive evidence suggesting that efficiency methods can provide a systemic interpretation of innovation performance, the *EIS* has remained methodologically blind. Both the "*Summary Innovation Index" (SII)* of the *EIS* and the overall "*Global Innovation Index score*" (*GII*) still follow a "the bigger, the better" rationale. The underlying logic behind composite measures such as the *SII* or the *GII* is that the larger the amount of these synthetic indicators, the superior will also be the ability of the system to produce and diffuse innovations. However, this does not mean that the system is able to keep certain proportionality between the invested resources and the actual results in terms of innovation, leading to a partial explanation on the functioning of innovation systems. To capture this relevant economic effect, we introduce the concept of innovation performance, which measures the efficiency of innovation systems.

If we are to characterize and measure innovation systems, it is critical to select the right set of input and output indicators capturing their complex characteristics (Katz, 2016), which is not an easy task. The number and kind of indicators to be used is difficult to deduce systemically from innovation theory (Grupp and Schubert, 2010: 68; Dziallas and Blind, 2019). The literature on innovation studies has discerned a series of activities or functions which are accomplished within the frame of an innovation system, and which are required for its proper functioning (e.g. Galli and Teubal, 1997; McKelvey, 1997; Brenner and Broekel, 2011). Such a focus on the activities of an innovation system emphasizes 'what happens in the system', providing a dynamic perspective on its functioning. Each of the activities may thus be considered a partial determinant of the development and diffusion of innovations.

In this regard, Hekkert et al. (2007) propose a set of seven functions to be applied when mapping an innovation system. These seven functions are: (i) entrepreneurial activities, (ii) knowledge development, (iii) knowledge diffusion through networks, (iv) guidance of the search, (v) market formation, (vi) resources mobilization, and (vii) creation of legitimacy/counteract resistance to change. Johnson (2001) also follows a similar approach, by relating the functions accomplished by an innovation system with the activities developed in it. Accordingly, she identifies the following functions: (i) supply incentives for companies to engage in innovative work, (ii) supply resources, (iii) guide the direction of search, (iv) recognize the potential for growth of innovation, (v) facilitate the exchange of information and knowledge, (vi) stimulate/create markets, (vii) reduce social uncertainty, and (viii) counteract the resistance to change. The work by Edquist (2005, 2011) needs to be stressed here, as he provides a systemic and holistic rationale for the activities required by an innovation system. Edquist introduces a list of ten activities, which are structured into four thematic categories, and which represent those factors that influence, support, hinder, ease and promote the development of innovation processes (see Table 1).

The functions listed above are not independent, but rather reinforce — or offset — one another (see Acs et al., 2014). The underlying logic behind the previous views on innovation functions is that the more activities developed in a system, the larger will be its ability to produce and diffuse innovations (i.e. namely, to act as a fully equipped system). However, the literature has not managed to elucidate as to yet which could be the indicators that could help characterize each of the previous functions, which remains for further exploration.

Since our purpose in this paper is limited to national innovation systems in Europe, a departing point is the set of indicators provided by the *EIS*. Edquist et al. (2018) discuss at length which subset of the *EIS* indicators are most sensible for measuring innovation performance. The set of key indicators they substantiate as being more appropriate for such a purpose is presented in Table 2. From their point of view, despite the indicators included in the *EIS* are related to innovation, many of them refer to environmental or contextual factors (e.g. population with tertiary education, new doctorate graduates), or to the impact of innovations on the economy as a whole (e.g. employment in knowledge-intensive activities, medium and high-tech product exports), which are beyond the actual management of any innovation system.

To ease comparison with their results, in this paper we use the same set of indicators, corresponding to what Edquist et al. (2018) name

Table 1
Key activities/functions of innovation systems.

I. Provision of knowledge inputs to the innovation process
1. <u>Provision of R&D results,</u> and thus creation of new knowledge, primarily in engineering, medicine and natural sciences.
2. <u>Competence building, *e.g.*</u> through individual learning (educating and training the labour force for innovation and R&D activities) and organizational learning. This includes formal learning as well as informal learning.

II. Demand-side activities
3. <u>Formation of new product markets (*e.g.* public procurement for innovation)</u>.
4. <u>Articulation of new product quality requirements</u> emanating from the demand side.

III. Provision of constituents
5. <u>Creating and changing organizations</u> needed for developing new fields of innovation. Examples include enhancing entrepreneurship to create new firms and intrapreneurship to diversify existing firms, and creating new research organizations, policy organizations, etc.
6. <u>Networking through markets and other mechanisms</u>, including interactive learning among different organizations (potentially) involved in the innovation processes. This implies integrating new knowledge elements developed in different spheres of the SI and coming from outside with elements already available in the innovating firms.
7. <u>Creating and changing institutions</u> – e.g., patent laws, tax laws, environment and safety regulations, R&D investment routines, cultural norms, etc. – that influence innovating organizations and innovation processes by providing incentives for and removing obstacles to innovation.

IV. Support services for innovating firms
8. <u>Incubation activities</u> such as providing access to facilities and administrative support for innovating efforts.
9. <u>Financing of innovation processes</u> and other activities that may facilitate commercialisation of knowledge and its adoption.
10. <u>Provision of consultancy services</u> relevant for innovation processes, e.g., technology transfer, commercial information, and legal advice.

Source: Adapted from Edquist (2011).

[6] The *Global Innovation Index* (*GII*) includes 81 indicators for 143 countries. In it, all indicators are classified as innovation inputs or outputs and a sub-index is calculated for each. Rather than using the arithmetic mean for aggregation, a weighted mean is employed, but the weights for the individual input and output indicators are unreported. The *GII* also provides an "*overall GII score*" which follows the same logic as the *SII* of the *EIS*.

Table 2
The "*standard (baseline) model*" for measuring innovation systems.

	Innovation output indicators
2.2.1	SMEs innovating in-house (% of SMEs)
2.3.3	Community trademarks per billion GDP (in PPP€)
2.3.4	Community designs per billion GDP (in PPP€)
3.1.1	SMEs introducing product or process innovations (% of SMEs)
3.1.2	SMEs introducing marketing or organizational innovations (% of SMEs)
3.2.2	Contribution of medium and high-tech products exports to the trade balance
3.2.3	Knowledge-intensive services exports (as % of total service exports)
3.2.4	Sales of new to market and new to firm innovations (as % of turnover)
	Innovation input indicators
1.3.1	R&D expenditure in the public sector (% of GDP)
1.3.2	Venture capital (% of GDP)
2.1.1	R&D expenditure in the business sector (% of GDP)
2.1.2	Non-R&D innovation expenditures (% of turnover)

Source: Adapted from Edquist et al. (2018).

"standard (baseline) model", according to the following criteria (p. 199).[7]:

- *Innovation inputs*: variables referring to the resources (human, material and financial; private as well as governmental) used not only to create innovations but also to bring them to the market.
- *Innovation outputs*: variables referring to new products and processes, new designs and community trademarks, as well as marketing and organizational innovations, which are connected to the market, and which can either be new to the world, the industry and/or to the firm.

Accordingly, a scalar measure of *innovation inputs* corresponds to the arithmetic mean of the four indicators selected by Edquist et al. (2018). Increasing the number of possible input indicators (corresponding to alternative models studied by these authors) does not result in substantial changes in the value of the arithmetic mean. We have calculated alternative definitions by considering up to seven innovation inputs instead of four. The correlations for year 2010 are 0.9069, 0.9245 for year 2013, and 0.8753 for year 2016.[8] In turn, *innovation performance* is measured using the DEA-TOPSIS method applied to the previous set of eight innovation output indicators and four innovation inputs.

Given the heterogeneity of innovation systems in Europe, the set of 12 indicators identified by Edquist et al. (2018) allows to assess those activities which need to be undertaken by all innovation systems, despite, with different intensities. As we discuss in the method section (see Section 3) the DEA approach will attribute the most favorable weights to each of these indicators in each country, depending on its structural characteristics, so as to maximize its relative efficiency. The next section discusses the methodological details behind the measurement of innovation performance using DEA-TOPSIS methods.

[7] Edquist et al. (2018) carry out a sensitivity analysis of what they label "the standard model" of innovation performance, by considering as many as 12 innovation outputs and 7 innovation inputs, and provide evidence of the high correlations between the standard model and the "extended" version. The interested reader can consult this study for the specific list of additional inputs and outputs considered in the extended model.

[8] Moreover, as shown in the empirical section, we perform a sensitivity analysis of the innovation performance results obtained for each country using the baseline model, by changing the number of the input and output indicators. We conclude its robustness to changes in the number of indicators included in the analysis. Hence the choice of the number of inputs in the calculation of innovation scale does not statistically change the results on innovation performance.

3. Methodology

Charnes et al. (1978) introduced DEA to assess the relative performance of a group of observations. Their original ratio-form formulation, known as *CCR* in the literature, computes the productivity of each observation relative to those of their remaining counterparts. However, the original formulation restricts productivity measurement to either the partial (radial) output orientation that increases outputs given a level of inputs, or its counterpart input orientation that reduces inputs for a given a level of outputs. Later, Chambers et al. (1996) proposed a more flexible measure of relative productivity by allowing for both output increases and input reductions.

This can be formalized in the context of the present study as follows. Let us denote by $j = 1, \ldots, J$ the set of countries observed in $t = 1, \ldots, T$ time periods−years. Countries use innovation resources (human, material and financial), each represented by the elements of the following−input−vector: $x_i^t = (x_{1i}^t, \ldots, x_{Ni}^t) \in R_+^N$, to generate innovation outputs such as new products and processes, new designs, etc., represented by the output vector $y_i^t = (y_{1i}^t, \ldots, y_{Mi}^t) \in R_+^M$. The relative productivity of an innovation system i in the direction defined by the vector $g^t = \left(g_x^t, g_y^t\right) \neq 0$, can be calculated by solving the following program:

$$\min_{\nu_n^t, \mu_m^t, \omega^t} - \sum_{m=1}^{M} \mu_m^t y_{im}^t + \sum_{n=1}^{N} \nu_n^t x_{in}^t + \omega^t = \overrightarrow{D}^t\left(x_i^t, y_i^t; g^t\right) \quad (1)$$

s.t.

$$\frac{\sum_{m=1}^{M} \mu_m^t y_{jm}^t - \omega^t}{\sum_{n=1}^{N} \nu_n^t x_{jn}^t} \leq 1, \quad j = 1, \ldots, J,$$

$$\sum_{n=1}^{N} \nu_n^t g_n^t + \sum_{m=1}^{M} \mu_m^t g_m^t = 1,$$

$$\nu_n^t \geq 0, \;\; \mu_m^t \geq 0,$$

where ν_n^{t*} and μ_m^{t*} denote the optimal input and output weights, and ω^{t*} is a scalar that informs about the nature of returns to scale at the benchmark variable returns to scale frontier−see Fukuyama (2003). Note that the weights ν_n^{t*} and μ_m^{t*} define aggregator functions for the inputs and outputs and that the objective function in (1) represents the supporting hyperplane bounding the productivity levels across the group of J countries.

