ISSN 2281-4299

DIPARTIMENTO DI INGEGNERIA INFORMATICA
AUTOMATICA E GESTIONALE ANTONIO RUBERTI

SAPIENZA
UNIVERSITÀ DI ROMA

A Bootstrap Approach for Bandwidth Selection in Estimating Conditional Efficiency Measures

Luiza Badin
Cinzia Daraio
Léopold Simar

Technical Report n. 2, 2018

A Bootstrap Approach for Bandwidth Selection in Estimating Conditional Efficiency Measures

Luiza Bădin*
luiza.badin@csie.ase.ro

Cinzia Daraio§
daraio@diag.uniroma1.it.

Léopold Simar§,**
leopold.simar@uclouvain.be

March 05, 2018

Abstract

Conditional efficiency measures are needed when the production process does not depend only on the inputs and outputs, but may be influenced by external factors and/or environmental variables (Z). They are estimated by means of a nonparametric estimator of the conditional distribution function of the inputs and outputs, conditionally on values of Z. For doing this, smoothing procedures and smoothing parameters, the bandwidths, are involved. So far, Least Squares Cross Validation (LSCV) methods have been used, which have been proven to provide bandwidths with optimal rates for estimating conditional distributions. In efficiency analysis, the main interest is in the estimation of the conditional efficiency score, which typically depends on the boundary of the support of the distribution and not on the full conditional distribution. In this paper, we show indeed that the rate for the bandwidths which is optimal for estimating conditional distributions, may not be optimal for the estimation of the efficiency scores. We propose hence a new approach based on the bootstrap which overcomes these difficulties. We analyze and compare, through Monte Carlo simulations, the performances of LSCV techniques with our bootstrap approach in finite samples. As expected, our bootstrap approach shows generally better performances and is more robust to the various Monte Carlo scenarios analyzed. We provide in an Appendix the Matlab code performing our experiments.

Key Words: Data Envelopment Analysis (DEA)/Free Disposal Hull (FDH); Conditional Efficiency; Bandwidth; Bootstrap; Monte Carlo.

*Department of Applied Mathematics, Bucharest University of Economic Studies and *Gh. Mihoc-C. Iacob* Institute of Mathematical Statistics and Applied Mathematics, Bucharest, Romania. Financial support from the Romanian National Authority for Scientific Research and Innovation, CNCS-UEFISCDI, project number PN-II-RU-TE-2014-4-2905, is gratefully acknowledged.

§Department of Computer, Control and Management Engineering A. Ruberti (DIAG) Sapienza University of Rome, Italy. Financial support from the Project Sapienza 2015 Awards, N. 6H15XNFS, FILAS RU 2014-1186, PRIN 2015 (2015RJARX7) and Sapienza 2017 Awards, N. PH11715C8239C105 is gratefully acknowledged.

**Institut de Statistique, Biostatistique et de Sciences Actuarielles, Université Catholique de Louvain, Voie du Roman Pays 20, B 1348 Louvain-la-Neuve, Belgium. Financial support from the Inter-university Attraction Pole, Phase VII (No.P7/06) of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

1 The Background

In order to boost the economic performance of productive units, one is interested in developing management strategies that can lead to increasing the units efficiency (or technical efficiency). An important aspect is the explanation of differences in the efficiency levels achieved by economic producers which are facing different environmental and external conditions (see Bădin et al. 2012, 2014). Nonparametric conditional frontier models include exogenous variables or environmental factors that may influence the production process, having a compound impact, affecting on the one hand, the range of values for input×output vectors including the shape of the boundaries and, on the other hand, the distribution of the efficiencies. The conditional frontier approach became very popular due to its direct and natural manner of defining conditional efficiency scores and so, providing a formal way for handling heterogeneity. First introduced by Cazals et al. (2002) and further extended by Daraio and Simar (2005, 2007a, 2007b), the approach is based on defining a Data Generating Process (DGP) including a probabilistic model that characterizes the production process in the presence of heterogeneous conditions.

Summing up, recent theoretical and empirical studies on conditional frontier models highlight the importance of conditional efficiency measures as a general fully nonparametric way to treat appropriately the presence of environmental factors in a production process (see Bădin et al., 2012, 2014 and the references therein). The conditional approach has been applied to university rankings (see e.g. Daraio et al., 2015) but also in macroeconomic setups (see e.g. Mastromarco and Simar, 2015). The approach has also been used to handle quality variables in the production process (see e.g. for the efficiency in the hospital sector, Varabyova et al. 2016a, 2016b and Varabyova and Schreyögg, 2017).

To have an idea of the variety of fields where conditional efficiency measures have been used, see Table 1 which shows references of applications in regional innovation, environment, water, municipalities, public services and culture.

The bandwidths for the conditioning variables play a crucial role in the process of estimating these measures since they "tune" the localization for computing the conditional efficiencies (FDH and/or DEA). Statistical theory, so far, was based on results from Hall et al. (2004), Li and Racine (2008). These approaches use Least Squares Cross Validation (LSCV) techniques sharing some nice optimality properties. They have been adapted to frontier analysis in Bădin et al. (2010) and involve the estimation of a nonstandard conditional Probability Density Function (PDF), nonstandard in the sense that, e.g. in the output orientation, the condition is on the value considered for the external factor and on an inequality for the input vector. Recent works from Simar et al. (2016) have stressed the possibility of improving the properties of the conditional efficiency scores by estimating the joint PDF of the inputs and outputs by conditioning only on the external

Table 1: *A selective survey on empirical applications of conditional efficiency measures.*

Field of application	*Reference*
Regional innovation	Broekel (2012), Broekel and Schlump (2009), Filippetti and Peyrache (2015)
Education	Haelermans and De Witte (2012), Daraio et al. (2015), Cordero et al. (2017)
Environment	Halkos and Tzeremes (2010, 2013a, 2014), Halkos and Managi (2016)
Water	Carvalho and Marques (2011), Zschille (2015), Fuentes et al. (2015), Guerrini et al. (2016)
Health	Halkos and Tzeremes (2011), Cordero et al. (2015) Varabyova et al. 2016a, 2016b , Varabyova and Schreyögg, 2016
Banking	Tzeremes (2015), Matousek and Tzeremes (2016), Bădin et al. (2012)
Macroeconomics	Mastromarco and Simar (2015)
Agriculture	Kourtesi et al. (2012) Serra and Lansink (2014), Baležentis and De Witte (2015)
Public services	Verschelde and Rogge (2012), De Witte and Geys (2011)
Culture	Halkos and Tzeremes (2013b,c)
Airports	D'Alfonso et al. (2015)
Municipalities	Cordero et al. (2016)
Mutual funds	Daraio and Simar (2006), Bădin and Daraio (2011), Bădin et al. (2014)

factors. Also, the latter involves much less computational burden. On the other hand, results from Li et al. (2013), suggest to estimate optimal bandwidths for conditional distribution directly, by evaluating a criterion based on the Cumulative Distribution Function (CDF). Their Monte Carlo experiments indicate the superiority of this approach, relative to the indirect one passing through the estimation of the conditional PDF, then correcting the order of the bandwidths when the final objective is to estimate the CDF. All these techniques are asymptotically equivalent and provide bandwidths having optimal rates. Since the latest results from Li et al. (2013) and Simar et al. (2016) suggest rather the use of the joint CDF of inputs and outputs and conditioning only on the external/environmental factors, we will follow this strategy in our paper.

The main goal of the paper is to describe in details the impact of the bandwidth choice on the object of interest, i.e. the efficiency score, which is determined by the upper (lower) boundary of the support of the conditional CDF, when output (input) oriented measures are estimated, and not by the conditional CDF itself in its full range. Therefore the objective to get an optimal estimation of conditional efficiency scores may be thus quite different from the objective of estimating the full conditional CDF, which is the target of all the LSCV approaches developed so far. In this paper, we show that the rate for the bandwidths which is optimal for estimating conditional distributions, may not be optimal for the estimation of the efficiency scores. As explained in the paper, the situation can be different if the external variables Z are separable or not. In real empirical applications one does not know in advance if separability holds. We propose in this paper a new approach, based on the bootstrap, which overcomes these difficulties, providing, by construction, optimal bandwidths in all the situations. It is based on a bootstrap estimator of the mean square error of the efficiency

estimators themselves.

The paper is organized as follows. Section 2 summarizes the definitions and introduces the notations for marginal and conditional efficiency scores and their nonparametric estimators. Section 3 introduces the elements which allow us to develop our new approach for bandwidth selection. In Section 4 we analyze through a Monte Carlo study the finite sample performances of the available bandwidth selection methods. As expected, our bootstrap approach shows generally better performances and is more robust to the various Monte Carlo scenarios analyzed. Section 5 summarizes the main contribution of the paper and concludes, while the Matlab code is reported in Appendix A.

2 The Production Process in the Presence of Environmental Factors

This section can easily be skipped by readers familiar to conditional efficiency scores and their nonparametric estimators. Details could be found, e.g. in Simar and Wilson (2007, 2011) and in Bădin et al. (2012, 2014).

The statistical model for production analysis in the presence of environmental factors is the set of assumptions describing the Data Generating Process (DGP) of triples (X, Y, Z) where $X \in \mathbb{R}^p$ is the vector of inputs, $Y \in \mathbb{R}^q$ is the vector of outputs and $Z \in \mathcal{Z} \subseteq \mathbb{R}^d$ is the vector of external factors or environmental variables. We consider $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space on which the random variables X, Y, Z are defined and we denote by $\mathcal{P}$ the support of the joint distribution of (X, Y, Z). The elements of Z are neither inputs nor outputs and are typically not under the control of the manager; they characterize heterogeneity conditions, but they may influence the production process in different ways, as explained below.

Let $f_{XYZ}(x, y, z)$ denote the pdf of (X, Y, Z) on $\mathcal{P}$. This joint density can always be decomposed as

$$f_{XYZ}(x, y, z) = f_{XY|Z}(x, y|z) f_Z(z), \tag{2.1}$$

where the notations are self-explanatory. Let Ψ^z denote the support of $f_{XY|Z}(x, y|z)$; it is the support of (X, Y) given that $Z = z$. Thus it is the attainable set for units facing external conditions $Z = z$:

$$\Psi^z = \{(x, y)|x \text{ can produce } y \text{ if } Z = z\}. \tag{2.2}$$

The variables Z can affect the production process either (i) only through Ψ^z the support of (X, Y), or (ii) only through the density $f_{XY|Z}(x, y|z)$, affecting only the probability of a unit to reach its optimal boundary, or (iii) through both Ψ^z and $f_{XY|Z}(x, y|z)$.

Let Ψ be the marginal support of (X, Y). By definition we have

$$\Psi = \{(x, y)|x \text{ can produce } y\} = \{(x, y)|f_{XY}(x, y) > 0\} = \bigcup_{z \in \mathcal{Z}} \Psi^z, \tag{2.3}$$

and by construction, $\Psi^z \subseteq \Psi$, for all $z \in \mathcal{Z}$.

If the joint support of (X, Y, Z) can be written as a cartesian product $\mathcal{P} = \Psi \times \mathcal{Z}$, then Z will not have an impact on the boundaries of Ψ and $\Psi^z = \Psi$ for all $z \in \mathcal{Z}$ (this is called the "separability condition" in this literature). In this very particular case, the only potential influence of Z on the production process, might be on the distribution of the efficiencies. In this case, the usual two-stage approaches are valid for investigating, with appropriate tools, the potential effects of Z (see Simar and Wilson 2007, 2011 for details). Otherwise, measuring the distance of a unit (x, y) to the boundary of Ψ is of little economic interest, since it ignores the heterogeneity introduced by Z on the attainable set of values for (X, Y). The conditional nonparametric approach does not rely on this restrictive assumption, conditional efficiencies derived below are defined in terms of the support of the conditional distribution which characterizes the production process when $Z = z$.

