Robust frontier estimation from noisy data: a Tikhonov regularization approach

Abdelaati Daouia[a], Jean-Pierre Florens[a], Léopold Simar[a,b,*]

[a] *Toulouse School of Economics, University of Toulouse Capitole, France.*
[b] *Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Belgium.*

Abstract

In stochastic frontier models, the regression function defines the production frontier and the regression errors are assumed to be composite. The actually observed outputs are assumed to be contaminated by a stochastic noise. The additive regression errors are composed from this noise term and the one-sided inefficiency term. The aim is to construct a robust nonparametric estimator for the production function. The main tool is a robust concept of partial, expected maximum production frontier, defined as a special probability-weighted moment. In contrast to the deterministic one-sided error model where robust partial frontier modeling is fruitful, the composite error problem requires a substantial different treatment based on deconvolution techniques. To ensure the identifiability of the model, it is sufficient to assume an independent Gaussian noise. In doing so, the frontier estimation necessitates the computation of a survival function estimator from an ill-posed equation. A Tikhonov regularized solution is constructed and nonparametric frontier estimation is performed. The asymptotic properties of the obtained survival function and frontier estimators are established. Practical guidelines to effect the necessary computations are described via a simulated example. The usefulness of the approach is discussed through two concrete data sets from the sector of Delivery Services.

Keywords: Deconvolution, Nonparametric estimation, Probability-weighted moment, Production function, Robustness, Stochastic frontier, Tikhonov regularization.

JEL: C1, C14, C13, C49

1. Introduction

1.1. Deterministic frontier estimation

In deterministic nonparametric frontier models, the data

$$Y_j = \varphi(X_j) - U_j, \quad j = 1 \ldots, n,$$

are observed, where $X_j \in \mathbb{R}_+^p$ represents a vector of input factors (*e.g.*, labor, energy, capital) used to produce an output $Y_j \in \mathbb{R}_+$ in a certain firm j, with $\varphi(\cdot)$ being the production function and $U_j \geqslant 0$ being the inefficiency term. In contrast to standard regression models, the observation errors (U_j) are assumed to be one-sided instead of centred, and hence the regression function φ describes some frontier or boundary curve.

*Corresponding author

Email addresses: abdelaati.daouia@tse-fr.eu (Abdelaati Daouia), jean-pierre.florens@tse-fr.eu (Jean-Pierre Florens), leopold.simar@uclouvain.be (Léopold Simar)

For a fixed level of inputs-usage $x \in \mathbb{R}_+^p$, the frontier point $\varphi(x)$ gives the achievable maximum output. A closed form expression of $\varphi(x)$ has been introduced by Cazals *et al.* (2002) in terms of the non-standard conditional distribution of Y given $X \leqslant x$. If $(\Omega, \mathcal{A}, \mathbb{P})$ denotes the probability space on which the random vector $(X, Y) \in \mathbb{R}_+^{p+1}$ is defined and

$$F_{Y|X}(y|x) = \mathbb{P}(Y \leqslant y \,|\, X \leqslant x) \quad \text{with} \quad F_X(x) := \mathbb{P}(X \leqslant x) > 0,$$

then $\varphi(x)$ can be characterized as the conditional right endpoint

$$\varphi(x) = \sup\{y \geqslant 0 \,|\, F_{Y|X}(y|x) < 1\}. \tag{1}$$

This frontier function is isotonic nondecreasing in x. It is actually the lowest monotone function which envelops the upper extremity, say $\varphi^u(x)$, of the support of (X, Y) at $X = x$. Generally speaking, $\varphi(x)$ equals $\sup_{x' \leqslant x} \varphi^u(x')$. Production econometrics leads to the natural assumption that the true upper support boundary $\varphi^u(x)$ is itself nondecreasing, and hence it coincides with $\varphi(x)$. However, one can use only the observations in a local strip around x to estimate $\varphi^u(x)$ because of its local nature. Consideration of $\varphi(x)$ is advantageous as it offers estimation at a faster rate. By replacing $F_{Y|X}(y|x)$ in (1) with its empirical version

$$\widehat{F}_{Y|X}(y|x) = \sum_{i=1}^{n} \mathbb{1}(X_i \leqslant x, Y_i \leqslant y) / \sum_{i=1}^{n} \mathbb{1}(X_i \leqslant x),$$

where $\mathbb{1}(\cdot)$ stands for the indicator function, Cazals *et al.* (2002) recover the pioneering Free Disposal Hull (FDH) estimator of Deprins *et al.* (1984):

$$\widehat{\varphi}(x) = \sup\{y \geqslant 0 \,|\, \widehat{F}_{Y|X}(y|x) < 1\} = \max_{i: X_i \leqslant x} Y_i. \tag{2}$$

Its full asymptotic theory has been elucidated in a general setting from the perspective of extreme value theory in Daouia *et al.* (2010, 2014). The major drawback of $\widehat{\varphi}(x)$ is its lack of robustness to outliers. To remedy this vexing defect, Cazals *et al.* (2002) have proposed to first estimate a *partial* frontier well inside the joint support of (X, Y) and then to shift the obtained estimate to the true *full* support boundary. Their main tool is the *partial* production function of order $m \in \{1, 2, \ldots\}$ defined as

$$\psi_m(x) = \mathbb{E}\big[\max(Y^1, \ldots, Y^m) | X \leqslant x\big] = \int_0^\infty \left(1 - [F_{Y|X}(y|x)]^m\right) dy,$$

where $(Y^1, \ldots, Y^m)$ are i.i.d. random variables generated by the conditional distribution of Y given $X \leqslant x$. The quantity $\psi_m(x)$ gives the expected maximum achievable output among a fixed number of m firms using less inputs than x. It converges to the true production function $\varphi(x)$ as $m \to \infty$. Likewise, its empirical counterpart

$$\widehat{\psi}_m(x) = \int_0^\infty \left(1 - [\widehat{F}_{Y|X}(y|x)]^m\right) dy = \widehat{\varphi}(x) - \int_0^{\widehat{\varphi}(x)} [\widehat{F}_{Y|X}(y|x)]^m \, dy$$

achieves the envelopment FDH frontier $\widehat{\varphi}(x)$ as $m \to \infty$. Then, the use of $\widehat{\psi}_m(x)$ instead of $\widehat{\varphi}(x)$, for an appropriate large value of m, could help the practitioners to achieve their objective of 'robustification'. Yet, this device is not without disadvantages. One of the main criticisms on the partial frontier $\psi_m(x)$ and its estimate $\widehat{\psi}_m(x)$ is their failure to fulfill the monotonicity property of the efficient full frontier $\varphi(x)$. This property, referred to as non-negative marginal productivity, is a minimal requirement from the point of view of the axiomatic production theory. Another desirable property of any benchmark partial frontier, as argued by Wheelock and Wilson (2008), is to closely parallel the true production frontier. However, due to the conditioning by $X \leqslant x$, both $\psi_m(x)$ and $\widehat{\psi}_m(x)$ diverge from $\varphi(x)$ as the input level x increases. Also, for large values of m, the estimates $\widehat{\psi}_m(x)$ coincide with the non-robust FDH frontier $\widehat{\varphi}(x)$ for small values of x. In all of these aspects, Daouia *et al.* (2018) have recently suggested a better alternative by transforming first the $(p+1)$-dimensional vector (X, Y) and its independent copies (X_i, Y_i) into the dimensionless random variables

$$Y^x = Y\,\mathbb{1}(X \leqslant x) \quad \text{and} \quad Y_i^x = Y_i\,\mathbb{1}(X_i \leqslant x)\,, \quad i = 1, \ldots, n, \tag{3}$$

whose common *unconditional* distribution function $F_{Y^x}(\cdot)$ is closely related to $F_{Y|X}(\cdot|x)$:

$$F_{Y^x}(y) = \big\{1 - F_X(x)[1 - F_{Y|X}(y|x)]\big\}\,\mathbb{1}(y \geqslant 0). \tag{4}$$

A nice property of these transformed univariate random variables lies in the fact that

$$\begin{aligned} \varphi(x) &\equiv \sup\{y \geqslant 0 \,|\, F_{Y^x}(y) < 1\}, \\ \widehat{\varphi}(x) &\equiv \sup\{y \geqslant 0 \,|\, \widehat{F}_{Y^x}(y) < 1\} = \max(Y_1^x, \ldots, Y_n^x), \end{aligned} \tag{5}$$

where $\widehat{F}_{Y^x}(y) = (1/n)\sum_{i=1}^n \mathbb{1}(Y_i^x \leqslant y)$. Then, Daouia *et al.* (2018) define the new concept of partial m-frontier

$$\varphi_m(x) = \mathbb{E}\big[\max(Y_1^x, \ldots, Y_m^x)\big] = \int_0^\infty \big(1 - [F_{Y^x}(y)]^m\big)\,dy, \tag{6}$$

as the expected maximum of m independent copies of Y^x. This *anchor* production function is identical to the expectation of the FDH estimator based on the m-tuple of observations $Y_1^x, \ldots, Y_m^x$. It achieves the optimal frontier $\varphi(x)$ when $m \to \infty$. Its empirical version

$$\widehat{\varphi}_m(x) = \int_0^\infty \big(1 - [\widehat{F}_{Y^x}(y)]^m\big)\,dy = \widehat{\varphi}(x) - \int_0^{\widehat{\varphi}(x)} [\widehat{F}_{Y^x}(y)]^m\,dy \tag{7}$$

converges to the FDH frontier $\widehat{\varphi}(x)$ as $m \to \infty$. Both the *unconditional* expected maximum output frontiers $\varphi_m(x)$ and their estimators $\widehat{\varphi}_m(x)$ share the fundamental property of monotonicity. Their superiority over the conditional versions $\psi_m(x)$ and $\widehat{\psi}_m(x)$ was also established from a robustness theory point of view. Furthermore, they do not suffer from border and divergence effects for small or large levels of inputs. Also, the selection problem of an appropriate trimming number m in $\widehat{\varphi}_m(x)$ tends to be easier than in $\widehat{\psi}_m(x)$. The asymptotic distributional properties of both partial and full frontier estimators have been derived as well under the deterministic frontier model $Y_j = \varphi(X_j) - U_j$, $j = 1 \ldots, n$.

1.2. Stochastic frontier estimation

For a practitioner it would be more realistic to assume that the outputs are contaminated by an additive stochastic error. That is, the actually observed outputs are

$$Z_j = Y_j + \varepsilon_j, \quad j = 1 \ldots, n, \tag{8}$$

instead of Y_j, where ε_j denotes a stochastic noise. This results in the composite-error model

$$Z_j = \varphi(X_j) - U_j + \varepsilon_j, \quad j = 1 \ldots, n. \tag{9}$$

The issue of frontier estimation in such a model goes back to the works of Aigner *et al.* (1977) and Meeusen and van den Broeck (1977). Typically, it is assumed that φ has a parametric structure (like Cobb-Douglas or translog), ε_j is normally distributed and U_j is generated by some specified parametric one-sided distribution (often Half-normal, exponential, truncated normal or gamma). Parametric techniques of estimation include modified least-squares and maximum likelihood methods, see for instance Greene (2008) for a survey. More recently some attempts have been proposed to relax the parametric restrictions. Of course, a fully nonparametric frontier model allowing convolution of inefficiency and a two-sided noise is not identifiable as shown by Hall and Simar (2002). Some amount of structure is then required to allow identification. One approach is to leave only φ unspecified, while specifying a parametric density for inefficiency and an independent Gaussian noise, both being homoskedastic. This semi-parametric approach has been investigated by Fan *et al.* (1996). It is appealing but still very restrictive: both the homoskedasticity assumption and the choice of a parametric density for U may be problematic and could introduce misspecification and statistical inconsistency. Alternatively, Kumbhakar *et al.* (2007), Simar and Zelenyuk (2011) and Simar *et al.* (2017) suggest the use of local maximum likelihood or least-squares techniques, allowing heteroskasticity and functional forms for the local parameters. The main drawbacks of these approaches are the computational burden to select optimal bandwidths and the fact that they still rely on local parametric specifications for the distribution of U_j.

In this paper we adopt a different strategy based on deconvolution techniques. In the standard deconvolution problem (8), the observed data Z_j are used to estimate the unknown density of the latent signal Y_j. Most of the literature in this area supposes that the noise ε_j has a known density (*e.g.*, Gaussian with known variance) and presents kernel estimation methods such as, for instance, Carroll and Hall (1988), Stefanski and Carroll (1990), and Fan (1991a, 1991b). Meister (2006) deals with the density estimation problem based on a normally distributed error ε_j whose variance is unknown. These ideas have been applied by Horrace and Parmeter (2011) to the case of an unknown homoskedastic inefficiency term and independent Gaussian noise with unknown variance, but the frontier remains fully specified by a parametric model. More recently in Kneip *et al.* (2015), the unspecified inefficiency distribution is estimated by simple histograms, then a simultaneous estimation of the boundary and the variance of the Gaussian noise is performed via a penalized likelihood method. This procedure provides, however, estimators with disappointing rates of convergence.

The alternative approach that we propose to address the deconvolution problem is by applying a Tikhonov regularization technique [see, *e.g.*, Engl *et al.* (2000)] in conjunction with the 'robustified' concept of *unconditional* expected maximum production frontiers $\{\varphi_m(x)\}$ described in (6). For a prespecified predictor value of interest x, we only assume that the density of ε given $X \leqslant x$ is known (*e.g.*, Gaussian with known variance). Before estimating the frontier point $\varphi(x)$ under the composite-error model (9), a main tool is to first estimate the regular partial frontier $\varphi_m(x)$ which tends to $\varphi(x)$ as the trimming order $m \to \infty$. In doing so, this necessitates the computation of an estimator for the distribution function $F_{Y^x}(\cdot)$ in (4) from an ill-posed equation. A Tikhonov regularized solution is constructed in Section 2 and its asymptotic properties are derived in Section 3, including the rate of convergence of quadratic risk and the asymptotic normality of scalar products. Section 4 describes how to estimate in a second stage both $\varphi_m(x)$ and $\varphi(x)$. We establish the asymptotic normality for the $\varphi_m(x)$ estimator and the rate of convergence of quadratic risk for the $\varphi(x)$ estimator. In Section 5 we first indicate how to implement the procedure with a simulated example, then we estimate the production frontier from two concrete datasets in the sector of Delivery Services. Section 6 concludes and the Appendix collects the proofs.

2. Deconvolution and Tikhonov regularization

Throughout this paper we consider the stochastic model described in (9), namely

$$Y_j = \varphi(X_j) - U_j, \quad j = 1 \ldots, n,$$

where $(X_j, Y_j) \in \mathbb{R}_+^p \times \mathbb{R}_+$ are i.i.d. random vectors, with $\varphi(\cdot)$ being the unknown frontier function and $U_j \geqslant 0$ being the inefficiency term, but the actually observed outputs are $Z_j = Y_j + \varepsilon_j$ instead of Y_j, where ε_j denotes a stochastic noise satisfying the condition that

(C.1) ε_j is independent of Y_j given $X_j \leqslant x$,

for a prespecified level of inputs x such that $F_X(x) > 0$. We also assume that

(C.2) the density of ε_j given $X_j \leqslant x$ is fully known.

Our objective is to first estimate the distribution function $F_{Y^x}(\cdot)$ of the dimensionless variable Y^x defined in (3), or equivalently, its survival function $S_{Y^x} := 1 - F_{Y^x}$ from the noisy data $\{(X_j, Z_j)|j = 1, \ldots, n\}$, and then to use the corresponding plug-in frontier estimators $\widehat{\varphi}(x)$ and $\widehat{\varphi}_m(x)$ described in (5) and (7).