When country i under evaluation maximizes productivity at the scale given by ω^{t*}, it corresponds to the minimum feasible *distance* to the supporting hyperplane, and therefore the objective function is zero: $\overrightarrow{D}^t\left(x_i^t, y_i^t; g^t\right) = 0$, implying that i defines the reference frontier. The greater the optimal value (distance to the frontier), the more inefficient the innovation system. Notice that we can deem the solution to (1) as a relative inefficiency measure because the set of J constraints normalize the productivity to one. Therefore, when an innovation system is efficient, its productivity is maximal and equal to one in the $i = j$ constraint corresponding to itself. The scale parameter ω^{t*} in program (1) plays a critical role in our analysis by capturing the nature of returns to scale of innovation inputs. Fukuyama (2003: 114) extends the analysis by Banker et al. (1984) on the scale properties of the radial output and input formulations under variable returns to scale to the directional distance function. In this case, the nature of returns to scale at the reference frontier can be ascertained through the following values:

i) Decreasing Returns to Scale (DRS) prevails for $(x_i^t, y_i^t) \Leftrightarrow \omega^{t*}(x_i^t, y_i^t; g^t) > 0$ for all optimal solutions.
ii) Increasing Returns to Scale (IRS) prevails for $(x_i^t, y_i^t) \Leftrightarrow \omega^{t*}(x_i^t, y_i^t; g^t) < 0$ for all optimal solutions.

iii) Constant Returns to Scale (CRS) prevails for $(x_i^t, y_i^t) \Leftrightarrow \omega^{t^*}(x_i^t, y_i^t; g^t) = 0$ for some optimal solution.

Since innovation systems subject to either decreasing or increasing returns to scale cannot maximize productivity by definition, it is relevant to determine the productivity loss caused by a suboptimal scale. This loss can be quantified by calculating the relative productivity of those countries whose $\omega^{t^*} \neq 0$ when solving program (1), with respect to those that maximize productivity according to $\nu_n^{t^*}$ and $\mu_m^{t^*}$ by exhibiting constant returns to scale (i.e., those with $\omega^{t^*} = 0$, representing most productive scale sizes in the terminology of Banker et al. (1984)). For the scale inefficient countries exhibiting decreasing or increasing returns this comparison can be perfomed by solving program (1) for a second time, but on this ocassion including only the subset of scale efficient countries with $\omega^{t^*} = 0$ in the $j = 1, \ldots, J$ restrictions.[9] Let us denote by $\overrightarrow{D}^t_{CRS}(x_i^t, y_i^t; g^t)$ the distance to the supporting hyperplane defined by the scale efficient countries characterized by constant returns to scale, then the productivity loss due to a suboptimal scale is measured through the following scale inefficiency (*SI*) measure:

$$SI(x_i^t, y_i^t; g^t) = \overrightarrow{D}^t_{CRS}(x_i^t, y_i^t; g^t) - \overrightarrow{D}^t(x_i^t, y_i^t; g^t). \tag{2}$$

All countries whose scale parameter ω^{t^*} is equal to zero when solving program (1) are scale efficient by definition, because both distances coincide, i.e, $SI(x_i^t, y_i^t; g^t) = 0$. The remaining countries exhibiting decreasing or increasing returns incur in scale inefficiency, i.e., $SI(x_i^t, y_i^t; g^t) > 0$, and its value corresponds to the productivity loss due to a suboptimal scale, measured by the distance between the constant and the variable returns to scale benchmarks. Afterwards, it is possible to identify the nature of returns to scale causing the scale inefficiency, either decreasing or increasing, by looking at the values of ω^{t^*} in program (1), as presented in i) (DRS) and ii) (IRS) above. Regarding the conclusions of our study, in the empirical section we show that most national innovation systems in Europe are prone to decreasing returns to scale, and calculate the magnitude of the resulting scale inefficiency.

As it can be observed, the previous definition of innovation performance following program (1) is far from that provided by the *EIS* through the Summary Innovation Index (*SII*), which corresponds to the arithmetic mean of the complete set of *EIS* indicators: $SII_i^t = \sum_{l=1}^{25} \overline{\upsilon}_l \, \hat{\iota}_{li}^t$, $\overline{\upsilon}_l = 1/25, \ \hat{\iota}_{li}^t \in [0,1]$, where $\hat{\iota}_{li}^t$ is the max-min normalized value of each l indicator (either input or output as no distinction is made). The Global Innovation Index (*GII*) provides a counterpart with the same structure and interpretation of the *SII*, which is called "*GII score*". A second measure provided by the *GII* is the "*Innovation Efficiency Ratio*" *(IER)*, whose formulation corresponds to a productivity measure, as it defines the ratio of an innovation output sub-index over an innovation input sub-index:

$$IER_i^t = \sum_{m=1}^{M} \mu_m y_{mi}^t \Bigg/ \sum_{n=1}^{N} \nu_n x_{ni}^t, \ \ \mu_m > 0, \ \ \nu_n > 0. \tag{3}$$

According to the *GII*, this ratio "serves to highlight those economies that have achieved more with less … providing an insight *that should be neutral to the development stages of economies*" (italics added, p. 419). This remark about the development stage can be actually interpreted as regardless the amount of innovation inputs (*scale*) of the innovation system, since developed countries inevitably show larger innovation systems (measured by the *SII* or the *GII* score). The main difference with program (1) is that the *IER* does not rely on optimizing techniques to compare performance across observations and to obtain individual input and output weights, ν_n and μ_m.[10] Moreover, the *IER* in (3) defines as an absolute productivity index that is characterized by constant returns to scale, and therefore cannot capture the nature of returns to scale that the innovation system faces, as shown by the optimal value of the scale parameter ω^{t^*}—note in particular that this is the only difference between expression (3) and the set of $j = 1, \ldots, J$ restrictions in (1). Hence, the *GII* constitutes a productivity index with fixed weights that is comparable to the efficiency measure under constant returns to scale, $\overrightarrow{D}^t_{CRS}(x_i^t, y_i^t; g^t)$, that we have presented in order to calculate scale inefficiency, $SI(x_i^t, y_i^t; g^t)$. However, as shown in the empirical section, the general assumption of constant returns to scale imposed by (3) is not warranted when variables returns to scale are allowed, as in (1).

3.1. Relevant drawbacks of standard DEA methods

Standard DEA approaches have been extensively used in the literature to assess innovation performance, as it has been proven an effective method to successfully identify efficient benchmarks. However, it presents several weaknesses that compromise its applicability from a policymaking perspective. The most relevant ones concerning our present study are the following:

1) Inability to discriminate among efficient observations: One relevant weakness of standard DEA, particularly when variable returns to scale are brought into the analysis, is that a significant subset of observations are identified as efficient, with $\overrightarrow{D}^t(x_i^t, y_i^t; g^t) = 0$.
2) Sensitivity to extreme observations representing questionable benchmarks: When searching for the most favorable optimal weights, a large number of observations are deemed efficient by default.
3) Vulnerability to rank reversals: Optimal weights are not unique, compromising the stability of ranks when additional observations or variables are included. Eventually, adding an observation with the same innovation indicators into the analysis may result in rank reversals (Wang and Luo, 2009; Soltanifar and Shahghobadi, 2014).

There have been several proposals that partially address these shortcomings. A first set focuses on the ranking issue, and aim at qualifying the scores of the efficient observations. This ranges from simple super-efficiency models to elaborated cross-efficiency proposals (Aparicio and Zofío, 2020; Balk et al., 2021). However, these methods do not solve the extreme observations problem and the need to identify a credible set of efficient observations. A second set is characterized by the introduction of weight restrictions. These restrictions are based on prior information on their relative importance such as expert opinion (e.g. engineering shadow prices), which helps to improve the discriminatory power of the method and reduce weight dispersion (see Cooper et al., 2011). However, there still remains the issue of multiple−single valued, efficient units. Finally, both methods are prone to the rank reversal problem given the multiplicity of optimal weights.

[9] From a computationally perspective this is equivalent to solving program under the general assumption of constant returns to scale, dropping the scale parameter ω^* in the objective function and the set of restrictions.

[10] The innovation input sub-index of the *GII* is comprised of five input pillars that capture elements of the national economy that enable innovative activities (Dutta et al., 2017: 11–12): institutions, human capital and research, infrastructure, market sophistication, and business sophistication. In turn, the innovation output sub-index provides information about two output pillars: knowledge and technology outputs, and creative outputs. While the *SII* is the arithmetic mean of the *EIS* indicators, the *GII* is a weighted average: "the five input pillars each have a fixed weight of 0.10; the two output pillars each have a fixed weight of 0.25" (Dutta et al., 2017: 70). However, the weights for the individual inputs and outputs within each pillar remain unreported.

3.2. The DEA-TOPSIS evaluation of national innovation systems

A method capable of jointly addressing all these limitations, allowing to establish a meaningful and robust ranking of observations, to identify credible benchmarks that policymakers can agree on, and immune to rank reversal, is the DEA version of the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). This technique creates virtual ideal (anti-ideal) production units with the maximum observed values of outputs and minimum observed values of inputs (and vice versa), and calculates for each unit two efficiency scores, namely with respect to the ideal (optimistic) and anti-ideal (pessimistic) frontiers respectively. Central to the method is the idea that decision makers can learn from both best and worst practice. Wang and Luo (2006) combine DEA and TOPSIS using the standard (radially oriented) constant returns to scale measures, showing that both methods can be integrated, so as to provide a robust ranking of observations, using undisputed benchmarks such as the ideal and anti-deal observations. Later on Wu (2006) and Chen (2012) qualified the initial proposal by improving the interpretability of the models and solving apparent inconsistencies related to conflicting orientations and efficiency values that question the possibility of aggregating the DEA-TOPSIS best and worst relative efficiencies into a relative closeness ratio (*RC*); i.e., a composite performance index.

We rely on this notion but generalize the methodology by considering as an efficiency measure that associated to the directional distance function approach, and introducing variable returns to scale. First, as the directional distance functions embeds the partially oriented standard measures, previous proposals can be obtained by setting the specific directional vectors g^t to match the input or output orientations (Färe and Grosskopf, 2000), and removing the scale parameter. Second, considering variables returns to scale allows us to explore the existence of decreasing, increasing or constant returns to scale at the individual country level.

In the present study, the DEA-TOPSIS method consists of a three-step process. In the first step the efficiency scores of the ideal and anti-ideal innovation systems, with respect to those actually observed, are calculated. In the second step, an optimistic model maximizes the relative efficiency of the evaluated unit under the condition that the best relative efficiency of the ideal unit remains unchanged. A pessimistic (or aggressive) model minimizes the relative efficiency of the unit while keeping the worst relative efficiency of the anti-ideal unchanged. The last step consists in calculating the *Relative Closeness Innovation Index* (*RCII*), which relates both measures of efficiency and ranks countries depending on their relative innovation performance.

Using TOPSIS terminology, we start out defining the "*Ideal Innovation System* (*IIS*)" as that producing the largest amount of outputs with the least amount of inputs in period t, and vice versa for the "*Anti-ideal Innovation System* (*AIS*)". Note that both the *IIS* and the *AIS* are virtual units created from real observed values.

$$IIS^t = \left(y^t_{IISm}, x^t_{IISn}\right) = \left(\max_j\left(y^t_{mj}\right), \min_j\left(x^t_{nj}\right)\right), \quad \forall\ n,\ m,$$
$$AIS^t = \left(y^t_{AISm}, x^t_{AISn}\right) = \left(\min_j\left(y^t_{mj}\right), \max_j\left(x^t_{nj}\right)\right), \quad \forall\ n,\ m.$$

Recalling program (1), and choosing as directional vector the common value corresponding to the mean of the input and output indicators: $g^t = (g^t_x, g^t_x) = (\overline{x}^t, \overline{y}^t)$—using a common value renders the efficiency scores comparable, in the first step we calculate the highest relative productivity or efficiency of the *IIS*, with respect to those actually observed across countries. This value is the solution to the following model:

$$\min_{\nu^t_n, \mu^t_m, \omega^t} -\sum_{m=1}^{M}\mu^t_m y^t_{IISm} + \sum_{n=1}^{N}\nu^t_n x^t_{IISn} + \omega^t_{IIS} = \overrightarrow{D}^t\left(x^t_{IIS}, y^t_{IIS}; g^t\right) \tag{4}$$

s.t.

$$-\sum_{m=1}^{M}\mu^t_m y^t_{jm} + \sum_{n=1}^{N}\nu^t_n x^t_{jn} + \omega^t \geq 0, \quad j = 1, ..., J,$$
$$\sum_{n=1}^{N}\nu^t_n \overline{x}^t_n + \sum_{m=1}^{M}\mu^t_m \overline{y}^t_m = 1,$$
$$\nu^t_n \geq 0,\ \ \mu^t_m \geq 0.$$

We remark that unless one of the observed countries performs as well as the *IIS*, $\overrightarrow{D}^t\left(x^t_{IIS}, y^t_{IIS}; g^t\right) < 0$, indicating that the (maximum) amounts of outputs and (minimum) amounts of inputs must be reduced and increased, respectively, to reach the maximum productivity across the observed countries. Moreover, those countries defining the optimal supporting reference hyperplane for the *IIS* in (4) can be identified as those performing best by maximizing the productivity.

