

# Data envelope fitting with constrained polynomial splines

Abdelaati Daouia<sup>a</sup>, Hohsuk Noh<sup>b</sup> and Byeong U. Park<sup>c</sup>

<sup>a</sup> Toulouse School of Economics, University of Toulouse, France  
& Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, Belgium (abdelaati.daouia@tse-fr.eu)

<sup>b</sup> Department of Statistics, Sookmyung Women's University, South Korea  
& Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, Belgium (word5810@gmail.com)

<sup>c</sup> Department of Statistics, Seoul National University, South Korea  
(bupark@stats.snu.ac.kr)

## Abstract

Estimation of support frontiers and boundaries often involves monotone and/or concave edge data smoothing. This estimation problem arises in various unrelated contexts, such as optimal cost and production assessments in econometrics and master curve prediction in the reliability programs of nuclear reactors. Very few constrained estimators of the support boundary of a bivariate distribution have been introduced in the literature. They are based on simple envelopment techniques which often suffer from lack of precision and smoothness. Combining the edge estimation idea of Hall, Park and Stern with the quadratic spline smoothing method of He and Shi, we develop a novel constrained fit of the boundary curve which benefits from the smoothness of spline approximation and the computational efficiency of linear programs. Using cubic splines is also feasible and more attractive under multiple shape constraints; computing the optimal spline smoother is then formulated into a second-order cone programming problem. Both constrained quadratic and cubic spline frontiers have a similar level of computational complexity to the unconstrained fits and inherit their asymptotic properties. The utility of this method is illustrated through applications to some real datasets and simulation evidence is also presented to show its superiority over the best known methods.

AMS 2000 subject classification: 62G05, 62P20, 62P30

*Key words* : Boundary curve; Concavity; Least majorant; Linear programming; Monotone smoothing; Multiple shape constraints; Polynomial spline; Second-order cone programming.

## 1 Introduction

Frontier modeling, that is, estimating the topological extremity of the support of a bivariate density function, is one of the basic tools in statistical applications. This has been well reflected by the expanding recent literature on data edge and data envelope analysis. A

number of semi- and non-parametric techniques have been proposed, including extreme values (de Haan and Resnick (1994), Hall *et al.* (1997), Gijbels and Peng (2000), and Girard and Jacob (2003, 2004)), projections (*e.g.* Jacob and Suquet (1995)), piecewise polynomials (Härdle *et al.* (1995), and Korostelev and Tsybakov (1993)), or local polynomials (Hall and Park (2004), Hall *et al.* (1998), and Knight (2001)).

In this article we focus on the less-discussed problem of estimating boundary curves that are believed or required to be monotone. This problem has increasing usage in education, finance, management, physics, public policy, and other arenas. It is also closely related to edge estimation in image reconstruction (see, *e.g.*, Korostelev and Tsybakov (1993) and Park (2001) for the literature therein). Typical examples appear, for instance, in medicine where the probability of contracting a certain disease depends monotonically on certain factors, or in environmental pollution where monotonicity applies to the ozone level as a function of the inversion base temperature (Croux *et al.* (2012)). In Section 5 we shall discuss economical applications where monotonicity or monotone concavity are requisite theoretical axioms. Also, we shall discuss concrete examples in production theory and nuclear reactors' reliability.

Most of the works on frontier estimation do not rely on the monotonicity constraint and require large samples to provide good results. There are mainly two known methods for preserving monotonicity: the free disposal hull (FDH) and the data envelopment analysis (DEA). The FDH estimator is the lowest “stair-case” monotone curve covering all the data points (see, *e.g.*, Korostelev *et al.* (1995)). When the joint support is in addition convex, the DEA estimator is defined as the least concave majorant of the FDH frontier (see, *e.g.*, Gijbels *et al.* (1999)). Although FDH and DEA estimators are very simple in nature, their full statistical aspects have been elucidated only during the last decade. See, for instance, Jeong and Park (2006), Kneip *et al.* (2008), Daouia *et al.* (2010) and Park *et al.* (2010) for recent contributions. An improved version of the FDH estimator, referred to as the linearized FDH (LFDH) and obtained by drawing the polygonal line smoothing the staircase FDH curve, has been considered in Hall and Park (2002) and Jeong and Simar (2006).

Although the FDH, LFDH and DEA estimators are very easy to implement and provide the fitted values at the observed predictor with monotonicity, they undersmooth the data and underestimate the true support boundary. These vexing defects are more exacerbated in case of small samples as those explored in our applications. Typically, the development of the asymptotic theory of these estimators requires the assumption that the unknown frontier function,  $\varphi$ , is at least continuously differentiable. It is then natural to incorporate such information into the estimation procedure. The idea of this paper is to combine spline smoothing with monotonic boundary estimation. It is well known that  $\varphi$  and  $\varphi'$  can

be uniformly approximated by polynomial splines and their derivatives (see, *e.g.*, Dierckx (1993) and Schumaker (2007)). We propose to estimate the data edge with a monotone spline function defined on a suitably chosen set of knots, which envelopes the full data and minimizes the area under its graph. A similar idea can be found in Hall *et al.* (1998), where the boundary curve is rather modeled by a single polynomial of known degree, and without the inherent monotonicity constraint. The argument of polynomial approximation is very attractive in terms of both pragmatic and didactic advantages. Spline functions extend the advantages of polynomials to have greater flexibility as they are piecewise polynomials with specified continuity constraints. They also afford the possibility of imposing monotonicity and addressing a wide variety of settings, especially the range of applications which are likely to have non-polynomial boundary curves.

The first proposed estimator in this work is derived by minimizing the integration of a polynomial spline, and both monotonicity and data envelopment can be characterized by linear constraints. In this way, the minimization problem can be efficiently solved by a very simple linear programming algorithm. A similar estimator was considered in the context of regression smoothing with monotonicity by He and Shi (1998), who suggested using a constrained least absolute deviation principle in the space of polynomial splines to impose monotonicity. Our approach is different from theirs at least in the following two important aspects: it has the additional data envelopment constraint and its optimization criterion is the integration of splines rather than the  $L_1$ -type loss function. However, we share with He and Shi the elegant idea of using quadratic splines on a selected knot mesh to impose the monotone constraint efficiently. Higher-order splines are more appealing for smoothness, but monotonicity can no longer be characterized as linear constraints at the knots.

Yet, using cubic splines to estimate a smooth monotone support boundary is also possible. The key argument is that the necessary and sufficient condition for a cubic spline smoother to be nondecreasing can be characterized as second-order cone constraints. The envelopment constraint remains a linear one, and hence is a second-order cone constraint as well. Therefore, the estimation of the unknown parameters can be formulated into a standard second-order conic programming problem. This method inherits the attractive properties of cubic spline approximations and the computational efficiency of convex optimization. We refer to Alizadeh and Goldfarb (2003) for an influential paper in the second-order cone programming (SOCP) literature. Nowadays, SOCP has come into use in various areas of statistics, especially in high dimensional statistical learning and constrained nonparametric function estimation. Concerning the latter, which is related to our problem, Wang and Li (2008) have used SOCP for isotonic smoothing spline regression, and Papp and Alizadeh (2014) have suggested recently to apply this technique in a more general manner to some

shape constrained estimation problems including density and regression estimation.

Our estimated frontier curves are sometimes positively biased and sometimes not, depending on the slope and curvature of the sample boundary and on the selected locations of knots. This marks a substantial difference of our construction from the traditional data envelopment techniques that provide overall negatively biased estimates even when the data points concentrate near or on the boundary curve. This also highlights the important feature of our estimators whose deviation errors have some sign balance leading to better bias properties as indicated by the Monte-Carlo evidence in Section 4.

The choice of the number and location of knots for regularizing both estimated quadratic and cubic spline functions is a major issue in practice, but the shape constraint makes this selection easier than the unconstrained smoothing problem. Indeed, as monotonicity reduces sharp changes in the slope and curvature of the estimated frontier, typically a very small number of knots will suffice for the success of our methods. Considering a set of knots equally spaced in percentile ranks, an adequate number of inter-knot segments can be determined by analogy to the popular Akaike information criterion (AIC) and Bayesian information criterion (BIC). Both these selection criteria work remarkably well as demonstrated below in various simulated scenarios in Section 4.

In some applications, concavity is also an important characteristic of the monotone function being fitted (see, *e.g.*, Gijbels *et al.* (1999) and the references therein). Both linear and second-order cone programming problems can then easily be expanded to include the additional concavity constraints that are linear. This is much harder to do with other methods except for the piecewise linear DEA approach.

Section 2 describes in detail the constrained quadratic and cubic spline smoothing methods, including computation via linear programming and second-order cone programming as well as knot selection procedures. Section 3 presents some indicative rates of strong uniform convergence for the unconstrained frontier estimates. We show that the monotone quadratic spline fit inherits the asymptotic rate of its unrestricted version, and the same holds true for the cubic spline fit under both separate and simultaneous monotonicity and concavity constraints. Section 4 provides a comparison with the best known frontier estimation methods through Monte Carlo experiments. Section 5 discusses several applications and illustrates our methods in measuring the reliability of nuclear reactors and the production performance of air controllers and electric utility companies. Section 6 concludes and the Appendix provides necessary mathematical proofs.

## 2 Constrained Polynomial Spline Smoothing

Suppose that we have  $n$  pairs of observations  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , from a bivariate distribution with a density  $f(x, y)$  in  $\mathbb{R}^2$ . The support  $\Psi$  of  $f$  is assumed to be of the form

$$\begin{aligned}\Psi = \{(x, y) | y \leq \varphi(x)\} &\supseteq \{(x, y) | f(x, y) > 0\}, \\ \{(x, y) | y > \varphi(x)\} &\subseteq \{(x, y) | f(x, y) = 0\},\end{aligned}\tag{1}$$

where  $\varphi$  is a monotone increasing and/or concave function whose graph corresponds to the locus of the curve above which the density  $f$  is zero. Without much loss of generality, we restrict ourselves to  $x_i \in [0, 1]$  and  $y_i \geq 0$ . We are interested in estimating  $\varphi$  based on the sample  $\{(x_i, y_i), i = 1, \dots, n\}$  by making use of its spline approximation.

As we shall discuss below in Section 2.1, a monotone quadratic spline smoother of  $\varphi$  can easily be defined using either the B-spline basis or the truncated power function basis. In spite of the computational expedience of the resulting estimator using linear programming, revealing its asymptotic properties becomes a tedious matter in the case of multiple shape constraints because of the discontinuity at knots of the second derivatives of the underlying piecewise quadratic polynomials. Fortunately, making use of cubic splines, we have been able to come up with a satisfactory solution in Section 2.2. The asymptotics of the obtained fit can fully be elucidated under both separate and simultaneous monotonicity and concavity constraints, but its implementation requires the more complex second-order cone programming. It is also possible to incorporate multiple simultaneous shape constraints into the estimation procedure by using any higher-order polynomial spline referring to the work of Papp and Alizadeh (2014). However, one can only use the truncated power function basis for the implementation of the estimators based on cubic or higher-order splines (see the discussion below in Remark 1 of Section 2).

The present paper is the first work to actually implement the idea of constrained boundary regression via polynomial spline approximations. A closely related work in quantile regression theory is He and Shi (1998), where quadratic B-spline monotonicity was inaugurated as pointed out in Section 1. The idea of monotone data smoothing can also be found in a number of standard mean regression problems including Turlach (2005) and Hazelton and Turlach (2011), where penalized spline smoothing with shape constraints is utilized. Lu *et al.* (2007) estimated the mean function of counting processes with panel count data by making use of monotone polynomial splines, and Wang and Meyer (2011) tested the monotonicity or convexity of a function by resorting to regression splines as well. Recently, Pya and Wood (2014) used splines to fit each component in additive models with monotonicity or concavity constraints.

## 2.1 Quadratic splines and linear programming

Denote a partition of  $[0, 1]$  by  $0 = t_0 < t_1 < \dots < t_{k_n} = 1$ . Let  $N = k_n + p$  and  $\pi(x) = (\pi_0(x), \dots, \pi_{N-1}(x))^T$  be any basis of spline functions of order  $p + 1$  based on the knot mesh  $\{t_j\}$  (see, *e.g.*, Schumaker (2007)). To characterize monotonicity as linear constraints at the knots, we choose to use quadratic splines that correspond to  $p = 2$ . We then estimate the frontier function  $\varphi(x)$  by  $\hat{\varphi}_n(x) = \pi(x)^T \hat{\alpha}$ , where  $\hat{\alpha}$  minimizes

$$\int_0^1 \pi(x)^T \alpha \, dx = \sum_{j=0}^{N-1} \alpha_j \int_0^1 \pi_j(x) \, dx \quad (2)$$

over  $\alpha \in \mathbb{R}^N$  subject to envelopment and monotonicity constraints, or equivalently,

$$\pi(x_i)^T \alpha \geq y_i \quad i = 1, \dots, n, \quad \text{and} \quad \pi'(t_j)^T \alpha \geq 0 \quad j = 0, 1, \dots, k_n, \quad (3)$$

with  $\pi'$  being the continuous, piecewise linear derivative of  $\pi$ . It is easily seen that  $\hat{\alpha}$  is identical to the maximum likelihood estimator of the parameter  $\alpha$  in the special case where the true frontier itself is a spline function  $\varphi(x; \alpha) = \pi(x)^T \alpha$ , and data are independent and have a uniform density on the region  $\Psi = \{(x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq \varphi(x; \alpha)\}$ , with  $\varphi(x; \alpha)$  being nondecreasing in  $x$ .

From a computational point of view, minimizing (2) under (3) is an *inequality form* linear program as the objective and constraint functions are all affine and the problem has no equality constraints. Note that the monotonicity adds  $k_n + 1$  linear constraints to  $n$  of them already in use. As the number of knots  $k_n$  is usually small, the added computational burden is negligible. By introducing slack variables  $s = (s_1, \dots, s_{n+k_n+1})^T$  for the inequalities and expressing the variable  $\alpha$  as the difference of two nonnegative variables  $\alpha^+$  and  $\alpha^-$ , this minimization problem can be solved by applying any linear programming algorithm to

$$\begin{aligned} & \text{minimize} && \int_0^1 \pi(x)^T (\alpha^+ - \alpha^-) \, dx \\ & \text{subject to} && \pi(x_i)^T (\alpha^+ - \alpha^-) = y_i + s_i \quad i = 1, \dots, n, \\ & && \pi'(t_j)^T (\alpha^+ - \alpha^-) = s_{n+1+j} \quad j = 0, \dots, k_n, \\ & && \alpha^+ \geq 0, \alpha^- \geq 0, s \geq 0, \end{aligned}$$

which is a linear program in *standard form*, with variables  $\alpha^+$ ,  $\alpha^-$ , and  $s$  (see *e.g.* Boyd and Vandenberghe (2004)). The inequalities here have to be understood componentwise.

