I N S T I T U T   D E   S T A T I S T I Q U E

B I O S T A T I S T I Q U E   E T

S C I E N C E S   A C T U A R I E L L E S

(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Nonparametric Least Squares Methods for Stochastic Frontier Models

SIMAR, L., VAN KEILEGOM, I. and V. ZELENYUK

Nonparametric Least Squares Methods for Stochastic Frontier Models

Léopold Simar[*]

leopold.simar@uclouvain.be

Ingrid Van Keilegom [*,§]

ingrid.vankeilegom@uclouvain.be.

Valentin Zelenyuk[**]

v.zelenyuk@uq.edu.au

March 21, 2014

Abstract

When analyzing productivity and efficiency of firms, stochastic frontier models are very attractive because they allow, as in typical regression models, to introduce some noise in the Data Generating Process. Most of the approaches so far have been using very restrictive fully parametric specified models, both for the frontier function and for the components of the stochastic terms. Recently, local MLE approaches were introduced to relax these parametric hypotheses. However, the high computational complexity of the latter makes them difficult to use, in particular if bootstrap-based inference is needed. In this work we show that most of the benefits of the local MLE approach can be obtained with less assumptions and involving much easier, faster and numerically more robust computations, by using nonparametric least-squares methods. Our approach can also be viewed as a semi-parametric generalization of the so-called "modified OLS" that was introduced in the parametric setup. If the final evaluation of individual efficiencies requires, as in the local MLE approach, the local specification of the distributions of noise and inefficiencies, it is shown that a lot can be learned on the production process without such specifications. Even elasticities of the mean inefficiency can be analyzed with unspecified noise distribution and a general class of local one-parameter scale family for inefficiencies. This allows to discuss the variation in inefficiency levels with respect to explanatory variables with minimal assumptions on the Data Generating Process. Our method is illustrated and compared with other methods with a real data set.

Key Words: Stochastic Frontier Analysis, Nonparametric Frontiers, Efficiency and productivity Analysis, Local Polynomial Least-Squares.

JEL Classification: C1,C14, C13

1   The Background and The Basic Model

In productivity and efficiency analysis, researchers are primarily interested in two aspects: (i) the estimation of a function characterizing the production frontier and its characteristics (marginal productivity, elasticities, etc.), and (ii) the explanation of variation in inefficiency.[1] One of the most popular approaches for studying these aspects is referred to as stochastic frontier analysis (SFA), introduced by Aigner, Lovell and Schmidt (1977) and Meusen and van den Broek (1977) (henceforth ALSMB). The SFA paradigm has a very appealing feature relative to other methods—it allows the presence of both an inefficiency term modeling the distance of an observation to the optimal frontier and the more traditional error term (as in most regression models) allowing for noise. In this paper we work in this paradigm but in a non-parametric context.

Specifically, we consider a set of i.i.d. random variables (X_i, Y_i, Z_i), for $i = 1, \ldots, n$ where $X_i \in \mathbb{R}^p$ are the inputs, $Z_i \in \mathbb{R}^d$ represent a set of heterogeneous conditions (these could be environmental or regulatory conditions that are not inputs in the usual sense but that may influence the production process and that are not necessarily controlled by the producer) and $Y_i \in \mathbb{R}$ is the output that can be produced. The joint pdf of (X, Z, Y) can be decomposed into a joint marginal for (X, Z) and a conditional pdf for Y given (X, Z). The conditional of Y given $X = x$ and $Z = z$ is characterized through

$$Y = m(x, z) - U + V, \tag{1.1}$$

where $m(x, z)$ is the production frontier, $U|X = x, Z = z \sim D^+(\mu_U(x, z), var_U(x, z))$ with $D^+(\cdot, \cdot)$ being a positive random variable with mean $\mu_U(\cdot, \cdot)$ and variance $var_U(\cdot, \cdot)$, and $V|X = x, Z = z \sim D(0, var_V(x, z))$ with $D(0, \cdot)$ being a real random variable with mean zero and variance $var_V(\cdot, \cdot)$. We also assume that, conditionally on (X, Z), U and V are independent random variables, where V has a symmetric distribution around zero, while U is a positive random variable whose skewness will be of central interest.

As in parametric SFA models, the observation of Y is adjusted by some possible inefficiency level U and by some statistical noise V. The two components U and V are unobserved random variables which may vary with the inputs X as well as with the other variables Z. The particular and very common case where the variables Z do not influence the technology but only the inefficiency or noise corresponds to the so-called "separability condition" de-

---

[1]We focus the presentation in the output oriented case, where we search for the maximal level of the output, given the level of the inputs, but we could have a similar approach when we try to model a cost function that would search for the minimal costs given the level of the outputs.

scribed in Simar and Wilson (2007). In our set-up, this would correspond to the particular case where $m(x, z) = m(x)$, which we do not impose.

By contrast with parametric approaches, we assume that the production frontier $m(\cdot, \cdot)$ is completely *unknown* to the researcher. The general goal of estimation is to obtain information about the production technology (scale elasticities, marginal productivity of inputs, etc.), and about inefficiency (whether it is present and how it is related to various factors amongst (x, z)), given the sample of i.i.d. observations $\mathcal{S}_n = \{(X_i, Z_i, Y_i) \mid i = 1, \ldots, n\}$.

Note that if we assume a particular distribution for U and V, with homoskedasticity of both terms, and if we impose a parametric specification for m, then we obtain the parametric SFA setup sparked by ALSMB. This classical, fully parametric, homoskedastic SFA setup is a good start for explorations, yet might be very restrictive, imposing a lot of structure which is hardly possible to check in practice. Typical specification is a log-linear model for the production function m, normal distribution $N(0, \sigma_V^2)$ for the noise and some specified one-sided distribution for U, such as $\text{Expo}(\lambda)$ or $|N(0, \sigma_U^2)|$ (see Kumbhakar and Lovell, 2000 for a nice survey on these fully parametric models). Various extensions have been developed in recent years to reduce various restrictions. Among the earliest of such extensions is the work by Kneip and Simar (1996), who used Nadaraya-Watson estimator, and by Fan, Li and Weersink (1996) who used the quasi-likelihood method by replacing the true production function relationship $m(x)$, in ALSMB-type likelihood formulation, with its Nadaraya-Watson estimate.

One of the most recent non-parametric extensions of ALSMB is due to Kumbhakar et al. (2007), who proposed using local MLE for estimating (1.1) without a parametric assumption on m, yet still using semi-parametric assumptions about U and V. Typically, they use a local parametric stochastic assumption on U and V, like e.g. $U|X = x, Z = z \sim |N(0, \sigma_U^2(x, z))|$ and $V|X = x, Z = z \sim N(0, \sigma_V^2(x, z))$, but keeping the variance functions $\sigma_U^2(x, z)$ and $\sigma_V^2(x, z)$ as unspecified functionals. Their approach has appealing theoretical properties and impressive MC performances, even when the parametric stochastic part of the assumptions about U and V are misspecified. However, the high computational complexity, due to the need to optimize the local likelihood function in many parameters, makes the local MLE rather disadvantageous, especially when bootstrap-based inference is needed.

In our work, we show that much of positives from the local MLE approach of Kumbhakar et al. (2007) can be obtained with less assumptions and, more importantly, involving much easier, faster and more robust estimation. Our approach can be viewed as a non- or semi-parametric version of the "modified OLS" (MOLS) method that was introduced as an

2

alternative to MLE method for SFA in parametric setups. The basic idea of the method is to first estimate an average production function and some moments of the convoluted error term. We show then how we can, under rather wide assumptions, identify the local asymmetry of this error allowing to analyze the local presence of inefficiency. In particular we will see that elasticities of $\mu_U(x, z)$ can be analyzed with unspecified noise distribution and for inefficiencies belonging to a general class of local one-parameter scale family. Then, upon adding appropriate local, and so flexible, parametric assumptions on the moments of U and V, we can identify the frontier itself. These assumptions will provide estimators of the frontier that are more robust than the one obtained with the local likelihood approach, because we use only local moment restrictions on U and V and not their full local distribution.

At this point it is useful to note that the interests of researchers in production efficiency measurement can be more broadly classified into, at least, four different problems: (i) the estimation and inference for production relationship (output vs. inputs), (ii) the inference on existence of inefficiency, (iii) the estimation and inference of the stochastic frontier level and/or of efficiency scores for particular observations (e.g., firms) and/or the corresponding efficiency ranking, and (iv) the analysis of determinants of inefficiency. Only the problem (iii) will require some localized parametric assumptions (especially for the cross-sectional data) while the problems (i) and (ii) can be done non-parametrically, and problem (iv) can be handled with unspecified noise distribution and inefficiencies belonging to a general class of local one-parameter scale family. Analogously to the case when one needs to dress a black tuxedo only for some special occasions while he can dress what is most comfortable (although not so glamorous) on others, our strategy will be to impose any local parametric assumptions only when we cannot do otherwise, while we will try to do the rest non-parametrically.

The paper is organized as follows: Section 2 discusses the estimation of the average production function and of the moments of the corresponding error terms. Section 3 gives the basis of the semi-parametric version of the MOLS approach, describing what we need to specify in order to recover the frontier level and the individual inefficiencies. Section 4 focuses on the analysis of determinants of inefficiency levels in a robust and flexible nonparametric setup, providing a test of significancy of the covariates on the efficiency level. Section 5 illustrates the estimation procedure with a real data example and Section 6 concludes. The technical details (asymptotic properties of the nonparametric estimators used in Section 3 and 4) are described in an heuristic way in the Appendix A.

2   Average Production Function

A relatively easy task is to estimate the so-called "average production function" relationship (i.e., when the presence of inefficiency term U is simply ignored), and this can be done in a fully non-parametric way. Indeed, letting $\varepsilon = V - U + \mu_U(x, z)$, and $r_1(x, z) = m(x, z) - \mu_U(x, z)$ we can rewrite (1.1) as

$$Y = r_1(x, z) + \varepsilon. \tag{2.1}$$

Now, we note that $E(\varepsilon|x, z) = 0$, and $Var(\varepsilon|x, z) = var_U(x, z) + var_V(x, z) \in (0, \infty)$, where $var_U(x, z)$ $(var_V(x, z))$ is a function of (x, z) that represents the conditional variance of $U|x, z$ (of $V|x, z$), resp.).

