Assessing the robustness of composite indicators: The case of the Global Innovation Index

Khatab AlQararah (khatab.alqararah@gmail.com)

Research

Keywords: Composite Indicators, Global Innovation Index, Data Envelopment Analysis, Principal Component Analysis, Random Forests, Innovation performance

Posted Date: April 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1505335/v1

Version of Record: A version of this preprint was published at Journal of Innovation and Entrepreneurship on September 18th, 2023. See the published version at https://doi.org/10.1186/s13731-023-00332-w.

Abstract

This research paper introduces a methodology to assess the robustness of the Global Innovation Index (GII), by comparing the rankings provided in it with those achieved using alternative data-driven methodologies such as Data Envelopment Analysis (DEA) and Principal Component Analysis (PCA). With it, the paper aims to reduce the level of subjectivity in the construction of composite indicators with regard to weight generation and indicator aggregation. The paper relies on PCA as a weighting-aggregation scheme to reproduce the 21 sub-pillars of the GII before the application of DEA to calculate the relative efficiency score for every country. By using the PCA-DEA model, a final ranking is produced for all countries. The paper uses Random Forests (RF) classification to examine the robustness of the new rank. The comparison between the new rank and that of the GII suggests that the countries positioned at the top or the bottom of the GII rank are less sensitive toward the modification than those in the middle of the GII, the rank of which is not robust against the modification of the construction method. The PCA-DEA model introduced in this paper provides policymakers with an effective tool to monitor the performance of national innovation policies from the perspective of their relative efficiency. Ultimately, this would enhance the innovation performance of their respective countries.

1. Introduction

For long, continuous efforts have been deployed to improve Composite Indicators (CIs) as a tool for measuring national innovation systems' performance (Bandura, 2011; Dutta & Lanvin, 2013; Edquist et al., 2018; Corrente et al., 2021; Alnafrah, 2021). Innovation, according to the Oslo Manual (OECD/Eurostat, 2018, p. 20) is defined as 'new or improved product or business process (or combination thereof) that differs significantly from the firm's previous products or business processes and that has been introduced on the market or brought into use by the firm'. In this regard, national innovation systems represent the combination and the harmony of several facilitators and parties such as institutional structures, infrastructures, or any supporting activities and policies that orchestrate together to facilitate and create an appropriate environment to foster innovation (Lundvall, 2007). The need for innovation and knowledge-driven growth is no longer exclusively for developed countries, as developing countries are also acting to design policies aiming to enhance their innovation performance to achieve better economic and environmental growth (Broughel & Thierer, 2019; Nikolaidis et al., 2013).

CIs represent the aggregation of a set of individual indicators into a single index, that aims to measure a multi-dimensional concept (OECD, 2004). In the last decades, a growing trend has emerged among policymakers, media, and all areas of research regarding CIs and the rankings derived from them – examples of CIs include education, health, government quality, innovation – (Bandura, 2011; Hatefi & Torabi, 2010; Greco et al., 2019; Barbero et al., 2021; Crespo & Crespo, 2016). These ranks have gained a significant impact on the perceived image of how decision-making units (e.g., countries, regions, cities, universities) are relatively performing (Cherchye et al., 2008; Zabala-Iturriagagoitia et al., 2007a). Not only do they influence people's desire to visit or to invest in these countries, but also the legitimacy of the policies adopted in these countries to support the development of certain activities (i.e., education, economy, health, etc.). In the case of innovation systems, increasing efforts have been devoted to developing tools and indicators to assess economic and political potential, in order to provide countries with scientific evidence that helps them to develop policies and activities towards the development, diffusion, and adoption of innovation (Edquist, 2011), and CIs play a fundamental role in such endeavors. Nevertheless, several concerns emerge from this general context, namely: how much does the ranking provided by a CI represent the actual performance of a particular decision-making unit? And how robust are the results against the methodology used in such CIs? This paper aims to contribute to clarifying these research questions, by focusing on the Global Innovation Index (GII).

The literature has evidenced, how the underlying methodology followed in the construction of CIs has a direct influence on its final results and rankings (Zabala-Iturriagagoitia et al., 2007a; Grupp & Schubert, 2010; Edquist et al., 2018; Greco et al., 2019; OECD, 2008). However, the CIs used to measure such issues as innovation, education, health, etc. have remained blind to this evidence. The examination of more methodologies to assess whether the results of CIs change depending on the way in which data is processed, weighted, or aggregated is referred to as the robustness of the CI (OECD, 2008). Out of the multiple CIs that have been developed to assess innovation potential, this paper examines the robustness of the 2019 edition of the GII. To do that, this paper applies alternative data-driven techniques such as Data Envelopment Analysis (DEA), Principal Component Analysis (PCA), and Random Forests (RF) to produce a new rank for the GII countries to examine the robustness of the GII.

The structure of this paper can be summarized as follows: section 2 presents the relevance of the GII and its main characteristics, as well as the main concerns about its structure and methodology. Moreover, it highlights the need to combine different perspectives to deal with the complexity of a phenomenon such as innovation. Section 3 presents the main methodological challenges that face the construction of CIs. Section 4 discusses the methodology of the new model (PCA-DEA) introduced in this paper. Section 5 presents the results obtained by the PCA-DEA model, as well as the explanation of the outcomes. Also, it conducts a comparison between the GII rank and the rank of the new model. Section 6 discusses the main conclusions and the contribution of the paper.

2. The Global Innovation Index And The Need For Different Perspectives

The GII is a global comparative study that involves 129 countries using detailed evidence of 80 indicators in the 2019 edition, accounting for around 95% of the global Gross Domestic Product. It provides a tool to measure national innovation performance using two sub-indices, innovation input, and innovation output. There are five pillars under the input sub-index, which consists of Institutions, Human Capital and Research, Infrastructure, Market sophistication, and Business sophistication, while the innovation output sub-index consists of two pillars, Knowledge and Technology Output and Creative Output. Under each pillar, there are several sub-pillars and then under each sub-pillar, there are several indicators (Dutta et al., 2019, p. 307).

The GII represents a significant source of insight about countries' levels in relation to sustainable innovation and aims to evaluate the environment supporting innovation at the national level. At the same time, it also helps to determine conditions for the diffusion of innovation and its importance for a country's development (Dutta et al., 2019). The GII is distinguished from the other composite innovation indices, as it emphasizes various components related to intangible assets, such as trademarks, business models, and industrial designs (Dutta & Lanvin, 2013). Owing to this feature, the GII can be used as a leading reference for researchers, business executives, and policymakers toward creative intangible asset-related innovation.

The GII is not designed to be the ultimate tool to reinforce rankings with respect to economies and countries. Rather, it provides a foundation to be able to continually evaluate the factors of innovation which can provide much insight for economies. In its 2019 edition, it provided a rich database of detailed metrics to redefine innovation policies, which are imperative for policymakers. There are many layers in measuring and understanding innovation, due to its multidimensional nature, which can help identify core and best practices and focus on holistic policies (Decancq & Lugo, 2013; Zabala-Iturriagagoitia et al., 2007a; Edquist et al., 2018). The intricate data allows for economies to monitor performance over time and thus standardize developments concerning other economies, allowing comparisons and the identification of best practices, which further supports the aim of the GII. Additionally, the GII 2019 report highlights that it should be "regarded as a sound attempt, matured over 12 years of constant refinements, to pave the way for better and more informed innovation policies worldwide" (Dutta et al., 2019, p. 387). With this is in mind, it is important to recognize that the GII is not the be-all and end-all in rankings within innovation, but a resource that can be used to assist countries in policymaking.

The GII has persistently preserved a similar methodology in all the editions published over the last 10 years, with very slight changes in the weights attributed to its indicators. As a result, it has presented consistency in the results, yielding similar rankings with very small variations across countries. In particular, the aggregation methods followed by the GII are the arithmetic average and weighted average. Also, it attributes the rationality of assigning the weights only for "statistical coherence" and "highest correlation" between indicators, sub-pillars, and pillars (Dutta et al., 2019, pp. 371–373). In the construction of every CI, assigning weights is an imperative part of the process, and the level of subjectivity should be treated with caution since it has a bearing consequence on the final results (Greco et al., 2019). In the case of the GII methodology, the allocation and production of weights imply several concerns within the methodology. First, the GII reveals how weights have been generated at the level of the seven pillars, but it has withheld how weights have been generated to its indicators and sub-pillars. Second, the 2019 edition of the GII specifically adjusted the weights for three indicators to provide "statistical coherence". Third, it needs to be taken into consideration that five of the indicators included in the 2019 edition were collected by subjective means such as using qualitative data, i.e., stakeholder interviews, which might unravel multiple layers of subjectivity (Dutta et al., 2019). Fourth, in the process of data aggregation, the arithmetic average has been used in multiple layers, which is an unreliable statistical tool to employ to combine large amounts of data, since the arithmetic average disregards any significance of certain variables over the others. Interestingly, the GII does not indicate any methodological rationale for using the arithmetic average in the process of aggregation. Furthermore, when aggregating 80 indicators by using averages, there is a high probability that two or more indicators are correlated, creating an additional bias in the results. Retrospectively, duplicating weights may distort the soundness of the outcomes. Considering these methodological concerns, the application of methodologies such as PCA is highly recommended to produce these relative weights (Adler & Yazhemsky, 2010), as it can help to solve this problem, by grouping indicators along with a set of consistent components, each being given a relative weight.