2.1. Deconvolution problem

Let $S_{Z^x}(\cdot)$ and $S_{\varepsilon^x}(\cdot)$ denote the survival functions of the random variables

$$Z^x = Z\,\mathbb{1}(X \leqslant x) \quad \text{and} \quad \varepsilon^x = \varepsilon\,\mathbb{1}(X \leqslant x).$$

It is easily seen that, for all $z \in \mathbb{R}$,

$$S_{Z^x}(z) - S_{\varepsilon^x}(z) = \left[S_{Z|X}(z|x) - S_{\varepsilon|X}(z|x)\right] F_X(x),$$

where $S_{Z|X}(z|x) := \mathbb{P}(Z > z|X \leqslant x)$ and $S_{\varepsilon|X}(z|x) := \mathbb{P}(\varepsilon > z|X \leqslant x)$. On the other hand, since $S_{Y|X}(y|x) = 1$ for all $y \leqslant 0$, simple calculations lead to the following equation defining the convoluted conditional survivor function of Z, for any $z \in \mathbb{R}$,

$$S_{Z|X}(z|x) - S_{\varepsilon|X}(z|x) = \int_{-\infty}^{z} S_{Y|X}(z-\epsilon|x) \cdot f_{\varepsilon|X}(\epsilon|x)\, d\epsilon,$$

with $S_{Y|X}(y|x) := 1 - F_{Y|X}(y|x)$ and $f_{\varepsilon|X}(\cdot|x)$ being the density function of ε given $X \leqslant x$. It follows that, for all $z \in \mathbb{R}$,

$$\begin{aligned} S_{Z^x}(z) - S_{\varepsilon^x}(z) &= \int_0^{\infty} S_{Y|X}(y|x) \cdot F_X(x) \cdot f_{\varepsilon|X}(z-y|x)\, dy \\ &= \int_0^{\infty} S_{Y^x}(y) \cdot f_{\varepsilon|X}(z-y|x)\, dy. \end{aligned} \tag{10}$$

By the assumption **(C.2)** that the density $f_{\varepsilon|X}(\cdot|x)$ is fully known, the problem reduces to solving the integral equation (10) in $S_{Y^x}(\cdot)$ in terms of $S_{Z^x}(z) - S_{\varepsilon^x}(z)$. Given that

$$S_{\varepsilon^x}(z) = \begin{cases} S_{\varepsilon|X}(z|x)\, F_X(x) & \text{if} \quad z \geqslant 0 \\ S_{\varepsilon|X}(z|x)\, F_X(x) + 1 - F_X(x) & \text{if} \quad z < 0, \end{cases}$$

our estimator would then be obtained by plugging the empirical $\widehat{S}_{n,Z^x}(\cdot)$ and $\widehat{S}_{n,\varepsilon^x}(\cdot)$ survivors given by

$$\begin{aligned} \widehat{S}_{n,Z^x}(z) &= \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}(Z_j^x > z), \quad \text{where} \quad Z_j^x := Z_j\, \mathbb{1}(X_j \leqslant x), \quad \text{for each} \quad j = 1, \ldots, n, \\ \widehat{S}_{n,\varepsilon^x}(z) &= S_{\varepsilon|X}(z|x)\, \widehat{F}_X(x) + [1 - \widehat{F}_X(x)]\, \mathbb{1}(z < 0), \quad \text{with} \quad \widehat{F}_X(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}(X_j \leqslant x). \end{aligned}$$

Technically we have actually to solve the integral equation (10) in $S_{Y^x}(\cdot)$ in terms of $\widehat{S}_{n,Z^x}(\cdot)$ and $\widehat{S}_{n,\varepsilon^x}(\cdot)$. This is a deconvolution problem which is well known to be an ill-posed inverse problem. One way to see this is by trying to solve (10) via characteristic functions. Denote by $\psi_{Y^x}(\cdot)$ and $\psi_{Y|X}(\cdot|x)$ the characteristic functions for Y^x and Y given $X \leqslant x$, respectively. We have

$$\psi_{Y^x}(t) = -\int_{-\infty}^{\infty} e^{ity}\, dS_{Y^x}(y) \equiv F_X(x) \cdot \psi_{Y|X}(t|x),$$

where $i^2 = -1$. If $\psi_{Y^x}(t)$ would be known, then by Fourier inversion [see, *e.g.*, Lukacs (1960)], $S_{Y^x}(y)$ could be computed as

$$S_{Y^x}(y) = 1 - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1 - e^{-ity}}{it}\, \psi_{Y^x}(t)\, dt.$$

Thus what we need is to determine $\psi_{Y^x}(t)$. Since Y is independent of ε given $X \leqslant x$, we directly get

$$\psi_{Y|X}(t|x) = \frac{\psi_{Z|X}(t|x)}{\psi_{\varepsilon|X}(t|x)},$$

where $\psi_{Z|X}(\cdot|x)$ and $\psi_{\varepsilon|X}(\cdot|x)$ are the characteristic functions of Z and ε given $X \leqslant x$, respectively. This leads to

$$\psi_{Y^x}(t) \equiv F_X(x) \cdot \frac{\psi_{Z|X}(t|x)}{\psi_{\varepsilon|X}(t|x)},$$

with $\psi_{\varepsilon|X}(\cdot|x)$ being known. Therefore good estimates of $\psi_{Y^x}(t)$ could be obtained from accurate estimates of $\psi_{Z|X}(t|x)$ and the empirical marginal distribution function $\widehat{F}_X(x)$ of X. However, for large values of $|t|$, $\psi_{\varepsilon|X}(t|x)$ converges to zero making the estimation of $\psi_{Y^x}(t)$ notoriously difficult even if good estimates of $\psi_{Z|X}(t|x)$ are available.

In order to regularize this ill-posed problem, we can use for instance truncation methods or deconvoluted kernel methods [see, *e.g.*, Fan (1991a) for details]. Yet, given that the support of $S_{Y^x}(\cdot)$ is bounded with upper endpoint $\varphi(x)$, it is more natural to solve the equation (10) by staying in the space of survivor functions. Of course this remains an ill-posed inverse problem as described below, and so some regularization will be needed. Built on the ideas of Hall and Meister (2007) and Carrasco and Florens (2009), we propose in the next section to pose the deconvolution problem in the usual Tikhonov regularization framework. The main advantage of this procedure is its computational expedience. We refer to Van Rooij and Ruymgaart (1999), and the references therein, for more thorough discussion of the rationale for this elegant device.

2.2. Tikhonov regularization

Our main objective here is to estimate the unconditional survivor function S_{Y^x}. Survivor functions are typically assumed to belong to some Hilbert space of square-integrable functions with respect to appropriate weight functions:

(H.1) $S_{Y^x}(\cdot) \in \mathcal{E} := L^2\big([0,\infty)\big)$,

where $\mathcal{E}$ is endowed with the norm $||g|| = \big(\int_0^\infty g^2(u)du\big)^{1/2} < \infty$, for all $g \in \mathcal{E}$. Note that this condition is explicitly satisfied in our setup, since $||S_{Y^x}(\cdot)||^2 \leqslant \int_0^\infty S_{Y^x}(u)du \leqslant \varphi(x)$. We shall need also the following condition on the noise density:

(H.2) $f_{\varepsilon|X}(\cdot|x)$ is square integrable, *i.e.*, $\int_{-\infty}^\infty f^2_{\varepsilon|X}(u|x)\,du < \infty$.

Consider now the Hilbert space $\mathcal{F} = L^2(\mathbb{R}, w)$, where w stands for a probability measure. It is easily seen that $S_{Z^x}(\cdot) - S_{\varepsilon^x}(\cdot) \in \mathcal{F}$ by using (10) and applying the fact that

$$\int_{-\infty}^{\infty}\left[\int_0^\infty S_{Y^x}(y)\cdot f_{\varepsilon|X}(z-y|x)\,dy\right]^2 w(z)\,dz \leqslant \int_{-\infty}^{\infty}\int_0^\infty S^2_{Y^x}(y)\, f^2_{\varepsilon|X}(z-y|x)\,w(z)\,dy\,dz \tag{11}$$

in conjunction with **(H.2)**, which also implies that

$$\int_{-\infty}^{\infty}\int_0^\infty f^2_{\varepsilon|X}(z-y|x)\,w(z)\,dy\,dz < \infty. \tag{12}$$

Hence, the basic integral equation (10) involves the operator $K : \mathcal{E} \to \mathcal{F}$ defined as

$$KS_{Y^x} = S_{Z^x} - S_{\varepsilon^x}, \tag{13}$$

or equivalently

$$\big(KS_{Y^x}\big)(z) = \int_0^\infty S_{Y^x}(y)\cdot f_{\varepsilon|X}(z-y|x)\,dy. \tag{14}$$

This is an integral operator with kernel $f_{\varepsilon|X}(z-y|x) \in L^2\big(\mathbb{R}\times\mathbb{R}, \mathbb{1}(x \geqslant 0)w(z)\big)$, in view of (12). As such, K is a Hilbert-Schmidt integral operator allowing a discrete Singular Value Decomposition (SVD), which is also compact [see, *e.g.*, Kreiss (1999) for mathematical details and Carrasco *et al.* (2007) for econometrics considerations]. We also note that this operator is injective, since $\phi \in \mathcal{E}$ such that $K\phi = 0$ implies $\phi = 0$. This follows immediately when looking to (14).

Let now K^* be the adjoint operator of K. By definition, for any $g \in \mathcal{E}$ and $\psi \in \mathcal{F}$, the scalar products $\langle Kg, \psi\rangle_{\mathcal{F}}$ in $\mathcal{F}$ and $\langle g, K^*\psi\rangle_{\mathcal{E}}$ in $\mathcal{E}$ are equal, so we have

$$\begin{aligned}\langle KS_{Y^x}, \psi\rangle_{\mathcal{F}} &= \int_{-\infty}^{\infty} \psi(z)\left\{\int_0^{\infty} S_{Y^x}(y)\cdot f_{\varepsilon|X}(z-y|x)\,dy\right\} w(z)dz \\ &= \int_0^{\infty} S_{Y^x}(y)\left\{\int_{-\infty}^{\infty} \psi(z)\, f_{\varepsilon|X}(z-y|x)\, w(z)dz\right\} dy \\ &= \langle S_{Y^x}, K^*\psi\rangle_{\mathcal{E}}.\end{aligned}$$

This allows to identify the adjoint operator as

$$\big(K^*\psi\big)(y) = \int_{-\infty}^{\infty} \psi(z)\, f_{\varepsilon|X}(z-y|x)\, w(z)dz. \tag{15}$$

It is well known that the equation (13) is ill-posed in the sense that the problem

$$S_{Y^x} = \underset{S\in\mathcal{E}, S:\mathbb{R}\to[0,1]}{\operatorname{argmin}} ||KS - (S_{Z^x} - S_{\varepsilon^x})||^2 \tag{16}$$

has not a well-defined solution. The rationale behind this ill-posedness can be explained by making use of the discrete SVD of K. We know that there exists a sequence $(\lambda_j, \phi_j, \zeta_j)_{j\geqslant 1}$, with $\lambda_j \in \mathbb{R}_+$ and $\{\phi_j\}_j$ (resp. $\{\zeta_j\}_j$) being an orthonormal basis of $\mathcal{E}$ (resp. of $\mathcal{F}$), such that for all j,

$$K\phi_j = \lambda_j\zeta_j, \quad K^*\zeta_j = \lambda_j\phi_j.$$

Since K is compact, the sequence of λ_j's may be ranked so that $\lambda_1 \geqslant \lambda_2 \geqslant \ldots > 0$, with zero being an accumulation point of the λ_j's. Note that $K^*K\phi_j = \lambda_j^2\phi_j$, which indicates that the inverse of the operator K^*K will be unstable as many of the λ_j are near zero. Accordingly, there is no hope to solve the normal equations $K^*KS_{Y^x} = K^*(S_{Z^x} - S_{\varepsilon^x})$ [first order conditions] coming from the problem (16). This also appears when writing equivalently

$$S_{Z^x} - S_{\varepsilon^x} = \sum_j \langle S_{Z^x} - S_{\varepsilon^x}, \zeta_j\rangle\zeta_j, \tag{17}$$

$$S_{Y^x} = \sum_j \langle S_{Y^x}, \phi_j\rangle\phi_j, \tag{18}$$

which leads to $KS_{Y^x} = \sum_j \lambda_j\langle S_{Y^x}, \phi_j\rangle\zeta_j$. Solving (16) would yield the identification of S_{Y^x} via the equations

$$\langle S_{Y^x}, \phi_j\rangle = \frac{\langle S_{Z^x} - S_{\varepsilon^x}, \zeta_j\rangle}{\lambda_j},$$

and then we would obtain S_{Y^x} by (18). Again, this is rather unstable when λ_j are near zero. Instead, the Tikhonov regularization consists in replacing the ratios $1/\lambda_j$ in the latter equation by $\lambda_j/(\alpha + \lambda_j^2)$ for some $\alpha > 0$. It is not hard to verify that this corresponds to replace the least squares problem (16) by the following regularized version

$$S^{\alpha}_{Y^x} = \operatorname*{argmin}_{S\in\mathcal{E}, S:\mathbb{R}\to[0,1]} \left\{||KS - (S_{Z^x} - S_{\varepsilon^x})||^2 + \alpha||S||^2\right\},$$

where the parameter $\alpha > 0$ is actually introduced in order to regularize the behavior of $(K^*K)^{-1}$. The solution is given by the normal equations

$$\alpha S^{\alpha}_{Y^x} + K^*KS^{\alpha}_{Y^x} = K^*(S_{Z^x} - S_{\varepsilon^x}),$$

so that the regularized solution has the usual form

$$S^{\alpha}_{Y^x} = (\alpha I + K^*K)^{-1}K^*(S_{Z^x} - S_{\varepsilon^x}). \tag{19}$$

This motivates us to estimate S_{Y^x} by the empirical version $\widehat{S}^{\alpha}_{Y^x} : \mathbb{R} \to [0, 1]$ defined as

$$\begin{aligned}\widehat{S}^{\alpha}_{Y^x} &= \operatorname*{argmin}_{S\in\mathcal{E}, S:\mathbb{R}\to[0,1]} \left\{||KS - (\widehat{S}_{n,Z^x} - \widehat{S}_{n,\varepsilon^x})||^2 + \alpha||S||^2\right\} \\ &= (\alpha I + K^*K)^{-1}K^*(\widehat{S}_{n,Z^x} - \widehat{S}_{n,\varepsilon^x}). \end{aligned} \tag{20}$$

The practical computation of this estimator requires first the characterization of K^*K. For any $g \in \mathcal{E}$, we have

$$(Kg)(z) = \int_0^{\infty} g(\xi) f_{\varepsilon|X}(z - \xi|x)\, d\xi,$$

and hence

$$(K^*Kg)(y) = \int_{-\infty}^{\infty} f_{\varepsilon|X}(z - y|x) \left\{\int_0^{\infty} g(\xi) f_{\varepsilon|X}(z - \xi|x)\, d\xi\right\} w(z)dz = \int_0^{\infty} g(\xi)\, c(y, \xi)\, d\xi,$$

where

$$c(y, \xi) := \int_{-\infty}^{\infty} f_{\varepsilon|X}(z - y|x) f_{\varepsilon|X}(z - \xi|x) w(z)dz \tag{21}$$

defines the kernel of K^*K, which is symmetric in both y and ξ. The normal equation to be solved can then be formulated, for any $y \geqslant 0$, as

$$\alpha\widehat{S}^{\alpha}_{Y^x}(y) + \int_0^{\infty} \widehat{S}^{\alpha}_{Y^x}(\xi)\, c(y, \xi)\, d\xi = b(y|x), \tag{22}$$

with

$$b(y|x) = \int_{-\infty}^{\infty} (\widehat{S}_{n,Z^x}(z) - \widehat{S}_{n,\varepsilon^x}(z)) f_{\varepsilon|X}(z - y|x) w(z)dz. \tag{23}$$

A numerical solution to this equation can be achieved in practice via discretization. Consider a regular grid of fixed values y_j for $j = 1, \ldots, k$, with constant spacing Δ_y, covering the range of Y. We shall need to choose a value of k sufficiently large so that the numerical error due to discretization has lower order than the statistical error. Equation (22) can then be formulated into the discrete version

$$\alpha S_i + \Delta_y \sum_{j=1}^{k} S_j\, C_{ij} = b_i, \quad i = 1, \ldots, k,$$

where $S_i = \widehat{S}^{\alpha}_{Y^x}(y_i)$, $b_i = b(y_i|x)$ and $C_{ij} = c(y_i, y_j)$. Equivalently, this translates in terms of simplified matrix notations to

$$\alpha \boldsymbol{S} + \Delta_y \boldsymbol{C}\boldsymbol{S} = \boldsymbol{b}$$

whose exact solution is

$$\boldsymbol{S} = (\alpha \boldsymbol{I_k} + \Delta_y \boldsymbol{C})^{-1}\boldsymbol{b}, \tag{24}$$

with $\boldsymbol{S} = (S_1, \ldots, S_k)^T$, $\boldsymbol{b} = (b_1, \ldots, b_k)^T$, and $\boldsymbol{C}$ being the $k \times k$ matrix of coefficients C_{ij}. Thus, the regularized estimator of the survivor function is very easy and fast to compute over the chosen grid of values for Y. Note that the evaluation of $\boldsymbol{b}$ and $\boldsymbol{C}$ only requires the calculation of univariate integrals. It should also be clear that the regularization comes from the Tikhonov step and not from the projection on the finite set of values for Y that we can choose as large as we want.