2.1   Nonparametric estimation of $r_1(x, z)$

With model (2.1) we are back to a standard nonparametric regression problem and we can use standard methods to estimate $r_1(x, z) = E(Y|x, z)$ from the sample $\mathcal{S}_n$. Our preference falls onto the local polynomial least squares (LPLS) estimator, which is a fully non-parametric approach with good asymptotic properties and is relatively easy and fast to compute (e.g., when compared to the local MLE approach). One of the useful features of the LPLS is that, similarly as in the local MLE, in one estimation this approach can produce consistent and asymptotically normal estimates of $r_1(x, z)$, which we will denote here as $\widehat{r}_1(x, z)$, and estimates of its kth-order partial derivatives (if the order of the polynomial is chosen to be at least of order k). If we are interested in estimation of partial derivatives of the average production relationship (e.g., for estimating marginal productivity of inputs, scale elasticity, etc.), then it might be worth using the local quadratic or cubic fit, but for notational simplicity we will focus our discussion on the local linear and local constant fits that usually can do the job very well.

It is important to emphasize here that $\widehat{r}_1(x, z)$ will be an estimate not of the production frontier $m(x, z)$ but of $r_1(x, z) = m(x, z) - \mu_U(x, z)$, which we refer to as the average production function. Since $\mu_U(x, z) \geq 0$, we have $r_1(x, z) \leq m(x, z)$, $\forall(x, z) \in \mathbb{R}^{p+d}$. Thus, clearly, $r_1(x, z) = m(x, z)$, $\forall(x, z) \in \mathbb{R}^{p+d}$, if and only if $\mu_U(x, z) = 0$, i.e., if and only if there is no inefficiency. Hence, if there is some inefficiency, then $\widehat{r}_1(x, z)$ would be a downward-biased estimator of $m(x, z)$ at any level of inputs. Moreover, the bias is given exactly by $\mu_U(x, z)$, and so the bias is varying with (x, z) unless $\mu_U(x, z)$ is a constant equal to $E(U)$. We will see below that if $U|x, z$ belongs to a local one parameter scale family, $\mu_U(x, z)$ can be estimated non-parametrically at a multiplicative constant. This would allow to test if $\mu_U(x, z)$ is con-

stant or if some of the covariates are significant, but also to analyze non-parametrically its elasticities, and this, without specifying neither the particular scale family nor the distribution of the noise. In the case of constant $\mu_U(x, z)$, most of the characteristics of $r_1(x, z)$ are directly transferred to $m(x, z)$, except its level.[2]

Indeed, from a sample of i.i.d. data $\{(Y_i, X_i, Z_i) \; : \; i = 1, ...n\}$ we can estimate the local linear estimate of $r_1(x, z)$ by solving, for any given value (x, z)

$$(\alpha_{x,z}, \beta_{x,z}) = \arg\min_{\alpha, \beta} \sum_{i=1}^{n} [Y_i - (\alpha + \beta'((X_i, Z_i) - (x, z)))]^2 \, K_h\left(\frac{(X_i, Z_i) - (x, z)}{h}\right),$$

where $K_h\big(((X_i, Z_i) - (x, z))/h\big)$ stands (with some abuse of notations) for a product kernel for the $p + d$ components of (X, Z) with h denoting the $p + d$ bandwidths. Then we have

$$\widehat{r}_1(x, z) = \alpha_{x,z} \tag{2.2}$$

$$\widehat{\nabla} r_1(x, z) = \beta_{x,z}, \tag{2.3}$$

where the second equation provides an estimate of the gradient of $r_1(x, z)$ at (x, z). Usually, the bandwidths are selected by least-squares cross validation. The interesting feature is that these estimators do not require any nonlinear optimization procedure, they can be obtained by simple linear algebra (see Fan and Gijbels, 1996 or Li and Racine, 2007). Under mild regularity conditions and appropriate choice of the bandwidths, these estimators have nice properties (consistency, asymptotic normality, etc.), see e.g. Theorems 3.5 and 3.6 in Pagan and Ullah (1999) for local constant case and Theorem 2.7 in Li and Racine (2007) for local linear case. Fan and Gijbels (1996) describe in details the general properties of local polynomial modeling.

2.2 The conditional moments of ε

Importantly, even without specifying a particular choice for the local distributions of U and of V, we can also rather easily estimate the conditional moments of ε. To summarize, under the symmetry assumption on V, we have

$$E(\varepsilon|x, z) = 0, \tag{2.4}$$

$$E(\varepsilon^2|x, z) = var_U(x, z) + var_V(x, z) > 0, \tag{2.5}$$

$$E(\varepsilon^3|x, z) = -E\left[\left(U - \mu_U(x, z)\right)^3|x, z\right]. \tag{2.6}$$

---

[2]Note however, that even in the most general case, some important information about the average production relationship (e.g., marginal average productivity of inputs, average scale elasticity, etc.) is still contained in $r_1(x, z)$ and so can be inferred from $\widehat{r}_1(x, z)$.

We will extend the idea of the Modified OLS (MOLS), originated in the full parametric, homoskedastic stochastic frontier models (see Olson et al., 1980) for our semi-parametric setup. The idea is to exploit the fact that the residuals in (2.1) may help estimating the conditional moments of ε. Note that for most of the distributions of $U|x, z$ on the real positive line have a positive skewness and so $E(\varepsilon^3|x, z) \leq 0$ in most of the cases. However it is not necessarily the case for all distributions; we will come back to this later.

Let us denote for $j = 2$ and 3, $r_j(x, z) = E(\varepsilon^j|x, z)$. From the estimation of $r_1(x, z) = E(Y|x, z)$ above we obtain the residuals. In particular, at any data point we have

$$\widehat{\varepsilon}_i = Y_i - \widehat{r}_1(X_i, Z_i), \ i = 1, ..., n. \tag{2.7}$$

Now by using appropriate nonparametric techniques (e.g., local constant or local linear), the regression functions $r_2(x, z)$ and $r_3(x, z)$ can be consistently estimated from the sets of data points $\{(\widehat{\varepsilon}_i^2, X_i, Z_i)|i = 1, ..., n\}$ and $\{(\widehat{\varepsilon}_i^3, X_i, Z_i)|i = 1, ..., n\}$, respectively. So we define

$$\widehat{r}_2(x, z) = \sum_{i=1}^{n} W_{i,h}(x, z)\, \widehat{\varepsilon}_i^2 = \sum_{i=1}^{n} W_{i,h}(x, z)\big(Y_i - \widehat{r}_1(X_i, Z_i)\big)^2 \tag{2.8}$$

and

$$\widehat{r}_3(x, z) = \sum_{i=1}^{n} W_{i,h}(x, z)\, \widehat{\varepsilon}_i^3 = \sum_{i=1}^{n} W_{i,h}(x, z)\big(Y_i - \widehat{r}_1(X_i, Z_i)\big)^3 \tag{2.9}$$

where $W_{i,h}(x, z)$ depends on the chosen method (local linear or local constant) and on the vector of bandwidths h. For the Nadaraya-Watson estimator (local constant) the expression is as follows[3]

$$W_{i,h}(x, z) = \frac{K\left(\frac{X_i-x}{h_x}, \frac{Z_i-z}{h_z}\right)}{\sum_{j=1}^{n} K\left(\frac{X_j-x}{h_x}, \frac{Z_j-z}{h_z}\right)}, \tag{2.10}$$

where again, with some abuse of notations, $K\left(\frac{X_i-x}{h_x}, \frac{Z_i-z}{h_z}\right)$ represent a product kernel with the p factors for X and the d factors for Z, h_x, h_z being the $p + d$ individual bandwidths for these components.

In Appendix A we give an heuristic proof of the consistency and asymptotic normality of these estimators in the local constant case, under similar regularity properties and appropriate choice of the bandwidths as the ones required for estimating $r_1(x, z)$. To summarize we obtain for the local constant case

$$(nh^{p+d})^{1/2}\big(\widehat{r}_2(x, z) - r_2(x, z)\big) \xrightarrow{\mathcal{L}} N(0, s_2^2(x, z)) \tag{2.11}$$

---

[3]For the local linear case, the expressions for the estimation of the regression along with its gradients is easier to present in matrix notation as soon as $p + d > 1$, see e.g. Li and Racine (2007), p.81.

and

$$(nh^{p+d})^{1/2}\big(\widehat{r}_3(x,z) - r_3(x,z)\big) \xrightarrow{\mathcal{L}} N(0, s_3^2(x,z)) \qquad (2.12)$$

where for the asymptotic developments we assume, as it is often the case, that all the bandwidths h_x, h_z have the same order as h and where the variances are given in the appendix. Similar results could be obtained through the same lines for local linear estimators but at a cost of notational complexity.

Remark 2.1. *If we want to impose that $U|x,z$ is positively skewed then we have $r_3(x,z) \leq 0$. We can impose this restriction on our nonparametric estimator $\widehat{r}_3(x,z)$ without affecting its asymptotic properties.[4] Indeed as $n \to \infty$, the probability of having these constraints active converges to zero. In finite samples however, it may matter as illustrated in our empirical example. So imposing such restriction is not without practical effects. Using local exponential estimators, see Ziegelmann (2002), is an alternative but here we would lose the numerical flexibility of local constant and local linear estimators. See also footnote 7.*

Now from (2.5)–(2.6), we see the link of $r_2(x,z)$ and $r_3(x,z)$ with the conditional moments of U and V, so plugging these estimates in the equations will provide information on the moments of U and V. But for this we must add some information on their local distributions.

We will see that in order to identify the frontier level and some important parameters of the model, e.g., $m(x,z)$, $E(U|x,z)$, $var_U(x,z)$ and $var_V(x,z)$ we will need to select a local parametric model for both the density of $U|x,z$ and of $V|x,z$.

