T E C H N I C A L

R E P O R T

0418

NONPARAMETRIC STOCHASTIC FRONTIERS: A LOCAL MAXIMUM LIKELIHOOD APPROACH

S.C. KUMBHAKAR, B.U. PARK, L. SIMAR and E.G. TSIONAS

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Nonparametric Stochastic Frontiers: A Local Maximum Likelihood Approach

Subal C. Kumbhakar
Department of Economics
State University of New York at Binghamton

Byeong U. Park*
Department of Statistics
Seoul National University

Léopold Simar†
Institut de Statistique
Université Catholique de Louvain

Efthymios G. Tsionas
Department of Economics
Athens University of Economics and Business

May 4, 2004

Abstract

This paper proposes a nonparametric approach for stochastic frontier (SF) models based on local maximum likelihood techniques. The SF model is presented as encompassing some anchorage parametric model in a nonparametric way. First, we derive asymptotic properties of the estimator for the general case (local linear approximations). Then the results are tailored to a SF model where the convoluted error term (efficiency plus noise) is the sum of an half normal and a normal random variable. The parametric anchorage model is a linear production function and an homoscedastic error term. The local approximation is thus local linear for the production function and local constant for the parameters of the error terms. The performance of our estimator is first established with a simulated data set and then with real data on milk production in Spanish dairy farms. The methods appear to be robust, numerically stable and particularly useful for investigating a production process and the derived efficiency scores.

*Research of B. U. Park was supported by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.

†Research support from the "Interuniversity Attraction Pole", Phase V (No. P5/24) from the Belgian Government (Belgian Science Policy) is acknowledged.

1 Introduction

The economic function of a firm is to bid resources away from alternative uses. As a result of such resource transfer, aggregate output may be increased or decreased. If inefficiency exists, an increase in output can be achieved by reallocating resources to more efficient uses. Given the seriousness of the issue (viz., the economic, political and social implications of inefficiency), it is essential that the measurement of efficiency/performance should be theoretically valid and subject to unambiguous interpretations.

Econometric measurement of efficiency of firms goes back to Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). When analyzing the performance of a firm that is observed to produce an output level $y \in \mathbb{R}$ using input quantities $x \in \mathbb{R}^d$, one typically compares the observed output level with the maximum possible output that can be obtained from the production frontier (defined as $f(x) = \max\{y : y \in P(x)\}$ where $P(x)$ describes the set of outputs that are feasible for each input vector $x \in \mathbb{R}^d$). The estimation of the production frontier is obtained from a random sample of observed firms $\{(X_i, Y_i) \,|\, i = 1, \ldots, n\}$. Then an efficiency score for a given production plan (x, y) is derived from the distance of this point to the estimated production frontier. The same approach can also be applied when a cost frontier is under the analysis. In the latter we seek the minimal cost achievable for a given level of output(s).

Since the publication of the seminal papers by Aigner *et al.* (1977) and Meeusen and van den Broeck (1977), econometric estimation of parametric stochastic frontier (SF) functions has become a standard practice in efficiency measurement studies. However, in this approach, the estimation relies heavily on the particular choices of the functional form of the production/cost frontier (Cobb-Douglas, Translog,..., etc.) as well as the specific distributional assumptions on the error term (a convolution of a one-sided inefficiency term and of a two-sided noise term). Typically these parametric models are written as

$$Y_i = \beta_0 + \beta^T X_i - u_i + v_i,\ i = 1, \ldots, n, \tag{1.1}$$

where $u_i > 0$ is the inefficiency term and $v_i \in \mathbb{R}$ represents random noise. The estimation technique is straightforward and is mostly based on the maximum likelihood principle.[1] Of course, in practice we cannot be confident about the validity of these parametric assumptions that are used to estimate the model. The parametric form of the frontier function might be wrong due to several reasons. For example, the parametric functional form might be wrong, the stochastic specifications of the error components (particularly for the inefficiency component) might be wrong, among others.

[1]Other methods, such as the COLS and MOLS (see Kumbhakar and Lovell, 2000) are often used.

An alternative to the parametric SF is the deterministic nonparametric approach where no specific parametric assumptions are made on the model. The frontier is defined as the upper boundary of the attainable set, say $\Psi = \{(x, y) \,|\, x \text{ can produce } y\}$. In these nonparametric approaches the statistical properties of envelopment estimators like the DEA and FDH[2] (Farrell, 1957, Charnes, Cooper and Rhodes, 1978 and Deprins, Simar and Tulkens, 1989), rely on the so called "deterministic" assumption, viz.,

$$\text{Prob}((X_i, Y_i) \in \Psi) = 1. \tag{1.2}$$

The latter implies that no noise is allowed in these deterministic frontier (DF) models. The introduction of noise in a full nonparametric setup is problematic due to identification problems (see Hall and Simar, 2002 and Simar, 2003). Statistical inference in now available in these nonparametric DF models (see Simar and Wilson, 2000 for a recent survey and Kneip, Simar and Wilson, 2003 for the asymptotic properties of DEA) but assumption (1.2) is too strong in many practical situations where we might expect measurement error, random shocks,..., etc. Recently Cazals, Florens and Simar (2002) and Aragon, Daouia and Thomas (2002) have proposed robust versions of the FDH estimator, robust to extremes values and/or outliers since they do not envelop all the data. But these approaches still rely heavily on the deterministic assumption (1.2), where no noise is allowed.

In the presence of panel data, Park and Simar (1994), Park, Sickles and Simar (1998), (2003a) and (2003b), in a series of papers, consider the semiparametric estimation of SF panel models under various assumptions on the joint distribution of the random firm effects and the regressors and on various dynamic specifications. The nonparametric part of these models concerns the distribution of the inefficiency terms. However, the estimators in these panel models are based on the linearity of the efficient frontier.

Fan, Li and Weersink (1996) propose a two-step pseudolikelihood estimator in a semiparametric model where the production frontier is not specified, but distributional assumptions are imposed on the stochastic components as in Aigner *et al.* (1977). An average production frontier is then estimated through standard kernel methods, the shift for the frontier is obtained through a moment condition, as in the MOLS approach (see Kumbhakar and Lovell, 2000) and the remaining parameters of the stochastic components are estimated by maximizing a pseudolikelihood function.

Our purpose in this paper is to propose a general nonparametric approach for stochastic frontier models. The method is based on the local maximum likelihood principle (see Tibshirani and Hastie, 1987, or Fan and Gijbels, 1996), which is nonparametric in the sense that the parameters of a given local polynomial model are localized with respect to the covariates

[2]DEA is for Data Envelopment Analysis and FDH stands for Free Disposal Hull.

of the model. As pointed out by Gozalo and Linton (2000), localizing can be viewed as a way of nonparametrically encompassing a parametric "anchorage" model. The idea to use local likelihood for stochastic frontier models was first suggested by Kumbhakar and Tsionas (2002) for a particular case of the model proposed here. In this paper we develop the general theory with all the asymptotics.

The paper is organized as follows. Section 2, presents the model and the theory with the main asymptotic results. Section 3 analyses the practical side, viz., how to compute the local estimators and determine the bandwidth. Section 4 reports results from both simulated and real data, and finally Section 5 concludes the paper. Regularity conditions and proofs are given in Section 6.

2 Main Results

2.1 The model

We consider a set of i.i.d. random variables (X_i, Y_i), for $i = 1, \dots, n$ with $X_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$. The joint pdf of (X, Y) is decomposed into a marginal pdf for X: $\mathrm{pdf}(x) = p(x)$ and a conditional pdf for Y given X: $\mathrm{pdf}(y|x) = g(y\,;\,\theta(x))$, where $\theta(x) \in \mathbb{R}^k$ is unknown and has to be estimated. The function g is assumed to be known.