Marginal Farrell efficiency measures (output-oriented case[1]) for a unit operating at the level (x, y) can be defined as

$$\lambda(x, y) = \sup\{\lambda > 0|(x, \lambda y) \in \Psi\}, \tag{2.4}$$

whereas when the environmental conditions are $Z = z$, the same unit has a conditional efficiency scores (introduced by Cazals et al., 2002 and Daraio and Simar, 2005) defined as

$$\lambda(x, y|z) = \sup\{\lambda > 0|(x, \lambda y) \in \Psi^z\}. \tag{2.5}$$

In order to derive the nonparametric estimators below, the efficiency measure in (2.4) and (2.5) are better defined in terms of our probability model. It has been shown that under the free disposability assumption, we have for the marginal case

$$\lambda(x, y) = \sup\{\lambda > 0|H_{XY}(x, \lambda y) > 0\}, \tag{2.6}$$

where $H_{XY}(x, y) = \Pr(X \leq x, Y \geq y)$ is the marginal probability of finding a unit dominating the production plan (x, y). This can be factored as $\Pr(X \leq x)\Pr(Y \geq y|X \leq x) = F_X(x)S_{Y|X}(y|x)$, where the latter conditional survival function is nonstandard due to the condition $X \leq x$. For (x, y) such that x is in the interior of its support (i.e. $F_X(x) > 0$) the efficiency score can equivalently be defined as

$$\lambda(x, y) = \sup\{\lambda > 0|S_{Y|X}(\lambda y|X \leq x) > 0\}. \tag{2.7}$$

[1]We follow the presentation for the output-oriented case, but this can be easily translated to the input-oriented, the hyperbolic and the directional distance cases. See the recent survey Simar and Wilson (2015) and the references therein.

We have the analog decompositions for the conditional efficiency scores:

$$\lambda(x,y|z) = \sup\{\lambda > 0 | H_{XY|Z}(x, \lambda y|z) > 0\}, \tag{2.8}$$

where $H_{XY|Z}(x,y|z) = \Pr(X \leq x, Y \geq y | Z = z)$ is the conditional probability of finding a firm dominating the production plan (x,y), facing the same environmental conditions z. Along the same line as above this can also be written as

$$\lambda(x,y|z) = \sup\{\lambda > 0 | S_{Y|X,Z}(\lambda y | X \leq x, Z = z) > 0\}, \tag{2.9}$$

where here $S_{Y|X,Z}(\lambda y | X \leq x, Z = z) = \Pr(Y \geq y | X \leq x, Z = z)$, noting the different condition for the inputs X and for the external factors Z.

If we have a sample of observations $\mathcal{S}_n = \{(X_i, Y_i, Z_i)\}_{i=1}^n$, the nonparametric envelopment estimators of $\lambda(x,y)$ and $\lambda(x,y|z)$ can be obtained by plugging in the nonparametric estimator of the corresponding distribution H_{XY} and $H_{XY|Z}$ respectively. The nonparametric estimator of H_{XY} is given by

$$\widehat{H}_{n,XY}(x,y) = \frac{\sum_{i=1}^n \mathbb{1}(X_i \leq x, Y_i \geq y)}{n}, \tag{2.10}$$

where $\mathbb{1}(A) = 1$ if A is true and zero otherwise. For the conditional version we have to smooth over the values of Z_i in a neighborhood of z, because usually we do not have for all z, observations with exact values $Z_i = z$, so we have

$$\widehat{H}_{n,XY|Z}(x,y|Z=z) = \frac{\sum_{i=1}^n \mathbb{1}(X_i \leq x, Y_i \geq y) K_h(Z_i, z)}{\sum_{i=1}^n K_h(Z_i, z)}, \tag{2.11}$$

where $K_h(Z_i, z)$ are appropriate kernel functions and h is a vector of d bandwidths, one for each component of z. We know that the resulting FDH estimators of the efficiency scores are given by the simple expressions

$$\widehat{\lambda}_n(x,y) = \max_{i|X_i \leq x} \left\{ \min_{j=1,\ldots,q} \frac{Y_i^{(j)}}{y^{(j)}} \right\}, \tag{2.12}$$

$$\widehat{\lambda}_n(x,y|z) = \max_{i|X_i \leq x, ||Z_i - z|| \leq h} \left\{ \min_{j=1,\ldots,q} \frac{Y_i^{(j)}}{y^{(j)}} \right\}, \tag{2.13}$$

where the inequality $||Z_i - z|| \leq h$ has to be understood component by component $|Z_i^{(j)} - z^{(j)}| \leq h^{(j)}$. By comparing (2.12) and (2.13), we see clearly that the conditional efficiency estimate is a localized version of the marginal one, where the localization is in the Z-space and it is tuned by the bandwidths.[2] It has been pointed in Daraio and Simar (2005) that for being able to estimate

[2]For simplicity we focus in this paper on the FDH estimators because our target is to compare different ways for selecting the bandwidths h, but similar expressions have been derived for the DEA estimators and for the different orientations. When DEA is involved, the estimated attainable sets are convexified, so they imply to solve a linear optimization program in both cases. See Simar and Wilson (2015) for a detailed list of references.

the conditional upper boundary of the support of $H_{XY|Z}$ and so detect potential effect of Z, the nonparametric estimator has to be based on kernels for Z with compact support, like Epanechnikov or Quartic kernels, the Gaussian kernel being not allowed.[3]

The statistical properties of these estimators have been established. If we denote by $h^{(j)}$, $j = 1, \ldots, d$ the d components of the vector h and by $\bar{h}$ their product $\prod_{j=1}^{d} h^{(j)}$, at any fixed $(x, y) \in \Psi$ and under mild regularity conditions, we have as $n \to \infty$ with $h^{(j)} \to 0$ such that $n\bar{h} \to \infty$,

$$n^{1/(p+q)}\big(\lambda(x,y) - \widehat{\lambda}_n(x,y)\big) \xrightarrow{\mathcal{L}} Q_1(\eta_1) \tag{2.14}$$

$$(n\bar{h})^{1/(p+q)}\big(\lambda(x,y|z) - \widehat{\lambda}_n(x,y|z)\big) \xrightarrow{\mathcal{L}} Q_2(\eta_2), \tag{2.15}$$

where for $k = 1, 2$, Q_k is a Weibull distribution with parameters η_k described in Park et al. (2000) for FDH and in Jeong et al. (2010) for conditional FDH.

So we see in (2.13) the crucial role of the bandwidths in computing the conditional efficiencies in practice because it determines the localization in the data set of points where the "local" FDH is computed (the same is true for the conditional DEA and conditional variants of the FDH). We know that the optimal bandwidth for estimating a conditional distribution is of order $n^{-1/(d+4)}$ and, as pointed by Jeong et al. (2010), this deteriorates the rate of convergence of the conditional FDH to $(n^{4/(d+4)})^{-1/(p+q)}$, since the real number of observations used to compute $\widehat{\lambda}_n(x, y|z)$ is not of the order n but of order $n\bar{h} = n^{4/(d+4)}$. This indicates that any analysis with a large number of inputs/outputs and environmental factors will require large data sets.

Note also that the estimator of the conditional efficiency could equivalently be obtained by looking at the support of conditional survival function defined after (2.9) and estimated by

$$\widehat{S}_{n,Y|X,Z}(y|X \leq x, Z = z) = \frac{\widehat{H}_{n,XY|Z}(x, y|Z = z)}{\widehat{H}_{n,XY|Z}(x, 0|Z = z)}. \tag{2.16}$$

The estimator of $\lambda(x, y|z)$ is unique for a given h, but the derivation of the optimal bandwidths by LSCV methods will be different according to the chosen approach. First of all, we observe first that when estimating the survival function $S_{Y|X,Z}(y|x, z)$, we must select the optimal bandwidths h_x for each selected value of x, inducing numerical burden. In addition, by using the unique optimal bandwidths h derived from the estimation of $H_{XY|Z}(x, y|z)$, the resulting estimators of the efficiency scores have the expected monotonicity properties in x (as shown in Simar et al., 2016).

Remark 2.1. The use of discrete ordered or unordered variables Z

All the analysis done so far is valid even if some components of Z are categorical or discrete ordered

[3]This comes from the fact that in solving (2.8) where $H_{XY|Z}(x, y|z)$ is replaced by $\widehat{H}_{n,XY|Z}(x, y|Z = z)$, with e.g. a Gaussian kernel, $K_h(Z_i, z)$ in (2.11) is formally > 0 for all Z_i, so the $\sup\{\lambda > 0 | \widehat{H}_{n,XY|Z}(x, \lambda y|Z = z) > 0\}$ does not depend on z and $\widehat{\lambda}_n(x, y|z) = \widehat{\lambda}_n(x, y)$ (see also Remark 2.1 below).

variables. The only point is that special kernels handling these variables have to be used (see e.g. Li et al., 2013 and the references therein). Here bandwidths ℓ_z take often their values in $[0, c]$ where c depends on the chosen kernel. If $\ell_z = 0$, we obtain a separate analysis for each value of the corresponding discrete variable and for the maximum value c, there is no difference between the different groups corresponding to the different values of this variable, there is no smoothing at all. This may be useful when we are interested in the estimation of the full CDF (either $S_{Y|XZ}$ or $H_{XY|Z}$). However, when only the full frontier efficiency $\lambda(x, y|z)$ is of interest, it is easy to show that unless the optimal bandwidth $\ell_z = 0$, there is no effect on the support of $\widehat{S}_{n,Y|XZ}$ or of $\widehat{H}_{n,XY|Z}$, i.e. for any value of z, $\widehat{\lambda}_n(x, y|z) = \widehat{\lambda}_n(x, y)$. This is analog to what happens for continuous Z when using, e.g., Gaussian kernels (see Footnote 3). So smoothing in these discrete components is meaningless when we estimate $\lambda(x, y|z)$; we have to do separate analysis for each group ($\ell_z = 0$). This peculiar behavior was already pointed by Daraio et al. (2018) in another context.

3 Bandwidth Selection for Estimating $\lambda(x, y|z)$

3.1 The setup

Hereafter we denote our estimator $\widehat{\lambda}^h_n(x, y|z)$ to make explicit that the estimator depends on the chosen bandwidth h. Using the optimal bandwidth for estimating the conditional probabilities $H_{XY|Z}$ is certainly a good idea but estimating the support of $H_{XY|Z}(x, y|z)$ on a ray defined by y is more specific and nothing ensures that optimal value for this specific problem will be the same as the optimal value for estimating $H_{XY|Z}$ on its full range. A second issue, already pointed by Jeong et al. (2010), is related to the fact that for a given h, the real target of our estimator $\widehat{\lambda}^h_n(x, y|z)$ defined in (2.13), is not rigorously $\lambda(x, y|z)$ but rather

$$\lambda^h(x, y|z) = \sup\{\lambda > 0 | (x, \lambda y) \in \Psi^{z,h}\}, \tag{3.1}$$

where

$$\Psi^{z,h} = \left\{(x, y) \middle| H^h_{XY|Z}(x, y|z) = \Pr\left(X \leq x, Y \geq y \middle| ||Z - z|| \leq h\right) > 0\right\}. \tag{3.2}$$

In addition, it is clear that $\Psi^{z,h} = \cup_{||\tilde{z}-z|| \leq h} \Psi^{\tilde{z}}$.

Therefore, for all points (x, y) in Ψ^z, the error of estimation can be decomposed as

$$\widehat{\lambda}^h_n(x, y|z) - \lambda(x, y|z) = \left[\widehat{\lambda}^h_n(x, y|z) - \lambda^h(x, y|z)\right] + \left[\lambda^h(x, y|z) - \lambda(x, y|z)\right], \tag{3.3}$$

where the first difference in the brackets is due to the estimation error in the localized problem and the second difference is a non-random error due to the localization. The latter can be seen as a kind of "bias" introduced by the localization (we are not targeting the appropriate target). We need the following assumption to control the size of this bias:

Assumption 3.1. *For all z and all $(x, y) \in \Psi^z$, $\lambda^h(x, y|z) - \lambda(x, y|z) = O(||h||)$ as $||h|| \to 0$.*

This amounts to an assumption of differentiability of $\lambda(\cdot,\cdot|z)$ as a function of z and is analog to the Assumption 2 in Jeong et al. (2010), as will become clear from what follows. Note that if Z is separable and has no effect on the frontier ($\Psi^z = \Psi$ for all z) Assumption 3.1 is trivially satisfied for all h since in this case $\lambda^h(x,y|z) = \lambda(x,y|z) = \lambda(x,y)$ and so, when Z is separable, it turns out that the localization bias $O(||h||) \equiv 0$ for all h. Of course we do not know if Z is separable or not and this creates an additional problem, as explained below.[4]

3.2 Optimal order of the bandwidths: the problem

3.2.1 Z is not separable

Now looking at (3.3), we see that when the bandwidths are such that $||h|| \to 0$ and $n\bar{h} \to \infty$, the error of estimation has an order $O_p\left(n\bar{h}\right)^{-1/(p+q)}) + O(||h||))$. So the first part requires bandwidths as large as possible but the second term requires bandwidths as small as possible.