As previously argued, the GII has presented stability in their rankings for a long period of time, which may not aid in dynamically redefining policies, as the challenges facing innovation systems evolve over time. As things go, it is important to recognize that some countries that follow the direction of the GII, may reap the benefits of its ranking, but this does not necessarily aid the economic growth of that country, as the GII paints all countries with the same brush, disregarding each country on its own merits (Jankowska et al., 2017; Crespo & Crespo, 2016). Therefore, the literature has already tested new methods for data processing within the GII. For instance, Cui et al. (2020) used RF and Artificial Neural Networks to recalculate the GII scores for years 2019 and 2020 by using only 14 indicators. Along the same line, Pence et al. (2019) estimated the GII countries' scores, using the Artificial Neural Networks by choosing 27 indicators of the GII 2016. Consequently, the possibility to approximate the GII countries' score by using data-driven approaches with fewer parameters may help countries to monitor the change of their innovation performance in a shorter cycle.

The use of unweighted averaged between the GII sub-indices to calculate the general index hosts several issues. Firstly, it needs to be considered that the output sub-index disregards the number of innovation inputs utilized. Secondly, it does not seem appropriate to weight the contribution of the output sub-index equally to the contribution of the input sub-index, since the number of the GII pillars on both sides is not equal. Furthermore, the use of unweighted average to aggregate the data at the level pillars/sub-indices, unnecessarily suggests the same significance for all of them within all countries. Accordingly, implying that all countries can be clustered in one big group measured in the same way may yield an unfair scale (Stavbunik & Pelucha, 2019). Additionally, the GII has numerous missing data points, and hence, estimating or imputing these data points will lead to higher precision in results (Cui et al., 2020; Omer et al., 2020).

The GII produces efficiency scores. However, in determining the GII data structure (input and output), the efficiency of an innovation system can be paired with productivity i.e., the amount of output a system can generate using a certain amount of input (Jankowska et al., 2017). Hence, in order to compare the efficiency of two systems, it is necessary to measure the output that can be generated by utilizing the same input or less (Dutta et al., 2019). In stipulation to this notion, there is always room to improve the GII methodology by using data-driven techniques such as DEA, which seems a very appropriate technique to measure the efficiency of a national innovation system (Omrani et al., 2019; Barbero et al., 2021; Hatefi & Torabi, 2010; Alnafrah, 2021; Zabala-Iturriagagoitia et al., 2007b), given that the generation of the weights by using the DEA is taking place without prior intervention, and the efficiency score for each country is relative to all the other countries.

3. Methodological Challenges In The Construction Of Composite Indicators

CIs play an essential role in formulating the shape of the innovation policies and the awareness of them (Edquist et al., 2018). Meanwhile, it is crucial to set the characteristics of the policies that can fulfill the needs of the system (Zabala-Iturriagagoitia et al., 2007a). Therefore, it is alarming to accept and use them simplistically without any scrutiny and discussion (Grupp & Schubert, 2010). Accordingly, there are at least three fundamental challenges that need to be considered when creating CIs: (1) the relative weights associated with the indicators included in it, (2) the methodology used in the aggregation of these indicators, and (3) the robustness of its results (Munda, 2012; OECD, 2008; Greco et al., 2019; Freudenberg, 2003; Saisana et al., 2005). The misuse of the first two factors has a direct impact on the robustness of the final outcomes.

3.1) Composite indicator weighting

In the process of constructing a CI, selecting or developing the most sensible weighting scheme is critical, due to the strong effect of the weights on the final ranking (Greco et al., 2019; Yang et al., 2018). For example, in the case of the Technological Achievement Index introduced by Desai et al. (2002), changing the weights of some indicators reflected noticeable changes in the overall ranking with its consequent (political) implications (Cherchye et al., 2008). According to the OECD (2018), a "weight" represents a coefficient associated with a certain variable in the construction of CIs. In other words, weights represent the contribution of a variable to the overall CI, or to the sub-indices that constitute that overall CI. Therefore, the selection of the weights represents a major challenge in the process of constructing CIs. This challenge is frequently referred to as the 'Index problem' (Freudenberg, 2003; Cox et al., 1992).

To address this challenge a range of weighting schemes have been developed in the literature. For instance, (1) "No or equal weights" suggests that equal weights are assigned for all indicators, which is equivalent to saying that no weights are assigned. This is often labeled in the literature as an Attributes-Based Weighting System (Freudenberg, 2003). This approach has been among others used by the European Innovation Scoreboard (European Commission, 2019). However, this technique neglects the variation of the relative importance among indicators. Another solution is to adopt (2) Budget Allocation Processes. In this scheme, a set of 'n' points are given to a group of experts so they can distribute them across a group of indicators (Moldan & Billharz, 1997). However, a problem of inconsistency might occur if the number of indicators is larger than 10, or if the group of experts is not carefully selected (Saisana & Tarantola, 2002). This scheme is for example used in the estimation of the weights used in the e-Business Readiness Index introduced by Pennoni et al. (2005). The critical aspect of this scheme is that it involves a high level of subjectivity. Another alternative is to (3) adopt data-driven weight-assigning techniques. In this case, and contrary to the previously mentioned weighting schemes that contain a high level of subjectivity in the arbitrariness of weights selection, statistical techniques such as PCA and DEA are claimed to be desirable as they entail a high level of objectivity in the decision making (Decancq & Lugo, 2013; OECD, 2008).

3.2) Composite indicator aggregation

As argued above, characterizing and measuring complex phenomena through a 'simplistic' CI may lead to flawed results and conclusions. An alternative would be to develop several indices that would measure the same phenomenon from different perspectives. However, this alternative would increase the complexity of the interpretation of the results, particularly for non-specialized audiences such as the media, or policymakers. They would rather rely on a single number that includes all the indicators, which provides a simple understanding of a complex phenomenon even if this conclusion may be biased and hence lead to wrong (policy) implications (Saltelli, 2007). As a result, the way the CI has been produced is subject to debate. On the other hand, the discussion of the extent to which a CI does genuinely represent the complexity of the phenomenon is also contested (Greco et al., 2019).

An obvious debate emerges in the literature between aggregators versus non-aggregators. Aggregators support building synthetic indices to describe a whole (complex) phenomenon, by combining indicators using a certain aspect to produce a bottom line, which will result in a meaningful outcome (Cobb et al., 1995; Osberg & Sharpe, 2002; Gadrey & Jany, 2003). In turn, non-aggregators consider that the previous aggregation results will be statistically meaningless, and the process must stop at the level of having a set of indicators without combining them to a single outcome, their objection to aggregation being the arbitrariness of the process of weighting and combining (Henderson, 1974; Atkinson et al., 2002). Nevertheless, it is worth stating that most of the widespread indices such as the Human Development Index (UNDP, 2019) adopt a methodological framework that uses aggregation.

Aggregation is the last step in the process of constructing a CI. According to the OECD (2008) aggregation methods can be broken down into three different categories: linear, geometric, and multi-criteria. Moreover, there is yet another categorization of aggregation to be considered, namely, 'compensatory' or 'non-compensatory' (Munda, 2005). Compensatory approaches occur when there is a trade-off between two perceptions of weights (Paruolo et al., 2013). Understandably, this trade-off might cause a fixed compensability between pairs of dimensions. This could happen when one of the dimensions might cause a loss for another (OECD, 2008). In this context, Munda (2012) argues that in a hypothetical sustainability index, the dimension of economic growth could compensate for the loss in the environmental dimension. To conclude, the Linear or Additive utility-based approach is the most frequently used among the approaches of compensatory aggregation (Saisana & Tarantola, 2002).

The non-compensatory multi-criteria approach is less frequently used, due to the simplicity of applying compensatory approaches (Greco et al., 2019). Nevertheless, the Condorcet-Kemeny-Young-Levenglick non-compensatory approach is regarded as a plausible alternative instead of the frequently practiced linear aggregation. This approach has been applied to the Environmental Sustainability Index introduced by Esty et al. (2005), and it has given remarkable differences in the rankings produced by the original compensatory approach (Munda & Nardo, 2005).

