Policy evaluation and efficiency: a systematic literature review

Anna Mergoni\textsuperscript{a,}\textdagger and Kristof De Witte\textsuperscript{a,b}

\textsuperscript{a}LEER, KU Leuven, Naamsestraat 69, Leuven B-3000, Belgium
\textsuperscript{b}UNU-MERIT, Maastricht University, Boschstraat 24, Maastricht 6211 AX, The Netherlands
E-mail: anna.mergoni@kuleuven.be [Mergoni]; kristof.dewitte@kuleuven.be [De Witte]

Received 9 March 2021; received in revised form 28 April 2021; accepted 17 May 2021

Abstract

This paper provides a systematic literature review of studies investigating the effect of an intervention on the efficiency of a decision-making unit, when efficiency is computed using nonparametric frontier approaches. This paper offers a guide for future research by identifying patterns in (1) the fields of application, (2) applied efficiency models, and (3) analysis of efficiency determinants. Our findings indicate that, despite the prominent role of frontier techniques in the analysis of public sector performances and the importance of the effectiveness and the policy perspective, these two approaches have long been kept separate. Nevertheless, the combination of efficiency and effectiveness is fundamental to evaluate public interventions and to detect inefficiencies at the policy level, especially in key sectors such as education, health, and environment.

Keywords: systematic literature review; policy evaluation; efficiency; DEA

1. Introduction

In the evaluation of complex (public) activities (e.g., schools, hospitals, airports, water utilities, etc.) efficiency and effectiveness are two complementary aspects. On the one hand, efficiency analysis evaluates the performance of complex decision-making units (DMUs), accounting for the ability of transforming input into output (Farrell, 1957). On the other hand, effectiveness studies measure the performance of DMUs with respect to given goals (Golany, 1988; Cherchye et al., 2019). When the objectives are specified by a policy, effectiveness studies assess the impact of the intervention on a number of variables of interest, revealing the ability of the policy of influencing them (for a review of classical policy evaluation techniques, see Abadie and Cattaneo, 2018). Using the words of Peter Drucker, effectiveness analysis indicates whether we are doing the right thing, while efficiency reveals whether we are doing it right (Drucker, 1977; Asmild et al., 2007; Førsund, 2017). Therefore, the combination of these two perspectives is fundamental to evaluate public interventions and to detect inefficiency at the policy level (Kornbluth, 1991).

\textdagger Corresponding author.
Using a systematic literature review, this paper investigates the role of frontier evaluation techniques in policy evaluation studies. The objective is to collect information on how the literature has studied the impact of external interventions on the performance of DMUs, when the performance is evaluated in terms of distance to an ideal best performance frontier. Historically, frontier evaluation approaches offered a mathematical formulation to the concept of technical efficiency (Førsund et al., 1980). As defined by Koopmans (1951) and Farrell (1957), a combination of input and output is efficient if it is not possible to increase the level of any output without increasing also the level of at least one input, or to decrease the level of any input without decreasing the current level of at least one output. In operational terms, two main approaches for frontier estimation can be distinguished, parametric and nonparametric approaches, according to the criterion used to specify the (functional) form. Despite both approaches are characterized by a number of models, the most representative models are the “stochastic frontier model” (SFA; Aigner and Chu, 1968; Aigner et al., 1977) and the “data envelopment analysis” (DEA; Charnes et al., 1978; Banker et al., 1984) model, respectively, for the parametric and nonparametric frameworks. However, the boundary between these strands is becoming blurred, as an increasing number of the deterministic models include stochastic components and vice versa (Daraio and Simar, 2007).

This paper focuses on the nonparametric efficiency literature, which includes DEA (Deprins et al., 1984), free disposal hull (FDH) (Tulkens, 2006), network (Färe, 1991; Färe et al., 2007), Malmquist (Färe et al., 1994), slack-based (Tone, 2001), bootstrap (Simar & Wilson, 1998), order-m (Cazals et al., 2002), order-alpha (Aragon et al., 2005), and conditional (Daraio and Simar, 2005) models. The continuous development of alternative techniques demonstrates the flexibility and the success of these approaches (for a review, see Seiford, 1996; Gattoufi et al., 2004; Cook and Seiford, 2009; Emrouznejad and Yang, 2018). In particular, the nonparametric frontier techniques are popular in public sector evaluations as they allow to measure the performance of complex DMUs accounting for the multiple dimensions involved, they rely only on a small set of assumptions and do not require price information (De Witte et al., 2020). Besides, the endogenous weighting of variables ensures that every unit is evaluated in the best possible light ensuring fairness and objectivity (Cherchye et al., 2007b).

Policy evaluation studies aim to determine the impact of an exogenous intervention on some outcome of interest (Heckman and Vytlacil, 2007). Traditionally, this has been done using classical econometric and statistical techniques, such as regression analyses (Draper and Smith, 1998). However, since the work of Rubin (1974), one of the main challenges in the econometric literature has been overcoming correlational evidence and assessing causal relationships. In the absence of an experimental setting, this often requires the construction of sophisticated identification strategies. For example, matching techniques are based on the idea of comparing treated and untreated units with similar covariates (Rosenbaum and Rubin, 1985; Rosenbaum, 1989; Abadie and Imbens, 2011), instrumental variables approaches solve the problem of missing or unknown control variables (Wright, 1928; Sargan, 1958; Angrist et al., 1996), difference-in-differences estimations use data with a time dimension to control for unobserved variables with a fixed trend (Donald and Lang, 2007; Lechner, 2011), or regression discontinuity designs exploit the setting where the treatment is determined according to a forcing variable and rely on the fact that locally the assignment is as good as random (Lee and Lemieux, 2010); more recently, machine learning approaches have been developed to improve predictions and detect heterogeneous effects (Athey and Imbens,
By identifying causal linkages, policy evaluation studies appraise the effectiveness of policy interventions, therefore, they play a fundamental role in the promotion of evidence-based policy making.

This paper provides the first systematic literature review of studies investigating the effect of an intervention (or policy) on the efficiency (or performance) of a DMU, computed via a nonparametric frontier approach. Despite the prominent role of frontier techniques in the analysis of public sector performances and the importance of the effectiveness and the policy perspective, these two approaches have long been kept separate (Førsund, 2017). With this review we aim to promote the use of nonparametric frontier techniques for policy evaluation, by cataloguing how this has been done in previous works.

The remainder of the paper continues as follows. First, we specify the criteria for the systematic review of the literature. Second, we classify the papers according their empirical application. Third, we discuss the methodologies implemented in previous works for the estimation of the efficiency scores. Next, we analyze the methodologies implemented to perform the policy evaluation and to investigate the determinants of the efficiency scores. Doing so, specific attention is given to the issues of endogeneity and causality. Finally, we provide a discussion and we indicate future lines of research.

2. Criteria of paper selection

This systematic review focuses on the role of nonparametric frontier analysis in policy evaluation. To implement the review, we rely on three main academic libraries: EBSCOhost, Scopus, and Web of Science (WOS). EBSCOhost is a multidisciplinary platform that includes a wide set of databases, among which is the Education Resource Information Center, sponsored by the Institute of Education Sciences of the U.S. Department of Education. Scopus is the largest abstract and citation database of peer-reviewed literature and it covers the fields of science, mathematics, engineering, technology, health and medicine, social sciences, and arts and humanities. WOS is the world’s leading academic citation indexing database and search service and covers the science, social science, arts, humanities disciplines. The use of these three sources provides the widest coverage, delivering a comprehensive and interdisciplinary scientific output.

To include all relevant papers, we have implemented the search using the research string: “data envelopment analysis” AND “policy evaluation,” using the “all fields” filter. We have chosen “data envelopment analysis” as our keyword for two main reasons: first, DEA is the first and most common nonparametric frontier model; second, it is a baseline for subsequent approaches. For these reasons the papers relying on FDH, Malmquist, network models, or on other nonparametric frontier approaches always refer also to DEA. To guarantee relevance, we restricted our search only to the papers published in English scholarly peer-reviewed journals. As time span we consider papers published between 1957 and January 2021. The former year provides a natural starting point for our research, as it is the year of publication of the seminal paper by Farrell (1957). This search returned in 205 documents from Scopus, 78 from EBSCOhost and 14 from WOS. Further, we performed a full-text analysis to exclude papers that did not use any DEA or DEA-related model or did not implement any policy evaluation. Finally, we added 17 papers retrieved via a snowball method. The snowball method is a technique to include relevant documents using key documents
as a starting point (Wohlin, 2014). In the remainder of the systematic literature review, we rely on this set of 81 papers.

3. Application-wise focus of the literature

To set the scene, we first graphically represent when the 81 papers were published. Figure 1 shows that after the early 2000s, there has been a strong increase in the number of published papers. The growth emerges since the seminal work by Simar and Wilson (1998) and Cazals et al. (2002), as these papers provide tools to obtain bias-corrected scores and to account for the influence of environmental variables on the efficiency scores.

The efficiency literature has a long tradition of analyzing policy changes in various subject fields. Categorizing the 81 detected papers in Section 2, we distinguish 10 homogeneous clusters of applications. We classified the papers in the application fields based on the policy under evaluation and on the DMUs considered. In Table 1, we relate the application fields to the *Journal of Economic Literature* classification system.1

Among the 10 distinguished application fields, 3 main sectors can be identified: public sector policies, environment-related policies, and private sector policies. From Table 1, we observe that public sector policies are the dominant application field, with 52 papers. This main field involves six categories. Education is the absolute protagonist (with 22 papers), followed by local government (13 papers), health (10 papers), labor (4 papers), transportation (3 papers), and economic growth (1 paper). The second most represented field is composed of 22 papers investigating the role of environment-related policies. In particular, the field is composed of nine papers on environmental policy, seven papers on agriculture, and six papers on energy. Finally, 11 papers are clustered in the category private sector interventions to foster competitiveness and innovation.

