Journal of the Korean Statistical Society 43 (2014) 503–512

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Frontier estimation using kernel smoothing estimators with data transformation

Hohsuk Noh

Sookmyung Women's University, Republic of Korea

ARTICLE INFO

Article history:
Received 5 March 2014
Accepted 23 July 2014
Available online 13 August 2014

AMS 2000 subject classifications:
primary 62G05
secondary 62P20

Keywords:
Bandwidth selection
Concavity
Constrained kernel smoothing estimator
Priestly–Chao estimator
Monotonicity
Transformed data
Linear programming

ABSTRACT

In economics, a production frontier function is a graph that shows the maximum output of production units such as firms, industries, or economies, as a function of their inputs. Practically, estimating production frontiers often requires imposition of constraints such as monotonicity or monotone concavity. However, few constrained estimators of production frontier have been proposed in the literature. They are based on simple envelopment techniques which often suffer from lack of precision and smoothness. Motivated by this observation, we propose a smooth constrained nonparametric frontier estimator respecting constraints by considering kernel smoothing estimators from a transformed data. It is particularly appealing to practitioners who would like to use smooth estimates that, in addition, satisfy theoretical axioms of production. The utility of this method is illustrated through application to one real dataset and simulation evidences are also presented to show its superiority over the most known methods.

1. Introduction

In economics, a production frontier specifies the maximum possible production level of output for a given input. Hence, it naturally defines production efficiency in the context of that production set: a point on the frontier implies efficient use of the available input, whereas a point below the curve implies inefficiency. Due to importance and necessity of evaluating efficiencies of production units in various areas, various statistical techniques of estimating frontier functions have been proposed so far.

Practically, estimating frontier functions often requires imposition of constraints such as monotonicity or monotone concavity. Popular constrained nonparametric frontier estimators are the free disposal hull (FDH), the linearized FDH (LFDH) and the data envelopment analysis (DEA) estimators. The FDH and DEA estimators are the (upper) boundary curve of the smallest set that envelops all the data with the restriction of free disposability (FDH and DEA) and convexity (DEA) (see, Gijbels, Mammen, Park, & Simar, 1999 and Korostelev, Simar, & Tsybakov, 1995). The linearized FDH (LFDH) estimator is a linearized version of the FDH estimator, which is obtained by linear interpolation of appropriate FDH-efficient points (see, *e.g.*, Hall & Park, 2002, and Jeong & Simar, 2006). Although the FDH and DEA estimators have easy implementation without the need of choosing smoothing parameters and respect desirable constraints of a frontier curve such as monotonicity or monotone concavity, all of these lack smoothness of an estimated curve. To our knowledge, in many situations it is quite natural to assume that frontier curves change smoothly with respect to inputs except at some input points with special meaning. However, estimated curves by FDH and DEA have many unnecessary nonsmooth points. For example, if we see an

E-mail address: word5810@gmail.com.

http://dx.doi.org/10.1016/j.jkss.2014.07.005
1226-3192/

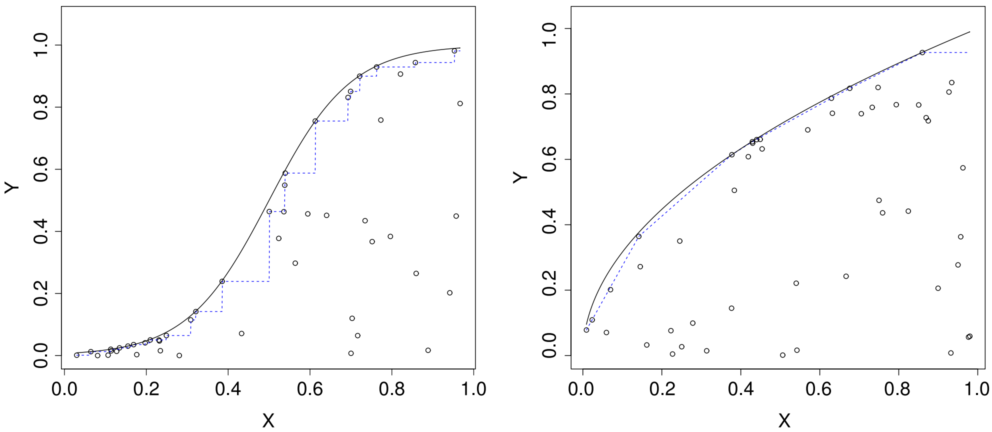

Fig. 1. Examples of FDH and DEA estimates—the true frontier curve (solid black), FDH estimate (dotted blue in the left panel) and DEA estimate (dotted blue in the right panel).

FDH estimate in Fig. 1, the estimated curve repeatedly stays flat and jumps and the location of jumps is just X-locations of FDH points with no appropriate interpretation. A similar problem occurs with the DEA estimator in Fig. 1. At every X-location of DEA points, we have sudden changes of the slope of the estimated curve, for which it is very hard to provide justification. Considering that the unknown frontier function is often assumed to be continuously differentiable and seems to be so in practice, we are motivated to develop a smooth constrained nonparametric frontier estimator less suffering from non-smoothness and underestimation problems.