3. Properties of the estimated survival function

This section presents (i) an indicative rate of convergence of the quadratic risk for the regularized estimator $\widehat{S}^{\alpha}_{Y^x}$, (ii) sufficient conditions for its asymptotic normality, and (iii) practical guidelines to select the regularization parameter α via an iterated Tikhonov technique.

3.1. Quadratic risk $\mathbb{E}(||\widehat{S}^{\alpha}_{Y^x} - S_{Y^x}||^2)$

To derive the rate of convergence of the quadratic risk for the regularized estimator $\widehat{S}^{\alpha}_{Y^x}(\cdot)$, we shall need the extra "source regularity assumption" on the signal $S_{Y^x}(\cdot)$, which is standard in the Tikhonov regularization framework:

(H.3) For some $\beta \in (0, 2]$, $S_{Y^x} \in \text{Range}\big(K^*K\big)^{\beta/2}$.

This means that $S_{Y^x} = \big(K^*K\big)^{\beta/2}\delta$, for a certain function $\delta \in \mathcal{E}$. This Hölder type condition is needed to control $S^{\alpha}_{Y^x} - S_{Y^x}$, the bias introduced by the regularization. We give in Appendix A.2 two examples illustrating **(H.3)**.

Theorem 1. *Let $x \in \mathbb{R}^p_+$ be fixed such that $0 < F_X(x) < 1$. Under Assumptions* **(H.1)**, **(H.2)** *and* **(H.3)**, *as $n \to \infty$ with $\alpha = O\big(n^{-1/(\beta+1)}\big)$,*

$$\mathbb{E}\left(||\widehat{S}^{\alpha}_{Y^x} - S_{Y^x}||^2\right) = O\big(n^{-\beta/(\beta+1)}\big).$$

The key element in the proof is decomposing the quadratic risk into a squared bias term $||S^{\alpha}_{Y^x} - S_{Y^x}||^2$ and a variance term $\mathbb{E}\left(||\widehat{S}^{\alpha}_{Y^x} - S^{\alpha}_{Y^x}||^2\right)$.

Lemma 1. *Under Assumptions* **(H.1)**, **(H.2)** *and* **(H.3)**, *we have for all $\alpha > 0$,*

$$||S^{\alpha}_{Y^x} - S_{Y^x}||^2 = O(\alpha^{\beta}).$$

Lemma 2. *Under Assumptions* **(H.1)** *and* **(H.2)**, *we have for all* $\alpha > 0$,

$$\mathbb{E}\left(||\widehat{S}^{\alpha}_{Y^x} - S^{\alpha}_{Y^x}||^2\right) = O\Big(\frac{1}{\alpha n}\Big).$$

The analysis of the variance term does not require Assumption **(H.3)**. This condition is only needed to control the bias term. If it is not satisfied, a different proof may show that the bias term goes to zero when $\alpha \to 0$ [see, for instance, Kreiss (1999), Section 16.5]. It follows then that $\mathbb{E}\left(||\widehat{S}^{\alpha}_{Y^x} - S_{Y^x}||^2\right) \to 0$ if $\alpha \to 0$ and $\alpha n \to \infty$.

Remark 1. The so-called source condition **(H.3)** eliminates pathological cases leading to very low rates of convergence such as, for instance, $\log(n)$ [see Carrasco and Florens (2009)]. Note that this assumption is not incompatible with normal errors as shown in Appendix A.2 through a simple example. In case of a normal error, it just requires that the signal be sufficiently regular. This source condition may be derived from a degree of ill-posedness (linked to the smoothness of the error distribution) and from a regularity of the unknown distribution [more details can be found in *e.g.* Carrasco *et al.* (2014)]. Intuitively, the degree of ill-posedness depends on the rate of decline of $\psi_{\varepsilon|X}(t|x)$, and the degree of regularity depends on the rate of decline of $\psi_{Y|X}(t|x)$. Assumption **(H.3)** warrants a compatibility condition between these two rates.

Remark 2. In what concerns the squared bias term, even if $||S^{\alpha}_{Y^x} - S_{Y^x}||^2 \to 0$ as $\alpha \to 0$, we need to restrict the family of S_{Y^x} in order to define a speed of convergence of this regularization bias. More generally, the theory of inverse problems characterizes families $\mathcal{G}$ of suitable functions g such that $||S^{\alpha}_{Y^x} - S_{Y^x}||^2 = O(g(\alpha))$ with $g(\alpha) \to 0$ as $\alpha \to 0$. Then, under the assumption that $S_{Y^x} \in \mathcal{G}$, we have

$$||\widehat{S}^{\alpha}_{Y^x} - S_{Y^x}||^2 = O\Big(\frac{1}{\alpha n} + g(\alpha)\Big).$$

An optimal α is obtained by solving $1/n = \alpha g(\alpha)$, and hence the optimal rate of convergence can be derived. To ease the presentation, we restrict to a Hölder condition $S_{Y^x} \in \text{Range}\big(K^*K\big)^{\beta/2}$ by Assumption **(H.3)**. This results in a family $\mathcal{G}$ of functions $g(\alpha) = \alpha^{\beta}$ which satisfies the condition $||S^{\alpha}_{Y^x} - S_{Y^x}||^2 = O(\alpha^{\beta})$. A weaker characterization of the family $\mathcal{G}$ may also be given in terms of Hilbert scale [see Carrasco *et al.* (2014)]. Other types of functions $g(\cdot)$ have been suggested in the literature as can be seen from Dunker *et al.* (2014) and the references therein.

Remark 3. It should also be pointed out that, due to the qualification of the Tikhonov regularization, we restrict ourselves to the case $\beta \leqslant 2$. Values of $\beta > 2$ involve some mathematical difficulties that would require more complicated methods such as, for instance, iterated Tikhonov.

3.2. Asymptotic normality of $\sqrt{n}\big(\widehat{S}^{\alpha}_{Y^x} - S_{Y^x}\big)$

By standard theory of empirical processes, it is not hard to verify that $\sqrt{n}\big[(\widehat{S}_{n,Z^x} - \widehat{S}_{n,\varepsilon^x}) - (S_{Z^x} - S_{\varepsilon^x})\big]$ converges in the Hilbert space $\mathcal{E}$ to a zero mean Gaussian process with a variance operator Σ defined as

$$(\Sigma g)(z) = \int_{-\infty}^{\infty} \Gamma(y, z) g(y) dy, \tag{25}$$

where

$$\Gamma(y,z) = \big(S_{Z^x}(y) - S_{\varepsilon^x}(y)\big)\cdot\big[-\big(S_{Z^x}(z) - S_{\varepsilon^x}(z)\big) - S_{\varepsilon|X}(z|x)\big]$$
$$+F_X(x)\cdot\big[S_{Z|X}(y\vee z|x) - S_{\varepsilon|X}(y|x)S_{Z|X}(y|x)\big].$$

Then, we obtain from (19) and (20) that

$$\sqrt{n}\big(\widehat{S}^{\alpha}_{Y^x} - S^{\alpha}_{Y^x}\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0,(\alpha I + K^*K)^{-1}K^*\Sigma K(\alpha I + K^*K)^{-1}\big), \quad n\to\infty,$$

for any fixed $\alpha > 0$. Hence, we get for all $\delta \in \mathcal{E}$,

$$\Big\langle \sqrt{n}\big(\widehat{S}^{\alpha}_{Y^x} - S^{\alpha}_{Y^x}\big),\delta\Big\rangle \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\big\langle(\alpha I + K^*K)^{-1}K^*\Sigma K(\alpha I + K^*K)^{-1}\delta,\delta\big\rangle\right),$$

or equivalently,

$$\frac{\Big\langle \sqrt{n}\big(\widehat{S}^{\alpha}_{Y^x} - S^{\alpha}_{Y^x}\big),\delta\Big\rangle}{\big\langle(\alpha I + K^*K)^{-1}K^*\Sigma K(\alpha I + K^*K)^{-1}\delta,\delta\big\rangle^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{26}$$

For these results to remain valid when the regularization parameter $\alpha = \alpha(n) \to 0$ as a function of the sample size n, we shall need to verify a Lyapunov type condition [see Carrasco *et al.* (2007), Proposition 3.2]. More specifically, since

$$\widehat{S}_{n,Z^x}(z) - \widehat{S}_{n,\varepsilon^x}(z) = \frac{1}{n}\sum_{j=1}^{n}\eta_j(z),$$

where

$$\eta_j(z) := \mathbb{1}(Z^x_j > z) - S_{\varepsilon|X}(z|x)\,\mathbb{1}(X_j \leqslant x) - [1 - \mathbb{1}(X_j \leqslant x)]\mathbb{1}(z<0),$$

we have

$$K^*\big(\widehat{S}_{n,Z^x} - \widehat{S}_{n,\varepsilon^x}\big) \equiv \frac{1}{n}\sum_{j=1}^{n}K^*\eta_j,$$

and hence the needed Lyapunov condition [see Carrasco *et al.* (2007), Assumption (3.27), p.83] reads as follows:

(K.1) There exists $\mathfrak{d} > 0$ such that

$$\frac{\mathbb{E}\big\langle(\alpha I + K^*K)^{-1}K^*\eta_j,\delta\big\rangle^{2+\mathfrak{d}}}{||\sqrt{\mathrm{Var}(\eta_j)}(\alpha I + K^*K)^{-1}\delta||^{2+\mathfrak{d}}} = O(1).$$

Of course, the denominator of (26) should be bounded in order to achieve the $\sqrt{n}$ speed of convergence. This is obtained under an additional regularity condition on δ by following the arguments of Carrasco *et al.* (2014). We need that

(K.2) $\delta \in \mathrm{Range}\big(K^*K\big)^{1/2}$,

or equivalently that $\delta = \big(K^*K\big)^{1/2}\mu$ for some element $\mu \in \mathcal{E}$ [see Proposition 3.2 in Carrasco *et al.* (2014) for more details].

Under the same assumption **(K.2)**, the asymptotic normality remains still valid when eliminating the bias due to regularization. Indeed, on one hand, the numerator of (26) can be written as $\left\langle \sqrt{n}\big(\widehat{S}^{\alpha}_{Y^x} - S_{Y^x} - (S^{\alpha}_{Y^x} - S_{Y^x})\big), \delta \right\rangle$ and, on the other hand, it can be shown under our regularity assumption on δ that $\left\langle \sqrt{n}\big(S^{\alpha}_{Y^x} - S_{Y^x}\big), \delta \right\rangle \to 0$ when $n \to \infty$. To verify this result we have

$$\left\langle \sqrt{n}\big(S^{\alpha}_{Y^x} - S_{Y^x}\big), \delta \right\rangle = \left\langle \sqrt{n}\big(K^*K\big)^{1/2}\big(S^{\alpha}_{Y^x} - S_{Y^x}\big), \mu \right\rangle,$$

for some $\mu \in \mathcal{E}$. As it follows from Lemma 1 that

$$\left\langle \sqrt{n}\big(K^*K\big)^{1/2}\big(S^{\alpha}_{Y^x} - S_{Y^x}\big), \mu \right\rangle^2 = O\big(n\alpha^{(\beta+1)\wedge 2}\big),$$

we gain one unit in the exponent of α due to the factor K^*K. Hence, to get the desired convergence $\left\langle \sqrt{n}\big(S^{\alpha}_{Y^x} - S_{Y^x}\big), \delta \right\rangle \to 0$ as $n \to \infty$, we need a smaller order for α than the optimal size derived above in Theorem 1 to satisfy $\alpha \to 0$, $\alpha n \to \infty$ and $n\alpha^{(\beta+1)\wedge 2} \to 0$. To summarize, we obtain the following result which will be particularly useful below to derive the asymptotic normality of our estimator of the robust order-m frontier.

Theorem 2. *Under Assumptions* **(H.1)**, **(H.2)** *and* **(H.3)**, *we have for $\alpha \to 0$ such that $n\alpha \to \infty$ and $n\alpha^{(\beta+1)\wedge 2} \to 0$, and for all $\delta \in \mathcal{E}$ satisfying* **(K.1)** *and* **(K.2)**,

$$\frac{\left\langle \sqrt{n}\big(\widehat{S}^{\alpha}_{Y^x} - S_{Y^x}\big), \delta \right\rangle}{\langle (\alpha I + K^*K)^{-1}K^*\Sigma K(\alpha I + K^*K)^{-1}\delta, \delta\rangle^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \quad n \to \infty,$$

where Σ is described in (25).

3.3. Selection of the regularization parameter

Several approaches were proposed in the literature on inverse problems to select the regularization parameter. Prominent among these approaches is the iterated Tikhonov technique described below. The solution will select an optimal α having the appropriate order $n^{-1/(\beta+1)}$. By iterating the Tikhonov principle a second time, we can define

$$\widehat{S}^{\alpha,(2)}_{Y^x} = \arg\min_{S \in \mathcal{E}} \left\{ ||KS - (\widehat{S}_{n,Z^x} - \widehat{S}_{n,\varepsilon^x})||^2 + \alpha ||S - \widehat{S}^{\alpha}_{Y^x}||^2 \right\}$$

which leads, by solving the first order condition, to the following estimator of the unconditional survivor function

$$\widehat{S}^{\alpha,(2)}_{Y^x} = (\alpha I + K^*K)^{-1}\big(K^*(\widehat{S}_{n,Z^x} - \widehat{S}_{n,\varepsilon^x}) + \alpha \widehat{S}^{\alpha}_{Y^x}\big).$$

By convoluting this estimator with the noise, we obtain an estimator of the survivor function of the noisy signal Z^x,

$$\widehat{S}^{\alpha,(2)}_{Z^x}(z) - \widehat{S}_{n,\varepsilon^x}(z) = \int_0^{\infty} \widehat{S}^{\alpha,(2)}_{Y^x}(y)\, f_{\varepsilon|X}(z - y|x)\, dy.$$

Finally, the optimal value of α is given by the solution

$$\begin{aligned}\widehat{\alpha} &= \arg\min_{\alpha>0} ||\widehat{S}^{\alpha}_{Y^x}||^2 ||\widehat{S}^{\alpha,(2)}_{Z^x} - \widehat{S}_{n,Z^x}||^2 \\ &= \arg\min_{\alpha>0} \int_0^{\infty} \big(\widehat{S}^{\alpha}_{Y^x}(y)\big)^2 dy \int_{-\infty}^{\infty} \big[\widehat{S}^{\alpha,(2)}_{Z^x}(z) - \widehat{S}_{n,Z^x}(z)\big]^2 w(z)dz.\end{aligned}$$

It is not hard to verify that this solution satisfies $\widehat{\alpha} = O(n^{-1/(\beta+1)})$ for $\beta \in (0,2]$. We refer to Engl *et al.* (2000, Proposition 4.37) for a proof and to Fève and Florens (2014) for more thorough discussion.