3.3) Robustness of composite indicator

The robustness analysis aims to guarantee the quality and authenticity of a CI during the different levels of its construction, such as the theoretical and methodological framework development (Saisana et al., 2005; Corrente et al., 2021). Therefore, in the absence of robustness analysis, CIs may draw conclusions that would deliver a misleading message to the audience (Billaut et al., 2010; Saisana et al., 2005; Barbero et al., 2021; Corrente et al., 2021). However, the significance of robustness analysis is often neglected by most widespread CIs (OECD, 2008). To assure a high level of robustness in the image reflected by the CI, some techniques such as uncertainty analysis or sensitivity analysis should be used (OECD, 2008).

Uncertainty analysis refers to the magnitude of changes that may occur in the final outputs (i.e., conclusions) of a CI as a result of changes in the construction stages, such as weights assigning or aggregation (Greco et al., 2019; Grupp & Schubert, 2010; OECD, 2008). In turn, sensitivity analysis deals with how much variance of the final outcomes is yielded due to these uncertainties (Saisana et al., 2005; Grupp & Mogee, 2004). Hereby, it can be concluded that there is paramount demand to assess the influence of the methodological issues amid the construction of CIs. One of the methods that can be beneficial to examine the robustness of CIs, is the classification of the RF. Such classification can be an indicator of the level of a unit (country) ranking.

4. Methodology

4.1) The dataset

This research uses the dataset provided by the GII report 2019 (Dutta et al., 2019), which consists of 80 indicators (53 inputs and 27 outputs) collected from 129 countries. The data is available only in PDF format, so we had to transfer it into an MS-Excel sheet manually. It is important to acknowledge that the reliability of the data combines with the element of being presented by well-known international agencies (WIPO, INSEAD, Cornell University) and recognized by institutes and established bodies such as the UN Economic and Social Council as a reliable benchmark for measuring innovation (WIPO, 2021). Furthermore, GII indicators are derived from distinguished socio-economic data such as government effectiveness, tertiary education enrolment, global R&D companies' expenditure, and ICT usage.

When analyzing the data of the 2019 GII, there are only sixteen indicators without missing data for all countries. The remaining indicators have different percentages of missing data points, where the highest percentage of missing data for one indicator (High-tech net exports) is 51.2%. Among the 64 indicators with missing data, 27 indicators have less than 5% missing data points, 11 indicators have 5–10% missing data points, 12 indicators have between 10–20% missing data points, 7 indicators have 20–30% missing data points, and the remaining 7 indicators have more than 30% missing data points.

To deal with the missing data, several steps have been taken. Firstly, the countries have been divided into four groups based on the level of income: High-, Upper-middle-, Lower-middle-, and Low-income according to the World Bank (Dutta et al., 2019). Following this, the mean of each indicator was calculated for each group. Secondly, the countries have been divided into four groups again based on the Human Development Index (UNDP, 2019). Following this, the mean of each indicator was calculated for each group. Thirdly, each country was grouped with the five nearest neighboring countries, then the mean of each indicator was calculated for each group. Finally, the mean of the three means for each indicator was taken as an estimation for the missing data point. However, the previous steps did not solve the problem entirely, leaving the data with 40 missing data points, which were finally imputed by linear regression modeling.

4.2) Data Envelopment Analysis

DEA is a linear nonparametric programming model developed by Charnes et al. (1978) in the field of operations research, which is often referred to as the CCR Model. The idea behind this model is to evaluate the relative efficiency for homogenous Decision-Making Units (DMUs) (i.e., companies, banks, universities, countries, etc.) by giving them scores between 0 and 1. Specifically, DMUs with a score = 1 are considered as "efficient" and DMUs with a score < 1 are considered as "inefficient". In the 1980s, Banker et al. (1984) extended the method to develop a model that deals with multiple inputs and several outputs, which the literature refers to as the BCC-Model. DEA relies on a frontier created from the observed DMUs by utilizing the so-called best-practice, based on the minimum extrapolation principle (Thanassoulis, 2001).

Shen et al.(2013) conclude that DEA provides several features to the field of CIs such as: (1) it is a means to combine multiple indicators for countries without any prior awareness about the tradeoffs (i.e., the weights), (2) the country itself obtains its own best possible indicators weights, (3) if a country is underperforming compared to other countries, this cannot be attributed to the unfair weighting scheme, since every country has been put in its most beneficial position vis a vis all the other countries, (4) and any other weighting scheme would have generated lower weighting scores for that particular country. Additionally, DEA evaluates the relative efficiency of every country, taking into consideration the performance of all other countries (Cherchye et al., 2008). For the above-mentioned features, DEA has been broadly utilized to examine CIs to name but a few: Technology Achievement Index (Cherchye et al., 2008), the Macro-economic Performance Index (Ramanathan, 2006), the Human Development Index (Despotis, 2005) and the Knowledge Economy composite indicator (Guaita Martínez et al., 2021).

In the context of CI construction, the literature has broadly suggested an adjustment for the classic DEA formulation by considering all the indicators to be treated as outputs (Hermans et al., 2008; Cherchye et al., 2008; Martin et al., 2017; Guaita Martínez et al., 2021). This adjustment is known as the "Benefit of doubt" approach (Cherchye et al., 2008), and it shifts all input variables to become outputs, compromising the inputs with a dummy variable equal to one. It was initially adopted by Melyn & Moesen (1991) as a method to construct CIs to evaluate macroeconomic performance. This approach is to be considered if the underlying structure of the evaluated composite phenomenon is not definitive or if there is disagreement regarding the construction methodology, or if the input indicators are considered to be "achievements" (Cherchye et al., 2007). All these concerns are valid for any CI that endeavors to measure innovation performance. For example, Crespo & Crespo (2016) by applying a fuzzy-set qualitative comparative analysis conclude that none of the GII input pillars is a necessary condition for anticipating high innovation performance. Meanwhile, in the high-income countries, only two of the pillars (Infrastructure and Human capital and research) are sufficient to secure better innovation performance. Over and above, Jankowska et al. (2017), Edquist et al. (2018), and Barbero et al. (2021) among others, evidence that the common assumption that the higher GII input indicators, the higher GII output indicators is not confirmed.

For the above-stated, this paper relies on the DEA, using the benefit of doubt approach (see Eq. 1), with one dummy input variable equal to 1, and 21 output variables. Particularly, these 21 output variables will be generated by considering the linear combination of the indicators under each sub-pillar of the GII (i.e.

$V_q = (u_1.I_1 + u_2.I_2 + \cdots + u_n.I_n)$, I_n= indicators under sub-pillar $V_q, q = (1, \ldots, 21), n =$ number of indictors under each sub-pillar, u_n= the weight generated for indicator n by using the PCA one component loadings for the sub-pillar that the indicator belongs to. Eventually, the linear combination of indicators under the sub-pillar Political environment will produce variable 1, and the linear combination of indicators under the sub-pillar Regulatory environment will produce variable 2, etc. (See Table 1). This PCA-DEA approach has been introduced by (Adler & Yazhemsky, 2010).

$$I^{*} = \quad I^{*}(w) = \max_{I_K,\ k \in \{1,\ldots,M\}} \sum_{q=1}^{Q} I_{qk}\ w_q \qquad \text{Equation (1).}$$

$$CI_c^{*} = \quad \max_{w_{qc}} \sum_{q=1}^{Q} I_{qc}\ w_{qc} \qquad \text{Equation (2).}$$

$$S.t.$$

$$\sum_{q=1}^{Q} I_{qk}\ w_{qk} \ \leq 1$$

$$w_{qk} \geq 0$$

$$\forall\, k = 1, \ldots, M\ ; \forall\, q = 1, \ldots, Q$$

Where I_{qc} is the normalized value of the qth individual variable $(q = 1, \ldots, Q)$ for the country $c\ (c = 1, \ldots M)$ and w_{qc} the corresponding weight (Cherchye et al., 2004). Whilst I^{*} is the “benchmark performance” (i.e., the hypothetical country that maximizes the overall performance (OECD, 2008)).

However, due to the nature of the DEA, all efficient countries will obtain the same efficiency score equal to one (i.e., DMUs that lay at the frontier). Consequently, at least for these countries, the ultimate desired ranking will not be entirely discriminating. This limitation of DEA is known in the literature as the “discrimination power problem” (Adler & Yazhemsky, 2010; Barbero et al., 2021; Hatefi & Torabi, 2010). To address this problem, a sequence of sub-DEAs will be executed over the efficient countries only, by dividing the output variables for these countries into subsets according to the GII pillars. For example, the first sub-DEA will be performed over the output variables: Political environment, Regulatory environment, and Business environment, with a dummy input equal to 1. This will be repeated seven times for the seven pillars. Finally, the total efficiency score for each country will be the average of the seven sub-DEAs scores.