1The categories of the United Nations’ International Standard Industrial Classification (ISIC) of all economic activities (used in Daraio et al., 2020) are not suited for our review, as there are many ISIC categories for which no “policy evaluation and efficiency” exists.

© 2021 The Authors.
*International Transactions in Operational Research* published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies
Table 1
Papers categorized by application field

| Panel 1: Public sector policies | # | JEL | Application | Papers |
|---------------------------------|---|-----|-------------|--------|
|                                 | 22| I2  | Education   | Glass et al. (2006), Soares de Mello et al. (2006), Waldo (2007), Cherchy et al. (2007a), Chang et al. (2009), Coupet and Barnum (2010), Kantabutra and Tang (2010), Zhang et al. (2011), Montoneri et al. (2012), Sav (2012), Tochkov et al. (2012), Haelermans and De Witte (2012), De Witte et al. (2013a, 2013b), Schubert and Yang (2016), Carrington et al. (2018), Yang et al. (2018), Cordero et al. (2018), Santín & Sicilia (2018), Papadimitriou and Johnes (2019), Johnes and Virmani (2020), D’Inverno et al. (2021) |
|                                 | 13| H7  | Local government | Zabala-Iturriagagoitia et al. (2007), López et al. (2008), Shi and Yang (2008), Halkos and Tzeremes (2010b), De Witte and Saal (2010), Halkos and Tzeremes (2011b), Broersma et al. (2013), Asatryan and De Witte (2015), Cordero et al. (2017), Vieira et al. (2018), Wang et al. (2019), Ma et al. (2020), Titl and De Witte (2021) |
|                                 | 10| I1  | Health       | Puenpatom and Rosenman (2008), Zhang et al. (2008), Halkos and Tzeremes (2011a), Sahin et al. (2011), Seth and Patel (2014), Hitt and Tambe (2016), McCallion et al. (2000), Sevigny et al. (2020), Maragos et al. (2020), Ferreira et al. (2020) |
|                                 | 4 | J2  | Labor        | Feroz et al. (2001), Althin and Behrenz (2004), Sofi and Sharma (2015), Cordero and Tzeremes (2018) |
|                                 | 3 | R4  | Public transport | López et al. (2009), Mallikarjun et al. (2014), Park and Kim (2021) |
|                                 | 1 | O4  | Growth       | Halkos and Tzeremes (2009) |

| Panel 2: Environment-related policies | # | JEL | Application | Papers |
|---------------------------------------|---|-----|-------------|--------|
|                                       | 9 | Q5  | Environment | Halkos and Tzeremes (2010a), Rogge (2012), Ke et al. (2015), Whittaker et al. (2017), Pan et al. (2019), Ding et al. (2020), Czyżewski et al. (2020), Zhang et al. (2020), Ma et al. (2020) |
|                                       | 7 | Q1  | Agriculture | Whittaker (1994), Färe and Whittaker (1995), Whittaker et al. (2003), Murillo et al. (2007), Whittaker et al. (2017), Priscilla and Chauhan (2019), Kaliba et al. (2021) |
|                                       | 6 | Q4  | Energy      | Morfeldt and Silveira (2014), Halkos et al. (2015), Lin and Zhu (2019), Baláž et al. (2020), Zhang et al. (2020), Xin-gang et al. (2020) |

| Panel 3: Private sector policies | # | JEL | Application | Papers |
|---------------------------------|---|-----|-------------|--------|
|                                 | 11| O3  | Competitiveness, innovation, and technology | Halkos and Tzeremes (2007), Shi and Yang (2008), Hsu and Hsueh (2009), Kao and Hwang (2010), Jiménez-Sáez et al. (2011), Jiménez-Sáez et al. (2013), Halkos and Tzeremes (2013), Castillo and Salem (2013), Du et al. (2014), Wang et al. (2019), He et al. (2020) |

Note: # is the symbol of cardinality, that is, the number of elements in a given set.

In what follows, we discuss for each application category the main research questions, the specific policies evaluated, and the level of intervention. Besides, we highlight the patterns in the relevant characteristics and the time trends.
3.1. Public sector policies

3.1.1. Education
Policy evaluations using nonparametric efficiency models have been most frequently applied to the education field. The systematic literature review revealed 22 papers. Within the education field, the majority of applications focuses on higher education (Glass et al., 2006; Soares de Mello et al., 2006; Waldo, 2007; Chang et al., 2009; Coupet and Barnum, 2010; Kantabutra and Tang, 2010; Zhang et al., 2011; Montoneri et al., 2012; Sav, 2012; Tochkov et al., 2012; De Witte et al., 2013a, 2013b; Schubert and Yang, 2016; Carrington et al., 2018; Yang et al., 2018; Papadimitriou and Johnes, 2019), followed by secondary education (Haelermans and De Witte, 2012; Cordero et al., 2018; D’Inverno et al., 2021) and primary education (Cherchye et al., 2007a; Santín & Sicilia, 2018). Only Johnes and Virmani (2020) follow the students across different levels (excluding higher education). Within the education applications, the policy evaluated includes funding policies (Chang et al., 2009; Tochkov et al., 2012; Schubert and Yang, 2016; Carrington et al., 2018; D’Inverno et al., 2021), policy to improve teaching quality (Montoneri et al., 2012; Santín & Sicilia, 2018; Yang et al., 2018) or research quality (Chang et al., 2009; Zhang et al., 2011), or both (De Witte et al., 2013a, 2013b). Also studies investigating the impact of merger (Glass et al., 2006; Papadimitriou and Johnes, 2019) or the difference between public or private institutions (Cherchye et al., 2007a; Waldo, 2007; Kantabutra and Tang, 2010; Sav, 2012; Johnes and Virmani, 2020) are reported. Three main geographical areas of investigation are reported: Europe (Glass et al., 2006; Cherchye et al., 2007a; Waldo, 2007; Haelermans and De Witte, 2012; Tochkov et al., 2012; De Witte et al., 2013a, 2013b; Schubert and Yang, 2016; Cordero et al., 2018; Santín & Sicilia, 2018; Papadimitriou and Johnes, 2019; D’Inverno et al., 2021), China and Taiwan (Chang et al., 2009; Kantabutra and Tang, 2010; Zhang et al., 2011; Montoneri et al., 2012; Yang et al., 2018), and the United States (Coupet and Barnum, 2010; Sav, 2012). Three studies fall out of these geographical clusters: Carrington et al. (2018), Soares de Mello et al. (2006), and Johnes and Virmani (2020), who focus, respectively, on Australia, Brazil, and on the comparison of a set of developing countries (Peru, India, Ethiopia, Vietnam).

3.1.2. Local government
A second subfield of public sector applications focuses on local government policy changes. Although there is a wider literature on local government efficiency (Narbón-Perpiñá and De Witte, 2018a, 2018b; Daraio et al., 2020), the systematic literature review revealed 13 papers that combine nonparametric efficiency analysis and local government policy evaluation.

The policy investigated by these studies are wide and heterogeneous, from the role of direct democracy (Asatryan and De Witte, 2015) to the effect of public investments for innovation competitiveness and development (Zabala-Iturriagagoitia et al., 2007; Shi and Yang, 2008; Vieira et al., 2018; Wang et al., 2019), from policies centered more on sustainability (Natesan and Marathe, 2017; Ma et al., 2020) to welfare and social assistance policies (López et al., 2008; Halkos and Tzeremes, 2011b; Broersma et al., 2013).

From the perspective of the level of analysis, a consistent portion of these studies involves the evaluation of regional policies (Shi and Yang, 2008; Halkos and Tzeremes, 2010b, 2011b;
Vieira et al., 2018; Natesan and Marathe, 2017; Titl and De Witte, 2021), with a special attention given to the study case of Spain (Zabala-Iturriagagoitia et al., 2007; López et al., 2008), while the majority of the other studies focuses on the municipal and provincial levels (Broersma et al., 2013; Asatryan and De Witte, 2015; Cordero et al., 2017; Wang et al., 2019; Ma et al., 2020).

3.1.3. Health
The third most researched field within public sector policy evaluations are the health applications, with 10 papers. Despite the key position played by the health sector and the importance of efficiency measurement in the health care provision, application of frontier techniques to hospitals, nursing homes, and, more generally, to health management organizations are relatively recent (Worthington, 2004) and the use of efficiency techniques for the evaluation of public health policies still constitute an unexplored potential (Lobo et al., 2014). Except for the pioneer study by McCallion et al. (2000), investigating the role of the size on the efficiency of Northern Irish hospitals, literature on health policy and efficiency dates back to 2008 with the study of Zhang et al. (2008) and Puenpatom and Rosenman (2008). These studies investigate, respectively, the effect of Medicare prospective paying system on the efficiency in skilled nursing homes (Zhang et al., 2008) and the extension of health insurance programs on the efficiency of provincial public hospitals in Thailand (Puenpatom and Rosenman, 2008) or in India (Seth and Patel, 2014). In line with the latter research, Sahin et al. (2011) studied a health transformation program based on the introduction of a general health insurance in Turkish hospitals. Other studies compare different health systems (Halkos and Tzeremes, 2011a), funding systems (Ferreira et al., 2020), evaluate the introduction of information technology on the performance of hospitals (Hitt and Tambe, 2016), compare different national policies (Maragos et al., 2020), or responses to HIV (Sevigny et al., 2020). Despite the heterogeneity in the health policy considered, the large majority of studies share their unit of analysis, that is, the hospital level (McCallion et al., 2000; Puenpatom and Rosenman, 2008; Halkos and Tzeremes, 2011a; Sahin et al., 2011; Hitt and Tambe, 2016).

3.1.4. Other topics
Three additional topics can be classified: papers within the fields of labor, public transport, and growth.