4. Estimation of partial and full production frontiers

We can now use the estimator $\widehat{S}^{\alpha}_{Y^x}$ of S_{Y^x} defined in (20) to estimate the partial order-m production function $\varphi_m(x)$ as well as the true full frontier function $\varphi(x)$.

4.1. Estimation of the expected maximum output frontier

By definition (6) we have

$$\begin{aligned}\varphi_m(x) &= \int_0^{\infty} \big\{1 - [1 - S_{Y^x}(y)]^m\big\}\, dy \\ &\equiv \int_0^{\tau} \big\{1 - [1 - S_{Y^x}(y)]^m\big\}\, dy, \qquad (27)\end{aligned}$$

for any arbitrary large positive number τ satisfying $\tau \geqslant \varphi(x)$. Then, by substituting $\widehat{S}^{\alpha}_{Y^x}$ in place of S_{Y^x}, we get the trimmed estimator

$$\widehat{\varphi}^{\alpha}_m(x) = \int_0^{\tau} \big\{1 - [1 - \widehat{S}^{\alpha}_{Y^x}(y)]^m\big\}\, dy. \qquad (28)$$

In practice, it suffices to prespecify any trimming number τ larger than the maximum of the observed contaminated outputs $Z_1, \ldots, Z_n$. Next we establish an indicative rate of convergence of the quadratic risk for the estimator $\widehat{\varphi}^{\alpha}_m(x)$.

Theorem 3. *Under the conditions of Theorem 1, we have for any fixed $m \geqslant 1$,*

$$\mathbb{E}\,|\widehat{\varphi}^{\alpha}_m(x) - \varphi_m(x)|^2 = O\big(n^{-\beta/(\beta+1)}\big).$$

If $m = m(n) \to \infty$ with $m = O\left(n^{\beta/(2(\beta+1))}\right)$, then $\mathbb{E}\,|\widehat{\varphi}^{\alpha}_m(x) - \varphi_m(x)|^2 = m^2 O\big(n^{-\beta/(\beta+1)}\big)$.

By making use of the results of Section 3.2, we can also derive the asymptotic normality of the m-frontier estimator $\widehat{\varphi}^{\alpha}_m(x)$.

Theorem 4. *Let $m \geqslant 1$ be a fixed integer. Under the conditions of Theorem 2 with $\beta > 1$, if $\delta := \mathbb{1}(\cdot \leqslant \tau)F^{m-1}_{Y^x}$ satisfies* **(K.1)** *and* **(K.2)**, *then*

$$\frac{\sqrt{n}\big(\widehat{\varphi}^{\alpha}_m(x) - \varphi_m(x)\big)}{m\,\langle(\alpha I + K^*K)^{-1}K^*\Sigma K(\alpha I + K^*K)^{-1}\delta, \delta\rangle^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \quad n \to \infty.$$

4.2. Estimation of the full production frontier

For estimating the true frontier function itself we need an extra regularity condition on the behavior of the unconditional survivor function $S_{Y^x}(y) \equiv F_X(x)[1 - F_{Y|X}(y|x)]$ near the frontier point $\varphi(x)$. This condition indicates the rate at which the survivor function reaches the value 0 when $y \uparrow \varphi(x)$:

(K.3) For some constants $\ell_x > 0$ and $\rho_x > 0$,

$$S_{Y^x}(y) = \ell_x\big(\varphi(x) - y\big)^{\rho_x} + o\big((\varphi(x) - y)^{\rho_x}\big) \quad \text{as} \quad y \uparrow \varphi(x).$$

Remark 4. [Hidden extreme-value condition] The rationale for Assumption **(K.3)** relies on an interesting connection between our class of expected maximum production functions $\{\varphi_m(x) : m \geqslant 1\}$ defined in (6) and the popular FDH estimator $\widehat{\varphi}(x)$ described in (2) and (5). It is immediate from (5) and (6) that

$$\mathbb{E}\big[\widehat{\varphi}(x)\big] = \mathbb{E}\big[\max(Y_1^x, \ldots, Y_n^x)\big] \equiv \varphi_n(x), \quad \text{for all} \quad n \geqslant 1. \tag{29}$$

Equivalently, for any trimming number $m \geqslant 1$, $\varphi_m(x)$ is identical to the expectation of the FDH estimator based on the m-tuple $\{Y_i^x = Y_i \mathbb{1}(X_i \leqslant x)\,,\, i = 1, \ldots, m\}$. As elucidated in Daouia *et al.* (2010, Theorem 2.1), there exists $b_n > 0$ such that $b_n^{-1}(\widehat{\varphi}(x) - \varphi(x))$ converges in distribution (as $n \to \infty$) if and only if, for some $\rho_x > 0$,

$$\lim_{t\to\infty} \{1 - F_{Y|X}(\varphi(x) - 1/tz|x)\}/\{1 - F_{Y|X}(\varphi(x) - 1/t|x)\} = z^{-\rho_x} \quad \text{for all} \quad z > 0$$

[regular variation with exponent $-\rho_x$, notation $1 - F_{Y|X}(\varphi(x) - \frac{1}{t}|x) \in \mathrm{RV}_{-\rho_x}$]. As also pointed out in Daouia *et al.* (2010, Remark 2.1), this necessary and sufficient condition for the standard FDH estimator $\widehat{\varphi}(x)$ to converge in distribution is equivalent to the representation

$$F_X(x)[1 - F_{Y|X}(y|x)] = L_x\left(\{\varphi(x) - y\}^{-1}\right)(\varphi(x) - y)^{\rho_x} \quad \text{as} \quad y \uparrow \varphi(x),$$

or equivalently

$$S_{Y^x}(y) = L_x\left(\{\varphi(x) - y\}^{-1}\right)(\varphi(x) - y)^{\rho_x} \quad \text{as} \quad y \uparrow \varphi(x), \tag{30}$$

where $L_x(\cdot) \in \mathrm{RV}_0$ stands for a slowly varying function. As a matter of fact, Assumption **(K.3)** is hidden in Condition (30). Both conditions are equivalent when $L_x\left(\{\varphi(x) - y\}^{-1}\right) \sim \ell_x$, or equivalently, $L_x\left(\{\varphi(x) - y\}^{-1}\right) = \ell_x + o(1)$, as $y \uparrow \varphi(x)$. While the necessary and sufficient condition (30) is sometimes difficult to verify, the sufficient von Mises assumption **(K.3)** may be more helpful. Under this sufficient condition, it is shown in Daouia *et al.* (2010, Corollary 2.1) that $b_n \sim (n\ell_x)^{-1/\rho_x}$ as $n \to \infty$.

We know by Theorem 2.1(iii) in Daouia *et al.* (2010) that the convergence in distribution of the FDH estimator $\widehat{\varphi}(x)$ implies the convergence of moments. More precisely, given (30), $\lim_{n\to\infty} \mathbb{E}\big\{b_n^{-1}(\varphi(x) - \widehat{\varphi}(x))\big\}^k = \Gamma(1 + k\rho_x^{-1})$, for all integer $k \geqslant 1$. In particular, it follows from (29) and for $k = 1$ that

$$\varphi(x) - \varphi_n(x) = \varphi(x) - \mathbb{E}\big[\widehat{\varphi}(x)\big] = b_n\Gamma(1 + \rho_x^{-1}) + o\big(b_n\big), \quad n \to \infty.$$

We also know from Remark 4 that, under the sufficient condition **(K.3)**, we have $b_n \sim (n\ell_x)^{-1/\rho_x}$ as $n \to \infty$. Therefore, using m instead of n, we get

$$\varphi(x) - \varphi_m(x) = (m\ell_x)^{-1/\rho_x}\Gamma(1+\rho_x^{-1}) + o\big(m^{-1/\rho_x}\big), \quad m \to \infty. \tag{31}$$

This result will be crucial for our setup to derive the speed of convergence of the quadratic risk for the regularized estimator $\widehat{\varphi}^{\alpha}_m(x)$ when it estimates $\varphi(x)$ itself, with $m = m(n) \to \infty$ as $n \to \infty$.

Remark 5. [Connection with the joint density and intuitive meaning for the exponent ρ_x] In the econometrics and statistical literatures on frontier analysis, it is common to assume that the joint density $f(x, y)$ of (X, Y) is an algebraic function of the distance $(\varphi(x) - y)$ from the efficient frontier, that is

$$f(x, y) = c_x \{\varphi(x) - y\}^{\gamma_x} + o(\{\varphi(x) - y\}^{\gamma_x}) \quad \text{as} \quad y \uparrow \varphi(x), \tag{32}$$

for some constants $c_x > 0$ and $\gamma_x > -1$. For nonparametric approaches to frontier estimation, we refer to Hardle *et al.* (1995), Hall *et al.* (1998), Gijbels and Peng (2000), Park *et al.* (2000), Hwang *et al.* (2002) and Daouia *et al.* (2010, 2012), to cite a few. In all parametric approaches, the shape parameter γ_x of the joint density as well as c_x are assumed to be known. The traditional assumption (32) is actually hidden in Condition (30) and is more stringent than Condition **(K.3)**. It is obtained by considering the class of slowly varying functions $L_x(\cdot)$ satisfying $L_x\left(\{\varphi(x) - y\}^{-1}\right) = \ell_x$ as $y \uparrow \varphi(x)$. Then, if $\ell_x > 0$, $\rho_x > p$ and $\varphi(x)$ are differentiable as functions of x with first partial derivatives of $\varphi(x)$ being strictly positive, one can easily recover the usual assumption (32), with

$$\gamma_x = \rho_x - (p+1),$$

by simply differentiating both sides of (30) [see also Daouia *et al.* (2010, Corollary 2.2)]. Accordingly, the joint density has an interesting connection with the regular variation exponent ρ_x and the dimension $(p+1)$: When $\rho_x > p+1$, the joint density decays to zero at a speed of power $\rho_x - (p+1)$ of the distance from the frontier point $\varphi(x)$. When $\rho_x = p+1$, the density has a sudden jump at the frontier. Finally, when $\rho_x < p+1$, the density rises up to infinity at a speed of power $\rho_x - (p+1)$ of the distance from the frontier.

Next, we establish the rate of convergence of the quadratic risk $\mathbb{E}\left[(\widehat{\varphi}^{\alpha}_m(x) - \varphi(x))^2\right]$ by applying Theorem 3 in conjunction with (31).

Theorem 5. *Under* **(K.3)** *and the conditions of Theorem 1, if* $m = \left(\rho_x^{-1}\, n^{\frac{\beta}{\beta+1}}\right)^{\frac{\rho_x}{2(1+\rho_x)}}$, *then*

$$\mathbb{E}\left[(\widehat{\varphi}^{\alpha}_m(x) - \varphi(x))^2\right] = O\big(n^{-\frac{\beta}{(\beta+1)}\frac{1}{(1+\rho_x)}}\big).$$

Under the conditions of this theorem we have that

$$r_n\, |\widehat{\varphi}^{\alpha}_m(x) - \varphi(x)| = O_p(1),$$

where $r_n = n^{\kappa}$ with $\kappa \geqslant \frac{\beta}{2(\beta+1)(1+\rho_x)}$. Under the common assumption in most nonparametric frontier estimation approaches that the joint density of (X, Y) has a jump at its efficient support boundary, we have

$(1+\rho_x) = p+2$ (see Remark 5). Hence, if for instance $\beta = 2$, we get the rate $r_n \geqslant n^{\frac{1}{3(p+2)}}$, which is still polynomial in n.

Remark 6. [Tuning parameters selection] Critical to the quality of the stochastic frontier approximation is the selection of the trimming number m. We propose to choose m by an analogy to the method motivated in the deterministic frontier model, in Section 2.4 of Daouia *et al.* (2018), by substituting the stochastic frontier estimator $\widehat{\varphi}_m^\alpha$ in place of $\widehat{\varphi}_m$ and using the observed contaminated outputs Z_j instead of Y_j. This method is applied below in Section 5 through simulated and real data sets. The regularization parameter $\alpha = \widehat{\alpha}$ is obtained following the guidelines described in Section 3.3. As shown in the examples below, these selection techniques aim to balance the robustness of the estimate to outliers (not too large m) with the desire of reaching the full sample frontier (sufficiently large m).

Remark 7. [Isotonized estimators] The regularized estimator $\widehat{S}^\alpha_{Y^x}$ may not automatically inherit the monotonicity property of the true survival function S_{Y^x}. One way to monotonize this unconstrained estimator is by using the isotonic version

$$\breve{S}^\alpha_{Y^x} = [\overline{S}^\alpha_{Y^x} + \underline{S}^\alpha_{Y^x}]/2,$$

with

$$\overline{S}^\alpha_{Y^x}(y) = \sup_{y' \geqslant y} \widehat{S}^\alpha_{Y^x}(y') \quad \text{and} \quad \underline{S}^\alpha_{Y^x}(y) = \inf_{y' \leqslant y} \widehat{S}^\alpha_{Y^x}(y'),$$

where y and y' run over $\mathbb{R}_+$. Both $\overline{S}^\alpha_{Y^x}$ and $\underline{S}^\alpha_{Y^x}$ are monotone non-increasing such that $\underline{S}^\alpha_{Y^x} \leqslant \widehat{S}^\alpha_{Y^x} \leqslant \overline{S}^\alpha_{Y^x}$. While $\overline{S}^\alpha_{Y^x}$ is the smallest monotone function that lies above the unconstrained estimator $\widehat{S}^\alpha_{Y^x}$, $\underline{S}^\alpha_{Y^x}$ is the largest monotone function that lies below $\widehat{S}^\alpha_{Y^x}$. As a matter of fact, any convex combination of these envelope estimators would have sufficed as a definition of $\breve{S}^\alpha_{Y^x}$, but we do not see any reason to bias the restricted estimator one way or the other. By substituting in (28) the monotone estimator $\breve{S}^\alpha_{Y^x}$ in place of the unconstrained version $\widehat{S}^\alpha_{Y^x}$, we get the refined frontier estimator

$$\breve{\varphi}^\alpha_m(x) = \int_0^\tau \left\{1 - [1 - \breve{S}^\alpha_{Y^x}(y)]^m\right\} dy.$$

The isotonic estimator $\breve{S}^\alpha_{Y^x}$ reduces the sup-norm error in the following sense

$$\sup_{y \geqslant 0} \left|\breve{S}^\alpha_{Y^x}(y) - S_{Y^x}(y)\right| \leqslant \sup_{y \geqslant 0} \left|\widehat{S}^\alpha_{Y^x}(y) - S_{Y^x}(y)\right|.$$

This can easily be checked by applying Lemma 3.1 of Daouia and Simar (2005). A deeper study of the properties of such a projection-type technique of isotonization can be found in Daouia and Park (2013). In particular, it follows from their generic Theorem 1 that the monotonized estimator $\breve{S}^\alpha_{Y^x}(y)$ inherits the same asymptotic first-order properties of the unrestricted estimator $\widehat{S}^\alpha_{Y^x}(y)$ if the latter is asymptotically equicontinuous as a process indexed by y. Establishing the asymptotic equicontinuity of this process will lead to further investigations that are outside the scope of the present paper.