As is typical in nonparametric estimation, the selection of knots is critical to the performance of the spline smoother  $\hat{\varphi}_n$ . It is usual to pick a set of knots equally spaced in percentile ranks by taking  $t_j = x_{[jn/k_n]}$ , the  $j/k_n$ th quantile of the values of  $x_i$  for  $j = 1, \dots, k_n - 1$

(see, *e.g.*, He and Shi (1998)). However, considering the special connection of our estimator  $\hat{\varphi}_n$  with the conventional FDH frontier estimator, one can propose an easier way of choosing the knot mesh. The monotone envelopment FDH estimator of  $\varphi$ , defined explicitly as

$$\varphi_n(x) = \max\{y_i, i \in I_x\}$$

with  $I_x = \{i : x_i \leq x\}$ , represents the lowest nondecreasing function that covers the data points  $(x_1, y_1), \dots, (x_n, y_n)$ . Then, a monotone function envelopes all the  $(x_i, y_i)$ 's if and only if it envelopes the extreme FDH points  $(x_i, y_i)$  such that  $y_i = \varphi_n(x_i)$ . Therefore, the spline smoother  $\hat{\varphi}_n$  is identical to the smallest monotone majorant of the FDH function  $\varphi_n$  in the space of quadratic splines, and hence the envelopment constraint in (3) reduces to

$$\pi(\mathcal{X}_i)^T \alpha \geq \mathcal{Y}_i \quad i = 1, \dots, \mathcal{N},$$

where  $(\mathcal{X}_1, \mathcal{Y}_1), \dots, (\mathcal{X}_{\mathcal{N}}, \mathcal{Y}_{\mathcal{N}})$  are the observations  $(x_i, y_i)$  lying on the FDH boundary. This might suggest using the set of knots  $\{t_j = \mathcal{X}_{[j\mathcal{N}/k_n]}, j = 1, \dots, k_n - 1\}$  among the FDH points from the intuition that the  $X$ -locations of FDH points are more appropriate as knots than those of other usual observations. On the other hand, since the number of knots  $k_n$  determines the complexity of the spline approximation, we may view the choice of  $k_n$  as model selection through the minimization of the following two information criteria:

$$AIC(k) = \log \left( \sum_{i=1}^n |y_i - \hat{\varphi}_n(x_i)| \right) + (k + 2)/n, \quad (4)$$

$$BIC(k) = \log \left( \sum_{i=1}^n |y_i - \hat{\varphi}_n(x_i)| \right) + \log n \cdot (k + 2)/(2n). \quad (5)$$

The first one is similar to the famous Akaike information criterion (Akaike, 1973) and the second one to the Bayesian information criterion (Schwartz, 1978).

These information criteria can be motivated by considering the following deterministic frontier model:

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

where  $x_i$  is uniformly distributed on  $[0, 1]$ , and the error term  $u_i \geq 0$  (referred to as inefficiency in applied econometrics) is exponential with mean  $\lambda$ . Here, the regression boundary is a spline function  $\varphi(x; \alpha) = \pi(x)^T \alpha$ , where the basis vector  $\pi(x)$  depends on the number of inter-knot segments  $k$  and the smoothness degree  $p$ . If we see each basis function as a covariate, the choice of the number  $k$  can be viewed as a model selection problem. Denote the maximum likelihood estimators of the model parameters  $\alpha$  and  $\lambda$  as  $\check{\alpha}$  and  $\check{\lambda}$ , respectively.

Then, following the typical definition of AIC and BIC, which is expressed as a penalized log likelihood function and ignoring constants, we have

$$\begin{aligned} -2 \log \text{likelihood}(\check{\alpha}, \check{\lambda}) + 2(k+p) &\propto \log \left( \sum_{i=1}^n (\pi(x_i)^T \check{\alpha} - y_i) \right) + \frac{k+p}{n}; \\ -2 \log \text{likelihood}(\check{\alpha}, \check{\lambda}) + \log n(k+p) &\propto \log \left( \sum_{i=1}^n (\pi(x_i)^T \check{\alpha} - y_i) \right) + \log n \frac{k+p}{2n}. \end{aligned}$$

If we take into account the fact that the maximum likelihood estimator  $\check{\alpha}$  becomes asymptotically the same as the minimizer of  $\int_0^1 \pi(x)^T \alpha dx$ , under the unique envelopment constraint, then we arrive at the formulations (4) and (5) when  $p = 2$ . Both these proposed criteria work reasonably well in our simulations and real data analysis.

The asymptotic theory in the next section shows that the optimal number of knots is of the order of  $n^{1/(3\gamma+1)}$ , where  $\gamma > 0$  stands for the sharpness degree, or equivalently, the quantity  $(\gamma - 1)$  describes the rate at which the joint density  $f$  of the data tends to zero (in case  $\gamma > 1$ ) or to infinity (in case  $\gamma < 1$ ) at the boundary. When the density has sudden jumps at the boundary (in case  $\gamma = 1$ ), the optimal number of knots is of order  $n^{1/4}$ . Given that the selection of smoothing parameters is typically a hard problem in nonparametric boundary regression [see, *e.g.*, Hall *et al.* (1998), Hall and Park (2004) and Daouia *et al.* (2010)], our method benefits from an important advantage of having a simple and effective smoothing parameter selector.

Let  $\tilde{\varphi}_n(x) = \pi(x)^T \tilde{\alpha}$  be the unconstrained B-spline estimator, where  $\tilde{\alpha} \in \mathbb{R}^N$  minimizes (2) only subject to the data envelopment constraints in (3). We shall show that, when the true frontier function  $\varphi$  is strictly increasing,  $\tilde{\varphi}_n$  is monotone and thus equals  $\hat{\varphi}_n$  for sufficiently large  $n$ . This means that we get the monotonicity “free of charge” and the asymptotic properties of  $\tilde{\varphi}_n$  hold automatically for  $\hat{\varphi}_n$ . The basic argument is to use the following proposition and show  $\tilde{\varphi}'_n$  converges to  $\varphi'$  uniformly at all knots. The latter uniform convergence of  $\tilde{\varphi}'_n$  at all knots implies the uniform convergence on the whole interval  $[0, 1]$  due to the fact that  $\tilde{\varphi}'_n$  is piecewise linear. The strong uniform convergence of  $\tilde{\varphi}_n$  and its derivative is proved in the next section. The asymptotic rates of uniform convergence are also provided there.

**Proposition 1.** *Suppose  $\varphi$  has a continuous and strictly positive derivative  $\varphi'$  on  $[0, 1]$  and  $\max_{1 \leq j \leq k_n} |\tilde{\varphi}'_n(t_j) - \varphi'(t_j)| = o(1)$  almost surely. Then*

$$\mathbb{P}[\tilde{\varphi}_n = \hat{\varphi}_n, n \rightarrow \infty] = 1.$$

When the monotone boundary curve is also known or required to be concave, the linear programming problem of (2) and (3) can easily be expanded to include the additional



concavity constraints

$$\pi''(t_j^*)^T \alpha \leq 0 \quad j = 1, \dots, k_n, \quad (6)$$

that are linear, where  $\pi''$  denotes the second derivative of  $\pi$  and  $t_j^*$  stands for the midpoint of  $(t_{j-1}, t_j]$ . The second derivative of a quadratic spline exists on each inter-knot interval and is constant. We then estimate  $\varphi$  by  $\hat{\varphi}_n^*(x) = \pi(x)^T \hat{\alpha}^*$ , where  $\hat{\alpha}^* \in \mathbb{R}^N$  minimizes (2) subject to (3) and (6).

Arguably, the best known estimator for preserving both concavity and monotonicity is the DEA frontier. This piecewise linear estimator being, by construction, the smallest concave nondecreasing curve covering all the  $(x_i, y_i)$ 's, it is necessarily enveloped by the constrained smoother  $\hat{\varphi}_n^*$ , and the envelopment constraints in the linear program are equivalent to

$$\pi(\mathcal{X}_i^*)^T \alpha \geq \mathcal{Y}_i^* \quad i = 1, \dots, \mathcal{M},$$

where  $(\mathcal{X}_1^*, \mathcal{Y}_1^*), \dots, (\mathcal{X}_{\mathcal{M}}^*, \mathcal{Y}_{\mathcal{M}}^*)$  denote the observations  $(x_i, y_i)$  lying on the DEA frontier.

Regarding the choice of knots for computing the optimal concave spline  $\hat{\varphi}_n^*$ , an easy option for it is just applying the same scheme as for  $\hat{\varphi}_n$  by replacing the FDH points  $(\mathcal{X}_i, \mathcal{Y}_i)$  with the DEA points  $(\mathcal{X}_i^*, \mathcal{Y}_i^*)$  in the selection criteria (4) and (5). Typically, the number of DEA points is very small compared to the sample size. As such, our experience with real and simulated data indicates that the strategy of just using all the DEA points as knots is also working quite well for datasets of modest size.

The constrained fit  $\hat{\varphi}_n^*$  is similar to the unconstrained estimate  $\tilde{\varphi}_n$  in terms of computational complexity and computing expedience using linear program. Asymptotically, both smoothers coincide with probability one under the conditions of the following proposition.

**Proposition 2.** *Suppose the conditions of Proposition 1 hold. If  $\varphi$  has a continuous and strictly negative second derivative  $\varphi''$  on  $[0, 1]$  and  $\max_{1 \leq j \leq k_n} |\tilde{\varphi}_n''(t_j^*) - \varphi''(t_j^*)| = o(1)$  almost surely, then  $\mathbb{P}[\tilde{\varphi}_n = \hat{\varphi}_n^*, n \rightarrow \infty] = 1$ .*

Unfortunately, posing the question of strong uniform convergence of  $\tilde{\varphi}_n''$  on the subintervals  $(t_{j-1}, t_j)$  involves some mathematical difficulties that we have not yet succeeded in overcoming. We shall need higher-order splines to obtain the uniform convergence of the second derivative  $\tilde{\varphi}_n''$  as established below in Theorem 1.

All of the methods described above, including computation of  $\tilde{\varphi}_n$ ,  $\hat{\varphi}_n$  and  $\hat{\varphi}_n^*$  via linear programming and knot selection, have been implemented for the R package *npbr* (Daouia *et al.* (2014)). Next, we propose an entirely satisfactory cubic spline based approach that can handle separate as well as simultaneous monotonicity and concavity constraints.

## 2.2 Cubic splines and second-order conic programming

Denote a partition of  $[0, 1]$  by  $0 = t_0 < t_1 < \dots < t_{k_n} = 1$ . Let  $N = k_n + p$  and  $\pi(x) = (\pi_0(x), \pi_1(x), \dots, \pi_{N-1}(x))^T = (1, x, \dots, x^p, (x - t_1)_+^p, \dots, (x - t_{k_n-1})_+^p)^T$  be a vector of truncated power basis based on the knot mesh  $\{t_j\}$  with  $a_+ = \max\{0, a\}$ . We choose here to use cubic splines that correspond to  $p = 3$ . We then estimate  $\varphi(x)$  by  $\tilde{\varphi}_n(x) = \pi(x)^T \tilde{\alpha}$ , where  $\tilde{\alpha}$  minimizes

$$\int_0^1 \pi(x)^T \alpha \, dx = \sum_{j=0}^{N-1} \alpha_j \int_0^1 \pi_j(x) \, dx \quad (7)$$

over  $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{N-1})^T \in \mathbb{R}^N$  subject to the envelopment constraint

$$\pi(x_i)^T \alpha \geq y_i, \quad i = 1, \dots, n.$$

This defines the unconstrained cubic spline estimator of the frontier function. Given that the second derivative of cubic splines is a linear spline, the concavity constraint can be characterized as linear constraints at the knots  $t_j$  themselves instead of the midpoints  $t_j^*$  in the case of quadratic splines, that is,

$$\pi''(t_j)^T \alpha \leq 0, \quad j = 0, 1, \dots, k_n.$$

In contrast, since the first derivative of cubic splines is a quadratic spline, the monotonicity constraint can no longer be formulated into linear constraints at the knots. Yet, it is possible to come up with an alternative appealing representation of monotonicity as standard second-order cone constraints thanks to the following proposition.

**Proposition 3** (Karlin and Studden, 1966). *Let  $p(x) = p_0 + p_1x + p_2x^2$  be a quadratic polynomial. Then  $p(x) \geq 0$  for all  $x \in [0, 1]$  if and only if there exists  $y_0 \geq 0$  such that  $(p_0 + p_2 + y_0, p_0 - p_2 - y_0, p_1 - y_0)^T \in \mathcal{Q}_3$ , where  $\mathcal{Q}_{k+1} = \{(z_0, \dots, z_k) : z_0 \geq \|(z_1, \dots, z_k)^T\|_2\}$  is the  $(k + 1)$ -dimensional second order cone, with  $\|\cdot\|_2$  being the  $L_2$  norm.*

This well-known characterization of nonnegative quadratic polynomials easily extends to a characterization of monotone cubic splines. Indeed, suppose that we have a cubic spline  $f(x) = \sum_{j=0}^3 \alpha_j x^j + \sum_{j=4}^{k_n+2} \alpha_j (x - t_{j-3})_+^3$ . Then the monotonicity constraint means that for all  $j = 1, \dots, k_n$ ,

$$f'((t_j - t_{j-1})z + t_{j-1}) \geq 0 \quad \text{for all } z \in [0, 1].$$

This inequality can be re-expressed as

$$\begin{aligned}
& \alpha_1 + 2t_{j-1}\alpha_2 + 3t_{j-1}^2\alpha_3 + \sum_{l=1}^{j-1} 3\alpha_{l+3}(t_{j-1} - t_l)^2 \\
& + \left\{ 2(t_j - t_{j-1})\alpha_2 + 6(t_j - t_{j-1})t_{j-1}\alpha_3 + \sum_{l=1}^{j-1} 6\alpha_{l+3}(t_j - t_{j-1})(t_{j-1} - t_l) \right\} z \\
& + \left\{ 3\alpha_3(t_j - t_{j-1})^2 + \sum_{l=1}^{j-1} 3\alpha_{l+3}(t_j - t_{j-1})^2 \right\} z^2 \\
& = p_{0j} + p_{1j}z + p_{2j}z^2 \geq 0 \quad \text{for all } z \in [0, 1], \tag{8}
\end{aligned}$$

with obvious definitions for  $p_{0j}$ ,  $p_{1j}$  and  $p_{2j}$ . When  $j = 1$ , we define all the summations in (8) to be zero. Let  $u = (\alpha^T, (y^0)^T)^T$  where  $y^0 = (y_1^0, \dots, y_{k_n}^0)^T$  and  $y_j^0 \geq 0$  for  $j = 1, \dots, k_n$ . When  $j = 2, \dots, k_n$ , note that  $p_{0j} + p_{2j} + y_j^0 = (d_j^T, e_{j,k_n}^T)^T u := c_j^T u$ , where

$$d_j = (0, 1, 2t_{j-1}, 6t_{j-1}^2 - 6t_{j-1}t_j + 3t_j^2, 3((t_{j-1} - t_1)^2 + (t_j - t_{j-1})^2), \dots, 3(t_j - t_{j-1})^2, 0, \dots, 0)^T \in \mathbb{R}^N$$

and  $e_{j,k_n}$  is a standard unit vector of  $\mathbb{R}^{k_n}$  with the  $j$ th element being 1. Moreover,  $p_{0j} - p_{2j} - y_j^0 = (B_{1j}^T, -e_{j,k_n}^T)^T u := A_{1j}^T u$ , where

$$B_{1j} = (0, 1, 2t_{j-1}, 6t_{j-1}t_j - 3t_j^2, 3((t_{j-1} - t_1)^2 - (t_j - t_{j-1})^2), \dots, -3(t_j - t_{j-1})^2, 0, \dots, 0)^T \in \mathbb{R}^N$$

and  $p_{1j} - y_j^0 = (B_{2j}^T, -e_{j,k_n}^T)^T u := A_{2j}^T u$ , with

$$B_{2j} = (0, 0, 2(t_j - t_{j-1}), 6(t_j - t_{j-1})t_{j-1}, 6(t_j - t_{j-1})(t_{j-1} - t_1), \dots, 0, 0, \dots, 0)^T \in \mathbb{R}^N.$$

Therefore, according to Proposition 3, the monotonicity constraint means that there exist  $y^0 = (y_1^0, \dots, y_{k_n}^0)^T$  with  $y_j^0 \geq 0$  such that  $\|(A_{1j}^T u, A_{2j}^T u)^T\|_2 \leq c_j^T u$  for every  $j = 2, \dots, k_n$  (called second-order cone constraints), where  $u = (\alpha^T, (y^0)^T)^T$ . It is not hard to verify that the second-order cone constraint holds when  $j = 1$ . This is the key argument for the estimation of the unknown parameters  $\alpha$  of the constrained nondecreasing cubic spline to be formulated into a second-order cone programming (SOCP) problem.