However, if explanation of efficiency levels by the elements of x and/or of z is our main concern, or if we want to test if $\mu_U(x,z)$ is a constant with respect to any of the components of x,z, then we do not need to specify the distribution of $V|x,z$ and we only have to assume that the density of $U|x,z$ belongs to a one-parameter scale family, without specifying which member of the family is chosen. This is interesting since this approach is robust not only to complex (and usually unknown) regression functions but also to different (and usually unknown) distributional assumptions about the inefficiency term and the noise term in the SFA model.

---

[4]If we use local constant estimator, it means censoring at zero (as e.g. in (3.8) below), but for local linear estimators, it would be, in cases where $\widehat{r}_3(x,z) > 0$, a constrained optimization when deriving the fit and the gradients. So we impose only that $\alpha \leq 0$ and we let the β free when minimizing the local least-squares criterion and only for those (x,z) where the constraint on α is violated.

3 Estimation of Frontier and of Individual Inefficiency

Our goal now is to go beyond the average production function estimation and tackle the problem of estimation and inference for the stochastic frontier itself, and about efficiency scores measured with respect to this frontier for particular observations. From our discussion above it must be clear that for implementing this task we need to have information on $\mu_U(x, z) = E(U|x, z)$ and this is where (for the cross-sectional data framework) we need to make local parametric assumptions on the types of distributions of $U|x, z$ and $V|x, z$.

Since the introduction of ALSMB model, the standard way to disentangle the effect of inefficiency and noise in cross-sectional frontier estimation is to make some assumptions about distribution of V and distribution of U. After agreeing to make such assumptions, we face a variety of choice problem: there are many distributions that could be used for the inefficiency and the noise terms. A natural choice will be to select the most popular choices made by ALSMB, where they assumed homoskedastic centered Normal for V and either exponential or half-normal (and homoskedastic) for U. The choice of one parameter density for U is flexible enough and avoids some of the identification problem encountered with two-parameters densities, such as the Gamma (Greene, 1990) or the truncated-normal (Stevenson, 1980), which happen when the two parameters are fully free in variation. The main identification issue is that for some combination of values of these 2 parameters, the density of U is hard to distinguish from a Gaussian, and in this case, the noise cannot be easily identified from the efficiency level (see Ritter and Simar, 1997, for details).

Because of its dominance in empirical research, we chose here to present the results for the case of a convolution of a Normal and Half-normal that was first considered by Olson et al. (1980) in parametric homoskedastic context (similar results can also be obtained by using the convolution of a Normal and an Exponential distribution as well as other suitable convolutions). This is the same paradigm as the one used in Kumbhakar et al. (2007). We assume that

$$V|x, z \sim N(0, \sigma_V^2(x, z)), \tag{3.1}$$

$$U|x, z \sim |N(0, \sigma_U^2(x, z))|, \tag{3.2}$$

where we also assume that, conditionally on (x, z), U and V are independent. The flexibility of the model comes from its nonparametric localization, because we let the variance functions

unspecified (unknown functionals). As a result, we would have

$$\mu_U(x, z) = E(U|x, z) = \sqrt{\frac{2}{\pi}}\sigma_U(x, z) \tag{3.3}$$

$$r_2(x, z) = E(\varepsilon^2|x, z) = \sigma_V^2(x, z) + \left(\frac{\pi - 2}{\pi}\right)\sigma_U^2(x, z) \tag{3.4}$$

$$r_3(x, z) = E(\varepsilon^3|x, z) = \sqrt{\frac{2}{\pi}}\left(1 - \frac{4}{\pi}\right)\sigma_U^3(x, z) \leq 0. \tag{3.5}$$

Rearranging the system of equations given in (3.4)-(3.5), and solving it first for $\sigma_U^3(x, z)$ and $\sigma_V^2(x, z)$ we get

$$\sigma_U^3(x, z) = \sqrt{\frac{\pi}{2}}\left(\frac{\pi}{\pi - 4}\right)r_3(x, z), \tag{3.6}$$

$$\sigma_V^2(x, z) = r_2(x, z) - \left(\frac{\pi - 2}{\pi}\right)\sigma_U^2(x, z). \tag{3.7}$$

The consistent nonparametric estimates $\widehat{r}_2(x, z)$ and $\widehat{r}_3(x, z)$ derived above can be plugged in these equations to provide consistent estimates of the two variance functions at each point of interest (x, z). Specifically, we have

$$\widehat{\sigma}_U(x, z) = \max\left\{0, \left[\sqrt{\frac{\pi}{2}}\left(\frac{\pi}{\pi - 4}\right)\widehat{r}_3(x, z)\right]^{1/3}\right\} \geq 0, \tag{3.8}$$

$$\widehat{\sigma}_V^2(x, z) = \widehat{r}_2(x, z) - \widehat{\sigma}_U^2(x, z)\left(\frac{\pi - 2}{\pi}\right). \tag{3.9}$$

The max operator in (3.8) is needed if we do not impose the non-positivity constraint of $\widehat{r}_3(x, z)$ (see Remark 2.1 above).[5] Using these estimates, we can also obtain the estimates of the efficiency scores for each observation, for example, by using the method of Jondrow et al (1982)—after generalizing it to the heteroskedastic case, by estimating $E(U_i|\varepsilon_i, x_i, z_i)$ instead of $E(U_i|\varepsilon_i)$. However, one must be cautious in interpreting such efficiency scores, as they are "predicted values" that are based on unobserved ε_i replaced with its estimate for one particular realization i. As a result, their confidence intervals are usually fairly wide (see Simar and Wilson, 2010, for more discussion on it).

Another useful information can also be inferred from a consistent estimate of the conditional mean of inefficiency term, conditional on (x, z), given by

$$\widehat{\mu}_U(x, z) = \sqrt{\frac{\pi}{2}}\widehat{\sigma}_U(x, z). \tag{3.10}$$

---

[5]Note that if homoskedasticity of the variance functions is assumed, we can estimate the constant variance functions e.g. by averaging $\widehat{\sigma}_U^2(X_i, Z_i)$ and $\widehat{\sigma}_V^2(X_i, Z_i)$ over all the observations.

Estimates of $\widehat{\mu}_U(x, z)$ at every combination of interest (x, z) can then be used to recover a consistent estimate of the stochastic frontier, $m(x, z)$, via

$$\widehat{m}(x, z) = \widehat{r}_1(x, z) + \widehat{\mu}_U(x, z). \tag{3.11}$$

Note that asymptotic properties of $\widehat{m}(x, z)$, $\widehat{\mu}_U(x, z)$, $\widehat{\sigma}_U^2(x, z)$ and $\widehat{\sigma}_V^2(x, z)$ are obviously inherited from the asymptotic properties of $\widehat{r}_1(x, z)$, $\widehat{r}_2(x, z)$ and $\widehat{r}_3(x, z)$ described above.

4   Robust Analysis of Determinants of Inefficiency

Note that in the traditional SFA setup, when statistical noise is symmetric while inefficiency term is asymmetric, all the information about the inefficiency is essentially contained in the skewness of the composite error. Therefore, studying inefficiency boils down into studying the conditional skewness with respect to (x, z), which can be done via a fully non-parametric regression approach.

Indeed we show now that to perform such analysis, we do not need to specify the distribution of $V|x, z$ and we can use any one-parameter scale family for the density $U|x, z$, without specifying which member of the family is chosen. To be precise, the density of $U|x, z$ belongs to the one parameter scale family if it can be written as

$$f_U(u|x, z) = \frac{1}{\sigma_U(x, z)} g\Big(\frac{u}{\sigma_U(x, z)}\Big), \tag{4.1}$$

where $g(\cdot)$ is any density on $\mathbb{R}_+$ and $\sigma_U(x, z) \in (0, \infty)$ is the local shape parameter. Examples of this are the Exponential, the Half-normal, the Gamma with fixed shape parameter and the Weibull with fixed shape parameter.[6] In this family it is easy to show that for all $j \geq 1$

$$E(U^j|x, z) = \sigma_U^j(x, z) k_j, \tag{4.2}$$

as long as the jth moment of g, $k_j = \int_0^\infty t^j\, g(t)\, dt$, exists. As a result, we also have

$$E(\varepsilon^3|x, z) = -E\left[\big(U - \mu_U(x, z)\big)^3 \big|\, x, z\right] = -c\sigma_U^3(x, z). \tag{4.3}$$

where $c = (k_3 - 3k_2 k_1 + 2k_1)$ is a constant that can be computed if g is known.[7]

---

[6]Note that for the Gamma and the Weibull, the shape parameter does not need to be known, but it has to be a constant w.r.t. (x, z).

[7]Note that for the Exponential, the Half-normal and the Gamma cases, $c > 0$ and so $U|x, z$ is positively skewed. However for the Weibull, as pointed by Carree (2002), the c becomes negative when the shape parameter is bigger than 3.60235. Since we restrict our one-parameter scale family to a fixed (unknown) shaped parameter, this fix the skewness of $U|x, z$.

