Modern Management based on Big Data III
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220079

Nonparametric Estimation of the Production Frontier Using a Data-Fitting Technique

Yu ZHAO [a,1]

[a] *School of Management, Department of Management, Tokyo University of Science*

Abstract. We propose a novel nonparametric approach for estimating the production frontier based on a data-fitting technique. The proposed approach allows for stochastic noise and provides decision-makers with a general and flexible estimation procedure to support and facilitate the decision-making process in the stochastic context. A major feature of the proposed approach is that the estimation procedure is completely nonparametric and easy to implement. Similar to other existing nonparametric approaches, our proposed approach results in an estimate of the piece-wise linear production frontier. In contrast to the existing ones, the evaluation of each data point is performed within a unit-specific data range. We also propose a naive method for determining the data ranges of each data point. The performance of our proposed approach is examined using various simulated scenarios. For each scenario, we compare our proposed approach with the existing methods, including the data envelopment analysis (DEA), the stochastic nonparametric envelopment of data (StoNED), and the stochastic frontier analysis (SFA). The simulation results suggest that our approach performs better than the existing methods in the single input and single output case. Our proposed approach can also be easily extended to a multi-input setting. Moreover, the proposed naive method on data ranges also shows its flexibility and usefulness in the simulated examples.

Keywords. production frontier, nonparametric estimation, data-fitting technique, downside deviation, stochastic noise

1. Introduction

Estimating production frontiers is essential for performance benchmarking and productivity analysis. The current approaches include the data envelopment analysis (DEA,e.g., [1,2]), the stochastic nonparametric envelopment of data (StoNED,e.g., [3, 4]), and the stochastic frontier analysis (SFA,e.g., [5,6]). DEA is a nonparametric method for measuring the relative efficiencies of peer decision-making units (DMUs). Because DEA ignores stochastic noise, the frontier estimated by DEA is entirely deterministic, suggesting that any deviations from the frontier (e.g., gauging the distance to the boundary of the production technology) can be considered a measure of pure inefficiency. By contrast, SFA is known as a general stochastic parametric approach. It accounts for

[1]Corresponding Author: Yu Zhao, School of Management, Department of Management, Tokyo University of Science, Japan; E-mail: yu.zhao@rs.tus.ac.jp

stochastic noise by treating all deviations from the frontier as aggregations of both inefficiency and noise. However, compared with the flexibility of nonparametric measurements, SFA relies heavily on an accurately pre-specified functional form for production technology. To combine both advantages of DEA (i.e., nonparametric estimation) and SFA (i.e., stochastic noise), StoNED has been introduced in the literature on efficiency analysis [7]. Because StoNED formulates a quadratic programming problem, solving large-scale problems requires more efficient computational algrithms [8].

This study proposes a nonparametric approach for estimating the production frontier. The proposed approach is based on a data-fitting technique and accounts for stochastic noise. The commonly used regression technique is easy to implement using ordinary least squares regression (OLS). A significant feature of the regression technique lies in that the estimated coefficients characterized a hyperplane passing the barycenter of all data points. Contrary to the regression lines (or hyperplanes), the production frontier identifies a set of maximum producible outputs at a given level of inputs. To obtain an upper bound of the observed level of outputs, we modify the OLS problem as a minimization problem of downside deviations [9] by restricting all residuals to be nonnegative. However, in practice, an infinite number of upper bounds could exist as it is challenging to observe the same DMU multiple times. Such an issue can be interpreted as a missing data problem. Methodologically, we propose a naive method that considers the similarity of the observed data point, and the upper bound of the observed point is estimated by using its neighborhood points. After the upper bound of each data point is estimated, the following procedure becomes a problem of choosing correctly estimated coefficients that can be used to construct a non-decreasing concave function. We show that such a problem can be solved by Afriat's inequalities [10]. Finally, we can derive an estimator of the piece-wise linear production frontier by using the lower bound of the maximum producible outputs corresponding to the selected estimated coefficients.

Recently, methods such as DEA, SFA, and StoNED have become important analytical tools in expert and intelligent systems that facilitate decision-making by policy-makers or regulators (for recent relevant studies, see, e.g., [11,12,13,14]). This study provides an alternative approach for supporting the decision-making process from the methodological perspective. Moreover, our proposed approach overcomes several drawbacks of the existing approaches and provides decision-makers with a general and flexible modeling approach in the stochastic context. We will discuss this in detail in Sections 2 and 4.

The remainder of the paper is structured as follows. Section 2 reviews previous related research on the production frontier estimation. Section 3 provides methodological details. In Section 4, we use six simulated scenarios to examine the flexibility and usefulness of the proposed estimation procedure. Section 5 concludes the paper.

