UNIVERSIDAD DE EXTREMADURA

TESIS DOCTORAL

NUEVAS ALTERNATIVAS DE MEDICIÓN DE LA EFICIENCIA: EVALUACIÓN MEDIANTE DATOS SIMULADOS Y APLICACIONES EN EL SECTOR PÚBLICO

NEW ALTERNATIVES FOR EFFICIENCY MEASUREMENT: ASSESSMENT WITH SIMULATED DATA AND APPLICATIONS IN THE PUBLIC SECTOR

Cristina Polo Fernández

PROGRAMA DE DOCTORADO EN ECONOMÍA Y EMPRESA

2018

UNIVERSIDAD DE EXTREMADURA

THESIS

NEW ALTERNATIVES FOR EFFICIENCY MEASUREMENT: ASSESSMENT WITH SIMULATED DATA AND APPLICATIONS IN THE PUBLIC SECTOR

Cristina Polo Fernández

PhD PROGRAM IN ECONOMICS AND BUSINESS

2018

UNIVERSIDAD DE EXTREMADURA

TESIS DOCTORAL

Título:

NUEVAS ALTERNATIVAS DE MEDICIÓN DE LA EFICIENCIA: EVALUACIÓN MEDIANTE DATOS SIMULADOS Y APLICACIONES EN EL SECTOR PÚBLICO

NEW ALTERNATIVES FOR EFFICIENCY MEASUREMENT: ASSESSMENT WITH SIMULATED DATA AND APPLICATIONS IN THE PUBLIC SECTOR

Autora:

Cristina Polo Fernández

Conforme, los directores:

Dr. D. José Manuel Cordero Ferrera **Dr. D. Julián Ramajo Hernández**

PROGRAMA DE DOCTORADO EN ECONOMÍA Y EMPRESA

Badajoz, 2018

“If you can’t explain it simply, you don’t understand it well enough”

A. Einstein, physicist.

A mi familia.

To my family.

RESUMEN EN ESPAÑOL

La provisión de servicios públicos del modo más eficiente se ha convertido en un objetivo fundamental en el ámbito de la economía pública, especialmente desde la entrada de la reciente crisis financiera. En particular, existe un creciente interés por obtener medidas de la eficiencia relativa de las unidades de gestión pública que proporcionan un mismo servicio a los ciudadanos. La presente tesis doctoral se centra en analizar una cuestión de máxima relevancia dentro de este ámbito, como es la consideración del efecto que puedan tener las denominadas variables exógenas, contextuales o ambientales. Estas variables están al margen del control de las unidades evaluadas, pero pueden influir en gran medida sobre el nivel de producción y el consumo de recursos, por lo tanto, deben ser tenidas en cuenta en la estimación de las medidas o índices de eficiencia.

El tema analizado cuenta con una larga tradición en la literatura previa, tanto en el ámbito de las aproximaciones paramétricas como no paramétricas. De hecho, en las últimas décadas se han desarrollado diferentes opciones metodológicas para tal fin. Sin embargo, hasta el momento sigue sin existir consenso entre los investigadores sobre qué alternativa es la mejor opción desde el punto de vista teórico (Cordero et al., 2008), lo que sumado al hecho de que cada una de ellas ofrece resultados muy distintos al ser aplicadas con datos reales (Muñiz, 2002; Yang y Pollit, 2009), convierte la elección entre las diferentes alternativas en una tarea extremadamente compleja (Huguenin, 2015).

Las principales aportaciones de esta tesis doctoral están fundamentalmente estructuradas en dos líneas claramente especializadas. El primer capítulo está centrado en la comparación de varias alternativas que incorporan el efecto de las variables exógenas en la estimación de la eficiencia a través de un experimento de Monte Carlo. Los dos capítulos siguientes componen la parte empírica de la tesis, donde se estudia cómo medir la eficiencia técnica en dos ámbitos diferentes de la economía pública tomando en consideración la influencia de las variables de contexto. Todas estas contribuciones han sido publicadas en revistas indexadas en la clasificación *Journal of Citation Reports*, tras someterse a un proceso de evaluación anónimo.