The localized version of the parametric model in Aigner et al. (1997) is a particular case of our model. In this case, the conditional pdf for Y given $X = x$ would, for instance, be characterized through:

$$Y = r(X) - u + v \tag{2.1}$$

where $r(x)$ is the frontier function, $u|X = x \sim |\mathcal{N}(0, \sigma_u^2(x))|$ and $v|X = x \sim \mathcal{N}(0, \sigma_v^2(x))$, u and v being independent conditionally on X. Here $\theta(x) = (r(x), \sigma_u^2(x), \sigma_v^2(x))^T$ is a 3-dimensional local parameter. In our approach here we will consider local polynomial approximations for $\theta(x)$. For simplicity of presentation, we treat only the cases where the orders of local polynomials are equal for all the components of $\theta(\cdot)$. In practice, one may prefer to use different orders of polynomials for different components as in our numerical illustrations below. The theory for the latter cases may be obtained along the same lines of development as for the equal order cases.

The conditional log-likelihood can thus be written as

$$L(\theta) \triangleq \sum_{i=1}^{n} \log g(Y_i\,;\,\theta(X_i)) \tag{2.2}$$

In the next section we consider the order-m local polynomial estimator of $\theta(x)$ when x is univariate and then, in the following section, we derive the local linear estimator of $\theta(x)$ when x is multivariate.

2.2 Univariate case

Here we consider the case when $d = 1$. Let x be a fixed interior point in the support of $p(x)$ and let $q \equiv \log g$. Denote, for $j = 0, 1, \ldots, m$, $\theta_j \equiv (\theta_{j1}, \ldots, \theta_{jk})^T$. Then the conditional local log-likelihood for the m-th order local polynomial fit is given by

$$L_n(\theta_0, \theta_1, \ldots, \theta_m) \triangleq \sum_{i=1}^{n} q(Y_i\,;\, \theta_0 + \theta_1(X_i - x) + \ldots + \theta_m(X_i - x)^m) K_h(X_i - x) \tag{2.3}$$

where $K_h(\cdot) = (1/h)K(\cdot/h)$, K being a kernel function and h the appropriate bandwidth. Thus, the log-likelihood depends on the local x.

Then, the local polynomial estimator $\widehat{\theta}(x)$ is given by

$$\widehat{\theta}(x) = \widehat{\theta}_0(x), \tag{2.4}$$

where

$$(\widehat{\theta}_0(x), \ldots, \widehat{\theta}_m(x)) = \arg\max_{\theta_0, \ldots, \theta_m} L_n(\theta_0, \theta_1, \ldots, \theta_m). \tag{2.5}$$

We now derive the asymptotic distribution of $\widehat{\theta}(x)$. For this we have to introduce some additional notations. We define for $v \in \mathbb{R}^k$,

$$\begin{aligned} q_1(y, v) &\triangleq \frac{\partial}{\partial v} q(u, v) \\ q_2(y, v) &\triangleq \frac{\partial^2}{\partial v \partial v^T} q(u, v) \\ \rho(x) &\triangleq -E\left[q_2(Y_1\,;\, \theta(X_1)) | X_1 = x\right]. \end{aligned}$$

We also introduce the following matrices and vectors

$$N \triangleq \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_m \\ \mu_1 & \mu_2 & \cdots & \mu_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_m & \mu_{m+1} & \cdots & \mu_{2m} \end{pmatrix} \quad \text{where } \mu_j \triangleq \int u^j K(u) du,$$

$$S \triangleq \begin{pmatrix} \kappa_0 & \kappa_1 & \cdots & \kappa_m \\ \kappa_1 & \kappa_2 & \cdots & \kappa_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_m & \kappa_{m+1} & \cdots & \kappa_{2m} \end{pmatrix} \quad \text{where } \kappa_j \triangleq \int u^j K^2(u) du,$$

$$\theta^{(m+1)}(x) \triangleq \begin{pmatrix} \theta_1^{(m+1)}(x) \\ \vdots \\ \theta_k^{(m+1)}(x) \end{pmatrix} \quad \text{where } \theta_j^{(m+1)}(x) \triangleq \frac{\partial^{m+1}}{\partial x^{m+1}}\theta_j(x),$$

$$\gamma \triangleq \begin{pmatrix} \mu_{m+1} \\ \vdots \\ \mu_{2m+1} \end{pmatrix}.$$

Finally, we define

$$v(x) \triangleq E\left[q_1(Y_1\,;\,\theta(X_1))q_1^T(Y_1\,;\,\theta(X_1))\,|\,X=x\right].$$

Now we can state our theorem for the univariate case:

Theorem 2.1 *Under regularity conditions (see Section 6.1), it follows that*

$$\sqrt{nh}\left[(e_0^T N^{-1}SN^{-1}e_0)\rho(x)^{-1}v(x)\rho(x)^{-1}/p(x)\right]^{-1/2} \times \left(\widehat{\theta}(x)-\theta(x)-\frac{h^{m+1}}{(m+1)!}(e_0^T N^{-1}\gamma)\theta^{(m+1)}(x)+o_p(h^{m+1})\right) \xrightarrow{d} \mathcal{N}(0,I_k), \quad (2.6)$$

where $e_0=(1,0,\ldots,0)^T$ is the k-dimensional unit vector.

We notice, as for most of kernel based nonparametric estimators, the role of the bandwidth h which balances between the bias and the variance of the estimator. An optimal asymptotic value of h can be derived from the above theorem[3] (e.g., minimizing the asymptotic MSE would lead to $h \asymp n^{-1/(2m+3)}$)

Note also as a special case, when $m=1$ (local linear estimator) we have:

$$e_0^T N^{-1}\gamma = \frac{\mu_2^2-\mu_1\mu_3}{\mu_0\mu_2-\mu_1^2} = \mu_2$$

$$e_0^T N^{-1}SN^{-1}e_0 = \frac{\kappa_0\mu_2^2-2\kappa_1\mu_1\mu_2+\kappa_2\mu_1^2}{(\mu_0\mu_2-\mu_1^2)^2} = \kappa_0$$

where the two equalities on the right are valid for appropriate symmetric kernels, with $\mu_0=1$ and $\mu_1=\mu_3=0$.

2.3 Multivariate case

In this case, $\theta(x)$ is a function of $x \in \mathbb{R}^d$ and we consider only the local linear fit. For simplicity of presentation, we assume further that the multivariate kernel K satisfies

$$\int K(u)\,du = 1 \quad \text{and} \quad \int uu^T K(u)\,du = \mu_2 I_d.$$

[3] In practice the bandwidth choice is based on cross validation. We address this issue in Section 3.

In the multiple covariate case, the conditional local linear log-likelihood is given by

$$L_n(\theta_0, \Theta_1) \triangleq \sum_{i=1}^{n} q\left(Y_i\,;\, \theta_0 + \Theta_1(X_i - x)\right) K_H(X_i - x).$$

Here θ_0 is a $(k \times 1)$ vector, Θ_1 is a $(k \times d)$ matrix, H is a bandwidth matrix which we assume is positive definite and symmetric, and $K_H(u) = |H|^{-1} K(H^{-1}u)$. For instance we could chose a multivariate product kernel as

$$K(u) = K_1(u_1) \ldots K_1(u_d)$$

where $K_1(\cdot)$ is a symmetric univariate probability density. In this case

$$\int uu^T K(u)\,du = \left(\int u_1^2 K_1(u_1)\,du_1\right) I_d.$$

The local linear estimator $\widehat{\theta}(x)$ is given by

$$\widehat{\theta}(x) = \widehat{\theta}_0(x) \tag{2.7}$$

where $\widehat{\theta}_0(x)$ and $\widehat{\Theta}_1(x)$ maximize $L_n(\theta_0, \Theta_1)$ with respect to θ_0 and Θ_1.