By Proposition 1 in Jeong et al. (2010), we need $||h|| \propto n^{-\gamma}$ with $\gamma < 1/d$ to ensure there are enough observations in the h-neighborhood of z (roughly speaking, the cardinality of the number of data points such that $||Z_i - z|| \leq h$ tends to infinity with probability one when $n \to \infty$).

Since we do not have explicit expression for the second component of (3.3) and the Weibull distribution linked to the first term contains unknown parameters, the best we can do is to determine the order of the optimal bandwidth by balancing the order of the two error terms. Simple algebraic manipulations lead to the following optimal order for each component of the bandwidth vector:

$$h^{(j)\star} \propto n^{-\gamma}, \text{ with } \gamma = \frac{1}{d+p+q}. \tag{3.4}$$

It is easy to check that in this case the two terms of (3.3) have an order $n^{-1/(d+p+q)}$. As is often the case in nonparametric smoothing techniques (where we usually balance between the square of the bias and the variance), choosing a smaller order for the bandwidths, say $h^{(j)} \propto n^{-\gamma}$ with $\gamma > 1/(d+p+q)$ (but keeping $\gamma < 1/d$), the second term in (3.3) is negligible as now being $o(n^{-1/(d+p+q)})$. We then obtain the asymptotic results described in (2.15) for the conditional efficiency estimators but at a rate which is now $(n\bar{h})^{1/(p+q)} = n^{(1-d\gamma)/(p+q)}$ where $\gamma > 1/(d+p+q)$, for eliminating the "bias" term. This argument is similar to the argument in Jeong et al. (2010), when we note that if $\gamma > 1/(d+p+q)$, our Assumption 3.1 leads to Assumption 2 in Jeong et al. (2010).

So we see that the optimal order of the bandwidth is different from the order resulting from the usual LSCV techniques focusing on the best estimation of the probabilities $H_{XY|Z}$. Let us denote $h^{(j)}_{\text{LSCV}}$ this resulting bandwidth, so we have, as pointed above, $h^{(j)}_{\text{LSCV}} \propto n^{-1/(d+4)}$. Since having a good estimation of the underlying probabilities remains an attractive idea, we suggest to

[4] Daraio et al. (2018) suggest a procedure to test the hypothesis of separability but the test requires anyway the selection of a bandwidth, as discussed here.

correct the bandwidths obtained by LSCV to achieve the optimal order. We denote this approach hereafter CORR, to be specific, we define

$$\begin{aligned} h^{(j)}_{\text{CORR}} &= n^{1/(d+4)} n^{-1/(d+p+q)} h^{(j)}_{\text{LSCV}} \\ &= n^{\frac{p+q-4}{(d+4)(d+p+q)}} h^{(j)}_{\text{LSCV}}. \end{aligned} \tag{3.5}$$

Of course, the resulting rates of convergence to the Weibull distribution using $h^{(j)}_{\text{LSCV}}$ and $h^{(j)}_{\text{CORR}}$ will be different. The rates $(n\bar{h})^{1/(p+q)}$ in (2.15) are $n^{4/((d+4)(p+q))}$ for the LSCV case and, as pointed above, $n^{1/(d+p+q)}$ for the corrected case. Obvious algebra, looking at the correction factor in (3.5), indicates that we achieve the same rate for $p+q=4$, better rates for $h^{(j)}_{\text{CORR}}$ if $p+q>4$ but slower rates if $p+q<4$. Note that our CORR approach is able to handle the bias due to localization, which is ignored in the LSCV criterion. Note also that the CORR approach only provide a bandwidth with optimal order, which is not necessarily the optimal bandwidth.

3.2.2 Z is separable

As pointed above when Z is separable there is no localization bias and only the estimation error has to be controlled for. Clearly we need large values of h, large enough so that $\widehat{\lambda}^h_n(x,y|z) \equiv \widehat{\lambda}_n(x,y)$, i.e., the conditional efficiency estimator has to coincide with the FDH estimator, with usual rate of convergence $n^{-1/(p+q)}$, as given in (2.14). Indeed when Z is separable, there is no need to compute the conditional estimator since for all z, $\lambda(x,y|z) = \lambda(x,y)$.

Of course in practice, when we compute $\widehat{\lambda}^h_n(x,y|z)$, we do not know if Z is separable or not. Hence, we need a data driven technique to select the appropriate bandwidth, that should be "large". The existing approaches we could use are based on the LSCV approaches which in general provides small bandwidths of order $n^{-1/(d+4)}$, which is evidently not appropriate. Since here the localization bias is identically equal to zero, the CORR approach is fully inappropriate in all the cases. But there is only one particular case where the LSCV method could be asymptotically optimal and this is when the component of Z is completely independent of the production process (Z is independent of (X,Y)). Indeed, in this case, Hall et al. (2004) have shown that the LSCV will provide for such component a bandwidth $h^{(j)} \to \infty$, when $n \to \infty$. However, for all the other components of Z which are separable, if they have some impact on the distribution of the efficiencies (so Z is not independent of (X,Y)), the corresponding elements $h^{(j)}_{\text{LSCV}} \to 0$ when $n \to \infty$, hence the LSCV approaches will provide too small values for the bandwidths.

But of course remember that we do not know in advance if the separability assumption holds and testing approaches (like Daraio et al., 2018) will require the use of bandwidths. That is the reason why we search to develop a new method allowing to estimate the mean squared error of the estimator of conditional efficiencies for a given bandwidth vector and valid in every possible situation. This is the main objective of the bootstrap approach described in the next section.

3.3 A bootstrap-based solution

The idea of using a bootstrap approach as an alternative was already mentioned in Jeong et al. (2010), but considering only the first part of the error (estimation part) described in (3.3) and they did not provide details on how to implement the idea in practice. Of course the "bias" part should also play a crucial role and we will suggest a way (adapted to the two situations described above, i.e., Z being separable or not) to handle these two parts simultaneously.

Now, to clarify the presentation of the bootstrap, we need to specify which is the sample of reference when computing the estimators. So hereafter, we denote our estimator as $\widehat{\lambda}^h(x, y|z; \mathcal{S}_n)$ to clearly indicate the chosen bandwidth h and the sample $\mathcal{S}_n$ used in its computation.

For a given point of interest (x, y, z), the Mean Squared Error (MSE) for a given bandwidth h is defined as

$$MSE_{x,y,z}(h) = \mathbb{E}_{\mathcal{S}_n}\left[\left(\widehat{\lambda}^h(x, y|z; \mathcal{S}_n) - \lambda(x, y|z)\ \right)^2\right], \tag{3.6}$$

where the only random part is linked to the randomness of the sample $\mathcal{S}_n$. The error of estimation can be decomposed, as above, in the two components

$$MSE_{x,y,z}(h) = \mathbb{E}_{\mathcal{S}_n}\left[\left(\widehat{\lambda}^h(x, y|z; \mathcal{S}_n) - \lambda^h(x, y|z) + \lambda^h(x, y|z) - \lambda(x, y|z)\right)^2\right]. \tag{3.7}$$

The first part is the estimation error in the localized problem and can be approximated by the subsampling-bootstrap (see Jeong and Simar, 2006 and Simar and Wilson 2011), whereas we suggest below a way to approximate the "bias" term $\lambda^h(x, y|z) - \lambda(x, y|z)$.

Due to the Assumption 3.1, it may be reasonable to estimate the bias by the first term of its Taylor expansion which can be written as

$$\lambda^h(x, y|z) - \lambda(x, y|z) = \sum_{j=1}^{d} \left.\frac{\partial \lambda^h(x, y|z)}{\partial h^{(j)}}\right|_{h^{(j)}=0} h^{(j)} + o(||h||) \tag{3.8}$$

Clearly this approximation is valid even if Z is separable, in this case the corresponding derivatives being equal to zero. The problem of course is that the function $\lambda^h(x, y|z)$ is unknown, so, in the bootstrap world, we will use an estimate by using a fixed pilot bandwidth h_0, such that $h_0 \to 0$ as $n \to \infty$. In practice, we could choose h_{LSCV}, since we do not have localization-bias issues when estimating $\lambda^h(x, y|z)$. We suggest to use the conditional FDH estimator, so we use the next proxy for the derivative

$$\left.\frac{\partial \lambda^h(x, y|z)}{\partial h^{(j)}}\right|_{h^{(j)}=0} \approx \lim_{\varepsilon \to 0^+} \frac{\widehat{\lambda}^{h_0^{(j)}+\varepsilon}(x, y|z, \mathcal{S}_n) - \widehat{\lambda}^{h_0^{(j)}-\varepsilon}(x, y|z, \mathcal{S}_n)}{2\varepsilon}, \tag{3.9}$$

where $\varepsilon > 0$ is a tuning parameter to fix the neighborhood of $h_0^{(j)}$ which should converge to zero faster than h_0 (we used $\varepsilon = n^{-1/2}$ in our Monte Carlo experiments, but the results are very stable to this choice).

Note that the FDH estimators are quite discontinuous, and since the estimates depend only on one data point, tilting the bandwidth with ε could produce high jumps in the numerator even with small value of ε. One could use as well a continuous version of the FDH, which is the linearized free disposal hull (LFDH) proposed in Jeong and Simar (2006), but, even if this estimator is continuous, it is still not differentiable, therefore some smoothing techniques are needed to define better estimates for the local derivatives. We will use standard and fast B-splines techniques, but other smoothing techniques could be used as well.

3.3.1 The bootstrap algorithm

For the part of the MSE which comes from the estimation error, we can use subsampling. So the bootstrap estimate of $MSE_{x,y,z}(h)$ can be obtained through the following steps.

1. Compute the "true" value of $\lambda^h(x,y|z)$ in the bootstrap world: here we need a pilot bandwidth h_0 that we choose as h_{LSCV} for the reason explained above, this gives $\widehat{\lambda}^{h_0}(x,y|z;\mathcal{S}_n)$.[5]

2. Select a subsample size $m < n$, then for $b = 1,\ldots,B$, draw without replacement a subsample of size m from the $\mathcal{S}_n$.[6] This provides a sample $\mathcal{S}^\star_{m,b} = \left\{(X^\star_{i,b}, Y^\star_{i,b}, Z^\star_{i,b})\right\}_{i=1}^m$ and we compute $\widehat{\lambda}^{\star,h}_b(x,y|z;\mathcal{S}^\star_{m,b})$.

3. The bootstrap analog of $MSE_{x,y,z}(h)$ for the selected value of m, is given by

$$\widehat{MSE}_{x,y,z}(h,m) = \frac{1}{B}\sum_{b=1}^{B}\left[\left(\frac{m}{n}\right)^{1/(p+q)}\left(\widehat{\lambda}^{\star,h}_b(x,y|z;\mathcal{S}^\star_{m,b}) - \widehat{\lambda}^{h_0}(x,y|z;\mathcal{S}_n)\right)\right.$$
$$\left. + \left(\frac{\widehat{\lambda}^{h_0^{(j)}+\varepsilon}(x,y|z,\mathcal{S}_n) - \widehat{\lambda}^{h_0^{(j)}-\varepsilon}(x,y|z,\mathcal{S}_n)}{2\varepsilon}\right)h\right]^2, \qquad (3.10)$$

   where the factor m/n is there for correcting the rate of convergence due to the different sample size $m < n$ used with $\mathcal{S}^\star_{m,b}$ (see Simar and Wilson, 2011a for details). We will discuss below how to choose m in practice.