The literature on labor involves four studies, two on the effect of some policies on labor outcomes, such as employment (Feroz et al., 2001; Althin and Behrenz, 2004) and two that focus on the impact of labor policies on productivity (Sofi and Sharma, 2015; Cordero and Tzeremes, 2018).

Three additional studies investigate the relation between efficiency and (public) transport interventions. Specifically, López et al. (2009) and Mallikarjun et al. (2014) investigate the impact of public subsidies on transport infrastructures while Park and Kim (2021) study airports efficiency.

We classify Halkos and Tzeremes (2009) in a residual category. This study cannot be considered as a local government study as it evaluates the effect of the European Economic Monetary Union on the growth rate of the member states.

© 2021 The Authors.
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies
3.2. Environment-related policies

3.2.1. Environment
The systematic literature review distinguished nine papers evaluating the impact of environmental policies in a nonparametric frontier context. The first article dates back to Halkos and Tzeremes (2010a) who investigate the efficiency of environmental policies on biodiversity, probably inspired by the work of Nielsen et al. (2007). The late development of this literature can be attributed to the difficulty in choosing the appropriate dimensions for the evaluation of the environmental performances.

The fact that the concept of environmental performance involves different dimension results in a number of multidisciplinary studies that relate environmental issues with agricultural efficiency (Whittaker et al., 2017), marine economy (Ding et al., 2020), sustainable development (Czyżewski et al., 2020; Ma et al., 2020), or energy efficiency (Pan et al., 2019; Zhang et al., 2020). In particular, the bond between the energy and the environmental studies perspective is confirmed by the fact that all the energy papers in this review deal with environmental thematic (for more details, see Subsection 3.2.3).

Regarding the level of investigation, first studies focused on the construction of indexes to evaluate environmental performance at country level (Halkos and Tzeremes, 2010b; Rogge, 2012), but this approach was quickly abandoned in favor of studies on a lower level (as counties and provinces). This change of direction has probably been driven by the growing data availability, which is in turn driven by the rising attention for environmental concerns. With this respect, it is interesting to note that most studies were geographically located in Europe (Pan et al., 2019; Czyżewski et al., 2020) or China (Ke et al., 2015; Ding et al., 2020; Ma et al., 2020; Zhang et al., 2020).

3.2.2. Agriculture
As the second main category in the context of environment-related policies, we distinguished policies for the primary production sectors. In the review, we observe seven papers that use nonparametric efficiency models to evaluate agricultural policies. This literature is characterized by an analysis at the farm level and by the attention to environmental issues. In particular, the use of production efficiency to evaluate (agricultural) policy goes back to Whittaker (1994) who suggested to use DEA for (agricultural) policy analysis. This recommendation was reinforced by Färe and Whittaker (1995) who developed an intermediate input model to the evaluation of dairy farm and called for the use of production efficiency tools for the evaluation of agricultural policy. More recently, Whittaker et al. (2003, 2017) investigate the viability of a tax reduction policy to decrease the use of agricultural fertilizer of U.S. farms, Murillo et al. (2007) analyze the impact on efficiency and environmental adaption of the subsidy system of the Common Agricultural Reform on 1992 for animal-oriented farms in Spain, Priscilla and Chauhan (2019) focus on the role of cooperation on dairy farms in Manipur, and Kaliba et al. (2021) assess the advantage of using climate-adaptive sorghum for sorghum producers in Tanzania.

3.2.3. Energy
The rise of environmental concerns is also reflected in the energy efficiency literature. In particular, the six papers we observe in the systematic literature review analyzed the effect on energy efficiency
of policies involving the use of renewable energy sources (Halkos et al., 2015), incentives for sustainability and emission reductions (Lin and Zhu, 2019; Baláž et al., 2020; Zhang et al., 2020), and energy conservation (Xin-gang et al., 2020). The geographical distribution of these studies suggests that particularly Europe (Morfeldt and Silveira, 2014; Halkos et al., 2015; Baláž et al., 2020) and China (Lin and Zhu, 2019; Xin-gang et al., 2020; Zhang et al., 2020) are the regions with most attention to environmental efficiency.

3.3. Private sector policies

3.3.1. Competitiveness and innovation

With 11 relevant publications, the competitiveness and innovation applications constitute the third main field of application in absolute terms. As a number of local government studies also focused on the evaluation of public investments for innovation and competitiveness, we need to clarify the difference between these two categories. Studies that analyze the impact of policies for innovation at municipal or regional level have been categorized as local government studies, while the ones considering the impact of similar policy at firm level have been set in the category competitiveness and innovation. The only exception is the paper by Halkos and Tzeremes (2013), which investigates the impact of different national cultures on innovation in Europe.

Among the policy to promote competitiveness and innovation of firms, we observe a heterogeneous set of interventions. From science and technology policies, as in Du et al. (2014), to policy manufacturing policies to innovate the logistic industries, as in He et al. (2020), while other studies focus on the effect of policies for competitiveness (Halkos and Tzeremes, 2007; Wang et al., 2019). In particular, it is interesting to distinguish R&D projects as they constitute the biggest slice (Shi and Yang, 2008; Hsu and Hsueh, 2009; Jiménez-Sáez et al., 2011, 2013).

4. Method-wise focus of the literature

The idea of technical efficiency of a unit as the ratio of its actual production over the potential production, for a given set of inputs, dates back to Farrell (1957). Starting from this seminal paper, parametric and nonparametric frontiers models developed the idea of measuring inefficiency in terms of the relative ability of transforming inputs into valuable outputs with respect to the best performing peers.

Regarding the nonparametric frontier models, the reference point for the DEA literature is the Charnes, Cooper and Rhodes (CCR) model proposed by Charnes et al. (1978), which assumes an underlying convex production function characterized by constant return to scale and is, in its original formulation, output-oriented.\(^2\) In the literature on efficiency for policy evaluation the Banker, Charnes and Cooper (CCR) model encountered great success, in particular in the field of education and firms’ productivity. However, only few studies rely entirely on the CCR model (for education

\(^2\)Output-oriented models measure the efficiency as the ability of producing as much output as possible given set of inputs, while in input-oriented models efficiency is measured as the ability of producing a given amount of output using as less input as possible (Maragos et al., 2020).
Even more flexible is the FDH model proposed by Deprins et al. (1984), which abandons the convexity assumption to introduce a step-like frontier. With respect to DEA, FDH models are particularly appealing under uncertainty as they rely solely on the assumption that production possibilities satisfy free disposability (Cherchye et al., 2000; Tulkens, 2006). However, for what concerns the practical implementation, nonconvex models are still not as popular as their convex counterparts and they remain potentially underutilized in efficiency estimation (Cook and Seiford, 2009). This is reflected also in the works on efficiency for policy evaluation. Among the papers included in this review only five implemented an FDH approach (Cherchye et al., 2007a; Haelermans and De Witte, 2012; Asatryan and De Witte, 2015; Cordero et al., 2017; D’Inverno et al., 2021). A possible explanation can be traced back to the fact that nonconvex models are computationally more demanding, as they require to solve binary mixed integer programming problems instead of the classical linear problems, and the solution to these problems is less likely to be covered in the commonly available software packages (Kerstens and Van de Woestyne, 2018).

As for the orientation, both input-oriented and output-oriented models have been implemented in the literature, adapting to the characteristics of the empirical applications in question. Interestingly, we note that in educational applications output-oriented models are more common, as they highlight the perspective of educational outcome production (in line with the suggestion of Worthington, 2001), while in health application input-oriented models are implemented to give more emphasis on the ability of spending less resources. In the other application fields, it is not possible to identify similar patterns.

The literature is characterized by a continuous number of innovations and the classical DEA approaches have been used as a starting point for successive models. In the context of policy evaluation particular emphasis is given to the Malmquist index, which allows to measure efficiency changes over time (or across programs) and to the Network DEA model, which allow detection of the inefficiency of internal processes employed by the units. The Malmquist index had particular success in the context of educational policies as in Sahin et al. (2011), Zhang et al. (2011), Sav (2012), Schubert and Yang (2016), and D’Inverno et al. (2021), while the network model was exploited in the context of firm’s performance (Kao and Hwang, 2010; He et al., 2020).

Since the seminal work of Simar and Wilson (1998), bootstrap and bias-corrected techniques have become common across different sectors. These techniques have been used to analyze the sensitivity of efficiency scores relative to sample variations of the estimated frontier and aim to mitigate them. Therefore, they are particularly useful in the presence of noisy input or output measures, which are common in the field of environment (Halkos and Tzeremes, 2010a; Czyżewski et al., 2020), innovation (Castillo and Salem, 2013; Halkos and Tzeremes, 2013), and education (Tochkov et al., 2012; Papadimitriou and Johnes, 2019). A further innovation was provided by Cazals et al. (2002) and Daraio and Simar (2007) with the introduction of robust and conditional models as they allow to investigate the effect of possible environmental variables on the efficiency scores, without assuming the separability condition. Despite attractive features of the conditional

© 2021 The Authors.
*International Transactions in Operational Research* published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies
efficiency approach (see De Witte and Kortelainen, 2013), they have been used merely in education (Cherchye et al., 2007a; Haelermans and De Witte, 2012; De Witte et al., 2013b; Cordero et al., 2018; D’Inverno et al., 2021), local government (Asatryan and De Witte, 2015; Cordero et al., 2017; Titl and De Witte, 2021), and innovation (Halkos and Tzeremes, 2013) applications. The absence of a wider diffusion can be partly attributed to the fact that conditional models rely on a sophisticated theoretical apparatus and partly to the fact that at the moment no statistical software provide built-in procedure to implement conditional analysis.