Define $\rho(\cdot)$ and $v(\cdot)$ as in the case $d = 1$. Let

$$B_H(x) \triangleq \begin{pmatrix} \mathrm{tr}(\theta_1''(x)H^2) \\ \vdots \\ \mathrm{tr}(\theta_k''(x)H^2) \end{pmatrix}$$

where $\theta_j(x)$ is the j-th component of $\theta(x)$ and $\theta_j''(x)$ is the $(d \times d)$ Hessian matrix of $\theta_j(x)$. Now we can state our theorem (the proof is omitted):

Theorem 2.2 *Under the regularity conditions (see Section 6.1), it follows that*

$$\begin{aligned}(n|H|)^{1/2} &\left[\frac{1}{p(x)} \int K^2(u)\,du \cdot \rho(x)^{-1} v(x) \rho(x)^{-1}\right]^{-1/2} \\ &\times \left(\widehat{\theta}(x) - \theta(x) - \frac{1}{2}\mu_2 B_H(x)\right) \xrightarrow{d} \mathcal{N}(0, I_k).\end{aligned} \tag{2.8}$$

3 Practical Computations

3.1 Computation of the local linear estimator

We will illustrate how the computations could be performed in the case of a local linear fit choosing for the local convolution of the noise and the inefficiency term, the convolution of a normal and an half normal, in the spirit of Aigner et al (1977) and Meeusen and van den

Broeck (1977). Other choices are also available, such as the convolution of a normal and an exponential, but to save space, we present the normal-half normal case. So the parametric "anchorage" model is the following:

$$Y = \beta_0 + \beta^T X - u + v, \tag{3.1}$$

where $u \sim |\mathcal{N}(0, \sigma_u^2)|$ and $v \sim \mathcal{N}(0, \sigma_v^2)$, u and v being mutually independent and both independent of X.

Our nonparametric localized model is tailored as

$$Y = r(X) - u + v, \tag{3.2}$$

where $u|X = x \sim |\mathcal{N}(0, \sigma_u^2(x))|$ and $v|X = x \sim \mathcal{N}(0, \sigma_v^2(x))$, u and v being independent conditionally on X. The conditional probability density function of $\varepsilon = v - u$ is given by[4]

$$f(\varepsilon|X = x) = \frac{2}{\sigma(x)}\,\varphi\left(\frac{\varepsilon}{\sigma(x)}\right)\Phi\left(-\varepsilon\,\frac{\lambda(x)}{\sigma(x)}\right)$$

where $\sigma^2(x) = \sigma_u^2(x) + \sigma_v^2(x)$ and $\lambda(x) = \sigma_u(x)/\sigma_v(x)$. Finally, $\varphi(\cdot)$ and $\Phi(\cdot)$ are the pdf and CDF of a standard normal variable.

Thus, $\theta(x) = (r(x), \sigma^2(x), \lambda(x))^T$ and the conditional pdf of Y given X is

$$g(y\,;\,\theta(x)) = \frac{2}{\sigma(x)}\,\varphi\left(\frac{y - r(x)}{\sigma(x)}\right)\Phi\left(-(y - r(x))\,\frac{\lambda(x)}{\sigma(x)}\right). \tag{3.3}$$

The conditional local log-likelihood is given by

$$L(\theta) \propto \sum_{i=1}^{n}\left[-\frac{1}{2}\log\sigma^2(X_i) - \frac{1}{2}\frac{(Y_i - r(X_i))^2}{\sigma^2(X_i)} + \log\Phi\left(-(Y_i - r(X_i))\,\frac{\lambda(X_i)}{\sqrt{\sigma^2(X_i)}}\right)\right], \tag{3.4}$$

where the constants have been eliminated. Then conditional local log-likelihood for the local linear fit at the point x, is given by

$$\begin{aligned} L_n(\theta_0, \Theta_1) \propto \sum_{i=1}^{n}\Bigg[&-\frac{1}{2}\log(\sigma_0^2 + \sigma_1^{2\,T}(X_i - x)) - \frac{1}{2}\frac{(Y_i - r_0 - r_1^T(X_i - x))^2}{\sigma_0^2 + \sigma_1^{2\,T}(X_i - x)} \\ &+ \log\Phi\left(-(Y_i - r_0 - r_1^T(X_i - x))\,\frac{\lambda_0 + \lambda_1^T(X_i - x)}{\sqrt{\sigma_0^2 + \sigma_1^{2\,T}(X_i - x)}}\right)\Bigg] K_H(X_i - x), \end{aligned} \tag{3.5}$$

[4]In the case of the estimation of a cost function where $\varepsilon = v + u$, we would obtain:

$$f(\varepsilon|X = x) = \frac{2}{\sigma(x)}\,\varphi\left(\frac{\varepsilon}{\sigma(x)}\right)\Phi\left(\varepsilon\,\frac{\lambda(x)}{\sigma(x)}\right)$$

where $\theta_0 = (r_0, \sigma_0^2, \lambda_0)^T$ and the $(3 \times d)$ matrix of parameters Θ_1 is given by $\Theta_1^T = (r_1, \sigma_1^2, \lambda_1)$ with r_1, σ_1^2 and λ_1 being $(d \times 1)$ vectors. Let

$$(\widehat{\theta}_0(x), \widehat{\Theta}_1(x)) = \arg\max_{\theta_0, \Theta_1} L_n(\theta_0, \Theta_1), \tag{3.6}$$

then our local linear estimator of the model is given by $\widehat{\theta}_0(x)$. From $\widehat{\sigma}_0^2(x)$ and $\widehat{\lambda}_0(x)$, values of $\widehat{\sigma}_u^2(x)$ and $\widehat{\sigma}_v^2(x)$ can also be derived.

The estimation of the individual efficiency score for a particular point (x, y) might be obtained by following the Jondrow, Lovell, Materov and Schmidt (1982) procedure. Denoting $\varepsilon(x) = y - r(x)$, it can be shown in the local parametric model chosen here that

$$u(x) \,|\, \varepsilon(x), X = x \sim \mathcal{N}^+(\mu^*(x), \sigma^{2*}(x))$$

i.e., a truncated (positive values) normal, where

$$\begin{aligned} \mu^*(x) &= \frac{-\varepsilon(x)\sigma_u^2(x)}{\sigma^2(x)} \\ \sigma^{2*}(x) &= \frac{\sigma_u^2(x)\sigma_v^2(x)}{\sigma^2(x)}. \end{aligned}$$

In particular, we can compute

$$E(u(x) \,|\, \varepsilon(x), X = x) = \frac{\sigma(x)\lambda(x)}{1 + \lambda^2(x)} \left[\frac{\varphi(-\varepsilon(x)\lambda(x)/\sigma(x))}{\Phi(-\varepsilon(x)\lambda(x)/\sigma(x))} - \frac{\varepsilon(x)\lambda(x)}{\sigma(x)} \right].$$

As in Jondrow et al. (1982) a point estimator of the individual efficiency score for an observation (X_i, Y_i) could be obtained from $\widehat{u}_i = \widehat{E}[u(X_i) \,|\, \widehat{\varepsilon}(X_i), X = X_i]$, where $\widehat{\varepsilon}(X_i) = Y_i - \widehat{r}_0(X_i)$, viz.,

$$\widehat{u}_i = \frac{\widehat{\sigma}_0(X_i)\widehat{\lambda}_0(X_i)}{1 + \widehat{\lambda}_0^2(X_i)} \left[\frac{\varphi(-\widehat{\varepsilon}(X_i)\widehat{\lambda}_0(X_i)/\widehat{\sigma}_0(X_i))}{\Phi(-\widehat{\varepsilon}(X_i)\widehat{\lambda}_0(X_i)/\widehat{\sigma}_0(X_i))} - \frac{\widehat{\varepsilon}(X_i)\widehat{\lambda}_0(X_i)}{\widehat{\sigma}_0(X_i)} \right]. \tag{3.7}$$

We know of course, as in the full parametric case, that this is a rather poor predictor of u_i, since it is based on a single "noisy" observation $\widehat{\varepsilon}(X_i)$. As usual in frontier models, if the variables are measured in logs, a point estimate of the efficiency is then provided by $\widehat{eff}_i = \exp(-\widehat{u}_i) \in [0, 1]$.