The main idea of our proposal stems from the works of Hall and Huang (2001), who reweighted the data to impose monotonicity on kernel regression methods and Du, Parmeter, and Racine (2013), who extended the idea to the multivariate and multi-constraint setting. Their common idea is to consider a transformed data by introducing weights for each response and compute the ordinary kernel regression estimator from the transformed data as the final estimator. The weights are decided to make the resulting estimator from the transformed data obey the given constraints (on an unspecified conditional mean), and fit well the given data. Inspired by this insight, we propose a smooth frontier function estimator by developing an appropriate weight-choosing scheme for frontier estimation. Specifically, we choose the weights to make the resulting estimator obey typical constraints in frontier estimation such as monotonicity or monotone concavity, envelop all the data and minimize the area under the frontier curve. We follow Hall, Park, and Stern's (1998) idea that finding a curve beyond the data points under which the area is minimized gives a good estimator in frontier analysis, and some theoretical justification of the idea can be found in Hall, Park, and Stern (1998). As typical in kernel smoothing methods, performance of the estimator depends on the choice of the bandwidth. We propose that an appropriate bandwidth can be determined by analogy to the popular Bayesian information criterion (BIC). This criterion is shown to work well in our simulation studies in Section 3. Actually, the idea of adapting the mean regression method of Du et al. (2013) and Hall and Huang (2001) to frontier estimation was previously considered in Parmeter and Racine (2013). However, our method is distinctively different from theirs in the minimization criterion and the strategy of bandwidth selection. The details about the difference are discussed in Section 2.3.

The rest of the paper is organized as follows. Section 2 describes in detail the proposed smooth frontier estimator, including computation via linear programming as well as bandwidth selection method. Section 3 provides comparison with the most known frontier estimation methods in productivity analysis such as the FDH, LFDH and DEA estimators through Monte Carlo experiments. Section 4 illustrates the utility of our method for the performance analysis of electric utility companies. Finally, the conclusion is given in Section 5.

2. Methodology

Our interest is to estimate from the data the boundary curve of the support of a bivariate distribution with density $f(x, y)$ in $\mathbb{R}^2$. Before introducing our estimator, we will describe some details of the setting for it. We assume that the support Ψ of f is of the form

$$\Psi = \{(x, y)|y \le \varphi(x)\} \supseteq \{(x, y)|f(x, y) > 0\}, \{(x, y)|y > \varphi(x)\} \subseteq \{(x, y)|f(x, y) = 0\}, \tag{1}$$

where φ is an increasing or concavely increasing function whose graph corresponds to the locus of the curve above which the density f is zero. Further, we assume that the data $(x_i, y_i),\ i = 1, \ldots, n$ from the density $f(x, y)$ is obtained from the following data generating process:

$$y_i = \varphi(x_i) - u_i,$$

where $u_i \geq 0$ is an inefficiency factor following a certain distribution that has the non-negative support. Note that we are working under the framework of deterministic frontier models, which assume that the actual observations deviate from the frontier function only by inefficiency of each production unit, not by noise. Hence, once we have an estimate $\hat{\varphi}(\cdot)$ of $\varphi(\cdot)$, we will evaluate technical efficiency (TE) of each production (x_i, y_i) by the Shephard output efficiency (see Shephard, 1970):

$$\text{TE of production } (x_i, y_i) = \frac{\hat{\varphi}(x_i) - \hat{u}_i}{\hat{\varphi}(x_i)} = \frac{y_i}{\hat{\varphi}(x_i)},$$

where $\hat{u}_i = \hat{\varphi}(x_i) - y_i$ is the estimated inefficiency.

2.1. Definition of the estimator

Suppose that we have n sample pairs of input and output $(x_i, y_i),\ i = 1, \ldots, n$, from a bivariate distribution with a density $f(x, y)$ in $\mathbb{R}^2$. Without much loss of generality, we restrict ourselves to the case where $x_i \in [0, 1]$ and $y_i \geq 0$. We estimate the frontier function φ based on the sample $\{(x_i, y_i),\ i = 1, \ldots, n\}$ by using the class of kernel regression smoothers that can be written as linear combinations of the output y_i, in other words,

$$\tilde{\varphi}(x) = \sum_{i=1}^{n} A_i(x) y_i, \tag{2}$$

where the weight functions A_i depend only on the x_i's, not on the y_i's. Examples of the kernel smoothers of the form (2) include the Nadaraya–Watson estimator (Nadaraya, 1965, and Watson, 1964), the Priestley–Chao estimator (Priestley & Chao, 1972), the Gasser–Müller estimator (Gasser & Müller, 1979), and the local linear estimator (Fan, 1992), among others. For the Nadaraya–Watson estimator, $A_i = K_i/(\sum_{j=1}^{n} K_j)$ where $K_i(x) = h^{-1}K\{(x - x_i)/h\}$, the kernel function K is a bounded and symmetric probability density and h is a bandwidth; for the Priestley–Chao estimator,

$$A_i(x) = \begin{cases} 0, & i = 1 \\ (x_i - x_{i-1})K_i(x), & i \neq 1, \end{cases}$$

where here and in the next example it is assumed that the pairs (x_i, y_i) have been ordered so that $x_1 \leq \cdots \leq x_n$; for the Gasser–Müller estimator, $A_i(x) = \int_{\mathcal{S}(i)} K_i(x - u)du$, where $\mathcal{S}(i)$ denotes the interval $[z_{i-1}, z_i]$ and $z_i = (x_i + x_{i+1})/2$; for the local linear estimator,

$$A_i(x) = \frac{S_2(x) - S_1(x)(x_i - x)}{S_2(x)S_0(x) - S_1(x)^2} K_i(x),$$

where $S_j(x) = \sum_{i=1}^{n} (x_i - x)^j K_i(x)$.