Table 1
Output variables.

Sub-pillar	Weight	Variable	Sub-pillar	Weight	Variable
Political environment		**V1**	**Trade, competition, & market scale**		**V12**
Political and operational stability	0.969		Applied tariff rate, weighted avg.	0.716	
Government effectiveness	0.969		Intensity of local competition	0.809	
			Domestic market scale, bn PPP$	0.552	
Regulatory environment		**V2**			
Regulatory quality	0.964		**Knowledge workers**		**V13**
Rule of law	0.958		Knowledge-intensive employment, %	0.914	
Cost of redundancy dismissal, salary	0.478		Firms offering formal training, % firms	0.332	
			GERD performed by business, % GDP	0.817	
Business environment		**V3**	GERD financed by business, %	0.831	
Ease of starting a business	0.842		Females employed w/advanced degrees	0.883	
Ease of resolving insolvency	0.842		**Innovation linkages**		**V14**
			University/industry research collaboration	0.92	
Education			State of cluster development	0.873	
Expenditure on education, % GDP	0.488	**V4**	GERD financed by abroad,	0.098	
Government funding/pupil, secondary	0.453		JV-strategic alliance deals/bn PPP$ GDP	0.624	
School life expectancy, years	0.894		Patent families 2 + offices/bn PPP$ GDP	0.76	
PISA scales in reading, maths, & science	0.881		**Knowledge absorption**		**V15**
Pupil-teacher ratio, secondary	0.761		Intellectual property payments	0.718	
Tertiary education		**V5**	High-tech imports, % total trade	0.343	
Tertiary enrolment, % gross	0.807		ICT services imports, % total trade	0.514	
Graduates in science & engineering, %	0.725		FDI net inflows, % GDP	0.736	
Tertiary inbound mobility, %	0.437		Research talent, % in business enterprise	0.623	
Research & development		**V6**	**Knowledge creation**		**V16**
Researchers, FTE/mn pop	0.915		Patents by origin/bn PPP$ GDP	0.82	
Gross expenditure on R&D, %	0.942		PCT patents by origin/bn PPP$ GDP	0.871	
GDP Global R&D companies, avg. exp. top 3	0.927		Utility models by origin/bn PPP$ GDP	0.325	
QS university ranking, average score top 3	0.881		Scientific & technical articles/bn PPP$ GDP	0.559	
			Citable documents H-index	0.798	
(ICTs)		**V7**			
ICT access	0.914		**Knowledge impact**		**V17**
ICT use	0.917		Growth rate of PPP$ GDP/worker	0.074	
Government's online service	0.928		New businesses/th pop. 15–64	0.409	
E-participation	0.918		Computer software spending, % GDP	0.71	
			ISO 9001 quality certificates/bn PPP$ GDP	0.691	
General infrastructure		**V8**	High- & medium-high-tech manufactures,	0.801	
Electricity output, kWh/mn pop	0.849		**Knowledge diffusion**		**V18**
Logistics performance	0.853		Intellectual property receipts	0.731	
Gross capital formation, % GDP	0.117		High-tech net exports, % total trade	0.377	
			ICT services exports, % total trade	0.697	
Ecological sustainability		**V9**	FDI net outflows, % GDP	0.776	

Sub-pillar	Weight	Variable	Sub-pillar	Weight	Variable
GDP/unit of energy use	0.646				
Environmental performance	0.885		**Intangible assets**		**V19**
ISO 14001 environmental certificates	0.692		Trademarks by origin/bn PPP$ GDP	0.496	
			Industrial designs by origin/bn PPP$ GDP	0.553	
Credit		**V10**	ICTs & business model creation	0.899	
Ease of getting credit	0.726		ICTs & organizational model creation	0.881	
Domestic credit to private sector, % GDP	0.77				
Microfinance gross loans, % GDP	0.257		**Creative goods & services**		**V20**
			Cultural & creative services exports	0.834	
Investment		**V11**	National feature films/mn pop	0.607	
Ease of protecting minority investors	0.645		Entertainment & Media market/th pop	0.621	
Market capitalization, % GDP	0.61		Printing & other media, % manufacturing	0.63	
Venture capital deals/bn PPP$ GDP	0.738		Creative goods exports, % total trade	0.082	
			Online creativity	0.816	**V21**
			Generic top-level domains (TLDs)/th pop	0.792	
			Country-code TLDs/th pop	0.878	
			Wikipedia edits/mn pop	0.709	
			Mobile app creation/bn PPP$ GDP	0.816	

4.3) Random Forests

RF is a non-parametric supervised learning statistical method, introduced by Breiman (2001). It has been proved to be a reliable method for classification problems (Hastie et al., 2009; Hamidi & Berrado, 2018). RF develops a random bootstrap of a set of data, performing multiple decision trees according to identified features (variables), eventually by so-called 'Bagging' to vote for the best classification (Hastie et al., 2009). The relationship between CIs and RF emerged recently in the fields of Data mining and Machine learning, to examine the robustness of the classifiers (Setiawan et al., 2019). In this paper, the idea behind the use of RF is to assure the robustness of the PCA-DEA results, by using the 21 variables in Table 1 to classify the countries and see to what level this classification matches the PCA-DEA results. Another quality RF can provide, is the ability to assess the 'importance' of every variable in the production of the classification (i.e., what are the variables that played an effective role in the classification of the countries?).

Before running the RF, all countries have been divided into three groups: (1) Countries in the top quartile of the PCA-DEA ranking and labelled as 'Efficient'. (2) Countries in the lowest quartile of the PCA-DEA ranking and labelled as 'Highly inefficient'. (3) Countries in the two remaining middle quartiles labelled as 'Inefficient".

5. Results

5.1) Principal component analysis

The purpose of PCA is to generate the weights for the linear combination V_q . Thus, a one-component PCA has been performed for every GII sub-pillar separately. The loading of every indicator on that component is considered to be the weight of that indicator in the linear combination. The results show that indicators in the sub-pillar Political environment have gained an equal weight of 0.97. This case of "equal weights" among all the indicators in a given sub-pillar has occurred in several sub-pillars such as Business Environment, Research & Development, Information & Communication Technologies, Investment, and Online creativity. These sub-pillars are across the board of the GII input and output. Meanwhile, Indicators such as GERD financed by abroad, Growth rate of PPP$ GDP/worker, and Creative goods exports have gained considerably low weights of less than 0.1. Table 1 provides the value of weights associated with each indicator.

5.2) Data envelopment analysis

After assigning the relative weights of the 21 variables (sub-pillars), we applied the benefit of doubt DEA. The efficiency scores yielded by the model PCA-DEA are presented in Table 2. It shows that 31 countries obtain a relative efficiency score = 1 and the rest of the countries obtain relative efficiency scores < 1. Concluding that, if all the GII indicators were to be considered as outputs (i.e., the benefit of doubt), 31 out of 129 countries tend to perform efficiently with regards to national innovation performance. On the other hand, the national innovation systems in the remaining 98 countries that obtained a score < 1 are considered to be performing inefficient at different levels, the closer the score to 1, the better the country is performing. The case of equal scores for the 31 countries is acknowledged because of the discrimination power problem of the DEA (see Section 4.2).

Table 2
PCA-DEA efficiency scores.