Now this value can be computed for any value of h and any value of (x,y,z) but the latter depends on the realization of the random variable (X,Y,Z). So what is more appropriate is the Average Mean Square Error (AMSE) defined as follows

$$AMSE(h) = \mathbb{E}_{X,Y,Z}\left[MSE_{X,Y,Z}(h)\right]. \qquad (3.11)$$

[5] It is important to notice that we use a pilot bandwidth h_0 for defining the true value in the bootstrap world. If we would rather use the current value of h, it is easy to verify that the minimum of $\widehat{MSE}_{x,y,z}(h,m)$ in (3.10) would be zero with $h = 0$.

[6] As explained in Jeong and Simar (2006), with or without replacement does not matter asymptotically, but Simar and Wilson (2011a) report better performance in small samples of the drawings without replacement.

In practice, as often suggested in this kind of problems, we will use the empirical version of the AMSE, and use the bootstrap approximation defined above for a given value of subsample size m. So our criterion to evaluate a particular bandwidth vector h is given by

$$\widehat{AMSE}(h,m) = \frac{1}{n}\sum_{i=1}^{n} \widehat{MSE}_{X_i,Y_i,Z_i}(h,m), \tag{3.12}$$

where $\widehat{MSE}_{X_i,Y_i,Z_i}(h,m)$ is given by evaluating (3.10) at each data point (X_i, Y_i, Z_i). So we will search for the value of h minimizing $\widehat{AMSE}(h,m)$.

3.3.2 Selection of the subsample size m

Jeong and Simar (2006) have proven that the subsampling is consistent for any value $m = n^\beta$, where $\beta \in (0,1)$. But the quality of the approximation in finite samples depends on β, although the results are very stable for small variations in the value of m. To select a value for m, we follow the procedure suggested in Simar and Wilson (2011a), which is inspired from Politis et al. (2001). We compute for a given m, $\hat{h}_m$ the optimal value for h and the achieved value of the optimum $\widehat{AMSE}(\hat{h}_m, m)$. Then we redo the exercise for a grid of values for m and search the values of m where the quantity of interest, i.e. $\widehat{AMSE}(\hat{h}_m, m)$ is less volatile. This can be viewed by some appropriate plots of $\widehat{AMSE}(\hat{h}_m, m)$ versus m, but can be automated in a Monte Carlo experiment along the lines described in Politis et al. (2001). Note that in our setup here, the FDH estimates depends on only one point in the reference samples, $\mathcal{S}^\star_{m,b}$ and $\mathcal{S}_n$, used in (3.10). So for $h = h_0$, the probability that $\widehat{\lambda}^{\star,h}_b(x,y|z;\mathcal{S}^\star_{m,b}) = \widehat{\lambda}^{h_0}(x,y|z;\mathcal{S}_n)$ is equal to m/n. Therefore if we choose m too large, we will bias to zero the sampling part of the estimate $\widehat{MSE}_{x,y,z}(h,m)$ for h near h_0. On the other hand, too small m will give too much noise in $\widehat{\lambda}^{\star,h}_b(x,y|z;\mathcal{S}^\star_{m,b})$ and too many undefined local FDH estimators. In our Monte Carlo experiments we used a grid of values for m in the range $[0.10n, 0.35n]$. Again the procedure is quite robust to the choice of the preceding grid.

In practice, we use the various bootstrap subsample sizes in the selected grid $m_1 < m_2 < \ldots < m_J$, and then measure the volatility corresponding to m_j by computing the standard deviations of the achieved optimal $\widehat{AMSE}$ corresponding to the values of $m = m_{j-k}, \ldots, m_j, \ldots, m_{j+k}$ where k is a small integer (e.g., $k = 1, 2$, or 3) and $j = (k+1), \ldots, (J-k)$. The subsample size m would then be chosen as the m_j yielding the smallest measure of volatility. This involves some computational burden, but the FDH estimators are so fast to compute that for a given sample of data, it remains quite reasonable.

We have observed in our Monte Carlo experiments a great stability of the results w.r.t. the choice of m and that often the chosen value was not far from $n^{3/4}$, which can be viewed, in our experiments, as a rough approximation of the optimal m. In our Monte Carlo experiment this simple rule provided in all the scenarios very good results, as good as the full search of optimal m

over the grid. While this is not a proof of any optimality, for sure with this simple rule for defining m, the subsampling is consistent.

3.3.3 Approximating the derivatives

If we choose the conditional FDH estimates for defining the derivative at each data point (X_i, Y_i, Z_i) we have

$$\widehat{der}_i = \frac{\widehat{\lambda}^{h_0^{(j)}+\varepsilon}(X_i, Y_i|Z_i, \mathcal{S}_n) - \widehat{\lambda}^{h_0^{(j)}-\varepsilon}(X_i, Y_i|Z_i, \mathcal{S}_n)}{2\varepsilon}, \tag{3.13}$$

where $\varepsilon = n^{-1/2}$. But as pointed above the estimates of the derivatives can still show high jumps due to the discontinuity of the FDH estimator. So we suggest to smooth the obtained estimates $\widehat{der}_i$ over the values of Z_i to correct this disappointing behavior of the FDH. We use in our Monte Carlo experiment a smoothing technique based on penalized B-splines (see e.g. Eilers and Marx, 1996). We define

$$\overline{der}_i = g(Z_i),\ i = 1, \ldots, n, \tag{3.14}$$

where $g(z)$ is a penalized B-spline estimate of the regression of $\widehat{der}$ on Z, obtained from the data points $\{Z_i, \widehat{der}_i\}_{i=1}^n$, where the FDH estimators are used to define the $\widetilde{\lambda}$.[7] It appears from our Monte Carlo experiments, that this smoothing provides the desired stabilization and improves substantially the performances of the bootstrap algorithm. So we have the explicit formula

$$\widehat{AMSE}(h, m) = \frac{1}{nB}\sum_{i=1}^{n}\sum_{b=1}^{B}\left[\left(\frac{m}{n}\right)^{1/(p+q)}\left(\widehat{\lambda}_b^{\star,h}(X_i, Y_i|Z_i; \mathcal{S}^\star_{m,b}) - \widehat{\lambda}^{h_0}(X_i, Y_i|Z_i; \mathcal{S}_n)\right) + h\,\overline{der}_i\right]^2. \tag{3.15}$$

We can denote by h_{BOOT} the minimizers of (3.15), and we can then select the optimal m according to the rule described above. We will also refer to the Rule Of Thumb (ROT) for selecting h by minimizing (3.15) where m is fixed to our simple rule $m = n^{3/4}$.

4 Monte Carlo Experiments

Under the assumptions made above, we have for any given h a consistent estimator of the localization bias, and we know (Jeong and Simar, 2006) that subsampling provides consistent estimator of the estimation error. So we conjecture that the bootstrap we propose above provides for all h a consistent estimator of the $AMSE(h)$. The Monte Carlo experiments confirm this conjecture. Still, a formal proof of the consistency of the bootstrap algorithm has to be done but this is out of the scope of this paper.

[7] We used the Matlab programs provided by Eilers and Marx (1996), with B-splines defined on 20 intervals, cubic splines, with penalty on the first differences (with penalty term= 100), to target the continuity of the derivatives. As known in the related literature the results are rather stable with respect to the spline tuning parameters.

So we will compare through Monte Carlo simulations the performances, in finite samples, of the nonparametric estimator $\widehat{\lambda}_n^h(x, y|z)$, obtained by using different approaches for selecting the bandwidth h. We have two methods based on LSCV techniques using h_{LSCV} and its corrected version h_{CORR} as defined in the preceding section and the method based on bootstrap, h_{BOOT}. We also provide some empirical evidence that the bootstrap is able to estimate the $AMSE(h)$ by comparing for some particular given sample, its bootstrap estimate with the true value of $AMSE(h)$ evaluated from an independent Monte Carlo experiment.[8]

To achieve this we consider 4 scenarios (denoted A, B, C and D) which correspond to 4 different situations on how Z interacts with the production process crossing the fact that Z is separable or not with the fact that Z influences the distribution of inefficiencies or not. We limit (for the ease of the presentation and because of the computational burden of the bootstrap approach in a simulation setup) the presentation in the case of one input X, one output Y and one external factor Z. Our DGP is inspired from the scenarios proposed in Simar and Wilson (2011b). The basic marginal frontier (i.e. ignoring the possible effect of Z) is given by the function $g(x) = \sqrt{1-(x-1)^2}$ for $x \in (0, 1)$. We simulate n independent uniformly distributed input values $X_i \sim \text{Unif}(0, 1)$, external factors Z_i independent $Z_i \sim 4\,\text{Beta}(2, 2)$ (the factor 4 is to scale the beta density so that $Z_i \in (0, 4)$) and we generate the inefficiencies according to a half-normal distribution $U_i|Z_i \sim \text{N}^+(0, \sigma_U^2(Z_i))$. The effect of Z on the frontier and the variance function $\sigma_U^2(Z_i)$ will vary according the 4 cases as described below. In all the 4 scenarios we expect a good behavior of the bootstrap method which is supposed to be adapted to all the situations. We described shortly what we expect from the methods using LSCV bandwidths, h_{LSCV} and h_{CORR}.

Case A : We consider the "basic" model, without any effect of the environmental factor, assuming full independence between Z and X, Y, U and setting $\sigma_U(Z) = 0.5$:

$$Y_i = g(X_i) \times \exp(-U_i), \text{ where } U_i \text{ does not depends on } Z_i. \tag{4.1}$$

So here we have separability and full independence: here the usual marginal FDH estimator should be used since no conditioning is required, so The FDH is our benchmark in this case . We expect that when selecting the bandwidth, the LSCV criterion would provide reasonable solutions since, as the theory tells us, in this case the resulting bandwidth h_{LSCV} should be large. The CORR should not be appropriate since there is no localization bias.

Case B : In this case, we assume separability, but we impose $\sigma_U(Z) = Z/4$, such that the distribution of inefficiency U depends on Z, with higher probability of being inefficient when Z increases

[8] In each of the scenarios, we simulated M random samples of specified size n, and over a grid of values for h we evaluate the "true" $AMSE(h)$ given in (3.11) by replacing the operator $\mathbb{E}_{X,Y,Z}$ by its Monte Carlo average; by the Law of large numbers, if $M \to \infty$, we can approach this true value as close as we want. In practice we used $M_0 = 5000$.

(the average value of $\sigma_U(Z)$ remains equal to 0.5, as in Case A). So we have:

$$Y_i = g(X_i) \times \exp(-U_i), \text{ where } U_i \text{ depends on } Z_i. \tag{4.2}$$

Here again, as in Casa A, the benchmark is the FDH (no conditioning is needed) and we expect a poor behavior of methods based on LSCV, at least when n is large, because the dependency between Z and (X, Y) will induce bandwidth converging to zero when $n \to \infty$, where the optimal values for h should be large.

Case C : This is a non-separable case since Z influences the position of the frontier, but we keep $\sigma_U(Z) = 0.5$, so no effect of Z on the distribution of inefficiencies:

$$Y_i = g(X_i) \times \exp\{-(Z_i - 2)\} \times \exp(-U_i), \text{ where } U_i \text{ does not depends on } Z_i. \tag{4.3}$$

Here, the effect of Z is only on the boundary of the attainable (X, Y), the shift is multiplicative and more important when Z decreases. We expect here a better behavior of the CORR approach (with respect to the basic LSCV), since it takes into account the bias of localization.

Case D : In this non-separable case, Z influences both the frontier and the inefficiency U, as above $\sigma_U(Z) = Z/4$:

$$Y_i = g(X_i) \times \exp\{-(Z_i - 2)\} \times \exp(-U_i), \text{ where } U_i \text{ depends on } Z_i. \tag{4.4}$$

In this case, Z has a compound effect on both the boundary of the attainable set and the distribution of inefficiency. We expect similar behavior as in Case C.

4.1 Quality of the bootstrap approximation of the $AMSE(h)$

By an intensive Monte Carlo simulation we can approximate the "true" optimal value of h and the values of $AMSE(h)$ as a function of h. It is computed for a given h in a grid of values, for a given sample size n, and for a given scenario, as

$$AMSE(h) = \frac{1}{M_0} \sum_{m_0=1}^{M_0} \frac{1}{n} \sum_{i=1}^{n} \left[\widehat{\lambda}^h(X_i, Y_i | Z_i; \mathcal{S}_n^{(m_0)}) - \lambda(X_i, Y_i | Z_i) \right]^2 \tag{4.5}$$

where $\mathcal{S}_n^{(m_0)}$ is one particular simulated sample, for $m_0 = 1, \ldots, M_0$, and we choose $M_0 = 5000$ in practice.