In summary, cubic spline smoothing under monotonicity and/or concavity constraints requires solving the following typical convex programming problem with respect to  $u = (\alpha^T, (y^0)^T)^T$ :

$$\begin{aligned}
& \text{minimize} && \left( \int_0^1 \pi_0(x) dx, \dots, \int_0^1 \pi_{N-1}(x) dx, 0_{k_n} \right)^T u && (9) \\
& \text{subject to} && \|A_j u\|_2 \leq c_j^T u, \quad j = 1, \dots, k_n, && \text{(monotonicity constraint)} \\
& && (\pi''(t_j)^T, 0_{k_n}) u \leq 0, \quad j = 0, \dots, k_n, && \text{(concavity constraint)} \\
& && (\pi(x_i)^T, 0_{k_n}) u \geq y_i, \quad i = 1, \dots, n, && \text{(envelopment constraint)} \\
& && [O_{k_n, N}, I_{k_n}] u \geq 0_{k_n}. &&
\end{aligned}$$

Here,  $A_j = (A_{1j}, A_{2j})^T$ ,  $0_{k_n}$  is a zero vector of size  $k_n$ ,  $O_{k_n, N}$  is a  $k_n \times N$  zero matrix and  $I_{k_n}$  is the identity matrix of dimension  $k_n$ . Since both concavity and envelopment constraints are linear ones, and hence are second-order cone constraints, the problem (9) is identical to a standard SOCP problem (see the review article by Alizadeh and Goldfarb (2003)). This optimization model is solvable with minimal running time using available off-the-shelf softwares. We may take any solution to be our estimate. For our numerical simulations and real data analysis, we used CVX, a MATLAB-based free package for specifying and solving convex programs because of its user-friendly nature and efficient implementation (for details, refer to Grant and Boyd (2008, 2013)).

**Remark 1.** One may wonder whether a similar formulation for our cubic spline smoother is possible via the more popular B-spline basis. Such a characterization is indeed still valid under the single concavity constraint, but Proposition 3 can no longer be helpful when imposing the monotonicity constraint due to the recursive nature of B-spline basis definition. As a different attempt of using monotone polynomial splines in conjunction with a vector of B-spline basis functions  $\pi(x) = (\pi_0(x), \pi_1(x), \dots, \pi_{N-1}(x))^T$ , Lu *et al.* (2007) and Pya and Wood (2014) considered the following sufficient but not necessary condition to make the resulting cubic B-spline  $\sum_{j=0}^{N-1} \alpha_j \pi_j(x)$  monotone:

$$\alpha_0 \leq \alpha_1 \leq \dots \leq \alpha_{N-1}. \quad (10)$$

Despite the computational expedience of the resulting cubic spline smoother using linear programming, we favor the use of the second-order conic formulation (9) as we cannot completely characterize monotonicity by relying only on the sufficient constraint (10).

When only the monotonicity constraint is of interest, we estimate the frontier function  $\varphi(x)$  by  $\check{\varphi}_n(x) = \pi(x)^T \check{\alpha}$ , where  $\check{\alpha} \in \mathbb{R}^N$  is the solution of the SOCP problem (9) without the concavity constraint. Under the monotonicity and concavity constraints, we estimate  $\varphi(x)$  by  $\check{\varphi}_n^*(x) = \pi(x)^T \check{\alpha}^*$ , where  $\check{\alpha}^*$  is the solution of the full optimization problem (9). Next, we show that both restricted estimators  $\check{\varphi}_n$  and  $\check{\varphi}_n^*$  inherit the same asymptotic properties as the unrestricted version  $\tilde{\varphi}_n$ .

**Proposition 4.** *Let  $\tilde{\varphi}_n(x) = \pi(x)^T \tilde{\alpha}$  be the unconstrained cubic spline estimator, where  $\tilde{\alpha} \in \mathbb{R}^N$  minimizes (7) only subject to data envelopment. If  $\varphi$  has a continuous and strictly positive derivative  $\varphi'$  on  $[0, 1]$  with  $\sup_{x \in [0, 1]} |\tilde{\varphi}'_n(x) - \varphi'(x)| = o(1)$  almost surely, then*

$$\mathbb{P}[\tilde{\varphi}_n = \check{\varphi}_n, n \rightarrow \infty] = 1.$$

*If in addition  $\varphi$  has a continuous and strictly negative second derivative  $\varphi''$  on  $[0, 1]$  with  $\sup_{x \in [0, 1]} |\tilde{\varphi}''_n(x) - \varphi''(x)| = o(1)$  almost surely, then*

$$\mathbb{P}[\tilde{\varphi}_n = \check{\varphi}_n^*, n \rightarrow \infty] = 1.$$

The next section provides indicative asymptotic rates of uniform convergence for the derivatives  $\tilde{\varphi}'_n$  and  $\tilde{\varphi}''_n$  of the unconstrained estimate. In what regards the knot selection process, the same powerful way of regularizing the quadratic spline estimates  $\hat{\varphi}_n$  and  $\hat{\varphi}_n^*$ , described in Section 2.1, can be applied to the cubic spline versions  $\check{\varphi}_n$  and  $\check{\varphi}_n^*$ , respectively. The asymptotic theory in the following section shows that the optimal number of knots is of order  $n^{1/(4\gamma+1)}$  for cubic spline smoothing, and hence a smaller number of knots is typically needed relative to quadratic spline smoothing.

### 3 Some asymptotic results

Due to the arguments made earlier in Propositions 1 and 4, the asymptotic properties of the unconstrained spline estimate  $\tilde{\varphi}_n$  carry over automatically to the monotonic quadratic spline frontier  $\hat{\varphi}_n$  and to both constrained cubic spline smoothers  $\check{\varphi}_n$  and  $\check{\varphi}_n^*$ . As a matter of fact, spline smoothing does not appear to have been considered before even in the literature on unconstrained frontier estimation. In this section we initiate a study of such estimation procedures.

Let  $0 = t_0 < t_1 < \dots < t_{k_n} = 1$  be a knot sequence. We consider splines of general order  $(p + 1)$ . We first note that any spline function of order  $(p + 1)$  based on the B-spline basis and knot meshes  $t_j$  can be re-expressed as  $\pi(x)^T \alpha$ , where  $\pi(x) = (1, x, \dots, x^p, (x - t_1)_+^p, \dots, (x - t_{k_n-1})_+^p)^T$ ,  $\alpha$  is a  $(k_n + p)$ -dimensional vector and  $a_+ = \max\{0, a\}$ . Thus, the problem is to minimize  $\int_0^1 \pi(x)^T \alpha dx$  subject to  $y_i \leq \pi(x_i)^T \alpha$  for all  $1 \leq i \leq n$ .

Let  $q_j(x) = x^j - (j + 1)^{-1}$  for  $1 \leq j \leq p$  and  $q_j(x) = (x - t_{j-p})_+^p - (p + 1)^{-1}(1 - t_{j-p})^{p+1}$  for  $p + 1 \leq j \leq p + k_n - 1$ . The unconstrained spline estimator is then  $\tilde{\varphi}_n(x) = \pi(x)^T \hat{\alpha}$ , where  $(\hat{\alpha}_j : 1 \leq j \leq p + k_n - 1)$  minimizes  $\max_{1 \leq i \leq n} [y_i - \sum_{j=1}^{p+k_n-1} \alpha_j q_j(x)]$  and

$$\hat{\alpha}_0 = \max_{1 \leq i \leq n} \left[ y_i - \sum_{j=1}^p \hat{\alpha}_j x_i^j - \sum_{k=1}^{k_n-1} \hat{\alpha}_{p+k} (x_i - t_k)_+^p \right].$$

Below, we demonstrate the uniform rates of convergence of  $\tilde{\varphi}_n$  and its derivatives, which is also of independent interest. We consider the general setting where the density function  $f$  of the data  $(x_i, y_i)$  may have sudden jumps at its support boundary, decay to zero or rise up to infinity at a speed of power  $\gamma - 1$  ( $\gamma > 0$ ) of the distance from the boundary. More specifically, we assume

(A1)  $f(x, y) = 0$  for all  $y > \varphi(x)$  and  $f(x, y) = \gamma[\varphi(x) - y]^{\gamma-1} \mu(x) + o((\varphi(x) - y)^{\gamma-1})$  as  $y \uparrow \varphi(x)$  for some  $\gamma > 0$ , where the function  $\mu$  is bounded away from zero on  $[0, 1]$ ;

(A2)  $\varphi$  has a bounded  $(p + 1)$ th order derivative on  $[0, 1]$ ;

(A3)  $\max_{1 \leq j \leq k_n} (t_j - t_{j-1}) / \min_{1 \leq j \leq k_n} (t_j - t_{j-1})$  is bounded.

Condition (A1) has been considered in Härdle *et al.* (1995), Hall *et al.* (1998), Gijbels and Peng (2000), Hwang *et al.* (2002) and Daouia *et al.* (2010), to name a few. Note that the case  $\gamma \leq 1$  corresponds to sharp or fault-type boundaries. When  $\gamma > 1$ , the density decays to zero smoothly as it approaches the support frontier. The smoothness of the frontier function  $\varphi$  is given in (A2). A similar assumption was used in the usual and more-discussed problems of nonparametric central and/or quantile regression estimation (see, *e.g.*, He and Shi 1998). Condition (A3) is standard in spline smoothing, it can also be found in earlier work by He and Shi (1994). For simplicity, we use  $a_n \sim b_n$  to mean that  $a_n/b_n$  and  $b_n/a_n$  are bounded.

**Theorem 1.** *Assume that the conditions (A1)–(A3) hold. If  $k_n \sim (n/\log n)^{1/[(p+1)\gamma+1]}$ , then with probability one*

$$\sup_{x \in [0,1]} |\tilde{\varphi}_n^{(m)}(x) - \varphi^{(m)}(x)| = O\left((n^{-1} \log n)^{(p-m)/[(p+1)\gamma+1]}\right), \quad 0 \leq m \leq p-1.$$

As a direct consequence of Theorem 1, we get indicative asymptotic rates of global convergence of  $\tilde{\varphi}_n$  and its derivative  $\tilde{\varphi}'_n$ :

$$\begin{aligned} \sup_{x \in [0,1]} |\tilde{\varphi}_n(x) - \varphi(x)| &= O\left((n^{-1} \log n)^{p/[(p+1)\gamma+1]}\right) \quad a.s.; \\ \sup_{x \in [0,1]} |\tilde{\varphi}'_n(x) - \varphi'(x)| &= O\left((n^{-1} \log n)^{(p-1)/[(p+1)\gamma+1]}\right) \quad a.s. \end{aligned}$$

The obtained rates depend on both the sharpness degree  $\gamma$  and the smoothness order  $p$ . The smaller  $\gamma$ , the faster the attainable uniform convergence rates are.

**Remark 2.** In the standard mean and/or median regression, Stone (1982) and He and Shi (1998) established that the optimal rates of uniform convergence are  $(n^{-1} \log n)^{p/(2p+1)}$  for the regression curve estimate ( $m = 0$ ) and  $(n^{-1} \log n)^{(p-1)/(2p+1)}$  for its derivative ( $m = 1$ ) when the number of knots  $k_n \sim (n/\log n)^{1/(2p+1)}$ . In this case, the proposed boundary smoothing method attains better convergence rates than the classical central regression if and only if  $\gamma < 2p/(p+1)$ . This is what happens in the irregular setting  $\gamma \leq 1$  for all  $p \geq 2$ .

In regions where the boundary is strictly increasing, we obtain rates of convergence of the monotone quadratic and cubic spline estimates  $\hat{\varphi}_n$  and  $\check{\varphi}_n$  as well, as can easily be seen from Propositions 1 and 4, respectively.

**Corollary 1.** *Suppose that the boundary function  $\varphi$  has a continuous and strictly positive derivative  $\varphi'$  on  $[0, 1]$ . If (A1)–(A3) hold and  $k_n \sim (n/\log n)^{1/[(p+1)\gamma+1]}$  with  $p = 2$  for  $\hat{\varphi}_n$*

and  $p = 3$  for  $\check{\varphi}_n$ , then we have with probability one

$$\begin{aligned} \sup_{x \in [0,1]} |\hat{\varphi}_n(x) - \varphi(x)| &= O((n^{-1} \log n)^{2/(3\gamma+1)}); \\ \sup_{x \in [0,1]} |\hat{\varphi}'_n(x) - \varphi'(x)| &= O((n^{-1} \log n)^{1/(3\gamma+1)}), \\ \sup_{x \in [0,1]} |\check{\varphi}_n(x) - \varphi(x)| &= O((n^{-1} \log n)^{3/(4\gamma+1)}); \\ \sup_{x \in [0,1]} |\check{\varphi}'_n(x) - \varphi'(x)| &= O((n^{-1} \log n)^{2/(4\gamma+1)}). \end{aligned}$$

Hence, the higher-order spline  $\check{\varphi}_n$  is more appealing than  $\hat{\varphi}_n$  for both smoothness and speed of convergence. In small samples, however, the accuracy of the quadratic spline smoother is quite respectable in comparison with the cubic version as indicated by the Monte-Carlo evidence in the next section.