Remark 8. [Case of a Gaussian noise with unknown variance] The concern was raised by a referee that the noise distribution is fully known (*i.e.* ε_j given $X_j \leqslant x$ is Gaussian with known variance). Let us comment briefly the case where ε_j given $X_j \leqslant x$ has a normal distribution with zero mean and unknown variance $\sigma^2(x)$. The identification follows from *e.g.* Theorem 2.1 in Schwarz and Van Bellegem (2010), because the support of Y given $X \leqslant x$ is a subset of $\mathbb{R}_+$. We do not enter here into the important issue of simultaneous estimation of the frontier function $\varphi(x)$ and the variance parameter $\sigma^2(x)$. We only describe a way to estimate $\sigma^2(x)$ in a first stage before applying our estimation procedure of $\varphi(x)$ in a second stage. Very few estimators have been proposed in the literature to address the problem of simultaneous estimation, but they tend to be either much more computationally expensive and/or very disappointing in terms of rates of convergence [see *e.g.* Kneip *et al.* (2015)].

Attempts to estimate $\sigma^2(x)$ have been proposed, for instance, by Meister (2006) and Butucea *et al.* (2008). The identification in these works comes from the assumption that the tails of the characteristic function of Y given $X \leqslant x$ decay at a slower rate than the tails of the characteristic function of the normal distribution, which seems unsuitable for our framework. In addition, these approaches are very difficult to implement in practice.

Instead, we follow in our experiments with simulated and real data an alternative heuristic path based on a sensitivity analysis with several choices of $\sigma^2(x)$. The basic idea is to first compute the estimate $\widehat{S}^{\alpha}_{Y^x}$ by selecting a value for $\sigma^2(x)$, potentially different from the true variance, and then evaluate the $L^2(\mathbb{R}, w)$ distance in $\mathcal{F}$ between the re-convoluted estimator of $S_{Z^x}(z)$, obtained by plugging in (10) our regularized estimate $\widehat{S}^{\alpha}_{Y^x}$, and the observed empirical $\widehat{S}_{n,Z^x}$:

$$\nabla_{L^2}(x) = \int_{-\infty}^{\infty} \left(\widehat{S}^{\alpha}_{Z^x}(z) - \widehat{S}_{n,Z^x}(z) \right)^2 w(z) dz. \tag{33}$$

A small value of this distance should indicate a better fit of the observed data, and hence an appropriate estimate of $\sigma^2(x)$ would be obtained by minimizing the criterion $\nabla_{L^2}(x)$. The merits of this approach are not justified theoretically here, but our experience with simulated data below indicates that the resulting estimates of $\sigma^2(x)$ are quite reasonable. We shall also discuss this idea on a concrete application to two real datasets in Section 5.2.

5. Numerical illustrations

Section 5.1 comments on some implementation details and reports some simulation illustrations. Section 5.2 explores the new unconditional m-frontier estimates for two datasets in the sector of Delivery Services.

5.1. Simulated example

We simulate the data (X_i, Y_i) following a uniform distribution on the region $D = \{(x, y) \,|\, 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant \varphi(x)\}$, where $\varphi(x) = 2 + x$ and $x \in [0, 1]$. Then we introduce the noise $\varepsilon_i \sim N(0, \sigma^2)$ producing the observed production levels $Z_i = Y_i + \varepsilon_i$. We use in all our simulations the sample size $n = 200$ and the two values $\sigma = 0.20$, 0.40.

In the Hilbert space $\mathcal{F} = L^2(\mathbb{R}, w)$, we choose a uniform weight function w on the interval $[\tau_1 = -20, \tau_2 = 20]$. With this choice and the Gaussian noise, it is not hard to verify that the kernel function defining K^*K and given in (21) has the form

$$c(y,\xi) = \frac{1}{\tau_2 - \tau_1}\phi_N(y-\xi; 0, \sqrt{2}\sigma)\left[\Phi_N\Big(\tau_2; \frac{y+\xi}{2}, \frac{\sigma}{\sqrt{2}}\Big) - \Phi_N\Big(\tau_1; \frac{y+\xi}{2}, \frac{\sigma}{\sqrt{2}}\Big)\right],$$

where $\phi_N(\cdot; a, b)$ and $\Phi_N(\cdot; a, b)$ denote, respectively, the pdf and cdf of $\mathcal{N}(a, b^2)$. The one-dimensional integrals involving z, *e.g.* for computing $b(y|x)$ in (23), are performed by a trapezoidal rule over a grid of 801 equi-spaced values in $[\tau_1, \tau_2]$. The solution of the linear system in (24) was done for a grid of $k = 801$ points y_j in $[0, \max(Z_i|X_i \leqslant x) + 3\sigma]$. The estimators of the survivor and frontier functions have to be evaluated at fixed values of x: we selected here a grid of 91 equidistant values in $[0.1, 1]$. When $\sigma = 0.20$, the 91 optimal values of α obtained by the iterative Tikhonov method are ranging form $0.0184 * 10^{-3}$ upto $0.4914 * 10^{-3}$ with an average of $0.2081 * 10^{-3}$. We then computed the m-frontier estimates over various values of the trimming order m ranging from 1 up to 5000, and for each value we computed the percentage of points left above the m-frontier. We selected an appropriate high value of m by looking to the place where an "elbow" effect appears in the curve of the percentage of points above the m-frontier as a function of m. The evolution of this percentage, graphed in Figure 1, indicates that the curve becomes almost horizontal from $m = 600$. Figure 2 shows the resulting frontier estimates for various trimming orders m around the selected value 600, namely $400 \leqslant m \leqslant 900$. We can see that the results are rather stable with respect to the choice of m in the displayed range. The final frontier estimate $\widehat{\varphi}_m^\alpha$ of order $m = 600$ is graphed in Figure 3 along with its isotonized version $\breve{\varphi}_m^\alpha$.

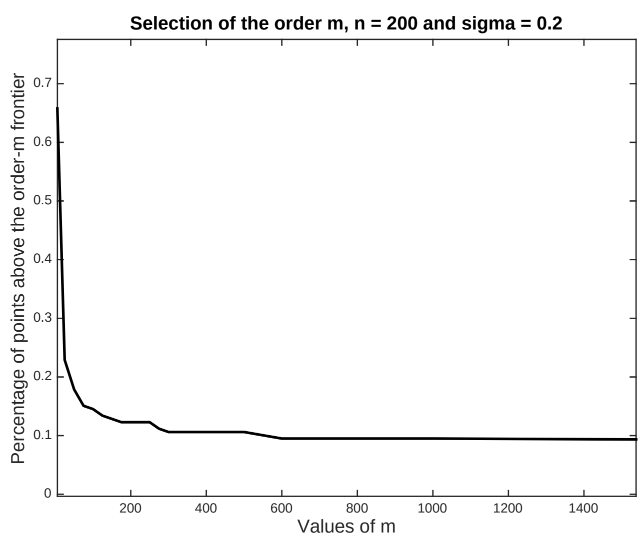

Figure 1: *Simulated example with $\sigma = 0.20$. The evolution of the percentage of observations left outside the m-frontier.*

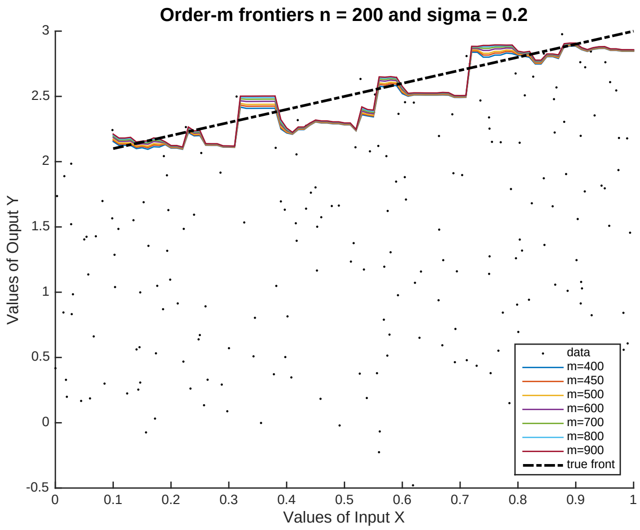

Figure 2: *Simulated example with* $\sigma = 0.20$. *Order-*m *frontier estimates for some selected values of* m *along with the true frontier.*

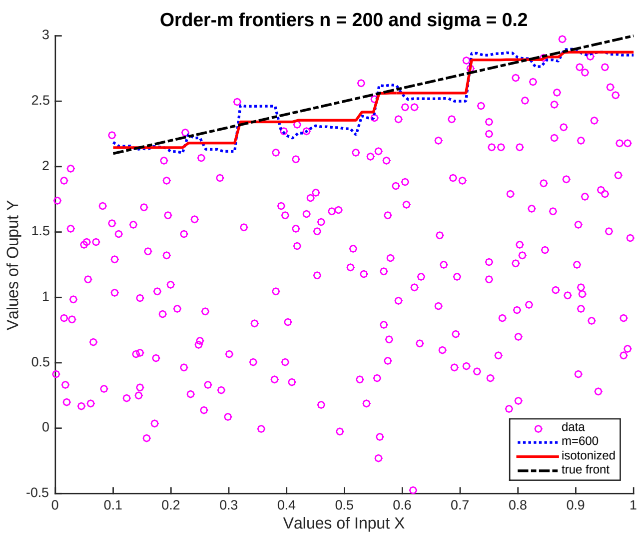

Figure 3: *Simulated example with* $\sigma = 0.20$. *Final* m*-frontier estimate along with its isotonized version and the true frontier.*

When $\sigma = 0.40$, the 91 optimal Tikhonov regularization parameters α range from $0.0206 * 10^{-3}$ to $1.5537 * 10^{-3}$ with an average of $0.6223 * 10^{-3}$. The obtained order-m frontier estimates, the percentage curve which suggests the choice of $m = 325$, and the final frontier estimates are shown in Figures 4 to 6.

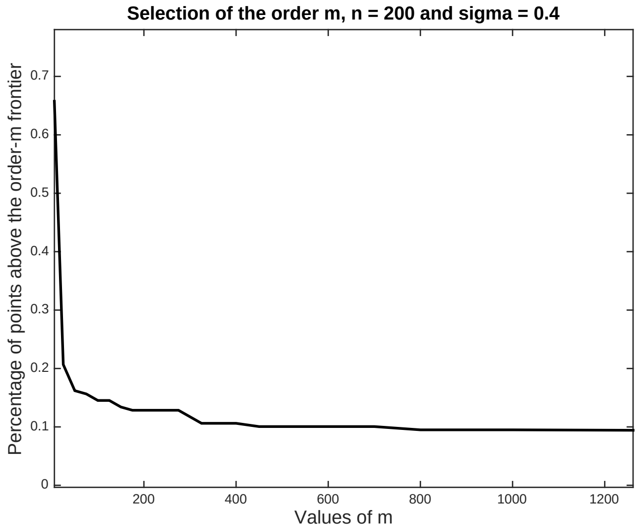

Figure 4: *Simulated example with* $\sigma = 0.40$*. The evolution of the percentage of observations left outside the* m*-frontier.*

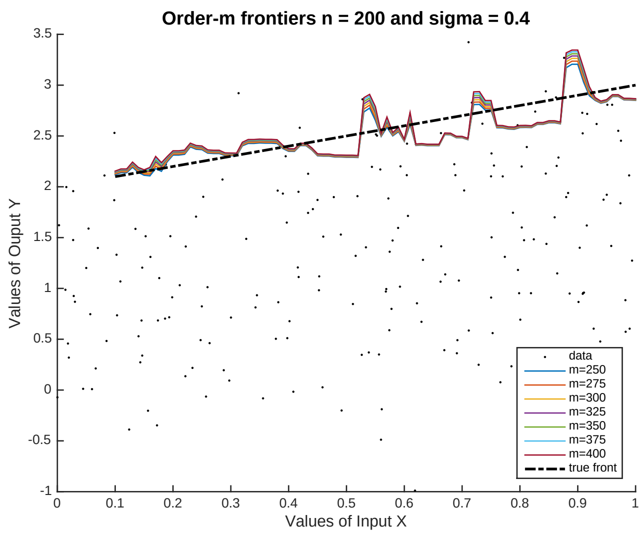

Figure 5: *Simulated example with* $\sigma = 0.40$*. Order-*m *frontier estimates for selected values of* m *and the true frontier.*

Figure 6: *As before with the final m-frontier estimate and its isotonized version.*

The calculation above has been done with the noise distribution being fully known. It is clear that a misspecified value of the variance σ^2 would have an impact on the frontier estimation: if we select a value $\sigma_{\text{mis}} \neq \sigma$, the corresponding estimated frontier is expected to lie above the estimate obtained with the well-specified variance if $\sigma_{\text{mis}} < \sigma$ (little or almost no noise in this case, so the estimate tends to envelop more data points). The opposite is expected if $\sigma_{\text{mis}} > \sigma$. This is illustrated, when $\sigma = 0.20$, in Figure 7 for $\sigma_{\text{mis}} = 0.5\sigma$ and in Figure 8 for $\sigma_{\text{mis}} = 2\sigma$ (note that the new values of m have been selected here according to the same rule as above). The correct frontier estimate, obtained by using the well-specified variance $\sigma = 0.20$, can be visualised in Figure 3. By comparing Figures 3 and 7, it may be seen that the choice of a too small σ results in a frontier estimate slightly above the correct one. By contrast, as visualised more clearly from Figures 3 and 8, the choice of a too large σ results in a frontier estimate below the correct one.

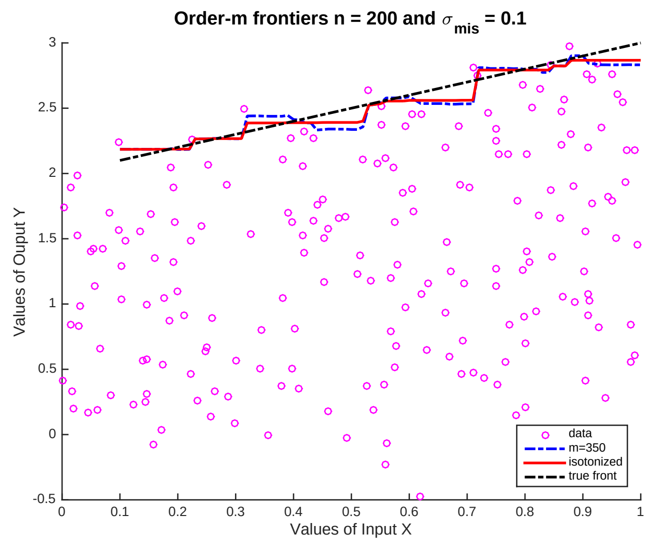

Figure 7: *The final m-frontier estimate (dashed blue) and its isotonized version (solid red), when the true* $\sigma = 0.20$ *and the misspecified* $\sigma_{mis} = 0.5\sigma = 0.10$*. Here* $m = 350$ *with 8% of data points above the frontier.*

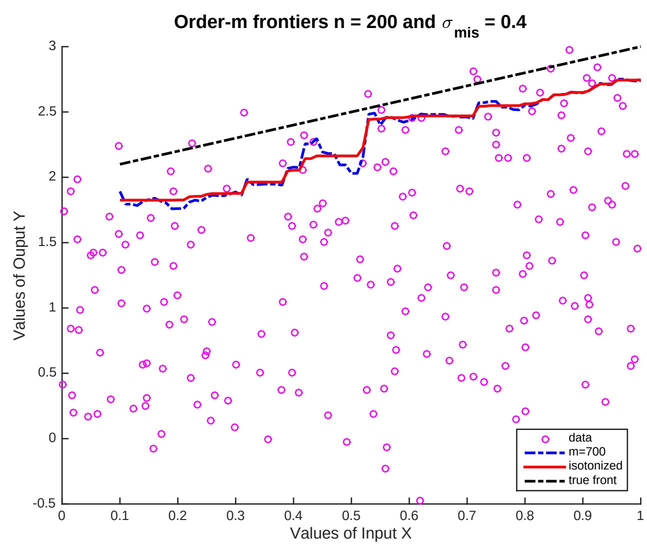

Figure 8: *The final m-frontier estimate (dashed blue) and its isotonized version (solid red), when the true* $\sigma = 0.20$ *and the misspecified* $\sigma_{mis} = 2\sigma = 0.40$*. Here* $m = 700$ *with 17% of data points above the frontier.*

When the variance is unknown, it is then crucial to have a rule for selecting the value of σ in order to avoid the misspecification effects. As described above in Remark 8, we propose to use a heuristic method based on the evaluation of the quality of the fit via the criterion $\nabla_{L^2}(x)$ defined in (33). We calculate this distance at each point of the chosen grid of values of x, before computing the average $\overline{\nabla}_{L^2}$ of $\nabla_{L^2}(x)$ over the 91 values of the grid. A small value of $\overline{\nabla}_{L^2}$ should indicate a better fit of the observed data. Table 1 displays the averaged estimates $\overline{\nabla}_{L^2}$ for several values of the misspecified σ_{mis} (in the first column), for the sample sizes $n = 200, 400$ and 800 (in columns 2-4), and for the true values $\sigma = 0.20$ (in the top part of the table) and $\sigma = 0.40$ (in the bottom part of the table). We see clearly in both cases that the procedure provides reasonable estimates of σ. Of course, this tentative conclusion follows from only one sample, and this does not prove the consistency of the procedure. Still the results are promising, which merits to be fully explored from a theoretical point of view in a separate research.