To have an idea of the quality of the approximation of our bootstrap algorithm, in each scenario we provide the picture of the bootstrap approximation with the selected value of the subsample size derived by our algorithm. We can first do this for one particular random sample, to see if we have any chance (as we expect) to have similar behavior of the curves in the 4 cases. Of course this

is only for qualitative evaluation but below we analyze the results for many such random samples. The results are displayed in Figure 1 to 4 for the four scenarios with $n = 200$.

First we see that for the 4 cases, the true values behave as expected, with the optimal values for h going to ∞ for case A and B (Z is separable, so no conditioning is needed and the FDH is the optimal estimator; in practice, since the range of Z is equal to 4, we achieve already this with $h \geq 4$). For the other cases, as expected, conditioning is needed and the optimal bandwidth is around 0.3–0.35. Their bootstrap approximations for one sample of size $n = 200$ are quite reasonable leading to optimal bandwidth (for this arbitrary sample) of the same order of magnitude as the true one.

Figure 1: *Case A, $n = 200$. Left panel: Monte Carlo (true) values of $AMSE(h)$. Right panel: its Bootstrap estimation for one particular sample.*

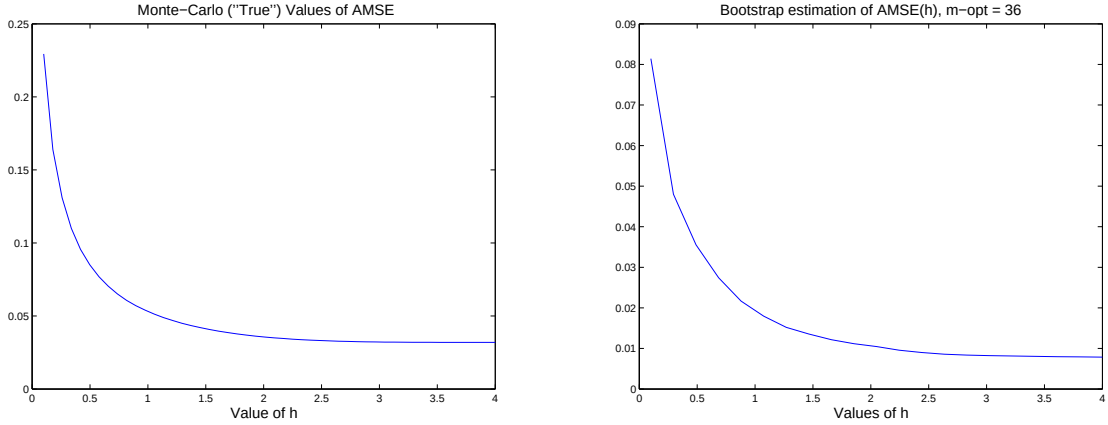

Figure 2: *CaseB, $n = 200$. Left panel: Monte Carlo (true) values of $AMSE(h)$. Right panel: its Bootstrap estimation for one particular sample.*

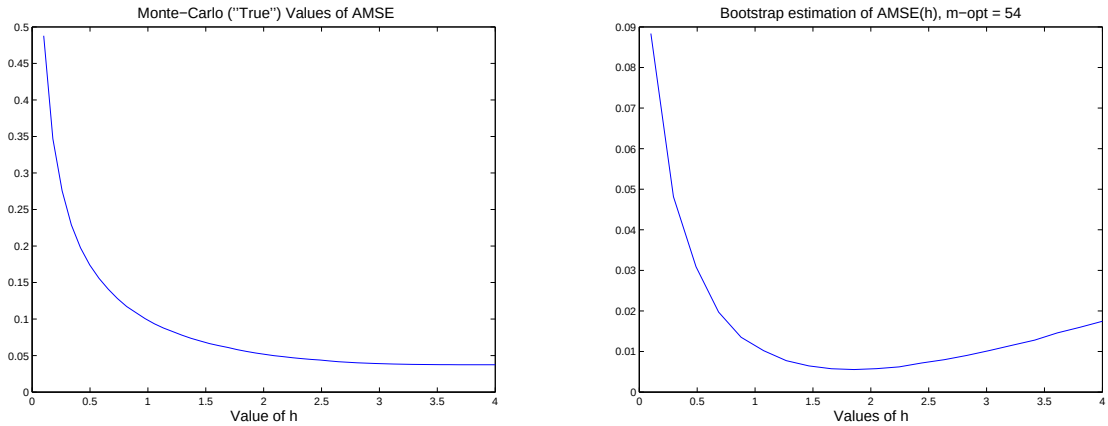

Figure 3: *Case C, $n = 200$. Left panel: Monte Carlo (true) values of $AMSE(h)$. Right panel: its Bootstrap estimation for one particular sample.*

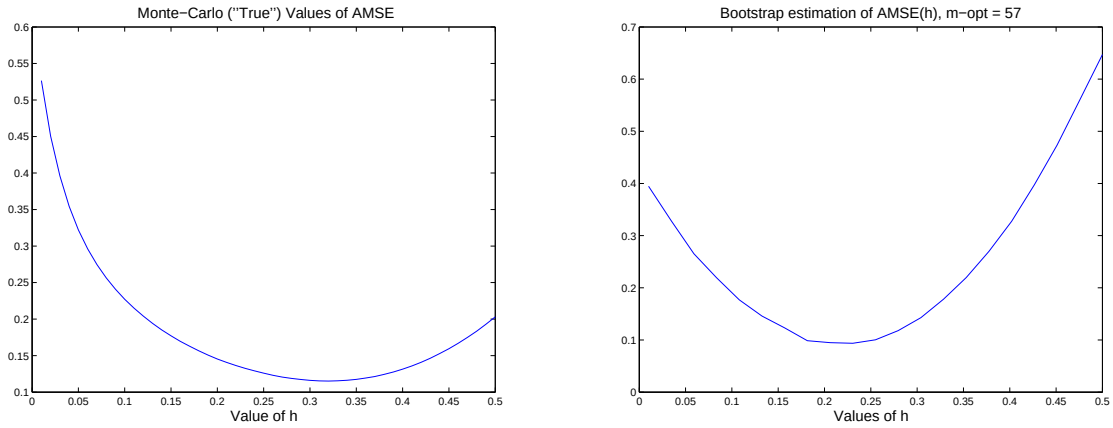

Figure 4: *Case D, $n = 200$. Left panel: Monte Carlo (true) values of $AMSE(h)$. Right panel: its Bootstrap estimation for one particular sample.*

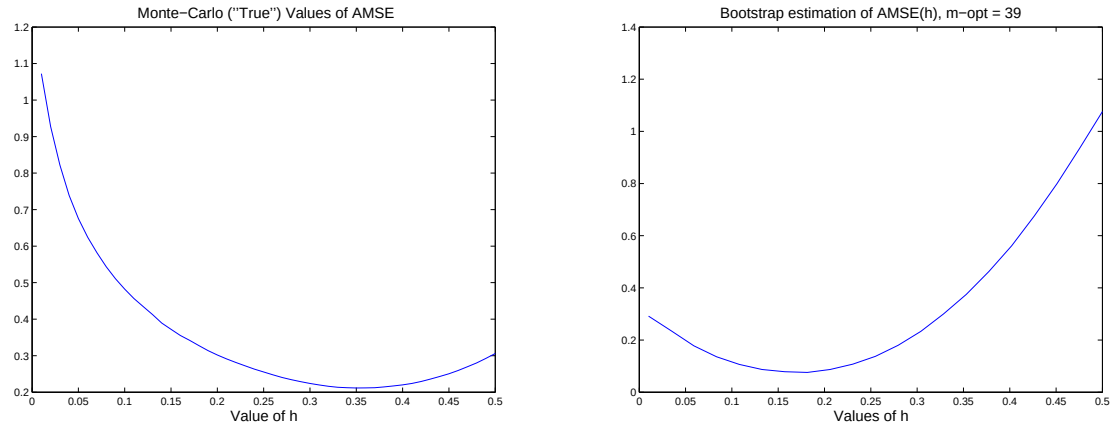

Table 2 gives for the 4 cases, and for different sample sizes the "true' optimal bandwidths obtained through the Monte Carlo approximation, and the square root of the corresponding optimal AMSE (the square roots are in the same units as the efficiency scores). The latter will serve as benchmark to evaluate and compare the performance of the different approaches in the next subsection.

Table 2: *The optimal "true" value of h and the corresponding values of RAMSE obtained by Monte Carlo simulation across the $M_0 = 5000$ trials.*

	$n = 100$		$n = 200$		$n = 400$		$n = 800$	
	h_{opt}	RAMSE	h_{opt}	RAMSE	h_{opt}	RAMSE	h_{opt}	RAMSE
Case A	4.00	0.2334	3.92	0.1787	4.00	0.1366	4.00	0.1035
Case B	4.00	0.2575	4.00	0.1935	4.00	0.1474	4.00	0.1097
Case C	0.38	0.4004	0.32	0.3393	0.26	0.2858	0.21	0.2398
Case D	0.42	0.5411	0.35	0.4600	0.29	0.3906	0.24	0.3274

4.2 Performance of various approaches

In this subsection we compare and analyze the behavior of the different methods to select the bandwidth, by simulating in the 4 cases described above, $MC = 200$ samples of various sizes, $n = 100, 200, 400$ and 800. For one particular sample, the computational burden is reasonable depending on the sample size and the number of values of m over which the subsampling is performed (we fixed the number of bootstrap replications at 100, but similar results where found with larger values on some pilot experiments). For one particular sample with $n = 200$ and 21 values of the grids for both m and h, the code provided in the Appendix took less than 3 minutes with an Intel 3.4 GHz machine (see details in Appendix A) but it raises quickly with n increasing. So the Monte Carlo evaluations here are limited to 200 replications. Under the heading "BOOT_{ROT}", we also report the results of the Rule Of Thumb (ROT) for selecting $m = n^{3/4}$, which decreases the computing time by a factor 5. The resulting optimal bandwidths with the corresponding evaluation of the $AMSE$ are displayed in Tables 3 to 6. In fact we report, as above, the $RAMSE$, i.e. the square root of the $AMSE$.

Table 3: *Case A. Optimal value of h and the corresponding values of RAMSE, obtained through MC=200 Monte Carlo trials.*

	$n = 100$		$n = 200$		$n = 400$		$n = 800$	
	h_{opt}	RAMSE	h_{opt}	RAMSE	h_{opt}	RAMSE	h_{opt}	RAMSE
FDH	—	0.2301	—	0.1778	—	0.1364	—	0.1031
LSCV	6.09	0.2638	5.58	0.2014	4.72	0.1567	3.89	0.1273
CORR	3.29	0.3037	2.75	0.2359	2.12	0.1909	1.59	0.1628
BOOT	2.66	0.2548	2.77	0.1957	2.74	0.1497	2.54	0.1210
BOOT_{ROT}	2.61	0.2581	2.78	0.1965	2.72	0.1493	2.58	0.1196

Table 4: *Case B. Optimal value of h and the corresponding values of RAMSE, obtained through MC=200 Monte Carlo trials.*

	$n = 100$		$n = 200$		$n = 400$		$n = 800$	
	h_{opt}	RAMSE	h_{opt}	RAMSE	h_{opt}	RAMSE	h_{opt}	RAMSE
FDH	—	0.2564	—	0.1878	—	0.1476	—	0.1065
LSCV	1.44	0.3655	1.12	0.3024	0.92	0.2594	0.76	0.2096
CORR	0.78	0.4666	0.55	0.4084	0.41	0.3605	0.31	0.3110
BOOT	1.95	0.3385	1.48	0.2844	1.24	0.2391	1.02	0.1886
BOOT_{ROT}	1.83	0.3456	1.44	0.2863	1.23	0.2397	1.03	0.1883

Table 5: *Case C. Optimal value of h and the corresponding values of RAMSE, obtained through MC=200 Monte Carlo trials.*