Often, practitioners are actually more interested in the determinants of the inefficiency rather than the inefficiency or the frontier per se. Sometimes, researchers are even satisfied with at least the direction (sign) of the influence, although perhaps ideal information would be about the elasticities of the inefficiency w.r.t. certain variables, since they do not depend on units of measurement of the variables involved. To facilitate our further discussion, let ψ_ℓ denote an element of (x, z),

4.1 Estimation of the elasticity of $\mu_U(x, z)$ w.r.t. ψ_ℓ

The (partial) elasticity measure of $\mu_U(x, z)$ w.r.t. ψ_ℓ, which we denote by $\xi_{\psi_\ell}(x, z)$, is given by

$$\xi_{\psi_\ell}(x, z) = \frac{\partial \mu_U(x, z)}{\partial \psi_\ell} \frac{\psi_\ell}{\mu_U(x, z)} \tag{4.4}$$

assuming that $\mu_U(x, z) \neq 0$. Using (4.2) with $j = 1$ we immediately get

$$\xi_{\psi_\ell}(x, z) = \frac{\partial \sigma_U(x, z)}{\partial \psi_\ell} \frac{\psi_\ell}{\sigma_U(x, z)} \tag{4.5}$$

provided that $\sigma_U(x, z) \neq 0$ (i.e., that there is some inefficiency at (x, z)). Although $\partial \sigma_U(x, z)/\partial \psi_\ell$ is not directly estimable, an advantage of the one parameter scale family is that we can still recover it from estimate of $E(\varepsilon^3|x, z)$. Indeed, using (4.2) when $j = 3$ we get

$$\frac{\partial E(\varepsilon^3|x, z)}{\partial \psi_\ell} \frac{\psi_\ell}{E(\varepsilon^3|x, z)} = 3c\,\sigma_U^2(x, z) \frac{\partial \sigma_U(x, z)}{\partial \psi_\ell} \frac{\psi_\ell}{c\,\sigma_U^3(x, z)} \tag{4.6}$$

$$= 3\,\xi_{\psi_\ell}(x, z)$$

Therefore,

$$\xi_{\psi_\ell}(x, z) = \frac{1}{3} \frac{\partial E(\varepsilon^3|x, z)}{\partial \psi_\ell} \frac{\psi_\ell}{E(\varepsilon^3|x, z)} \tag{4.7}$$

and so, a consistent non-parametric estimate of $\xi_{\psi_\ell}(x, z)$ can be obtained by replacing the true moment $E(\varepsilon^3|x, z)$ and its partial derivative $\partial E(\varepsilon^3|x, z)/\partial \psi_\ell$ with their non-parametric estimates. So we have

$$\widehat{\xi}_{\psi_\ell}(x, z) = \frac{1}{3} \frac{\widehat{\partial r_3}(x, z)}{\partial \psi_\ell} \frac{\psi_\ell}{\widehat{r}_3(x, z)} \tag{4.8}$$

where $\widehat{r}_3(x, z)$ and $\widehat{\partial r_3}(x, z)/\partial \psi_\ell$, $\ell = 1, ..., p + d$ are the estimates provided in Section 2. Of course, we need that $\widehat{r}_3(x, z) \neq 0$ for the particular combination of interest (x, z).[8]

---

[8]Note that in practice it may indeed happen that $\widehat{r}_3(x, z) = 0$ (no inefficiency) at some combination of (x, z), causing the elasticity at that combination to be undefined and possibly an error message in a computing software.

11

In Appendix A, we provide a proof of consistency and asymptotic normality of the estimator $\widehat{\xi}_{\psi_\ell}(x, z)$ when using the local constant estimator (Nadaraya-Watson). To be specific, under similar regularity conditions as in Section 2 and under appropriate choice of the size of the bandwidths, we have

$$(nh^{p+d+2})^{1/2}\big(\widehat{\xi}_{\psi_\ell}(x, z) - \xi_{\psi_\ell}(x, z)\big) \xrightarrow{\mathcal{L}} N(0, s^2_{\xi_\ell}(x, z)), \tag{4.9}$$

where the expression of the variance and the required assumptions are described in the appendix. We remark that we have lost a power 2 in the rate of convergence, to get rid of the bias term.

It is worth noting here that the estimator in (4.8) depends on the estimate of $r_3(x, z)$ and on its first partial derivative w.r.t. the variable of interest only. These were obtained in Section 2 where no parametric assumptions were made on the distributions of $U|x, z$ and $V|x, z$ (except for symmetry and zero mean for the latter). We observe also that the expression of the estimate is generally valid under the assumption that $U|x, z$ belongs to the one-parameter scale family defined in (4.1), but we do not need to know to which member it belongs for estimating the elasticities. This provides a formidable robustness of the resulting estimation (4.8).

In general, estimation results may reveal that inefficiency is pertinent for some particular ranges of (x, z) and insignificant for others. To verify such questions, one can estimate asymptotic confidence intervals, using the knowledge about the asymptotic distribution of the estimator for the regression function and its partial derivatives and using bootstrap methods for evaluating the variances.

4.2 Testing significance of ψ_ℓ's on $\mu_U(x, z)$

If we want to test if some elements of x, z are independent of $E(U|x, z)$, we can, as suggested in Racine (1997), test if the elements of the vector of partial derivatives of $\mu_U(x, z)$ are equal to zero. We see by (4.2) that in the local one-parameter scale family this happens if and only if the corresponding partial derivatives of $r_3(x, z)$ are equal to zero. Suppose, without loss of generality that we want to test that the first $q \geq 1$ elements ψ_ℓ, $\ell = 1, \ldots, q$ are significant. Racine (1997) suggests the use of the test statistics

$$T = \frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=1}^{q} \left[\frac{\widehat{\partial r_3}(X_i, Z_i)}{\partial \psi_\ell}\right]^2. \tag{4.10}$$

We reject the null hypothesis when T is too large. The asymptotic distribution of $\widehat{\partial r_3}(x, z)/\partial \psi_\ell$ is normal and is given in the Appendix A (see equation (A.8)) and the test is consistent,

but the sampling distribution of T under the null is complicated so the bootstrap is used to approximate the p-value. In our case of heteroskedasticity, the wild bootstrap would rather be used. The algorithm is given in details in Racine (1997), where an option to build a pivotal test statistics by using a double bootstrap is also described.

5   Empirical Example

The goal of this section is to illustrate and compare several approaches for a real data set. Here, we use data on GDP (PPP adjusted), labor and capital of 57 countries that were also used in several recent studies about patterns of economic growth in the world, most prominently in a study of Kumar and Russell (2002), where more details of the data can be found.[9] Recall that in their seminal work, Kumar and Russell (2002) used an alternative to ours estimator, referred to as Data Envelopment Analysis (DEA), to estimate the world frontier and the efficiency levels of individual countries. To its advantage, the DEA is also a non-parametric estimator, however its key weakness is the assumption of no (or ignorance of) possible noise in the data—and this is the drawback addressed by our approach, while still retaining the advantage of being non-parametric. Another weakness of the standard DEA that we avoid here is the fact that the estimation of the frontier is not taking into account possible dependency of the inefficiency on some variables that also may (or may not) influence the frontier. For relating inefficiency to explanatory variables in DEA framework, one can use approach recently proposed by Simar and Wilson (2007), where efficiency score, estimated via DEA at the first stage, are then regressed on explanatory variables at the second stage. The idea is somewhat similar to our approach, yet it is much more restrictive than ours. Indeed, in addition to allowing for statistical noise in the data, a very important advantage of our approach is that it does not need an assumption of "separability", which is crucial in the approach of Simar and Wilson (2007). It is also worth noting that the work of Simar and Wilson (2007) originally employed parametric regression at the second stage, while the first stage was non-parametric. This was relaxed in Park et al. (2008), who developed theory for local likelihood with truncated regression and illustrated it with analysis of relationship between DEA estimated efficiency scores of countries, as the dependent variable, and their capital per worker levels, using the same data from Kumar and Russell (2002). This gives us an opportunity to make some comparisons of qualitative conclusions from our approach

---

[9]Also see Henderson and Russell (2005), Simar and Zelenyuk (2006), Henderson and Zelenyuk (2007), Badunenko et al. (2008), Park et al. (2008), etc.

and from the two-stage efficiency analysis in Park et al. (2008).

Formally, we will be estimating the following regression relationship,

$$Y_i = m(X_i) + V_i - U_i, \ i = 1, \ldots, n$$

where Y_i is the natural log of GDP per worker of country i in 1990, X_i is the natural log of capital per worker of country i in 1990, V_i is the statistical noise term, for which we assume that $E(V_i|X_i) = 0$ and $Var(V_i|X_i) \in (0, \infty)$ for all i, while $U_i \geq 0$ is the inefficiency term. We also assume that $Var(U_i|X_i) \in (0, \infty)$ for all i and that U_i is independent from V_i, conditionally on X_i. Note also that while in our production model we originally assume the output (GDP) is produced by two inputs (labor and capital), by dividing all variables through by one of inputs (labor), we impose assumption of constant returns to scale (CRS) on the technology. In general, we do not need a CRS assumption for our approach to work and we do so here only for convenience. Indeed, besides ability to illustrate our results with simple two-dimensional figures, with CRS assumption our results are then more comparable to other works in the area, in particular to Kumar and Russell (2002) and Park et al. (2008) who imposed the same assumption.[10]

The result of the estimations are shown in four panels of Figures 1. The top left panel depicts the data points and the estimation results for the average production function and the frontier from several estimation methods. In this panel, the thin dashed line depicts results of standard parametric (and homoskedastic) MOLS (a la OSW), which happened to coincide here with estimates of parametric average production function—because the parametric estimate of $E(U_i)$ happened to be zero for this data set. In other words, the standard parametric MOLS approach suggests that there is no inefficiency in the sample, i.e., suggesting that *all* countries are efficient and all the deviations are to be attributed to statistical noise. This is, of course, completely opposite from conclusions of Kumar and Russell (2002), where all the deviations from frontier (estimated via DEA) were attributed to inefficiency, which, in turn were then found to play an important role in shaping the dynamics of distributions of income per worker in the world. For the sake of reference, we also depict the DEA estimate of the technology frontier with the solid piece-wise linear line in top left panel of Figure 1.