2. Literature review

This section briefly reviews major existing approaches for estimating production frontiers in the field of productivity analysis and efficiency measurement. Understanding the level of efficiency and productivity is essential for fostering continuous improvement of the production and operation management in a production activity (e.g., firm, school, government, nonprofit organization, etc.). During the past several years, the term "best practice" has been widely used in both the theoretical and practical benchmarking literature. A popular idea for identifying best practices is to model the frontier of the production technology, which is a mathematical description of the relationship between inputs and outputs. The theoretical literature on production theory imposes several basic axioms of production technology, such as free disposability, convexity, and returns to scale (see, e.g., [10,15,16,17]). The frontier estimated in this study is non-decreasing and concave, which is consistent with the theoretical axioms of production theory.

The production frontier estimation can be implemented with DEA, SFA, or StoNED. Since the first DEA model was proposed in 1978 [1], there has been a rapid growth in the number of theoretical and practical DEA studies in various areas such as agriculture, banking, supply chain, transportation, and public policy, among others (see, e.g., [18,19, 20,21]). A major reason for this popularity lies in that the DEA method can measure various frontiers according to the different assumptions of returns to scale. Examples include the constant and variable returns-to-scale DEA frontiers (see, e.g., [1,2]). Note that the frontier estimated in this study is implicitly based on the variable returns-to-scale assumption. However, by dropping off the intercept term (i.e., α in Section 3), one can easily impose the constant returns-to-scale assumption. Compared with the DEA frontier, our approach allows the existence of stochastic noise while the DEA frontier envelopes all of the observed data as tightly as possible, which is purely deterministic.

On the other hand, both SFA and StoNED account for stochastic noise, and these methods allow the data outside the estimated frontier are outside by pure chance. According to the survey [22], the SFA frontier has also been widely used in many areas of applications, and the top five SFA research areas are banking, insurance, container ports, hospital/health care, and agriculture (see also [23] for a systematical review of empirical applications in SFA). A major drawback of SFA is that the SFA frontier is constructed by a pre-specified functional form such as the Cobb-Douglas, translog and generalized McFadden. In contrast to SFA, the StoNED frontier can be estimated nonparametrically (more precisely, semi-nonparametrically, as additional parametric assumptions will be required if one uses the method of moments or quasi-likelihood estimation to derive the frontier estimates). However, compared with the number of applications in DEA or SFA, there are few empirical studies using the StoNED, and most focus on banking, energy, and agriculture (see, e.g., [24,25,26]).

A significant feature of StoNED is that the convex nonparametric least-squares [3] is used to estimate the frontier, which is a problem with quadratic constraints (QCP). Several packages and solvers can be used for estimating the StoNED frontier, such as the "benchmarking" package in R, the "pyStoNED" package in python, QCP solvers in GAMS, and the "CVX" toolbox in Matlab. From a computing perspective, however, the standard estimation procedure of the StoNED frontier suffers from the computational burden even with relatively small sample sizes [27]. Moreover, previous studies on improving the computational performance of StoNED require the assumption of the additive composite error structure [8]. If the composite error structure is multiplicative (e.g., the simulated example using the univariate Cobb-Douglas model in Section 4), the StoNED estimation becomes a nonlinear programming problem, and the computational merits of the algorithm in [8] no longer hold.

Another existing stochastic nonparametric approach is called chance-constrained data envelopment analysis (CCDEA, see, e.g., [28,29,30]). The CCDEA approach shares with SFA and StoNED that some of the data variations may be noise. However, the estimation problem requires strong assumptions on the noise terms, and the computational burden is bigger [31]. In this study, we only compare the performance of our proposed approach with DEA, SFA, and StoNED. Once the frontier is estimated, the existing metrics for the efficiency or productivity in the benchmarking literature, such as distance functions (e.g., [32,33]) or Malmquist-type indices (e.g., [34,35,36]), can be applied for further analysis.

3. Method

We propose a novel nonparametric approach for estimating the production frontier with consideration of inefficiencies and random noise. To introduce the basic idea, we begin with the description of the theoretical model for a single input and single output case.

Considering a sample of n DMUs, each of them produces a single output $y_i \in \mathbb{R}_+$ with a single input $x_i \in \mathbb{R}_+$ for $i = 1, \ldots, n$. The production function is denoted by $f : \mathbb{R}_+ \to \mathbb{R}_+$. We assume that f is a continuous, non-decreasing, and concave function. Let $u_i \in \mathbb{R}_+$ and $v_i \in \mathbb{R}$ be the nonnegative inefficiency term and random noise, respectively. Formally,

$$\begin{aligned} y_i &= f(x_i) - u_i + v_i \\ &= f(x_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \end{aligned}$$

where $\varepsilon_i = v_i - u_i$ is a composite error term. We also assume that u_i and v_i are independent of each other as well as of x_i for all $i = 1, \ldots, n$.