The results in Corollary 1 and those in Corollary 2 below are only indicative rates of strong uniform convergence for our spline estimators. These indicative rates of global convergence are not superior to the optimal pointwise convergence rates of the traditional FDH and DEA estimators unless we assume a certain degree of smoothness for the frontier function. For example, in the “fault-type boundary” case  $\gamma = 1$ , we shall need to assume that  $p \geq 2$  under the monotonicity constraint and  $p \geq 4$  under the monotone concavity constraint. Although this is an obvious criticism of our asymptotic results at this stage, theory can be improved by providing better rates of convergence. Doubtless, further work will yield new refinements. The two selling points of our procedures that we emphasize in this paper are remarkable smoothness and better bias properties for small, moderate and even large samples. We shall discuss in Section 5, in detail and in a comprehensive way, the importance of smoothness and especially differentiability of  $\varphi$  in the econometric literature on frontier analysis. Also, the deviation errors of our spline smoothers have sign balance by construction, leading to better bias properties as evidenced in Section 4.

We also provide the uniform convergence of  $\check{\varphi}_n''$  on the unit interval  $[0, 1]$ , which enables us to get the concavity of the cubic spline  $\check{\varphi}_n^*$  free of charge. According to Theorem 1, if the conditions (A1)-(A3) hold and  $k_n \sim (n/\log n)^{1/[(p+1)\gamma+1]}$  with  $p = 3$ , then

$$\sup_{x \in [0,1]} |\check{\varphi}_n''(x) - \varphi''(x)| = O((n^{-1} \log n)^{1/(4\gamma+1)})$$

with probability one. By applying Proposition 4 in conjunction with this result, we obtain indicative rates of convergence of the constrained smoother  $\check{\varphi}_n^*$  and its derivatives  $\check{\varphi}_n^{* \prime}$  and  $\check{\varphi}_n^{* \prime \prime}$  in regions where the frontier is strictly increasing and concave.

**Corollary 2.** *Suppose that the boundary function  $\varphi$  has a strictly positive derivative  $\varphi'$  and a continuous and strictly negative second derivative  $\varphi''$  on  $[0, 1]$ . If the conditions (A1)-(A3)*

hold and  $k_n \sim (n/\log n)^{1/[(p+1)\gamma+1]}$  with  $p = 3$ , we have with probability one

$$\begin{aligned} \sup_{x \in [0,1]} |\check{\varphi}_n^*(x) - \varphi(x)| &= O\left((n^{-1} \log n)^{3/(4\gamma+1)}\right), \\ \sup_{x \in [0,1]} |\check{\varphi}_n^{*'}(x) - \varphi'(x)| &= O\left((n^{-1} \log n)^{2/(4\gamma+1)}\right), \\ \sup_{x \in [0,1]} |\check{\varphi}_n^{*''}(x) - \varphi''(x)| &= O\left((n^{-1} \log n)^{1/(4\gamma+1)}\right). \end{aligned}$$

## 4 Monte Carlo evidence

Some numerical evidence is given in this section to demonstrate the superiority of the proposed spline smoothers  $\hat{\varphi}_n$ ,  $\check{\varphi}_n$ ,  $\hat{\varphi}_n^*$  and  $\check{\varphi}_n^*$  over the best known constrained and unconstrained frontier estimators based on data envelopment ideas. Those were the popular linearized FDH (LFDH) and DEA estimators described in Section 1, and the modern local-polynomial smoothing estimator of Hall *et al.* (1998). Specifically, the local linear frontier estimator is defined by

$$\begin{aligned} \hat{\varphi}_{n,LL}(x) = \min \left\{ z : \text{there exists } \theta_1 \text{ such that } y_i \leq z + \theta_1(x_i - x) \right. \\ \left. \text{for all } i \text{ such that } x_i \in (x - h, x + h) \right\}. \end{aligned}$$

Hall and Park (2004) proposed a bootstrap procedure for selecting the bandwidth  $h$  in  $\hat{\varphi}_{n,LL}$ .

To evaluate finite-sample performance of the constrained spline smoothed estimators in comparison with the various frontier estimates described above, we have undertaken some simulation experiments. The experiments all employ the model  $y_i = \varphi(x_i)v_i$ , where  $x_i$  is uniform on  $[0, 1]$  and  $v_i$ , independent of  $x_i$ , is Beta( $\beta, \beta$ ) with values of  $\beta = 0.5, 1$  and  $3$  (corresponding, respectively, to a joint density of the  $(x_i, y_i)$ 's tending to infinity, having a jump or converging to zero as it approaches the frontier points). The frontier function  $\varphi$  is either linear  $\varphi_a(x) = x$ , concave  $\varphi_b(x) = x^{1/2}$ , or  $\varphi_c(x) = \exp(-5+10x)/(1+\exp(-5+10x))$ .

All the experiments were performed over  $N = 200$  independent samples of size  $n = 25, 50, 100$  and  $200$ . In Tables 1, 2 and 3 we report the simulation results devoted to accuracy of the seven estimation methods: LFDH, DEA, QS (monotone quadratic spline,  $\hat{\varphi}_n$ ), CS (monotone cubic spline,  $\check{\varphi}_n$ ), LL (local linear), QS-C (monotone and concave quadratic spline,  $\hat{\varphi}_n^*$ ) and CS-C (monotone and concave cubic spline,  $\check{\varphi}_n^*$ ). To assess the 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

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

where  $\{z_i, i = 0, \dots, I\}$  is an equispaced grid with width  $1/I$  over  $[0, 1]$  with  $I = 1000$ ,  $\hat{\varphi}^{(j)}(\cdot)$  is the estimated frontier function from the  $j$ -th data sample and  $\bar{\varphi}(z_i) = N^{-1} \sum_{j=1}^N \hat{\varphi}^{(j)}(z_i)$ . To guarantee a fair comparison among the different methods, we used the smoothing parameter which minimizes the MISE for the spline estimators (QS and CS) and the local linear estimator (LL). Regarding the monotone and concave spline estimators (QS-C and CS-C), we just used all the DEA points as knots for simplicity as explained earlier in Section 2.

Additionally, to see how our automatic selection procedures for  $k$  perform in practice, we compared the results when the number of knots is selected by (4) and (5) for the estimators  $\hat{\varphi}_n$  and  $\check{\varphi}_n$ . We call those versions of  $\hat{\varphi}_n$  QS-A and QS-B, respectively, and those versions of  $\check{\varphi}_n$  CS-A and CS-B, respectively. Also included was the local linear frontier estimator with the optimal bandwidth being chosen by the bootstrap procedure proposed in Hall and Park (2004), called LL-B. For the sake of conciseness, we only present the results for QS-B and CS-B. The results for QS-A and CS-A were qualitatively similar to those for QS-B and CS-B, so they are omitted. Actually, to initiate Hall and Park (2004)'s bootstrap procedure, we need to set a pilot bandwidth, which we have found to be quite critical to the quality of their procedure. To see how the local linear frontier estimator performs empirically at the best, we used the bandwidth which minimizes the MISE as the pilot bandwidth. Two typical realizations of the experiment, with  $\varphi \in \{\varphi_b, \varphi_c\}$ ,  $\beta = 0.5$  and  $n = 50$ , are shown in Figures 2 and 3.

Overall, the spline-based estimators (QS, CS, QS-B, CS-B, QS-C and CS-C) show better performance than all the other estimators regardless of the boundary type (which depends on  $\beta$ ) and the sample size. It is clear that the spline-based estimators enjoy the benefit of smooth approximation in reducing the bias when the true function is smooth as shown in Figures 2 and 3. Moreover, both selection criteria of the number of knots seem to work quite well in practice. It is remarkable that our splined-based estimators (QS-B and CS-B) with empirically chosen knots are performing better than the local linear estimator with the theoretically MISE-optimal bandwidth.

When the true frontier function is concave and monotone ( $\varphi = \varphi_a$  or  $\varphi_b$ ), we observe that the DEA and the concave spline estimators (QS-C and CS-C) have lower IVARs compared

to the other estimators. The reason is that they enjoy both monotonicity and concavity properties, reducing thus the unnecessary sampling variability. Moreover, the concave spline estimators have the added advantage over the DEA estimator of reducing the bias thanks to their modeling flexibility and function approximation power as illustrated in Figure 2.

When the frontier function  $\varphi$  is simply monotone but not concave, the local linear estimator (LL) seems to be a useful alternative to the LFDH estimator if the bandwidth is judiciously chosen. However, the LL frontier lacks of smoothness and has no guarantee of being monotone even if the true frontier is so. Accordingly, following the curvature of the monotone frontier  $\varphi$ , its LL estimator is likely to exhibit substantial bias when the number of data points is not large enough, especially at the sample boundaries, as shown in the left panel of Figure 4. A simple way to remedy to this drawback is to impose the extra condition  $\theta_1 \geq 0$  in the definition of  $\hat{\varphi}_{n,LL}(x)$  to get

$$\hat{\varphi}_{n,LL2}(x) = \min \left\{ z : \text{there exists } \theta_1 \geq 0 \text{ such that } y_i \leq z + \theta_1(x_i - x) \right. \\ \left. \text{for all } i \text{ such that } x_i \in (x - h, x + h) \right\}.$$

This version reduces the vexing border defect of the LL estimator as illustrated in the right panel of Figure 4. We actually utilized the improved version  $\hat{\varphi}_{n,LL2}(x)$  in all our simulations instead of  $\hat{\varphi}_{n,LL}(x)$ . Yet, it may be seen from Figure 3 and Table 3 that our spline-based estimators are clearly superior to both the  $\hat{\varphi}_{n,LL2}(x)$  estimator (computed with the theoretically MISE-optimal bandwidth) and the LFDH estimator.

## 5 Data examples

In this section, we first discuss the importance of the monotonicity and concavity shape constraints in frontier analysis, and then illustrate the utility of the proposed spline-based estimators through three concrete data examples.

A popular field of application where the problem of constrained frontier estimation naturally occurs is concerned with the analysis of economic efficiency of firms in a variety of industries. The data typically consist of an input factor  $x_i \in \mathbb{R}_+$  used to produce an output  $y_i \in \mathbb{R}_+$  in a certain firm  $i$ . The conventional microeconomic theory is based on the assumption of optimizing behaviour. Thus it is assumed that producers optimize their production choices from a technical or engineering perspective by avoiding wasting resources. This means that, theoretically, producers operate somewhere on the boundary, rather than on the interior, of their production possibility set  $\Psi = \{(x, y) \in \mathbb{R}_+^2 \mid x \text{ can produce } y\}$ . This optimal boundary, referred to as production, profit or cost frontier, is interpreted as the set of the most efficient firms. The economic performance of a firm is then defined in terms of its

ability to operate close to or on the efficient frontier. From a statistical viewpoint, the efficient frontier may be viewed as the upper extremity of the joint support  $\Psi$  of the population of firms density in the input and output space. This support boundary is often described by the graph of a frontier function  $\varphi(x)$ , which gives the maximal level of output attainable by a firm operating with input-usage  $x$ . The efficiency of a production unit located at  $(x, y)$  may then be estimated via the distance between the attained output  $y$  and the optimal level  $\varphi(x)$ .

Econometric considerations lead to the general production axiom of free disposability of inputs and outputs, that is, if  $(x, y) \in \Psi$  then  $(x', y') \in \Psi$  for any  $x' \geq x$  and  $y' \leq y$ . The monotonicity of  $\varphi(x)$ , referred to as non-negative marginal productivity, is justified by this free disposability assumption. The set  $\Psi$  of physically attainable points  $(x, y)$  is also often assumed to be convex, that is, every convex combination of feasible production units is also feasible. This justifies the concavity property of the frontier function  $\varphi(x)$ , known in economics as diminishing marginal returns. Thus, the differentiability of  $\varphi(x)$  across the support of the covariate is a desirable microeconomic feature if one is interested in measuring the responsiveness of output to changes in the input [see, *e.g.*, Parmeter and Racine (2013)]. This is particularly important if one is interested in returns to scale, which is defined as the sum of input elasticities, which in turn rely on the second derivative  $\varphi''(x)$ . Although the concavity of  $\varphi(x)$  is widely used in economics, it is not always valid. The production set  $\Psi$  might admit increasing returns to scale, *i.e.*, the outputs increase faster than the inputs. Hence, the monotonicity of  $\varphi(x)$  is a minimal requirement in production theory [see, *e.g.*, Färe, Grosskopf and Lovell (1985)].

It should also be clear that the monotonicity of the support boundary is an immediate consequence of the frequently encountered property of stochastic monotonicity of the distribution  $F(\cdot|x)$  of  $y_i$  conditional on  $x_i = x$ . This stronger hypothesis is of genuine interest in many applications beyond frontier and efficiency analysis in economics. It describes the increasingness effect in the data, or more specifically,

$$\text{for each } y \in \mathcal{Y}, \quad F(y|x) \leq F(y|x') \quad \text{whenever } x \geq x' \quad \text{for } x, x' \in \mathcal{X},$$

with  $\mathcal{Y}$  and  $\mathcal{X}$  being, respectively, the supports of the variables  $y_i$  and  $x_i$ . This property is equivalent to the monotonicity of all regression quantile curves, and hence implies the boundary curve's monotonicity. It is quite natural in production theory as the chance of producing less than a value  $y$  decreases if a firm uses more inputs. A description of many other applied settings where this regression dependency of  $y_i$  on  $x_i$  naturally appears can be found in Lee, Linton and Whang (2009).

The first motivating application that we consider in this section is concerned with the

reliability of nuclear reactors. An accurate knowledge of the change in fracture toughness of the reactor pressure vessel materials as a function of the temperature is of prime importance in a nuclear power plant lifetime program. Physical considerations lead to the natural hypothesis of stochastic monotonicity and particularly to the assumption that the master curve prediction, that is, the set of materials having optimal fracture toughness, is monotonely increasing. The scatterplot of 254 non-irradiated representative steels obtained from the US Electric Power Research Institute (EPRI), along with the three monotone regression quartile curves, are given in Figure 1 (left panel). Here, it is important to know both lower and upper limits of fracture toughness of each material as a function of temperature. This translates into estimating both optimal support boundaries. Although our focus in the sections above was only on the estimation of the upper support extremity, similar considerations evidently apply to the estimation of the lower boundary. It may be seen from Figure 5, which shows various estimates of the upper and lower frontier functions, that the spline-based estimates (QS-B and CS-B) suggest better capability of fitting edge data. It may also be noted that the lower frontier estimator via CS-B exhibits, as is to be expected, more smoothness and indicates a convex and monotonely increasing shape for the master curve prediction.