Motivated by the works of Du et al. (2013) and Hall and Huang (2001) and to impose necessary constraints on the estimator, we consider a generalization of (2)

$$\hat{\varphi}(x|p) = \sum_{i=1}^{n} p_i A_i(x) y_i, \tag{3}$$

and its sth derivative $\hat{\varphi}^{(s)}(x|p) = \sum_{i=1}^{n} p_i A_i^{(s)}(x) y_i$. Note that the generalized estimator (3) can be seen as the ordinary kernel smoother from the transformed data $(x_i, y_i^*) = (x_i, p_i y_i),\ i = 1, \ldots, n$. Our main idea is to choose the weight p_i appropriately so that the estimator (3) has all desirable properties of frontier function such as enveloping the data and respecting monotonicity and concavity constraints. Basically, following the idea of Hall et al. (1998), we choose $p = \hat{p}$ to minimize the area under the frontier curve $\hat{\varphi}(x|p)$, $A(p) = \int_0^1 \hat{\varphi}(x|p)dx = \sum_{i=1}^{n} p_i y_i \left(\int_0^1 A_i(x)dx\right)$. The constrained estimator is then obtained by choosing the weights p which minimize $A(p)$ subject to the following constraints:

$$\hat{\varphi}(x_i|p) - y_i = \sum_{i=1}^{n} p_i A_i(x_i) y_i - y_i \geq 0, \quad i = 1, \ldots, n; \text{ (envelopment constraints)} \tag{4}$$

$$\hat{\varphi}^{(1)}(x|p) = \sum_{i=1}^{n} p_i A_i^{(1)}(x) y_i \geq 0, \quad x \in \mathcal{M}; \text{ (monotonicity constraints)} \tag{5}$$

$$\hat{\varphi}^{(2)}(x|p) = \sum_{i=1}^{n} p_i A_i^{(2)}(x) y_i \leq 0, \quad x \in \mathcal{C}, \text{ (concavity constraints)} \tag{6}$$

where $\mathcal{M}$ is the collection of points where monotonicity is imposed, and $\mathcal{C}$ is the collection of points where concavity is imposed. When the sample size is large, we simply take all the sample points $\{(x_i, y_i),\ i = 1, \ldots, n\}$ for both $\mathcal{M}$ and $\mathcal{C}$. However, in case of small sample, it is necessary to add some points to $\mathcal{M}$ and $\mathcal{C}$ to ensure monotonicity and concavity, respectively. For such purpose, in our simulations we augment $\mathcal{M}$ and $\mathcal{C}$ by a equispaced grid of length 101 over the observed support of X ($[\min_i x_i, \max_i x_i]$). Since all the constraints are linear in p as well as the optimization criterion $A(p)$, estimating p can be cast as linear programming, which can be solved very fast using standard off-the-shelf software (a few hundredths of a second with R). For our simulations, we use the R package *Rglpk* to solve linear programming, and all the R codes for the simulations are available upon request.

Remark. We find that when the bandwidth h is small, the computation of $\hat{p}$ becomes sometimes numerically unstable and the l_1-norm of the $\hat{p}$ vector tends to be very large. For numerical stability, we consider one additional constraint $\sum_{i=1}^n |p_i| \le C$ for some large positive C. In our simulations, we use this version of the estimator with C being 10,000. We observe that the choice of C does not result in significant difference of the estimator as long as it is large enough. Finally, the optimization problem (3) with (4)–(6) and the additional constraint is again a linear programming problem. This can be easily shown using the decomposition $p = p^+ - p^-,\ p^+ = (p_1^+, \ldots, p_n^+)^\top,\ p^- = (p_1^-, \ldots, p_n^-)^\top\ (p_i^+, p_i^- \ge 0,\ i = 1, \ldots, n)$.

2.2. Bandwidth selection

It is known that smaller bandwidth is appropriate for modeling complex curves, whereas larger bandwidth is suitable for modeling simple curves. Therefore, bandwidth selection can be viewed as determining how complex the estimated curve is supposed to be according to the data. Based on this insight, our main idea for bandwidth selection is to use Bayesian Information Criterion (BIC), which is widely used to determine appropriate model complexity.

To get an idea of BIC for frontier estimation, we first consider a problem of choosing appropriate degree of a polynomial frontier model and then move on to our case. The polynomial frontier model of interest is

$$y_i = \varphi(x_i|\theta) - u_i, \quad i = 1, \ldots, n,$$

where the inefficiency u_i follows an exponential distribution with mean λ, $\varphi(x|\theta) = \theta_0 + \theta_1 x + \cdots + \theta_k x^k$ ($\theta = (\theta_0, \theta_1, \ldots, \theta_k)^\top$) and x_i follows a uniform distribution on [0,1]. In this case, the likelihood function for θ and λ is given as

$$L(\theta, \lambda) = \prod_{i=1}^n \frac{1}{\lambda} \exp\left(-\frac{1}{\lambda}(\varphi(x_i|\theta) - y_i)\right) I(\varphi(x_i|\theta) - y_i \ge 0),$$

where $I(\cdot)$ is an indicator function. From this likelihood, we derive the maximum likelihood estimators $\hat{\theta}$ and $\hat{\lambda}$ of θ and λ:

$$\hat{\theta} = \underset{\theta:\varphi(x_i|\theta)-y_i\ge 0 \text{ for all } i}{\operatorname{argmin}} \sum_{i=1}^n \varphi(x_i|\theta);$$