	$n = 100$		$n = 200$		$n = 400$		$n = 800$	
	h_{opt}	RAMSE	h_{opt}	RAMSE	h_{opt}	RAMSE	h_{opt}	RAMSE
FDH	—	10.3021	—	11.6621	—	13.1221	—	14.4115
LSCV	0.68	0.7421	0.55	0.6065	0.49	0.5360	0.42	0.4670
CORR	0.36	0.4196	0.27	0.3626	0.22	0.3047	0.17	0.2556
BOOT	0.28	0.4216	0.24	0.3616	0.20	0.3014	0.18	0.2490
BOOT_{ROT}	0.30	0.4169	0.24	0.3610	0.20	0.3031	0.16	0.2532

Table 6: *Case D. Optimal value of h and the corresponding values of RAMSE, obtained through MC=200 Monte Carlo trials.*

	$n = 100$		$n = 200$		$n = 400$		$n = 800$	
	h_{opt}	RAMSE	h_{opt}	RAMSE	h_{opt}	RAMSE	h_{opt}	RAMSE
FDH	—	21.00881	—	23.3558	—	25.8223	—	28.2463
LSCV	0.64	0.8417	0.51	0.6548	0.45	0.5807	0.38	0.4901
CORR	0.34	0.5685	0.25	0.5030	0.20	0.4455	0.15	0.3710
BOOT	0.27	0.5793	0.23	0.4981	0.20	0.4387	0.17	0.3507
BOOT_{ROT}	0.29	0.5685	0.23	0.4952	0.20	0.4397	0.16	0.3575

From the results of our Monte Carlo experiments summarized in Tables 3–6, we learn the following. For Case A, Z is independent of (X, Y) and so is separable as expected, the benchmark is the FDH: no need to condition, but when conditioning is used, the LSCV criterion gives rather good estimator of the bandwidths and so of the resulting efficiency scores (RAMSE, not far from the

benchmark values of the FDH). The CORR, as expected, behaves less good (there is no localization bias here). The good news is that the BOOT approach and its ROT version behave even slightly better than the LSCV.

Case B (Z is separable but influences the production process through inefficiency) seems to be the most difficult case for all the approaches, compared with the benchmark solution (no conditioning and use of the FDH estimator). The bootstrap approaches seems to be more robust, giving the smallest RAMSE among its competitors.

On the contrary, if Z is not separable (Case C and Case D), the FDH is nonsense, as shown by the achieved RAMSE. Here as expected, for both cases, the LSCV approach does not work well even for large n, because it does not take into account for the correction for the localization bias. We can see that by correcting the rate, the CORR method does much better. However, here again, the bootstrap approaches BOOT and BOOT_{ROT} seem to be able to adapt the procedure to the actual situation: the achieved RAMSE are smaller that for CORR, in particular when n increases.

Although there is a price to pay in the numerical burden which is typical for bootstrap based methods (unless we use an appropriate Rule Of Thumb for selecting m), our new bootstrap approach appears to be able to handle all the cases. So we think it will be suitable for practical applications where the practitioner does not *a priori* know whether the separability condition holds or not.

5 Conclusions

This paper provides a complete investigation on the practical aspects of bandwidth selection procedures for the particular framework of conditional efficiency estimation. It proposes a new approach based on bootstrap for selecting the bandwidth, crucial for obtaining reliable estimates of conditional efficiency in empirical studies. The theoretical developments are complemented by practical implementation details and the Matlab codes are provided.

Traditional LSCV methods focus on the estimation of an appropriate conditional distribution, but not on the estimation of its support. In case the external variable is non separable (has an effect on the frontier), the order of the resulting bandwidth obtained by LSCV may be not optimal. In these cases, we suggest to correct the LSCV bandwidth. But in case of separability, this correction is not recommended. Our new approach is able to handle all these cases. Indeed, by developing the component of the MSE for conditional efficiency estimators, we have been able to disentangle the role of the bandwidth for the underlying localization bias and the estimation process itself. Our approach suggests how to estimate these two parts.

By Monte Carlo techniques, we have illustrated that our new procedure is robust with respect to various separable/non-separable DGPs improving the quality of the estimation. In a practical situation where the researcher has no information about the separability condition, the bootstrap

method is an attractive approach which allows us to handle all the cases correctly. To be more specific, it gives an approach able to handle separable and non separable cases doing as well as the traditional LSCV (and its CORR version) when the latter are appropriate, and much better when they are not.

If one is interested to estimate partial frontiers like order-m or order-α, then the target would no longer be the support of the distribution and the story may be quite different in these cases. In particular, LSCV remains a reasonable way to determine the bandwidth for fitting the quantile, but these issues are kept for future work. Another track for future research is to derive better estimates of the derivatives, in the light of the techniques developed e.g. in Park et al. (2008).