While program (4) maximizes the productivity of the *IIS* with respect those observed across countries−minimizing the distance to the best actually observed innovation systems, its negative counterpart is calculated by maximizing the distance of the *AIS* to the worst performing countries. To determine this worst case bound, one solves the following model:

$$\max_{\nu^t_n, \mu^t_m, \omega^t} -\sum_{m=1}^{M}\mu^t_m y^t_{AISm} + \sum_{n=1}^{N}\nu^t_n x^t_{AISn} + \omega^t_{AIS} = \overrightarrow{D}^t\left(x^t_{AIS}, y^t_{AIS}; g^t\right) \tag{5}$$

s.t.

$$-\sum_{m=1}^{M}\mu^t_m y^t_{jm} + \sum_{n=1}^{N}\nu^t_n x^t_{jn} + \omega^t \leq 0, \quad j = 1, ..., J,$$
$$\sum_{n=1}^{N}\nu^t_n \overline{x}^t_n + \sum_{m=1}^{M}\mu^t_m \overline{y}^t_m = 1,$$
$$\nu^t_n \geq 0,\ \ \mu^t_m \geq 0.$$

On this occasion, those countries defining the optimal supporting reference hyperplane for the *AIS* in (3) can be identified as those performing worst by exhibiting the lowest productivity, which implies that their individual constraints in the set of the $j = 1, \ldots, J$ restrictions are, once again, saturated.

Therefore, the whole purpose of programs (4) and (5) is to establish the best and worst reference benchmarks (hyperplanes) across the observed innovation systems included in the *EIS*, relying on the virtual ideal and anti-ideal innovations system as reference benchmarks to identify them (see Fig. 1). Once these reference hyperplanes have been obtained they can be used to calculate the optimistic and pessimistic performance of each country with respect to them.

The second step of the TOPSIS method evaluates the performance of country i with respect to these best and worst benchmarks. Starting with the ideal benchmark, relative efficiency can be determined by solving an equivalent program to (4) but ensuring that the efficiency of the ideal reference remains constant, thereby restricting the set of available optimal hyperplanes to those previously identified. That is:

$$\min_{\nu^t_n, \mu^t_m, \omega^t} -\sum_{m=1}^{M}\mu^t_m y^t_{im} + \sum_{n=1}^{N}\nu^t_n x^t_{in} + \omega^t = \overrightarrow{D}^t_{IIS}\left(x^t_i, y^t_i; g^t\right) \tag{6}$$

s.t.

$$-\sum_{m=1}^{M}\mu^t_m y^t_{jm} + \sum_{n=1}^{N}\nu^t_n x^t_{jn} + \omega^t \geq 0, \quad j = 1, ..., J,$$
$$-\sum_{m=1}^{M}\mu^t_m y^t_{IISm} + \sum_{n=1}^{N}\nu^t_n x^t_{IISn} + \omega^t = \overrightarrow{D}^t\left(x^t_{IIS}, y^t_{IIS}; g^t\right),$$
$$\sum_{n=1}^{N}\nu^t_n \overline{x}^t_n + \sum_{m=1}^{M}\mu^t_m \overline{y}^t_m = 1,$$
$$\nu^t_n \geq 0,\ \ \mu^t_m \geq 0.$$

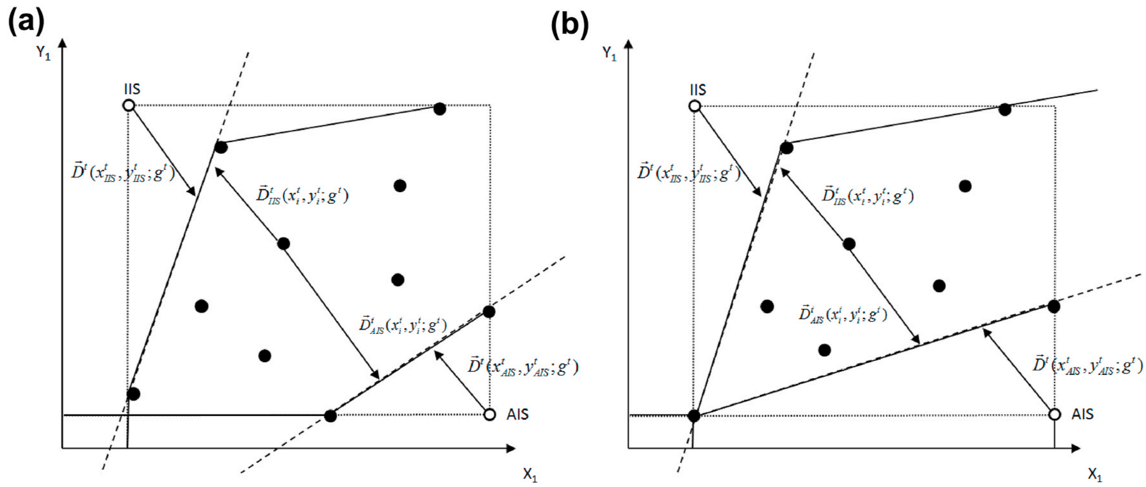

Fig. 1. A graphical representation of the DEA-TOPSIS method with one input and one output.

Consequently, if $\overrightarrow{D}^t_{IIS}(x^t_i, y^t_i; g^t) = 0$, the innovation system of country i defines the reference hyperplane for the *IIS* in program (4); otherwise $\overrightarrow{D}^t_{IIS}(x^t_i, y^t_i; g^t) > 0$, and the shortest the distance the better the country under evaluation performs with respect to the benchmark peers identified by the *IIS*. Next, the counterpart to this program, representing the pessimistic approach measuring how distant is an innovation system to the worst reference hyperplane identified by the *AIS* in (5), is given by:

$$\max_{\nu^t_n, \mu^t_m, \omega^t} - \sum_{m=1}^{M} \mu^t_m y^t_{im} + \sum_{n=1}^{N} \nu^t_n x^t_{in} + \omega^t = \overrightarrow{D}^t_{AIS}(x^t_i, y^t_i; g^t) \tag{7}$$

s.t.

$$- \sum_{m=1}^{M} \mu^t_m y^t_{jm} + \sum_{n=1}^{N} \nu^t_n x^t_{jn} + \omega^t \leq 0, \quad j = 1, ..., J,$$

$$- \sum_{m=1}^{M} \mu^t_m y^t_{AISm} + \sum_{n=1}^{N} \nu^t_n x^t_{AISn} + \omega^t = \overrightarrow{D}^t(x^t_{AIS}, y^t_{AIS}; g^t),$$

$$\sum_{n=1}^{N} \nu^t_n \overline{x}^t_n + \sum_{m=1}^{M} \mu^t_m \overline{y}^t_m = 1,$$

$$\nu^t_n \geq 0, \quad \mu^t_m \geq 0.$$

Again, if $\overrightarrow{D}^t_{AIS}(x^t_i, y^t_i; g^t) = 0$, the innovation system of country i defines the worst reference hyperplane for *AIS* in program (5). But now, since the rest of the innovation systems perform better that these observations, their outputs and inputs are to be reduced and increased, respectively, to reach the worst hyperplane, with $\overrightarrow{D}^t_{AIS}(x^t_i, y^t_i; g^t) < 0$. Therefore the longest the distance in absolute values to the reference hyperplane, the better the performance of the country under evaluation with respect to the worst benchmarks.

Fig. 1. a provides a graphical illustration of the four different distances calculated in order to facilitate its interpretation. Fig. 1. b shows the extreme case of one innovation system defining both the best and worst frontiers (i.e., the point at the bottom-left). This result implies that an innovation system can perform well on some dimensions, but poorly on others. Results for these observations should be interpreted with caution and studied on a case-by-case basis.[11]

[11] Shen et al. (2016) establish sufficient conditions to ensure that the efficient and inefficient frontiers do not intersect. In our empirical application we obtain this result in very few occasions.

3.3. A robust indicator for innovation systems performance

The last step defines a performance composite index in the vein of the relative closeness (*RC*) ratio proposed by the TOPSIS method. Models (3) and (5) measure the best possible relative efficiencies of the *IIS* and those of the actually observed systems, while models (4) and (6) measure the worst possible relative efficiencies as compared to the *AIS*. As both performance indicators may lead to different conclusions, a robust indicator capturing both dimensions into a single scalar is defined. Hence, the *Relative Closeness Innovation Index* (*RCII*) is defined as follows:

$$RCII^t_i = \frac{\left| \overrightarrow{D}^t_{AIS}(x^t_i, y^t_i; g^t) - \overrightarrow{D}^t(x^t_{AIS}, y^t_{AIS}; g^t) \right|}{\left| \overrightarrow{D}^t_{AIS}(x^t_i, y^t_i; g^t) - \overrightarrow{D}^t(x^t_{AIS}, y^t_{AIS}; g^t) \right| + \left| \overrightarrow{D}^t(x^t_{IIS}, y^t_{IIS}; g^t) - \overrightarrow{D}^t_{IIS}(x^t_i, y^t_i; g^t) \right|}, \tag{8}$$

where the distance differences are considered in absolute terms given the negative values that $\overrightarrow{D}^t_{AIS}(x^t_i, y^t_i; g^t)$ and $\overrightarrow{D}^t(x^t_{IIS}, y^t_{IIS}; g^t)$ may adopt in both the numerator and denominator.

The monotonicity properties of this indicator in the present directional distance function are satisfactory. For a given innovation system under evaluation, its $RCII^t_i$ index is: i) increasing in the distance between its efficiency with respect to the worst reference frontier and that of the anti-ideal innovation system: $\overrightarrow{D}^t_{AIS}(x^t_i, y^t_i; g^t) - \overrightarrow{D}^t(x^t_{AIS}, y^t_{AIS}; g^t)$–i.e. the farther away from the worst frontier the better; and ii) decreasing in the distance between its efficiency with respect to the best reference frontier and that of the ideal innovation system: $\overrightarrow{D}^t(x^t_{IIS}, y^t_{IIS}; g^t) - \overrightarrow{D}^t_{IIS}(x^t_i, y^t_i; g^t)$–i.e. the closer to the best frontier the better. In summary, the shorter the distance to the best reference frontier and the longer the distance from the worst reference frontier, the greater the value of $RCII^t_i$.

In the following section both the standard DEA and the DEA-TOPSIS approaches will be used to characterize the nature of returns to scale and to assess the performance of national innovation systems in Europe.