For any observed point (x_i, y_i), the tangent line of $f(x_i)$ represents the upper bound of x_i. Consider the following minimization problem:

$$\min \sum_{i=1}^{n} (\alpha + \beta x_i - y_i)^2 \tag{1}$$

$$\text{s.t.} \quad \alpha + \beta x_i - y_i \geq 0, \quad i = 1, \ldots, n \tag{2}$$

$$\beta \geq 0. \tag{3}$$

Let $e_i := \alpha + \beta x_i - y_i$ for all $i = 1, \ldots, n$ and let $(\hat{e}_i, \hat{\alpha}, \hat{\beta})$ be the optimal solution to the problem (1)–(3). We call $\hat{e}_i$ the estimated downside deviation of the point (x_i, y_i), and $y_i + \hat{e}_i (= \hat{\alpha} + \hat{\beta} x_i)$ is then referred to as the estimated upper bound of (x_i, y_i). It can be proved that the estimated upper bound is independent of the choice of $(\hat{e}_i, \hat{\alpha}, \hat{\beta})$ (For details of the proof, see Theorem 1 in [9]). The difference between the problem (1)–(3) and the ordinary least squares regression is illustrated below: In Figure 1, we plot a sample of 100 points that are produced with a common production technology $y = x^{0.5}$ where the composite error term follows from $u \sim Exp[\mu = 1/6]$ with μ representing the expected inefficiency and $v \sim N(0, 1/6)$. The red line passing through the barycenter of all points is the regression line estimated by the ordinary least squares. By contrast, the blue dashed line results from the problem (1)–(3) and envelopes all points.

Our purpose is to estimate the unknown function f in a nonparametric approach. We consider the problem of estimating the tangent line of $f(x)$ at any given (x_i, y_i) such as

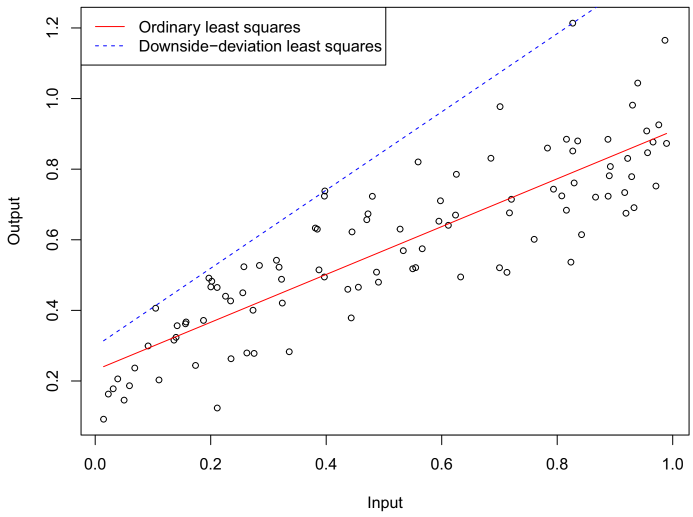

Figure 1. Illustration of the downside-deviation least squares

$$\min \sum_{i \in G_i} (\alpha + \beta x_i - y_i)^2 \tag{4}$$

$$\text{s.t. } \alpha + \beta x_i - y_i \geq 0, \quad i \in G_i \tag{5}$$

$$\beta \geq 0, \tag{6}$$

where G_i is an index set for the range of the observed point (x_i, y_i) and $\bigcup_{i=1}^{n} G_i = \{1, \ldots, n\}$. The set G_i can be determined based on a priori information about the production function. In this paper, we propose a naive method for determining G_i for the purpose of clarification. For some pre-assigned number $1 \leq k \leq n$, define

$$G_{min}(i) := \max(i - k, 1), \quad i \in \{1, \ldots, n\} \tag{7}$$

$$G_{max}(i) := \min(i + k, n), \quad i \in \{1, \ldots, n\}. \tag{8}$$

The set G_i is then represented by $[G_{min}(i), G_{max}(i)]$. If $k = n$, the problem (4)–(6) coincides with the problem (1)–(3).

The problem (4)–(6) implicitly estimates the upper bound for any observed point (x_i, y_i) at a given range G_i. Let $(\tilde{\alpha}, \tilde{\beta})$ be the optimal solution of (x_i, y_i). Solving the problem (4)–(6) for n observed points leads to n estimated coefficients (i.e., $(\tilde{\alpha}_i, \tilde{\beta}_i), i = 1, \ldots, n$). To further impose the concavity assumption, we apply the following Afriat's theorem [10]:

Theorem 3.1 (Afriat's inequalities)**.** *For n observations and m inputs $\boldsymbol{x} = (x_1, \ldots, x_m)^\top$, the following statements hold:*
(1) There exists a continuous concave function $f : \mathbb{R}^m \to \mathbb{R}$ that satisfies $y_i = f(\boldsymbol{x}_i)$ in a

finite number of points $i = 1, \dots, n$.
(2) There exists finite coefficients $\alpha_i, \boldsymbol{\beta}_i = (\beta_{1i}, \dots, \beta_{mi})^\top$ such that $y_i = \alpha_i + \boldsymbol{\beta}_i^\top \boldsymbol{x}_i$ for $i = 1, \dots, n$, that satisfy the following system of inequalities:

$$\alpha_i + \boldsymbol{\beta}_i^\top \boldsymbol{x}_i \le \alpha_h + \boldsymbol{\beta}_h^\top \boldsymbol{x}_i, \quad \textit{for } i, h = 1, \dots, n \textit{ and } i \ne h.$$