En particular, el capítulo 1 evalúa el modelo condicional no paramétrico desarrollado por Daraio y Simar (2005, 2007) frente a las metodologías tradicionales de una y dos etapas. La herramienta de análisis es un experimento de Monte Carlo, en el que se generan datos simulados empleando una forma funcional flexible de tipo translogarítmica para describir

la función de producción. Según los resultados obtenidos con diferentes criterios estadísticos de comparación, se concluye que el modelo condicional es mucho más preciso que sus antecesores cuando se pretenden obtener índices de eficiencia incorporando las variables contextuales al análisis y, en concreto, si se desean identificar las unidades eficientes situadas en la frontera en muestras de tamaño relativamente pequeño. Basándonos en estos resultados, este enfoque es el elegido en los dos capítulos siguientes para estimar medidas del desempeño de diferentes unidades de producción encargadas de la provisión de servicios públicos.

En la primera de los dos estudios empíricos (capítulo 2 de esta tesis), se estiman índices de eficiencia global para una muestra de 154 municipios catalanes en el período comprendido entre 2005 y 2012. En el análisis se incorporan como variables de contexto diferentes indicadores representativos de las principales características socio-económicas y demográficas municipales. Además, se adapta la metodología a un contexto dinámico a través del cual es posible analizar la evolución temporal de la eficiencia a lo largo del período evaluado. Los resultados muestran que la eficiencia promedio de las unidades experimentó un marcado descenso durante los años previos a la crisis económica, aunque esa tendencia se revierte a partir de 2010. Asimismo, se comprueba que las tres variables exógenas consideradas (el PIB *per cápita*, la tasa de desempleo y la capacidad de consumo) tienen un efecto significativo sobre los niveles de eficiencia de los municipios. Por último, se observa que los municipios de mayor tamaño poblacional resultan ser los más eficientes en el suministro de servicios públicos a lo largo de todo el período.

En el segundo trabajo empírico con datos reales (capítulo 3) se emplea la versión robusta del modelo condicional para evaluar la eficiencia de una amplia muestra de escuelas de enseñanza secundaria de 36 países que participaron en PISA 2012, con el propósito de construir un ranking y analizar el efecto de las divergencias entre los diferentes sistemas educativos. En este caso, la heterogeneidad entre las unidades viene determinada por factores representativos tanto del contexto escolar específico de cada centro, como de las características del sistema educativo y la cultura del país al que pertenece. Con el fin de profundizar en el análisis, se desarrolla una extensión muy reciente de la metodología propuesta que consiste en un modelo de regresión de segunda etapa cuyo objetivo es “limpiar” los índices de eficiencia del efecto de las variables exógenas. Los principales resultados ponen de manifiesto una mayor heterogeneidad entre países que entre las

escuelas. En concreto, las diferencias fundamentales detectadas en las estimaciones de la eficiencia vienen explicadas esencialmente por indicadores económicos y las variables representativas de los valores culturales del país.

Para concluir, los resultados de esta tesis pueden considerarse una contribución útil para analistas en el ámbito de la medición de la eficiencia en el sector público. De hecho, se ponen de manifiesto las cualidades del modelo condicional frente a otras alternativas previas para incorporar el efecto de los factores exógenos en la estimación de los índices de eficiencia.

CONTENTS

INTRODUCTION

The provision of public services in the most efficient way possible has become a universal target of central importance in economic policy, especially since the arrival of the recent financial crisis. A central concern is to measure the relative efficiency of different public entities providing the same public service. The traditional productivity literature defines measures of organisational efficiency as the distance of the unit under scrutiny from a frontier function, which is estimated using the best observed practice of the set of other similar units.