4. Empirical results

4.1. Assessing innovation performance through standard DEA

Recalling the discussion in Section 2 on the characterization of innovation systems, we first implement the DEA-TOPSIS analysis for the baseline model including the selected four input and eight output indicators proposed by Edquist et al. (2018). Before undertaking the analysis, we explore the efficiency levels and returns to scale characteristics of the innovations systems of the standard DEA method following program (1) – see Section 3. This allows identifying the initial benchmark innovation systems under variable returns to scale, which in the first step of the DEA-TOPSIS method are potential candidates to define the best reference hyperplanes for the *IIS*. We also calculate the magnitude of the scale inefficiency resulting from decreasing or increasing returns to scale, implying that innovation systems exhibit suboptimal scales.

Table 3 presents the results for the directional distance function in years 2010, 2013, and 2016. The weaknesses of the standard DEA becomes immediately clear, resulting in a large number of innovations systems being efficient and thereby pointing to the low discriminatory power in ranking analyses of standard DEA. Out of the 31 countries considered in the *EIS,* only 6, 7 and 9 countries are respectively deemed inefficient in these three years; i.e. $\overrightarrow{D}^t(x_i^t, y_i^t; g^t) \geq 0$. Estonia, Croatia, Lithuania, The Netherlands and Poland present evidence of inefficient innovation systems in at least two years, while Sweden is consistently inefficient in all years considered. These results should not come as a surprise. On the one hand, Sweden invests heavily in innovation, placing this country at the top of the rank by the amount of innovation inputs (i. e. largest aggregate input indicator). The direct effect is that it also ranks high up in terms of aggregate output indicators. This is why it is generally in the first position of the *SII*, computed as the arithmetic mean of all *EIS* indicators. However, even if Sweden is defined as an *innovation leader* by the *EIS*, when an undisputed measure of relative efficiency or productivity like (1) is brought into the analysis, Swedish performance can be described as relatively poor, belonging to the scarcely populated set of inefficient countries.

In relation to the high number of efficient countries, and thanks to the parameter characterizing the nature of returns to scale in program (1), ω^{t*}, we identify that many of them are indeed scale inefficient: 12 in 2010, 10 in 2013, and 19 in 2016. Fukuyama (2003) shows that maximum productivity is attained by those observations defining reference hyperplanes characterized by constant returns to scale ($\omega^{t*}= 0$). Contrarily, those countries that, being efficient, exhibit increasing or decreasing returns to scale ($\omega^{t*}<0$ and $\omega^{t*}> 0$, respectively) incur in scale inefficiencies and therefore could increase their productivity by either increasing or reducing the amounts of innovation inputs employed. Table 3 reports the magnitude of the scale inefficiency caused by either decreasing or increasing return to scale, defined as the difference between distances under constant and variables returns to scale, $SI(x_i^t, y_i^t; g^t) = \overrightarrow{D}^t_{CRS}(x_i^t, y_i^t; g^t) - \overrightarrow{D}^t(x_i^t, y_i^t; g^t)$, see expression (2). Comparing the values of $SI(x_i^t, y_i^t; g^t)$ and $\overrightarrow{D}^t(x_i^t, y_i^t; g^t)$ we see that scale inefficiency is remarkably larger than technical inefficiency. Indeed, the average value of $SI(x_i^t, y_i^t; g^t)$ doubles that of $\overrightarrow{D}^t_{CRS}(x_i^t, y_i^t; g^t)$, and even triples the value in 2016: 0.087 vs. 0.030. This shows the relevance of the productivity loss resulting from a suboptimal scale, either because of increasing or decreasing returns. We now confirm that this loss is due to the latter.

As shown in the last three columns of Table 3, decreasing returns to scale predominate over increasing returns as the main cause of scale inefficiency (only in 4 cases ω^{t*} is negative indicating the existence of increasing returns, IRS). This is confirmed in Table 4 summarizing the nature of returns to scale. The widespread and increasing existence of decreasing returns causing scale inefficiency is confirmed. In fact, there is a shift towards decreasing returns over the whole period. While in 2010 the number of innovation systems (efficient and inefficient) with increasing returns represented 6.5% of countries ($n = 2$), those with decreasing returns accounted for 35.5% ($n = 11$). Yet, in 2016, 58.1% of the innovations systems exhibit decreasing returns ($n = 18$). As for inefficient countries, the existence of decreasing returns in their reference benchmarks is certain, since over 80% of them are subject to this type of returns in all three years.[12]

These results show the positive correlation between scale inefficiency and innovation inefficiency, implying that countries with a high innovation scale tend to overinvest in innovation inputs when compared to less developed innovation systems. To corroborate this relationship we have calculated the correlation between scale inefficiency $SI(x_i^t, y_i^t; g^t)$ and the value of the scale parameter ω^{t*}in absolute terms (since both negative (IRS) and positive (DRS) values result in scale inefficiency): $\rho(SI(x_i^t, y_i^t; g^t),\ |\omega^{t*}|) = 0.534$, which is significant at the 1% level. Hence, we conclude that the existence of decreasing returns to scale results in relevant scale inefficiencies of the innovation systems.

This result proves that the forces that prevent the materialization of all investments into innovation outputs are highly related scale inefficiencies, because of lower rates of return to those investments. There is a need to reflect on the cause of this result, because once a country reaches high levels of innovation scale due to the large investments made in it, even if the returns on those investment are low, it is counterintuitive to prescribe a downsizing (i.e. reducing the amount of innovation inputs) of the innovation system from a policy perspective. Yet, the other option of increasing the innovation outputs associated to those investments is also unrealistic given the existence of decreasing returns. Nevertheless, for inefficient countries, a general recommendation is to learn from other countries that, using similar amounts of innovation inputs (i.e. and which are in the same development stage of their innovation system), exhibit better innovation performance. We thus conclude that a serious reconsideration of the way in which resources are being invested should be made, since our results point to clear saturations in some of the activities of the system (see Acs et al., 2014), in contrast to others which could be still subject to increasing returns to scale of innovation inputs. This calls for the realization of possible economies of scope (or diversification versus specialization), related to changes in the mix of innovation outputs, which need to be explored as possible sources of further productivity gains (e.g. Morita 2003).

4.2. Assessing innovation performance through DEA-TOPSIS

The first step in the DEA-TOPSIS methodology is to calculate the highest (lowest) relative productivity or efficiency of the *IIS* (*AIS*), with respect to those actually observed, solving programs (3) and (4) respectively. Table 5 provides the distances of the *IIS* and *AIS* with respect to the best and worst reference benchmarks (hyperplanes) across the observed sample of countries (see Fig. 1). As for the nature of returns to scale at the best and worst innovation frontiers, inspecting the reference countries for the *IIS* for the three years we learn that in most cases they exhibit constant returns (83.3%), with a scale parameter $\omega^{t*} = 0$ in Table 3.[13] Unsurprisingly, this implies that the reference hyperplane for the *IIS* is defined by benchmarks that maximize productivity since they are not subject to increasing or decreasing returns. Conversely, the reference hyperplane for the *AIS* is made up of a

[12] Only Romania (2010), Latvia (2013) and Lithuania (2016) exhibit increasing returns.

[13] Reference countries for the *IIS* are Austria (2010, 2013), Cyprus (2010), Germany (2010, 2013), Greece (2013), Italy (2010, 2013), Luxemburg (2010, 2013, 2016), Malta (2010, 2013, 2016), Slovakia (2016), Spain (2013), United Kingdom (2016), and Switzerland (2016).

Table 3
Efficiency values of the innovation system, scale inefficiency and returns to scale, program (1).

Country	$\overrightarrow{D}^t(x_i^t, y_i^t; g^t)$			$\overrightarrow{D}^t_{CRS}(x_i^t, y_i^t; g^t)$			$SI(x_i^t, y_i^t; g^t)$			ω^{t^*}			Returns to scale		
	2010	2013	2016	2010	2013	2016	2010	2013	2016	2010	2013	2016	2010	2013	2016
Austria	0.000	0.000	0.000	0.000	0.000	0.083	0.000	0.000	0.083	0.000	0.000	0.709	CRS	CRS	DRS
Belgium	0.051	0.000	0.000	0.131	0.000	0.122	0.080	0.000	0.122	0.898	0.000	0.197	DRS	CRS	DRS
Bulgaria	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	CRS	CRS	CRS
Croatia	0.034	0.000	0.139	0.064	0.000	0.140	0.030	0.000	0.001	0.313	0.000	0.033	DRS	CRS	DRS
Cyprus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	CRS	CRS	CRS
Czech Republic	0.000	0.000	0.000	0.239	0.048	0.129	0.239	0.048	0.129	0.500	0.208	0.293	DRS	DRS	DRS
Denmark	0.000	0.000	0.161	0.038	0.000	0.179	0.038	0.000	0.019	0.148	0.000	0.214	DRS	CRS	DRS
Estonia	0.000	0.189	0.148	0.168	0.433	0.325	0.168	0.243	0.177	0.228	1.184	0.864	DRS	DRS	DRS
Finland	0.000	0.034	0.000	0.081	0.210	0.203	0.081	0.176	0.203	0.342	0.976	0.748	DRS	DRS	DRS
France	0.000	0.029	0.000	0.000	0.055	0.184	0.000	0.027	0.184	0.000	0.758	0.807	CRS	DRS	DRS
Germany	0.000	0.000	0.000	0.000	0.078	0.278	0.000	0.078	0.278	0.000	0.342	0.331	CRS	DRS	DRS
Greece	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	CRS	CRS	CRS
Hungary	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	CRS	CRS	CRS
Ireland	0.000	0.000	0.000	0.063	0.000	0.000	0.063	0.000	0.000	0.662	0.000	0.000	DRS	CRS	CRS
Italy	0.000	0.000	0.000	0.000	0.000	0.048	0.000	0.000	0.048	0.000	0.000	0.385	CRS	CRS	DRS
Latvia	0.000	0.000	0.050	0.000	0.000	0.056	0.000	0.000	0.006	0.000	0.000	−0.028	CRS	CRS	IRS
Lithuania	0.000	0.066	0.034	0.013	0.112	0.103	0.013	0.046	0.069	−0.120	−0.167	0.391	IRS	IRS	DRS
Luxembourg	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	CRS	CRS	DRS
Malta	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	CRS	CRS	CRS
Netherlands	0.090	0.073	0.000	0.102	0.144	0.000	0.012	0.071	0.000	0.596	0.610	0.000	DRS	DRS	CRS
Poland	0.000	0.037	0.135	0.000	0.045	0.147	0.000	0.008	0.012	0.000	0.110	0.065	CRS	DRS	DRS
Portugal	0.000	0.000	0.000	0.000	0.000	0.185	0.000	0.000	0.185	0.000	0.000	0.318	CRS	CRS	DRS
Romania	0.046	0.000	0.000	0.047	0.000	0.000	0.002	0.000	0.000	−0.123	0.000	0.000	IRS	CRS	CRS
Slovakia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	CRS	CRS	CRS
Slovenia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.053	0.000	0.000	DRS	CRS	CRS
Spain	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	CRS	IRS	CRS
Sweden	0.236	0.187	0.121	0.357	0.285	0.524	0.121	0.097	0.403	0.898	1.427	1.493	DRS	DRS	DRS
United Kingdom	0.016	0.000	0.000	0.021	0.000	0.021	0.005	0.000	0.021	0.067	0.000	0.431	DRS	CRS	DRS
Iceland	0.000	0.000	0.028	0.000	0.000	0.298	0.000	0.000	0.270	0.000	0.000	0.678	CRS	CRS	DRS
Norway	0.000	0.000	0.122	0.000	0.000	0.339	0.000	0.000	0.217	0.000	0.000	0.522	CRS	CRS	DRS
Switzerland	0.000	0.000	0.000	0.000	0.379	0.258	0.000	0.379	0.258	0.000	0.685	0.285	CRS	DRS	DRS
Average	0.015	0.020	0.030	0.043	0.058	0.117	0.027	0.038	0.087	0.305	0.613	0.460			
Maximum	0.236	0.189	0.161	0.357	0.433	0.524	0.239	0.379	0.403	0.898	1.427	1.493			
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	−0.195	−0.167	−0.028			
Stand. Dev.	0.046	0.049	0.055	0.082	0.115	0.134	0.057	0.085	0.113	0.364	0.499	0.362			

Notes: $g^t = \left(g_x^t, g_y^t\right) = (\overline{x}^t, \overline{y}^t)$; * Descriptive statistics for the scale parameter ω^{t^*} exclude scale efficient countries for whom it is zero in the reported optimal solution.
Legend: DRS: Decreasing returns to scale; IRS: Increasing returns to scale; CRS: Constant returns to scale.