In the single input and single output case ($m = 1$), the concavity assumption is

$$\alpha_i + \beta_i x_i \le \alpha_h + \beta_h x_i, \quad \text{for } i, h = 1, \dots, n \text{ and } i \ne h. \tag{9}$$

Let $I(i,i) := \{(\tilde{\alpha}_i, \tilde{\beta}_i) \mid i = 1, \dots, n\}$ be the set of optimal solutions of the problem (4)–(6) for all observed points. Define

$$\bar{\bar{I}}(i,i) := \begin{cases} I(i,i) & \tilde{\alpha}_i + \tilde{\beta}_i x_i \le \tilde{\alpha}_h + \tilde{\beta}_h x_i \\ I(h,h) & \tilde{\alpha}_i + \tilde{\beta}_i x_i > \tilde{\alpha}_h + \tilde{\beta}_h x_i \end{cases} \tag{10}$$

for $i, h = 1, \dots, n$ and $i \ne h$. The set $\bar{\bar{I}}(i,i)$ contains the estimated coefficients that can be used to constructed a concave function f.

Using $(\tilde{\alpha}_i, \tilde{\beta}_i) \in \bar{\bar{I}}(i,i)$, the estimated outputs are computed by $\tilde{y}_i = \tilde{\alpha}_i + \tilde{\beta}_i x_i$, $\forall i$. We then construct the production frontier with the lower bound of the estimated outputs:

$$\tilde{f}(x) := \min\{\alpha + \beta x \mid \alpha + \beta x_i \ge \tilde{y}_i, \ \forall i\}. \tag{11}$$

Figure 2 illustrates how the production frontier can be determined using the proposed estimation procedure. In Figure 2, we randomly generated 100 data points using the function $y = x^{0.5}$ (i.e., the solid black line). By using a pre-assigned number $k = 4$, we solve the problem (4)–(6) for each data point and the size of range G_i belongs to the set $\{5, 6, 7, 8, 9\}$ (See, Eqs. (7)–(8)). By further imposing the concavity assumption, the upper bounds for determining the frontier are finally reduced to five lines (i.e., the dashed lines). Consequently, the frontier is estimated as the lower bound of those five dashed lines (i.e., the red lines). It can be seen that our proposed approach estimates a piece-wise linear production frontier.

Let J be the number of points satisfying $y_i \le \tilde{f}(x_i)$, $\forall i$. Then $\rho := J/n$ represents the ratio of the observed points enveloped by the estimated frontier in the whole observed points. If $k = n$, we have $\rho = 100\%$. Suppose the interest is in estimating a frontier that envelops 80% observed points. In that case, one can start by selecting $k = 1$ and repeat the estimation procedure by changing the value of k until $\rho \ge 80\%$ is achieved.

4. Simulations

In this section, we compare the performance of the proposed approach with the existing methods: DEA, StoNED, and SFA. We consider six scenarios similar to the previous studies [4,37]. To clarify the significant difference between our proposed approach and existing ones, we report the results of the index ρ and the standard mean squared error (MSE) of each method. The notations used in this section are defined below:

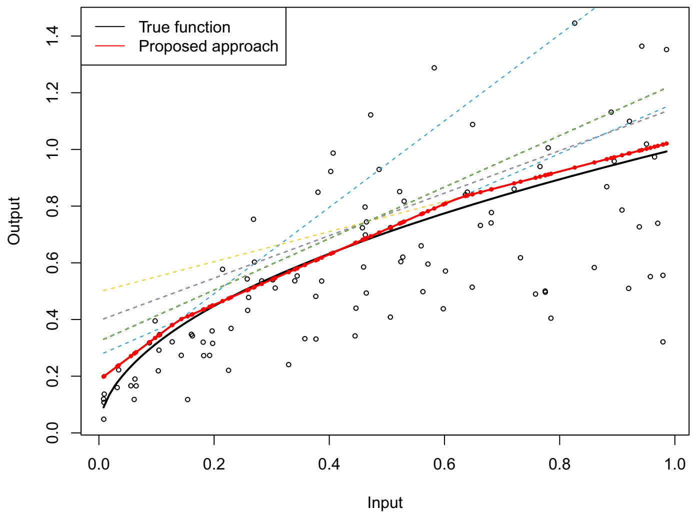

Figure 2. Illustration of the estimated piece-wise linear production frontier

$$\rho_{\hat{f}} = \frac{\text{number of the points under the estimated frontier } \hat{f}}{\text{total number of the points}} \times 100\%,$$

where $\hat{f}$ denotes the true function f or the frontier function estimated by our proposed method, StoNED, SFA, or DEA. The MSE statistic is

$$MSE_{\hat{f}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}(x_i) - f(x_i))^2.$$