Based on the ideas originally developed by Koopmans (1951) and Debreu (1951), Farrell (1957) introduced the seminal concept of technical efficiency by developing a radial measure which permits to quantify the level of efficiency of the producers in terms of percentages. Nowadays, a wide spectrum of methodologies exists to measure efficiency. Although they can be classified according to different criteria, the most common is the distinction between parametric and nonparametric approaches. In the former, the production technology is defined by specifying *a priori* a functional form with constant parameters to be estimated (i.e. Cobb-Douglas, translog, etc.). Stochastic-type frontiers are generally used where two components are identified in the residuals: inefficiency and a noise term. In contrast, nonparametric approaches formulate the technological features from certain assumptions about the structure of the production technology (e.g. free disposability, convexity, etc).

Since Aigner and Chu's (1968) pioneering work, a large number of studies have been carried out with parametric models. One of the main advantages is that parameters can be obtained and interpreted by applying econometric techniques, thus the estimators of the inefficiency have well-known statistical properties. Nonetheless, some of the principal disadvantages of that approach are the treatment of the function when more of one output wants to be incorporated to the model and the errors caused by misspecification of the functional shape of the production function (Kuosmanen et al., 2015). The stochastic production function was independently proposed by Aigner et al., (1977) and Meeusen and Van den Broeck, (1977) and it was usually estimated by maximum likelihood models. It is the well-known *Stochastic Frontier Analysis* (SFA).

Nonparametric models do not require the imposition of a concrete functional form for the production function, so defining a set of mathematical properties might be enough. This

feature implies a higher flexibility that justifies why these methods are widely used when the production process is very difficult to represent, as it often happens in the context of public services. The main problem is that any deviation from the efficient frontier is considered an inefficient performance of the producers, thus the estimations are very sensitive to outliers (or extreme points) and measurement errors. Mathematical programming problems are the usual techniques applied to estimate efficiency, among which the most representative are the Data Envelopment Analisys (DEA) proposed by Charnes et al. (1978) and the Free Disposal Hull (FDH) developed by Deprins et al. (1984).

In most empirical studies using data about public services, the results of the production process depend of variables that are beyond the control of producers commonly known as contextual, exogenous or environmental variables (or simply Z variables). These factors do not take part on the production process, but they have an impact over the production level as well as on use of inputs. Therefore, they must be taken into account in the efficiency estimation and must be specifically treated according to their particular features. This is the only way to reflect properly whether producers are actually performing efficiently or there are factors that do not allow them to achieve the production objectives that are feasible for others even doing their best.

Incorporating the effects of these variables in the estimation of efficiency measures has been one of the topics that has generated more controversial discussions in both the parametric and nonparametric literature, since there are multiple approaches that can be used to deal with them (Fried et al., 2008; Badin et al., 2014).

Under a parametric perspective, the most accepted approach among researches is the stochastic model proposed by Kumbhakar et al. (1991) and Battese and Coelli (1995), in which the inefficiency term depends on a vector of exogenous factors[1]. In addition, other models that permit to introduce the effect of these variables in the construction of the frontier have also been developed recently (Zhang et al., 2012). However, the main problem of these approaches is that they are not very flexible, since they impose a specific

[1] Kumbhakar et al. (1991) carried out that model in a cross-sectional data context, while Battese and Coelli (1995) adapted it to a panel data scenario.

functional form to the evaluated production process, which difficult its application on contexts with a hard modelling process.

In the nonparametric framework, multiple models have also been proposed to incorporate the effect of Z variables. The most well-known model is surely the proposed approximation by Banker and Morey (1986), adapting the original DEA formulation to explicitly consider the exogenous character of some variables. Despite its popularity, it presents some important limitations such as the fact that it requires knowing *a priori* the direction of the effect of the exogenous variables (Fried et al., 2002) or the possible skewness in the estimated efficiencies (Cordero et al., 2008).