The same top-left panel in Figure 1 also depicts the non-parametric estimate of the average production function (with thick dashed line) and semi-parametric estimates of frontier

---

[10]Clearly, in a detailed analysis, one may also want to condition for many other potentially important explanatory variables, yet we will limit our illustration to the case of one continuous variable for the sake convenience in graphical representation and comparisons.
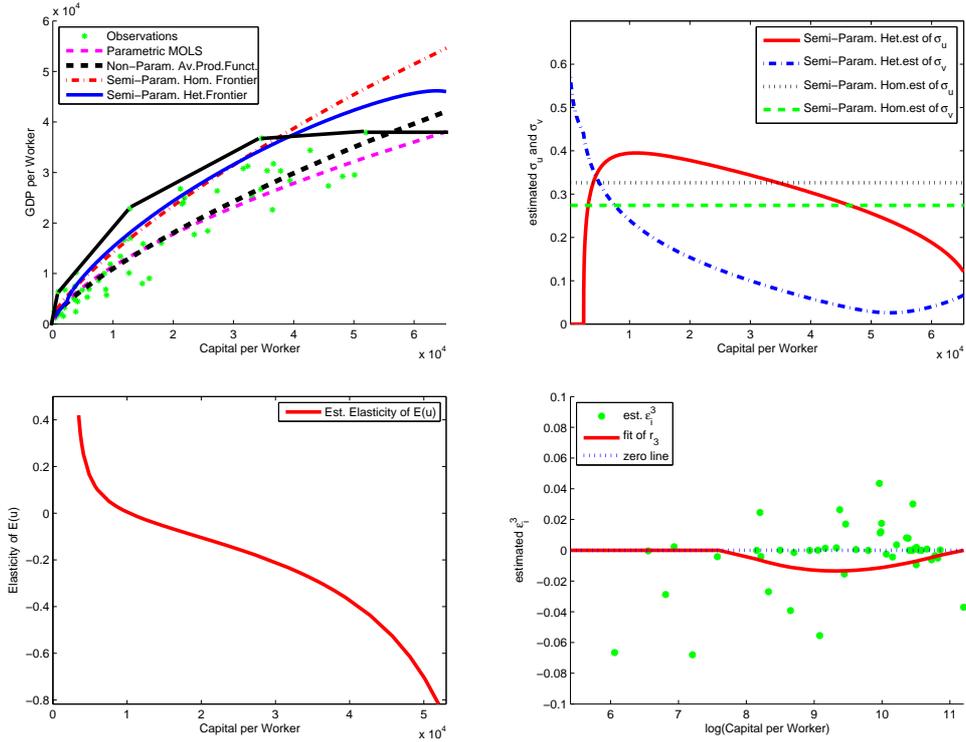
Figure 1: *Estimation results imposing non-positivity constraint for $\widehat{r}_3(x)$. On the top left panel, the solid piece-wise linear line is the traditional DEA estimate. Note that the bottom right panel as horizontal scale in log$s$.*

under homoskedasticity assumption (dash-dot line) and the one allowing for heteroskedasticity (smooth solid line). Observing these curves, one can see that the shape of the average production function estimated non-parametrically is similar, yet mostly above the parametric MOLS (which coincides here with the parametric estimate of average production function). More importantly, our semi-parametric estimates suggest there is substantial inefficiency across countries, which is very different conclusion than that from the parametric MOLS. Indeed, even the semi-parametric approach with assumption of *homoskedastic* variances of inefficiency (i.e., assuming $E(U_i|X_i) = E(U_i)$ is a constant) and of the noise suggested the presence of substantial inefficiency. Also note that our semi-parametric approach that allows for heteroskedastic variances of inefficiency and of the noise also suggested substantial inefficiency is present, although yielded usually smaller levels of inefficiency than those suggested by the homoskedastic semi-parametric approach.[11]   Thus, despite the similarity of para-

[11]We obtain our non- and semi- parametric estimates using all observations at the grid of points in the observed range. The plots are given for all the grid points except for the largest observation in capital per

15

metric and non-parametric estimates of the average production function, the frontiers and inefficiencies are here substantially different between the parametric and semi-parametric approaches, even in the fully homoskedastic case.

Interestingly, note that the semi-parametric heteroskedastic approach yields similar results relative to its simplified homoskedastic version for the lower half of the data, but results are quite different for the upper half of the data, where the heteroskedastic version is showing more curvature and less inefficiency. These results are somewhat intuitive: the heteroskedastic semi-parametric version is expected to be more adaptive to the changing variation in the data, and so it is more likely to yield more adequate (or less extreme) results than the homoskedastic semi-parametric approach. Thus, the results from semi-parametric approach, homoskedastic and heteroskedastic, support the claim of Kumar and Russell (2002) about substantial differences in inefficiency across countries, unlike the parametric homoskedastic MOLS that identified no inefficiency. However, unlike in Kumar and Russell (2002), we avoid the drawback of DEA by allowing for both, the inefficiency and the noise, yet also stay free from parametrization of the production function, as well as allow for influence of explanatory variables onto inefficiency and the output levels. Such differences may imply completely different distribution of inefficiency across the countries and thus different conclusions about the impact of efficiency change on patterns of economic growth dynamics than were found in Kumar and Russell (2002).

Let us now look closer at the top-right panel of Figure 1, which presents estimates of σ_U and σ_V from semi-parametric approach under the two assumptions: homoskedastic and heteroskedastic variances for the inefficiency and the noise. From this picture we see that both variants are able to disentangle the total deviation into σ_U and σ_V but they do it quite differently. An obvious advantage of the heteroskedastic variant is the ability to disentangle different natures of variability of σ_U and σ_V at different values of an explanatory variable (capital per worker here). Interestingly, for our sample, the heteroskedastic variant suggests that σ_U is at zero (no inefficiency) at the very low level of capital per worker, then rapidly increases and then gradually decreases for countries with very large capital per worker, indicating less inefficiency.[12] So, we remark that the "wrong skewness" problem appears only for the lowest values of x where the sign-constrained estimator of $\widehat{r}_3(x)$ gave values of $\widehat{\sigma}_U(x) = 0$. In fact this corresponds to a region where lies less than 20% of the data points in

_____

worker (Switzerland), which is substantially outlying from others and so the estimated curve, not surprisingly, passes through that point.

[12]The estimates and the plot of the conditional expectation of the inefficiency term is different from those of σ_U just by the constant $\sqrt{2/\pi} \approx 0.8$, for all points, and so we do not present it for the sake of space.

our sample. This is in accordance with results reported in the literature: with similar sample sizes, we may observe similar proportions of wrong skewness even in a fully parametric model when the data are simulated with the Normal-Half normal convolution (see Coelli, 1995 and Simar and Wilson, 2010 for details). This will be confirmed below, in the top right panel of Figure 2, where the estimation of $\widehat{r}_3(x)$ is unconstrained and so the truncation in (3.8) for defining $\widehat{\sigma}_U$ is active.

The heteroskedastic variant of the estimator also suggests that the dependency of σ_V on x for our sample is likely to be of quite different nature than the one we observed for σ_U w.r.t. x. Indeed, we see that the estimated σ_V is monotonically decreasing with x, with an asymptote about zero for observations with largest capital per worker. In a sense, one would expect that, at least on average, countries that are richer in capital per worker would tend to have lower uncertainty risk (captured here by σ_V) as well as lower inefficiency risk (captured here by σ_U) and that is what we see in our estimation results. Interestingly, note that Park et al. (2008) obtained somewhat similar conclusions, although using different estimation paradigm: they found that influence of capital per worker onto inefficiency of countries was pronounced not through the conditional mean but through the conditional variance.

One of the main novelties of our work is a fully non-parametric estimation of the impact of explanatory variables onto the conditional moment of inefficiency term—via the elasticity— and these estimates for our data set are depicted in bottom-left panel of Figure 1. Specifically, one can see that the estimated elasticity starts at about 0.4 for the countries with lowest capital per worker then monotonically decreases to negative territory, and all the way to about -0.8 for countries with high capital per worker levels. The decrease is rapid at the beginning, slows down in the mid range and accelerates to be rapid again at the right end of the range. Note that negative values imply negative relationship between the (estimated) conditional mean of inefficiency term, $E(U|\cdot)$, and the explanatory variable experiencing the change (capital per worker here). Thus, our estimates show that the larger the capital per worker in a country, the larger is the negative impact of its percentage change on the percentage change in the inefficiency term (on average, ceteris paribus). While the figure illustrates the entire sample, one can also obtain and analyze point estimates for particular countries: e.g., the elasticity estimate for USA in our case is -0.276, which implies that an increase in capital per worker by 1% for USA is expected to lead to about 0.28% decrease in the inefficiency term, on average and ceteris paribus, according to 1990 data.[13] Similar

---

[13]Also recall that $-E(U|\cdot)$ is the first order approximation (around zero) of $1 - E(\exp(-U)|\cdot)$, which is the percentage measure of inefficiency.

estimates of elasticities are obtained for most other developed countries, and for eleven countries they turned out to be undefined (due to $E(U|\cdot) = 0$) and so not presented on the figure. Recall that to estimate the elasticity we do not need to make parametric assumptions about the distribution of U or V, and only need to non-parametrically estimate r_3 and assume the quite general local one-parameter scale family, without specifying which distribution it is. Bottom-right panel of Figure 1 presents these estimates for our data set, for the case of constrained estimation—with constraint that r_3 is non-positive, as the theory suggests for most one-parameter scale distributions (see footnote 7).[14]

With the help of four panels of Figure 2 we now present estimation results for the same data and the same methods as those described in Figures 1, except that the non-parametric estimation of r_3 is done without non-positivity restriction, with standard local linear least squares method. Comparing results from Figure 1 with those in Figure 2, one can see substantial differences. First of all, comparison of bottom-right panels of the two figures reveals different estimates of r_3. Although the positive estimates of r_3 depicted in the bottom-right panel of Figure 2 are then censored to zero, ex post, when computing estimates of σ_U, the resulting estimates are substantially different when comparing to σ_U obtained using constrained local least squares to estimate r_3. These differences are then reflected in the estimates of the frontier and, most sensitively, in the estimation of elasticity of $E(U_i|\cdot)$ w.r.t. X_i, as the differences in left panels of Figure 1 and of Figure 2 suggest. This illustrates that constrained estimation of r_3 may produce, especially in small samples ($n = 57$), substantially different results than the non-constrained estimation and so should be preferred if justified by theoretical reasoning, as in our case.

Concluding our empirical illustration, we would like to note that since our approach addresses the drawbacks of DEA (allows for statistical noise in the data, etc.) while preserving its advantages (no or minimal parametric assumptions), many results and conclusions of the seminal work of Kumar and Russell (2002), as well as many other papers inspired by their approach, might need to be revisited to understand if qualitative conclusions change due to using more flexible method as ours. In doing so, one shall of course consider other variables and aspects that may potentially influence the technology frontier as well as the variances of the inefficiency and of the statistical noise terms and we leave this for more elaborate empirical studies to focus on in the future.