Our second example is concerned with the increase of the production activity of 123 American electric utility companies. The measurements for each company of the produced output and the total cost involved in the production, along with the three monotone quartile curves, are represented in Figure 1 (middle panel). Naturally, the true full econometric frontier, that is, the locus of the most efficient firms, is nondecreasing as well.

Another related application is concerned with the assessment of the performance of 37 European Air Controllers. The inefficiency of each controller can be measured by its “distance” from the efficient econometric boundary. The scatterplot of the controllers in the year 2000, along with the three regression quartile curves, are given in Figure 1 (right panel). Here, the activity of the controllers is described by one input (an aggregate factor of different kind of labor) and one output (an aggregate factor of the activity produced, based on the number of air movements controlled, the number flight hours controlled, etc.).

Figure 6 shows the efficient econometric frontier estimates which correspond to the production activity of the 123 American electric utility companies and the 37 European Air Controllers in the left and right panels, respectively. For each dataset, we plot the DEA, the QS-B, the CS-B and the QS-C estimates. Here, imposing concavity appears to improve on imposing monotonicity only since the obtained QS-C fits provide more appealing results in terms of stability and smoothness [the CS-C fits were very similar to the QS-C estimates, and hence they are not reported here]. This is consistent with the (non)parametric findings in the econometric literature where the production set (*i.e.* the joint support of data) is

mostly assumed to be convex [see, *e.g.*, Gijbels *et al.* (1999) for the electric utility companies and Daouia *et al.* (2008) for the Air Controllers]. Also, the inefficiency measures  $\{QS-C(x_i) - y_i\}$  (imposing both monotonicity and concavity) tend to reveal that a large number of air controllers and electricity plants are highly inefficient. Finally it should be clear that, although the DEA curve tends to produce similar estimates to the concave spline smoothers, it suffers from the vexing lack of differentiability especially if one is interested in marginal effects and elasticities.

## 6 Discussion

We proposed a novel approach of using polynomial spline fitting for the problem of constrained nonparametric boundary regression. The method allows to handle both single and multiple shape constrained estimation. We mainly considered monotone and/or concave frontier smoothing. Using cubic splines requires solving a convex programming problem with second-order conic constraints to characterize monotonicity, and only linear constraints to represent both envelopment and concavity constraints. The proposed constrained fits are similar to the unconstrained estimates in terms of computational complexity without sacrificing modeling flexibility and function approximation power. They also have the same asymptotic rate of strong uniform convergence.

Quadratic spline smoothing results in a simpler linearly constrained model, and the restricted fit is shown to inherit the asymptotic rate of uniform convergence of its unrestricted version only under the monotonicity constraint. The key advantage of the quadratic spline fit over the cubic spline estimate is its computational expedience using linear programming. Additionally, the latter can be implemented only with truncated power function basis rather than the more popular B-spline basis, whereas the former can be implemented using both bases. By contrast, the cubic spline smoother is the winner in terms of both smoothness and speed of global convergence. Although both approaches work quite well and either might be used in practice, we have a particular preference for the cubic spline fit as it exhibited slightly better performance than the quadratic spline estimate in our simulation studies. The quadratic spline smoothing method has been implemented for the R package *npbr* (Daouia *et al.* 2014) and all the MATLAB codes for the cubic spline smoothing procedure are available upon request.

Although we focus on the single input case in this paper, extending our methodology to the more general setting of multiple inputs is of genuine interest since many encountered applications, especially in efficiency and productivity analysis, involve multiple cost variables. One way to extend our results is to consider a tensor product of B-spline space and its

corresponding basis and develop a multivariate spline-based frontier estimator which satisfies shape and envelopment constraints. Based on the properties of derivatives of B-spline basis [see Section 5.2 in Schumaker (2007)], one may impose some desirable restrictions on the estimator of the frontier surface such as componentwise increasingness and concavity, and the resulting constrained estimator can be implemented via linear programming. This idea is not new. It was applied in classical mean regression in conjunction with multivariate Bernstein polynomials [see Chapter 3 in Wang (2011)]. However, two major obstacles remain to be solved. First, this type of multivariate extension is probably only of a theoretical value in the case of very high-dimensional problems, because the number of basis for the estimator inevitably becomes enormous. Second, the full concavity with respect to all coordinates can no longer reduce conveniently to negativity, but translates into the negative semidefiniteness of the Hessian. As such, more efforts are needed to bring such shape constraints under the umbrella of available computing techniques.

The question of whether similar work can also be done via alternative known monotone regression methods is another concern of interest. One may adapt, for instance, the monotone polynomials' prescription of Hawkins (1994) and Murray *et al.* (2013) to our specific setup of frontier analysis. The argument of a single polynomial approximation is very attractive as demonstrated earlier by Hall *et al.* (1998) in the unconstrained setting. Yet, the constrained polynomial splines we advocate have greater modeling flexibility and function approximation power. Moreover, as opposed to purely polynomial methods, spline fitting allows for easy handling of the additional concavity constraint. Another feasible option under multiple shape constraints is to adapt the elegant kernel regression device of Hall and Huang (2001) and Du *et al.* (2013). This idea has been recently implemented by Parmeter and Racine (2013) and improved in Noh (2014) by developing kernel-type boundary smoothers computed via quadratic programming, with suitably selected bandwidths. A drawback of this method so far is, however, that it does not ensure perfectly the desired monotonicity or monotone concavity constraints. Also, the asymptotic integrity of the proposed kernel estimators is not justified and does appear to be very challenging and even harder to derive compared to our procedure.

It should be also pointed out that the bias of our spline-based construction can be pronounced in the setting  $\gamma > 1$  of “non-sharp boundary”. It is therefore of high interest, both from a theoretical and a practical perspective, to correct our final spline estimates for their inherent bias in this case. Applying some promising methods for bias reduction at boundaries including extreme-value techniques, bootstrap and excess-mass ideas from Daouia *et al.* (2010), Peng (2004), Hall and Park (2002) and Polonik (1995) may lead to significantly improved estimation results, but we do not pursue this further in this paper. A thorough

bias reduction requires a precise second-order asymptotic expansion of limit distributions involving estimates of the sharpness  $\gamma$ , which is beyond the scope of this paper.

Finally, for knot selection we used the information criteria derived under the assumption that the inefficiency follows exponential distributions, but one may use other information criteria coming from other distributional assumptions. To the best of our knowledge, it is not known which criterion is better than others, neither how the distributional assumption affects theoretical properties of the resulting information criteria. This is an interesting open problem which deserves further investigation.

## Acknowledgements

The authors would like to thank two referees, the Associate Editor and the Joint Editor for their valuable suggestions, which have significantly improved the paper. A. Daouia's research was supported by the Toulouse School of Economics Individual Research Fund (IRF/Daouia-20125) and the Seventh Framework Programme of the European Union (IEF/273584/EMBAF-project). H. Noh's research was supported by IAP research network P7/06 of the Belgian Government (Belgian Science Policy), the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650 and the Sookmyung Women's University Research Grant 1-1309-0006. Research of B. U. Park was supported by the NRF Grant funded by the Korea government (MEST) No. 2010-0017437.

## Appendix

### A.1 Proof of Proposition 1

The key idea goes as in He and Shi (1998, p. 646). Let  $\varepsilon = \inf\{\varphi'(x), x \in [0, 1]\}$ . We have  $\varphi'(t_j) \geq \varepsilon > 0$  for each  $j = 1, \dots, k_n$ . Then, by the strong uniform convergence of  $\tilde{\varphi}'_n$  to  $\varphi'$  at knots, there exists an  $n_\varepsilon$  such that for all  $n > n_\varepsilon$  and all  $j = 1, \dots, k_n$ , we have  $\tilde{\varphi}'_n(t_j) > \varphi'(t_j) - \varepsilon/2 \geq \varepsilon/2$ . Hence, with probability one, we get  $\tilde{\varphi}'_n(t_j) > 0$  at all knots for all  $n > n_\varepsilon$ . The derivative  $\tilde{\varphi}'_n$  being piecewise linear, it follows that  $\tilde{\varphi}'_n(x) > 0$  at all  $x \in [0, 1]$ . Therefore  $\mathbb{P}[\tilde{\varphi}_n = \hat{\varphi}_n, \forall n > n_\varepsilon] = 1$ .

### A.2 Proof of Proposition 2

Let  $\eta = \sup\{\varphi''(x), x \in [0, 1]\} < 0$ . By the uniform convergence of  $\tilde{\varphi}''_n$  to  $\varphi''$  at midpoints, there exists an  $n_\eta$  such that for all  $n > n_\eta$  and all  $j = 1, \dots, k_n$ , we have  $\tilde{\varphi}''_n(t_j^*) < -\eta/2 + \varphi''(t_j^*) \leq \eta/2 < 0$ . Because  $\tilde{\varphi}''_n$  is piecewise constant, this implies that  $\tilde{\varphi}''_n(x) < 0$  at all

$x \in (t_{j-1}, t_j)$ , for each  $j = 1, \dots, k_n$ , and all  $n > n_\eta$ . Since  $\tilde{\varphi}_n$  is also monotone on  $[0, 1]$  for all  $n > n_\varepsilon$ , we get  $\mathbb{P}[\tilde{\varphi}_n = \hat{\varphi}_n^*, \forall n > n_\eta \vee n_\varepsilon] = 1$ .

### A.3 Proof of Proposition 4

Let  $\varepsilon = \inf\{\varphi'(x), x \in [0, 1]\} > 0$ . By the uniform convergence of  $\tilde{\varphi}'_n$  to  $\varphi'$  over  $[0, 1]$ , with probability one, there exists an  $n_\varepsilon$  such that for all  $n > n_\varepsilon$ , we have  $\tilde{\varphi}'_n(x) > \varphi'(x) - \varepsilon/2 \geq \varepsilon/2$  for all  $x \in [0, 1]$ . Hence, with probability one, we get  $\tilde{\varphi}'_n(x) > 0$  at all  $x \in [0, 1]$  and for all  $n$  large enough. Whence  $\mathbb{P}[\tilde{\varphi}_n = \check{\varphi}_n, n \rightarrow \infty] = 1$ .

Let  $\eta = \sup\{\varphi''(x), x \in [0, 1]\} < 0$ . By the strong uniform convergence of  $\tilde{\varphi}''_n$  to  $\varphi''$ , there exists an  $n_\eta$  such that for all  $n > n_\eta$  and all  $x \in [0, 1]$ , we have  $\tilde{\varphi}''_n(x) < -\eta/2 + \varphi''(x) \leq \eta/2 < 0$ . Then  $\tilde{\varphi}''_n(x) < 0$  at all  $x \in [0, 1]$  whenever  $n > n_\eta$ , which implies the concavity of  $\tilde{\varphi}_n$  on  $[0, 1]$ , with probability one. Therefore  $\mathbb{P}[\tilde{\varphi}_n = \check{\varphi}_n^*, \forall n > n_\eta \vee n_\varepsilon] = 1$ .

### A.4 Proof of Theorem 1

Let  $z_i = \nu_n(y_i - \varphi(x_i))$  and  $\delta_n = \max_{1 \leq j \leq k_n}(t_j - t_{j-1})$ , where  $\nu_n = n^{1/\gamma}(k_n \log n)^{-1/\gamma}$ . From (A3) we have  $k_n \delta_n \sim 1$ . Here and below,  $a_n \sim b_n$  means that  $a_n/b_n$  and  $b_n/a_n$  are bounded. Let  $\omega(f, \delta : [a, b]) = \sup_{0 \leq h \leq \delta} \sup_{a \leq x \leq b-h} |f(x+h) - f(x)|$  be a modulus of continuity of  $f$  on the interval  $[a, b]$ . According to an approximation theorem for spline functions (see Theorem 6.20 in Schumaker, 2007, for example), there exists a  $(k_n + p)$ -dimensional vector  $\alpha^*$  such that, for all  $m : 0 \leq m \leq p$ ,

$$\sup_{t_{j-1} \leq x \leq t_j} |\varphi^{(m)}(x) - \alpha^{*\top} \pi^{(m)}(x)| \leq C_0 \delta_n^{p-m} \omega(\varphi^{(p)}, \delta_n : [t_{j-p-1}, t_{j+p}]), \quad 1 \leq j \leq k_n, \quad (\text{A.1})$$

where  $C_0$  is a positive constant that depends only on  $p$  and we set  $t_{-p} = \dots = t_{-1} = 0$ ,  $t_{k_n+1} = \dots = t_{k_n+p} = 1$ . We note that  $\pi_{p+j}$ , the  $(p+j)$ th component of  $\pi$ , is not  $p$  times differentiable at  $x = t_{j-1}$ . In (A.1),  $\pi_{p+j}^{(p)}(t_{j-1})$  is understood as the  $p$ th right derivative of  $\pi$  at  $x = t_{j-1}$ . From (A.1) and the condition (A2), we get

$$\sup_{x \in [0, 1]} |\varphi^{(m)}(x) - \alpha^{*\top} \pi^{(m)}(x)| \leq C'_0 \delta_n^{p+1-m}, \quad 0 \leq m \leq p \quad (\text{A.2})$$

for some constant  $C'_0 > 0$ .