Other well-known alternative is the second-stage model, which consists of running a simple DEA in a first stage without considering exogenous variables and, subsequently, including the estimated efficiency scores as the dependent variable in a regression with contextual factors as explanatory variables. Some of the most relevant proposals adopting this framework are Ray (1991), Hoff (2007), Simar and Wilson (2007, 2011) and Banker and Natarajan (2008). Likewise, this approach can be extended by including additional stages in which the effect of contextual variables can be incorporated into the estimation of efficiency scores by adjusting the original values of inputs and/or outputs (e.g. Fried et al., 1999; Fried et al., 2002; Muñiz, 2002; Lozano-Vivas et al., 2002; Avkiran and Rowlands, 2008 or Cordero et al., 2010). The main problem of all these methods is that they require assuming the restrictive separability condition, i. e., that exogenous variables only affect the probability of being more or less efficient and not the shape of the frontier, which is difficult to be held in many cases[2]. This problem can be avoided using the conditional nonparametric model proposed by Cazals et al. (2002) and extended by Daraio and Simar (2005, 2007), which provides a solution based on the use of a probabilistic formulation that allows for a nonparametric estimation of the efficiency measures by incorporating the effect of the exogenous variables in a single stage.

The existence of this wide range of methodological alternatives has led many authors to wonder which approach should be applied to deal with contextual variables when performing an efficiency analysis. Until now, researchers have not reached an agreement

[2] Daraio et al. (2018) propose several tests to prove if separability actually exists between the input-output space and the space of the exogenous variables.

about what alternative performs better from a theoretical point of view (Cordero et al., 2008). This is a great concern, since they usually provide very different results in empirical applications (Muñiz, 2002; Yang and Pollit, 2009). Therefore, the election among the different methodologies becomes an extremely complex task (Huguenin, 2015).

In this PhD thesis, we are interested in exploring which alternative perform better. For that purpose, we rely on simulated data generated through a Monte Carlo experiment where the production technology is known and efficiency measures can be compared with the true efficiency level generated from an experimental design. In the literature, we can find many previous works using this tool to assess the performance of different approaches to incorporate the influence of contextual variables into efficiency measures. Muñiz et al. (2006), Cordero et al. (2009), Estelle et al. (2010) or Harrison et al. (2012) are some examples comparing nonparametric techniques, while, parametric models have been compared either between them or with nonparametric alternatives (Wang and Schmidt, 2002; Krüger, 2012). Nevertheless, to the best of our knowledge, there is no previous study in the literature comparing the performance of the conditional approach with traditional methods[3].

Therefore, the objectives of this thesis are fundamentally structured in two main lines. In the first chapter we compare the performance of the conditional DEA estimator with some traditional approaches such as the one-stage and two-stage methods using a flexible *translog* functional form to describe the production function. According to different accuracy criteria, we conclude that the conditional DEA clearly outperforms all the traditional alternatives. Subsequently, in the following two chapters, we present two empirical studies using this approach to measure the performance of different public organizations. All these chapters have been already published in journals included in the Journal of Citation Reports classification.

In the first empirical study, we estimate efficiency measures for a sample of Catalan municipalities over an eight-year period (2005-2012). The effect of a set of

[3] The accuracy of this approach was only previously tested in the original work in which it as developed by comparing efficiencies estimated with this method and the real efficiency generated in a Monte Carlo experiment (Daraio and Simar, 2007).

socioeconomic variables over the performance of the municipalities is analysed as well as the influence of the time by extending the model in a dynamic context. In the second application with real data we apply a robust nonparametric approach to analyze the performance of a set of secondary schools from 36 countries participating in PISA 2012. The effects of contextual factors at both school and country level are taken into account in the efficiency estimation. As an extension of that model, a second-stage regression analysis was carried out with the aim of calculating the "pure efficiency", i. e., the efficiency scores cleaned by the effect of the exogenous variables.