---

[14]Note that due to the nature of the dependent variable, the observations on this picture are very dispersed vertically, with many being very close to or at zero and our figure is zoomed such that the estimated curve is visibly different from zero, which in turn made many observations being beyond the scope of the figure.
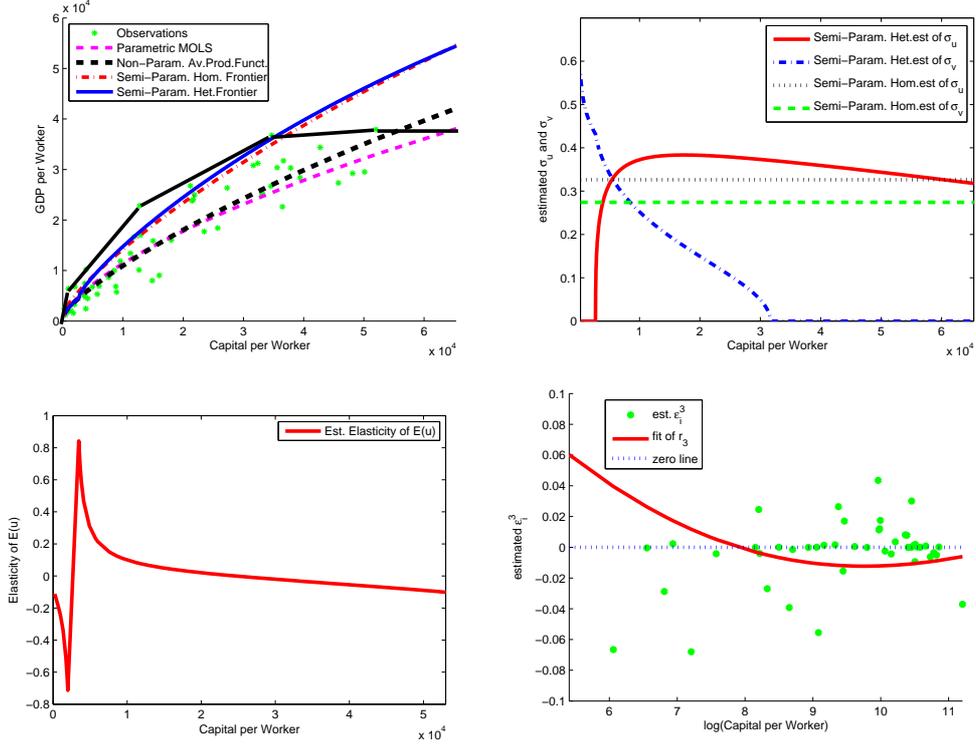
Figure 2: *Estimation results with unconstraint sign for $\widehat{r}_3(x)$. On the top left panel, the solid piece-wise linear line is the traditional DEA estimate. Note that the bottom right panel as horizontal scale in logs.*

6 Conclusions

In this work we aimed to improve upon the currently popular methods for estimating technology frontier and (in)efficiency levels. We proposed the semi-parametric generalization of the so-called "modified OLS" approach as an alternative to the parametric and the local MLE approaches. We show that most of the benefits of the local MLE approach can be obtained in our approach with less assumptions and involving much easier, faster and numerically more robust computations, by using nonparametric (or local polynomial) least-squares methods. We illustrated our method and compared it with other methods with a real data set. We also discuss the vital issue of explaining variation in inefficiency levels with respect to explanatory variables with minimal assumptions on the Data Generating Process.

In summary, with the approach we propose in this paper, we are able to combine the advantages of the two competing approaches—the freedom from parametric assumptions about the technology frontier, similar to DEA, and the ability of decomposing the total deviation from the frontier into noise and inefficiency, as in the standard SFA, but allowing

for heteroskedasticity of the inefficiency and of the statistical noise. We still have to use semi-parametric assumptions about the densities of the noise and inefficiency, to be able to do the decomposition of variance of inefficiency from variance of statistical noise, yet in our semi-parametric approach with heteroskedastic inefficiency, this assumption is, in a sense, localized, allowing inefficiency to be changing with (x, z), similarly to local MLE approach of Kumbhakar et al (2007), but much easier and much faster in computation. In addition, the analysis of the determinants of the inefficiency and a test for the significancy of covariates has been proposed in a model where neither the distribution of the noise nor the ditribution of the inefficiencies have to be fully specified.

The estimation results we obtained for our real data set illustration are quite interesting and intuitive, but also reveal information that is not obvious without proper estimation, as well as different from estimation results of more restrictive standard parametric SFA and somewhat different from the DEA approach.

A   Appendix: Technical Details

In this appendix we establish the asymptotic properties of $\widehat{r}_2(x,z)$, $\widehat{r}_3(x,z)$ and $\widehat{\xi}_{\psi_\ell}(x,z)$ described in Sections 2 and 4 when using a local constant estimator for $\widehat{r}_2(x,z)$ and $\widehat{r}_3(x,z)$. The estimators defined in (2.8) and (2.9) are weighted averages with weights given by (2.10) where the kernel is a product kernel $K(s,t) = k(s_1)\ldots k(s_p)k(t_1)\ldots k(t_d)$ with $k(\cdot)$ being a univariate second order kernel with $||k||_2^2 = \int k^2(\nu)d\nu < \infty$ and (h_x, h_z) is a $(p+d)$-dimensional bandwidth corresponding to each component of (X,Z). For the asymptotic properties we assume that all the bandwidths have the same order, so we simplify the notation by denoting each bandwidth by h (in practice, for $\ell = 1,\ldots,p+d$, $h_\ell = b_\ell h$ for some constant b_ℓ). So the weights will be written as

$$W_{i,h}(x,z) = \frac{K\left(\frac{X_i-x}{h}, \frac{Z_i-z}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{X_j-x}{h}, \frac{Z_j-z}{h}\right)}.$$

Throughout the appendix we also assume that $f_{XZ}(x,z) > 0$, where $f_{XZ}(x,z)$ is the joint density of (X,Z). Some regularity conditions (smoothness of the functions r_2 and r_3) will appear in the proofs below. We will also assume that the conditional moments of ε of order up to 7 are finite, i.e. $r_j(x,z) = E(\varepsilon^j|x,z) < \infty$ for $j = 2,\ldots,7$. For the asymptotic normality of $\widehat{r}_2(x,z)$ and $\widehat{r}_3(x,z)$ we need to work with a bandwidth h that satisfies $nh^{p+d}(\log n)^{-4} \to \infty$ and $nh^{p+d+4} \to 0$, whereas for $\widehat{\xi}_{\psi_\ell}(x,z)$ we require that $nh^{p+d+2} \to \infty$ and $nh^{p+d+6} \to 0$.

A.1   Asymptotic normality of $\widehat{r}_2(x,z)$

We have

$$\widehat{r}_2(x,z) = \sum_{i=1}^{n} W_{i,h}(x,z)(Y_i - \widehat{r}_1(X_i, Z_i))^2$$

$$= \sum_{i=1}^{n} W_{i,h}(x,z)\Big[(Y_i - \widehat{r}_1(x,z))^2$$

$$+ 2(Y_i - \widehat{r}_1(x,z))(\widehat{r}_1(x,z) - \widehat{r}_1(X_i, Z_i)) + (\widehat{r}_1(x,z) - \widehat{r}_1(X_i, Z_i))^2\Big]$$

$$= A + B + C. \tag{A.1}$$

Let us first consider term C above. First note that by Jensen's inequality,

$$(\widehat{r}_1(x,z) - \widehat{r}_1(X_i, Z_i))^2$$

$$\leq 3\left[(\widehat{r}_1(x,z) - r_1(x,z))^2 + (\widehat{r}_1(X_i, Z_i) - r_1(X_i, Z_i))^2 + (r_1(X_i, Z_i) - r_1(x,z))^2\right]$$

and we know (e.g., from Masry, 1996) that $\sup_{x,z} |\widehat{r}_1(x,z) - r_1(x,z)| = O_p\left((nh^{p+d})^{-1/2}\log n + h^2\right)$ and that $\sum_{i=1}^{n} W_{i,h}(x,z)(r_1(x,z) - r_1(X_i,Z_i))^2 = O(h^2)$ for second order kernels, provided that all partial derivatives of $r_1(x,z)$ w.r.t. x and z are finite. Hence,

$$C = O_p\left((nh^{p+d})^{-1}(\log n)^2 + h^2\right) = o_p\left((nh^{p+d})^{-1/2}\right)$$

provided that $(nh^{p+d})(\log n)^{-4} \to \infty$ and $nh^{p+d+4} \to 0$.