A Appendix: Matlab Code

We present below the Matlab code allowing to run all the examples used in the Monte Carlo scenarios described above. For the spline smoothing, we used the Matlab code provided by Eilers and Marx (1996) and downloadable from their own website. All these computations can be equivalently done in R. About computational burden, the example below is for Case B, with n=200, and B=100 bootstrap replications. It took 160 seconds on an Intel 3.4GHz (16G RAM) and 313 seconds on a MacPro, Intel 2.6GHz, (8G RAM).

```
% This version for UNIVARIATE Z
% prepared by  L. SIMAR, February 2018
% This is for one sample simulated according one of the scenarios of the paper

clearvars
close all

n=200;
B=100;% number of bootstrap loop (should be >=200, but works fine with B=100)
pen_spl=100;% penalty factor for splines
ord_pen=1; % order of the penalty in splines

% Ex='caseA';% full independence between Z and others
Ex='caseB';% Separability but inefficiency U depends on Z
% Ex='caseC';% Non separable Z influence the frontier BUT NOT U
% Ex='caseD';% Non separable Z influence the frontier AND U
fprintf('================= NEW RUN ==================\n')
fprintf(' You choose Case-Example %8s \n',Ex)
fprintf(' With sample size %4.0f \n',n)
fprintf(' Spline Tuning order = %2d and penalty = %6.3f \n',[ord_pen,pen_spl])
switch Ex
    case{'caseA','caseB'}
        % SEPARABILITY
        hmin=0.1;hmax=4;% this has to be adpated after some trials
    case{'caseC','caseD'}
        % NO-SEPARABILITY
        hmin=0.01;hmax=0.5;% this has to be adpated after some trials
    otherwise
        disp('select appropriate example')
        return
end
p=1;q=1;d=1;
disp('Bounds for the grid search in h')
disp([hmin, hmax])
nhV=21;
hV=linspace(hmin,hmax,nhV);

state = 9;% to allow reproducability
rng(state);
% # Generate the random sample of size n = (X,Y,Z)
```

```
% # Environment Z
Z=4*betarnd(2,2,n,1);
% # Input
X = unifrnd(0,1,n,1);
switch Ex
    case('caseA')% SEPARABILITY
        Yfr = sqrt(ones(n,1)-(X-1).^2);
        sigU = 0.5*ones(n,1); % U independent of Z
    case('caseB')% SEPARABILITY
        Yfr = sqrt(ones(n,1)-(X-1).^2);
        sigU = Z/4;
    case('caseC')% # NO separability
        Yfr = sqrt(ones(n,1)-(X-1).^2).*exp(-(Z-2));
        sigU = 0.5*ones(n,1); % U independent of Z
    case('caseD')% # NO separability
        Yfr = sqrt(ones(n,1)-(X-1).^2).*exp(-(Z-2));
        sigU = Z/4;
    otherwise
        disp('select appropriate example')
        return
end
% # Efficiencies
U = abs(normrnd(0,sigU));
% # # Values for OUTPUT Y
Y = Yfr.*exp(-U);
trueff=exp(U);
nobs=(1:n)';

t0=clock;
% Build the grid search for the subsample size
m_min=floor(0.15*n);
m_max=floor(0.35*n);
mgrid=21;
mpas=max(1,round((m_max-m_min)/(mgrid-1)));
mv=(m_min:mpas:m_max)';
nmv=length(mv);

hoptm=zeros(nmv,1);
MSEoptm=zeros(nmv,1);
MSEboot=zeros(nmv,nhV);
% Evaluation of the LSCV bandwidth and its corrected version
kappa = 1/(p+q);% rate of FDH
% % % LSCV bandwidth
fprintf('-------------------------------------------------------------------------------\n')
fprintf('\n LSCV on Non-smoothed CDF of H(X,Y|Z) (Li, Lin, Racine 2013) \n')
kernelZ='epan';
hz0=n^(-1/(d+4))*std(Z);%
W=[X,-Y]; % -Y because survivor function in Y
CV0=CCDFnonsmth_lscv_NEW(hz0,W,Z,n,d,kernelZ);
disp('starting values for h')
disp(hz0')
fprintf('Starting value of LSCV is CV = %15.8f \n', CV0)
LB=hz0/10; % lower bound for h
UB=Inf*ones(d,1);% upper bound for h
options=optimset('Algorithm','active-set','Display','iter','MaxFunEvals',5000,'MaxIter',100);

LSCV='YES';% type here YES or NO
if strcmp(LSCV,'YES')
    tic
[h,CV,exitflag]=fmincon(@(h)CCDFnonsmth_lscv_NEW(h,W,Z,n,d,kernelZ),hz0,[],[],[],[],LB,UB,[],options);
    toc
    if exitflag <= 0
        fprintf('Problem for minimization, exitfalg is %3.0f \n',exitflag)
    end
    disp('Selected Bandwidths')
    fprintf('Final Value of LSCV = %15.10f\n',CV)
    fprintf('\nFor estimation of H(x,y|z)\n  Starting h0    Lower    Upper   Final values for h \n')
    fprintf('%15.8f %15.8f %15.8f %15.8f \n',[hz0 LB  UB  h]')
    else
    h=1.44162963;% value found for Case B, seed=9 and n=200
    CV =0.0883316662;% value found for Case B, seed=9 and n=200
end
hz=h*n^(((1/kappa)-4)/((d+4)*(d +(1/kappa))));
hlscv=h;
hp=hz;% this is h_{CORR} in BDS4 paper
fprintf(' Final value of h_{corr} after rate correction hz = %8.6f \n',hz)

hoptmPEN=zeros(nmv,1);
```

```
ih_optmPEN=zeros(nmv,1);
MSEoptmPEN=zeros(nmv,1);
MSEbootPEN=zeros(nmv,nhV);
effzk_h=zeros(n,nhV);
biaskh=zeros(n,nhV);
FDHk=zeros(n,1);
effzk_hp=zeros(n,1);
effzk_lscv=zeros(n,1);
effzk_hpc1=zeros(n,1);
effzk_hpc2=zeros(n,1);
derkh=zeros(n,1);

epsh = 1/sqrt(n); % value of epsilon for derivative
% % % % Evaluation of the "bias" term (the derivative with hlscv)
hd=hlscv;
tic
for k=1:n
    xk=X(k,:);
    yk=Y(k,:);
    zk=Z(k,:);
    effk = fdhxkykNoScaling(X,Y,xk,yk);
    FDHk(k)=effk(3);% output

    % LSCV-corrected Bandwidth Conditional efficiency in original sample
    flagzkp = all(abs(Z - repmat(zk,n,d))<= hp',2);
    Xzkp = X(flagzkp,:);
    Yzkp = Y(flagzkp,:);
    nzkp =sum(flagzkp);
    effzkp = fdhxkykNoScaling(Xzkp,Yzkp,xk,yk);
    effzk_hp(k)=effzkp(3);% output

%   LSCV Bandwidth Conditional efficiency in original sample
    flagz0 = all(abs(Z - repmat(zk,n,d))<= hlscv',2);
    Xzk0 = X(flagz0,:);
    Yzk0 = Y(flagz0,:);
    effzk0 = fdhxkykNoScaling(Xzk0,Yzk0,xk,yk);
    effzk_lscv(k)=effzk0(3);% output

%    Conditional efficiency for hd+epsh in original sample for derivative
    flagzkpc = all(abs(Z - repmat(zk,n,d))<= (hd+epsh)',2);
    Xzkpc = X(flagzkpc,:);
    Yzkpc = Y(flagzkpc,:);
    effzkpc2 = fdhxkykNoScaling(Xzkpc,Yzkpc,xk,yk);
    effzk_hpc2(k)=effzkpc2(3);% output

%    Conditional efficiency for hd-epsh in original sample for derivative
    flagzkpc = all(abs(Z - repmat(zk,n,d))<= (hd-epsh)',2);
    Xzkpc = X(flagzkpc,:);
    Yzkpc = Y(flagzkpc,:);
    effzkpc1 = fdhxkykNoScaling(Xzkpc,Yzkpc,xk,yk);
    effzk_hpc1(k)=effzkpc1(3);% output

    derkh(k) = (effzk_hpc2(k) - effzk_hpc1(k))/(2*epsh);% Z UNIVARIATE (otherwise gradient vector)

%    Basic conditional eff in original sample with hc in the grid
%   (we need the full matrix effzk(i,ih) near the end for final evaluation)
    hhg=hV;
    for ih=1:nhV
        hc=hhg(ih);% current value of h in the grid
        % Conditional efficiency in original sample for the grid bandwidths
        flagzk = all(abs(Z - repmat(zk,n,d))<= hc',2);
        Xzk = X(flagzk,:);
        Yzk = Y(flagzk,:);
        nzk =sum(flagzk);
        flagxk = all( Xzk <= repmat(xk,nzk,d),2);
        nxk = sum(flagxk);
        if nxk > 0
            effzk = fdhxkykNoScaling(Xzk,Yzk,xk,yk);
            effzk_h(k,ih)=effzk(3);% output
        else
            fprintf('WARNING: no points below or = xk, h is too small increase h \n')
            return
        end
    end
end
toc
% Spline Smoothing of Derivatives: the next comes from Eilers and Marx (1996)
```

```
spder = pnormal(Z,derkh,20,3,ord_pen,pen_spl,1,1);
title('Estimate of derivatives as function of Z')
xlabel('values of Z_i')
derspline = spder.muhat;

% Bootstrap for each value of m in mgrid
Pnans=zeros(nmv,nhV);
for im=1:nmv
    mboot=mv(im);% current value of m (subsample size)
    fprintf('Value of mboot is %4.0f \n',mboot)
    for ih=1:nhV
        hc=hV(ih);% current value of h
        % BOOTSTRAP CALCULATIONS
        effzk_boot=zeros(n,B);
        numnans=0;
        parfor b=1:B
            % generate random labels
            rng(state+b-1);% for reproducability in parfor
            rN=randperm(n)';labelb=rN(1:mboot); % resampling WITHOUT replacement
            Xb=X(labelb,:);Yb=Y(labelb,:);Zb=Z(labelb,:);
            for k=1:n
                xk=X(k,:);
                yk=Y(k,:);
                zk=Z(k,:);
                flagbzk = all(abs(Zb - repmat(zk,mboot,d))<= hc',2);
                Xbzk = Xb(flagbzk,:);
                Ybzk = Yb(flagbzk,:);
                nbzk =sum(flagbzk);
                flagbxk = all( Xbzk <= repmat(xk,nbzk,d),2);
                nbxk = sum(flagbxk);
                if nbxk > 0
                    effbzk = fdhxkykNoScaling(Xbzk,Ybzk,xk,yk);
                    effzk_boot(k,b)=effbzk(3);
                else
                    % Jeong-Simar JMVA2006: add the point to bootsample if NaN
                    effzk_boot(k,b)=1;
                    numnans=numnans + 1;
                end
            end
        end % end bootstrap loop b=1:B
%       For the "Estimation-FDH" part correct for sample size to get the appropriate rate wrt n
        MSEbootPEN(im,ih) =  nanmean(nanmean((repmat(derspline*hc,1,B) +...
            (mboot/n)^(kappa)*(effzk_boot -repmat(effzk_lscv,1,B))).^2,2));
        Pnans(im,ih) = numnans/(n*B);% percentage of Nans
    end % end loop ih=1:nvh
    [MSE_opt,Imin]=min(MSEbootPEN(im,:));
    h_opt = hV(Imin);
    ih_optmPEN(im)=Imin;
    hoptmPEN(im)=h_opt;
    MSEoptmPEN(im)=MSE_opt;
%%   If wanted, see the Pictures and check that the bounds for the grid in h is OK
%    figure
%    plot(hV,MSEbootPEN(im,:),'r-','LineWidth',2)
%    title(['For m = ',num2str(mboot),' AMSE(h) as function of h'],'FontSize',14)
%    xlabel('Values of h','FontSize',14)
end % end loop im=1:nmv
fprintf('====================\n End of the Bootstrap \n====================\n')

CPU=etime(clock,t0);
fprintf(' Case-Example is %10s \n',Ex)
fprintf(' Sample size is n = %4.0f \n',n)
fprintf(' Elapsed time  = %15.4f  seconds for B = %6.0f \n',[CPU,B])
fprintf(' Seed = %10.4f \n',state)
fprintf('-----------------------------------------------------------\n')

wind=1; % define the window width choose 1 or 2 (only 2 if mgrid >20)
fprintf(' Window width for measuring volatility = %3.0f \n',2*wind +1)
jvec=zeros(nmv-2*wind,1);
CV1 = zeros(nmv-2*wind,1);
for j=1:nmv-2*wind
    jvec(j)=j+wind;
    CV1(j)= std(MSEoptmPEN((jvec(j)-wind:jvec(j)+wind)),0);% volatility of MSEopt
end
figure
plot(mv(jvec),CV1,'o','LineWidth',2)
title('volatility for MSEbootPEN','FontSize',14)
```

```
[~,jopt]=min(CV1);
imopt=jvec(jopt);
mopt_pen=mv(imopt);
HOPT_pen=hoptmPEN(imopt);
MSEOPT_pen=MSEoptmPEN(imopt);
ihOPT_pen = ih_optmPEN(imopt);
fprintf('PEN: mOPT ihOPT   Hopt    MSEopt     RMSEopt \n')
fprintf(' %3.0f %3.0f %9.6f %9.6f %9.6f \n',[mopt_pen ihOPT_pen HOPT_pen MSEOPT_pen sqrt(MSEOPT_pen)])
figure
plot(hV,MSEbootPEN(imopt,:),'-')
title(['Bootstrap estimation of AMSE(h), m-opt = ',num2str(mopt_pen)],'FontSize',14)
xlabel('Values of h','FontSize',14)

% FINAL EVALUATION of THE n efficiency score and estimation of the AMSE for this one sample
effzk_hopt = effzk_h(:,ihOPT_pen);% WE NEED HERE the full MATRIX effzk_h: (n x nhV) computed above!
BIASZ_hp= nanmean(effzk_hp - trueff);
BIASZ_lscv= nanmean(effzk_lscv - trueff);
BIAS_FDH= nanmean(FDHk - trueff);
BIASZ_pen= nanmean(effzk_hopt - trueff);

MSEZ_hp= nanmean((effzk_hp - trueff).^2);
MSEZ_lscv= nanmean((effzk_lscv - trueff).^2);
MSE_FDH= nanmean((FDHk - trueff).^2);
MSEZ_pen= nanmean((effzk_hopt - trueff).^2);

fprintf('==========================\n')
fprintf(' Bandwidths selected :\n')
fprintf(' h_{LSCV} = %10.6f \n',hlscv)
fprintf(' h_{CORR} = %10.6f \n',hp)
fprintf(' h_{BOOT} = %10.6f \n',HOPT_pen)
fprintf('============================================================\n')
fprintf('  Bias, AMSE and RAMSE within that paricular sample ! \n')
fprintf('---------------  Bias            AMSE            RAMSE------ \n')
fprintf('FDH    :%15.6f %15.6f %15.6f \n' ,[BIAS_FDH   MSE_FDH sqrt(MSE_FDH)])
fprintf('LSCV   :%15.6f %15.6f %15.6f \n' ,[BIASZ_lscv   MSEZ_lscv sqrt(MSEZ_lscv)])
fprintf('CORR   :%15.6f %15.6f %15.6f \n' ,[BIASZ_hp   MSEZ_hp sqrt(MSEZ_hp)])
fprintf('BOOT   :%15.6f %15.6f %15.6f \n' ,[BIASZ_pen   MSEZ_pen sqrt(MSEZ_pen)])
fprintf('============================================================\n')
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function eff = fdhxkykNoScaling(X,Y,xk,yk)
% FDH FOR A FIXED POINT (xk,yk) with reference set (X,Y)
%            X   : Matrix of input(s)  (n x p)
%            Y   : Matrix of output(s) (n x q)
%            xk  : vector of input(s)  (1 x p)
%            yk  : vector of input(s)  (1 x d)
%        eff  :  results (1 x 3)
%                     Number of dominating units
%                     Efficiency score in input
%                     Efficiency score in output
% Written by L. SIMAR (april, 2002)
[n,p] = size(X);
q= size(Y,2);
xi=ones(n,1)*xk;
yi=ones(n,1)*yk;
flagx=(X<=xi);
flagy=(Y>=yi);
flagxy=[flagx ,flagy];
flag=all(flagxy,2);
ydi=Y(flag,:);xdi=X(flag,:);
ndi=size(xdi,1);
Status=ndi-1;
if ndi==0
   %  INPUT orientation
   flagy=(Y>=yi);
   flagy=all(flagy,2);
   XM=X(flagy,:);
   nxm=size(XM,1);
   if nxm==0
       I_eff=1;
   else
      xkv=ones(nxm,1)*xk;
      ratioxk=XM./xkv;
      if p==1
```

```
I_eff=min(ratioxk,[],1);
else
I_eff=min(max(ratioxk,[],2),[],1);
       end
   end
   %  OUTPUT orientation
   flagx=(X<=xi);
   flagx=all(flagx,2);
   YM=Y(flagx,:);
   nym=size(YM,1);
   if nym==0
      O_eff=1;
   else
      ykv=ones(nym,1)*yk;
      ratioyk=YM./ykv;
      if q==1
O_eff=max(ratioyk,[],1);
else
O_eff=max(min(ratioyk,[],2),[],1);
      end
   end
   eff=[Status,I_eff,O_eff];
   return
end

ratioxi=xdi./(ones(ndi,1)*xk);
I_eff=min(max(ratioxi,[],2),[],1);
ratioyi=ydi./(ones(ndi,1)*yk);
O_eff=max(min(ratioyi,[],2),[],1);
eff=[Status,I_eff,O_eff];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function  CV=CCDFnonsmth_lscv_NEW(hz,Y,Z,n,d,kernelz)
%  Evaluate the LSCV criterion for estimating a conditional CDF: F(y|Z=z)for bandwidth h=hz
%   see details in Li, Lin, Racine, JBES2013 paper
%  Written by L. Simar, August, 2014
%
%  kerz specify the UNIVARIATE kernels used for Z
%  We use product kernels for  d >1
%    hz is COLUMN vectors

% Some kernel function
Kepan = @(u) (abs(u) <=1).*(1 - u.^2)*3/4; % |u| <= 1
Kgaus = @(u) exp(-u.^2/2)/sqrt(2*pi); % u\in R
Kquar = @(u) (abs(u) <=1).*(1 - u.^2).^2 *15/16;  % |u| <= 1
Kunif = @(u) 0.5*(abs(u) <=1);   % |u| <= 1

wz=ones(n,1); % weight function to avoid dividing by zero below
Dhz=diag(ones(d,1)./hz); % this is diag matrix d x d
CVi=zeros(n,1);
for i=1:n
    zi=Z(i,:);
    yi=Y(i,:);
    %     leave-one out sample
    lo=[(1:i-1)'; (i+1:n)'];
    Yi=Y(lo,:);
    Zi=Z(lo,:);
    %     Kernel for Z
    tempz=(Zi-repmat(zi,n-1,1)); % this is a (n-1) x d matrix
    tempzh=tempz*Dhz;
    switch lower(kernelz)
    case ('gauss')
        kerzi= Kgaus(tempzh)*Dhz;
    case ('quart')
        kerzi= Kquar(tempzh)*Dhz;
    case ('epan')
        kerzi= Kepan(tempzh)*Dhz;
    case ('unif')
        kerzi= Kunif(tempzh)*Dhz;
    otherwise
        disp('Specify corect Kernel method for Z ''Epan'' or ''Quart''')
        CV=NaN;
        return
    end
    kerz=prod(kerzi,2); % Product kernel: a (n-1) x 1 vector
    mzi=mean(kerz); % this is \hat f_{-(i)}(zi) (Leave-i-Out)
    Fnsi=zeros(n-1,1); % non-smoothed F
    Ii1 = all(repmat(yi,n-1,1) <= Yi,2); % this is (n-1) x 1 vector
```

```
    for k1=1:n-1
        j1=lo(k1); % this insure that j1 ne i (the i of the outer loop)
        yj1=Y(j1,:);
        Ii2 = all(Yi <= repmat(yj1,n-1,1),2);
        numyzi= kerz.*Ii2;
        Fnsi(k1) = mean(numyzi)/(mzi); % this is non-smoothed hatF_{-i} (yj1 | xi)
    end
    CVi(i) = nanmean((Ii1 - Fnsi).^2);
end
CV= nanmean(CVi);
```

References

[1] Baležentis, T., De Witte, K. (2015). One-and multi-directional conditional efficiency measurement: Efficiency in Lithuanian family farms. *European Journal of Operational Research*, 245(2), 612–622.