Country	PCA-DEA	Country	PCA-DEA	Country	PCA-DEA	Country	PCA-DEA
Australia	1.0000	Arab Emirates	0.9825	Georgia	0.8402	Honduras	0.7314
Austria	1.0000	Belgium	0.9649	Moldova	0.8397	Togo	0.7311
Bulgaria	1.0000	Mongolia	0.9474	Kuwait	0.8384	Senegal	0.7246
Canada	1.0000	Latvia	0.9361	Brazil	0.8358	Nepal	0.7245
China	1.0000	Viet Nam	0.9310	Costa Rica	0.8321	Niger	0.7187
Cyprus	1.0000	Thailand	0.9240	Montenegro	0.8276	Dominican	0.7145
Denmark	1.0000	Poland	0.9236	Panama	0.8213	Lebanon	0.7138
Estonia	1.0000	Turkey	0.9211	Philippines	0.8099	Guatemala	0.7131
Finland	1.0000	Belarus	0.9194	Rwanda	0.8096	Bangladesh	0.7128
France	1.0000	Jamaica	0.9129	Indonesia	0.8078	Benin	0.7090
Germany	1.0000	Croatia	0.9071	South Africa	0.8072	Ghana	0.7048
Greece	1.0000	Hungary	0.9057	Bosnia and Herzegovina	0.8052	Uganda	0.7042
Hong Kong	1.0000	Russian	0.9046	Morocco	0.7995	Nigeria	0.7035
Iceland	1.0000	Slovakia	0.9017	Botswana	0.7938	Zambia	0.7033
Ireland	1.0000	Bahrain	0.8942	Trinidad and Tobago	0.7931	Algeria	0.7014
Israel	1.0000	Lithuania	0.8927	Tunisia	0.7895	Burkina Faso	0.6994
Italy	1.0000	Oman	0.8925	Armenia	0.7888	Cameroon	0.6993
Japan	1.0000	Mauritius	0.8923	El Salvador	0.7799	Yemen	0.6956
Luxembourg	1.0000	Malaysia	0.8918	Peru	0.7773	Mali	0.6934
Malta	1.0000	Serbia	0.8851	Ecuador	0.7751	Burundi	0.6874
Netherlands	1.0000	Chile	0.8845	Bolivia	0.7727	Tajikistan	0.6780
New Zealand	1.0000	Brunei Darussalam	0.8830	Co^te d'Ivoire	0.7685	Malawi	0.6709
Norway	1.0000	Saudi Arabia	0.8792	Sri Lanka	0.7684	Madagascar	0.6604
Portugal	1.0000	Uruguay	0.8769	Pakistan	0.7665	Mozambique	0.6506
Republic of Korea	1.0000	Ukraine	0.8750	Kenya	0.7658	Tanzania	0.6431
Singapore	1.0000	Romania	0.8749	Kyrgyzstan	0.7616	Zimbabwe	0.6248
Slovenia	1.0000	Argentina	0.8701	Namibia	0.7597	Ethiopia	0.6089
Sweden	1.0000	Kazakhstan	0.8690	Egypt	0.7551		
Switzerland	1.0000	Paraguay	0.8641	Jordan	0.7524		
United Kingdom	1.0000	Azerbaijan	0.8639	India	0.7465		
USA	1.0000	Albania	0.8632	North Macedonia	0.7410		
Czech Republic	0.9982	Colombia	0.8570	Cambodia	0.7403		
Spain	0.9939	Qatar	0.8537	Guinea	0.7350		
Iran	0.9879	Mexico	0.8481	Nicaragua	0.7339		

However, to reach the ultimate final ranking for all the countries, the case of 31 equal efficiency scores must be resolved. To do that, only for the top 31 countries, we examine the efficiency of each country under each GII pillar separately. In practice, this implies running the previous benefit of doubt DEA model for every country seven times (one for every pillar). In each pillar, we consider the output variables to be the variables from Table 1 under that specific pillar and the input being equal to 1. This step evaluates the efficiency of every country with regards to every GII pillar one by one. Eventually, the average of these seven sub-DEAs for every country will be the discriminating value to form the final ranking among these 31 countries. As far as the GII pillars are individually concerned, the results show that the performance of the national innovation system in Switzerland is the best, followed by the USA and the Republic of Korea (see Table 3). Lastly, by replacing the score values = 1 in Table 2 with the average value of the seven sub-DEAs for these 31 efficient countries, a final ranking for all countries can be generated (see Table 4).

Table 3
The results of seven sub-DEAs to discriminate the 31 relative efficient countries.

Rank	Country	DEA Pillar 1	DEA Pillar 2	DEA Pillar 3	DEA Pillar 4	DEA Pillar 5	DEA Pillar 6	DEA Pillar 7	DEA (Avg)
1	Switzerland	0.984	0.9289	1	0.8256	1	0.933	0.9884	0.9514
2	USA	0.9833	0.9027	0.967	1	1	0.8437	0.8467	0.9348
3	Republic of Korea	0.9654	1	1	0.7318	0.9756	0.9943	0.8755	0.9347
4	Germany	0.958	0.9352	0.9502	1	0.9378	0.8326	0.884	0.9283
5	Netherlands	1	0.9404	0.9752	0.5491	1	1	0.9235	0.9126
6	Hong Kong, China	0.9858	0.9729	0.9459	1	0.9818	0.6658	0.8346	0.9124
7	Japan	0.9671	0.9643	0.9703	0.781	0.9417	0.913	0.8483	0.9122
8	United Kingdom	0.9444	0.9326	0.9978	0.7346	0.8922	0.9159	0.8868	0.9006
9	China	0.8062	0.9376	0.7953	0.8984	0.8488	1	1	0.8980
10	Sweden	0.98	0.9609	0.995	0.6534	1	0.7223	0.9534	0.8950
11	Denmark	0.9766	1	1	0.8212	0.8802	0.6706	0.8572	0.8865
12	Luxembourg	0.9595	0.8792	0.9728	0.4276	0.9001	1	1	0.8770
13	Cyprus	0.9001	0.8244	0.8841	0.9073	0.7265	1	0.7573	0.8571
14	Finland	1	0.9949	0.951	0.5536	0.9157	0.6589	0.8934	0.8525
15	Israel	0.8915	1	0.8839	0.4531	0.943	0.837	0.9493	0.8511
16	Australia	0.9641	0.9942	0.9593	0.8068	0.7509	0.6833	0.7856	0.8492
17	Ireland	0.9449	0.9322	1	0.3929	1	0.8857	0.7842	0.8486
18	Singapore	1	1	0.9591	0.7099	0.9284	0.5775	0.702	0.8396
19	New Zealand	1	0.9362	0.965	0.8772	0.7372	0.4788	0.8545	0.8356
20	Canada	0.9864	0.9557	0.9294	0.6499	0.7833	0.6598	0.8161	0.8258
21	France	0.9039	0.9164	0.9722	0.5292	0.8007	0.6963	0.884	0.8147
22	Norway	0.9975	0.9362	1	0.7033	0.8119	0.4532	0.7965	0.8141
23	Estonia	0.8735	0.9524	0.9257	0.471	0.676	0.8557	0.9289	0.8119
24	Iceland	0.9416	0.8948	1	0.5208	0.7512	0.4795	1	0.7983
25	Italy	0.9006	0.8907	0.9154	0.4406	0.7691	1	0.6486	0.7950
26	Slovenia	0.9536	0.9353	0.8328	0.3105	0.9871	0.609	0.8342	0.7804
27	Portugal	0.9228	0.9156	0.9098	0.5202	0.6731	0.6078	0.8166	0.7666
28	Austria	0.9024	0.9239	0.909	0.4826	0.8499	0.4874	0.77	0.7607
29	Malta	0.7943	0.8566	0.9676	0.4003	0.7553	0.488	0.9122	0.7392
30	Bulgaria	0.7716	0.8139	0.8286	0.3974	0.5901	0.9501	0.6884	0.7200
31	Greece	0.7981	1	0.8622	0.5237	0.5324	0.5989	0.5424	0.6940

In pursuance of the comparison, Table 4 shows the PCA-DEA ranking of all countries compared to the GII 2019, where the negative sign of difference indicates moving backward and the positive sign indicates moving forward in the new ranking. The comparison adopts two aspects (1) the absolute value of the difference between the country's position in the GII rank and the PCA-DEA rank. (2) the distribution of the absolute value of difference over the GII rank.

The comparison shows a noticeable difference in the ranking between the GII 2019 and the PCA-DEA score. Specifically, the average absolute value of the difference between the two ranks is 10.9 positions. Whereas some countries change their positioning of more than 40 positions of ranking, such as India and Macedonia, other countries such as Switzerland maintain the same position. Moreover, the results show that countries that lay in the middle of the GII ranking have made the greatest difference in the absolute value in their positions (i.e., the greatest change in the ranking both ways, forward and backward). Whilst countries that lay at the top or bottom of the GII made less change in rankings (seeFigure 1).

Table 4
Comparison between the final PCA-DEA rank and GII 2019 rank for all the countries.