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n (\varphi(x_i|\hat{\theta}) - y_i).$$

Finally, from the usual definition of BIC in a form of penalized likelihood, we have a BIC for frontier model ignoring constants:

$$-2\log\ \text{likelihood}(\hat{\theta}, \hat{\lambda}) + \log n \cdot (k+1) \propto \log\left(\sum_{i=1}^n \left(\varphi(x_i|\hat{\theta}) - y_i\right)\right) + \frac{\log n}{2n}(k+1) \equiv \text{BIC}(k). \tag{7}$$

Note that the factor $(k+1)$, which represents the number of coefficients, enters into the definition of BIC as model complexity. We choose k which minimizes the right-hand side of (7) as appropriate degree.

Now, we will extend the BIC (7) to the one for our estimator. For such purpose, we only need to appropriately define the model complexity of our frontier estimator. First, note that once the weights p are chosen for a given bandwidth h, the generalized estimator (3) can be seen as the ordinary kernel smoother from a transformed data $(x_i, y_i^*(h)) = (x_i, \hat{p}_i(h) y_i),\ i = 1, \ldots, n$. Then, according to basic theory about kernel smoothers (see Buja, Hastie, & Tibshirani, 1989), the model complexity of the kernel smoother $\hat{\varphi}(x|\hat{p}(h)) = \sum_{i=1}^n A_i(x) y_i^*(h)$ with respect to the transformed data can be defined as the trace of its smoothing matrix, $\text{tr}(S(h))$, where

$$S(h) = \begin{pmatrix} A_1(x_1) & \cdots & A_n(x_1) \\ \vdots & \ddots & \vdots \\ A_1(x_n) & \cdots & A_n(x_n) \end{pmatrix}.$$

Therefore, we define a BIC for our estimator by replacing $\varphi(x_i|\hat{\theta})$ and $k+1$ in (7) with $\hat{\varphi}(x_i|\hat{p}(h))$ and $\text{tr}(S)$, respectively:

$$\text{BIC}(h) = \log\left(\sum_{i=1}^n (\hat{\varphi}(x_i|\hat{p}(h)) - y_i)\right) + \frac{\log n \cdot \text{tr}(S(h))}{2n}, \tag{8}$$

the minimizer of which is the optimal bandwidth we will choose.

2.3. Comparison with the previous method

Previously, Parmeter and Racine (2013) considered an estimator similar to ours based on the generalization (3). The idea to impose constraints such as (4)–(6) on the estimator is the same for both estimators. However, our estimator is distinctively different from theirs in the minimization criterion and the strategy of bandwidth selection.

To choose the vector of weights p, Parmeter and Racine (2013) chose $p = \hat{p}$ to minimize the distance from p to the uniform weights $p_i = 1$ for all i using the distance metric $D(p) = (p - p_u)^\top (p - p_u)$ subject to the constraints (4)–(6), where p_u is an n dimensional vector with all the elements being one. Since the estimator (3) with the weights p_u corresponds to the mean regression estimator, minimizing $D(p)$ leads to capturing the shape of the "middle" of the cloud of data points. This is not natural for capturing the shape of the frontier function, which may be different from the shape of the conditional mean function. Due to this, we replace $D(p)$ with $A(p)$, which is the area under the frontier curve and is more suitable to capture the shape of the boundary in frontier estimation.

Additionally, Parmeter and Racine (2013) chose the optimal bandwidth for the mean regression estimator (2) with the data $\{(x_i, y_i),\ i = 1, \ldots, n\}$ and reused it for the frontier estimator. However, generally the shape of mean regression curve $m(x) = E(Y|X = x)$ is different from that of the frontier curve. Hence, reusing the optimal bandwidth of mean regression curve estimation for frontier curve estimation is not recommended. For our estimator, we develop a new bandwidth selection procedure (8) from model selection point of view and use it for numerical simulations and real data analysis. The results from there show that our bandwidth selection method is sensible in frontier estimation.

In common with Parmeter and Racine's (2013) method, our method is not robust to outliers as it should cover all the data points. For stable performance of our method, the appropriate outlier detection method should be applied before the frontier estimation.

3. Numerical illustrations

In this section, we do some numerical simulations in order to demonstrate superiority of the proposed estimator $\hat{\varphi}$ over the most known constrained frontier estimators rooted in data envelopment ideas. Those were the FDH, linearized FDH (LFDH) and DEA estimators described in Section 1, and the estimator proposed by Parmeter and Racine (2013) (PR estimator hereafter). For our estimator and the PR estimator, the sort of the unconstrained nonparametric estimator $\tilde{\varphi}$ should be specified in advance. Concerning our estimator, we use the Priestley–Chao estimator with the Gaussian kernel for the convenience of the evaluation of the integral $\int A_i(x)dx$. Differently, for the PR estimator we use the Nadaraya–Watson estimator with the same kernel. We may use the Priestley–Chao (or Gasser–Müller) estimator considering some computational advantage of evaluating derivatives of $A_i(x)$ but we insist on our choice because of its better performance. The reason can be attributed to the fact that the PR estimator has its root in mean regression estimators, among which the Priestley–Chao estimator is not a good one, especially in terms of asymptotic variance (see Jones, Davies, & Park, 1994). For bandwidth selection, we use the leave-one-out cross-validation for the PR estimator. For our estimator, we employ two methods (the leave-one-out cross-validation and the proposed BIC-type criterion) to illustrate effectiveness of our bandwidth selection strategy BIC(h) in (8).