For the purpose of clarification, we consider the following univariate Cobb-Douglas model:

$$y_i = x_i^{0.5}\exp(-u_i)\exp(v_i), \tag{12}$$

where $x_i \sim Uni[0,1]$, $u \sim Exp[\mu = 1/6]$ with μ representing the expected value of u_i and $v \sim N(0, \sigma_v^2)$ where $\sigma_v^2 = p \times \mu$ and p represents the noise-to-signal ratio. If $p = 0$, the stochastic noise will be assumed away, which is not the interest of our study. Among the various parametric production functional forms, the Cobb-Douglas model characterizes a concave production function and is commonly used in the SFA literature. Following the previous studies [4,37], we use the univariate Cobb-Douglas frontier as the true frontier and consider six different scenarios for comparing the performance of our proposed approach with DEA, SFA, and StoNED.

Table 1 reports the results of the index $\rho_{\hat{f}}$. Because of the existence of stochastic noise, the generated random sample may appear above the true frontier. ρ_{true} shows the percentage of data under the true frontier. For example, scenario (a) randomly generated 100 points, and 76% of those are under the true frontier. Using this information, we

Table 1. Comparisons of the estimated production frontiers: Results of $\rho_{\hat{f}}$

Scenario	Description	ρ_{true}	$\rho_{proposed}$	ρ_{StoNED}	ρ_{SFA}	ρ_{DEA}
(a)	$n = 100, p = 1$	76	67 $(k = 5)$ 78 $(k = 4)$	55	81	100
(b)	$n = 100, p = 2$	62	54 $(k = 2)$ 69 $(k = 4)$	52	87	100
(c)	$n = 200, p = 1$	77	64 $(k = 7)$ 80 $(k = 9)$	50	85	100
(d)	$n = 200, p = 2$	69	64 $(k = 3)$ 70 $(k = 5)$	54	76	100
(e)	$n = 300, p = 1$	77	60 $(k = 10)$ 80 $(k = 11)$	49	87	100
(f)	$n = 300, p = 2$	67	52 $(k = 4)$ 73 $(k = 10)$	53	72	100

Table 2. Comparisons of the estimated production frontiers: Results of $MSE_{\hat{f}}$

Scenario	Description	$MSE_{proposed}$	MSE_{StoNED}	MSE_{SFA}	MSE_{DEA}
(a)	$n = 100, p = 1$	0.0006 $(k = 5)$ 0.0004 $(k = 4)$	0.0050	0.0009	0.0624
(b)	$n = 100, p = 2$	0.0032 $(k = 2)$ 0.0010 $(k = 4)$	0.0049	0.0297	0.1481
(c)	$n = 200, p = 1$	0.0074 $(k = 7)$ 0.0013 $(k = 9)$	0.0077	0.0017	0.0764
(d)	$n = 200, p = 2$	0.0039 $(k = 3)$ 0.0009 $(k = 5)$	0.0070	0.0038	0.4505
(e)	$n = 300, p = 1$	0.0042 $(k = 10)$ 0.0025 $(k = 11)$	0.0105	0.0026	0.0617
(f)	$n = 300, p = 2$	0.0058 $(k = 4)$ 0.0022 $(k = 10)$	0.0090	0.0016	0.2507

repeat the estimation procedure of our proposed method until $\rho_{proposed}$ gets closer to ρ_{true}. Moreover, we report two types of $\rho_{proposed}$ for each scenario: one is less than ρ_{true} and the other one is greater than ρ_{true}. As shown in Table 1, the frontier estimated by StoNED tends to be located on the lower side of the true frontier as $\rho_{StoNED} < \rho_{true}$ for all scenarios. By contrast, because $\rho_{SFA} > \rho_{true}$, the frontier estimated by SFA tend to be located on the upper side of the true frontier. The frontier estimated by DEA always envelops all data points because DEA ignores the stochastic noise. On the other hand, the frontier estimated by our proposed method can be very close to the true frontier if we choose the proper k. These observations can also be confirmed in Figure 3.

Table 2 further reports the results of $MSE_{\hat{f}}$. We obtain that $MSE_{proposed}$s of scenario (a)–(e) have smaller values than MSE_{SFA}, implying that our proposed method shows better performance than SFA in most scenarios. It is worth noting that we use the correct function to fit the SFA frontier for each scenario. Such an observation suggests that even the naive selection of k is useful in estimating the production frontier. Figure 3 further shows that our proposed method performs better than all other methods.