Table 1: *Simulated examples—the table give the values of $\overline{\nabla}_{L^2}$, an average measure of the quality of the fit of $\widehat{S}_{n,Z^x}$ for various sample sizes n and various misspecified values σ_{mis}. In bold the minimum of the averaged values within the same column (for a fixed sample size n).*

Case I: true $\sigma = 0.20$

σ_{mis}	$n = 200$	$n = 400$	$n = 800$
0.10	2.9841×10^{-5}	4.9058×10^{-6}	1.0818×10^{-4}
0.15	1.1826×10^{-5}	$\mathbf{1.3532 \times 10^{-6}}$	4.4450×10^{-6}
0.20	7.2351×10^{-6}	2.6118×10^{-6}	$\mathbf{2.5367 \times 10^{-6}}$
0.25	$\mathbf{5.0618 \times 10^{-6}}$	5.2855×10^{-6}	2.9511×10^{-6}
0.30	7.2159×10^{-6}	8.7844×10^{-6}	3.8063×10^{-6}

Case II: true $\sigma = 0.40$

σ_{mis}	$n = 200$	$n = 400$	$n = 800$
0.20	5.8825×10^{-5}	3.0659×10^{-5}	2.9828×10^{-4}
0.30	1.8649×10^{-5}	7.0668×10^{-6}	1.1967×10^{-5}
0.40	$\mathbf{1.0380 \times 10^{-5}}$	6.8141×10^{-6}	$\mathbf{3.3817 \times 10^{-6}}$
0.50	1.2041×10^{-5}	$\mathbf{5.1337 \times 10^{-6}}$	6.2561×10^{-6}
0.60	2.3734×10^{-5}	7.1865×10^{-6}	1.3726×10^{-5}

5.2. Real data examples

We consider the same datasets from the sector of Delivery Services as in the study of Daouia *et al.* (2018). The first dataset involves 2,326 European post offices observed in 2013, and the second dataset comprises 4,000 French post offices observed in 1994. For each post office j, the input X_j is the labor cost measured by the quantity of labor, and the output Y_j is the volume of delivered mail in number of objects. The scatterplots are displayed in the bottom of Figures 9 and 10.

In contrast to Daouia *et al.* (2018), here we consider that some noise may perturb the data of delivery post offices. We assume that the noise ε_j given $X_j \leqslant x$ has a normal distribution with zero mean and unknown variance $\sigma^2_{\varepsilon^x}$. As above we estimate the various m-frontiers at fixed values of x over a grid of 100 equidistant values covering most of the range of X. For selecting the values of $\sigma^2_{\varepsilon^x}$, we apply the device described in Remark 8 and tested in the two simulated examples above. We perform here the sensitivity analysis at different levels of the noise-to-signal ratio $\rho_{\text{nts}} \in \{0.01, 0.05, 0.10, 0.20, 0.40, 0.80\}$, so that $\sigma_{\varepsilon^x} = \rho_{\text{nts}} \times \text{std}(Z_i|X_i \leqslant x)$ at each given value of x, allowing thus for a heteroskedastic noise (though we assume ρ_{nts} to be constant over the values of x). We first estimate the survivor function $S_{Y^x}(y)$ for each x over a grid of $k = 1000$ values of y. In each case, the optimal regularization parameter is computed by the iterated technique described above. The one-dimensional integrals involving z are performed by a trapezoidal rule over a grid of 1000 equidistant values in $[\tau_1^x, \tau_2^x]$, where $\tau_1^x = \min_i(Z_i|X_i \leqslant x) - 2\sigma_{\varepsilon^x}$ and similarly, $\tau_2^x = \max_i(Z_i|X_i \leqslant x) + 2\sigma_{\varepsilon^x}$.

Table 2 reports the main results. First, it indicates that the obtained values of α slightly vary when changing the level of the noise. The table gives also the average $\overline{\nabla}_{L^2}$ of the 100 values of $\nabla_{L^2}(x)$ evaluated at each grid point. As small values of $\overline{\nabla}_{L^2}$ should indicate a better fit of the observed data, the analysis of $\overline{\nabla}_{L^2}$ as a function of ρ_{nts} can then be utilized to select a reasonable value of ρ_{nts}, and hence an appropriate value of σ_{ε^x} at each point x. In Case I of $n = 2326$, where the data seems more reliable (less hectic and extreme data points), this empirical rule determines a small value of the noise-to-signal ratio around $\rho_{\text{nts}} = 0.05$. In Case II of $n = 4000$, the empirical rule suggests $\rho_{\text{nts}} = 0.10$.

The selected values of m, displayed in columns 6, are quite stable. Unsurprisingly, we see a great difference between the two cases due to the obvious spread and over-dispersion of the data points in Case II. The choice of $\rho_{\text{nts}} = 0.05$ in Case I leads to $m_{\text{opt}} = 650$, while $\rho_{\text{nts}} = 0.10$ in Case II leads to $m_{\text{opt}} = 450$. Compared to the results obtained in the deterministic setting in Section 2.4 of Daouia *et al.* (2018), we get here lower trimming numbers m since a part of extreme data points is handled by the noise.

The final results are graphed in Figure 9 for Case I and in Figure 10 for Case II. In each figure we represent the percentage curve (top), some frontier estimates of trimming orders around the selected value m_{opt} (middle), and the m_{opt}-frontier estimate itself along with its isotonized version (bottom). In the top figures, the flatness of the percentage curves from the values m_{opt} might confirm the relevance of our choice of the noise-to-signal ratio ρ_{nts} in each case. The cloud of data points in case I is more concentrated than in case II, resulting in larger percentage of observations above the final estimate. This data concentration in case I generates an isotonized frontier estimator almost confounded with the unrestricted version in the bottom of Figure 9. Also, it may be seen from the figures in the middle that the results in both cases are rather stable to the choice of the trimming parameter in the selected range of values near m_{opt}.

Table 2: *Examples with the Delivery post offices—Estimates computed over 100 equi-spaced values of the input* x. ρ_{nts} *is the chosen global noise to signal ratio,* $\overline{\nabla}_{L^2}$ *is an average measure of the quality of the fit of* $\widehat{S}_{n,Z^x}$, m_{opt} *is the selected value of* m *and* %out *is the percentage of observations left above a smoothed version of the resulting* $\widehat{\varphi}^{\alpha}_{m_{\mathrm{opt}}}(x)$.

Case I: $n = 2326$

ρ_{nts}	$\alpha_{\min}$	$\bar{\alpha}$	$\alpha_{\max}$	$\overline{\nabla}_{L^2}$	m_{opt}	%out
0.01	2.124×10^{-3}	1.582×10^{-2}	5.387×10^{-2}	6.644×10^{-4}	600	31.40
0.05	2.194×10^{-4}	7.267×10^{-4}	5.181×10^{-3}	1.421×10^{-5}	650	31.32
0.10	5.801×10^{-5}	7.720×10^{-4}	8.729×10^{-3}	2.963×10^{-5}	700	33.35
0.20	1.452×10^{-5}	9.763×10^{-4}	1.047×10^{-2}	4.392×10^{-5}	800	35.85
0.40	8.644×10^{-7}	1.634×10^{-3}	1.909×10^{-2}	1.837×10^{-4}	800	42.76
0.80	1.210×10^{-4}	3.892×10^{-3}	2.125×10^{-2}	1.451×10^{-4}	900	51.31

Case II: $n = 4000$

ρ_{nts}	$\alpha_{\min}$	$\bar{\alpha}$	$\alpha_{\max}$	$\overline{\nabla}_{L^2}$	m_{opt}	%out
0.01	3.729×10^{-6}	5.340×10^{-6}	8.969×10^{-6}	2.711×10^{-4}	400	1.33
0.05	1.559×10^{-9}	1.684×10^{-7}	3.280×10^{-6}	6.786×10^{-6}	600	1.31
0.10	7.101×10^{-10}	4.774×10^{-7}	1.086×10^{-5}	6.687×10^{-6}	450	1.94
0.20	1.479×10^{-9}	7.166×10^{-6}	1.875×10^{-5}	1.397×10^{-5}	450	1.48
0.40	6.640×10^{-9}	3.483×10^{-7}	7.977×10^{-7}	4.317×10^{-5}	500	1.94
0.80	9.438×10^{-8}	3.528×10^{-6}	2.827×10^{-5}	1.027×10^{-3}	450	2.29

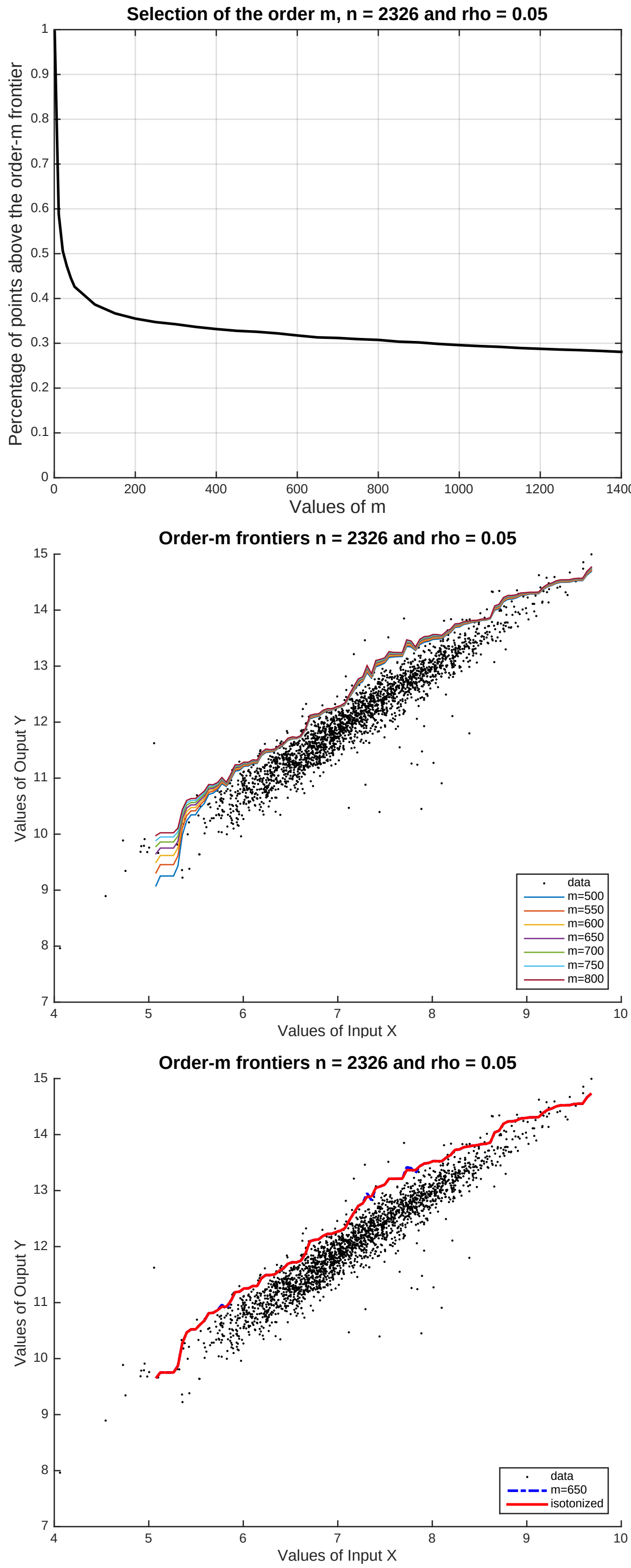

Figure 9: *Final results for Case I:* $n = 2326$ *and* $\rho_{\text{nts}} = 0.05$.

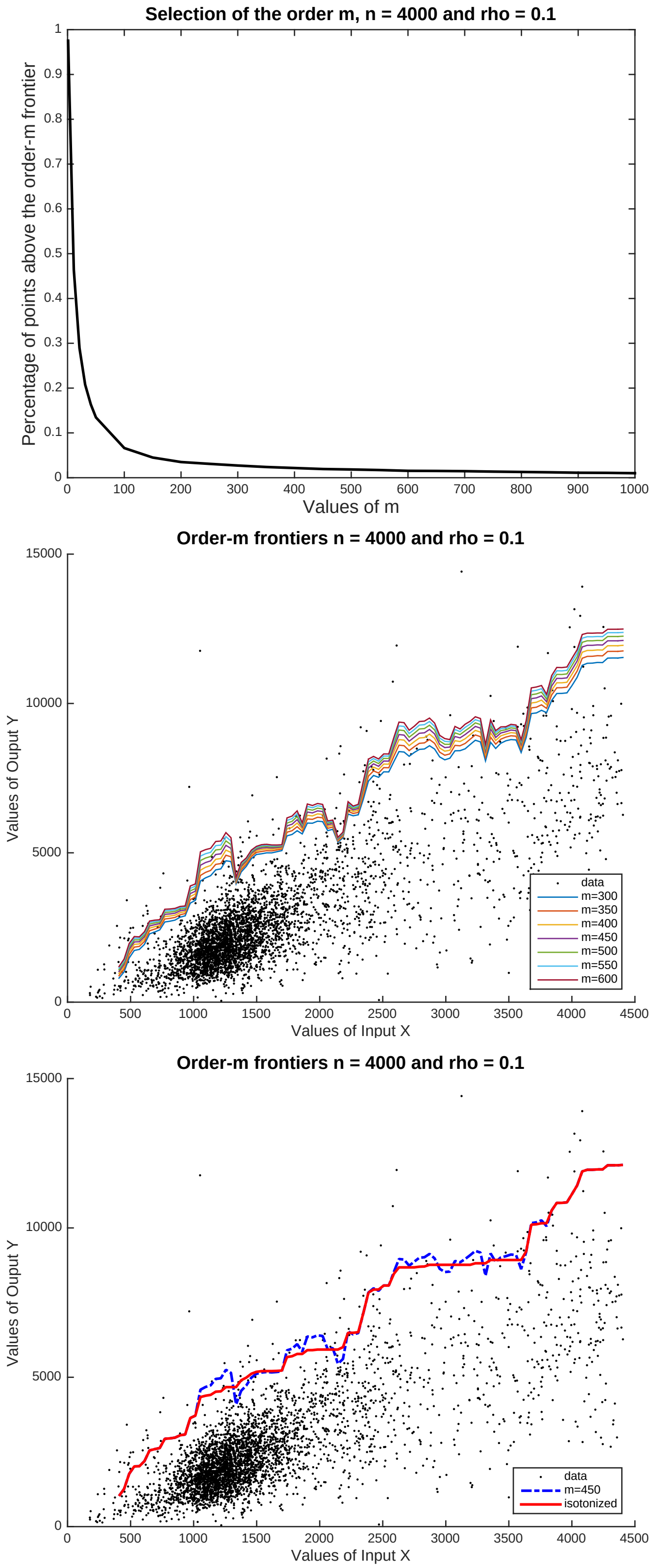

Figure 10: *Final results for Case II:* $n = 4000$ *and* $\rho_{\text{nts}} = 0.10$.