In the numerical illustrations below, we will customize the analysis to the case of a nonparametric model encompassing the anchorage parametric frontier model (3.1). So we will use a local linear model for the frontier and a local constant model for the parameters of the error term. In the above notation, this merits the particular choice of $\lambda_1 = \sigma_1^2 = 0$ and $\Theta_1 = r_1^T$ for the setup of (3.5). Solving (3.6) requires iterative algorithms, starting with some initial values. For example, we could choose $\Theta_1 = \widehat{\beta}^T$ and $\theta_0 = (\widehat{\beta}_0, \widehat{\sigma}^2, \widehat{\lambda})$, the parametric maximum likelihood estimator from the model (3.1), as the initial values of Θ_1 and θ_0.

The following is an alternative way to choose the starting values. Start with the local linear least squares estimator $\widehat{r}_0(x)$ and $\widehat{r}_1(x)$ of Gozalo and Linton (2000) and then correct the local intercept $\widehat{r}_0(x)$ for the moment condition along the lines of the parametric MOLS estimators (see, *e.g.*, Kumbhakar and Lovell, 2000). Here we suggest the use of the global parametric MLE $\widehat{\sigma}^2$ and $\widehat{\lambda}$ for the moment correction. Since the local linear $\widehat{r}_0(x)$ can be viewed as an estimator of $r_0(x) - E(u|X = x) = r_0(x) - \sqrt{2\sigma_u^2(x)/\pi}$, a sensible moment corrected estimator for the local intercept is obtained through $\widehat{r}_0^{MOLS}(x) = \widehat{r}_0(x) + \sqrt{2\widehat{\sigma}_u^2/\pi}$, where $\widehat{\sigma}_u^2 = \widehat{\sigma}^2\widehat{\lambda}^2/(1 + \widehat{\lambda}^2)$. In this case, the initial values for solving (3.6) are finally given by $\Theta_1 = \widehat{r}_1(x)^T$ and $\theta_0 = (\widehat{r}_0^{MOLS}(x), \widehat{\sigma}^2, \widehat{\lambda})$. In the optimization algorithms used in the numerical illustrations below, the latter choice of starting values proved to be very efficient and numerically stable.

3.2 Bandwidth selection

In practice we could choose a product kernel with a bandwidth $H = hV$ where V is the empirical variance-covariance matrix of the d covariates X_i. The choice of the bandwidth is then reduced to the selection of the scalar h. An alternative is to use the product kernel $h^{-d}\prod_{j=1}^{d} K(h^{-1}(x_j))$, so $H = hI_d$. In the numerical illustrations below we chose the more sensible d-dimensional vector of bandwidths h as

$$h = h_{base} s_X n^{-1/5}, \tag{3.8}$$

where s_X is the vector of empirical standard deviations of the d components of X. So the bandwidth is adjusted for different scales of the variables and different sample sizes. Then the cross validation criterion is evaluated for a grid of values for h_{base}.

The cross-validation proceeds as follows. For a given value of h_{base}, we compute

$$CV(h_{base}) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - (\widehat{r}_0^{(i)}(X_i) - \widehat{u}_i^{(i)})]^2 \tag{3.9}$$

where $\widehat{r}_0^{(i)}$ and $\widehat{u}^{(i)}$ are the leave-one-out version of the local linear estimators derived above. The optimal value for h_{base} is then easily found by an appropriate grid search.

If n is large, the evaluation in (3.9) could be performed on a random subsample of m units, where $m \ll n$ to reduce the computational burden. Also, a trimmed version of the average of (3.9) might be useful in the sense that it is less sensitive to potential numerical problems when computing the local ML many times.

4 Numerical Illustrations

4.1 Some simulated data sets

4.1.1 Example 1: A linear model

In this first case, we chose a standard linear model where the anchorage parametric model is true. We simulated a sample of size 100 from the model:

$$Y = 5 + 5X - u + v,$$

where $X \sim U(0,1)$, $u \sim |\mathcal{N}(0,2^2)|$ and $v \sim \mathcal{N}(0,2^2)$. The cross validation, as expected, gave a flat value of $CV(h)$ for values of h_{base} greater than, say, 6 (see Figure 1). Indeed, larger values of h_{base} makes our local maximum likelihood estimator equal to the full parametric MLE, as it should be in this particular example. A trimmed version (deleting the upper 5% of the values in the average of (3.9)) gives very similar results. The fit proposed in Figure 2 was obtained for a value of $h_{base} = 11$ and is very good. The statistical noise introduced by encompassing the true linear model did not affect our estimator.

Figure 1 around here

Figure 2 around here

4.1.2 Example 2: A quadratic model with heteroscedasticity

Here we simulate the data ($n = 100$) with a nonlinear model with heteroscedasticity for the efficiency term. Now, the chosen parametric anchorage model is wrong but as seen below, its localized estimated version fit the model pretty well. This illustrates the nonparametric nature of our estimator. The model is

$$Y = 5 + 15X - 8X^2 - u + v,$$

where $X \sim U(0,1)$, $u|X = x \sim |\mathcal{N}(0,(2+x)^2)|$ and $v \sim \mathcal{N}(0,2^2)$. The bandwidth selection procedure provided an optimal value of $h_{base} = 1.80$ as shown in Figure 3 and the excellent fit is displayed in Figure 4. The localized ML is able to capture the curvature and is not perturbed by the linear heteroscedasticity.

Figure 3 around here

Figure 4 around here

4.1.3 Example 3: A multivariate quadratic model with heteroscedasticity

This is a bivariate extension of the preceding example. The model is written as

$$Y = 5 + X_1 + 2X_2 - 0.25X_1X_2 - u + v,$$

where $X_1 \sim U(0,1)$, $X_2 \sim U(0,2)$, $u|X = x \sim |\mathcal{N}(0,(2+x)^2)|$ and $v \sim \mathcal{N}(0, 0.5^2)$. Here again $n = 100$. As shown in Figure 5, the optimal bandwidth is around $h_{base} = 9$. To give an idea of the quality of the fit in this multiple model, the left panel of Figure 6 displays the plot of the true frontier value against the fitted values. The right panel illustrates the quality of the prediction of the efficiency terms by showing the plot of the true simulated u_i against its predictor $\widehat{u}_i$, as provided by (3.7). Surprisingly, we can see that the difficult task of predicting one random variable u_i, conditionally on one observation of a noisy residual $\widehat{\varepsilon}(X_i)$ (see (3.7)) works pretty well globally. Of course, some individual prediction might be far from the truth.

Figure 5 around here *Figure 6 around here*

Table 1 provides detailed results for 25 units randomly chosen. The Table also provides the full parametric ML estimates of the frontier and of the efficiency terms. In all the cases, the localized version is more close to the true values.

Table 1 around here

4.2 A real data example

As a final illustration, we worked out an example with real data. The analysis is based on a balanced panel data set of 80 Spanish dairy farms for the years 1993 to 1998 (see Alvarez, Arias and Kumbhakar (2003) for details). These are all small family farms. We consider one output (liters of milk) and four variable inputs, viz., number of cows (Cows), kilograms of concentrates (Conc), hectares of land (Land), and labor (measured in man-equivalent units) (Labor). We also added a time trend variable (Trend) to capture technical change (a shift in the production technology over time). The production function is assumed to be Cobb-Douglas. All the variables are in logs, except the trend variable. The results from the parametric maximum likelihood method are given in Table 2. The covariance matrix of the estimators are computed from the inverse of the computed Hessian. The anchorage parametric model is

$$Milk_{it} = \beta_0 + \beta_1 Cows_{it} + \beta_2 Land_{it} + \beta_3 Labor_{it} + \beta_4 Conc_{it} + \beta_5 t - u_{it} + v_{it},$$

with the stochastic specification as in (3.1).