Finally, for the purpose of validation and to give robustness to the results, seven studies compare different DEA models in the fields of education (Glass et al., 2006; Waldo, 2007), health (Halkos and Tzeremes, 2011a; Seth and Patel, 2014), environment (Halkos and Tzeremes, 2010a), local government (Halkos and Tzeremes, 2010b), and innovation (Halkos and Tzeremes, 2013) and three implement both DEA and non-DEA approaches in the fields of local government (Zabaliturriagagoitia et al., 2007; Wang et al., 2019) and education (Tochkov et al., 2012). The combination of several techniques is especially common when the application fields are heterogeneous and interdisciplinary, as these characteristics hamper the establishment of wide accepted models.

5. Determinant-wise focus of the literature

To investigate the effect of the policy on efficiency, it is necessary to detect the determinants of the efficiency scores. Traditionally, this has been done using two-step procedures (starting from the pioneering paper by Byrnes et al. (1988)). In the first step the efficiency scores are computed, while in the second step the scores are used to make inference and investigate their determinants. Simar and Wilson (2007) pointed out that two-step procedures are biased when the separability condition is not fulfilled, that is, when the background variables influence the input or output mix. Conditional efficiency models (initially developed by Cazals et al., 2002) allow to overcome this criticism. In particular, Daraio and Simar (2007), Simar and Wilson (2011), and De Witte and Kortelainen (2013) suggest to nonparametrically regress the ratio of unconditional over conditional scores on the set variables that affect, but do not directly enter as inputs or outputs, the production process. Table 2 overviews the papers implementing this conditional approach.

Next, we consider the literature from the perspective of the model implemented in the second stage (Table 3). Interestingly, it is not possible to observe patterns between the suggested models and the application fields. What emerges, instead, is that three macrocategories can be singled out: the first and largest category is constituted by regression-based analysis; the second involves studies that aim to overcome correlational evidence to assess causal relations; the final category encompasses studies that do not implement special inference techniques, relying on simple descriptive statistics analysis.

Regression analyses constitute the main instrument of investigation (see Coupet and Barnum, 2010; Sofi and Sharma, 2015; Schubert and Yang, 2016; Vieira et al., 2018; Ding et al., 2020). However, standard regression analysis does not account for the fact that the efficiency score is bounded between 0 and 1, and therefore inference might be biased (for a discussion, see Hoff, 2007). To address this issue, semiparametric and nonparametric regression models have been implemented (see Murillo et al., 2007; Halkos and Tzeremes, 2011b). Besides, although heavily criticized by Simar and Wilson (2007), particular attention has been given to truncated models (censored Tobit regression
Table 2
Papers categorized by the nonparametric frontier models implemented

| #  | Panel 1: Return to scale | Papers categorized by the nonparametric frontier models implemented |
|----|--------------------------|-------------------------------------------------------------------|
| 10 | Constant Returns to Scale (CRS) | Soares de Mello et al. (2006), Chang et al. (2009), Coupet and Barnum (2010), Montoneri et al. (2012), Castillo and Salem (2013), Whittaker et al. (2017), Cordero and Tzeremes (2018), Papadimitriou and Johnes (2019), Ding et al. (2020), Zhang et al. (2020) |
| 18 | Variable Returns to Scale (VRS) | Puenpatom and Rosenman (2008), Zhang et al. (2008), Waldo (2007), Hsu and Hsueh (2009), Tochkov et al. (2012), Mallikarjun et al. (2014), Morfeldt and Silveira (2014), Seth and Patel (2014), Halkos et al. (2015), Ke et al. (2015), Hitt and Tambe (2016), Santin & Sicilia (2018), Carrington et al. (2018), Vieira et al. (2018), Czyżewski et al. (2020), Xin-gang et al. (2020), Johnes and Virmani (2020), Titl and De Witte (2021) |
| 9  | Compare CRS-VRS            | Glass et al. (2006), Halkos and Tzeremes (2011a, 2011b), Sahin et al. (2011), Zhang et al. (2011), Sav (2012), Halkos and Tzeremes (2013), Baláž et al. (2020), Czyżewski et al. (2020) |
| 8  | Scale efficiency           | Halkos and Tzeremes (2009), Halkos and Tzeremes (2010b, 2010a), Jiménez-Sáez et al. (2011), Jiménez-Sáez et al. (2013), Du et al. (2014), Schubert and Yang (2016), Natesan and Marathe (2017) |

| #  | Panel 2: Orientation | Papers categorized by the nonparametric frontier models implemented |
|----|----------------------|-------------------------------------------------------------------|
| 14 | Output-oriented      | Glass et al. (2006), Cherchye et al. (2007a), Chang et al. (2009), Kantabutra and Tang (2010), Zhang et al. (2011), Montoneri et al. (2012), Sav (2012), Tochkov et al. (2012), Haelermans and De Witte (2012), Asatryan and De Witte (2015), Natesan and Marathe (2017), Santin & Sicilia (2018), Cordero et al. (2018), Johnes and Virmani (2020) |
| 22 | Input-oriented       | McCallion et al. (2000), Glass et al. (2006), Waldo (2007), Halkos and Tzeremes (2009), Hsu and Hsueh (2009), Halkos and Tzeremes (2010b, 2010a), De Witte and Saal (2010), Puenpatom and Rosenman (2008), Zhang et al. (2008), Halkos and Tzeremes (2011a, 2011b), Sahin et al. (2011), Halkos and Tzeremes (2013), Hitt and Tambe (2016), Schubert and Yang (2016), Natesan and Marathe (2017), Cordero et al. (2017), Cordero and Tzeremes (2018), Vieira et al. (2018), D’Inverno et al. (2021), Titl and De Witte (2021) |
| 1  | Output/input         | Czyżewski et al. (2020)                                           |
| 3  | Nonoriented           | Glass et al. (2006), Mallikarjun et al. (2014), Morfeldt and Silveira (2014) |
| 5  | Slack-based model     | Seth and Patel (2014), Ke et al. (2015), Wang et al. (2019)(super), Baláž et al. (2020), Czyżewski et al. (2020) |
| 6  | Distance function     | Jiménez-Sáez et al. (2011), Jiménez-Sáez et al. (2013), Halkos et al. (2015), Pan et al. (2019), Lin and Zhu (2019), Zhang et al. (2020) |

| #  | Panel 3: Model       | Papers categorized by the nonparametric frontier models implemented |
|----|----------------------|-------------------------------------------------------------------|
| 12 | Banker, Charnes and Cooper (BCC) | Hsu and Hsueh (2009), Tochkov et al. (2012), Halkos and Tzeremes (2013), Du et al. (2014), Ke et al. (2015), Hitt and Tambe (2016), Schubert and Yang (2016), Vieira et al. (2018), Santin & Sicilia (2018), Baláž et al. (2020), Johnes and Virmani (2020) |
| 8  | Charnes, Cooper and Rhodes (CCR) | Soares de Mello et al. (2006), Montoneri et al. (2012), Castillo and Salem (2013), Halkos and Tzeremes (2013), Du et al. (2014), Schubert and Yang (2016), Cordero and Tzeremes (2018), Baláž et al. (2020) |
| 4  | Benefit of the Doubt  | Rogge (2012), Halkos et al. (2015), De Witte et al. (2013b), Titl and De Witte (2021) |
| 2  | Farrel (1957)         | Althin and Behrenz (2004), Murillo et al. (2007)                  |

Continued
Table 2
(Continued)

| #  | Panel 3: Model                              |
|----|--------------------------------------------|
| 10 | Malmquist                                  |
|    | McCallion et al. (2000), Sahin et al. (2011), Zhang et al. (2011), Sav (2012), Broersma et al. (2013), Jiménez-Sáez et al. (2013), Morfeldt and Silveira (2014), Sofi and Sharma (2015), Ke et al. (2015), Schubert and Yang (2016) |
| 6  | FDH                                        |
|    | Cherchye et al. (2007a), Halkos and Tzeremes (2011a), Haelemans and De Witte (2012), Asatryan and De Witte (2015), Cordero et al. (2017), D’Inverno et al. (2021) |
| 2  | Bootstrap DEA                              |
|    | Puenpatom and Rosenman (2008), Tochkov et al. (2012) |
| 8  | Robust                                     |
|    | Cherchye et al. (2007a), Haelemans and De Witte (2012), De Witte et al. (2013b), Asatryan and De Witte (2015), Cordero et al. (2017), Cordero et al. (2018), D’Inverno et al. (2021), Titl and De Witte (2021) |
| 11 | Conditional                                 |
|    | Cherchye et al. (2007a), Halkos and Tzeremes (2011a, 2011b), Haelemans and De Witte (2012), Halkos and Tzeremes (2013), De Witte et al. (2013b), Asatryan and De Witte (2015), Cordero et al. (2017), Cordero et al. (2018), D’Inverno et al. (2021), Titl and De Witte (2021) |
| 6  | Bias-corrected                              |
|    | Halkos and Tzeremes (2010b), Halkos and Tzeremes (2011b), Castillo and Salem (2013), Halkos and Tzeremes (2013), Papadimitriou and Johnes (2019), Cyżewski et al. (2020) |
| 6  | DEA (generic)                               |
|    | Feroz et al. (2001), Halkos and Tzeremes (2007), Shi and Yang (2008), De Witte and Saal (2010), Priscilla and Chauhan (2019), Kaliba et al. (2021) |
| 6  | Network DEA                                 |
|    | Kao and Hwang (2010) CCR type, Mallikarjun et al. (2014) unoriented, VRS, Yang et al. (2018) network two stage directional distance DEA, Ding et al. (2020) two-stage cooperative game, He et al. (2020) |
| 12 | Other DEA-based technique                   |
|    | Whittaker (1994), Soares de Mello et al. (2006), López et al. (2008), López et al. (2009), Whittaker et al. (2017), Halkos and Tzeremes (2009), Carrington et al. (2018), Ma et al. (2020), De Witte et al. (2013a), De Witte and Saal (2010), Sevigny et al. (2020), Ferreira et al. (2020) |
| 2  | DEA and SFA                                 |
|    | Sav (2012), Broersma et al. (2013)          |
| 6  | Compare different DEA-based method          |
|    | Glass et al. (2006), Waldio (2007), Halkos and Tzeremes (2010a, 2010b), Halkos and Tzeremes (2011a), Halkos and Tzeremes (2013) |
| 3  | Compare DEA with non-DEA approaches         |
|    | Zabala-Iturriagagoitia et al. (2007), Wang et al. (2019), Tochkov et al. (2012) |

Note: # is the symbol of cardinality, that is, the number of elements in a given set.

in Althin and Behrenz, 2004; Waldo, 2007; Hsu and Hsueh, 2009; Mallikarjun et al., 2014, and truncated regression in Puenpatom and Rosenman, 2008; Tochkov et al., 2012; Cyżewski et al., 2020). In the context of panel data, instead, panel models are implemented (see Zhang et al., 2011; Papadimitriou and Johnes, 2019).