6. Conclusions

A new approach is suggested to estimate nonparametrically and in a robust way stochastic frontier functions. We suppose that the noise has a given density (*e.g.* Gaussian) to ensure the identification of the model. For a prespecified level of inputs of interest x, the basic idea is to first transform the $(p+1)$-dimensional random vector (X, Y) into a dimensionless variable Y^x, and then employ deconvolution techniques in conjunction with a Tikhonov regularization to estimate the underlying unconditional survivor function S_{Y^x}. By integrating powers of the latter, we get robust estimators of the partial m-frontier functions as well as the true full production function (corresponding to the limiting case $m \to \infty$). As in most studies on deconvolution, we suppose in this first work that the variance of the noise is known, and derive under some regularity conditions the rate of convergence of quadratic risk for the proposed estimators as well as their asymptotic distributions. The practical implementation of the presented procedure is first described through a simulated example. Then we analyze the expected maximum production and the optimal production function itself in the sector of postal services by exploring two concrete datasets on delivery offices. Through this application we highlight the usefulness and the flexibility of our device even if the variance of the noise is unknown.

The difficult question of estimating simultaneously the frontier function and the variance parameter of the noise is a topic of interest for future research. The difficulty of this more general problem comes from the heavy dependence of the operator K, defining the integral equation, on the unknown variance. Yet, at this stage of our research, we suggest to apply a heuristic method, based on the evaluation of the quality of the fit of the observed data, in order to select in practice a reasonable estimate for $\sigma^2(x)$. The simulated data examples and the application to the two real datasets provide very promising results, but further theoretical research remains to be done for this idea to receive due appreciation.

Appendix

The proofs of all theoretical results are provided in Section A.1. Some examples illustrating the source condition **(H.3)** are presented in Section A.2.

A.1. Proofs

Proof of Theorem 1. The risk of our estimator can be decomposed into two terms:

$$\mathbb{E}\left(||\widehat{S}^{\alpha}_{Y^x} - S_{Y^x}||^2\right) = ||S^{\alpha}_{Y^x} - S_{Y^x}||^2 + \mathbb{E}\left(||\widehat{S}^{\alpha}_{Y^x} - S^{\alpha}_{Y^x}||^2\right),$$

where the first element is the square of a bias term introduced by the regularization and the second element is a variance term.

(i) Analysis of the bias term (Proof of Lemma 1): by making use of (13) we obtain

$$\begin{aligned} S^{\alpha}_{Y^x} - S_{Y^x} &= (\alpha I + K^*K)^{-1}K^*KS_{Y^x} - S_{Y^x} \\ &= (\alpha I + K^*K)^{-1}\big(K^*K - (\alpha I + K^*K)\big)S_{Y^x} \\ &= -\alpha(\alpha I + K^*K)^{-1}S_{Y^x}. \end{aligned}$$

Therefore, by using the SVD and the notations introduced above, we have $(\alpha I + K^*K)\phi_j = (\alpha + \lambda_j^2)\phi_j$, so that the eigenvalues of $(\alpha I + K^*K)^{-1}$ are $(\alpha + \lambda_j^2)^{-1}$. Since for any $\delta \in \mathcal{E}$, $||\delta||^2 = \sum_j \langle \delta, \phi_j \rangle^2$, it is easy to show that

$$||S^{\alpha}_{Y^x} - S_{Y^x}||^2 = \alpha^2 \sum_j \frac{\langle S_{Y^x}, \phi_j \rangle^2}{(\alpha + \lambda_j^2)^2}.$$

By Assumption **(H.3)**, and using the fact that $\big\langle (K^*K)^{\beta/2}\delta, \phi_j \big\rangle = \big\langle \delta, (K^*K)^{\beta/2}\phi_j \big\rangle$ (because K^*K is auto-adjoint), we obtain

$$\begin{aligned} ||S^{\alpha}_{Y^x} - S_{Y^x}||^2 &= \alpha^2 \sum_j \frac{\lambda_j^{2\beta}}{(\alpha + \lambda_j^2)^2} \langle \delta, \phi_j \rangle^2 \\ &\leqslant \alpha^{\beta} \sum_j \langle \delta, \phi_j \rangle^2 = O(\alpha^{\beta}). \end{aligned}$$

(ii) Analysis of the variance term (Proof of Lemma 2): We have

$$\widehat{S}^{\alpha}_{Y^x} - S^{\alpha}_{Y^x} = (\alpha I + K^*K)^{-1}K^*\big[(\widehat{S}_{n,Z^x} - \widehat{S}_{n,\varepsilon^x}) - (S_{Z^x} - S_{\varepsilon^x})\big].$$

First note that $\sqrt{n}\big[(\widehat{S}_{n,Z^x} - \widehat{S}_{n,\varepsilon^x}) - (S_{Z^x} - S_{\varepsilon^x})\big]$ converges in the Hilbert space $\mathcal{E}$ to a zero mean Gaussian process with a variance operator Σ described in (25) [see Cazals *et al.* (2002)]. This variance is "trace class", *i.e.*, its trace is finite, or for any basis of $\mathcal{E}$, $\sum_j \langle \Sigma\phi_j, \phi_j \rangle = \text{tr}\Sigma < \infty$. Second, we have

$$\begin{aligned} \mathbb{E}\left(||\widehat{S}^{\alpha}_{Y^x} - S^{\alpha}_{Y^x}||^2\right) &= \text{trVar}\big(\widehat{S}^{\alpha}_{Y^x} - S^{\alpha}_{Y^x}\big) \\ &= O\Big(\frac{1}{n}\text{tr}\big[(\alpha I + K^*K)^{-1}K^*\Sigma K(\alpha I + K^*K)^{-1}\big]\Big) \\ &= O\Big(\frac{1}{n}\sum_j \big\langle (\alpha I + K^*K)^{-1}K^*\Sigma K(\alpha I + K^*K)^{-1}\phi_j, \phi_j \big\rangle\Big) \\ &= O\Big(\frac{1}{n}\sum_j \frac{\lambda_j^2}{(\alpha + \lambda_j^2)^2} \langle \phi_j, \phi_j \rangle\Big). \end{aligned}$$

Since $\lambda_j^2/(\alpha + \lambda_j^2)^2 = O(1/\alpha)$, we get $\mathbb{E}\left(||\widehat{S}^{\alpha}_{Y^x} - S^{\alpha}_{Y^x}||^2\right) = O\Big(\frac{1}{\alpha n}\Big)$.

(iii) Bound for the risk: Finally, we obtain the following order of the risk for our estimator

$$\mathbb{E}\left(||\widehat{S}^{\alpha}_{Y^x} - S_{Y^x}||^2\right) = O\Big(\alpha^{\beta} + \frac{1}{n\alpha}\Big).$$

We see indeed that when $\alpha \to 0$, the contribution of the variance term increases but the contribution of the bias term decreases, so we need $\alpha \to 0$ and $\alpha n \to \infty$ to get consistency. As usual, an optimal value for α will be found in this situation by balancing the squared bias and the variance. This results in $\alpha = O\big(n^{-1/(\beta+1)}\big)$ giving a risk bounded by $O\big(n^{-\beta/(\beta+1)}\big)$. This completes the proof of the theorem. ■

Proof of Theorem 3. We have

$$\widehat{\varphi}_m^\alpha(x) - \varphi_m(x) = \int_0^\tau \big\{[1 - S_{Y^x}(y)]^m - [1 - \widehat{S}_{Y^x}^\alpha(y)]^m\big\}\, dy.$$

A Taylor expansion of $[1 - S_{Y^x}(y)]^m - [1 - \widehat{S}_{Y^x}^\alpha(y)]^m$ leads to

$$\widehat{\varphi}_m^\alpha(x) - \varphi_m(x) = m \int_0^\tau [1 - S_{Y^x}(y)]^{m-1}\big\{\widehat{S}_{Y^x}^\alpha(y) - S_{Y^x}(y)\big\}\, dy + r_{m,n}, \tag{A.1}$$

where

$$r_{m,n} = -\frac{1}{2} m(m-1) \int_0^\tau \big\{\widehat{S}_{Y^x}^\alpha(y) - S_{Y^x}(y)\big\}^2 [b_x(y)]^{m-2}\, dy,$$

with $[1 - \widehat{S}_{Y^x}^\alpha(y)] \wedge [1 - S_{Y^x}(y)] \leqslant b_x(y) \leqslant [1 - \widehat{S}_{Y^x}^\alpha(y)] \vee [1 - S_{Y^x}(y)]$. Since $0 \leqslant b_x(y) \leqslant 1$, we have $|r_{m,n}| \leqslant \frac{1}{2} m^2 ||\widehat{S}_{Y^x}^\alpha - S_{Y^x}||^2$, and hence

$$r_{m,n} = m^2 O_p\big(n^{-\beta/(\beta+1)}\big) \tag{A.2}$$

in view of Theorem 1. On the other hand, we have

$$\left| m \int_0^\tau [1 - S_{Y^x}(y)]^{m-1}\big\{\widehat{S}_{Y^x}^\alpha(y) - S_{Y^x}(y)\big\} dy \right|^2 \leqslant m^2\, \tau\, ||\widehat{S}_{Y^x}^\alpha - S_{Y^x}||^2.$$

Therefore

$$\begin{aligned}
|\widehat{\varphi}_m^\alpha(x) - \varphi_m(x)| &\leqslant \left| m \int_0^\tau [1 - S_{Y^x}(y)]^{m-1}\big\{\widehat{S}_{Y^x}^\alpha(y) - S_{Y^x}(y)\big\} dy \right| + |r_{m,n}| \\
&\leqslant m\, \tau^{1/2}\, ||\widehat{S}_{Y^x}^\alpha - S_{Y^x}|| + \frac{1}{2} m^2 ||\widehat{S}_{Y^x}^\alpha - S_{Y^x}||^2 \\
&\leqslant m\, \tau^{1/2}\, O_p\big(n^{-\beta/2(\beta+1)}\big) + m^2 O_p\big(n^{-\beta/(\beta+1)}\big).
\end{aligned}$$

Then for fixed m, we have $\mathbb{E}\,|\widehat{\varphi}_m^\alpha(x) - \varphi_m(x)|^2 = O\big(n^{-\beta/(\beta+1)}\big)$. If $m = m(n) \to \infty$ such that $m = O(n^{\beta/(2(\beta+1))})$, then $\mathbb{E}\,|\widehat{\varphi}_m^\alpha(x) - \varphi_m(x)|^2 = m^2 O\big(n^{-\beta/(\beta+1)}\big)$. ■

Proof of Theorem 4. By (A.1) we have

$$\sqrt{n}\big(\widehat{\varphi}_m^\alpha(x) - \varphi_m(x)\big) = m\sqrt{n} \int_0^\tau [1 - S_{Y^x}(y)]^{m-1}\big\{\widehat{S}_{Y^x}^\alpha(y) - S_{Y^x}(y)\big\}\, dy + \sqrt{n} r_{m,n},$$

where it follows from (A.2) that

$$\sqrt{n}\, r_{m,n} = m^2 \sqrt{n}\, O_p\big(n^{-\beta/(\beta+1)}\big) = m^2 O_p\big(n^{(1-\beta)/2(\beta+1)}\big). \tag{A.3}$$

On the other hand, the leading term in the decomposition can be written as

$$\begin{aligned}
&m\sqrt{n} \int_0^\infty \mathbb{1}(y \leqslant \tau) F_{Y^x}^{m-1}(y)\big\{\widehat{S}_{Y^x}^\alpha(y) - S_{Y^x}(y)\big\}\, dy \\
&= m < \sqrt{n}\big(\widehat{S}_{Y^x}^\alpha - S_{Y^x}\big), \mathbb{1}(\cdot \leqslant \tau) F_{Y^x}^{m-1} > .
\end{aligned}$$

Putting $\delta(y) := \mathbb{1}(y \leqslant \tau) F_{Y^x}^{m-1}(y)$, we have $\delta \in \mathcal{E}$. Under the conditions of Theorem 2, if δ satisfies **(K.1)**and **(K.2)**, then

$$\frac{m < \sqrt{n}\big(\widehat{S}_{Y^x}^\alpha - S_{Y^x}\big), \delta >}{m < (\alpha I + K^* K)^{-1} K^* \Sigma K (\alpha I + K^* K)^{-1} \delta, \delta >^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \quad n \to \infty,$$

or equivalently

$$\frac{\sqrt{n}\big(\widehat{\varphi}^{\alpha}_m(x)-\varphi_m(x)\big)-\sqrt{n}\,r_{m,n}}{m<(\alpha I+K^*K)^{-1}K^*\Sigma K(\alpha I+K^*K)^{-1}\delta,\delta>^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \quad n\to\infty.$$

Therefore, if

$$\frac{\sqrt{n}\,r_{m,n}}{m<(\alpha I+K^*K)^{-1}K^*\Sigma K(\alpha I+K^*K)^{-1}\delta,\delta>^{1/2}} = o_p(1), \quad n\to\infty, \tag{A.4}$$

we get immediately the asymptotic normality of $\sqrt{n}\big(\widehat{\varphi}^{\alpha}_m(x)-\varphi_m(x)\big)$. Since $m\geqslant 1$ is fixed and $\alpha\to 0$ as $n\to\infty$, then (A.4) holds as long as $\sqrt{n}r_{m,n}=o_p(1)$, which in turn happens if $\beta>1$ in view of (A.3). ■

Proof of Theorem 5. We have

$$\mathbb{E}\Big[\big(\widehat{\varphi}^{\alpha}_m(x)-\varphi(x)\big)^2\Big] \leqslant 2\mathbb{E}\Big[\big(\widehat{\varphi}^{\alpha}_m(x)-\varphi_m(x)\big)^2\Big]+2\big(\varphi_m(x)-\varphi(x)\big)^2.$$

Under Assumption **(K.3)**, we have seen in (31) that

$$\varphi(x)-\varphi_m(x)=(m\ell_x)^{-1/\rho_x}\Gamma(1+\rho_x^{-1})+o\big(m^{-1/\rho_x}\big), \quad m\to\infty.$$

Then $\big(\varphi_m(x)-\varphi(x)\big)^2 = O\big(m^{-2/\rho_x}\big)$, as $m\to\infty$. On the other hand, we have by Theorem 3 that $\mathbb{E}\Big[\big(\widehat{\varphi}^{\alpha}_m(x)-\varphi_m(x)\big)^2\Big]=O\big(m^2\,n^{-\beta/(\beta+1)}\big)$, for any sequence $m=m(n)\to\infty$ such that $m=O(n^{\beta/(2(\beta+1))})$. Thus

$$\mathbb{E}\Big[\big(\widehat{\varphi}^{\alpha}_m(x)-\varphi(x)\big)^2\Big]=O\left(m^2\,n^{-\beta/(\beta+1)}+m^{-2/\rho_x}\right), \quad n\to\infty.$$

While the variance term $m^2\,n^{-\beta/(\beta+1)}$ increases with m, the squared bias term m^{-2/ρ_x} (introduced by using a partial m-frontier to estimate the full frontier) decreases with m. Balancing both terms gives the following optimal order for m, as a function of n,

$$m=m(n)=\big(\rho_x^{-1}\,n^{\frac{\beta}{\beta+1}}\big)^{\frac{\rho_x}{2(1+\rho_x)}}.$$

The corresponding risk is given by $O\big\{n^{-\frac{\beta}{\beta+1}\frac{1}{1+\rho_x}}\big\}$. ■

A.2. Illustrating examples of Assumption **(H.3)**

We give here two examples illustrating our Assumption **(H.3)**. The first one provides a full analytical treatment, and the second one involves a complex expression that can easily be treated numerically.