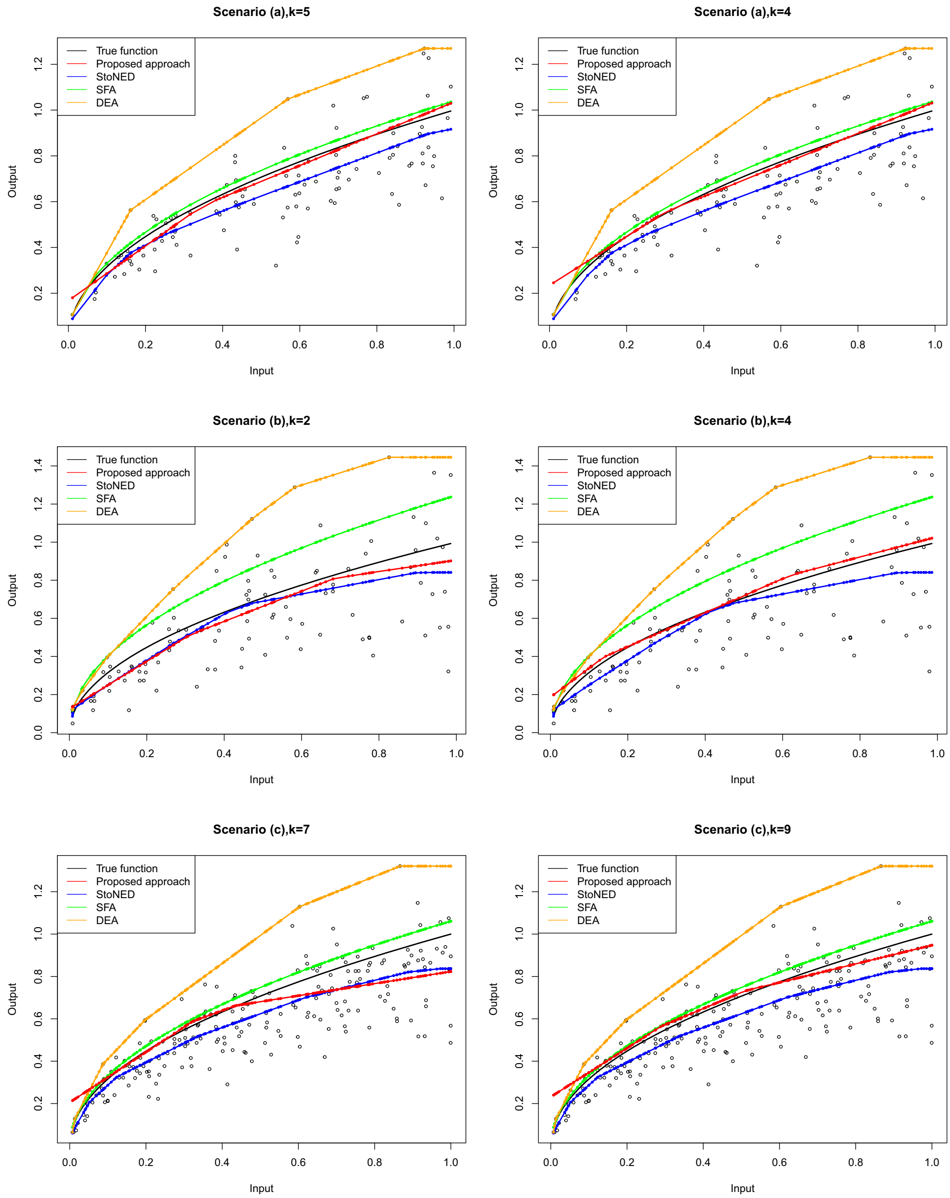

Figure 3. Estimated frontiers

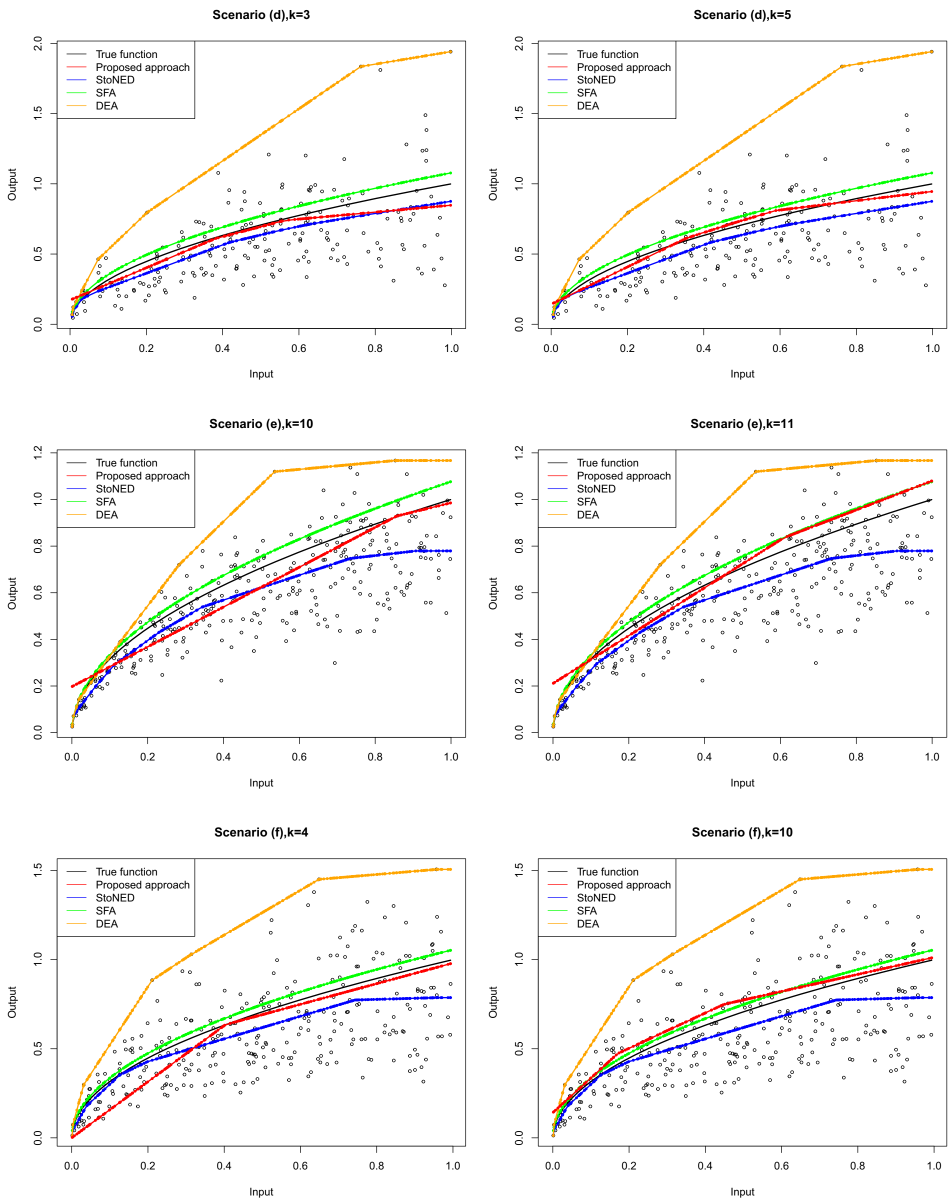

Figure 3. Estimated frontiers (continued)

5. Conclusions and future research

We propose a nonparametric approach for estimating a piece-wise linear production function based on a data-fitting technique. Previous studies usually use the parametric approach (e.g., SFA), the nonparametric approach (e.g., DEA), or the semi-nonparametric approach (e.g., StoNED). Our simulation results show that the proposed estimation procedure performs better than these existing ones. Compared with SFA, our proposed method does not rely on a pre-specific function and determines the production frontier in a completely nonparametric way. Compared with DEA, our proposed method accounts for the impact of stochastic noise. If the policy-makers or regulators have knowledge or information on the level of stochastic noise, the number k can be appropriately determined. Alternatively, researchers can obtain different estimates of production frontiers that envelop different data points by varying the number of k. Compared with StoNED, our proposed method is easy to implement and can be easily applied to large-scale problems. It is worth noting that StoNED requires additional parametric assumptions (e.g., method of moments, quasi-likelihood estimation) or the nonparametric kernel density estimator to determine the final estimate of the frontier. Another difference between StoNED and our proposed method is that the latter does not require further parametric assumptions.