A main drawback of the observed regression studies is the lack of a proper identification strategy, which impedes that the evidences supplied overcome correlation. Two main obstacles impede this overcoming. First, the interventions analyzed are not randomly assigned to the units, therefore, the treated units and nontreated units may differ in observed and unobserved characteristics. Second, the intrinsic multidimensionality involved in the computation of the efficiency scores makes it difficult to distinguish between the causal effect of the policy on the single inputs and outputs and the true causal effect of the policy on the efficiency. To overcome these issues, few attempts have been made in the literature by implementing classical econometric techniques for causal inference.
Table 3
Papers by methodology to investigate the determinants of efficiency

| #  | Methodology                  | Papers                                                                                                                                 |
|----|------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| 15 | Descriptive analysis         | McCallion et al. (2000), Glass et al. (2006), Murillo et al. (2007), De Witte and Saal (2010), Montoneri et al. (2012), Tochkov et al. (2012), De Witte et al. (2013a), Seth and Patel (2014), Halkos et al. (2015), Natesan and Marathe (2017), Carrington et al. (2018), Cordero et al. (2018), Sevigny et al. (2020), Wang et al. (2019), Johnes and Virmani (2020) |
| 9  | Censored models              | Althin and Behrenz (2004), Waldo (2007), Zhang et al. (2008), Puenpatom and Rosenman (2008), Hsu and Hsueh (2009), Tochkov et al. (2012), Mallikarjun et al. (2014), Santín & Sicilia (2018), Czyżewski et al. (2020) |
| 9  | Nonparametric models         | Cherchye et al. (2007a), Murillo et al. (2007), Halkos and Tzeremes (2011b, 2011a), De Witte et al. (2013b), Haelermans and De Witte (2012), Asatryan and De Witte (2015), Cordero et al. (2018), Titl and De Witte (2021) |
| 8  | Regression models            | Coupet and Barnum (2010), Sofi and Sharma (2015), Schubert and Yang (2016), Vieira et al. (2018), Ding et al. (2020); Flexible location scale regression: Cordero et al. (2017); Generalized Least Square regression: Cordero and Tzeremes (2018); Ridge regression: Xin-gang et al. (2020) |
| 2  | Panel analysis               | Panel regression: Zhang et al. (2011); Random effect: Papadimitriou and Johnes (2019)                                                                 |
| 6  | Propensity score             | Castillo and Salem (2013), Lin and Zhu (2019), Papadimitriou and Johnes (2019), Priscilla and Chauhan (2019), Ma et al. (2020), Kaliba et al. (2021) |
| 6  | Difference-in-Differences (DiD) | Hitt and Tambe (2016), Lin and Zhu (2019), Pan et al. (2019), Baláž et al. (2020), He et al. (2020), Ma et al. (2020) |
| 1  | Regression Discontinuity (RDD) | D’Inverno et al. (2021)                                                                                                           |
| 1  | Simultaneous equation        | Zhang et al. (2020)                                                                                                                   |

Note: # is the symbol of cardinality, that is, the number of elements in a given set. RDD, regression discontinuity design.

In particular, we observe that propensity score matching (Castillo and Salem, 2013; Papadimitriou and Johnes, 2019) and difference-in-differences (Hitt and Tambe, 2016; Pan et al., 2019; Baláž et al., 2020; He et al., 2020) are the more common techniques, and sometimes they are used as complementary tools (Lin and Zhu, 2019; Ma et al., 2020). Recent studies by D’Inverno et al. (2021) (and Mergoni et al. 2020) translated the idea of the identification strategy of the regression discontinuity design in the realm of frontier estimation. The value added of these papers is that they do not assume the separability condition and account for the endogeneity directly in the computation of the efficiency scores.

Finally, descriptive studies involve correlation analysis (Murillo et al., 2007; Montoneri et al., 2012; Tochkov et al., 2012) and the comparison of the distribution of the efficiency scores of particular subpopulations. In this context, the classical t-test and the ANOVA test are commonly used (Kantabutra and Tang, 2010; Zhang et al., 2011; Sav, 2012). Besides, the Mann–Whitney test and the Wilcoxon test are used to assess return to scale characterization (Puenpatom and Rosenman, 2008; Halkos and Tzeremes, 2009, 2010a, 2011a; Zhang et al., 2011), while the Jackknife and the Kruskal–Wallis tests are used to detect differences in the distribution of the efficiency scores (Waldo, 2007; Puenpatom and Rosenman, 2008).
5.1. Endogeneity

This subsection focuses on the topic of endogeneity, as it is crucial both in the policy evaluation and in the efficiency context to obtain unbiased results. On the one hand, endogeneity is defined by econometricians as the correlation between the error term and at least one of the regressors. Three main sources of endogeneity are reverse causality, omitted variable bias (also in the form of unobserved heterogeneity and selection bias), and measurement errors. On the other hand, in the context of frontier estimation, Orme and Smith (1996) defined endogeneity as the correlation between the inefficiency and the level of input and shown by that it may cause a bias in the efficiency scores’ estimation. This kind of endogeneity is also known as the input demand endogeneity or Olley and Pakes endogeneity, from their seminal paper (Olley and Pakes, 1992). Despite their importance, both types of endogeneity are still overlooked in the efficiency literature (Cordero et al., 2015).

To detect the effect of a policy on the efficiency, endogeneity issues play a central role, as it is likely that a nonrandom intervention affects both the input level and the efficiency of the units under analysis and, therefore, induces a correlation between these variables. Despite this potential threat, endogeneity concerns are discussed only to a limited extent in the papers collected. As reported in Table 4, the systematic literature review resulted in 16 papers that mentioned endogeneity, and only in nine papers that discuss it and indicate its causes. These numbers are surprisingly low considering that none of the investigated interventions in our review considers random assignments.

The empirical fields where endogeneity is more discussed are education (Waldo, 2007; Schubert and Yang, 2016; Santín & Sicilia, 2018; Papadimitriou and Johnes, 2019; D’Inverno et al., 2021), local government and environment (Halkos and Tzeremes, 2011b; Broersma et al., 2013; Asatryan and De Witte, 2015; Cordero et al., 2017; Ma et al., 2020), and environment (Pan et al., 2019; Czyżewski et al., 2020; Zhang et al., 2020). Interestingly, endogeneity is regarded, in any field of applications, only from a policy evaluation perspective, that is, it is discussed in terms of correlation between a regressor and the error term. Only Santín and Sicilia (2018) and D’Inverno et al. (2021) are an exception to this tendency, as they account also for the Olley and Pakes type of endogeneity, that is, they account for the endogenous variables that enter the construction of the efficiency score.
6. Discussion and future research

This systematic literature review analyzed the papers investigating the effect of policy interventions on performance, when performance is evaluated using nonparametric frontier models. Policy evaluation and nonparametric frontier approaches are widely used in the literature, respectively, to evaluate the effectiveness of (public) interventions and to evaluate the performance of (public) activities. With this review, we shed light on the studies that join these two perspectives with the purpose of determining good practices for future research.

Using three main search engines, Scopus, EBSCHOhost, and WOS, we were able to select 81 papers for our review. We categorized the fields of empirical application (Section 3). Moreover, we classified papers along the applied methodologies used to obtain the efficiency scores (Section 4) and to investigate their determinants (Section 5).

From this review, three main conclusions can be drawn. First, the literature analyzed is relatively recent. Despite nonparametric efficiency models dated back to the seminal paper by Charnes et al. (1978) and despite the first attempt to compare the efficiency scores of units operating under different programs by Charnes et al. (1981), the oldest paper of this review is Whittaker (1994), investigating the relation between farm size and government program benefits.

Second, the main areas of application involve mainly sectors of public interest such as education (22 papers), local government (13 papers), health (10 papers), and environment (9 papers). Together, these fields account for more than half of the papers included in the review. In turn, this suggests that the necessity for joining the efficiency and the effectiveness perspective is particularly relevant in the context of public policy.

Third, it appears that policy evaluation in the efficiency framework is a growing but still largely unexplored field. On the one hand, the analyzed policies are extremely heterogeneous, also in the context of the same application fields. Therefore, up to now it is impossible to implement meta-analyses to derive some pooled conclusions. On the other hand, causal results are scarce and not enough attention has been paid to the endogeneity issue. It emerges that the correlation between the policy intervention and the input level has not been sufficiently accounted for, yielding to possibly biased estimations. Besides, the lack of appropriate identification strategies prevent causal conclusions.