Country	PCA-DEA	GII 2019	Diff	Abs.Value	Country	PCA-DEA	GII 2019	Diff	Abs.Value
Switzerland	1	1	0	0	Colombia	66	67	1	1
USA	2	3	1	1	Qatar	67	65	-2	2
Republic of Korea	3	11	8	8	Mexico	68	56	-12	12
Germany	4	9	5	5	Georgia	69	48	-21	21
Netherlands	5	4	-1	1	Moldova	70	58	-12	12
Hong Kong	6	13	7	7	Kuwait	71	60	-11	11
Japan	7	15	8	8	Brazil	72	66	-6	6
United Kingdom	8	5	-3	3	Costa Rica	73	55	-18	18
China	9	14	5	5	Montenegro	74	45	-29	29
Sweden	10	2	-8	8	Panama	75	75	0	0
Denmark	11	7	-4	4	Philippines	76	54	-22	22
Luxembourg	12	18	6	6	Rwanda	77	94	17	17
Cyprus	13	28	15	15	Indonesia	78	85	7	7
Finland	14	6	-8	8	South Africa	79	63	-16	16
Israel	15	10	-5	5	Bosnia and Herzegovina	80	76	-4	4
Australia	16	22	6	6	Morocco	81	74	-7	7
Ireland	17	12	-5	5	Botswana	82	93	11	11
Singapore	18	8	-10	10	Trinidad and Tobago	83	91	8	8
New Zealand	19	25	6	6	Tunisia	84	70	-14	14
Canada	20	17	-3	3	Armenia	85	64	-21	21
France	21	16	-5	5	El Salvador	86	108	22	22
Norway	22	19	-3	3	Peru	87	69	-18	18
Estonia	23	24	1	1	Ecuador	88	99	11	11
Iceland	24	20	-4	4	Bolivia	89	110	21	21
Italy	25	30	5	5	Co^te d'Ivoire	90	103	13	13
Slovenia	26	31	5	5	Sri Lanka	91	89	-2	2
Portugal	27	32	5	5	Pakistan	92	105	13	13
Austria	28	21	-7	7	Kenya	93	77	-16	16
Malta	29	27	-2	2	Kyrgyzstan	94	90	-4	4
Bulgaria	30	40	10	10	Namibia	95	101	6	6
Greece	31	41	10	10	Egypt	96	92	-4	4
Czech	32	26	-6	6	Jordan	97	86	-11	11
Spain	33	29	-4	4	India	98	52	-46	46
Iran	34	61	27	27	Macedonia	99	59	-40	40
Arab Emirates	35	36	1	1	Cambodia	100	98	-2	2
Belgium	36	23	-13	13	Guinea	101	125	24	24
Mongolia	37	53	16	16	Nicaragua	102	120	18	18
Latvia	38	34	-4	4	Honduras	103	104	1	1
Viet Nam	39	42	3	3	Togo	104	126	22	22
Thailand	40	43	3	3	Senegal	105	96	-9	9
Poland	41	39	-2	2	Nepal	106	109	3	3

Country	PCA-DEA	GII 2019	Diff	Abs.Value	Country	PCA-DEA	GII 2019	Diff	Abs.Value
Turkey	42	49	7	7	Niger	107	127	20	20
Belarus	43	72	29	29	Dominican	108	87	-21	21
Jamaica	44	81	37	37	Lebanon	109	88	-21	21
Croatia	45	44	-1	1	Guatemala	110	107	-3	3
Hungary	46	33	-13	13	Bangladesh	111	116	5	5
Russian	47	46	-1	1	Benin	112	123	11	11
Slovakia	48	37	-11	11	Ghana	113	106	-7	7
Bahrain	49	78	29	29	Uganda	114	102	-12	12
Lithuania	50	38	-12	12	Nigeria	115	114	-1	1
Oman	51	80	29	29	Zambia	116	124	8	8
Mauritius	52	82	30	30	Algeria	117	113	-4	4
Malaysia	53	35	-18	18	Burkina Faso	118	117	-1	1
Serbia	54	57	3	3	Cameroon	119	115	-4	4
Chile	55	51	-4	4	Yemen	120	129	9	9
Brunei	56	71	15	15	Mali	121	112	-9	9
Saudi Arabia	57	68	11	11	Burundi	122	128	6	6
Uruguay	58	62	4	4	Tajikistan	123	100	-23	23
Ukraine	59	47	-12	12	Malawi	124	118	-6	6
Romania	60	50	-10	10	Madagascar	125	121	-4	4
Argentina	61	73	12	12	Mozambique	126	119	-7	7
Kazakhstan	62	79	17	17	Tanzania	127	97	-30	30
Paraguay	63	95	32	32	Zimbabwe	128	122	-6	6
Azerbaijan	64	84	20	20	Ethiopia	129	111	-18	18

5.3) Random Forests

To assure the robustness of the PCA-DEA results in this paper, the RF technique has been applied. To elaborate, all countries have been divided into three groups. The first group consists of the 31 efficient countries in the PCA-DEA rank, this group being labelled as “Efficient”. The second group consists of the lowest 30 countries in the PCA-DEA rank, this group is labelled as “Highly inefficient”. Lastly, the third group consists of the remaining 68 countries in the middle of the PCA-DEA rank, which is labelled as “Inefficient”. The application of the RF shows that the PCA-DEA ranking matches its classification to 88%. In detail, for 26 out of 31 efficient countries in the ranking of PCA-DEA, the RF also has classified them as efficient, while 5 countries shifted to be inefficient. As regards the highly inefficient countries, 26 out of the lowest 30 countries in the ranking of PCA-DEA, were classified as highly inefficient by the RF, while 4 countries were shifted to be inefficient. Finally, RF classified 62 out of the 68 countries in the middle of the PCA-DEA ranking as inefficient, while 3 countries shifted to efficient, and 3 countries shifted to highly inefficient (see Table 5).

Table 5
Random forests classification results.

PCA-DEA / RF	Efficient	Inefficient	Highly Inefficient	Class. Error
Efficient	26	5	0	0.161
Inefficient	3	62	3	0.088
Highly Inefficient	0	4	26	0.133
	OOB estimate of error rate = 11.63%			

Additionally, RF highlights the importance of each variable during the classification. For example, the input sub-pillars “Political environment”, “Education”, and “Research & development”’ play a significant role in determining the rank of the country, while the sub-pillars “Investment”, “Trade, competition, and market scale” play a minimal role. Likewise, the output side sub-pillars such as “Knowledge impact” and “Creative goods and services” are more important than “Knowledge diffusion” and “Intangible assets” (see Table 6).

Table 6
Random Forests variable importance.

Variables	Efficient	Inefficient	Quite inefficient	MDA	MDG
Political environment	8.497015	1.777996	8.967475	12.17111	4.61021
Regulatory environment	5.792929	4.276533	7.899029	9.89423	3.66083
Business environment	3.302171	1.435669	9.61357	9.697885	1.868765
Education	9.314394	10.4139	16.94575	19.68011	10.52525
Tertiary education	4.35787	4.614659	7.249219	8.483599	3.527415
Research & development	10.14069	9.15372	8.839339	13.27621	6.998675
(ICTs)	11.52129	4.257144	10.43641	15.4041	7.399558
General infrastructure	4.77182	2.200702	6.758043	8.224492	3.057214
Ecological sustainability	6.03089	6.773267	3.823822	10.08638	2.382411
Credit	5.143057	4.803702	4.742395	7.589029	2.14092
Investment	1.134616	0.378198	2.801854	2.201341	0.812804
Trade, competition, & market scale	-0.7832	0.125584	2.598789	1.412711	0.563952
Knowledge workers	6.424123	6.387756	8.779528	11.50626	5.25608
Innovation linkages	4.754297	0.894083	2.415186	5.456192	1.245409
Knowledge absorption	5.399193	1.692625	5.061141	7.34702	1.717166
Knowledge creation	6.388614	7.182269	5.601698	11.11525	3.099586
Knowledge impact	10.37929	2.141033	8.168362	11.59132	3.852483
Knowledge diffusion	1.874541	-1.05249	1.89918	1.735037	0.608923
Intangible assets	6.364096	-0.08515	6.124902	8.001001	2.472512
Creative goods & services	11.17375	3.464594	7.186516	12.19684	6.067561
Online creativity	9.78703	6.928621	8.007719	13.78468	6.214577

6. Conclusions

Innovation is one of the economic factors leading to societal progress, technological development, and economic growth. As a result, the measurement of innovation performance and capacity has received increasing attention, not only at the academic level but also at the policy and societal levels. Composite Indicators (CIs) such as the Global Innovation Index (GII) and the European Innovation Scoreboard are well recognized and accepted instruments for this task. Although they represent a vehicle of communication and an elevator for the awareness of innovation, they also play a significant role in shaping and directing innovation policies.

The literature dealing with the construction of the CIs has had an important evolution in recent years, with new contributions emphasizing that the methodologies underlying these CIs should be considered carefully, in terms of weighting, aggregation, and robustness. In this paper, we introduce a Data Envelopment Analysis (DEA) model that relies on the weights provided by the application of Principal Component Analysis (PCA). This PCA-DEA model is applied to the 2019 edition of the GII, which measures innovation at the national level in 129 countries worldwide. The paper aims to examine the sensitivity of the results provided by the GII, to variations in the methodology. In addition, the robustness of the results provided by this alternative methodology is tested using the Random Forest (RF) methodology. The rationale for using data-driven techniques such as PCA, DEA, and RF, lies in that they eliminate the subjective intervention during the construction of the CI.

The PCA-DEA model used in the paper provides the relative efficiency of innovation of the countries considered in the GII, which has enabled us to draw a final ranking. This PCA-DEA model relies on the “Benefit of doubt” technique, which measures the effectiveness of the national innovation systems, since all the countries are attributed a single input with a value equal to one (i.e., no minimization function for the input side).