To evaluate finite-sample performance of the proposed estimator in comparison with the frontier estimates described above, we do some simulation experiments similar to the ones in Girard and Jacob (2008). We consider a univariate random variable X uniformly distributed over the compact support $E = [0, 3]$. Besides, Y given $X = x$ is distributed on $[0, \varphi(x)]$ such that

$$P(Y > y|X = x) = \left(1 - \frac{y}{\varphi(x)}\right)^\gamma, \tag{9}$$

with $\gamma > 0$. The case $\gamma = 1$ corresponds to the situation where Y given $X = x$ is uniformly distributed on $[0, \varphi(x)]$. The cases $0 < \gamma < 1$ and $\gamma > 1$ represent the situations where the joint density of the (x_i, y_i)'s is tending to infinity and converging to zero as it approaches the frontier points, respectively. Note that the larger the γ is, the smaller the probability of (9) and the harder the estimation of the frontier function becomes. The behavior of the proposed estimator is investigated in different situations:

- Two frontiers are considered. The first one is

$$\varphi_1(x) = \begin{cases} -(1-x)^2 + 1, & 0 \le x < 1; \\ 1, & 1 \le x < 2; \\ (x-2)^2 + 1, & 2 \le x \le 3 \end{cases} \tag{10}$$

  is monotone increasing but not concave. The second one is $\varphi_2(x) = \sqrt{x}$, which is monotone increasing and concave.
- Four samples sizes are simulated $n \in \{25, 50, 100, 200\}$.
- Three exponents are used $\gamma \in \{0.5, 1, 2\}$.

When the true frontier function is monotone increasing, we only impose monotonicity constraints on our estimator and the PR estimator. If the frontier function is concave additionally, we impose both monotonicity and concavity constraints on them.

Table 1
Comparison when the true frontier is monotone increasing ($\varphi(x) = \varphi_1(x)$). All the results are multiplied by 100.

			FDH	LFDH	SF-BIC	SF-CV	PR
$\gamma = 0.5$	$n = 25$	IBIAS2	3.809	1.002	0.351	0.680	2.235
		IVAR	1.369	0.983	0.708	14.434	2.360
		IMSE	5.178	1.985	1.060	15.114	4.595
	$n = 50$	IBIAS2	1.783	0.459	0.108	0.227	0.752
		IVAR	0.719	0.445	0.221	4.823	0.835
		IMSE	2.502	0.904	0.330	5.051	1.587
	$n = 100$	IBIAS2	0.772	0.194	0.031	0.127	0.189
		IVAR	0.316	0.183	0.081	0.472	0.230
		IMSE	1.088	0.377	0.113	0.599	0.419
	$n = 200$	IBIAS2	0.321	0.075	0.007	0.036	0.087
		IVAR	0.136	0.071	0.023	0.070	0.062
		IMSE	0.456	0.145	0.030	0.106	0.150
$\gamma = 1$	$n = 25$	IBIAS2	8.422	3.835	2.455	3.410	2.657
		IVAR	1.785	1.696	1.478	11.476	3.328
		IMSE	10.207	5.531	3.933	14.886	5.984
	$n = 50$	IBIAS2	4.744	2.116	1.219	1.767	1.406
		IVAR	1.103	1.015	0.763	3.275	1.435
		IMSE	5.847	3.130	1.983	5.042	2.841
	$n = 100$	IBIAS2	2.609	1.177	0.612	0.910	0.641
		IVAR	0.622	0.536	0.436	2.339	0.613
		IMSE	3.231	1.713	1.047	3.248	1.254
	$n = 200$	IBIAS2	1.370	0.595	0.268	0.460	0.270
		IVAR	0.337	0.258	0.202	0.260	0.210
		IMSE	1.707	0.853	0.471	0.720	0.480
$\gamma = 2$	$n = 25$	IBIAS2	19.818	13.197	10.999	13.343	8.342
		IVAR	2.060	2.100	2.149	8.062	3.689
		IMSE	21.878	15.297	13.149	21.405	12.032
	$n = 50$	IBIAS2	13.469	8.621	7.436	8.503	6.032
		IVAR	1.475	1.625	1.445	4.503	2.113
		IMSE	14.944	10.246	8.881	13.006	8.145
	$n = 100$	IBIAS2	9.087	5.858	5.019	5.763	4.217
		IVAR	1.027	1.115	1.036	1.160	1.201
		IMSE	10.114	6.973	6.055	6.923	5.418
	$n = 200$	IBIAS2	6.001	3.749	3.267	3.695	2.680
		IVAR	0.698	0.704	0.656	0.680	0.646
		IMSE	6.699	4.453	3.923	4.375	3.326