To conclude, the last section of this thesis presents a summary of the main general conclusions derived from each chapter as well as some suggestions and guidelines for further research.

CHAPTER 1

A MONTE-CARLO COMPARISON OF CONDITIONAL NONPARAMETRIC METHODS AND TRADITIONAL APPROACHES TO INCLUDE EXOGENOUS VARIABLES

CHAPTER 1

Title:

"MONTE-CARLO COMPARISON OF CONDITIONAL NONPARAMETRIC METHODS AND TRADITIONAL APPROACHES TO INCLUDE EXOGENOUS VARIABLES"

Authors and affiliation:

José M. Cordero[1], Cristina Polo[1], Daniel Santín[2], Gabriela Sicilia[3]

[1] Universidad de Extremadura
[2] Universidad Complutense de Madrid
[3] Universidad Autónoma de Madrid

Journal:

Pacific Economic Review, 21: 4 (2016), pp: 483-497

DOI:

https://doi.org/10.1111/1468-0106.12194

Abstract:

The aim of this paper is to compare the performance of the conditional nonparametric approach with several traditional nonparametric methods to incorporate the effect of exogenous or environmental variables into the estimation of efficiency measures. To do this, we conduct a Monte Carlo experiment using a translog production function with one output, two discretionary inputs and two exogenous variables to generate simulated data. According to the values of different accuracy measures calculated to evaluate the performance of each method, the conditional data envelopment analysis clearly outperforms all the traditional alternatives.

CHAPTER 2

MEASURING EFFICIENCY IN CATALAN MUNICIPALITIES USING A DYNAMIC CONDITIONAL MODEL

CHAPTER 2

Title:

"MEASURING EFFICIENCY IN CATALAN MUNICIPALITIES USING A DYNAMIC CONDITIONAL MODEL"

Authors:

José M. Cordero[*], Carlos Díaz[*], Cristina Polo[*]

[*] Universidad de Extremadura

Journal:

Revista de Economía Aplicada, Tomo 25, Nº 75 (2017), pp: 29-51

DOI:

https://search.proquest.com/docview/1986332104?accountid=17204

Abstract:

The aim of this paper is to estimate efficiency measures for a sample of Catalan municipalities over an eight-year period (2005-2012) ranging from the years before the economic crisis to the start of recovery. To do this, we built a panel database for 154 municipalities with a population of from 5,000 to 50,000 inhabitants. The methodology used in the empirical analysis is a non-parametric dynamic conditional model that can be used to account for both the time dimension and a set of socioeconomic variables that might have an influence on the performance of the municipalities. The results show a remarkable decrease in efficiency levels during the years prior to the economic crisis, although this trend reversed as of 2010.

CHAPTER 3

EFFICIENCY MEASUREMENT AND CROSS-COUNTRY DIFFERENCES AMONG SCHOOLS: A ROBUST CONDITIONAL NONPARAMETRIC ANALYSIS

CHAPTER 3

Title:

"EFFICIENCY MEASUREMENT AND CROSS-COUNTRY DIFFERENCES AMONG SCHOOLS: A ROBUST CONDITIONAL NONPARAMETRIC ANALYSIS"

Authors:

José M. Cordero[1], Cristina Polo[1], Daniel Santín[2], Rosa Simancas[1]

[1] Universidad de Extremadura
[2] Universidad Complutense de Madrid

Journal:

Economic Modelling, *forthcoming*

DOI:

https://doi.org/10.1016/j.econmod.2018.05.001

Abstract:

Analyzing the efficiency of educational systems is one of the main focuses of the policy debate to promote national competitiveness and future economic growth. In this paper, we assess the performance of secondary schools from 36 countries (26 OECD countries and 10 partners) participating in PISA 2012. For this purpose, we apply a robust conditional nonparametric approach that allows us to incorporate the effect of contextual factors at both school and country level in the estimation of efficiency measures. Our results suggest that there is a greater heterogeneity across countries than across schools. Particularly, we find that differences in efficiency estimates are mainly explained by economic indicators and cultural values. In contrast, some factors previously identified as potential determinants of student achievement, like the existence of tracking or central examinations, do not seem to significantly affect the efficiency of secondary schools.