Define  $r_n(x) = \nu_n[\varphi(x) - \pi(x)^\top \alpha^*]$  and  $\Delta_j = \nu_n(\hat{\alpha}_j - \alpha_j^*)$  for  $0 \leq j \leq k_n + p - 1$ . For a vector  $d = (d_1, \dots, d_{k_n+p-1})$ , let

$$T(d) = \max_{1 \leq i \leq n} \left[ z_i - \sum_{j=1}^{p+k_n-1} d_j q_j(x_i) + r_n(x_i) \right]$$



where  $q_j$  are defined in Section 3. Thus, by definition  $\Delta = (\Delta_1, \dots, \Delta_{k_n+p-1})$  minimizes  $T(d)$ . We note that

$$\begin{aligned} \nu_n \sup_{x \in [0,1]} |\tilde{\varphi}_n(x) - \varphi(x)| &= \sup_{x \in [0,1]} |\pi(x)^T \Delta - r_n(x)| \\ &\leq \sum_{j=0}^p |\Delta_j| + \sum_{j=1}^{k_n-1} (1-t_j)^p |\Delta_{p+j}| + \sup_{x \in [0,1]} |r_n(x)| \\ &\leq |\Delta_0| + W_n + \sup_{x \in [0,1]} |r_n(x)|, \end{aligned} \quad (\text{A.3})$$

where  $W_n = \sum_{j=1}^p |\Delta_j| + c \sum_{j=p+1}^{p+k_n-1} \left(1 - \frac{j-p}{k_n}\right)^p |\Delta_j|$  and  $c$  denotes the generic positive constant that depends only on  $p$ . We also note that there exist constants  $C_m > 0$  for  $1 \leq m \leq p-1$  such that

$$\nu_n \sup_{x \in [0,1]} |\tilde{\varphi}_n^{(m)}(x) - \varphi^{(m)}(x)| \leq C_m k_n^m W_n + \sup_{x \in [0,1]} |r_n^{(m)}(x)|.$$

We take  $\delta_n \sim (n^{-1} \log n)^{1/((p+1)\gamma+1)}$ , so that  $\sup_{x \in [0,1]} |r_n(x)| = O(\nu_n \delta_n^{p+1}) = O(1)$  and  $\sup_{x \in [0,1]} |r_n^{(m)}(x)| = O(\nu_n \delta_n^{p+1-m}) = O((n^{-1} \log n)^{-m/((p+1)\gamma+1)})$  by (A.2). The main part of the proof is to verify that  $|\Delta_0|$  and  $W_n$  in (A.3) are of order  $O(k_n)$  with probability one, which gives the theorem. To show  $W_n = O(k_n)$  with probability one, we use the fact that there exists a random integer  $1 \leq J \leq p+k_n$  such that  $\Delta_j < 0$  for all  $j < J$  and  $\Delta_j \geq 0$  for all  $j \geq J$ . This can be proved along the lines in the proof of Theorem 2.1 of Hall *et al.* (1998). We also claim that there exists an absolute constant  $c_0 > 0$  such that, on the event  $J = j$ ,

$$\sup_{x \in I_j} \sum_{l=1}^{p+k_n-1} \Delta_l q_l(x) \leq -c_0 W_n / k_n, \quad (\text{A.4})$$

where  $I_j$  are the intervals defined below. The inequality (A.4) implies that, for a sufficiently large  $C > 0$ ,

$$\begin{aligned} P(W_n > C k_n) &\leq \sum_{j=1}^{p+k_n} P \left[ -\frac{1}{c_0} \sup_{x \in I_j} \sum_{l=1}^{p+k_n-1} \Delta_l q_l(x) > C, J = j \right] \\ &\leq \sum_{j=1}^{p+k_n} P \left[ T(\Delta) \geq \max_{i: x_i \in I_j} z_i + \frac{2}{3} c_0 C, J = j \right] \\ &\leq \sum_{j=1}^{p+k_n} P \left[ T(0) \geq \max_{i: x_i \in I_j} z_i + \frac{2}{3} c_0 C, J = j \right] \\ &\leq \sum_{j=1}^{p+k_n} P \left( \max_{i: x_i \in I_j} z_i \leq -\frac{1}{3} c_0 C \right). \end{aligned} \quad (\text{A.5})$$

The second inequality of (A.5) follows from the facts that  $\sup_{x \in [0,1]} |r_n(x)| = O(1)$  and  $T(\Delta) \geq \max_{i: x_i \in I_j} [z_i - \sum_{l=1}^{p+k_n-1} \Delta_l q_l(x_i) + r_n(x_i)]$ . The fourth inequality holds since  $T(0) \leq \sup_{x \in [0,1]} |r_n(x)|$ .

The event  $\max_{i: x_i \in I_j} z_i \leq -c_0 C/3$  occurs if and only if there is no  $(x_i, y_i)$  in the set  $A_n \equiv \{(x, y) : x \in I_j, y \in [\varphi(x) - \nu_n^{-1} c_0 C/3, \varphi(x)]\}$ . From (A.8) and (A.9) below it follows that

$$P[(x_i, y_i) \in A_n] = \int_{A_n} f(x, y) dy dx \geq c \cdot c_0^\gamma \cdot C^\gamma \frac{\log n}{n} (1 + o(1)) \quad (\text{A.6})$$

for some constant  $c > 0$ . This implies

$$\sum_{j=1}^{p+k_n} P\left(\max_{i: x_i \in I_j} z_i \leq -\frac{1}{3} c_0 C\right) \leq (p + k_n) n^{-c \cdot C^\gamma} \quad (\text{A.7})$$

for some constant  $c > 0$ , so that  $\sum_{n=1}^{\infty} P(W_n > C k_n) < \infty$  for sufficiently large  $C > 0$ . This proves  $W_n = O(k_n)$  with probability one.

To prove  $|\Delta_0| = O(k_n)$  with probability one, we note that

$$\begin{aligned} |\Delta_0| &\leq \left| \max_{1 \leq i \leq n} \left( z_i - \sum_{j=1}^{p+k_n-1} \Delta_j \pi_j(x_i) \right) \right| + \sup_{x \in [0,1]} |r_n(x)| \\ &\leq -\max_{1 \leq i \leq n} z_i + \max\{1, c\} W_n + O(1) \end{aligned}$$

for the constant  $c$  at (A.3). Since it holds that  $-\max_{i \leq i \leq n} z_i = O(k_n^{-1/\gamma})$  with probability one, we get that  $|\Delta_0| = O(k_n)$  with probability one.

It remains to prove the claim (A.4). Let  $\xi_j = (j+1)^{-1/j}$  for  $1 \leq j \leq p$  and  $\xi_{p+k} = t_k + (p+1)^{-1/p} (1-t_k)^{1+(1/p)}$  for  $1 \leq k \leq k_n - 1$ . Then,  $0 = \xi_0 < \xi_1 < \dots < \xi_{p+k_n-1} < \xi_{p+k_n} = 1$ . Write  $\xi_{j,1} = \xi_{j-1} + (\xi_j - \xi_{j-1})/3$  and  $\xi_{j,2} = \xi_{j-1} + 2(\xi_j - \xi_{j-1})/3$ . Let  $I_j$  denote the interval  $[\xi_{j,1}, \xi_{j,2}]$ . Then, there exists a constant  $c > 0$  such that

$$(\text{length of } I_j) \geq c, \quad 1 \leq j \leq p. \quad (\text{A.8})$$

Let  $g(t) = t + (p+1)^{-1/p} (1-t)^{1+(1/p)}$ . Then,  $g'$  is strictly increasing so that  $g'(t) \geq g'(0) = 1 - p^{-1} (p+1)^{1-(1/p)}$  for all  $t \in [0, 1]$  and  $g'(0) > 0$  for all  $p > 1$ . This means

$$\xi_{p+j+1} - \xi_{p+j} = g(t_{j+1}) - g(t_j) \geq g'(0)(t_{j+1} - t_j),$$

so that there exists a constant  $c > 0$  such that

$$(\text{length of } I_j) \geq c \delta_n, \quad p+1 \leq j \leq p+k_n. \quad (\text{A.9})$$

Here, we have used the condition (A3).

Since  $q_l(x) < 0$  if  $x < \xi_l$  and  $q_l(x) \geq 0$  if  $x \geq \xi_l$ , there exists a constant  $c > 0$  such that, for all  $x \in I_J = [\xi_{J,1}, \xi_{J,2}]$ ,

$$\begin{aligned} q_l(x) &\geq c(\xi_{J,1} - \xi_l) \quad \text{if } 1 \leq l \leq J-1, \\ q_l(x) &\leq -c(\xi_l - \xi_{J,2}) \leq \quad \text{if } l \geq \max\{J, p+1\} \text{ and } \xi_{J,2} \geq t_{l-p}, \\ q_l(x) &= -(1 - t_{l-p})^{p+1}/(p+1) \quad \text{if } l \geq \max\{J, p+1\} \text{ and } \xi_{J,2} < t_{l-p}, \\ q_l(x) &\leq -c(\xi_l - \xi_{J,2}) \quad \text{if } J \leq l < \max\{J, p+1\}. \end{aligned}$$

We note that, in case  $p+1 \leq l \leq p+k_n-1$ ,

$$-\frac{(1 - t_{l-p})^{p+1}}{p+1} \leq -c \left(1 - \frac{l-p}{k_n}\right)^{p+1}$$

for some constant  $c > 0$ . Define  $l(J) = \max\{J, p+1\}$  and for  $l(J) \leq l \leq p+k_n-1$

$$a(J, l) = (\xi_l - \xi_{J,2})I(\xi_{J,2} \geq t_{l-p}) + \left(1 - \frac{l-p}{k_n}\right)^{p+1} I(\xi_{J,2} < t_{l-p}).$$

Since  $\Delta_l < 0$  for all  $l < J$  and  $\Delta_l \geq 0$  for all  $j \geq J$ , there exists a constant  $c > 0$  such that, for all  $x \in I_J$ ,

$$\sum_{l=1}^{p+k_n-1} \Delta_l q_l(x) \leq -c \left[ \sum_{l=1}^{J-1} (\xi_{J,1} - \xi_l) |\Delta_l| + \sum_{l=J}^{l(J)-1} (\xi_l - \xi_{J,2}) |\Delta_l| + \sum_{l=l(J)}^{p+k_n-1} a(J, l) |\Delta_l| \right].$$

The second sum on the right hand side of the above inequality is set to be zero in case  $l(J) = J$ , i.e.,  $J \geq p+1$ . Now, we note that  $\xi_{J,1} - \xi_l \geq (\xi_J - \xi_{J-1})/3 \geq ck_n^{-1}$  for all  $1 \leq l \leq J-1$  and  $\xi_l - \xi_{J,2} \geq (\xi_J - \xi_{J-1})/3 \geq ck_n^{-1}$  for all  $J \leq l \leq p+k_n$ , where  $c > 0$  is a constant. Furthermore, there exists a constant  $c > 0$  such that

$$\begin{aligned} a(J, l) &\geq ck_n^{-1} \left[ I(\xi_{J,2} \geq t_{l-p}) + \left(1 - \frac{l-p}{k_n}\right)^p I(\xi_{J,2} < t_{l-p}) \right] \\ &\geq ck_n^{-1} \left(1 - \frac{l-p}{k_n}\right)^p. \end{aligned}$$

These entail that, for some constant  $c_0 > 0$ ,

$$\sum_{l=1}^{p+k_n-1} \Delta_l q_l(x) \leq -c_0 k_n^{-1} \left[ \sum_{l=1}^{l(J)-1} |\Delta_l| + \sum_{l=l(J)}^{p+k_n-1} \left(1 - \frac{l-p}{k_n}\right)^p |\Delta_l| \right] \leq -c_0 k_n^{-1} W_n$$

for all  $x \in I_J$ .

## References

- [1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, in *Second International Symposium of Information Theory*, eds. B. N. Petrov and F. Csaki, Budapest: Akademia Kiado, 267–281.
- [2] Alizadeh, F. and Goldfarb, D. (2003). Second-order cone programming, *Mathematical Programming Series B*, **95**, 3–51.
- [3] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*, Cambridge University Press, Cambridge.
- [4] Croux, C., Gijbels, I. and Prosdocimi, I. (2012). Robust estimation of mean and dispersion functions in extended generalized additive models. *Biometrics*, **68**, 31–44.
- [5] Daouia, A., Florens, J.P. and Simar, L. (2008). Functional Convergence of Quantile-type Frontiers with Application to Parametric Approximations, *Journal of Statistical Planning and Inference*, **138**, 708–725.
- [6] Daouia, A., Florens, J.P. and Simar, L. (2010). Frontier Estimation and Extreme Value Theory, *Bernoulli*, **16**, 1039–1063.
- [7] Daouia, A., Laurent, T. and Noh, H. (2014). *npbr*: A Package for Nonparametric Boundary Regression in R, v1.0. Available at <http://cran.r-project.org/package=npbr>.
- [8] de Haan, L. and Resnick, S. (1994). Estimating the home range, *Journal of Applied Probability*, **31**, 700–720.
- [9] Dierckx, P. (1993). *Curve and Surface Fitting With Splines*, Oxford, U.K.: Clarendon Press.
- [10] Du, P., Parmeter, C.F. and Racine, J.S. (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints, *Statistica Sinica*, **23**, 1347–1371.
- [11] Färe, R., Grosskopf, S., and C.A.K. Lovell (1985). *The measurement of Efficiency of Production*, Boston: Kluwer-Nijhoff Publishing.
- [12] Gijbels, I., Mammen, E., Park, B.U. and Simar, L. (1999). On estimation of monotone and concave frontier functions, *Journal of American Statistical Association*, **94**, 220–228.
- [13] Gijbels, I. and Peng, L. (2000). Estimation of a support curve via order statistics, *Extremes*, **3**, 251–277.
- [14] Girard, S. and Jacob, P. (2003). Extreme values and Haar series estimates of point process boundaries, *Scandinavian Journal of Statistics*, **30**, 369–384.
- [15] Girard, S. and Jacob, P. (2004). Extreme values and kernel estimates of point processes boundaries. *ESAIM: Probability and Statistics*, **8**, 150–168.

- [16] Grant, M. and Boyd, S. (2008). Graph implementations for nonsmooth convex programs, Recent Advances in Learning and Control (a tribute to M. Vidyasagar), V. Blondel, S. Boyd, and H. Kimura, editors, pages 95-110, Lecture Notes in Control and Information Sciences, Springer.
- [17] Grant, M. and Boyd, S. (2013). CVX: Matlab software for disciplined convex programming, version 2.0 beta.
- [18] Hall, P. and Huang, L.S. (2001). Nonparametric kernel regression subject to monotonicity constraints, *Annals of Statistics*, **29**, 624–647.
- [19] Hall, P., Nussbaum, M. and Stern, S.E. (1997). On the estimation of a support curve of indeterminate sharpness, *Journal of Multivariate Analysis*, **62**, 204–232.
- [20] Hall, P. and Park, B.U. (2002). New methods for bias correction at endpoints and boundaries, *Annals of Statistics*, **30**, 1460-1479.
- [21] Hall, P. and Park, B.U. (2004). Bandwidth choice for local polynomial estimation of smooth boundaries, *Journal of Multivariate Analysis*, **91**, 240–261.
- [22] Hall, P., Park, B.U. and Stern, S.E. (1998). On polynomial estimators of frontiers and boundaries, *Journal of Multivariate Analysis*, **66**, 71–98.
- [23] Härdle, W., Park, B.U. and Tsybakov, A.B. (1995). Estimation of non-sharp support boundaries, *Journal of Multivariate Analysis*, **43**, 205–218.
- [24] Hawkins, D.M. (1994) Fitting monotonic polynomials to data. *Computational Statistics*, **9**, 233–247.
- [25] Hazelton, M.L. and Turlach, B.A. (2011). Semiparametric regression with shape constrained penalized splines. *Computational Statistics and Data Analysis*, **55**, 2871–2879.
- [26] He, X. and Shi, P. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions, *Nonparametric statistics*, **3**, 299–308.
- [27] He, X. and Shi, P. (1998). Monotone B-Spline Smoothing, *Journal of American Statistical Association*, **93**, 643–650.
- [28] Hwang, J.H., Park, B.U. and Ryu, W. (2002). Limit theorems for boundary function estimators, *Statistics & Probability Letters*, **59**, 353–360.
- [29] Jacob, P. and Suquet, P. (1995). Estimating the edge of a Poisson process by orthogonal series, *Journal of Statistical Planning and Inference*, **46**, 215–234.
- [30] Jeong, S.-O. and Park, B.U. (2006). Large sample approximation of the distribution for convex-hull estimators of boundaries, *Scandinavian Journal of Statistics*, **33**, 139–151.
- [31] Jeong, S.-O. and Simar, L. (2006). Linearly interpolated FDH efficiency score for non-convex frontiers, *Journal of Multivariate Analysis*, **97**, 2141–2161.