$\boxed{\textit{Table 2 around here}}$

The estimated parameters are reported in Table 2. Only the Cows and Conc variables seems to be highly significant. The Labor and Trend variables appear to be less significant and the Land variable is far from being significant.

The localized version of these estimators requires the selection of the bandwidth. The CV criterion was selected at $m = 100$ randomly chosen points with the leave-one out formula given in (3.9). The results are shown in Figure 7: the optimal value is found to be around $h_{base} = 5$.

$\boxed{\textit{Figure 7 around here}}$

Figure 8, illustrates the variation of the estimates of the localized parameters, evaluated at the 480 data points by plotting the histograms. The histograms of s_0^2 and of λ_0 clearly indicate the presence of heteroscedasticity with an important heterogeneity in the shares of efficiency to noise (λ_0). The different histograms for the input coefficients (elasticities) $r_{1j}, j = 1, \ldots, 5$ are also illuminating. Since these estimates are observation-specific (thereby meaning that input elasticities vary across farms and over time) the variation in input elasticities show that the results from the parametric CD model (that assumes same elasticities for all observations) might be wrong. For example, the elasticity of the "Cows" variable (r_{11}) is evidently positive for all the farms but these are far from being the same for all farms. That is, we find heterogeneity in the estimated elasticity of the input "Cows" among the farms. The elasticity of the "Land" variable (r_{12}) has an average value that is not far from zero (this is also supported the standard MLE analysis). The elasticity of "Conc" (r_{14}) is positive but is less heterogeneous than the variable "Cows". Note also that the elasticity of "Labor" (r_{13}) is positive but its mean is not that much different from zero. Finally, the elasticity with respect to the "Trend" variable (r_{15}) (which is labeled as technical change) appears to be positive (meaning technical progress, i.e., an outward shift in the production function). But its mean values is quite small (smaller than what it's estimated value is in the parametric ML model).

$\boxed{\textit{Figure 8 around here}}$

In Table 3 we report the detailed results for 25 farms randomly drawn from the full data set. The last column reports the efficiency scores. Note that in this example the average efficiency score for the 480 farms is 0.8707.

$\boxed{\textit{Table 3 around here}}$

5 Conclusions

In this paper we have provided a nonparametric approach for estimating stochastic frontier models. The idea is based on local maximum likelihood techniques. The model can be presented as encompassing some anchorage parametric model in a nonparametric way. The estimation here is obtained by localizing the likelihood function to be maximized. The asymptotic properties of our estimator is established for the general setup of local linear approximations.

We tailored the results to a stochastic production frontier model where the convoluted/composed error term (inefficiency plus noise) is the sum of an half normal and a normal random variable. The parametric anchorage model is a linear production function and a homoscedastic error term. The local approximation is thus local linear for the production function and local constant for the parameters of the error terms.

The performance of the estimator is first illustrated with some simulated data sets. Then we apply it to a real data set (Spanish dairy farms) to test the flexibility of our method. We find enough variability in the estimated coefficient to suggest that the underlying production technology is heterogeneous so far as the input elasticities and the estimates of variances of the composed error are concerned. Based on this result, we conclude that estimated technology from a parametric model and elasticities, efficiencies, etc., derived therefrom might be wrong, especially if the technology is heterogeneous.

All these numerical illustrations indicate that the methods is robust, numerically stable and particularly useful for investigating a production process and the derived efficiency scores.

6 Regularity Conditions and Proof of Theorem 2.1

6.1 Regularity Conditions

The assumptions for Theorem 2.1 are as follows:

(A1) $q_2(u, t) < 0$ for $t \in \mathbb{R}^k$ and u in the range of Y;

(A2) $E[q_1(Y_1\,; \theta(X_1) \mid X_1] = 0$;

(A3) $q_3(u, t)$, the third partial derivatives of $q(u, t)$ with respect to t are continuous for $t \in \mathbb{R}^k$;

(A4) $p(x) > 0$, and the matrices $\rho(x)$ and $v(x)$ are positive definite;

(A5) p, all entries of ρ and v are continuous at x;

(A6) K has compact support with nonempty interior;

(A7) All entries of θ have $(m+1)$-th continuous derivatives at x;

(A8) The bandwidth h tends to zero as n goes to infinity, and satisfies $nh^{2m+3} < C$ for some positive constant C.

For Theorem 2.2 we need (A1)∼(A6) plus the following assumptions:

(A7') All entries of θ are twice partially continuously differentiable at x;

(A8') All entries of the bandwidth matrix H tends to zero as n goes to infinity, and satisfies $n|H|^5 < C$ for some positive constant C.

6.2 Proof of Theorem 2.1

We introduce additional notations:

$$\widetilde{\theta}(u) \triangleq \theta(x) + \theta'(x)(u-x) + \ldots + \theta^{(m)}(x)(u-x)^m/m!$$
$$\widehat{a}_j \equiv \widehat{a}_j(x) \triangleq (nh)^{1/2}h^j \left(\widehat{\theta}_j(x) - \frac{\theta^{(j)}(x)}{j!}\right), \quad j = 0, \ldots, m,$$

where

$$\theta^{(j)}(x) \triangleq \begin{pmatrix} \theta_1^{(j)}(x) \\ \vdots \\ \theta_k^{(j)}(x) \end{pmatrix} \quad \text{where } \theta_i^{(j)}(x) \triangleq \frac{\partial^j}{\partial x^j}\theta_i(x)$$

Then, $(\widehat{a}_0, \ldots, \widehat{a}_m)$ is the maximizer with respect to $(a_0, a_1, \ldots, a_m)$ of

$$\begin{aligned} L_n^*(a_0, a_1, \ldots, a_m) \;\; \triangleq \;\; & h\sum_{i=1}^n \Big[q\left(Y_i\,;\,\widetilde{\theta}(X_i) + (nh)^{-1/2}\left(a_0 + a_1(\frac{X_i - x}{h}) + \ldots + a_m(\frac{X_i - x}{h})^m\right)\right) \\ & -q(Y_i\,;\,\widetilde{\theta}(X_i))\Big] K_h(X_i - x). \end{aligned}$$

Now, defining

$$Z_i \triangleq \left(1, \left(\frac{X_i - x}{h}\right), \ldots, \left(\frac{X_i - x}{h}\right)^m\right)^T$$

and

$$\begin{aligned} W_n \;\; & \triangleq \;\; \left(\frac{h}{n}\right)^{1/2} \sum_{i=1}^n \Big[Z_i \otimes q_1(Y_i\,;\,\widetilde{\theta}(X_i))\Big] K_h(X_i - x) \\ B_n \;\; & \triangleq \;\; -\frac{1}{n}\sum_{i=1}^n \Big[(Z_i Z_i^T) \otimes q_2(Y_i\,;\,\widetilde{\theta}(X_i))\Big] K_h(X_i - x) \end{aligned}$$

where $\otimes$ is the Kronecker product between matrices, it can be shown that

$$L_n^*(a_0, a_1, \ldots, a_m) = W_n^T a - \frac{1}{2} a^T B_n a + o_p(1)$$

uniformly for a in any compact subset of $\mathbb{R}^{mk}$ (here $a = (a_0^T, \ldots, a_m^T)^T$)).