[2] Bădin, L., Daraio, C. (2011), Explaining Efficiency in Nonparametric Frontier Models. Recent developments in statistical inference, in *Exploring research frontiers in contemporary statistics and econometrics*, ed. by I. Van Keilegom and P.W. Wilson, Springer-Verlag Berlin Heidelberg.

[3] Bădin, L., Daraio, C. and L. Simar (2010). Optimal Bandwidth Selection for Conditional Efficiency Measures: a Data-driven Approach. *European Journal of Operational Research*, 201 (2), 633–640.

[4] Bădin, L., Daraio, C. and L. Simar (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research*, 223 (3), 818–833.

[5] Bădin, L., Daraio, C. and L. Simar (2014). Explaining inefficiency in nonparametric production models: the state of the art. *Annals of Operations Research*, 214 (1), 5–30.

[6] Broekel, T. (2012). Collaboration intensity and regional innovation efficiency in Germany - a conditional efficiency approach. *Industry and Innovation*, 19(2), 155–179.

[7] Broekel, T., Schlump, C. (2009). The importance of R&D subsidies and technological infrastructure for regional innovation performance-A conditional efficiency approach. *Papers in Evolutionary Economic Geography*, 9.

[8] Carvalho, P., Marques, R. C. (2011). The influence of the operational environment on the efficiency of water utilities. *Journal of environmental management*, 92(10), 2698–2707.

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

[10] Cordero, J. M., Alonso-Morn, E., Nuo-Solinis, R., Orueta, J. F., Arce, R. S. (2015). Efficiency assessment of primary care providers: a conditional nonparametric approach. *European Journal of Operational Research*, 240(1), 235–244.

[11] Cordero, J. M., Pedraja-Chaparro, F., Pisaflores, E. C., Polo, C. (2016). Efficiency assessment of Portuguese municipalities using a conditional nonparametric approach. *Journal of Productivity Analysis*, 1–24.

[12] Cordero, J. M., Santin, D., Simancas, R. (2017). Assessing European primary school performance through a conditional nonparametric model. *Journal of the Operational Research Society*, 68(4), 364–376.

[13] D'Alfonso, T., Daraio, C., Nastasi, A. (2015). Competition and efficiency in the Italian airport system: new insights from a conditional nonparametric frontier analysis. *Transportation Research Part E: Logistics and Transportation Review*, 80, 20–38.

[14] Daraio, C. and L. Simar (2005). Introducing Environmental Variables in Nonparametric Frontier Models: a Probabilistic Approach. *Journal of Productivity Analysis*, 24, 93–121.

[15] Daraio, C. and L. Simar (2006). A robust nonparametric approach to evaluate and explain the performance of mutual funds. *European Journal of Operational Research*, Vol 175 (1), 516–542.

[16] Daraio, C. and L. Simar (2007a). *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and applications*, Springer, New York.

[17] Daraio, C. and L. Simar (2007b). Conditional nonparametric frontier models for convex and non convex technologies: A unifying approach. *Journal of Productivity Analysis*, 28, 13–32.

[18] Daraio, C., Simar, L. and P. W. Wilson, (2018). Central limit theorems for conditional efficiency measures and tests of the separability condition in non-parametric, two-stage models of production. *The Econometrics Journal*, doi:10.1111/ectj.12103.

[19] Daraio, C., Bonaccorsi, A., Simar, L. (2015). Rankings and university performance: A conditional multidimensional approach. *European Journal of Operational Research*, 244(3), 918–930.

[20] De Witte, K., Geys, B. (2011). Evaluating efficient public good provision: Theory and evidence from a generalised conditional efficiency model for public libraries. *Journal of urban economics*, 69(3), 319–327.

[21] Eilers, P.H.C. and B.D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11,2,89–121.

[22] Filippetti, A., Peyrache, A. (2015). Labour productivity and technology gap in European regions: A conditional frontier approach. *Regional Studies*, 49(4), 532–554.

[23] Fuentes, R., Torregrosa, T., Ballenilla, E. (2015). Conditional order-m efficiency of wastewater treatment plants: the role of environmental factors. *Water*, 7(10), 5503–5524.

[24] Guerrini, A., Romano, G., Mancuso, F., Carosi, L. (2016). Identifying the performance drivers of wastewater treatment plants through conditional order-m efficiency analysis. *Utilities Policy*, 42, 20–31.

[25] Haelermans, C., De Witte, K. (2012). The role of innovations in secondary school performanceEvidence from a conditional efficiency model. *European Journal of Operational Research*, 223(2), 541–549.

[26] Halkos, G. E., Managi, S. (2016). Measuring the effect of economic growth on countries environmental efficiency: a conditional directional distance function approach. *Environmental and Resource Economics*, 1–23.

[27] Halkos, G. E., Tzeremes, N. G. (2010). Measuring biodiversity performance: A conditional efficiency measurement approach. *Environmental Modelling & Software*, 25(12), 1866-1873.

[28] Halkos, G. E., Tzeremes, N. G. (2011). A conditional nonparametric analysis for measuring the efficiency of regional public healthcare delivery: An application to Greek prefectures. *Health policy*, 103(1), 73–82.

[29] Halkos, G. E., Tzeremes, N. G. (2013a). A conditional directional distance function approach for measuring regional environmental efficiency: Evidence from UK regions. *European Journal of Operational Research*, 227(1), 182–189.

[30] Halkos, G. E., Tzeremes, N. G. (2013b). National culture and eco-efficiency: an application of conditional partial nonparametric frontiers. *Environmental Economics and Policy Studies*, 15(4), 423–441.

[31] Halkos, G. E., Tzeremes, N. G. (2013c). Modelling the effect of national culture on countries innovation performances: A conditional full frontier approach. *International review of applied economics*, 27(5), 656–678.

[32] Halkos, G. E., Tzeremes, N. G. (2014). Measuring the effect of Kyoto protocol agreement on countries environmental efficiency in CO2 emissions: an application of conditional full frontiers. *Journal of Productivity Analysis*, 41(3), 367–382.

[33] Hall, P., Racine, J.S. and Q. Li (2004). Cross-Validation and the Estimation of Conditional Probability Densities. *Journal of the American Statistical Association*, Vol 99, 486, 1015–1026.

[34] Jeong, S.O. , B. U. Park and L. Simar (2010). Nonparametric conditional efficiency measures: asymptotic properties. *Annals of Operations Research*, 173, 105–122.

[35] Jeong, S.O. and L. Simar (2006). Linearly interpolated FDH efficiency score for nonconvex frontiers. *Journal of Multivariate Analysis*, 97, 2141–2161.

[36] Kourtesi, S., Fousekis, P., Polymeros, A. (2012). Conditional efficiency estimation with environmental variables: evidence from Greek cereal farms. *Sci Bull-Econ Sci*, 11, 43–52.

[37] Li, Q., Lin, J. and J.S. Racine (2013). Optimal Bandwidth Selection for Nonparametric Conditional Distribution and Quantile Functions. *Journal of Business & Economic Statistics*, Vol 31 (1), 57–65.

[38] Li, Q. and J. Racine (2008). Nonparametric Estimation of Conditional CDF and Quantile Functions with Mixed Categorical and Continuous Data. *Journal of Business & Economic Statistics*, Vol 26 (4), 423–434.

[39] Mastromarco, C. and L. Simar (2015). Effect of Time and FDI on Catching-up: news insights from a Conditional Nonparametric Frontier Analysis. *Journal of Applied Econometrics*, 30, 826–847.

[40] Matousek, R., Tzeremes, N. G. (2016). CEO compensation and bank efficiency: An application of conditional nonparametric frontiers. *European Journal of Operational Research*, 251(1), 264–273.

[41] Minviel, J. J., De Witte, K. (2017). The influence of public subsidies on farm technical efficiency: A robust conditional nonparametric approach. *European Journal of Operational Research*, 259(3), 1112–1120.

[42] Park, B., Simar, L. and C. Weiner (2000). The FDH estimator for productivity efficiency scores: asymptotic properties. *Econometric Theory* 16, 855–877.

[43] Park, B., Simar, L. and V. Zelenyuk (2008). Local likelihood estimation of truncated regression and its partial derivatives: Theory and application. *Journal of Econometrics*, 146 (1), 185–198.

[44] Politis, D.N., J.P. Romano and M. Wolf (2001). On the asymptotic Theory of subsampling. *Statistica Sinica*, 11, 1105–1124.

[45] Serra, T., Lansink, A. O. (2014). Measuring the impacts of production risk on technical efficiency: A state-contingent conditional order-m approach. *European Journal of Operational Research*, 239(1), 237–242.

[46] Simar, L., Vanhems, A. and I. Van Keilegom (2016). Unobserved heterogeneity and endogeneity in nonparametric frontier estimation. *Journal of Econometrics*, 190(2), 360–373.

[47] Simar, L. and P.W. Wilson (2007). Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes. *Journal of Econometrics*, Vol 136, 1, 31–64.

[48] Simar, L. and P.W. Wilson (2011a). Inference by the m out of n bootstrap in Nonparametric Frontier Models. *Journal of Productivity Analysis*, 36, 33–53.

[49] Simar, L. and P.W. Wilson (2011b). Two-Stage DEA: *Caveat Emptor*. *Journal of Productivity Analysis*, 36, 205–218.

[50] Simar, L. and P.W. Wilson (2013). Estimation and inference in nonparametric frontier models: Recent developments and perspectives, Foundations and Trends®in Econometrics, Vol. 5: No 3-4, 183–337.

[51] Simar, L. and P.W. Wilson (2015). Statistical Approaches for Nonparametric Frontier Models: A Guided Tour. *International Statistical Review*, 83(1), 77–110.

[52] Tzeremes, N. G. (2015). Efficiency dynamics in Indian banking: A conditional directional distance approach. *European Journal of Operational Research*, 240(3), 807–818.

[53] Varabyova, Y., Blankart, C.R., Torbica, A. and J. Schreyögg (2016a). Comparing the Efficiency of Hospitals in Italy and Germany: Nonparametric Conditional Approach based on Partial Frontier. *Health Care Management Science*, DOI: 10.1007/s10729-016-9359-1.

[54] Varabyova, Y., Blankart, C.R. and J. Schreyögg (2016b). Using Nonparametric Conditional Approach to Integrate Quality into Efficiency Analysis: Empirical Evidence from Cardiology Departments. *Health Care Management Science*, DOI: 10.1007/s10729-016-9372-4.

[55] Varabyova, Y. and J. Schreyögg (2017). Integrating Quality into Nonparametric Analysis of Efficiency: A Simulation Comparison of Popular Methods. *Annals of Operations Reseacrh*, DOI 10.1007/s10479-017-2628-7.

[56] Verschelde, M., Rogge, N. (2012). An environment-adjusted evaluation of citizen satisfaction with local police effectiveness: Evidence from a conditional Data Envelopment Analysis approach. *European Journal of Operational Research*, 223(1), 214–225.

[57] Zschille, M. (2015). Consolidating the water industry: an analysis of the potential gains from horizontal integration in a conditional efficiency framework. *Journal of Productivity Analysis*, 44(1), 97–114.