In Tables 1 and 2 we present performance of the six estimation methods: FDH, LFDH, DEA, our smooth frontier (SF) estimator with two bandwidth selection methods (SF-BIC and SF-CV) and the PR estimator. As a measure of performance of each method, we consider the empirical mean integrated squared error (MISE), the empirical integrated squared bias (IBIAS2) and the empirical integrated variance (IVAR), which are given by

$$\begin{aligned} \text{MISE} &= \frac{1}{N}\sum_{j=1}^{N} \text{ISE}(\hat{\varphi}^{(j)}) := \frac{1}{N}\sum_{j=1}^{N}\left[\frac{1}{I}\sum_{i=0}^{I}\left(\hat{\varphi}^{(j)}(z_i) - \varphi(z_i)\right)^2\right] \\ &= \frac{1}{I}\sum_{i=0}^{I}\left(\varphi(z_i) - \bar{\hat{\varphi}}(z_i)\right)^2 + \frac{1}{I}\sum_{i=0}^{I}\left[\frac{1}{N}\sum_{j=1}^{N}(\hat{\varphi}^{(j)}(z_i) - \bar{\hat{\varphi}}(z_i))^2\right] \\ &\equiv \text{IBIAS2} + \text{IVAR}, \end{aligned} \tag{11}$$

where $\{z_i, i = 0, \ldots, I\}$ is an equispaced grid with width $1/I$ over $[0, 3]$ with $I = 1000$, $\hat{\varphi}^{(j)}(\cdot)$ is the estimated frontier function from the jth data sample and $\bar{\hat{\varphi}}(z_i) = N^{-1}\sum_{j=1}^{N}\hat{\varphi}^{(j)}(z_i)$.

Overall, our smooth frontier estimator with the bandwidth selected by (8) shows better performance than all the other estimators regardless of the boundary type (which depends on γ) and the sample size. Only when $\gamma = 2$ and $\varphi(x) = \varphi_1(x)$ (this corresponds to the case where boundary points are rare), the PR estimator performs slightly better than ours. When the bandwidth is chosen by the cross-validation, our estimator does not show as impressive performance as with the bandwidth from (8). This shows that bandwidth selection methods for regression setting cannot be directly applied to frontier setting, for which the appropriate selection method needs to be proposed. The superior performance of SF-BIC estimator suggests that our BIC-type criterion can work for such purpose as illustrated when $\gamma = 1$ in Fig. 2.

Compared with the PR estimator, our SF-BIC estimator shows better performance in most cases. Especially, when the number of data points near the boundary is relatively large ($\gamma = 0.5$), such performance difference becomes considerably bigger. This implies that the optimization criterion in Parmeter and Racine (2013), which does not utilize the information

Table 2
Comparison when the true frontier is monotone increasing and concave ($\varphi(x) = \varphi_2(x)$). All the results are multiplied by 100.

			DEA	SF-BIC	SF-CV	PR
$\gamma = 0.5$	$n = 25$	IBIAS2	0.326	0.133	0.809	4.070
		IVAR	0.262	0.238	5.738	1.098
		IMSE	0.589	0.372	6.547	5.169
	$n = 50$	IBIAS2	0.096	0.048	0.296	1.990
		IVAR	0.075	0.069	3.880	0.477
		IMSE	0.171	0.117	4.176	2.467
	$n = 100$	IBIAS2	0.029	0.029	0.090	1.077
		IVAR	0.024	0.023	2.027	0.179
		IMSE	0.053	0.052	2.117	1.257
	$n = 200$	IBIAS2	0.007	0.021	0.005	0.724
		IVAR	0.006	0.006	0.227	0.089
		IMSE	0.014	0.027	0.233	0.813
$\gamma = 1$	$n = 25$	IBIAS2	2.353	1.507	3.359	4.124
		IVAR	0.821	0.907	6.369	1.826
		IMSE	3.174	2.415	9.728	5.950
	$n = 50$	IBIAS2	0.978	0.588	1.266	2.230
		IVAR	0.353	0.395	3.226	0.867
		IMSE	1.330	0.984	4.493	3.097
	$n = 100$	IBIAS2	0.411	0.239	0.534	1.234
		IVAR	0.153	0.167	2.279	0.370
		IMSE	0.565	0.406	2.813	1.604
	$n = 200$	IBIAS2	0.163	0.096	0.111	0.801
		IVAR	0.061	0.066	0.103	0.161
		IMSE	0.224	0.163	0.214	0.963
$\gamma = 2$	$n = 25$	IBIAS2	12.460	10.204	14.349	8.854
		IVAR	1.702	1.996	6.614	2.973
		IMSE	14.162	12.200	20.963	11.827
	$n = 50$	IBIAS2	7.406	5.969	7.626	5.390
		IVAR	1.014	1.193	3.291	1.736
		IMSE	8.420	7.162	10.917	7.126
	$n = 100$	IBIAS2	4.379	3.485	4.260	3.341
		IVAR	0.632	0.727	2.162	0.968
		IMSE	5.011	4.212	6.421	4.309
	$n = 200$	IBIAS2	2.553	2.000	2.092	2.126
		IVAR	0.361	0.412	0.423	0.526
		IMSE	2.914	2.413	2.514	2.652