Example 1. Consider for a fixed value of x the survivor function

$$S_{Y^x}(y)=\begin{cases}1 & \text{for } y\in[0,1]\\ 1-(y-1)^3 & \text{for } y\in[1,2]\\ 0 & \text{for } y\in[2,\infty).\end{cases}$$

We choose ε given $X\leqslant x$ to be uniform on $[-1,1]$, and the weight function in $\mathcal{F}$ to be uniform on $[-1,1]$. It is not hard to check that this survival function belongs to $\text{Range}(K^*K)^{\beta/2}$ with $\beta=1$. For this, one has to prove that $S_{Y^x} \in \text{Range}(K^*)$ or, equivalently, that $S_{Y^x} = K^*\psi$ for some $\psi \in \mathcal{F}$. Indeed, an elementary calculus shows that if the function ψ is defined as

$$\psi(z) = \begin{cases} 0 & \text{for } z \leqslant 0 \\ 3z^2 & \text{for } z > 0, \end{cases}$$

then the survivor function can be written, for any $y \geqslant 0$, as

$$S_{Y^x}(y) = \int_{-\infty}^{\infty} \psi(z) \mathbb{1}(z - x \in [-1,1]) \, \mathbb{1}(z \in [-1,1]) \, dz.$$

Example 2. Here we consider a sophisticated analytical framework, where the distribution of ε given $X \leqslant x$ is $N(0,\sigma^2)$ and the weight function in $\mathcal{F}$ is given by a normal density with mean 0 and variance τ^2. We assume that the frontier point is $\varphi(x) = 0$, so that the survivor function starts from zero. By analytical developments using (21) we then arrive at

$$c(y,\xi) = (2\pi)^{-1} \frac{1}{a\sigma\tau} \exp\left\{-\frac{1}{2\sigma^2}\left(y^2 + \xi^2 - \frac{y^2+\xi^2}{a^2}\right)\right\},$$

where $a^2 = 2 + \sigma^2/\tau^2 > 2$. Now we have to show that there exists a function $\phi \in \mathcal{E}$ such that $K^*K\phi$ is a survivor function S_{Y^x}. If such is the case, we would have by construction $S_{Y^x} \in \text{Range}(K^*K)^{\beta/2}$ with $\beta = 2$. It can be shown after some analytical manipulations that

$$K^*K\phi(y) = (2\pi)^{-1} \frac{1}{a\sigma\tau} e^{-by^2} \tilde{\sigma} \Phi(\tilde{\mu}/\tilde{\sigma}) \sqrt{2\pi} \mathbb{E}\big[\phi(W)\big],$$

where $b = \frac{(a^2-2)}{2\sigma^2(a^2-1)} > 0$, $\Phi(\cdot)$ is the standard normal distribution function and $W \sim N^+(\tilde{\mu}, \tilde{\sigma}^2)$ is a truncated normal ($\geqslant 0$) random variable, with $\tilde{\mu} = y/(a^2-1)$ and $\tilde{\sigma} = a\sigma/\sqrt{a^2-1}$. By choosing, for instance, $\phi(u) = c_0 e^{-tu}$ for some constants $t, c_0 > 0$, the expectation in the last equation is $c_0 m_W(-t)$, where $m_W(\cdot)$ is the moment generating function of W. Thus, we find after some calculations that

$$K^*K\phi(y) = C(t,c_0) \exp\big\{ - by^2 - \frac{t}{a^2-1} y \big\} \Phi\left(\frac{y}{a\sigma\sqrt{a^2-1}} - \frac{a\sigma}{\sqrt{a^2-1}} t\right),$$

where $C(t,c_0) = c_0 \frac{1}{\tau\sqrt{2\pi(a^2-1)}} e^{(\sigma a t)^2/(2(a^2-1))}$ is a positive constant. Now, we have to prove that $S_{Y^x}(y) := K^*K\phi(y)$ is a survivor function on $\mathbb{R}_+$. For any t, the constant c_0 can be tuned to get $S_{Y^x}(0) = 1$. Clearly, $S_{Y^x}(y) > 0$ and $\lim_{y\to\infty} S_{Y^x}(y) = 0$. It remains to show that for an appropriate choice of t, we have $g(y) := S'_{Y^x}(y) < 0$ for all $y > 0$. Without loss of generality, we can fix $\sigma = 1$ and easily check that

$$\begin{aligned} g(y) = C(t,c_0) e^{-by^2 - \frac{t}{a^2-1} y} \Bigg[& \frac{1}{a\sqrt{a^2-1}} \Phi'\left(\frac{y}{a\sqrt{a^2-1}} - \frac{a}{\sqrt{a^2-1}} t\right) \\ & - \left(2by + \frac{t}{a^2-1}\right) \Phi\left(\frac{y}{a\sigma\sqrt{a^2-1}} - \frac{a\sigma}{\sqrt{a^2-1}} t\right)\Bigg], \end{aligned}$$

where $\Phi'(\cdot)$ is the standard normal density function. For instance, with $a = 4$ and $t = 0.9$, we find that $g(0) = -0.1376$ and the function $g(y)$ is non-positive for all $y \geqslant 0$, as shown on the left panel of Figure .11. In this case, the function $S_{Y^x}(y)$ is well decreasing and can be evaluated numerically. Its plot is displayed on the right panel of Figure .11.

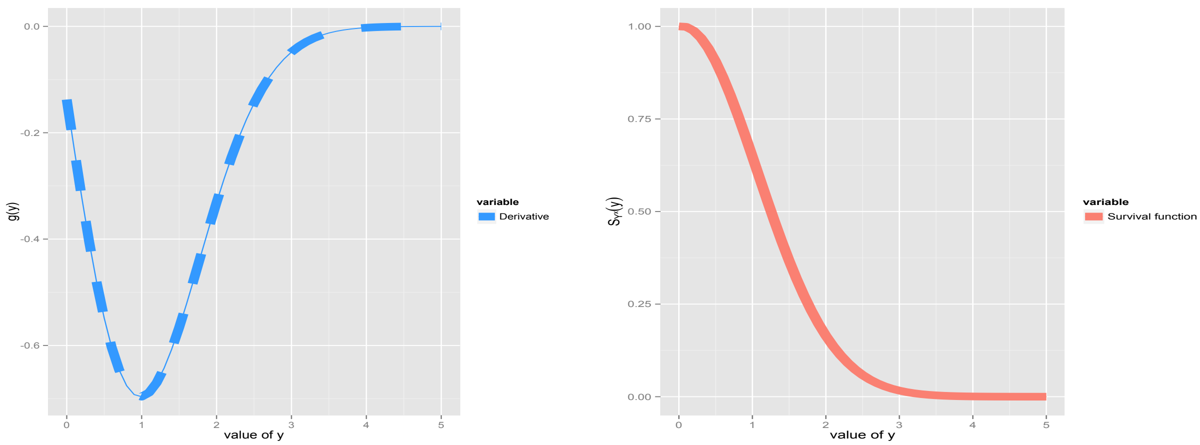

Figure .11: *Left panel—the function $g(y)$ giving the derivative of $S_{Y^x}(y)$. Right panel—the survivor function $S_{Y^x}(y)$.*

Acknowledgements

The authors would like to thank two referees, the Associate Editor and the Co-Editor for their valuable suggestions, which have significantly improved the paper. This research was supported by the Toulouse School of Economics Individual Research Fund (IRF/Daouia-20125), and by the French "Agence Nationale pour la Recherche" under grants ANR-08-BLAN-0106-01/EPI project and ANR-13-BSHE-0004-03. Support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is also gratefully acknowledged.

References

[1] Aigner, D.J., Lovell, C.A.K. and P. Schmidt (1977), Formulation and estimation of stochastic frontier models, *Journal of Econometrics*, 6, 21-37.

[2] Butucea, C., Matias, C. and C. Pouet (2008), Adaptativity in convolution models with partially known noise distribution, *Electronic Journal of Statistics*, 2, 897–915.

[3] Carrasco, M. and J.P. Florens (2009), Spectral method for deconvolving a density, *Econometric Theory*, 27, 3, 546–581.

[4] Carrasco, M., J.P. Florens and E. Renault (2007), Linear inverse problems in structural econometrics: estimation based on spectral decomposition and regularization. In J.J. Heckman and E.E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, North Holland Amsterdam, 5633–5746.

[5] Carrasco, M., J.P. Florens and E. Renault (2014), Asymptotic Normal Inference in Linera Inverse Problems, in J.S. Racine, L. Su and A. Ullah (Eds), *Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford Press, 65–96.

[6] Carroll, R.J. and P. Hall (1988), Optimal rates of convergence for deconvolving a density, *Journal of the American Statistical Association*, 83, 1184–1186.

[7] Cazals, C., J.P. Florens and L. Simar (2002), Nonparametric frontier estimation: a robust approach, *Journal of Econometrics*, 106, 1-25.

[8] Daouia, A., J.P. Florens and L. Simar (2010), Frontier estimation and Extreme value theory, *Bernoulli*, Vol. 16, No. 4, 1039–1063.

[9] Daouia, A., J.P. Florens and L. Simar (2012), Regularization of non-parametric frontier estimators, *Journal of Econometrics*, 168, 285–299.

[10] Daouia, A., J.P. Florens and L. Simar (2018), Robustified expected maximum production frontiers, Submitted. Available at `https://www.tse-fr.eu/people/abdelaati-daouia?tab=working-papers`.

[11] Daouia, A., S. Girard and A. Guillou (2014), A Γ-moment approach to monotonic boundary estimation, *Journal of Econometrics*, 78, 727–740.

[12] Daouia, A. and B.U. Park (2013), On Projection-Type Estimators of Multivariate Isotonic Functions, *Scandinavian Journal of Statistics*, 40, 363–386.

[13] Daouia, A. and L. Simar (2005), Robust Nonparametric Estimators of Monotone Boundaries, *Journal of Multivariate Analysis*, 96, 311–331.

[14] Deprins, D., Simar, L. and H. Tulkens (1984), Measuring labor inefficiency in post offices. In *The Performance of Public Enterprises: Concepts and measurements*. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243–267.

[15] Dunker, F., Florens, J.P., Hohage, T., Johannes, J. and Mammen, E. (2014). Iterative Estimation of Solutions to Noisy Nonlinear Operator Equations in Nonparametric Instrumental Regression, *Journal of Econometrics*, 178, 444–455.

[16] Engl, H.W., Hanke, M and A. Neubauer (2000), *Regularization of Inverse Problems*, Kluwer, Dordrecht.

[17] Fan, J. (1991a), Gobal behavior of deconvolution kernel estimates, *Statistica Sinica*, 541–551.

[18] Fan, J. (1991b), On the optimal rates of convergence for nonparametric deconvolution problems, *The Annals of Statistics*, 19, 3, 1257–1272.

[19] Fan, J., Q. Li and A. Weersink (1996), Semiparametric estimation of stochastic production frontier models, *Journal of Business and Economic Statistics*, 14, 460–468.

[20] Fève, F. and J.P. Florens (2014), Nonparametric analysis of Panel Data models with endogenous variables, in press *Journal of Econometrics*.

[21] Gijbels, I. and Peng, L. (2000), Estimation of a Support Curve via Order Statistics. *Extremes*, 3(3), 251–277.

[22] Greene, W.H. (2008), The Econometric Approach to Efficiency Analysis, in *The Measurement of Productive Efficiency*, 2nd Edition, Harold Fried, C.A.Knox Lovell and Shelton Schmidt, editors, Oxford University Press.

[23] Hall, P. and A. Meister (2007), A Ridge-Parameter Approach to Deconvolution, *Annals of Statistics*, 35, 1535–1558.

[24] Hall, P. and Park, B.U. and Stern, S.E. (1998), On polynomial estimators of frontiers and boundaries, *Journal of Multivariate Analysis*, 66(1), 71–98.

[25] Hall, P. and L. Simar (2002), Estimating a Changepoint, Boundary or Frontier in the Presence of Observation Error, *Journal of the American Statistical Association*, 97, 523–534.

[26] Hardle, W. and Park, B.U. and Tsybakov, A.B. (1995), Estimation of non-sharp support boundaries, *Journal of Multivariate Analysis*, 43, 205–218.

[27] Horrace, W.C. and Ch.F. Parmeter (2011), Semiparametric deconvolution with unknown variance, *Journal of Productivity Analysis*, 35, 2, 129–141.

[28] Hwang, J.H., B.U. Park and W. Ryu (2002), Limit theorems for boundary function estimators, *Statistics & Probability Letters*, 59, 353–360.

[29] Kneip, A., Simar, L. and I. Van Keilegom (2015), Frontier estimation in the presence of measurement error with unknown variance, *Journal of Econometrics*, 184, 379–393.

[30] Kumbhakar, S.C. , Park, B.U., Simar, L. and E.G. Tsionas (2007), Nonparametric stochastic frontiers: a local likelihood approach, *Journal of Econometrics*, 137, 1, 1–27.

[31] Kreiss, R. (1999), *Linear Integral Equations*, Applied Mathematical Sciences, New-York, Springer Verlag.

[32] Lukacs, E. (1960) *Characteristic Functions*, Charles Griffin & Company Ltd, London.

[33] Meeusen, W. and J. van den Broek (1977), Efficiency estimation from Cobb-Douglas production function with composed error, *International Economic Review*, 8, 435–444.

[34] Meister, A. (2006), Density estimation with normal measurement error with unknown variance, *Statistica Sinica*, 16, 195–211.

[35] Park, B. Simar, L. and Ch. Weiner (2000), The FDH Estimator for Productivity Efficiency Scores : Asymptotic Properties, *Econometric Theory*, 16, 855–877.

[36] Schwarz, M. and S. Van Bellegem (2010), Consistent density deconvolution under partially known error distribution, *Statistics and Probability Letters*, 80, 236–241.

[37] Simar, L., Van Keilegom, I. and V. Zelenyuk (2017), Nonparametric Least Squares Methods for Stochastic Frontier Models, *Journal of Productivity Analysis*, 47, 189–204.

[38] Simar, L. and V. Zelenyuk (2011), Stochastic FDH/DEA estimators for frontier analysis. *Journal of Productivity Analysis*, 36, 1–20.

[39] Stefanski, L.A. and R.J. Carroll (1990), Deconvoluting kernel density estimators, *Statistics*, 21, 169–184.

[40] Van Rooij, A.C.M. and F.H. Ruymgaart (1999), On inverse estimation, in *Asymptotics, Nonparametrics and Time Series*, 579–613, Dekker, New-York.

[41] Wheelock, D.C., and Wilson, P.W. (2008), Non-parametric, Unconditional Quantile Estimation for Efficiency Analysis with an Application to Federal Reserve Check Processing Operations, *Journal of Econometrics*, 145 (1-2), 209–225.