Next, we consider the second term B in (A.1). We clearly have

$$B \leq 2\sum_{i=1}^{n} W_{i,h}(x,z)|Y_i - \widehat{r}_1(x,z)| \times |\widehat{r}_1(x,z) - \widehat{r}_1(X_i,Z_i) - r_1(x,z) + r_1(X_i,Z_i)|$$

$$+ 2\sum_{i=1}^{n} W_{i,h}(x,z)|Y_i - \widehat{r}_1(x,z)| \times |r_1(x,z) - r_1(X_i,Z_i)|$$

$$= o_p\left((nh^{p+d})^{-1/2}\right) + O(h^2) = o_p\left((nh^{p+d})^{-1/2}\right)$$

provided that $nh^{p+d+4} \to 0$ and that $E\left(|\varepsilon|\,|x,z\right) < \infty$. This follows from the fact that

$$\sup_{||(x,z)-(x',z')||\leq h} |\widehat{r}_1(x,z) - \widehat{r}_1(x',z') - r_1(x,z) + r_1(x',z')| = o_p\left((nh^{p+d})^{-1/2}\right) + O(h^2).$$

Finally, consider the first term A in (A.1).

$$A = \sum_{i=1}^{n} W_{i,h}(x,z)Y_i^2 - \widehat{r}_1^2(x,z)$$

$$= \sum_{i=1}^{n} W_{i,h}(x,z)Y_i^2 - r_1^2(x,z) + (r_1(x,z) - \widehat{r}_1(x,z))(r_1(x,z) + \widehat{r}_1(x,z))$$

$$= \sum_{i=1}^{n} W_{i,h}(x,z)Y_i^2 - r_1^2(x,z) - 2r_1(x,z)(\widehat{r}_1(x,z)) - r_1(x,z)) + o_p\left((nh^{p+d})^{-1/2}\right)$$

$$= \sum_{i=1}^{n} W_{i,h}(x,z)Y_i^2 + r_1^2(x,z) - 2r_1(x,z)\sum_{i=1}^{n} W_{i,h}(x,z)Y_i + o_p\left((nh^{p+d})^{-1/2}\right)$$

$$= \sum_{i=1}^{n} W_{i,h}(x,z)(Y_i - r_1(x,z))^2 + o_p\left((nh^{p+d})^{-1/2}\right).$$

Hence, we finally have from (A.1) that

$$\widehat{r}_2(x,z) = \sum_{i=1}^{n} W_{i,h}(x,z)(Y_i - r_1(x,z))^2 + o_p\left((nh^{p+d})^{-1/2}\right). \tag{A.2}$$

Then, as is standard for nonparametric kernel estimators (see e.g. Pagan and Ullah, 1999, Theorem 3.5), Liapunov's central limit theorem for triangular arrays can be advocated (see

Serfling, 1980, p. 31), under an additional regularity assumption. Here, denoting $\eta = \varepsilon^2 - r_2(x,z)$, the conditions are $\int |k(\nu)|^{2+\delta} d\nu$ and $E(|\eta|^{2+\delta})$ are finite for some $\delta > 0$. So provided $nh^{p+d+4} \to 0$ we have

$$(nh^{p+d})^{1/2}(\widehat{r}_2(x,z) - r_2(x,z)) \xrightarrow{\mathcal{L}} N(0, s_2^2(x,z)), \tag{A.3}$$

where

$$s_2^2(x,z) = E\left[\eta^2 | x,z\right] ||k||_2^{2(p+d)} f_{XZ}^{-1}(x,z) = \left[r_4(x,z) - r_2^2(x,z)\right] ||k||_2^{2(p+d)} f_{XZ}^{-1}(x,z).$$

A.2   Asymptotic normality of $\widehat{r}_3(x,z)$

By similar arguments as for $\widehat{r}_2(x,z)$ above it is easy to show that

$$\begin{aligned}
\widehat{r}_3(x,z) &= \sum_{i=1}^n W_{i,h}(x,z)(Y_i - \widehat{r}_1(X_i, Z_i))^3 \\
&= \sum_{i=1}^n W_{i,h}(x,z)(Y_i - \widehat{r}_1(x,z))^3 + o_p\left((nh^{p+d})^{-1/2}\right).
\end{aligned}$$

Then we develop the leading term as

$$\sum_{i=1}^n W_{i,h}(x,z) \left[(Y_i - r_1(x,z))^3 + 3(Y_i - r_1(x,z))^2(r_1(x,z) - \widehat{r}_1(x,z))\right.$$

$$\left. + 3(Y_i - r_1(x,z))(r_1(x,z) - \widehat{r}_1(x,z))^2 + (r_1(x,z) - \widehat{r}_1(x,z))^3\right].$$

So we have using similar arguments as above

$$\begin{aligned}
\widehat{r}_3(x,z) &= \sum_{i=1}^n W_{i,h}(x,z)(Y_i - r_1(x,z))^3 \\
&\quad + 3\left[\sum_{i=1}^n W_{i,h}(x,z)(Y_i - r_1(x,z))^2 - r_2(x,z)\right](r_1(x,z) - \widehat{r}_1(x,z)) \\
&\quad - 3\, r_2(x,z) \sum_{i=1}^n W_{i,h}(x,z)(Y_i - r_1(x,z)) + o_p\left((nh^{p+d})^{-1/2}\right).
\end{aligned}$$

So we obtain at the end

$$\widehat{r}_3(x,z) = \sum_{i=1}^n W_{i,h}(x,z) \left[(Y_i - r_1(x,z))^3 - 3r_2(x,z)(Y_i - r_1(x,z))\right] + o_p\left((nh^{p+d})^{-1/2}\right).$$

Again we apply Liapunov's CLT. Here the regularity conditions are the same as for $r_2(x,z)$ with now $\eta = \varepsilon^3 - 3r_2(x,z)\varepsilon - r_3(x,z)$. So provided $nh^{p+d+4} \to 0$ we have

$$(nh^{p+d})^{1/2}(\widehat{r}_3(x,z) - r_3(x,z)) \xrightarrow{\mathcal{L}} N(0, s_3^2(x,z)), \tag{A.4}$$

where

$$s_3^2(x,z) = E\left[\eta^2|x,z\right]||k||_2^{2(p+d)}f_{XZ}^{-1}(x,z)$$
$$= \left[r_6(x,z) - 6r_2(x,z)r_4(x,z) - r_3^2(x,z) + 9r_2^3(x,z)\right]||k||_2^{2(p+d)}f_{XZ}^{-1}(x,z).$$

A.3   Asymptotic normality of $\widehat{\xi}_{\psi_\ell}(x,z)$

The elasticity estimator is defined in (4.8) as $\widehat{\xi}_{\psi_\ell}(x,z) = \dfrac{1}{3}\dfrac{\partial\widehat{r}_3(x,z)}{\partial\psi_\ell}\dfrac{\psi_\ell}{\widehat{r}_3(x,z)}$, where ψ_ℓ is an element in (x,z). We know already that $(nh^{p+d})^{1/2}(\widehat{r}_3(x,z) - r_3(x,z))$ converges to a normal, so $\widehat{r}_3(x,z) - r_3(x,z) = O_p\left((nh^{p+d})^{-1/2}\right)$, whereas $\partial\widehat{r}_3(x,z)/\partial\psi_\ell$ will converge at a slower rate. Therefore, $\widehat{\xi}_{\psi_\ell}(x,z)$ behaves asymptotically as

$$\tilde{\xi}_{\psi_\ell}(x,z) = \frac{1}{3}\frac{\partial\widehat{r}_3(x,z)}{\partial\psi_\ell}\frac{\psi_\ell}{r_3(x,z)}.$$

Now, we have the following relations:

$$\frac{\partial}{\partial\psi_\ell}\widehat{r}_3(x,z) = \frac{\partial}{\partial\psi_\ell}\sum_{i=1}^n W_{i,h}(x,z)(Y_i - \widehat{r}_1(X_i,Z_i))^3 = \sum_{i=1}^n\left(\frac{\partial}{\partial\psi_\ell}W_{i,h}(x,z)\right)(Y_i - \widehat{r}_1(X_i,Z_i))^3$$

$$= \sum_{i=1}^n\left(\frac{\partial}{\partial\psi_\ell}W_{i,h}(x,z)\right)\Big[(Y_i - r_1(X_i,Z_i))^3 + 3(Y_i - r_1(X_i,Z_i))^2(r_1(X_i,Z_i) - \widehat{r}_1(X_i,Z_i))$$

$$+ 3(Y_i - r_1(X_i,Z_i))(r_1(X_i,Z_i) - \widehat{r}_1(X_i,Z_i))^2 + (r_1(X_i,Z_i) - \widehat{r}_1(X_i,Z_i))^3\Big]$$

$$= T_1 + T_2 + T_3 + T_4.$$

We have seen above that $\sup_{x,z}|\widehat{r}_1(x,z) - r_1(x,z)| = O_p\left((nh^{p+d})^{-1/2}\log n + h^2\right)$, hence $T_2 + T_3 + T_4 = O_p\left((nh^{p+d})^{-1/2}\log n + h^2\right) = o_p\left((nh^{p+d+2})^{-1/2}\right)$, provided $\frac{\partial}{\partial\psi_\ell}r_2(x,z) < \infty$ and $nh^{p+d+6} \to 0$. So we have

$$\frac{\partial}{\partial\psi_\ell}\widehat{r}_3(x,z) = T_1 + o_p\left((nh^{p+d+2})^{-1/2}\right). \tag{A.5}$$

Let us consider $T_1 - \frac{\partial}{\partial\psi_\ell}r_3(x,z)$. We obtain

$$T_1 - \frac{\partial}{\partial\psi_\ell}r_3(x,z) = \frac{\partial}{\partial\psi_\ell}\sum_{i=1}^n W_{i,h}(x,z)\left[(Y_i - r_1(X_i,Z_i))^3 - r_3(x,z)\right]$$

$$= \frac{\partial}{\partial\psi_\ell}\sum_{i=1}^n W_{i,h}(x,z)\left[(Y_i - r_1(X_i,Z_i))^3 - r_3(X_i,Z_i)\right]$$

$$+ \sum_{i=1}^n\left(\frac{\partial}{\partial\psi_\ell}W_{i,h}(x,z)\right)r_3(X_i,Z_i) - \frac{\partial}{\partial\psi_\ell}r_3(x,z)$$

$$= T_{11} + T_{12}.$$

24

Let $K_h(s, t) = h^{-(p+d)} K(s/h, t/h)$. Then, by simple calculations and appropriate change in variables in the integrals we get successively:

$$
\begin{aligned}
T_{12} &= \int \frac{\partial}{\partial \psi_\ell} \frac{K_h(s - x, t - z)}{\widehat{f}_{XZ}(x, z)} r_3(s, t) d\widehat{F}_{XZ}(s, t) - \frac{\partial}{\partial \psi_\ell} r_3(x, z) \\
&= h^{-(p+d)} \frac{\partial}{\partial \psi_\ell} \int \frac{K(u, w)}{\widehat{f}_{XZ}(x, z)} r_3(x + hu, z + hw) d\widehat{F}_{XZ}(x + hu, z + hw) - \frac{\partial}{\partial \psi_\ell} r_3(x, z) \\
&= \frac{\partial}{\partial \psi_\ell} \frac{1}{\widehat{f}_{XZ}(x, z)} \int K(u, w) r_3(x + hu, z + hw) \widehat{f}_{XZ}(x + hu, z + hw) du dw - \frac{\partial}{\partial \psi_\ell} r_3(x, z) \\
&= \frac{\partial}{\partial \psi_\ell} r_3(x, z) + O_p(h^2) - \frac{\partial}{\partial \psi_\ell} r_3(x, z) = o_p \left((nh^{p+d+2})^{-1/2} \right), \quad \text{(A.6)}
\end{aligned}
$$

where the last equality holds provided that $nh^{p+d+6} \to 0$.