The present paper focuses on a single input and single output case to clarify the estimation procedure. Our approach can be easily extended to a multi-input and single-output setting by replacing the single variable x with a vector $\boldsymbol{x}$ of length m. The applicability of a multi-input and multi-output setting needs to be investigated. On the other hand, methods for the decision of k should be further investigated. We propose a naive method to decide the number k in this paper, and our simulation results prove the usefulness of the proposed naive method. Extensions to empirical data are necessary for examining the adaptability of the proposed approach in future research. Furthermore, the relationship among our proposed method, DEA, and StoNED also needs to be investigated. Because both DEA and StoNED estimate a piece-wise linear frontier, our proposed method may produce the same frontier as DEA or StoNED with some designed DMU-specific range G_is. Finally, we hope this study can make inroads into empirical practice and believe the proposed idea can be helpful to policy-makers and regulators.

References

[1] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. European journal of operational research. 1978;2(6):429-44.

[2] Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management science. 1984;30(9):1078-92.

[3] Kuosmanen T. Representation theorem for convex nonparametric least squares. The Econometrics Journal. 2008;11(2):308-25.

[4] Kuosmanen T, Kortelainen M. Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. Journal of productivity analysis. 2012;38(1):11-28.

[5] Aigner D, Lovell C, Schmidt P. Formulation and estimation of stochastic frontier production function model specifications. Journal of Productivity Analysis. 1977;7:399-415.

[6] Meeusen W, van Den Broeck J. Efficiency estimation from Cobb-Douglas production functions with composed error. International economic review. 1977:435-44.

[7] Kuosmanen T, Johnson A, Saastamoinen A. Stochastic nonparametric approach to efficiency analysis: A unified framework. In: Data envelopment analysis. Springer; 2015. p. 191-244.