Table 4
Characterizing the nature of innovation systems' returns to scale. Program (1).

2010					
N° efficient *ISs* (%): 25 (80.6%)			**N° inefficient *ISs* (%): 6 (19.4%)**		
IRS ($\omega^{t^*} < 0$)	**CRS** ($\omega^{t^*} = 0$)	**DRS** ($\omega^{t^*} > 0$)	**IRS** ($\omega^{t^*} < 0$)	**CRS** ($\omega^{t^*} = 0$)	**DRS** ($\omega^{t^*} > 0$)
1 (4.0%)	18 (72.0%)	6 (24.0%)	1 (16.7%)	0 (0.0%)	5 (83.3%)
2013					
N° efficient *ISs* (%): 24 (77.4%)			**N° inefficient *ISs* (%): 7 (22.6%)**		
IRS ($\omega^{t^*} < 0$)	**CRS** ($\omega^{t^*} = 0$)	**DRS** ($\omega^{t^*} > 0$)	**IRS** ($\omega^{t^*} < 0$)	**CRS** ($\omega^{t^*} = 0$)	**DRS** ($\omega^{t^*} > 0$)
0 (0.0%)	21 (87.5%)	3 (12.5%)	1 (14.3%)	0 (0.0%)	6 (85.70%)
2016					
N° efficient *ISs* (%): 22 (71.0%)			**N° inefficient *ISs* (%): 9 (29.0%)**		
IRS ($\omega^{t^*} < 0$)	**CRS** ($\omega^{t^*} = 0$)	**DRS** ($\omega^{t^*} > 0$)	**IRS** ($\omega^{t^*} < 0$)	**CRS** ($\omega^{t^*} = 0$)	**DRS** ($\omega^{t^*} > 0$)
0 (0.0%)	12 (54.5%)	10 (45.5%)	1 (11.1%)	0 (0.0%)	8 (88.9%)

Table 5
DEA-TOPSIS analysis. Step 1. Benchmark ideal and anti-ideal efficiencies. Programs (3) and (4).

Year	Benchmark	
	Ideal I.S. (*IIS*)	Anti-ideal I.S. (*AIS*)
	$\overrightarrow{D}^t_{IIS}(x_i^t, y_i^t; g^t)$	$\overrightarrow{D}^t_{AIS}(x_i^t, y_i^t; g^t)$
2010	−0.865	1.008
2013	−0.836	1.322
2016	−0.931	0.977

majority of countries (57.9%) that exhibit decreasing returns to scale $\omega^{t^*} > 0$, and therefore incur in scale inefficiencies.[14,15] Therefore, based on the nature of returns to scale at the supporting hyperplanes, we

[14] Reference countries for the *AIS* are the Czech Republic (2010), Estonia (2010, 2013, 2016), Finland (2013), Ireland (2013), Lithuania (2016), Hungary (2016), Poland (2016), Romania (2013), Slovenia (2013), Sweden (2010, 2016), United Kingdom (2013), Iceland (2016), and Switzerland (2010, 2013, 2016).

[15] The reference countries for the *IIS* and the *AIS* can be identified from the optimal solutions to programs (4) and (5), respectively. They corresponds to those countries in the $j = 1, \ldots, J$ restrictions whose constraints are satisfied as equalities (therefore equal to zero).

anticipate that countries exhibiting suboptimal scales (e.g. by using large amounts of inputs), will tend to rank poorly in terms of the relative closeness innovation index, *RCII*, because they share the scale characteristics of the worst performing benchmarks. We illustrate this supposition in Fig. 2 below.

The results provided by the TOPSIS method allow us to rank the 31 countries included in the *EIS* according to their *RCII* so the larger its value the better the relative performance of the innovation system. After solving programs (5) and (6), Table 6 shows the distances of every country with regard to the best and worst frontiers, the value of the *RCII* as defined in (7), and the ranking of innovation performance for every country in 2010, 2013 and 2016. The ranking is led by Luxembourg in 2010 and 2016 and by Austria in 2013. Other countries scoring high in their innovation performance are Italy, Greece, Germany and Malta. All these countries are characterized by being closer to the *IIS* and farther away from the *AIS*. It can also be observed that the relative performance of every innovation system is more clearly depicted by the DEA-TOPSIS method, as compared to the standard DEA approach, in which most countries were ranked as efficient (see previous section). The DEA-TOPSIS method also increases the discriminatory power of the evaluation by reducing the number of reference hyperplanes, yielding an individual ranking position.

From a policy perspective, we contend that this diversity of efficient countries provides a relevant range of potential benchmarks. Again, this does not mean all countries should be benchmarked with Malta or Greece, nor do the results imply that Greece or Malta are the most comprehensive innovation systems. We do however contend that there is something to be learnt from these high performing innovation systems, which can serve as example for others in similar stages of development.[16] This implies that every country should aim to benchmark itself against those structurally similar innovation systems. The solutions provided by the DEA-TOPSIS allow identifying which could be the most sensible benchmarks that every inefficient country could learn from. In the case of Sweden for example, as a relatively ‘small’ developed economy, representing a so called “innovation leader” by the *EIS*, but nevertheless showing a poor innovation performance, sensible targets could be Austria (benchmark in 2010 and 2013), Luxembourg (benchmark in 2010 and 2013) or Switzerland (benchmark in 2016), due to their similar innovation scale, but better innovation performance. In turn, for a larger economy like Spain, a country labelled as a “moderate innovator” by the *EIS* and showing an intermediate innovation performance, the DEA-TOPSIS identifies Germany (benchmark in 2010), Italy (benchmark in 2010) or the United Kingdom (benchmark in 2016) as potential references. Spain could learn from the innovation policies being implemented (and the results observed thereof) in these ‘peer’ countries, so identifying the mix of benchmarks should be welcome by policymakers in innovation policy circles. The identification of benchmarks represents one of the major advantages of DEA methods, and hence, its systematic exploration could be further explored.

To demonstrate that large spending in innovation inputs may result in diminishing innovation performance for EU countries because of the widespread existence of decreasing returns to scale, we regress innovation performance, using the calculated $RCII_i$, on innovation input scale, $Scale_i$—defined, once again, as the arithmetic mean of the four input indicators. Pooling the three years under study: 2010, 2013, 2016, the estimated coefficient for innovation input scale is −0.156, with a robust *t*-statistic of −2.96, which is significant at the 0.01 level; i.e., $RCII_i = 0.602–0.156\ Scale_i$. This confirms the existence of a mild but significant negative relationship between innovation performance and innovation scale. Particularly in the last year, 2016, when decreasing returns to scale are prevalent with 18 countries enduring this source of

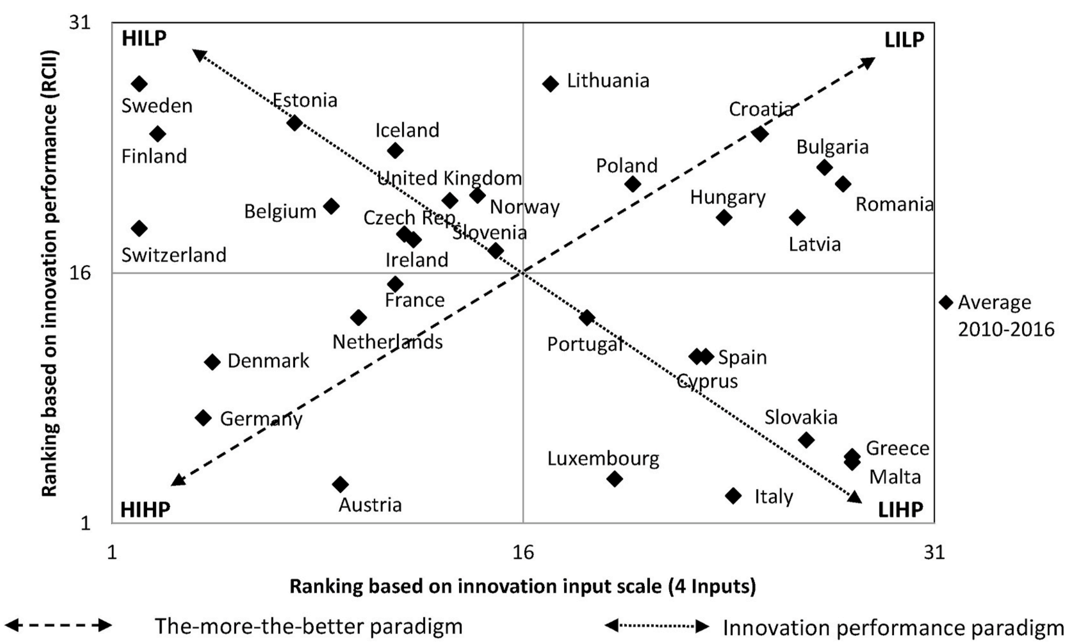

Fig. 2. Country categories based on the rankings of *innovation input scale* and *innovation performance* (Average for years 2010–2013-2016).

[16] The only caveat is the case illustrated in Fig. 1 b corresponding to innovation systems that define both the efficient and inefficient frontiers. This is only observed for Switzerland in 2016. However, this outcome does not alter substantially its raking position based on the RCII, 16th, which does not differ much from that of the two previous years.

Table 6
DEA-TOPSIS analysis.