This implies that

$$\begin{aligned} \widehat{a} &\equiv (\widehat{a}_0^T, \ldots, \widehat{a}_m^T)^T) \\ &= B_n^{-1} W_n + o_p(1) \end{aligned} \qquad (6.1)$$

First we approximate $E(W_n)$. Note that

$$E\left[q_1(Y_1\,;\,\theta(X_1))\,|\,X_1\right] = 0$$

from the regularity conditions. Thus,

$$\begin{aligned} (nh)^{-1/2} E(W_n) &= E\left[\left\{Z_1 \otimes q_1(Y_1\,;\,\widetilde{\theta}(X_1))\right\} K_h(X_1 - x)\right] \\ &= E\left[\{Z_1 \otimes q_2(Y_1\,;\,\theta(X_1))\}\,\{\widetilde{\theta}(X_1) - \theta(X_1)\} K_h(X_1 - x)\right] + O(h^{2m+2}) \\ &= E\left[\{Z_1 \otimes \rho(X_1)\}\,\{\theta(X_1) - \widetilde{\theta}(X_1)\} K_h(X_1 - x)\right] + O(h^{2m+2}) \\ &= \frac{1}{(m+1)!}\,\{\gamma \otimes \rho(x)\}\,\theta^{(m+1)}(x)\,p(x)\,h^{m+1} + o(h^{m+1}) \end{aligned}$$

Next, we approximate $Var(W_n)$.

$$\begin{aligned} Var(W_n) &= h\,Var\left[\left\{Z_1 \otimes q_1(Y_1\,;\,\widetilde{\theta}(X_1))\right\} K_h(X_1 - x)\right] \\ &= E\left[\left\{(Z_1 Z_1^T) \otimes \left(q_1(Y_1\,;\,\widetilde{\theta}(X_1)) q_1^T(Y_1\,;\,\widetilde{\theta}(X_1))\right)\right\} (K^2)_h(X_1 - x)\right] + O(h^{2m+3}) \\ &= E\left[\left\{(Z_1 Z_1^T) \otimes v(X_1)\right\} (K^2)_h(X_1 - x)\right] + O(h^{m+1}) \\ &= \{S \otimes v(x)\}\;p(x) + o(1). \end{aligned}$$

Then, we approximate B_n

$$\begin{aligned} B_n &= \frac{1}{n}\sum_{i=1}^{n}\left\{(Z_i Z_i^T) \otimes \rho(X_i)\right\} K_h(X_i - x) + O_p(h^{m+1} + n^{-1/2}\,h^{-1/2}) \\ &= \{N \otimes \rho(x)\}\,p(x) + o(1) + O_p(h^{m+1} + n^{-1/2}\,h^{-1/2}). \end{aligned}$$

Now from (6.1) we have

$$\widehat{a}_0 - (e_0^T \otimes I_k) B_n^{-1} E(W_n) = (e_0^T \otimes I_k) B_n^{-1} (W_n - E(W_n)) + o_p(1).$$

Thus

$$\left[(e_0^T \otimes I_k) B_n^{-1} Var(W_n) B_n^{-1} (e_0 \otimes I_k)\right]^{-1/2} \left(\widehat{a}_0 - (e_0^T \otimes I_k) B_n^{-1} E(W_n)\right) \xrightarrow{d} \mathcal{N}(0, I_k).$$

Using the properties of Kronecker product that $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ and $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$, it may be seen that

$$(e_0^T \otimes I_k) B_n^{-1} Var(W_n) B_n^{-1} (e_0 \otimes I_k) = (e_0^T N^{-1} S N^{-1} e_0) \rho(x)^{-1} v(x) \rho(x)^{-1} / p(x) + o_p(1)$$

and

$$(nh)^{-1/2} (e_0^T \otimes I_k) B_n^{-1} E(W_n) = \frac{h^{m+1}}{(m+1)!} (e_0^T N^{-1} \gamma) \theta^{(m+1)}(x) + o_p(h^{m+1}),$$

which completes the proof. ■

References

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

[2] Alvarez, A., Arias, C. and Kumbhakar, S. (2003), Empirical consequences of direction choice in technical efficiency analysis, manuscript, Department of Economics, University of Oviedo, Spain.

[3] Aragon, Y. and Daouia, A. and Thomas-Agnan, C. (2002), Nonparametric Frontier Estimation: A Conditional Quantile-based Approach, Discussion paper, GREMAQ et LSP, Université de Toulouse (http://www.univ-tlse1.fr/GREMAQ/Statistique/adt1202.pdf).

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

[5] Charnes, A., Cooper W.W. and E. Rhodes (1978), Measuring the inefficiency of decision making units, *European Journal of Operational Research* 2 (6), 429-444.

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

[7] Fan, Y. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

[8] Fan, Y., Li, Q. and Weersink, A. (1996), Semiparametric estimation of stochastic production frontier models, *Journal of Business and Economic Statistics*, Vol. 14, 4, 460–468.

[9] Farrell, M.J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society*, A(120), 253-281.

[10] Gozalo, P. and Linton, O. (2000), Local nonlinear least squares: using parametric information in nonparametric regression, *Journal of Econometrics*, 99, 63–106.

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

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

[13] Kneip, A, L. Simar and Wilson, P.W. (2003), Asymptotics for DEA Estimators in Nonparametric Frontier Models, Discussion paper 0317, Institut de Statistique, UCL (http://www.stat.ucl.ac.be).

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

[15] Kumbhakar, S.C. and Tsionas, E.G. (2002), Nonparametric stochastic frontier models, manuscript presented at the NAPWII, June 2002, Schenectady, New York.

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

[17] Park, B., Sickles R.C. and Simar, L. (1998), Stochastic Panel Frontiers : A Semiparametric Approach, *Journal of Econometrics*, 84, 273–301.

[18] Park, B., Sickles, R.C. and Simar, L. (2003a), Semiparametric Efficient Estimation of AR(1) Panel Data Models, *Journal of Econometrics*, 117, 279–309

[19] Park, B., Sickles, R.C. and Simar, L. (2003b), Semiparametric Efficient Estimation in Dynamic Panel Data Models, Discussion 0315, Institut de Statistique, UCL, Belgium (http://www.stat.ucl.ac.be).

[20] Park, B. and Simar, L. (1994), Efficient semiparametric estimation in a stochastic frontier model, *Journal of the American Statistical Association*, 89 (427),929–936.

[21] Simar, L. (2003), How to Improve the Performances of DEA/FDH Estimators in the Presence of Noise, Discussion paper 0323, Institut de Statistique, UCL, Louvain-la-Neuve, Belgium, (http://www.stat.ucl.ac.be).

[22] Simar L. and Wilson, P.W. (2000), Statistical Inference in Nonparametric Frontier Models: The State of the Art, *Journal of Productivity Analysis*, 13, 49–78.

[23] Tibshirani, R. and Hastie, T.J. (1987), Local likelihood estimation, *Journal of the American Statistical Association*, 82, 559–567.

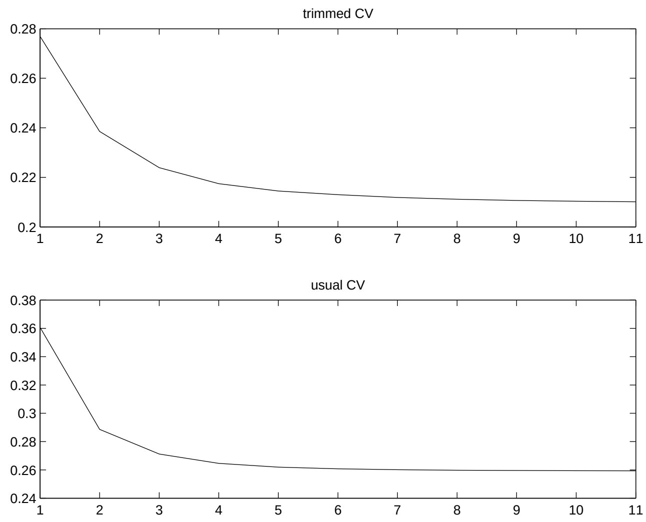

Figure 1: *Example 1, Linear model: CV criterion. Computing time for the grid search over 11 values of* h_{base}*, 339 seconds on a Pentium III, 450 Mghz machine.*