[8] Lee CY, Johnson AL, Moreno-Centeno E, Kuosmanen T. A more efficient algorithm for convex nonparametric least squares. European Journal of Operational Research. 2013;227(2):391-400.

[9] Sekitani K, Zhao Y. Performance benchmarking of achievements in the Olympics: An application of Data Envelopment Analysis with restricted multipliers. European Journal of Operational Research. 2021;294(3):1202-12.

[10] Afriat SN. Efficiency estimation of production functions. International economic review. 1972:568-98.

[11] Mohtashami A, Ghiasvand BM. Z-ERM DEA integrated approach for evaluation of banks & financial institutes in stock exchange. Expert Systems with Applications. 2020;147:113218.

[12] Henriques IC, Sobreiro VA, Kimura H, Mariano EB. Two-stage DEA in banks: Terminological controversies and future directions. Expert Systems with Applications. 2020;161:113632.

[13] Chen L, Wang YM. Limitation and optimization of inputs and outputs in the inverse data envelopment analysis under variable returns to scale. Expert Systems with Applications. 2021;183:115344.

[14] Molinos-Senante M, Maziotis A. Benchmarking the efficiency of water and sewerage companies: Application of the stochastic non-parametric envelopment of data (stoned) method. Expert Systems with Applications. 2021;186:115711.

[15] Hanoch G, Rothschild M. Testing the assumptions of production theory: a nonparametric approach. Journal of Political Economy. 1972;80(2):256-75.

[16] Diewert WE, Parkan C. Linear programming tests of regularity conditions for production functions. In: Quantitative studies on production and prices. Springer; 1983. p. 131-58.

[17] Varian HR. The nonparametric approach to production analysis. Econometrica: Journal of the Econometric Society. 1984:579-97.

[18] Liu JS, Lu LY, Lu WM, Lin BJ. A survey of DEA applications. Omega. 2013;41(5):893-902.

[19] Liu JS, Lu LY, Lu WM. Research fronts in data envelopment analysis. Omega. 2016;58:33-45.

[20] Emrouznejad A, Yang Gl. A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. Socio-economic planning sciences. 2018;61:4-8.

[21] Kaffash S, Azizi R, Huang Y, Zhu J. A survey of data envelopment analysis applications in the insurance industry 1993–2018. European journal of operational research. 2020;284(3):801-13.

[22] Lampe HW, Hilgers D. Trajectories of efficiency measurement: A bibliometric analysis of DEA and SFA. European journal of operational research. 2015;240(1):1-21.

[23] Daraio C, Kerstens K, Nepomuceno T, Sickles RC. Empirical surveys of frontier applications: a meta-review. International Transactions in Operational Research. 2020;27(2):709-38.

[24] Kuosmanen T, Kuosmanen N. Role of benchmark technology in sustainable value analysis: an application to Finnish dairy farms. Agricultural and Food Science. 2009;18(3-4):302-16.

[25] Kuosmanen T. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. Energy Economics. 2012;34(6):2189-99.

[26] Eskelinen J, Kuosmanen T. Intertemporal efficiency analysis of sales teams of a bank: Stochastic semi-nonparametric approach. Journal of Banking & Finance. 2013;37(12):5163-75.

[27] Johnson AL, Kuosmanen T. An introduction to CNLS and StoNED methods for efficiency analysis: Economic insights and computational aspects. In: Benchmarking for performance evaluation. Springer; 2015. p. 117-86.

[28] Land KC, Lovell CK, Thore S. Chance-constrained data envelopment analysis. Managerial and decision economics. 1993;14(6):541-54.

[29] Olesen OB, Petersen N. Chance constrained efficiency evaluation. Management science. 1995;41(3):442-57.

[30] Cooper WW, Huang Z, Lelas V, Li SX, Olesen OB. Chance constrained programming formulations for stochastic characterizations of efficiency and dominance in DEA. Journal of productivity analysis. 1998;9(1):53-79.

[31] Olesen OB, Petersen NC. Stochastic data envelopment analysis—A review. European journal of operational research. 2016;251(1):2-21.

[32] Shephard RW. Theory of Cost and Production Functions Princeton University Press. Princeton, New Jersey. 1970.

[33] Färe R, Grosskopf S. Theory and application of directional distance functions. Journal of productivity analysis. 2000;13(2):93-103.

[34] Malmquist S. Index numbers and indifference surfaces. Trabajos de Estadistica y de Investigacion Operativa. 1953;4(2):209-42.

[35] Pastor JT, Lovell CK. A global Malmquist productivity index. Economics Letters. 2005;88(2):266-71.

[36] Zhao Y, Morita H, Maruyama Y. The measurement of productive performance with consideration for allocative efficiency. Omega. 2019;89:21-39.

[37] Simar L, Zelenyuk V. Stochastic FDH/DEA estimators for frontier analysis. Journal of Productivity Analysis. 2011;36(1):1-20.