Country	2010				2013				2016			
	$\vec{D}^t_{IIS}(x^t_i,y^t_i;g^t)$	$\vec{D}^t_{AIS}(x^t_i,y^t_i;g^t)$	$RCII^t_i$	Rank	$\vec{D}^{2014}_{IIS}(x^t_i,y^t_i;g^t)$	$\vec{D}^t_{AIS}(x^t_i,y^t_i;g^t)$	$RCII^t_i$	Rank	$\vec{D}^t_{IIS}(x^t_i,y^t_i;g^t)$	$\vec{D}^t_{AIS}(x^t_i,y^t_i;g^t)$	$RCII^t_i$	Rank
Austria	0.000	−0.273	0.597	5	0.000	−0.739	0.711	1	0.027	−0.354	0.582	4
Belgium	0.500	−0.292	0.488	24	0.269	−0.460	0.617	14	0.618	−0.324	0.456	22
Bulgaria	0.397	−0.008	0.446	27	0.283	−0.203	0.577	22	0.259	−0.153	0.487	18
Croatia	0.377	−0.118	0.476	26	0.459	−0.596	0.597	18	0.931	−0.030	0.351	29
Cyprus	0.000	−0.262	0.595	6	0.110	−0.235	0.622	13	0.606	−0.668	0.517	14
Czech Republic	0.743	0.000	0.385	31	0.179	−0.483	0.640	12	0.208	−0.280	0.525	12
Denmark	0.257	−0.292	0.537	11	0.057	−0.275	0.641	11	0.240	−0.351	0.531	10
Estonia	0.169	0.000	0.494	22	0.372	0.000	0.523	27	0.644	0.000	0.383	26
Finland	0.694	0.000	0.393	30	0.260	−0.087	0.562	24	0.495	−0.336	0.479	19
France	0.344	−0.287	0.517	16	0.295	−0.354	0.597	17	0.254	−0.305	0.520	13
Germany	0.000	−0.393	0.618	3	0.000	−0.514	0.687	4	0.175	−0.199	0.515	15
Greece	0.115	−0.485	0.604	4	0.000	−0.733	0.711	2	0.293	−0.421	0.533	8
Hungary	0.281	−0.168	0.506	18	0.278	−0.351	0.600	16	0.438	0.000	0.417	24
Ireland	0.456	−0.369	0.510	17	0.587	0.000	0.482	31	0.170	−0.493	0.572	6
Italy	0.000	−0.465	0.630	2	0.000	−0.721	0.710	3	0.065	−0.573	0.609	3
Latvia	0.080	−0.127	0.546	9	0.474	−0.522	0.585	21	0.855	−0.083	0.372	28
Lithuania	0.282	−0.130	0.498	21	0.688	−0.099	0.483	30	0.940	0.000	0.343	31
Luxembourg	0.000	−0.615	0.652	1	0.000	−0.194	0.645	9	0.000	−0.820	0.659	1
Malta	0.000	−0.124	0.567	7	0.000	−0.476	0.683	6	0.000	−0.740	0.648	2
Netherlands	0.239	−0.310	0.544	10	0.303	−0.313	0.589	19	0.412	−0.522	0.528	11
Poland	0.121	−0.103	0.530	13	0.464	−0.040	0.512	28	0.434	0.000	0.417	23
Portugal	0.263	−0.290	0.535	12	0.160	−0.503	0.647	8	0.566	−0.335	0.467	20
Romania	0.485	0.000	0.427	28	0.277	−0.379	0.605	15	0.531	−0.290	0.464	21
Slovakia	0.156	−0.257	0.533	8	0.033	−0.568	0.685	5	0.000	−0.309	0.580	5
Slovenia	0.221	0.000	0.481	25	0.266	−0.660	0.643	10	0.198	−0.137	0.497	17
Spain	0.351	−0.224	0.503	19	0.000	−0.448	0.679	7	0.026	−0.273	0.566	7
Sweden	0.647	0.000	0.400	29	0.418	−0.155	0.541	26	0.692	0.000	0.376	27
United Kingdom	0.476	−0.294	0.493	23	0.501	0.000	0.497	29	0.000	−0.086	0.533	9
Iceland	0.352	−0.211	0.500	20	0.362	−0.370	0.585	20	0.922	0.000	0.345	30
Norway	0.424	−0.405	0.523	14	0.491	−0.430	0.569	23	0.928	−0.227	0.393	25
Switzerland	0.058	0.000	0.522	15	0.212	0.000	0.558	25	0.000	0.000	0.512	16

Notes.$g^t = \left(g^t_x, g^t_y\right) = (\bar{x}^t, \bar{y}^t)$

scale inefficiency (see Appendix 1).[17] This also explains previous results obtained in the literature that suggested this relationship but left it unexplained (e.g. Zabala-Iturriagagoitia et al., 2007b; Edquist et al., 2018).

Based on this negative relationship we propose an instrumental characterization of national innovation systems depending on their ranking position in terms of innovation performance and innovation scale. Tabulating the information as presented in Fig. 2, which compares the rankings in innovation inputs (x-axis) and innovation performance (y-axis), results in the following four categories of countries[18]:

- High innovation inputs and high innovation performance (HIHP): France, Netherlands, Denmark, Germany, Austria.
- High innovation inputs and low innovation performance (HILP): Sweden, Finland, Switzerland, Estonia, Belgium, Iceland, Czech Republic, Ireland, United Kingdom, Norway, Slovenia.
- Low innovation inputs and high innovation performance (LIHP): Portugal, Luxembourg, Spain, Cyprus, Slovakia, Italy, Malta, Greece.
- Low innovation inputs and low innovation performance (LILP): Lithuania, Poland, Croatia, Hungary, Bulgaria, Romania, Latvia.

Fig. 2 serves to illustrate the difference between the classic evaluation paradigm associated to "the-more-the-better", corresponding to the HIHP and LILP quadrants located in the positive diagonal, and our new approach emphasizing the relevance of studying innovation performance, focusing also on the LIHP and HILP categories; i.e. the quadrants in the negative diagonal. Arguably countries belonging to these categories deserve further scrutiny by falling outside what is traditionally expected. In this regard numerous countries that invest the lowest amount of resources—and hence exhibit low aggregate values of innovation inputs— present a considerable innovation performance—as measured by the *RCII* index, thereby belonging to the LIHP category (e.g. Portugal, Spain, Greece, Italy, Slovakia, and Malta). However, not all countries with a low innovation input scale achieve a high innovation performance (e.g. Lithuania, Poland, Croatia, Hungary, Bulgaria, Romania, Latvia), simply because those resources do not materialize in innovation outputs (LILP). Similarly, High input scale-Low performance (HILP) countries show also that large innovation performance does not necessarily go hand in hand, *per se*, with a large size of the innovation system. Finally, as previously remarked, it is also possible to observe that some countries can reconcile both perspectives in their innovation system (HIHP), constituting the reference benchmarks of both the traditional and new evaluation paradigms. These would be the ones that could be regarded as comprehensive innovation systems, as they count both with a high innovation input scale and a high innovation performance.

Ultimately, these results justify the need to distinguish between the size of an innovation system, measured through the quantity of inputs available to the system, and its innovation performance. Although the identification of the size of an innovation system is important, how efficiently the resources available in it are being applied should be as important. The amount of innovation inputs may provide an indication of the actual level of development of an innovation system, namely, whether it is in an embryonic stage, it is evolving or is fully developed, with low or high conditions for the development, emergence diffusion

[17] Estimation results for each year are available in the Supplemental On-line Appendix.

[18] Fig. 2 shows that despite the general negative relation between performance and scale attested by the regression results, it cannot be excluded that some individual countries exhibiting high innovation scales can simultaneously achieve high levels of innovation performance (i.e., those in the lower left quadrant and characterized as HIHP).

and absorption of innovations. However, the latter complements the conclusions achieved from a single perspective as innovation scale, and hence increases the robustness of the policy recommendations that can be derived from the use of scoreboard indicators.

We would also like to exert caution when interpreting the new results based on the *RCII*. Although a country may be at the top of the ranking in terms of innovation performance, by employing relatively few resources in relation to the attained results (e.g. Italy, Greece), it does not mean that it constitutes *per se* a reference benchmark for wider objectives such as enhancing economic development, reducing the gap in growth rates, social welfare, etc., which fall out of the quantitative scope of the present analysis (see Fagerberg and Srholec, 2008). On the other hand, a country that invests huge amounts of resources in its innovation system (e.g. Sweden, Finland) but whose use of resources is identified as inefficient compared to the peer group of best practice countries, cannot be seen as an example of best practice either. In terms of our categorization, those countries showing high innovation inputs but low performance (HILP) would be required to study the innovation systems of benchmark countries investing high in innovation inputs while simultaneously achieving high levels of performance (HIHP).

4.3. Testing the robustness of the DEA-TOPSIS results

One of the potential critiques that may emanate from the above results is the pertinence of the previous 12 indicators to assess the performance of each and every innovation system. To address robustness concerns and potential limits associated to the use of a single baseline or 'standard model'–including the selection of the 4 inputs and 8 outputs indicators, we have run the DEA-TOPSIS model for all possible combinations of 21 of the 27 indicators in the EIS (see Landete et al., 2017; Hou et al., 2018). This implies that we solve all sensible models, starting with the simplest one charactering a barebone innovation system with minimum requirements, including at least 2 inputs and 3 outputs, up to a full model that considers 21 indicators of the EIS (i.e. 9 inputs and 12 outputs). Table 7 compares the indicators considered in the "baseline model" introduced by Edquist et al. (2018) with the models with the minimum and maximum requirements considered in this combinatory analysis. This implies that a total of 4096 combinations are solved per country and per year to test whether the previously reported results are statistically robust, and hence, whether the policy recommendations derived from them are sensible. As a result, these 4096 models contemplate all possible scenarios, given the EIS indicators, and hence, avoid the potential bias that the standard model may not fit well the innovation structure of some countries.

We calculate all 4096 *RCII*s for every country and year, yielding the same number of rankings. Fig. 3 presents a box plot of the distribution of ranking positions (1st thru 31st) of each country for all models for year 2016.[19] The box in each "candle stick" represents the interval between the first and third quartiles of the ranking distribution (i.e. the interquantile range – IQR –, between Q_1 and Q_3), with its median represented by the horizontal line within it. The upper and lower "whiskers" represent thresholds one and a half times above and below the interquantile range: $Q1-1.5 \times IQR$ and $Q3+1.5 \times IQR$). Any ranking position below and above these values is represented by an asterisk and can be regarded as outliers in the distribution. The dispersion in the rankings within the interquartile ranges is rather low in general and no country exceeds a distance greater than 10 positions (Estonia presents the highest spread with 10 positions, 16th-26th). Finally, the upper and lower whiskers signal ranking positions one and half times the IQR, with the dots above and below representing outliers laying beyond those values.

Luxemburg consistently ranks in first place and therefore there is no

[19] The results for years 2010 and 2013 are available upon request. These boxplots have not been included due to space constrains.

Table 7
Indicators considered in the robustness analysis.

	Baseline model	Model with minimum requirements	Model with maximum requirements
Inputs			
1.1.1 New doctorate graduates per 1000 population aged 25-34			X
1.1.2 Population aged 25–34 having completed tertiary education (% share)			X
1.1.3 Population aged 25–64 involved in lifelong learning (% share)			X
1.2.1 International scientific co-publications per million population			
1.2.2 Scientific publications among the top 10% most cited publications worldwide (% of total scientific publications of the country)			
1.2.3 Foreign doctorate students (% of all doctorate students)			
1.3.1 Broadband penetration			
1.3.2 Opportunity-driven entrepreneurship (Motivational index)			X
2.1.1 R&D expenditure in the public sector (% of GDP)	X		X
2.1.2 Venture capital (% of GDP)	X		X
2.2.1 R&D expenditure in the business sector (% of GDP)	X	X	X
2.2.2 Non-R&D innovation expenditures (% of turnover)	X	X	X
2.2.3 Enterprises providing training to develop or upgrade ICT skills of their personnel (% of all enterprises)			X
Outputs			
3.1.1 SMEs introducing product or process innovations (% of SMEs)	X	X	X
3.1.2 SMEs introducing marketing or organizational innovations (% of SMEs)	X	X	X
3.1.3 SMEs innovating in-house (% of SMEs)	X		X
3.2.1 Innovative SMEs collaborating with others (% of SMEs)			
3.2.2 Public-private co-publications per million population			
3.2.3 Private co-funding of public R&D expenditures (% of GDP)			X
3.3.1 PCT patent applications per billion GDP (in PPS)			X
3.3.2 Trademark applications per billion GDP (in PPS)	X		X
3.3.3 Design applications per billion GDP (in PPS)	X		X
4.1.1 Employment in knowledge-intensive activities (% of total employment)			X
4.1.2 Employment in fast-growing enterprises (% of total employment)			X

(*continued on next page*)

Table 7 (*continued*)

	Baseline model	Model with minimum requirements	Model with maximum requirements
4.2.1 Exports of medium and high technology products as a share of total product exports	X		X
4.2.2 Knowledge-intensive services exports as % of total services exports	X		X
4.2.3 Sales of new-to-market and new-to-firm innovations as % of turnover	X	X	X

the rank correlation coefficient between this model and the ranking based on the median position for the remaining 4096 combinations. As few countries have the same median ranking position we employ Kendall's tau definition. The coefficient $\tau = 0.7392$ shows that a positive and rather large correlation exists, which is statistically significant at the standard confidence levels.[21] We can thus conclude that the results reported and discussed for the baseline model generally hold, as the ranking positions fall within the confidence intervals presented in Fig. 3, and not far from the median values. Indeed, these values can be consistently considered as representative ranking positions for benchmarking analyses of innovation systems and policymaking.