These conclusions are probably the reason that, despite the countless number of research articles testifying the great advantages in using DEA and similar nonparametric frontier methods to evaluate private and public sectors, the acceptance of DEA is still limited both in general interest journals in economics and in formal documents of government agencies or private companies. This pattern is especially marked in the United States, while in Europe, countries such as Belgium, Italy, Portugal, and Spain are starting to use DEA in a systematic way, following the pressure of an active group of academicians. The root of this underimplementation can be identified in three main interconnected issues. First, nonparametric frontier estimation techniques are not generally taught at graduate level, therefore its use and comprehension remain possible only for a niche of researchers. Second, the fact that the production process is considered as a black box is a double-edged sword. On the one hand, it allows the DEA models to be flexible and reliable, as they depend only on a small set of assumptions. On the other hand, it prevents a deep understanding of the transformation mechanisms and prevents an immediate interpretation of the results. Third, as it also emerged from this literature review, endogeneity issues are typically neglected in efficiency works. This
prevents the use DEA for the detection of causal results, therefore hampering a widespread acceptance of its use for policy evaluation.

As future lines of research, we indicate two main challenges. First, more methodological effort is needed to interpret the findings in a causal way. The comprehension of mechanisms to detect causal effects on efficiency is still a largely unexplored field. Despite the rapid evolution of policy evaluation (Abadie and Cattaneo, 2018) and machine learning (Athey and Imbens, 2019) techniques for causal inference, it is still not clear how to adapt them in the world of frontier estimations, in particular in the context of quasi-experiments.

Second, researchers should adapt the field of investigation to the current global challenges. In particular, three main directions of expansions can be indicated: sustainability, health, and education. Despite the growing attention of the DEA literature for environment- and sustainability-related topics (see the special issue by Chen et al., 2017) and the fact that environmental efficiency, eco-efficiency, emissions, pollution, sustainable development, environmental protection are among the main fields of current DEA studies (see the review of Emrouznejad and Yang, 2018, Section 5), researchers devoted only limited efforts in the investigation of their determinants and in the analysis of the role that environmental policy can play. To contribute to the development of adequate policies, the literature should make an effort to fill this gap. Besides, the perspective of sustainability should be used as a compass whenever possible. In particular, a priority should be given to investigate the efficiency of policies for the promotion of renewable energy and for the reduction of intensive farming, as these are the two key points in the climate change challenge. Another current global challenge concerns the health sectors. The covid-19 pandemic sheds light, on the one hand, on the fact that the actual health systems could not be able to cope with the needs of an aging and growing population, on the other hand, on the fact that an efficient health system is not only more sustainable economically but more effective in guaranteeing a high-quality service. Moreover, as the covid-19 pandemic deeply affected the education sector, it provides a unique opportunity to investigate the effects of distance teaching on the efficiency of the education production function.

Acknowledgments

We are grateful to Marc-André Chénier, Letizia Gambi, Silvia Palmaccio, and Viteslat Titl for their precious suggestions. Anna Mergoni acknowledges financial support from FWO (Fellowship 11G5520N).

References

Abadie, A., Cattaneo, M., 2018. Econometric methods for program evaluation. Annual Review of Economics 10, 465–503.
Abadie, A., Imbens, G., 2011. Bias-corrected matching estimators for average treatment effects. Journal of Business & Economic Statistics 29, 1, 1–11.
Aigner, D., Chu, S., 1968. On estimating the industry production function. The American Economic Review 58, 4, 826–839.
Aigner, D., Lovell, C., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 1, 21–37.
Althin, R., Behrenz, L., 2004. An efficiency analysis of Swedish employment offices. International Review of Applied Economics 18, 4, 471–482.
Angrist, J., Imbens, G., Rubin, D., 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 434, 444–455.

Aragon, Y., Daouia, A., Thomas-Agnan, C., 2005. Nonparametric frontier estimation: a conditional quantile-based approach. *Econometric Theory* 21, 2, 358–389.

Asatryan, Z., De Witte, K., 2015. Direct democracy and local government efficiency. *European Journal of Political Economy* 39, 58–66.

Asmild, M., Paradi, J., Reese, D., Tam, F., 2007. Measuring overall efficiency and effectiveness using DEA. *European Journal of Operational Research* 178, 1, 305–321.

Athey, S., Imbens, G., 2019. Machine learning methods that economists should know about. *Annual Review of Economics* 11, 685–725.

Baláž, V., Nežinský, E., Jeck, T., Filcák, R., 2020. Energy and emission efficiency of the Slovak regions. *Sustainability (Switzerland)* 12, 7.

Banker, R., Charnes, A., Cooper, W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 30, 9, 1078–1092.

Broersma, L., Edzes, A., van Dijk, J., 2013. Have Dutch municipalities become more efficient in managing the costs of social assistance dependency? *Journal of Regional Science* 53, 2, 274–291.

Byrnes, P., Färe, R., Grosskopf, S., Knox Lovell, C., 1988. The effect of unions on productivity: US surface mining of coal. *Management Science* 34, 9, 1037–1053.

Carrington, R., O’Donnell, C., Prasada Rao, D., 2018. Australian university productivity growth and public funding revisited. *Studies in Higher Education* 43, 8, 1417–1438.

Castillo, L., Salem, D., 2013. Overdraft facility policy and firm’s performance: an empirical analysis in Eastern European Union industrial firms. *Journal of Business Economics & Management* 14, 5, 886–902.

Cazals, C., Florens, J., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. *Journal of Econometrics* 106, 1, 1–25.

Chang, D.F., Wu, C.T., Ching, G., Tang, C.W., 2009. An evaluation of the dynamics of the plan to develop first-class universities and top-level research centers in Taiwan. *Asia Pacific Education Review* 10, 1, 47–57.

Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research* 2, 6, 429–444.

Charnes, A., Cooper, W., Rhodes, E., 1981. Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. *Management Science* 27, 6, 668–697.

Chen, C., Zhou, S., Zhu, J., 2017. New research frontiers in sustainability. *Omega* 66, 183–184.

Cherchye, L., De Witte, K., Ooghe, E., Nicaise, I., 2007a. Equity and efficiency in private and public education: a nonparametric comparison. Available at https://ssrn.com/abstract=1020922.

Cherchye, L., De Witte, K., Perelman, S., 2019. A unified productivity-performance approach applied to secondary schools. *Journal of the Operational Research Society* 70, 9, 1522–1537.

Cherchye, L., Kuosmanen, T., Post, T., 2000. What is the economic meaning of FDH? A reply to thrall. *Journal of Productivity Analysis* 13, 3, 263–267.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., 2007b. An introduction to “benefit of the doubt” composite indicators. *Social Indicators Research* 82, 1, 111–145.

Cook, W., Seiford, L., 2009. Data envelopment analysis (DEA)—thirty years on. *European Journal of Operational Research* 192, 1, 1–17.

Cordero, J., Pedraja-Chaparro, F., Pisañol, E., Polo, C., 2017. Efficiency assessment of Portuguese municipalities using a conditional nonparametric approach. *Journal of Productivity Analysis* 48, 1, 1–24.

Cordero, J., Polo, C., Santin, D., Simancas, R., 2018. Efficiency measurement and cross-country differences among schools: a robust conditional nonparametric analysis. *Economic Modelling* 74, 45–60.

Cordero, J., Santin, D., Sicilia, G., 2015. Testing the accuracy of DEA estimates under endogeneity through a Monte Carlo simulation. *European Journal of Operational Research* 244, 2, 511–518.

Cordero, J., Tzeremes, N., 2018. Financial crisis and hotels’ labour productivity growth: evidence from Spanish islands. *Applied Economics Letters* 25, 19, 1376–1382.

Coupet, J., Barnum, D., 2010. HBCU efficiency and endowments: an exploratory analysis. *International Journal of Educational Advancement* 10, 3, 186–197.
Czyżewski, B., Smedzik-Ambróży, K., Mrówczyńska-Kamińska, A., 2020. Impact of environmental policy on eco-efficiency in country districts in Poland: how does the decreasing return to scale change perspectives? *Environmental Impact Assessment Review* 84, 106431.

Daraio, C., Kerstens, K., Nepomuceno, T., Sickles, R., 2020. Empirical surveys of frontier applications: a meta-review. *International Transactions in Operational Research* 27, 2, 709–738.

Daraio, C., Simar, L., 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis* 24, 1, 93–121.

Daraio, C., Simar, L., 2007. *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. Springer Science & Business Media, Berlin.

De Witte, K., Johnes, G., Johnes, J., Agasisti, T., 2020. Preface to the special issue on efficiency in education, health and other public services. *International Transactions in Operational Research* 27, 4, 1819–1820.

De Witte, K., Mrówczyńska-Kamińska, A., 2020. Impact of environmental policy on eco-efficiency in country districts in Poland: how does the decreasing return to scale change perspectives? *Environmental Impact Assessment Review* 84, 106431.

Daraio, C., Kerstens, K., Nepomuceno, T., Sickles, R., 2020. Empirical surveys of frontier applications: a meta-review. *International Transactions in Operational Research* 27, 2, 709–738.

Daraio, C., Simar, L., 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis* 24, 1, 93–121.

Daraio, C., Simar, L., 2007. *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. Springer Science & Business Media, Berlin.

De Witte, K., Johnes, G., Johnes, J., Agasisti, T., 2020. Preface to the special issue on efficiency in education, health and other public services. *International Transactions in Operational Research* 27, 4, 1819–1820.

De Witte, K., Mrówczyńska-Kamińska, A., 2020. Impact of environmental policy on eco-efficiency in country districts in Poland: how does the decreasing return to scale change perspectives? *Environmental Impact Assessment Review* 84, 106431.

Daraio, C., Kerstens, K., Nepomuceno, T., Sickles, R., 2020. Empirical surveys of frontier applications: a meta-review. *International Transactions in Operational Research* 27, 2, 709–738.

Daraio, C., Simar, L., 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis* 24, 1, 93–121.

Daraio, C., Simar, L., 2007. *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. Springer Science & Business Media, Berlin.

De Witte, K., Johnes, G., Johnes, J., Agasisti, T., 2020. Preface to the special issue on efficiency in education, health and other public services. *International Transactions in Operational Research* 27, 4, 1819–1820.