- [32] Karlin, S. and Studden, W.J. (1966). *Tchebycheff Systems, with Applications in Analysis and Statistics*. Wiley Interscience Publishers.
- [33] Kneip, A., Simar, L. and Wilson, P.W. (2008). Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models, *Econometric Theory*, **24**, 1663–1697.
- [34] Knight, K. (2001). Limiting distributions of linear programming estimators, *Extremes*, **4**, 87–103.
- [35] Korostelev, A., Simar, L. and Tsybakov, A.B. (1995). Efficient estimation of monotone boundaries, *Annals of Statistics*, **23**, 476–489.
- [36] Korostelev, A. and Tsybakov, A.B. (1993). *Minimax theory of image reconstruction*. Volume 82 of Lecture Notes in Statistics, Springer-verlag, New-York.
- [37] Lee, S., Linton, O., and Whang, Y.-J. (2009). Testing for stochastic monotonicity, *Econometrica*, **77**(2), 585–602.
- [38] Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines, *Biometrika*, **94**, 705–718.
- [39] Murray, K., Müller, S. and Turlach, B.A. (2013). Revisiting fitting monotone polynomials to data, *Computational Statistics*, **28**, 1989–2005.
- [40] Noh, H. (2014). Frontier estimation using kernel smoothing estimators with data transformation, *Journal of the Korean Statistical Society*, to appear.
- [41] Papp, D. and Alizadeh, F. (2014). Shape constrained estimations using nonnegative splines, *Journal of Computational and Graphical Statistics*, **23**, 1211–1231.
- [42] Park, B.U. (2001). On nonparametric estimation of data edges, *Journal of the Korean Statistical Society*, **30**, 265–280.
- [43] Park, B.U., Jeong, S.-O. and Simar, L. (2010). Asymptotic distribution of conical-hull estimators of directional edges, *Annals of Statistics*, **38**, 1320–1340.
- [44] Parmeter, C.F. and Racine, J.S. (2013). Smooth constrained frontier analysis, in *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, 463–388, Springer-Verlag, New York.
- [45] Peng, L. (2004). Bias-corrected estimators for monotone and concave frontier functions, *Journal of Statistical Planning and Inference*, **119**, 263–275.
- [46] Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters-an excess mass approach, *Annals of Statistics*, **23**, 855–881.
- [47] Pya, N. and Wood, S.N. (2014). Shape constrained additive models, *Statistical Computing*, to appear.

- [48] Schumaker, L.L. (2007). *Spline Functions: Basic Theory*, 3rd edition, Cambridge University Press.
- [49] Schwartz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 461–464.
- [50] Stone, C. (1982). Optimal Global Rates of Convergence for Nonparametric Regression, *Annals of Statistics*, **10**, 1040–1053.
- [51] Turlach, B.A. (2005). Shape constrained smoothing using smoothing splines, *Computational Statistics*, **20**, 81–103.
- [52] Wang, J. C. and Meyer, M. C. (2011). Testing the monotonicity or convexity of a function using regression splines, *The Canadian Journal of Statistics*, **39**, 89–107.
- [53] Wang, J. (2011). *Shape restricted nonparametric regression with Bernstein polynomials*, Ph. D. dissertation.
- [54] Wang, X. and Li, F. (2008). Isotonic smoothing spline regression, *Journal of Computational and Graphical Statistics*, **17**, 21–37.

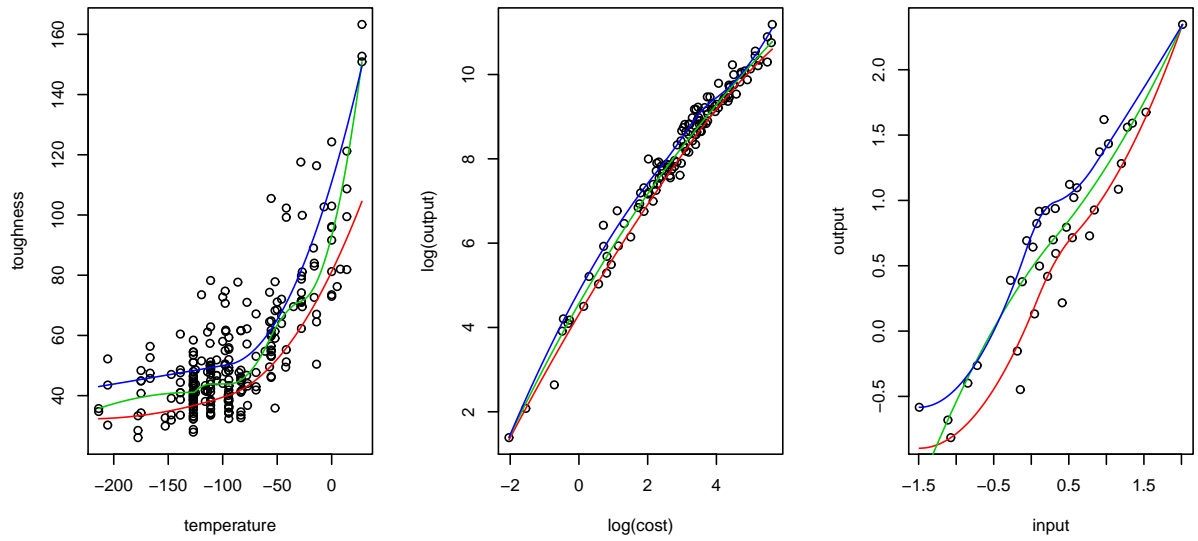


Figure 1: Scatterplots of the nuclear reactors' data (left), the American electric utility companies' data (middle) and the European Air Controllers' data (right), along with the quartile curves in each plot. The three regression quantiles are estimated by the method of He and Shi (1998).

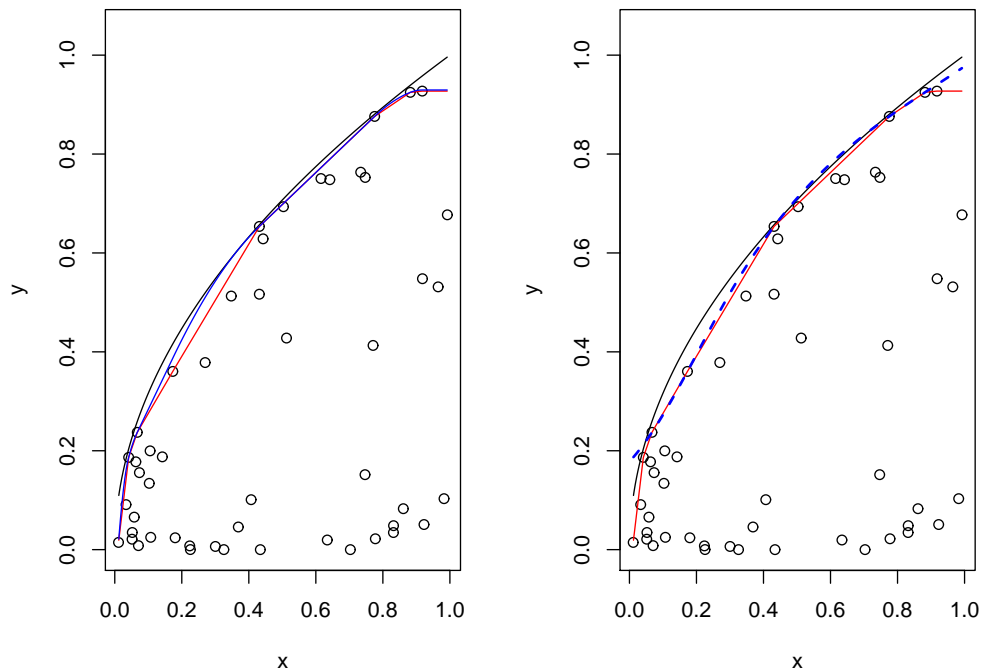


Figure 2: When  $n = 50$  and  $\beta = 0.5$ , the true frontier function ( $\varphi_b$ , black) and its three estimates: DEA (solid red), QS-C (solid blue, left panel) and CS-C (dotted blue, right panel)



Table 1: Comparison when the true frontier is linear ( $\varphi(x) = \varphi_a(x)$ ). All the results are multiplied by 100.

			DEA	QS	CS	LL	QS-C	CS-C	QS-B	CS-B	LL-B
$\beta = 0.5$	$n = 25$	IBIAS2	0.225	0.106	0.094	0.146	0.248	0.122	0.140	0.097	0.255
		IVAR	0.207	0.256	0.146	0.188	0.233	0.173	0.263	0.147	0.256
		IMSE	0.431	0.362	0.241	0.334	0.481	0.294	0.403	0.245	0.512
	$n = 50$	IBIAS2	0.077	0.026	0.022	0.049	0.078	0.037	0.032	0.024	0.078
		IVAR	0.082	0.062	0.049	0.066	0.081	0.064	0.089	0.052	0.081
		IMSE	0.158	0.087	0.071	0.114	0.159	0.101	0.122	0.077	0.159
	$n = 100$	IBIAS2	0.014	0.001	0.001	0.007	0.014	0.004	0.001	0.001	0.014
		IVAR	0.014	0.008	0.004	0.009	0.014	0.008	0.008	0.003	0.014
		IMSE	0.028	0.009	0.005	0.016	0.028	0.011	0.009	0.004	0.028
	$n = 200$	IBIAS2	0.004	0.000	0.000	0.002	0.004	0.001	0.000	0.000	0.004
		IVAR	0.006	0.001	0.001	0.004	0.006	0.003	0.001	0.000	0.006
		IMSE	0.009	0.001	0.001	0.005	0.009	0.004	0.001	0.001	0.009
$\beta = 1$	$n = 25$	IBIAS2	0.589	0.432	0.363	0.446	0.615	0.391	0.548	0.371	0.638
		IVAR	0.254	0.350	0.238	0.252	0.252	0.261	0.385	0.240	0.285
		IMSE	0.843	0.782	0.601	0.699	0.867	0.652	0.933	0.612	0.922
	$n = 50$	IBIAS2	0.217	0.131	0.108	0.149	0.222	0.130	0.184	0.110	0.223
		IVAR	0.099	0.113	0.078	0.080	0.099	0.089	0.163	0.078	0.099
		IMSE	0.317	0.244	0.186	0.230	0.321	0.220	0.347	0.188	0.322
	$n = 100$	IBIAS2	0.120	0.053	0.048	0.082	0.122	0.060	0.078	0.053	0.122
		IVAR	0.055	0.051	0.039	0.042	0.054	0.043	0.070	0.039	0.054
		IMSE	0.175	0.105	0.086	0.124	0.177	0.103	0.149	0.092	0.176
	$n = 200$	IBIAS2	0.031	0.012	0.010	0.022	0.032	0.016	0.016	0.011	0.032
		IVAR	0.015	0.012	0.009	0.012	0.015	0.010	0.019	0.009	0.015
		IMSE	0.047	0.023	0.019	0.035	0.047	0.026	0.035	0.020	0.047
$\beta = 3$	$n = 25$	IBIAS2	2.015	1.834	1.592	1.697	2.083	1.648	2.167	1.608	2.132
		IVAR	0.267	0.387	0.263	0.264	0.260	0.280	0.392	0.263	0.264
		IMSE	2.282	2.221	1.855	1.960	2.343	1.929	2.559	1.872	2.396
	$n = 50$	IBIAS2	1.298	1.162	1.042	1.123	1.353	1.073	1.409	1.053	1.352
		IVAR	0.164	0.200	0.159	0.158	0.160	0.167	0.236	0.159	0.161
		IMSE	1.461	1.362	1.202	1.281	1.513	1.240	1.645	1.213	1.513
	$n = 100$	IBIAS2	0.760	0.661	0.579	0.652	0.789	0.604	0.924	0.595	0.784
		IVAR	0.100	0.130	0.098	0.097	0.095	0.100	0.155	0.099	0.096
		IMSE	0.860	0.791	0.676	0.749	0.884	0.704	1.080	0.694	0.880
	$n = 200$	IBIAS2	0.490	0.418	0.367	0.415	0.504	0.383	0.635	0.379	0.503
		IVAR	0.063	0.086	0.061	0.058	0.062	0.063	0.113	0.062	0.062
		IMSE	0.553	0.504	0.428	0.473	0.566	0.446	0.748	0.442	0.564

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

			DEA	QS	CS	LL	QS-C	CS-C	QS-B	CS-B	LL-B	
$\beta = 0.5$	$n = 25$	IBIAS2	0.283	0.112	0.119	0.287	0.190	0.163	0.124	0.121	0.292	
		IVAR	0.185	0.257	0.187	0.189	0.197	0.194	0.265	0.189	0.199	
		IMSE	0.468	0.368	0.306	0.475	0.387	0.357	0.390	0.311	0.491	
	$n = 50$	IBIAS2	0.085	0.016	0.021	0.086	0.052	0.041	0.019	0.021	0.021	0.086
		IVAR	0.065	0.077	0.056	0.065	0.067	0.064	0.076	0.057	0.057	0.065
		IMSE	0.150	0.094	0.077	0.151	0.119	0.105	0.095	0.078	0.078	0.151
	$n = 100$	IBIAS2	0.022	0.003	0.004	0.022	0.012	0.009	0.005	0.003	0.003	0.022
		IVAR	0.018	0.018	0.013	0.018	0.017	0.016	0.017	0.013	0.013	0.018
		IMSE	0.040	0.021	0.017	0.040	0.029	0.026	0.022	0.017	0.017	0.040
	$n = 200$	IBIAS2	0.007	0.001	0.001	0.007	0.004	0.003	0.005	0.001	0.001	0.007
		IVAR	0.005	0.005	0.003	0.005	0.005	0.005	0.005	0.003	0.003	0.005
		IMSE	0.012	0.005	0.004	0.012	0.009	0.007	0.010	0.005	0.005	0.012
$\beta = 1$	$n = 25$	IBIAS2	0.849	0.505	0.498	0.866	0.641	0.573	0.584	0.508	0.889	
		IVAR	0.277	0.475	0.302	0.281	0.330	0.328	0.487	0.299	0.302	
		IMSE	1.126	0.981	0.800	1.148	0.971	0.901	1.071	0.807	1.191	
	$n = 50$	IBIAS2	0.297	0.144	0.143	0.302	0.210	0.186	0.167	0.147	0.147	0.301
		IVAR	0.109	0.146	0.117	0.110	0.121	0.112	0.158	0.115	0.115	0.109
		IMSE	0.406	0.290	0.260	0.412	0.331	0.298	0.325	0.263	0.263	0.410
	$n = 100$	IBIAS2	0.143	0.054	0.063	0.143	0.098	0.088	0.064	0.064	0.064	0.143
		IVAR	0.051	0.064	0.053	0.052	0.056	0.055	0.068	0.054	0.054	0.052
		IMSE	0.195	0.118	0.117	0.195	0.154	0.143	0.133	0.118	0.118	0.195
	$n = 200$	IBIAS2	0.054	0.017	0.020	0.054	0.036	0.030	0.019	0.019	0.019	0.054
		IVAR	0.019	0.024	0.017	0.020	0.022	0.020	0.022	0.018	0.018	0.019
		IMSE	0.072	0.040	0.038	0.072	0.055	0.050	0.041	0.037	0.037	0.072
$\beta = 3$	$n = 25$	IBIAS2	3.096	2.526	2.481	3.140	2.753	2.657	2.838	2.505	3.160	
		IVAR	0.288	0.438	0.332	0.289	0.341	0.346	0.408	0.323	0.292	
		IMSE	3.384	2.964	2.813	3.429	3.094	3.003	3.246	2.828	3.452	
	$n = 50$	IBIAS2	1.987	1.511	1.521	1.999	1.706	1.644	1.715	1.547	1.547	2.000
		IVAR	0.198	0.303	0.234	0.198	0.239	0.236	0.310	0.232	0.232	0.198
		IMSE	2.184	1.813	1.755	2.197	1.946	1.880	2.026	1.779	1.779	2.198
	$n = 100$	IBIAS2	1.287	0.939	0.952	1.295	1.091	1.047	1.117	0.968	0.968	1.295
		IVAR	0.104	0.169	0.138	0.104	0.131	0.126	0.172	0.129	0.129	0.104
		IMSE	1.391	1.108	1.089	1.400	1.222	1.173	1.289	1.097	1.097	1.399
	$n = 200$	IBIAS2	0.813	0.548	0.567	0.815	0.681	0.645	0.710	0.504	0.504	0.815
		IVAR	0.075	0.110	0.095	0.075	0.093	0.092	0.132	0.088	0.088	0.075
		IMSE	0.888	0.658	0.661	0.890	0.774	0.737	0.843	0.672	0.672	0.890

Table 3: Comparison when the true frontier is monotone but not concave ( $\varphi(x) = \varphi_c(x)$ ). All the results are multiplied by 100.