The comparison between the GII rank and PCA-DEA rank shows that the countries that lay in the middle of the GII rank have made a considerable change in their positions in terms of the absolute value of difference. Meanwhile, countries at the top or bottom of the GII rank have made a minimal change, despite the rank of some of the top countries such as Singapore has dropped 10 positions on the PCA-DEA model. This is explained by the Attributes-based weighting system (i.e., no weights or equal weights). The GII uses the unweighted average of the GII sub-indices to calculate the general index. Instead, we rely on the objective weights provided by the PCA to each of the GII sub-indices.

One of the conclusions of the paper is that the GII rank appears to be robust for the countries located at the top and the bottom of the ranking, but not for the countries in the middle. In other words, the positioning of the countries in the middle of the GII rank is more sensitive towards the modification of the construction method of the GII, compared to the countries at the top or the bottom. This can be attributed to the influence of the subjective means of generating and assigning the weights for the indicators such as the Budget Allocation scheme, since there is no consensus about the importance of an indicator over the others. In addition, what seems to be important for a certain group of countries is not necessarily important for other countries (Zabala-Iturriagagoitia et al., 2007a; Stavbunik & Pelucha, 2019).

As regards the new ranking provided by the alternative PCA-DEA model, the results of the RF methodology applied in the paper reveal that the new ranking is statistically robust. As a matter of fact, the RF attributes 88% of the countries to the right sub-groups. In addition, the RF classification also helped to assess the relative importance of each variable (sub-pillar) in explaining the factors that make every country to be efficient, inefficient, or highly inefficient.

This paper represents, to the best of our knowledge, the first contribution in which the robustness of the GII is assessed, using alternative methodologies such as the PCA and the DEA. Consistent with the extant literature, particularly related to the European Innovation Scoreboard, our results also reveal that the methodological election has a direct impact on the results provided by composite indices. Accordingly, further editions of the GII should seek to introduce alternative methodologies to the existing data, in order to increase the robustness and reliability of the rankings provided by the GII. This would allow emphasizing different dimensions of the innovation system, highlighting those areas in which each country may show relative strengths, and weaknesses. We contend that these alternative methodologies would provide additional information to policymakers, so more effective policies can be adopted on each innovation system.

Abbreviations

CI
Composite indicator
DEA
Data envelopment analysis
DMU
Decision-making unit
GII
Global innovation index
PCA
Principal component analysis
RF
Random forests

Declarations

Availability of data and materials.

The dataset used and analyzed during the current study is available in the 2019 edition of the Global Innovation Index report, [https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2019.pdf].

Competing interests

The author declares that he has no competing interests.

Funding

No source of funding.

Author’s contributions

Single author.

Acknowledgment

No acknowledgment.

Authors' information

Name: Khatab Alqararah

Email: khatab.alqararah@gmail.com

ORCiD: https://orcid.org/0000-0002-7837-5246

Counrty: Jordan

References

1. Adler, N., & Yazhemsky, E. (2010). Improving discrimination in data envelopment analysis: PCA–DEA. *European Journal of Operational Research*, 202, 273–284
2. Alnafrah, I. (2021). Efficiency evaluation of BRICS's national innovation systems based on bias-corrected network data envelopment analysis. *Journal of Innovation and Entrepreneurship*, 10(1), 1–28
3. Atkinson, T., Cantillon, B., Marlier, E., & Nolan, B. (2002). *Social indicators: The EU and social inclusion*. New York: Oxford University Press. doi:10.1093/0199253498.001.0001
4. Bandura, R. (2011). *Composite indicators and rankings: Inventory 2011. Technical report*. New York: United Nations Development Programme (UNDP)
5. Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1031–1142
6. Barbero, J., Iturriagagoitia, J. Z., & Zofio, J. L. (2021). Is more always better? On the relevance of decreasing returns to scale on innovation. *Technovation*, 107, 102314
7. Billaut, J. C., Bouyssou, D., & Vincke, P. (2010). Should you believe in the Shanghai ranking? *Scientometrics*, 84(1), 237–263
8. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32
9. Broughel, J., & Thierer, A. (2019). *Technological innovation and economic growth: A brief report on the evidence*. Arlington: Mercatus research
10. Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444
11. Cherchye, L., Moesen, W., & Van Puyenbroeck, T. (2004). Legitimately diverse, yet comparable: On synthesizing social inclusion performance in the EU. *Journal of Common Market Studies*, 42(5), 919–955
12. Cherchye, L., Moesen, W., Rogge, N., & Puyenbroeck, T. V. (2007). An introduction to 'Benefit of the Doubt' composite indicators. *Social Indicators Research*, 82(1), 111–145
13. Cherchye, L., Moesen, W., Rogge, N., Puyenbroeck, T., Saisana, M., Saltelli, A., & Tarantola, S. (2008). Creating composite indicators with DEA and robustness analysis: The case of the Technology Achievement Index. *Journal of the Operational Research Society*, 59, 239–251
14. Cobb, C. W., Halstead, T., & Rowe, J. (1995). *The genuine progress indicator: Summary of data and methodology*. San Francisco: Redefining Progress
15. Corrente, S., Bernabeu, A. G., Greco, S., & Makkonen, T. (2021). Robust measurement of innovation performances in Europe with a hierarchy of interacting composite indicators. *Economics of Innovation and New Technology*, 1–18. doi:https://doi.org/10.1080/10438599.2021.1910815
16. Cox, D. R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J., & Jones, D. R. (1992). Quality of life assessment: Can we keep it simple? *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 155(3), 353–393
17. Crespo, N., & Crespo, C. (2016). Global innovation index: Moving beyond the absolute value of ranking with a fuzzy-set analysis. *Journal of Business Research*, 69(11), 5265–5271
18. Cui, R., Sun, J., Li, Y., Yang, K., & Wu, X. (2020). Data-driven approach with artificial neural network for Global Innovation Index re-evaluation. *2020 6th International Conference on Big Data and Information Analytics (BigDIA)*, (pp. 88–93)
19. Decancq, K., & Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews*, 32(1), 7–34
20. Desai, M., Fukuda-Parr, S., Johansson, C., & Sagasti, F. (2002). Measuring technology achievement of nations and the capacity to participate in the network age. *Journal of Human Development*, 1(3), 301–311
21. Despotis, D. (2005). Measuring human development via data envelopment analysis: The case of Asia and the Pacific. *Omega*, 33(5), 385–390
22. Dutta, S., & Lanvin, B. (2013). *The Global Innovation Index 2013: The local dynamics of innovation.* Geneva, Ithaca, and Fontainebleau: Johnson Cornell University, INSEAD, and WIPO. From https://www.wipo.int/edocs/pubdocs/en/economics/gii/gii_2013.pdf
23. Dutta, S., Lanvin, B., & Vincent, S. W. (2019). *The Global Innovation Index 2019. Creating healthy lives—The future of medical innovation*. Fontainebleau, and Geneva: Ithaca. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2019.pdf From
24. Edquist, C. (2011). Systems of innovation perspectives and challenges. *African Journal of Science, Technology, Innovation and Development*, 2(3), 14–43
25. Edquist, C., Zabala-Iturriagagoitia, J. M., Barbero, J., & Zofío, J. L. (2018). On the meaning of innovation performance: Is the synthetic indicator of the Innovation Union Scoreboard flawed? *Research Evaluation*, 27(3), 196–211
26. Esty, D. C., Levy, M., Srebotnjak, T., & De Sherbinin, A. (2005). *Environmental sustainability index: Benchmarking national environmental stewardship*. New Haven: Yale Center for Environmental Law & Policy
27. European Commission (2019, June 17). *European Commission.* Retrieved March 30, 2020 from https://ec.europa.eu/commission/presscorner/detail/en/ip_19_2991
28. Freudenberg, M. (2003). Composite indicators of country performance: A critical assessment. *OECD Science, Technology and Industry Working Papers*
29. Gadrey, J., & Jany, F. C. (2003). *Les indicateurs de richesee et de developpement: Un bilan international en vue d'une initiative francaise*. Paris: DARES, Ministere du Travail
30. Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite. *Social Indicators Research*, 141, 61–94
31. Grupp, H., & Mogee, M. E. (2004). Indicators for national science and technology policy: How robust are composite indicators? *Research Policy*, 33(9), 1373–1384