5. Discussion

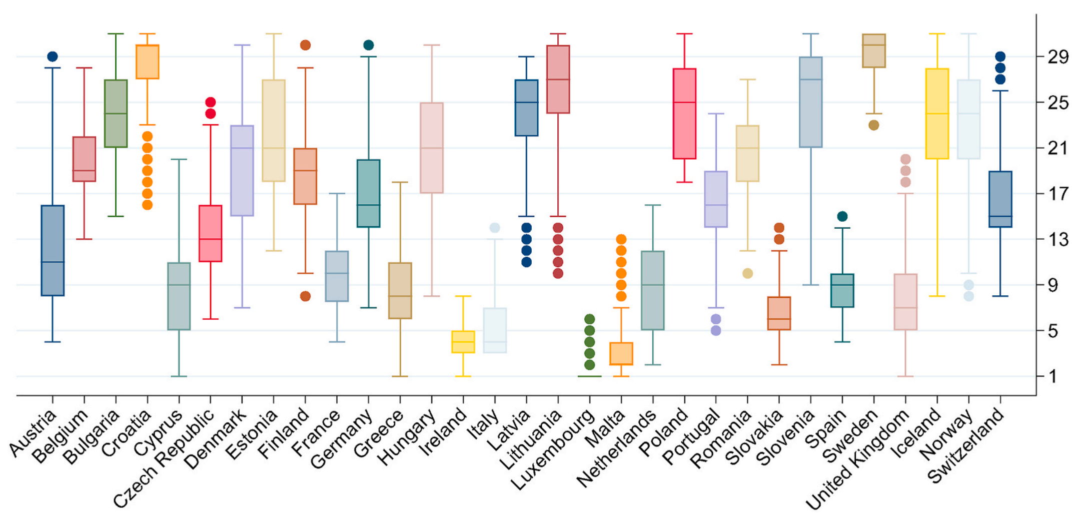

Fig. 3. Boxplots of the ranking distributions (from 1 to 31) for all 4096 models (year 2016).
Note: Boxplots summarizing the distribution of the ranking positions for each country. Each ranking position corresponds to one of the 4 096 chosen evaluation models that include alternative combinations of inputs and outputs (i.e., those between the minimum and maximum requirements models shown in Table 7).

spread between quartiles, but a few outliers below the 1st position. Some countries exhibiting comprehensive innovation systems that perform well under the baseline model with 12 indicators discussed in the previous section, i.e. "High Innovation Input and High Performance – HIHP" countries in Fig. 2, e.g. Netherlands and Denmark, always rank at the top, while Austria and Germany fare badly falling down to upper-middle positions for some models using less inputs and output indicators. However, the chart shows that considering our four group classification based on the preferred model and depicted in Fig. 2, results generally hold. Considering the opposite "Low Innovation Input-Low Performance – LILP" category, all countries belonging to this group rank in the worst positions, except the Czech Republic that improves its ranking to upper middle positions surpassing even Germany. Spain also behaves better than in the standard model. Within this group on the negative side, Croatia and Sweden are the two countries that perform the worst consistently, ranking between the 28th and 31st positions.[20]

In brief, we find that the relative position shown by most countries varies to a low extent, and while countries such as Sweden, Slovakia, Malta, Luxembourg, Italy, Spain or Ireland show really stable patterns across all models, in others, the variation is a bit larger, but still without entailing significant differences when compared to those previously reported for the base model. The reliability of the baseline model introduced by Edquist et al. (2018) can thus be confirmed by calculating

The policy evaluation related literature is largely in agreement about the need to combine different approaches, methodologies and indicators to avoid biased assessments of system/policy performance (Michelson, 2006). We also consider that the literature on innovation measurement should seek to combine different approaches, sometimes applying them to the same datasets, to avoid potential biases and thus, produce more credible policy recommendations. This paper aims to contribute in this respect by relating the amount of innovation inputs available to an innovation system and its innovation performance, through the notion of returns to scale. To achieve this goal we build upon Edquist et al. (2018), who leave this relevant question unexplored, but provide the conceptual framework for the evaluation of innovation performance. Hence, making use of production theory we can relate the size of the innovation system with its performance, calculate the nature of returns to scale, and show whether productivity gains can be realized by increasing the amount of innovation inputs invested in the system or

[20] Edquist et al. (2018) discuss the Swedish paradox in extent.

[21] We also compare our results with the efficiency scores reported by Edquist et al. (2018). In particular, we have calculated the correlation between the ranking associated to their bootstrapped efficiency scores calculated using the standard−input oriented−DEA formulation, and the ranking associated to the directional distance function (4) against the ideal benchmark hyperplanes. Thereby we only consider best performance in the comparison. Even though the assumption on returns to scale is different in both studies, constant and variable, Kendall's tau $\tau = 0.6798$, is again positive, large and statistically significant.

not. For this purpose we resort to robust Data Envelopment Analysis techniques (DEA-TOPSIS) that improve on existing DEA standard methods, as those employed by the previous authors.

Our results indicate that in many cases, countries often regarded as 'innovation leaders', using the labeling of the *EIS*, count with an innovation performance far from satisfactory. This suggests that innovation leading countries, while dedicating vast resources (i.e. high innovation inputs), often do not manage to produce as much output as it could be expected (i.e. low innovation performance), pointing to the existence of decreasing returns to scale in innovation. We confirm such hypothesis at the aggregate country level. This shows that the relative allocation of resources currently made in many countries does not yield the expected returns, which calls for a reconsideration of how innovation decisions are made in the policy field. Despite decreasing returns being observed in other production activities (see Madsen, 2007; Lang, 2009), innovation, and hence innovation policy, is still being regarded as a field in which the more resources invested, the better the relative performance of the innovation system is likely to be (Rodríguez Pose and Crescenzi, 2008). The evidence provided in this study shows otherwise; i.e. that the bigger is not always the better. Fragkandreas (2013) also discusses the existence of this paradox, in the sense that innovation seems not to pay-off for some territories in Europe, even highly innovative ones, as these grow at a slower pace than other peers. It is true that innovation may continue being the engine that provides the solutions to the grand challenges we are currently facing. However, the evidence provided by Strumsky et al. (2010) also raises some doubts as to whether innovation activities in developed economies will continue being as productive as they were in the past (see also Crafts, 2018).

These results might be explained by the complexity of innovation processes and the need to coordinate the activities articulated by innovation policies (Magro et al., 2014). Leading countries with large innovation systems in terms of invested inputs and a long tradition in the implementation of science, technology and innovation policies, tend to support new growth industries in the long-run, which imply higher risks in their innovation strategies. As a result, despite the innovation systems of these countries may be dynamic and with an innovation-friendly climate, the high levels of coordination required and the uncertainties involved reduce their levels of efficiency. We classify these countries in the "High Innovation Inputs-Low Performance" (HSLP) group – see Fig. 2.

On the contrary, countries with less developed innovation systems tend to absorb and adopt the embodied knowledge and the innovations of others, as their innovation systems do not count with the required conditions for such an effective generation and diffusion of in-house innovations. This strategy requires lower levels of development and hence lower input amounts in the operation of the system, while at the same time producing more efficient behaviors since risk is avoided and the 'new' knowledge is rapidly adopted. These countries are classified within the "Low Innovation Inputs -High Performance" (LIHP) group. It also needs to be noted that countries with fewer available resources have to pay much more attention to how these are being exploited. They cannot afford to squander the scarce resources dedicated to innovation activities, and as a result, their cautious behavior result in higher efficiencies.

Naturally, not all innovation systems share the same properties, as the strategies followed by the different countries in their innovation policies also differ. However, it seems reasonable to think that despite the different strategies followed, any policymaker would opt for an efficient allocation of their investments within the system in order to maximize the innovation outputs that are to be achieved with it. Related to this question and the alternative innovation strategies, another possible reason explaining innovation inefficiency is the existence of scope economies, by which output (or input) mixes in some countries might be sub-optimal when compared to those of the leading, best performing benchmarks.[22]

In all instances, and according to our evidence, it seems sensible to relate the amount of inputs invested in an innovation system and its performance, using this distinction to identify categories of innovation systems (see Fig. 2). This way, policymakers may consider the results of different and complementary analyses to obtain a wider picture of their respective innovation systems. From a quantitative perspective, we conclude that the approach followed by the *EIS*, the *GII*, and other approaches to measuring innovation and producing innovation-related rankings, thus offer a partial view of the actual state of development of any innovation system.

We conclude by acknowledging some limitations of our model and relevant research venues. As for the limitations, which can be raised also against the *EIS* and *GII*, we acknowledge that, as in any process of change, there are time lags between the investments made on the input side and achieving certain outputs. Accounting for these lags constitutes a challenging task due to the differences in the structural characteristics of every innovation system, and because, for example, radical innovations may require longer gestation periods than incremental innovations. As for relevant extensions related to our conclusions, further research could be developed to explore the existence of the above mentioned economies of scope (diversifications versus specialization) related to sub-optimal proportions in the usage of inputs and production of outputs. Also, an interesting extension would be the systematic identification of benchmarks for every country to provide policymakers with guidelines related to specific innovation models. Finally, the productivity analysis that we perform in levels (cross-section) could be complemented with the study of productivity trends through the so-called Malmquist productivity indices introduced by Färe et al. (1994). This will allow to determine long run trends in total factor productivity growth and identify its main drivers, particularly technological progress, efficiency change and the dynamic effects associated to returns to scale (Balk et al., 2020; Zabala-Iturriagagoitia et al., 2021).

Acknowledgments

The authors are indebted to the editor and three anonymous referees for their comments on an earlier draft. The authors acknowledge financial support from the Comunidad de Madrid Government, under grant H2019/HUM-5761 (INNJOBMAD-CM). Prof. José L. Zofío acknowledges financial support from the Spanish Ministry of Science and Innovation and the State Research Agency under grants EIN 2020-112260/AEI/10.13039/501100011033 and PID2019-105952GB-I00/AEI/10.13039/501100011033. Assoc. Prof. Jon Mikel Zabala-Iturriagagoitia acknowledges financial support from the Basque Government, Department of Education, Language Policy and Culture, under grant IT 885–16.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.technovation.2021.102314.

[22] For example, from an input perspective, either too much or too low relative investments in the public sector with respect to the private sector; from an output perspective, a sub-optimal balance between innovation activities carried out by SMEs with respect to larger firms, which in turn may hamper achieving larger export shares in foreign markets.

References

Acs, Z.J., Autio, E., Szerb, L., 2014. National systems of entrepreneurship: measurement issues and policy implications. Res. Pol. 43, 476–494.

Aparicio, J., Zofío, J.L., 2020. Economic cross-efficiency. Omega 100, 102374. https://doi.org/10.1016/j.omega.2020.102374.