			LFDH	QS	CS	LL	QS-B	CS-B	LL-B
$\beta = 0.5$	$n = 25$	IBIAS2	0.715	0.153	0.207	0.549	0.143	0.197	0.990
		IVAR	0.591	0.637	0.637	0.539	0.540	0.568	0.467
		IMSE	1.306	0.790	0.844	1.087	0.684	0.766	1.457
	$n = 50$	IBIAS2	0.258	0.029	0.035	0.192	0.031	0.034	0.610
		IVAR	0.216	0.154	0.148	0.170	0.184	0.150	0.132
		IMSE	0.474	0.183	0.184	0.362	0.214	0.185	0.743
	$n = 100$	IBIAS2	0.090	0.006	0.005	0.057	0.012	0.004	0.179
		IVAR	0.079	0.042	0.022	0.062	0.055	0.025	0.041
		IMSE	0.169	0.048	0.027	0.119	0.067	0.029	0.220
	$n = 200$	IBIAS2	0.037	0.001	0.001	0.019	0.004	0.000	0.055
		IVAR	0.034	0.011	0.006	0.012	0.019	0.004	0.011
		IMSE	0.071	0.012	0.006	0.031	0.023	0.004	0.065
$\beta = 1$	$n = 25$	IBIAS2	1.477	0.405	0.594	1.064	0.533	0.678	1.324
		IVAR	0.571	0.642	0.732	0.484	0.650	0.642	0.504
		IMSE	2.048	1.047	1.326	1.548	1.184	1.320	1.829
	$n = 50$	IBIAS2	0.788	0.125	0.221	0.545	0.186	0.252	0.726
		IVAR	0.329	0.321	0.305	0.265	0.351	0.346	0.243
		IMSE	1.117	0.446	0.526	0.810	0.538	0.599	0.969
	$n = 100$	IBIAS2	0.392	0.063	0.071	0.233	0.063	0.090	0.309
		IVAR	0.163	0.110	0.115	0.089	0.138	0.102	0.087
		IMSE	0.555	0.173	0.186	0.322	0.202	0.192	0.396
	$n = 200$	IBIAS2	0.169	0.019	0.023	0.087	0.016	0.027	0.111
		IVAR	0.075	0.052	0.032	0.040	0.062	0.034	0.036
		IMSE	0.244	0.070	0.055	0.127	0.078	0.061	0.147
$\beta = 3$	$n = 25$	IBIAS2	3.919	2.023	2.051	3.019	2.533	2.598	3.088
		IVAR	0.438	0.582	0.713	0.375	0.520	0.574	0.437
		IMSE	4.357	2.605	2.765	3.394	3.053	3.172	3.525
	$n = 50$	IBIAS2	2.587	1.054	1.036	2.321	1.405	1.591	2.318
		IVAR	0.270	0.325	0.474	0.214	0.337	0.311	0.215
		IMSE	2.857	1.379	1.510	2.535	1.742	1.903	2.532
	$n = 100$	IBIAS2	1.801	0.594	0.713	1.315	0.914	1.018	1.345
		IVAR	0.200	0.182	0.316	0.138	0.231	0.201	0.145
		IMSE	2.001	0.776	1.029	1.453	1.146	1.219	1.489
	$n = 200$	IBIAS2	1.245	0.319	0.509	0.886	0.605	0.679	0.890
		IVAR	0.134	0.127	0.186	0.089	0.156	0.129	0.092
		IMSE	1.378	0.446	0.695	0.975	0.761	0.808	0.982

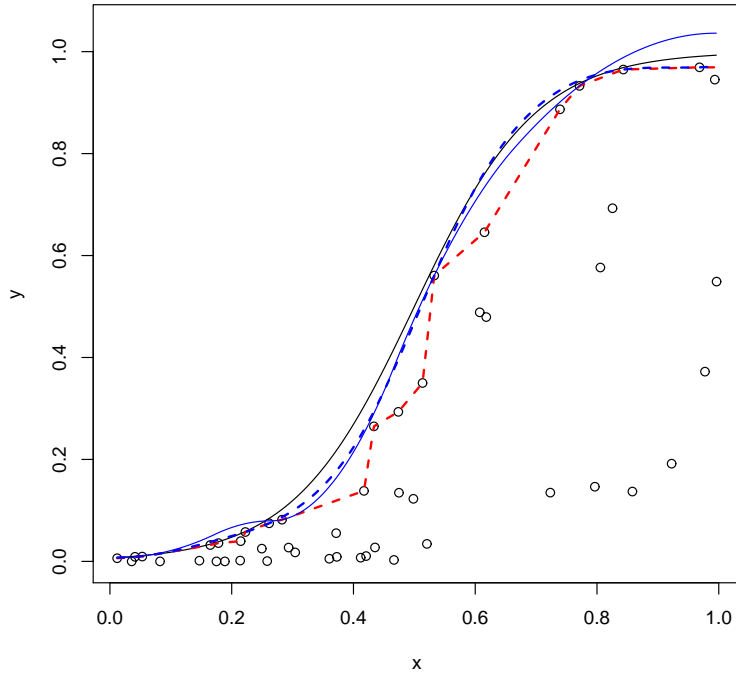


Figure 3: When  $n = 50$  and  $\beta = 0.5$ , the true frontier function ( $\varphi_c$ , black) and its three estimates: LFDH (dotted red), QS-B (solid blue) and CS-B (dotted blue)

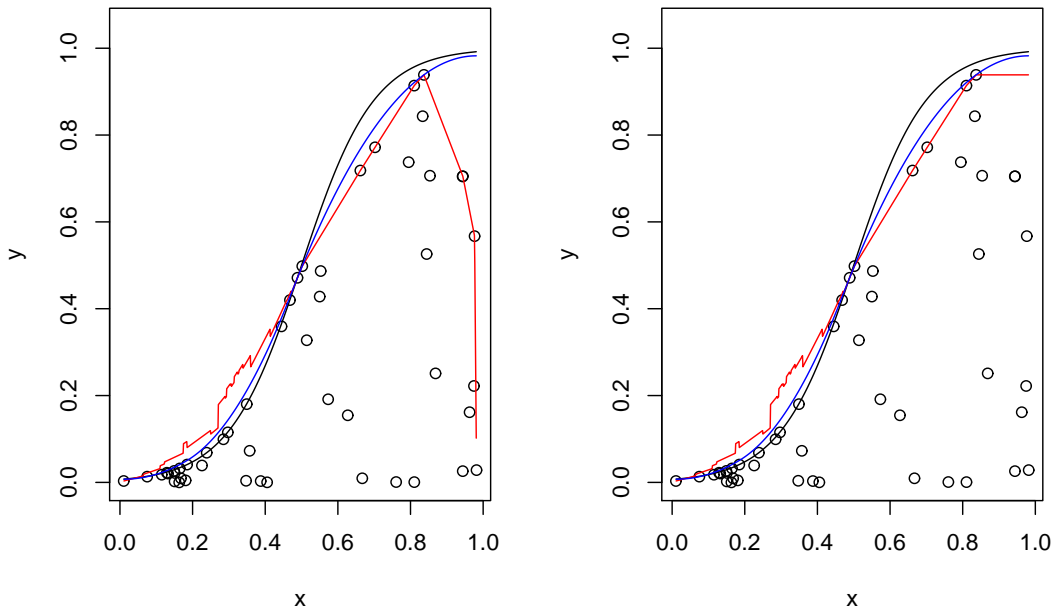


Figure 4: The problem of the local linear frontier estimator (left panel) and illustration of its improved version (right panel): LL (red curve), QS-B (blue curve) and the true frontier (black curve)

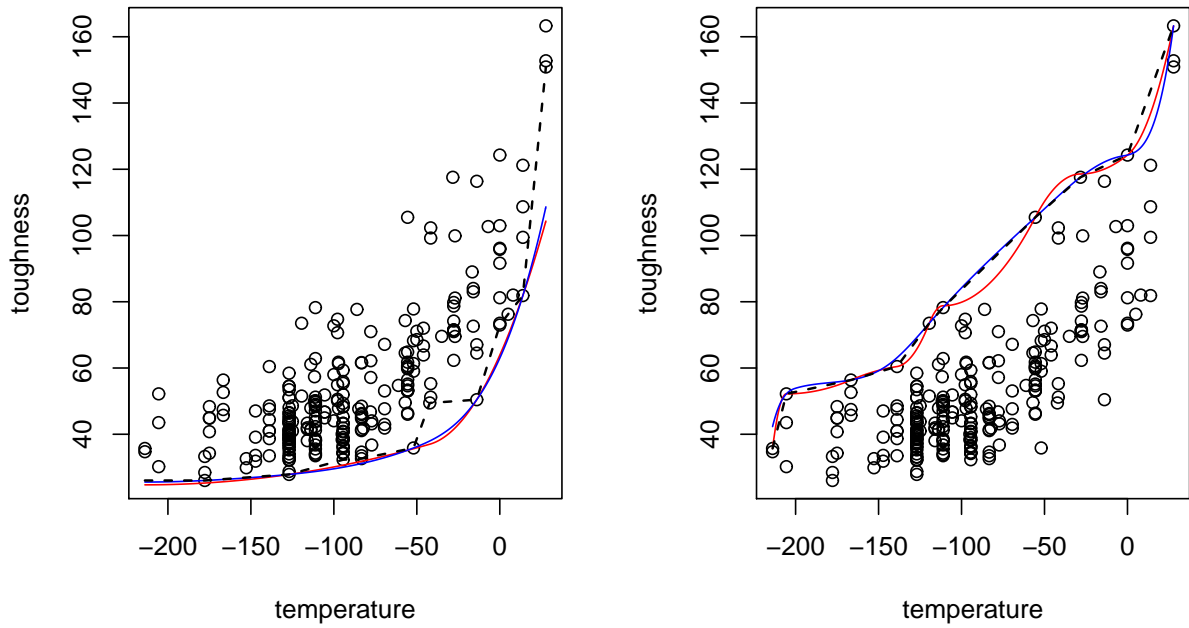


Figure 5: Scatterplot of the 254 nuclear reactors data, with three frontier estimates: LFDH (dotted black), QS-B (solid red) and CS-B (solid blue). From left to right, the estimates for the lower and the upper support boundaries.

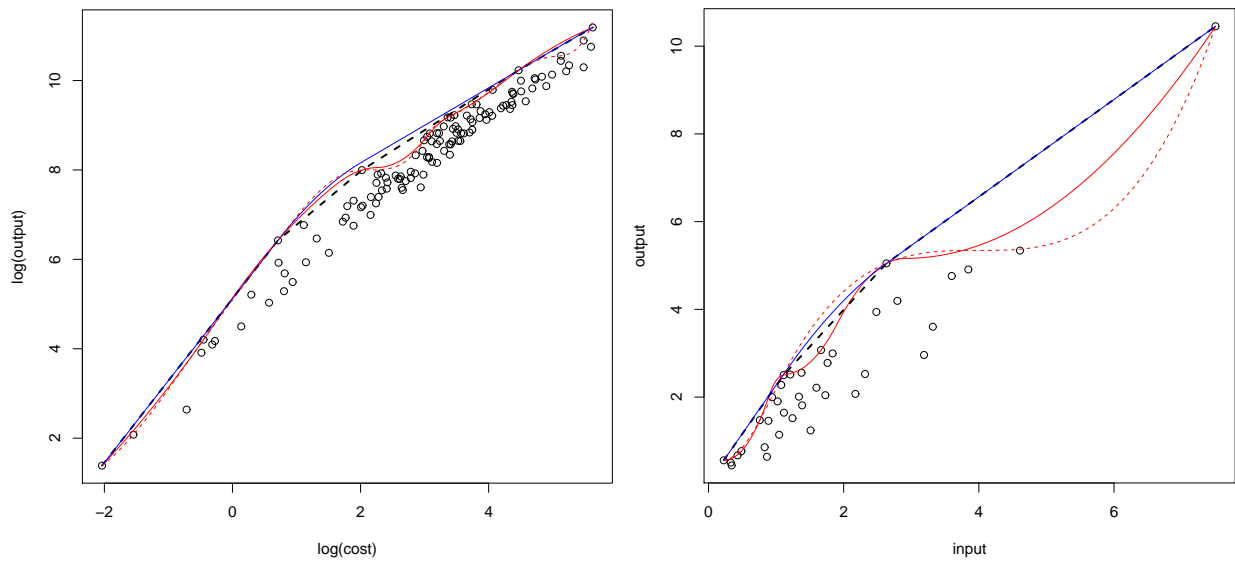


Figure 6: Scatterplots of the 123 American electric utility companies' data (left) and the 37 European Air Controllers' data (right) with three estimates of the efficient extremity in each plot: DEA (dotted black), QS-B (solid red), CS-B (dotted red) and QS-C (solid blue)