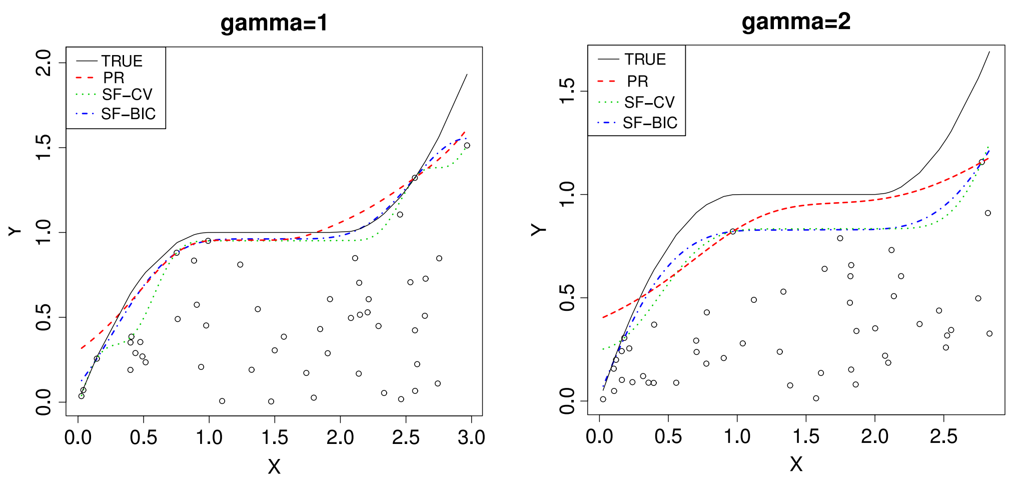

Fig. 2. When $n = 50$ and $\gamma = 1$ and 2, the true frontier function (φ_1) and its three estimates based on the given data: PR, SF-CV and SF-BIC.

about the boundary from the data, does not seem so appropriate in the frontier setting. As an example of such inappropriateness, see Figs. 3 and 4. However, in situations where the density of data is relatively low near the boundary ($\gamma = 1$ or 2), the PR estimator seems to work as well as our SF-BIC estimator. The reason may be that our method heavily relies on

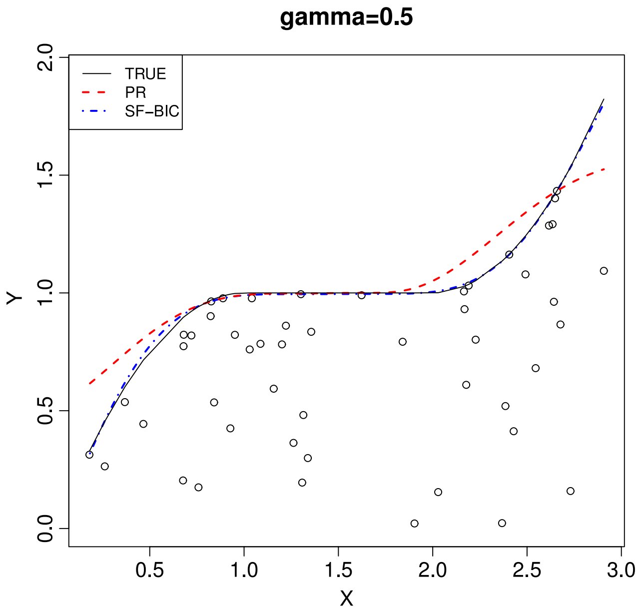

Fig. 3. When $n = 50$ and $\gamma = 0.5$, the true frontier function (φ_1) and its two estimates based on the given data: PR and SF-BIC.

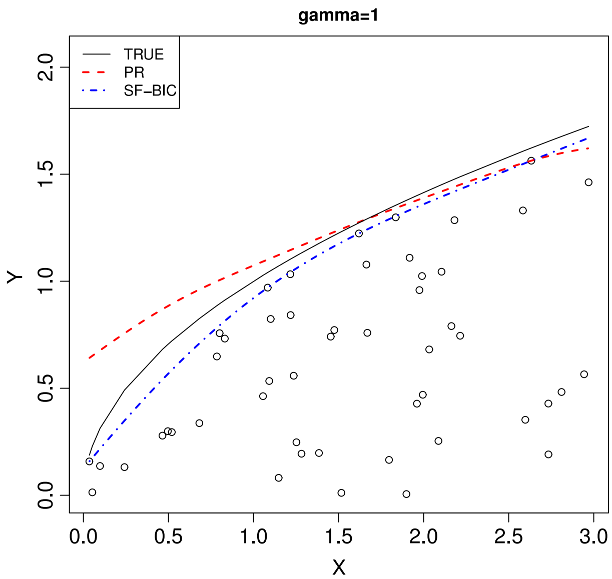

Fig. 4. When $n = 50$ and $\gamma = 1$, the true frontier function (φ_2) and its two estimates based on the given data: PR and SF-BIC.

the data which tends to mislead detection of the boundary in such case but PR estimator takes into account the shape of mean regression curve in estimation, which may have some similarity with the frontier curve depending on the case.