32. Grupp, H., & Schubert, T. (2010). Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy*, 39(1), 67–78
33. Guaita Martínez, J., Martín Martín,., Ostos, M. S., & Pardo, M. (2021). Constructing knowledge economy composite indicators using an MCA-DEA approach. *Economic Research-Ekonomska Istraživanja, 34*(1), 331–351
34. Hamidi, S., & Berrado, A. (2018). Segmentation of innovation determinants: Case of the Global Innovation Index. *12th International Conference on Intelligent Systems: Theories and Applications* (pp. 1–8). Rabat: Association for Computing Machinery. From https://doi.org/10.1145/3289402.3289548
35. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2 ed., Vol. 12). New York: Springer-Verlag. From https://hastie.su.domains/ElemStatLearn/printings/ESLII_print12_toc.pdf
36. Hatefi, M. S., & Torabi, S. A. (2010). A common weight MCDA–DEA approach to construct composite indicators. *Ecological Economics*, 70(1), 114–120
37. Henderson, D. W. (1974). *Social indicators: A rationale and research framework.* Ottawa: Cambridge University Press. From https://www.cambridge.org/core/journals/canadian-journal-of-political-science-revue-canadienne-de-science-politique/article/abs/dw-henderson-social-indicators-a-rationale-and-research-framework-ottawa-information-canada-1974-pp-vii-90/8DE50AF54F6D3452E0C7
38. Hermans, E., Bossche, F. V., & Wets, G. (2008). Combining road safety information in a performance index. *Accident Analysis and Prevention*, 40, 1337–1344
39. Jankowska, B., Jedrych, A., & Dabrowska, K. (2017). Efficiency of national innovation systems – Poland and Bulgaria in the context of the Global Innovation Index. *Comparative Economic Research*, 20(3), 77–94
40. Lundvall, B. A. (2007). National innovation systems—Analytical concept and development tool. *Industry and Innovation*, 14(1), 95–119
41. Martin, J., Mendoza, C., & Román, C. (2017). A DEA Travel–Tourism Competitiveness Index. *Social Indicators Research*, 130, 937–957
42. Melyn, W., & Moesen, W. (1991). *Towards a synthetic indicator of macroeconomic performance*. Leuven: Center for Economic Studies
43. Moldan, B., & Billharz, S. (1997). *Sustainability indicators: A report on the project on indicators of sustainable development*. New York: Chichester John Wiley & Sons
44. Munda, G. (2005). Multiple criteria decision analysis and sustainable development. In S. Greco, M. Ehrgott, & J. Figueira (Eds.), *Multiple criteria decision analysis. International series in operations research & management science* (233 vol., pp. 1235–1267). New York: Springer. https://doi.org/10.1007/978-1-4939-3094-4_27. From
45. Munda, G. (2012). Choosing aggregation rules for composite indicators. *Social Indicators Research*, 109(3), 337–354
46. Munda, G., & Nardo, M. (2005). *Non-compensatory composite indicators for ranking countries: A defensible setting-—EUR 21833 EN*. Ispra: Institute for the Protection and Security of the Citizen
47. Nikolaidis, Y., Fouskas, K., & Carayannis, E. G. (2013). Assisting regional policy by rapidly comparing enterprise innovation between regions. *Journal of Innovation and Entrepreneurship*, 2(1), 1–25
48. OECD. (2004). *The OECD-JRC handbook on practices for developing composite indicators*. Paris: The OECD committee on statistics
49. OECD. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. Paris: OECD
50. OECD/Eurostat. (2018). *Oslo Manual: Guidelines for Collecting, Reporting and Using Data on Innovation, Fourth Edition*. Paris: OECD Publishing
51. Omer, M., El-Amin, M., Nasr, A., & Ahmed, R. (2020). Modeling, visualization, and analysis of African innovation performance. *International Conference on Learning Representation (ICLR)*. Addis Ababa. Retrieved 3 11, 2022 from https://www.researchgate.net/publication/343735079_Modeling_Visualization_and_Analysis_of_African_Innovation_Performance/fulltext/5f3c9c16a6fd Visualization-and-Analysis-of-African-Innovation-Performance.pdf?_sg%5B0%5D=CEDIy0xN-y8EZs6
52. Omrani, H., Valipour, M., & Mamakani, S. J. (2019). Construct a composite indicator based on integrating common weight data envelopment analysis and principal component analysis models: An application for finding development degree of provinces in Iran. *Socio-Economic Planning Sciences*, 68, 100618
53. Osberg, L., & Sharpe, A. (2002). An index of economic well–being for selected OECD countries. *The Review of Income and Wealth*, 48(3), 291–316
54. Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(3), 609–634
55. Pence, Ä., Kalkan, A., & Cesmeli, M. (2019). Estimation of the country ranking scores on the Global Innovation Index 2016 using the artificial neural network method. *International Journal of Innovation and Technology Management*, 16(4), 1940007
56. Pennoni, F., Tarantola, S., & Latvala, A. (2005). *The European e-business readiness index.* Joint Research Centre (JRC). Retrieved March 20, 2020 from https://www.researchgate.net/publication/46446317_The_2005_European_e-Business_Readiness_Index
57. Ramanathan, R. (2006). Evaluating the comparative performance of countries of the Middle East and North Africa: A DEA application. *Socio-Economic Planning Sciences*, 40(2), 156–167
58. Saisana, M., & Tarantola, S. (2002). *State of the art report on current methodologies and practices for composite indicator development*. Ispra, Italy: European Commission, Joint Research Centre, Institute for the Protection and the Security of the Citizen, Technological and Economic Risk Management Unit
59. Saisana, M., Saltelli, A., Schulze, N., Tarantola, S., & Duchene, V. (2005). Uncertainty and sensitivity analysis for the Knowledge-based Economy Index. *Conference on Medium-Term Economic Assessment (CMTEA).* Sofia. Retrieved 3 11, 2022 from https://www.researchgate.net/profile/Michaela-Saisana/publication/242180137_Uncertainty_and_Sensitivity_Analysis_for_the_Knowledge-Based_Economy_Index/links/579723b708aed51475e66de4/Uncertainty-and-Sensitivity-Analysis-for-the-Knowledge-Based-Economy-Inde
60. Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research*, 81(1), 65–77

61. Setiawan, B., Djanali, S., Ahmad, T., & Aziz, N. M. (2019). Assessing centroid-based classification models for intrusion detection system using composite indicators. *Procedia Computer Science*, 161, 665–676
62. Shen, Y., Hermans, E., Brijs, T., & Wets, G. (2013). Data envelopment analysis for composite indicators:. *Social Indicators Research*, 114, 739–756
63. Stavbunik, Y., & Pelucha, M. (2019). The innovation system of Kazakhstan in the context of the Global Innovation Index. *The Innovation Journal: The Public Sector Innovation Journal*, 24(1), 1–24
64. Thanassoulis, E. (2001). *Introduction to the theory and application of data envelopment analysis* (1 ed.). Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-1-4615-1407-7
65. UNDP (2019). *Human development index.* New York: United Nations Development Programme. Retrieved 10 19, 2020 from https://hdr.undp.org/en/content/download-data
66. WIPO (2021). *www.wipo.int.* Retrieved 2 10, 2022 from https://www.wipo.int/pressroom/en/articles/2021/article_0008.html#:~:text=About%20the%20Global%20Innovation%20Index&text=The%20GII%20has%2
67. Yang, F. C., Kao, R. H., Chen, Y. T., Ho, Y. F., Cho, C. C., & Huang, S. W. (2018). A common weight approach to construct composite indicators: The evaluation of fourteen emerging markets. *Social Indicators Research*, 137(2), 463–479
68. 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*, 41(5), 661–672
69. Zabala-Iturriagagoitia, J. M., Voigt, P., Gutierrez-Gracia, A., & Jimenez-Saez, F. (2007b). Regional innovation systems: How to assess performance. *Regional Studies*, 41(5), 661–672

Figures

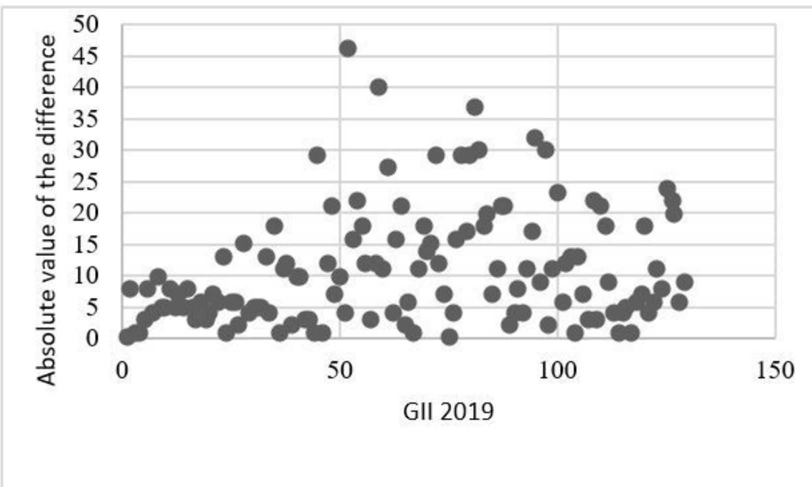

Figure 1

Distribution of absolute value of the difference over GII 2019.