SUMMARY OF RESULTS AND FINAL CONCLUSIONS

This doctoral thesis is focused on providing solutions to practitioners interested in assessing the efficiency of public service providers when their production is affected by external or contextual factors that are beyond their control. This has been a hot topic in the efficiency measurement literature during the last three decades, in which many methodological proposals have emerged with the aim of including the effect of such variables in the model. Our main emphasis is placed on the nonparametric conditional approach developed by Daraio and Simar (2005, 2007), which constitutes an appealing alternative for that purpose that have started to be applied in multiple empirical studies in different sectors during the last years.

Using data simulated through a Monte Carlo experiment, in the first chapter of this thesis we have compared the performance of this method with other traditional approaches in order to determine which alternative provides more accurate estimations. Our results indicate that the conditional model clearly outperforms all the other alternatives, especially when practitioners are interested in identifying the truly efficient units in relatively small samples. Therefore, we advocate for the use of this methodology in empirical studies.

In Chapters 2 and 3 we present two papers focused on measuring efficiency in the delivery of different public services using real data from different sectors in which contextual factors have a relevant role, thus they need to be taken into account in the estimation of efficiency measures. In particular, we have assessed the global performance of a sample of Spanish municipalities and, subsequently, schools operating in different educational systems around the world. The relevance of estimating efficiency in these two areas is supported by the existence of an extensive number of publications throughout the last decades (De Witte and López-Torres, 2017; Narbón-Perpiñá and De Witte, 2018a, 2018b).

The main purposes of the study presented in Chapter 2 are to analyze the effects of the economic crisis on the global efficiency of local governments in Spain and identify the main determinants of their performance. Our dataset comprises data about 154 municipalities from the Spanish region of Catalonia over the 2005-2012 period. We consider several variables representing the socio-economic and demographic characteristics of the municipalities and adapt the conditional approach to a dynamic

environment by including the time factor as an additional contextual variable, so it is possible to analyse the efficiency trend over the evaluated period. The results reveal that the average efficiency of units shows a remarkable decrease in efficiency levels during the years prior to the economic crisis, although this trend reversed as of 2010. In addition, three variables representing the socio-economic context (per capita income, unemployment rate and consumption capacity) are found to have a significant effect on municipal efficiency levels. Likewise, we notice that larger municipalities are more efficient in the provision of municipal services.

In Chapter 3 we have examined the performance of a huge set of schools around the world using PISA 2012 data with the aim of building a ranking of countries according to their performance as well as exploring the existing divergences in terms of efficiency across different educational systems. For that purpose, we have applied the robust version of the conditional approach that allows us to account for the existing heterogeneity both in the school context and due to some potential specific country-level features that might also affect school performance. In addition, we have performed a recent development of this approach which consists of implementing a second-stage regression model in order to "clean" the efficiency measures from the effect of the exogenous variables. The main findings suggest that there is a greater heterogeneity across countries than across schools since the country ranking is hardly affected by the consideration of school contextual factors. Specifically, the differences detected in efficiency estimates are mainly explained by economic indicators and cultural values.

From our point of view, the results of this thesis can be considered as a helpful contribution for practitioners in the field of efficiency measurement, since we provide evidence on the virtues of the conditional method over other alternatives to incorporate the effect of contextual factors into estimated measures of performance. Therefore, we should recommend researchers working with this type of data to use this approach. Nevertheless, it is also worth mentioning that the implementation of this method is somewhat complex, since there is no statistical program yet available to facilitate its implementation, thus it requires manually programming the codes in some specific statistical package like R or MATLAB. Moreover, when the size of the dataset is huge, as in our case in Chapter 3, the estimation process can be time consuming.