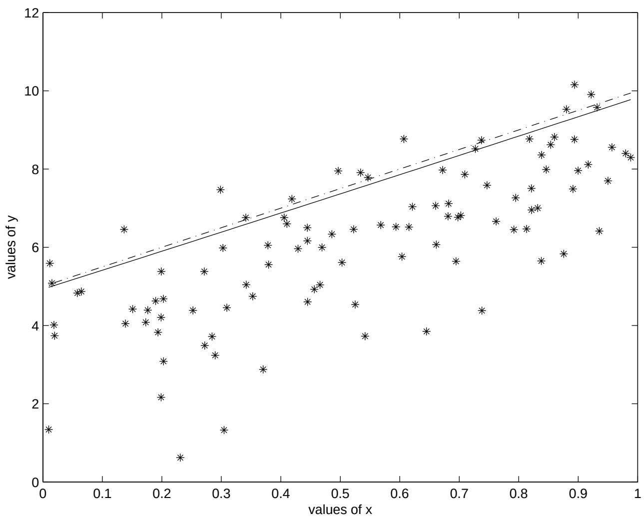

Figure 2: *Example 1, Linear model: The true frontier (dashed-dotted) and the local maximum likelihood fit (solid). Computation time, 30 seconds on a Pentium III, 450 Mghz machine.*

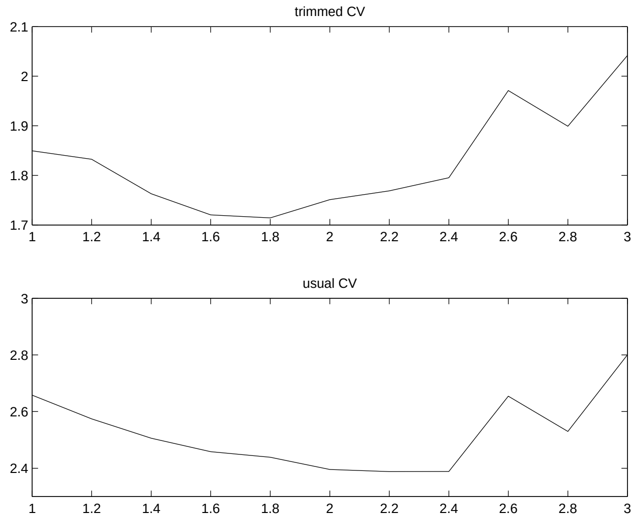

Figure 3: *Example 2, quadratic model with heteroscedasticity: CV criterion. Computing time for the grid search over 11 values of h_{base}, 426 seconds on a Pentium III, 450 Mghz machine.*

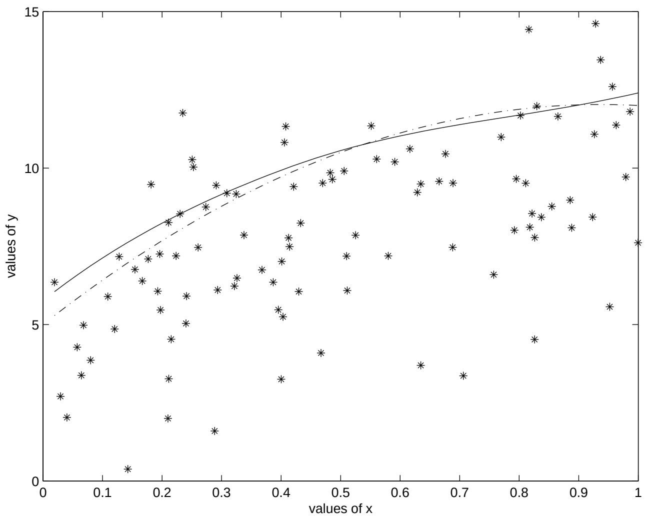

Figure 4: *Example 2, quadratic model with heteroscedasticity: The true frontier (dashed-dotted) and the local maximum likelihood fit (solid).*

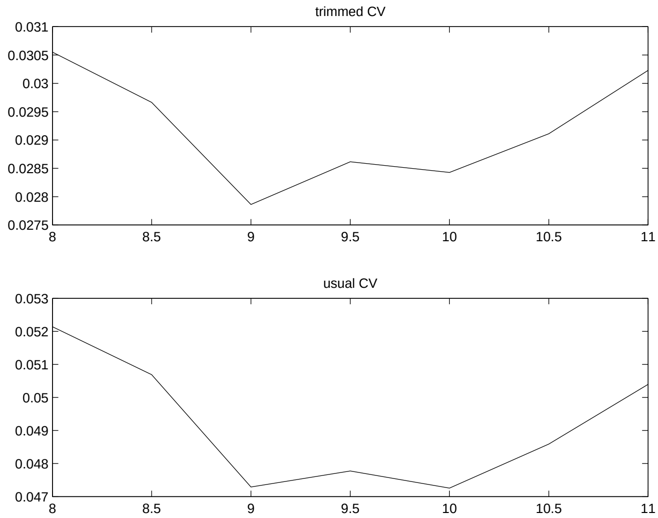

Figure 5: *Example 3: CV criterion. Computing time for the grid search over 7 values of* h_{base}*, 280 seconds on a Pentium III, 450 Mghz machine.*

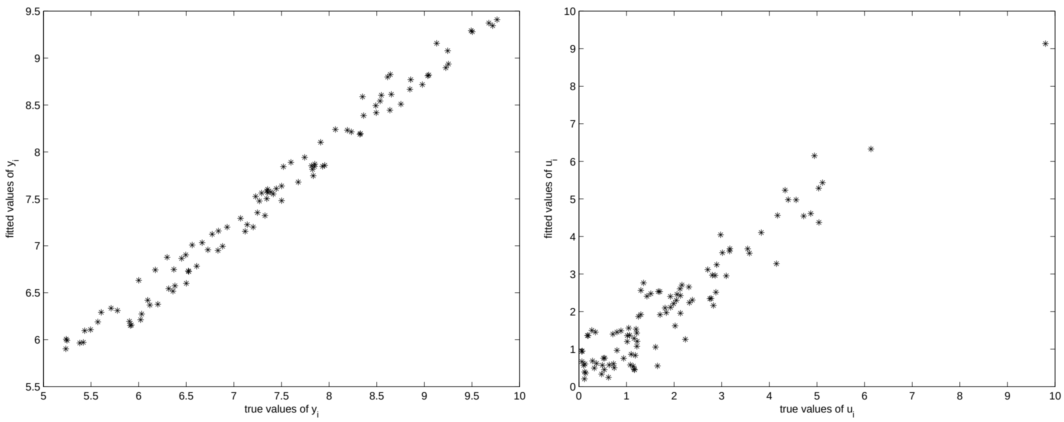

Figure 6: *Example 3: True against estimated frontier (left panel) and true against predicted efficiencies (right panel), for* $i = 1, \ldots, 100$.