Finally we consider T_{11}. For this we define $\gamma_i = (Y_i - r_1(X_i, Z_i))^3 - r_3(X_i, Z_i)$. Then,

$$
\begin{aligned}
T_{11} &= \frac{\partial}{\partial \psi_\ell} \left\{ n^{-1} \sum_{i=1}^n \frac{K_h(X_i - x, Z_i - z)}{\widehat{f}_{XZ}(x, z)} \gamma_i \right\} \\
&= n^{-1} \sum_{i=1}^n \widehat{f}_{XZ}^{-2}(x, z) \left\{ \frac{\partial}{\partial \psi_\ell} K_h(X_i - x, Z_i - z) \widehat{f}_{XZ}(x, z) - K_h(X_i - x, Z_i - z) \frac{\partial}{\partial \psi_\ell} \widehat{f}_{XZ}(x, z) \right\} \gamma_i \\
&= n^{-1} \sum_{i=1}^n \frac{\frac{\partial}{\partial \psi_\ell} K_h(X_i - x, Z_i - z)}{\widehat{f}_{XZ}(x, z)} \gamma_i - n^{-1} \sum_{i=1}^n \frac{K_h(X_i - x, Z_i - z)}{\widehat{f}_{XZ}^2(x, z)} \gamma_i \times \frac{\partial}{\partial \psi_\ell} \widehat{f}_{XZ}(x, z) \\
&= n^{-1} \sum_{i=1}^n \frac{\frac{\partial}{\partial \psi_\ell} K_h(X_i - x, Z_i - z)}{f_{XZ}(x, z)} \gamma_i + o_p \left((nh^{p+d+2})^{-1/2} \right), \quad \text{(A.7)}
\end{aligned}
$$

where we used the fact that $n^{-1} \sum_{i=1}^n \frac{K_h(X_i - x, Z_i - z)}{\widehat{f}_{XZ}^2(x, z)} \gamma_i = O_p \left((nh^{p+d})^{-1/2} \right)$ and $\frac{\partial}{\partial \psi_\ell} \widehat{f}_{XZ}(x, z) = O_p(1)$ provided we assume that $\frac{\partial}{\partial \psi_\ell} f_{XZ}(x, z) < \infty$. With our assumption that $f_{XZ}(x, z) > 0$, the leading term in (A.7) is a sample mean of iid random variables and so shall converge to a normal. Our next goal is to find the variance of this leading term, which we achieve below:

$$
\mathcal{V} = Var \left(n^{-1} \sum_{i=1}^n \frac{\frac{\partial}{\partial \psi_\ell} K_h(X_i - x, Z_i - z)}{f_{XZ}(x, z)} \gamma_i \right) = \frac{n^{-1}}{f_{XZ}^2(x, z)} Var \left(\frac{\partial}{\partial \psi_\ell} K_h(X - x, Z - z) \gamma \right)
$$

Now since $E(\gamma | X, Z) = 0$ this leads to

$$
\begin{aligned}
\mathcal{V} &= \frac{n^{-1}}{f_{XZ}^2(x, z)} E \left[\left(\frac{\partial}{\partial \psi_\ell} K_h(X - x, Z - z) \right)^2 Var(\varepsilon^3 | X, Z) \right] \\
&= \frac{n^{-1} Var(\varepsilon^3 | x, z)}{f_{XZ}^2(x, z)} \int \left(\frac{\partial}{\partial \psi_\ell} K_h(s - x, t - z) \right)^2 f_{XZ}(s, t) ds \, dt + o(1) \\
&= \frac{n^{-1} Var(\varepsilon^3 | x, z)}{f_{XZ}(x, z)} h^{-(p+d+2)} ||k||_2^{2(p+d-1)} ||k'||_2^2 + o(1),
\end{aligned}
$$

25

where $Var(\varepsilon^3|x,z) = r_6(x,z) - r_3^2(x,z)$.

Hence, from (A.5), (A.6) and (A.7) and provided $nh^{p+d+2} \to \infty$ and $nh^{p+d+6} \to 0$, we have for the derivatives of r_3 the following result

$$(nh^{p+d+2})^{1/2} \left(\frac{\partial}{\partial \psi_\ell} \widehat{r}_3(x,z) - \frac{\partial}{\partial \psi_\ell} r_3(x,z) \right) \xrightarrow{\mathcal{L}} N(0, s_4^2(x,z)), \tag{A.8}$$

where

$$s_4^2(x,z) = \frac{r_6(x,z) - r_3^2(x,z)}{f_{XZ}(x,z)} ||k||_2^{2(p+d-1)} ||k'||_2^2.$$

Finally, we obtain the asymptotic normality for the elasticities:

$$(nh^{p+d+2})^{1/2} \left(\widehat{\xi}_{\psi_\ell}(x,z) - \xi_{\psi_\ell}(x,z) \right) \xrightarrow{\mathcal{L}} N(0, s_\xi^2(x,z)), \tag{A.9}$$

where

$$s_\xi^2(x,z) = \frac{\psi_\ell^2}{9 f_{XZ}(x,z)} \left(\frac{r_6(x,z)}{r_3^2(x,z)} - 1 \right) ||k||_2^{2(p+d-1)} ||k'||_2^2.$$

References

[1] Aigner, D. J., C. A. K Lovell and P. Schmidt (1977), Formulation and Estimation of Stochastic Frontier Models, *Journal of Econometrics* 6, 21-37.

[2] Badunenko, O., D. Henderson and V. Zelenyuk (2008), Technological change and transition: relative contributions to worldwide growth during the 1990s, *Oxford Bulletin of Economics and Statistics*, 70, 461–492.

[3] Carree, M.A. (2002), Technological inefficiency and the skewness of the error component in stochastic frontier analysis, *Economics Letters*, 77, 101–107.

[4] Coelli, T. (1995), Estimators and hypothesis tests for a stochastic frontier function: a Monte Carlo analysis, *Journal of Productivity Analysis*, 6, 247–268.

[5] Fan, J. and I. Gijbels (1996), *Local Polynomial Modeling and Its Applications*, Chapman and Hall, London.

[6] Fan, Y., Li, Q. and A. Weersink (1996), Semiparametric Estimation of Stochastic Production Frontier, *Journal of Business and Economics Statistics* 14, 460-468.

[7] Greene, W. H. (1990), A Gamma-Distributed Stochastic Frontier Model, *Journal of Econometrics* 46, 141-163.

[8] Henderson, D.J. and R.R. Russell (2005), Human capital and convergence: a production-frontier approach, *International Economic Review*, Vol. 46, 1167–1205.

[9] Henderson, D.J. and Zelenyuk, V. (2007), Testing for (Efficiency) Catching-up, *Southern Economic Journal* 73(4), 1003-1019.

[10] Jondrow, J., Lovell, C.A.K., Materov, I.S., and Schmidt, P. (1982), On the estimation of technical inefficiency in stochastic frontier production models, *Journal of Econometrics* 19, 233-238.

[11] Kneip, A. and L. Simar (1996), A General framework for Frontier Estimation with Panel data, *Journal of Productivity Analysis*, 7, 187–212.

[12] Kumar, S. and Russell, R.R., (2002), Technological Change, Technological Catch-up, and Capital Deepening: Relative Contributions to Growth and Convergence. *American Economic Review* 92(3), 527-548.

[13] Kumbhakar, S. and C. A. K. Lovell (2000), *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge.

[14] Kumbhakar, S. C., B. U. Park, L. Simar and E. G. Tsionas (2007), Nonparametric Stochastic Frontiers: A Local Likelihood Approach, *Journal of Econometrics* 137(1), 1-27.

[15] Li, Q. and J.S. Racine (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

[16] Masry, E. (1996), Multivariate local polynomial regression for times series: uniform strong consistency and rates, *Journal of Times Series Analysis* 17, 571–599.

[17] Meeusen, W., and J. v. D. Broeck (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review* 18(2), 435-444.

[18] Olson, J.A., P. Schmidt, and D.M. Waldman (1980), "A Monte Carlo study of estimators of the stochastic frontier production function," *Journal of Econometrics* 13, 67-82.

[19] Pagan, A. and A. Ullah (1999), Nonparametric Econometrics, Cambridge University Press.

[20] Park, B. U., L. Simar and V. Zelenyuk (2008), Local Likelihood Estimation of Truncated Regression and its Partial Derivatives: Theory and Application, *Journal of Econometrics* 146(1), 185-198.

[21] Racine, J. S. (1997), Consistent Significance Testing for Nonparametric Regression, *Journal of Business and Economic Statistics*, 15, 3, 369–378.

[22] Ritter, C., and Simar, L. (1997), Pitfalls of Normal-Gamma Stochastic Frontier Models, *Journal of Productivity Analysis* 8, 167-182.

[23] Serfling, R.S. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New-York.

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

[25] Simar, L. and P.W. Wilson (2010), Inference From Cross-Sectional Stochastic Frontier Models. *Econometric Review*, 29, 1, 62–98.

[26] Simar, L. and V. Zelenyuk (2006), On Testing Equality of Two Distribution Functions of Efficiency Score Estimated via DEA, *Econometric Reviews*, 25(4), 497-522.

[27] Stevenson, R. E. (1980), Likelihood Functions for Generalized Stochastic Frontier Estimation, *Journal of Econometrics* 13(1), 57-66.

[28] Ziegelmann, F. (2002), Nonparametric Estimation of Volatility Functions: The Local Exponential Estimator, *Econometric Theory* 18, 985-991.