De Witte, K., Kortelainen, M., 2013. What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics* 45, 17, 2401–2412.

De Witte, K., Rogge, N., Cherchye, L., Van Puyenbroeck, T., 2013a. Accounting for economies of scope in performance evaluations of university professors. *Journal of the Operational Research Society* 64, 11, 1595–1606.

De Witte, K., Rogge, N., Cherchye, L., Van Puyenbroeck, T., 2013b. Economies of scope in research and teaching: a non-parametric investigation. *Omega* 41, 2, 305–314.

De Witte, K., Saal, D., 2010. Is a little sunshine all we need? On the impact of sunshine regulation on profits, productivity and prices in the dutch drinking water sector. *Journal of Regulatory Economics* 37, 3, 219–242.

Deprins, D., Simar, L., Tulkens, H., 1984. Measuring labor-efficiency in post offices. Technical Report, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Ding, L.L., Lei, L., Wang, L., Zhang, L.F., Calin, A., 2020. A novel cooperative game network DEA model for marine circular economy performance evaluation of China. *Journal of Cleaner Production* 253, 120071.

D’Inverno, G., Smet, M., De Witte, K., 2021. Impact evaluation in a multi-input multi-output setting: evidence on the effect of additional resources for schools. *European Journal of Operational Research* 290, 3, 1111–1124.

Donald, S., Lang, K., 2007. Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics* 89, 2, 221–233.

Draper, N., Smith, H., 1998. *Applied Regression Analysis*, Vol. 326. John Wiley & Sons, Hoboken, NJ.

Drucker, P., 1977. *An Introductory View of Management*. Harper & Row, New York City, NY.

Du, C., Xiao, Y., Lv, W., 2014. Performance evaluation of the science and technology policy in Chongqing. *Journal of Science and Technology Policy Management* 5, 2, 122–135.

Emrouznejad, A., Yang, G., 2018. A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences* 61, 4–8.

Färe, R., 1991. Measuring Farrell efficiency for a firm with intermediate inputs. *Academia Economic Papers* 19, 2, 329–340.

Färe, R., Grosskopf, S., Norris, M., Zhang, Z., 1994. Productivity growth, technical progress, and efficiency change in industrialized countries. *The American Economic Review* 84, 1, 66–83.

Färe, R., Grosskopf, S., Whittaker, G., 2007. Network DEA. In *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*. Springer, Berlin, pp. 209–240.

Färe, R., Whittaker, G., 1995. An intermediate input model of dairy production using complex survey data. *Journal of Agricultural Economics* 46, 2, 201–213.

Farrell, M., 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)* 120, 3, 253–281.

Feroz, E.H., Raab, R., Haag, S., 2001. An income efficiency model approach to the economic consequences of Osha cotton dust regulation. *Australian Journal of Management (University of New South Wales)* 26, 1, 69.

Ferreira, D.C., Nunes, A.M., Marques, R.C., 2020. Optimizing payments based on efficiency, quality, complexity, and heterogeneity: the case of hospital funding. *International Transactions in Operational Research* 27, 4, 1930–1961.

Førsund, F., 2017. Measuring effectiveness of production in the public sector. *Omega* 73, 93–103.

Førsund, F., Lovell, C., Schmidt, P., 1980. A survey of frontier production functions and of their relationship to efficiency measurement. *Journal of Econometrics* 13, 1, 5–25.

© 2021 The Authors. *International Transactions in Operational Research* published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.
Gattoufi, S., Oral, M., Reisman, A., 2004. Data envelopment analysis literature: a bibliography update (1951–2001). *Journal of Socio-Economic Planning Sciences* 38, 2–3, 159–229.

Glass, J., McCallion, G., McKillop, D., Rasaratnam, S., Stringer, K., 2006. Implications of variant efficiency measures for policy evaluations in UK higher education. *Socio-Economic Planning Sciences* 40, 2, 119–142.

Golany, B., 1988. An interactive MOLP procedure for the extension of dea to effectiveness analysis. *Journal of the Operational Research Society* 39, 8, 725–734.

Haelermans, C., De Witte, K., 2012. The role of innovations in secondary school performance—evidence from a conditional efficiency model. *European Journal of Operational Research* 223, 2, 541–549.

Halkos, G., Tzeremes, N., 2007. International competitiveness in the ICT industry: evaluating the performance of the top 50 companies. *Global Economic Review* 36, 2, 167–182.

Halkos, G., Tzeremes, N., 2009. Economic efficiency and growth in the EU enlargement. *Journal of Policy Modeling* 31, 6, 847–862.

Halkos, G., Tzeremes, N., 2010a. Measuring biodiversity performance: a conditional efficiency measurement approach. *Environmental Modelling and Software* 25, 12, 1866–1873.

Halkos, G., Tzeremes, N., 2010b. Measuring regional economic efficiency: the case of Greek prefectures. *Annals of Regional Science* 45, 3, 603–632.

Halkos, G., Tzeremes, N., 2011a. A conditional nonparametric analysis for measuring the efficiency of regional public healthcare delivery: an application to Greek prefectures. *Health Policy* 103, 1, 73–82.

Halkos, G., Tzeremes, N., 2011b. Modelling regional welfare efficiency applying conditional full frontiers. *Spatial Economic Analysis* 6, 4, 451–471.

Halkos, G., Tzeremes, N., 2013. Modelling the effect of national culture on countries’ innovation performances: a conditional full frontier approach. *International Review of Applied Economics* 27, 5, 656–678.

Halkos, G., Tzeremes, N., Tzeremes, P., 2015. A nonparametric approach for evaluating long-term energy policy scenarios: an application to the Greek energy system. *Journal of Economic Structures* 4, 1.

He, D., Yang, J., Wang, Z., Li, W., 2020. Has the manufacturing policy helped to promote the logistics industry? *PLoS ONE* 15, 7, 1–21.

Heckman, J., Vytlacil, E., 2007. Econometric evaluation of social programs, part I: causal models, structural models and econometric policy evaluation. *Handbook of Econometrics* 6, 4779–4874.

Hitt, L.M., Tambe, P., 2016. Health care information technology, work organization, and nursing home performance. *ILR Review* 69, 4, 834–859.

Hoff, A., 2007. Second stage DEA: comparison of approaches for modelling the DEA score. *European Journal of Operational Research* 181, 1, 425–435.

Hsu, F.M., Hsueh, C.C., 2009. Measuring relative efficiency of government-sponsored R&D projects: a three-stage approach. *Evaluation and Program Planning* 32, 2, 178–186.

Jiménez-Sáez, F., Zabala-Iturriagagoitia, J., Zofío, J., 2013. Who leads research productivity growth? Guidelines for R&D policy-makers. *Scientometrics* 94, 1, 273–303.

Jiménez-Sáez, F., Zabala-Iturriagagoitia, J., Zofío, J., Castro-Martínez, E., 2011. Evaluating research efficiency within national R&D programmes. *Research Policy* 40, 2, 230–241.

Johnes, G., Virmani, S., 2020. The efficiency of private and public schools in urban and rural areas: moving beyond the development goals. *International Transactions in Operational Research* 27, 4, 1869–1885.

Kaliba, A., Gongwe, A., Mazvimavi, K., Yigletu, A., 2021. Impact of adopting improved seeds on access to broader food groups among small-scale sorghum producers in tanzania. *SAGE Open* 11, 1.

Kantabutra, S., Tang, J., 2010. Efficiency analysis of public universities in Thailand. *Tertiary Education and Management* 16, 1, 15–33.

Kao, C., Hwang, S.N., 2010. Efficiency measurement for network systems: IT impact on firm performance. *Decision Support Systems* 48, 3, 437–446.

Ke, S., Chen, Z., Robson, M., Wen, Y., Tiana, X., 2015. Evaluating the implementation efficiency of the natural forest protection program in ten provinces of Western China by using data envelopment analysis (DEA). *International Forestry Review* 17, 4, 469–476.

Kerstens, K., Van de Woestyne, I., 2018. Enumeration algorithms for FDH directional distance functions under different returns to scale assumptions. *Annals of Operations Research* 271, 2, 1067–1078.
Koopmans, T., 1951. Efficient allocation of resources. *Econometrica: Journal of the Econometric Society* 19, 4, 455–465.

Kornbluth, J., 1991. Analysing policy effectiveness using cone restricted data envelopment analysis. *Journal of the Operational Research Society* 42, 12, 1097–1104.

Lechner, M., 2011. The estimation of causal effects by difference-in-difference methods. *Foundations and Trends (R) in Econometrics* 4, 3, 165–224.

Lee, D., Lemieux, T., 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48, 2, 281–355.

Lin, B., Zhu, J., 2019. Impact of energy saving and emission reduction policy on urban sustainable development: empirical evidence from China. *Applied Energy* 239, 12–22.

Lobo, M.S.C., Ozcan, Y.A., Lins, M.P.E., Silva, A.C.M., Fiszman, R., 2014. Teaching hospitals in Brazil: findings on determinants for efficiency. *International Journal of Healthcare Management* 7, 1, 60–68.

López, E., Gutiérrez, J., Gómez, G., 2008. Measuring regional cohesion effects of large-scale transport infrastructure investments: an accessibility approach. *European Planning Studies* 16, 2, 277–301.

López, E., Monzón, A., Ortega, E., Quintana, S., 2009. Assessment of cross-border spillover effects of national transport infrastructure plans: an accessibility approach. *Transport Reviews* 29, 4, 515–536.

Ma, J., Zhang, Z., Lu, C., Xue, B., 2020. Could the construction of sustainable development pilot zones improve the urban environment efficiency in China? *Discrete Dynamics in Nature and Society*, https://doi.org/10.1155/2020/7678525.

Mallikarjun, S., Lewis, H., Sexton, T., 2014. Operational performance of u.s. public rail transit and implications for public policy. *Socio-Economic Planning Sciences* 48, 1, 74–88.