4. Real data example

Our application of the proposed estimator is concerned with the increase of production activity of 123 American electric utility companies. The measurements of the produced output and the total cost of each company are represented in Fig. 5. Since it is known that the econometric frontier – that is, the locus of the most efficient firms – is nondecreasing and concave (see, *e.g.*, Gijbels et al., 1999), we plot our estimator (SF-BIC) and the PR estimator with monotonicity and concavity constraints along with DEA estimator. Fig. 5 shows that both DEA and our method seem to give reasonable estimates of the boundary of the production set. Considering that sudden changes of the slope of the DEA estimate do not have appropriate justification, it might be more sensible to refer to our estimate than the DEA estimate for further efficiency analysis. Finally,

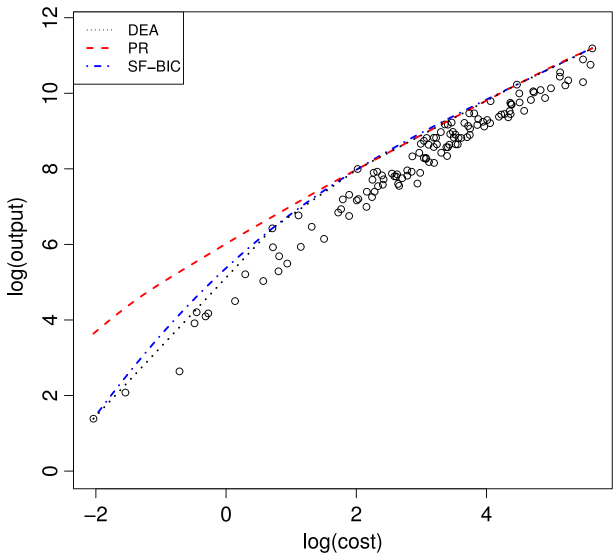

Fig. 5. Scatterplots of the 123 American electric utility companies' data with three estimates of the frontier function: SF-BIC, PR and DEA.

as in the simulations, the PR estimator does not seem to work well due to the inappropriateness of the optimization criterion, especially in the region where the cost is low.

5. Conclusion

We developed a novel smooth frontier estimator based on kernel smoothers for regression using the idea of constrained estimation considered in Du et al. (2013) and Hall and Huang (2001). All well-known kernel smoothers can be used for our method but we chose to use the Priestley–Chao estimator because of computational convenience. From a computational point of view, our method can be efficiently implemented via linear programming and hence the computation time is very fast. Further, our estimator appears to outperform the most known frontier methods such FDH, LFDH, and DEA estimators, when it is combined with the BIC-type bandwidth selection procedure proposed in this work. Our BIC criterion is motivated by assuming that the inefficiency factor follows a exponential distribution. However, one might consider other distributions such as half-normal or gamma distributions and use the resulting BICs for the same purpose. Although there are some similarities between our estimator and the one proposed in Parmeter and Racine (2013), the difference in the optimization criterion and the bandwidth selection method is shown to distinguish our estimator from theirs as the superior one in performance according to our simulations.

In our simulations, we observe that as the sample size grows, the accuracy of estimation of our estimator increases in all settings that we considered. However, consistency issue of our method should be investigated theoretically. In addition, it is worthwhile to investigate whether multivariate extension of our method is possible. We leave these as future works.

Acknowledgments

The author would like to thank two referees for their valuable suggestions, which have significantly improved the paper. This research was supported by the Sookmyung Women's University Research Grant 1-1309-0007.

References

Buja, A., Hastie, T., & Tibshirani, R. (1989). Linear smoothers and additive models. *Annals of Statistics*, *17*, 453–510.
Du, P., Parmeter, C. F., & Racine, J. S. (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statistica Sinica*, *23*, 1347–1371.
Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, *87*, 998–1004.
Gasser, T., & Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation* (pp. 23–68). New York: Springer-Verlag.
Gijbels, I., Mammen, E., Park, B. U., & Simar, L. (1999). On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association*, *94*, 220–228.
Girard, S., & Jacob, P. (2008). Frontier estimation via kernel regression on high power-transformed data. *Journal of Multivariate Analysis*, *99*, 403–420.
Hall, P., & Huang, L. S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics*, *29*, 624–647.
Hall, P., & Park, B. U. (2002). New methods for bias correction at endpoints and boundaries. *Annals of Statistics*, *30*, 1460–1479.
Hall, P., Park, B. U., & Stern, S. E. (1998). On polynomial estimators of frontiers and boundaries. *Journal of Multivariate Analysis*, *66*, 71–98.
Jeong, S.-O., & Simar, L. (2006). Linearly interpolated FDH efficiency score for nonconvex frontiers. *Journal of Multivariate Analysis*, *97*, 2141–2161.
Jones, M. C., Davies, S. J., & Park, B. U. (1994). Versions of kernel-type regression estimators. *Journal of the American Statistical Association*, *89*, 825–832.

Korostelev, A., Simar, L., & Tsybakov, A. B. (1995). Efficient estimation of monotone boundaries. *Annals of Statistics*, *23*, 476–489.
Nadaraya, E. A. (1965). On nonparametric estimates of density functions and regression curves. *Theory of Probability and its Applications*, *10*, 186–190.
Parmeter, C. F., & Racine, J. S. (2013). Smooth constrained frontier analysis. In *Recent advances and future directions in causality, prediction, and specification analysis* (pp. 463–488). New York: Springer-Verlag.
Priestley, M. B., & Chao, M. T. (1972). Nonparametric function fitting. *Journal of the Royal Statistical Society: Series B*, *34*, 385–392.
Shephard, R. W. (1970). *Theory of cost and production function*. Princeton, New Jersey: Princeton University Press.
Watson, G. S. (1964). Smooth regression analysis. *Sankhyā*, *26*, 175–184.