Likewise, the results of our empirical studies that can also be useful from a policy point of view, since they show clearly that the conclusions that can be derived from them depend on a great extent on whether contextual factors are taken into account or not in the estimation of efficiency measures of performance. Therefore, if policy makers are interested in implementing some policies according to these measures, i.e. distribute subsidies to the most or least efficient units, they should be aware that they should represent real targets of performance that can be achieved considering the context in which they are operating.

Finally, we would like to make clear that our research does not end at this point, since we are already exploring other alternative approaches that have emerged in the most recent literature to account for the effect of contextual variables such as the semi-nonparametric StoNEZD method developed by Johnson and Kuosmanen (2011) as well as the extension of this method proposed by Kuosmanen and Johnson (2017) to adapt it to the multi-output framework by using directional distance functions. Likewise, we are also working on the development of a new methodological approach in collaboration with other researchers based on a two-stage procedure that allows accounting for the exogenous variables in such a way that the separability condition is not assumed. Unfortunately, the assessment of this new approach is still unfinished, but the first results are quite promising and it will be presented to interested readers shortly.

REFERENCES

(Not included on any chapter)

Aigner, D.J., Chu, S. F. (1968). On estimating the industry production function. *The American Economic Review*, 58(4), 826-839.

Avkiran, N.K., Rowlands, T. (2008). How to better identify the true managerial performance: state of the art using DEA. *Omega*, 36(2), 317-324.

Cordero, J.M., Pedraja, F., Santin, D. (2010). Enhancing the inclusion of non-discretionary inputs in DEA. *Journal of Operational Research Society*, 61, 574-584.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19, 273-92.

Dios, R., De Haro, T., Martínez Paz, J.M. (2005): *Análisis de eficiencia de las industrias oleícolas. Un estudio con variables de entorno*. Ponencia presentada en el II Congreso de Eficiencia y Productividad, EFIUCO, Córdoba.

Fried, H., Schmidt, S., Yaisawarng, S. (1999). Incorporating the Operating Environment into a Nonparametric Measure of Technical Efficiency. *Journal of Productivity Analysis*, 12, 249-267.

Koopmans, T.C. (1951). *Activity analysis of production and allocation* (Nº. 13). New York: Wiley.

Kuosmanen, T. (2008). Representation theorem for convex nonparametric least squares. *Econometric Journal*, 11, 308-325.

Kuosmanen, T., Kortelainen, M. (2007). Stochastic nonparametric envelopment of data: Cross-sectional frontier estimation subject to shape constraints. Available at SSRN: https://ssrn.com/abstract=983882 or http://dx.doi.org/10.2139/ssrn.983882.

Kuosmanen, T., Johnson, A.L., Saastamoinen, A. (2015). Stochastic nonparametric approach to efficiency analysis: A unified framework. In Zhu, J. (Ed.). *Data Envelopment Analysis. A Handbook of Models and Methods*, Springer: New York, 191-244.

Kumbhakar, S.C., Ghosh, S., McGuckin, J.T. (1991). A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. *Journal of Business & Economic Statistics*, 9(3), 279-286.

Krüger, J.J. (2012). A Monte Carlo study of old and new frontier methods for efficiency measurement. *European Journal of Operational Research*, *222*(1), 137-148.

Lozano-Vivas, A., Pastor, J.T., Pastor, J.M. (2002). An efficiency comparison of European banking systems operating under different environmental conditions, *Journal of Productivity Analysis*, 18(1), 59-77.

Meeusen, W., Van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 435-444.

Wang, H.J., Schmidt, P. (2002). One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels. *Journal of Productivity Analysis*, 18, 129-144.

Zhang, R., Sun, K., Delgado, M.S., Kumbhakar, S.C. (2012). Productivity in China's high technology industry: Regional heterogeneity and R&D. *Technological Forecasting and Social Change*, 79(1), 127-141.