Maragos, E., Maravelakis, P., Linardis, A., 2020. A DEA evaluation of the successful implementation of HEALTH2020 policies. *Socio-Economic Planning Sciences*, https://doi.org/10.1016/j.seps.2020.100968.

McCallion, G., Colin Glass, J., Jackson, R., Kerr, C., McKillop, D., 2000. Investigating productivity change and hospital size: a nonparametric frontier approach. *Applied Economics* 32, 2, 161–174.

Montoneri, B., Lin, T., Lee, C., Huang, S.L., 2012. Application of data envelopment analysis on the indicators contributing to learning and teaching performance. *Teaching and Teacher Education* 28, 3, 382–395.

Morfeldt, J., Silveira, S., 2014. Capturing energy efficiency in European iron and steel production—comparing specific energy consumption and Malmquist productivity index. *Energy Efficiency* 7, 6, 955–972.

Murillo, C., Juan, C., Sperlich, S., 2007. An empirical assessment of the EU agricultural policy based on firm level data. *Jahrbucher fur Nationalokonomie und Statistik* 227, 3, 273–294.

Narbón-Perpiñá, I., De Witte, K., 2018a. Local governments’ efficiency: a systematic literature review—part I. *International Transactions in Operational Research* 25, 2, 431–468.

Narbón-Perpiñá, I., De Witte, K., 2018b. Local governments’ efficiency: a systematic literature review—part II. *International Transactions in Operational Research* 25, 4, 1107–1136.

Natesan, S., Marathe, R., 2017. Evaluation of MGNREGA: data envelopment analysis approach. *International Journal of Social Economics* 44, 4, 181–194.

Nielsen, S., Bayne, E., Schieck, J., Herbers, J., Boutin, S., 2007. A new method to estimate species and biodiversity intactness using empirically derived reference conditions. *Biological Conservation* 137, 3, 403–414.

Olley, G., Pakes, A., 1992. The dynamics of productivity in the telecommunications equipment industry. Technical Report, National Bureau of Economic Research, Cambridge, MA.

Orme, C., Smith, P., 1996. The potential for endogeneity bias in data envelopment analysis. *Journal of the Operational Research Society* 47, 1, 73–83.

Pan, X., Pan, X., Li, C., Song, J., Zhang, J., 2019. Effects of China’s environmental policy on carbon emission efficiency. *International Journal of Climate Change Strategies and Management* 11, 3, 326–340.

Papadimitriou, M., Johnes, J., 2019. Does merging improve efficiency? A study of English universities. *Studies in Higher Education* 44, 8, 1454–1474.

Park, J., Kim, J., 2021. The impact of airport managerial type and airline market share on airport efficiency. *Sustainability (Switzerland)* 13, 2, 1–16.

Priscilla, L., Chauhan, A., 2019. Economic impact of cooperative membership on dairy farmers in Manipur: a propensity score matching approach. *Agricultural Economics Research Review* 32, 117–123.

Puenpatom, R., Rosenman, R., 2008. Efficiency of Thai provincial public hospitals during the introduction of universal health coverage using capitation. *Health Care Management Science* 11, 4, 319–338.

© 2021 The Authors.

*International Transactions in Operational Research* published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies
Rogge, N., 2012. Undesirable specialization in the construction of composite policy indicators: the environmental performance index. *Ecological Indicators* 23, 143–154.
Rosenbaum, P., 1989. Optimal matching for observational studies. *Journal of the American Statistical Association* 84, 408, 1024–1032.
Rosenbaum, P., Rubin, D., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39, 1, 33–38.
Rubin, D., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66, 5, 688.
Sahin, I., Ozcan, Y., Ozgen, H., 2011. Assessment of hospital efficiency under health transformation program in Turkey. *Central European Journal of Operations Research* 19, 1, 19–37.
Santín, D., Sicilia, G., 2018. Using DEA for measuring teachers’ performance and the impact on students’ outcomes: evidence for Spain. *Journal of Productivity Analysis* 49, 1, 1–15.
Sargan, J., 1958. The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society* 26, 3, 393–415.
Sav, G., 2012. Frontier and envelopment evaluations of university graduation efficiencies and productivities: elements for performance-based funding. *Problems and Perspectives in Management* 10, 4, 71–79.
Schubert, T., Yang, G., 2016. Institutional change and the optimal size of universities. *Scientometrics* 108, 3, 1129–1153.
Seiford, L., 1996. Data envelopment analysis: the evolution of the state of the art (1978–1995). *Journal of Productivity Analysis* 7, 2, 99–137.
Seth, P., Patel, G., 2014. Measuring the efficiency of Rashtriya Swasthaya Bima Yojana using data envelopment analysis approach. *Journal of Health Management* 16, 2, 199–215.
Sevigny, E., Meylakhs, P., Feizollahi, M., Amini, M., 2020. Development of a global index measuring national policy commitments to HIV prevention and treatment among people who inject drugs. *International Journal of Drug Policy* 84, 102877.
Shi, P., Yang, J., 2008. The effectiveness of national R&D investment—an empirical investigation in China. *Problems and Perspectives in Management* 6, 4, 4–11.
Simar, L., Wilson, P., 1998. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science* 44, 1, 49–61.
Simar, L., Wilson, P., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics* 136, 1, 31–64.
Simar, L., Wilson, P., 2011. Inference by the m out of a bootstrap in nonparametric frontier models. *Journal of Productivity Analysis* 36, 1, 33–53.
Soares de Mello, J., Gomes, E., Meza, L., Soares de Mello, M., Soares de Mello, A., 2006. Engineering post-graduate programmes: a quality and productivity analysis. *Studies in Educational Evaluation* 32, 2, 136–152.
Sofi, I., Sharma, P., 2015. Does employment protection legislation matter in dualistic labor market? Panel evidence from Indian manufacturing sector. *Labor Studies Journal* 40, 2, 149–168.
Tittel, V., De Witte, K., 2021. How politics influence public good provision. *Socio-Economic Planning Sciences*, https://doi.org/10.1016/j.seps.2020.101000.
Tochkov, K., Nenovski, N., Tochkov, K., 2012. University efficiency and public funding for higher education in Bulgaria. *Post-Communist Economies* 24, 4, 517–534.
Tone, K., 2001. A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research* 130, 3, 498–509.
Tulkens, H., 2006. On FDH efficiency analysis: some methodological issues and applications to retail banking, courts and urban transit. In Chander, P., Drèze, J., Lovell, C.K., Mintz, J. (eds) *Public Goods, Environmental Externalities and Fiscal Competition*. Springer, Boston, MA, pp. 311–342.
Vieira, D., Hoffmann, V., Alberton, A., 2018. Public investment, competitiveness and development: a study into brazilian tourism destinations [investimentos públicos, competitividad e desenvolvimento: Um estudo em destinos turísticos brasileiros] [ inversiones públicas, competitividad y desarrollo: Un estudio en destaciones turísticas brasileñas]. *Revista de Administracao Publica* 52, 5, 899–917.
Waldo, S., 2007. Efficiency in Swedish public education: competition and voter monitoring. *Education Economics* 15, 2, 231–251.

© 2021 The Authors. *International Transactions in Operational Research* published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies
Wang, P., Lin, X., Li, S., 2019. Evaluation of government-supported provincial innovation competitiveness—evidence from China. *Growth and Change* 50, 2, 587–608.

Whittaker, G., 1994. The relation of farm size and government programme benefits: an application of data envelopment analysis to policy evaluation. *Applied Economics* 26, 5, 469–478.

Whittaker, G., Färe, R., Grosskopf, S., Barnhart, B., Bostian, M., Mueller-Warrant, G., Griffith, S., 2017. Spatial targeting of agri-environmental policy using bilevel evolutionary optimization. *Omega (United Kingdom)* 66, 15–27.

Whittaker, G., Färe, R., Srinivasan, R., Scott, D., 2003. Spatial evaluation of alternative nonpoint nutrient regulatory instruments. *Water Resources Research* 39, 4, WES11–WES19.

Wohlin, C., 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. ACM, New York, pp. 1–10.

Worthington, A., 2001. An empirical survey of frontier efficiency measurement techniques in education. *Education Economics* 9, 3, 245–268.

Worthington, A., 2004. Frontier efficiency measurement in health care: a review of empirical techniques and selected applications. *Medical Care Research and Review* 61, 2, 135–170.

Wright, P., 1928. *Tariff on Animal and Vegetable Oils*. Macmillan, New York.

Xin-gang, Z., Xin, M., Ying, Z., Pei-ling, L., 2020. Policy inducement effect in energy efficiency: an empirical analysis of China. *Energy* 211, https://doi.org/10.1016/j.energy.2020.118726.

Yang, G.L., Fukuyama, H., Song, Y.Y., 2018. Measuring the inefficiency of Chinese research universities based on a two-stage network DEA model. *Journal of Informetrics* 12, 1, 10–30.

Zabala-Iturriagagoitia, J., Voigt, P., Gutiérrez-Gracia, A., Jiménez-Sáez, F., 2007. Regional innovation systems: how to assess performance. *Regional Studies* 41, 5, 661–672.

Zhang, D., Banker, R., Li, X., Liu, W., 2011. Performance impact of research policy at the Chinese academy of sciences. *Research Policy* 40, 6, 875–885.

Zhang, N., Unruh, L., Wan, T., 2008. Has the medicare prospective payment system led to increased nursing home efficiency? *Health Services Research* 43, 3, 1043–1061.

Zhang, Y., Li, X., Jiang, F., Song, Y., Xu, M., 2020. Industrial policy, energy and environment efficiency: evidence from Chinese firm-level data. *Journal of Environmental Management* 260, https://doi.org/10.1016/j.jenvman.2020.110123.