Arundel, A., Hollanders, H., 2008. Innovation scoreboards: indicators and policy use. In: Nauwelaers, C., Wintjes, R. (Eds.), Innovation Policy in Europe. Measurement and Strategy. Edward Elgar, Cheltenham, pp. 29–52.

Balk, B.M., Koster, M.B.M., Kaps, C., Zofío, J.L., 2021. An evaluation of cross-efficiency methods: With an application to warehouse performance. Applied Mathematics and Computation 406, 126261. https://doi.org/10.1016/j.amc.2021.126261.

Banker, R.D., Charnes, A., Cooper, W.W., 1984. Some models for the estimation of technical and scale inefficiencies in data envelopment analysis. Manag. Sci. 30, 1078–1092.

Batterbury, S.C.E., 2006. Principles and purposes of European union cohesion policy evaluation. Reg. Stud. 40 (2), 179–188.

Balk, B.M., Barbero, J., Zofío, J.L., 2020. A toolbox for calculating and decomposing Total Factor Productivity indices. Comput. Oper. Res. 115, 104853.

Brenner, T., Broekel, T., 2011. Methodological issues in measuring innovation performance of spatial units. Ind. Innovat. 18 (1), 7–37.

Carayannis, E., Grigoroudis, E., 2014. Linking innovation, productivity, and competitiveness: implications for policy and practice. J. Technol. Tran. 39, 199–218.

Carayannis, E.G., Goletsis, Y., Grigoroudis, E., 2015. Multi-level multi-stage efficiency measurement: the case of innovation systems. Operational Research 15 (2), 253–274.

Castro-Martínez, E., Jiménez-Sáez, F., Ortega-Colomer, F.J., 2009. Science and technology policies: a tale of political use, misuse and abuse of traditional R&D indicators. Scientometrics 80 (3), 827–844.

Chambers, R.G., Chung, Y., Färe, R., 1996. Benefit and distance functions. J. Econ. Theor. 70, 407–419.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. Eur. J. Oper. Res. 2 (6), 429–444.

Chen, J.-X., 2012. A comment on DEA efficiency assessment using ideal and anti-ideal decision making units. Applied Economics and Computation 219, 583–591.

Cherchye, L., Moesen, W., Rogge, N., van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., Tarantola, S., 2008. Creating composite indicators with DEA and robustness analysis: the case of the technology achievement index. J. Oper. Res. Soc. 59 (2), 239–251.

Cooper, W.W., Ruiz, J.L., Sirvent, I., 2011. Choices and uses of DEA weights. In: Cooper, W.W., Seiford, L.M., Zhu, J. (Eds.), Handbook on Data Envelopment Analysis. Springer, New York, pp. 93–126.

Crafts, N., 2018. The productivity slowdown: is it the 'new normal'? Oxf. Rev. Econ. Pol. 34 (3), 443–460.

Dutta, S., Lanvin, B., Wunsch-Vincent, S., 2017. The Global Innovation Index 2017. Innovation Feeding the World, tenth ed. Available. https://www.globalinnovationindex.org/.

Dziallas, M., Blind, K., 2019. Innovation indicators throughout the innovation process: an extensive literature analysis. Technovation 80–81, 3–29.

Edquist, C. (Ed.), 1997. Systems of Innovations: Technologies, Institutions and Organizations. Pinter Publishers, London.

Edquist, C., 2005. Systems of innovation: perspectives and challenges. In: Fagerberg, J., Mowery, D., Nelson, R. (Eds.), The Oxford Handbook of Innovation. Oxford University Press, Oxford, pp. 181–208.

Edquist, C., 2011. Design of innovation policy through diagnostic analysis: identification of systemic problems (or failures). Ind. Corp. Change 20 (6), 1725–1753.

Edquist, C., Zabala-Iturriagagoitia, J.M., Barbero, J., Zofio, J.L., 2018. On the meaning of innovation performance: is the synthetic indicator of the Innovation Union Scoreboard flawed? Res. Eval. 27 (3), 196–211.

European Union, 2017. European innovation scoreboard 2017. Available. http://ec.europa.eu/growth/industry/innovation/facts-figures/scoreboards_en.

Fagerberg, J., Srholec, M., 2008. National innovation systems, capabilities and economic development. Res. Pol. 37 (9), 1417–1435.

Färe, R., Grosskopf, S., 2000. Theory and application of directional distance functions. J. Prod. Anal. 13 (2), 93–103.

Färe, R., Grosskopf, S., Norris, M., Zhang, Z., 1994. Productivity growth, technical progress, and efficiency change in industrialized countries. Am. Econ. Rev. 84, 66–83.

Fragkandreas, T., 2013. When innovation does not pay off: introducing the "European regional paradox". Eur. Plann. Stud. 21 (12), 2078–2086.

Fried, H., Lovell, C.A., Schmidt, S.S. (Eds.), 2008. The Measurement of Productive Efficiency and Productivity Growth. Oxford University Press, New York.

Fukuyama, H., 2003. Scale characterizations in a DEA directional technology distance function framework. Eur. J. Oper. Res. 144 (1), 108–127.

Furman, J.L., Porter, M.E., Stern, S., 2002. The determinants of national innovative capacity. Res. Pol. 31, 899–933.

Galli, R., Teubal, M., 1997. Paradigmatic shifts in national innovation systems. In: Edquist, C. (Ed.), Systems of Innovation: Growth, Competitiveness and Employment. Pinter Publishers, London, pp. 342–364.

Grupp, H., Mogee, M.E., 2004. Indicators for national science and technology policy: how robust are composite indicators? Res. Pol. 33, 1373–1384.

Grupp, H., Schubert, T., 2010. Review and new evidence on composite innovation indicators for evaluating national performance. Res. Pol. 39, 67–78.

Guan, J., Chen, K., 2012. Modeling the relative efficiency of national innovation systems. Res. Pol. 41, 102–115.

Hagedoorn, J., Cloodt, M., 2003. Measuring innovative performance: is there an advantage in using multiple indicators. Res. Pol. 32 (8), 1365–1379.

Hekkert, M.P., Suurs, R.A.A., Negro, S.O., Kuhlmann, S., Smits, R.E.H.M., 2007. Functions of innovation systems: a new approach for analyzing technological change. Technol. Forecast. Soc. Change 74, 413–432.

Hou, Q., Wang, M., Zhou, X., 2018. Improved DEA cross efficiency evaluation method based on ideal and anti-ideal points. Discrete Dynam Nat. Soc. 2018, 1604298.

Johnson, A., 2001. Functions in innovation system Approaches. In: Paper for DRUID's Nelson-Winter Conference. Aalborg, Denmark.

Jordan, G.B., 2010. A theory-based logic model for innovation policy and evaluation. Res. Eval. 19, 263–273.

Katz, J.S., 2016. What is a complex innovation system? PloS One 11 (6), e0156150.

Kou, M., Chen, K., Wang, S., Shao, Y., 2016. Measuring efficiencies of multi-period and multi-division systems associated with DEA: an application to OECD countries' national innovation systems. Expert Syst. Appl. 46, 494–510.

Landete, M., Monge, J.F., Ruiz, J.L., 2017. Robust DEA efficiency scores: a probabilistic/combinatorial approach. Expert Syst. Appl. 86, 1339–1351.

Lang, G., 2009. Measuring the returns of R&D—an empirical study of the German manufacturing sector over 45 years. Res. Pol. 38 (9), 1438-1445.

Lee, Y.-N., 2015. Evaluating and extending innovation indicators for innovation policy. Res. Eval. 24, 471–488.

Lovell, C.A.K., 2002. Performance assessment in the public sector. In: Fox, K.J. (Ed.), Efficiency in the Public Sector. Kluwer Academic Publishers, Dordrecht, pp. 11–35.

Lundvall, B.Å. (Ed.), 1992. National Systems of Innovation: towards a Theory of Innovation and Interactive Learning. Pinter Publishers, London.

Madsen, J.B., 2007. Are there diminishing returns to R&D? Econ. Lett. 95 (2), 161-166.

Magro, E., Navarro, M., Zabala-Iturriagagoitia, J.M., 2014. Coordination-mix: the hidden face of STI policy. Rev. Pol. Res. 31 (5), 367–389.

Mairesse, J., Mohnen, P., 2002. Accounting for innovation and measuring innovativeness: an illustrative framework and an application. Am. Econ. Rev. 92 (2), 226–230.

McKelvey, M., 1997. Using evolutionary theory to define systems of innovation. In: Edquist, C. (Ed.), Systems of Innovation: Growth, Competitiveness and Employment. Pinter Publishers, London, pp. 200–222.

Metcalfe, J.S., 1995. Technology systems and technology policy in an evolutionary framework. Camb. J. Econ. 19, 25–46.

Michelson, E.S., 2006. Approaches to research and development performance assessment in the United States: an analysis of recent evaluation trends. Sci. Publ. Pol. 33 (8), 546–560.

Morita, H., 2003. Analysis of economies of scope by data envelopment analysis: comparison of efficient frontiers. Int. Trans. Oper. Res. 10 (4), 393–402.

Nasierowski, W., Arcelus, F.J., 2012. About efficiency of innovations: what can be learned from the innovation union scoreboard index. Procedia - Social and Behavioral Sciences 58, 792–801.

Nelson, R.R. (Ed.), 1993. National Innovation Systems: A Comparative Analysis. Oxford University Press, New York.

Palmberg, C., 2006. The sources and success of innovations - determinants of commercialisation and break-even times. Technovation 26, 1253–1267.

Prajogo, D.I., Ahmed, P.K., 2006. Relationships between innovation stimulus, innovation capacity, and innovation performance. R&D Management 36 (5), 499–515.

Rodríguez Pose, A., Crescenzi, R., 2008. Research and development, spillovers, innovation systems, and the genesis of regional growth in Europe. Reg. Stud. 42 (1), 51–67.

Shen, W., Zhang, D., Liu, W., Yang, G., 2016. Increasing discrimination of DEA evaluation by utilizing distances to anti-efficient frontiers. Comput. Oper. Res. 75, 163–176.

Shephard, R.W., 1970. Theory of Cost and Production Functions. Princeton University Press, Princeton, New Jersey.

Soltanifar, M., Shahghobadi, S., 2014. Survey on rank preservation and rank reversal in data envelopment analysis. Knowl. Base Syst. 60, 10–19.

Strumsky, D., Lobo, J., Tainter, J., 2010. Complexity and the productivity of innovation. Syst. Res. Behav. Sci. 27, 496–509.

Wang, Y.-M., Luo, Y., 2006. DEA efficiency assessment using ideal and anti-ideal decision making units. Appl. Math. Comput. 173, 902–915.

Wang, Y.-M., Luo, Y., 2009. On rank reversal in decision analysis. Math. Comput. Model. 49 (5–6), 1221–1229.

Wu, D., 2006. A note on DEA efficiency assessment using ideal point: an improvement of Wang and Luo's model. Appl. Math. Comput. 183, 819–830.

Zabala-Iturriagagoitia, J.M., Jiménez-Sáez, F., Castro-Martínez, E., Gutiérrez-Gracia, A., 2007a. What indicators do (or do not) tell us about Regional Innovation Systems. Scientometrics 70 (1), 85–106.

Zabala-Iturriagagoitia, J.M., Voigt, P., Gutiérrez-Gracia, A., Jiménez-Sáez, F., 2007b. Regional innovation systems: how to assess performance. Reg. Stud. 41 (5), 661–672.

Zabala-Iturriagagoitia, J.M., Aparicio, J., Ortiz, L., Carayannis, E.G., Grigoroudis, E., 2021. The productivity of national innovation systems in Europe: catching up or falling behind? Technovation 102, 102215. https://doi.org/10.1016/j.technovation.2020.102215 (in press).