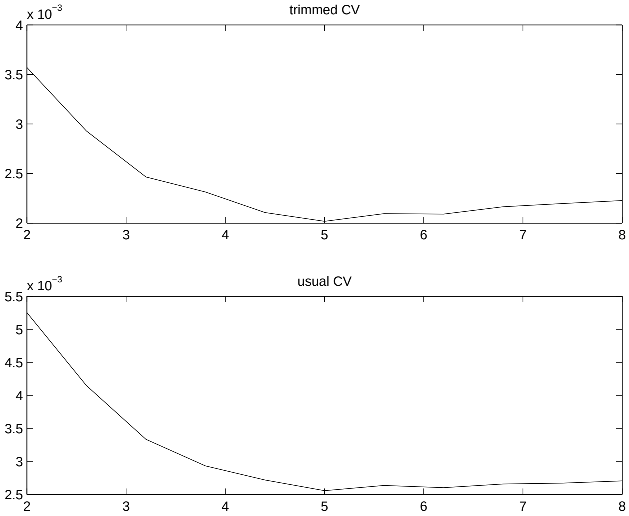

Figure 7: *Dairy data set: CV criterion. Computing time for the grid search over 11 values of h_{base}, 1042 seconds on a Pentium III, 450 Mghz machine.*

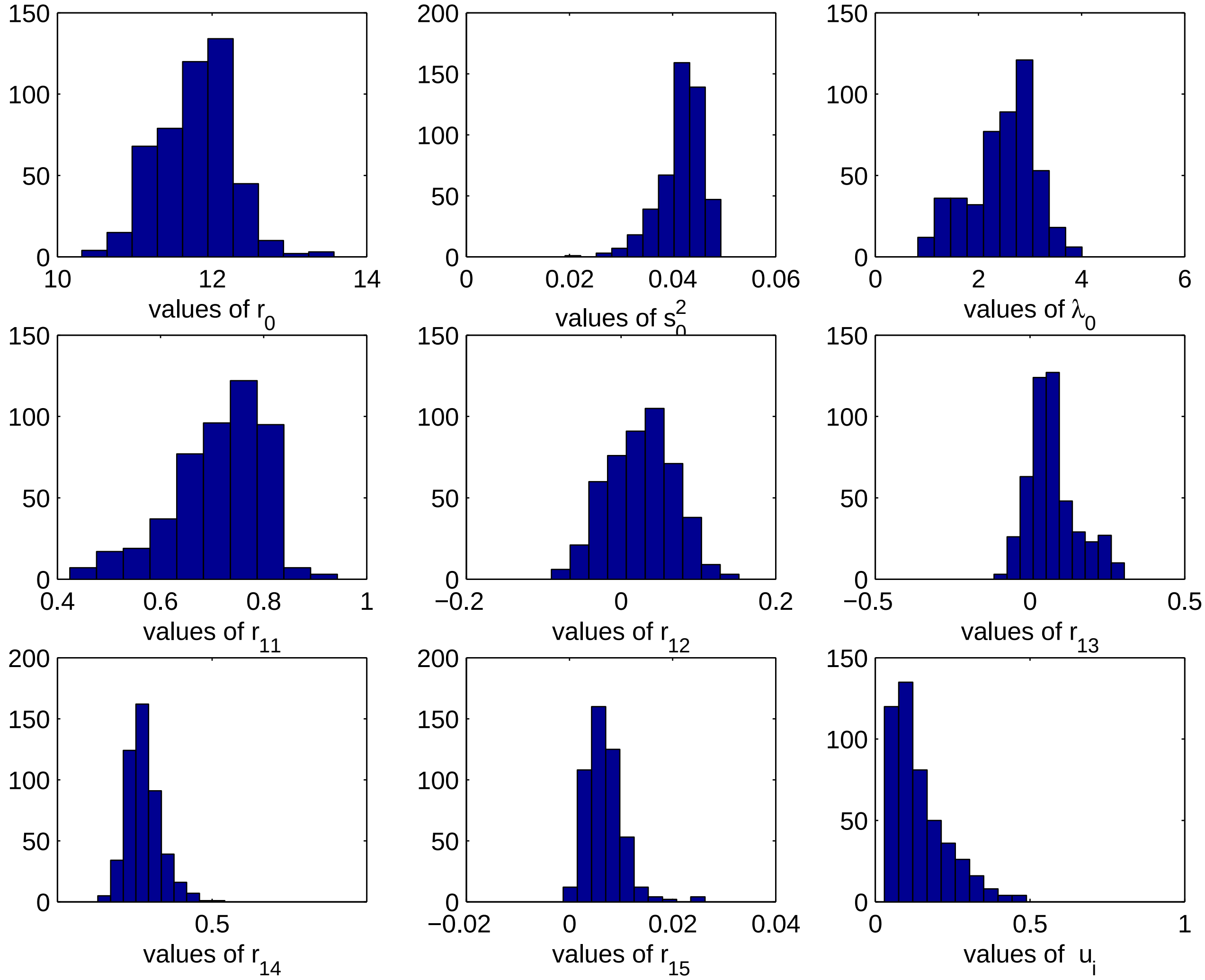

Figure 8: *Distribution of localized estimates over the 480 observed points. Computing time for all the computations, 3792 seconds on a Pentium III, 450 Mghz machine.*

X_1	X_2	Y	True front	Fitted front	MLE front	True u_i	$\hat{u}_i$	$\hat{u}_i^{MLE}$
0.3075	1.6208	7.1913	8.5491	8.6041	8.6337	1.0663	1.3686	1.4017
0.4619	1.3645	4.4324	8.1909	8.2318	8.2567	3.1721	3.6714	3.7096
0.8298	1.2496	4.9784	8.3290	8.1947	8.1834	2.7055	3.1157	3.1088
0.3001	0.7362	4.6990	6.7725	7.1242	7.1592	2.7503	2.3410	2.3864
0.7096	0.4056	5.8958	6.5207	6.7320	6.7402	1.0989	0.8618	0.8640
0.0746	0.5501	4.8044	6.1748	6.7434	6.7773	1.2518	1.8690	1.9138
0.7502	0.9631	3.1629	7.6764	7.6785	7.6811	5.0428	4.3723	4.3827
0.4689	0.8004	4.9061	7.0697	7.2923	7.3202	2.3832	2.3060	2.3416
0.9449	1.4046	8.3640	8.7541	8.5097	8.4782	1.1690	0.4489	0.4225
0.0140	0.6427	5.4701	6.2994	6.8784	6.9120	0.1809	1.3615	1.4012
0.0542	0.0969	5.2896	5.2480	5.9928	6.0166	0.5422	0.7614	0.7730
0.7384	1.5550	7.1718	8.8484	8.6676	8.6623	0.8001	1.4510	1.4477
0.6356	1.1068	5.3290	7.8492	7.8702	7.8835	2.0604	2.4585	2.4779
0.9310	0.1355	5.5749	6.2020	6.3777	6.3617	1.1841	0.8342	0.8185
0.4610	0.9924	6.6433	7.4457	7.6090	7.6371	0.8009	0.9679	0.9888
0.5129	1.6722	8.0822	8.8573	8.7693	8.7852	0.9395	0.7537	0.7552
0.4245	0.9219	6.0327	7.2682	7.4777	7.5081	0.7135	1.3997	1.4332
0.4159	1.5601	5.1533	8.5360	8.5424	8.5674	4.1503	3.2740	3.3118
0.3882	0.9835	7.1130	7.3552	7.5669	7.5990	0.1288	0.6037	0.6097
0.8555	1.3914	9.0144	8.6384	8.4453	8.4277	0.6221	0.2480	0.2320
0.8230	1.9701	2.8772	9.7633	9.4082	9.3803	6.1351	6.3280	6.3081
0.9139	0.7063	4.7090	7.3266	7.3218	7.3062	1.6680	2.5330	2.5194
0.8360	1.2440	5.4475	8.3240	8.1881	8.1760	2.3114	2.6550	2.6467
0.8681	1.0318	4.8018	7.9317	7.8462	7.8332	3.0942	2.9501	2.9404
0.7756	1.6301	7.7465	9.0357	8.8111	8.7992	1.6096	1.0540	1.0403

Table 1: *Example 3: Table of some fitted values*

Parameter	Estimates	Std error
Const.	5.7639	0.1237
Cows	0.7274	0.0336
Land	0.0055	0.0197
Labor	0.0529	0.0329
Conc	0.3506	0.0160
Trend	0.0063	0.0036
σ^2	0.0422	0.0049
λ	2.2792	0.4323

Table 2: *Parametric MLE for Dairy data*

Units	Cows	Land	Labor	Conc	Trend	Output	Fit. front.	$\hat{u}_i$	$\widehat{eff}_i$
126	3.1864	2.5649	0.4055	11.0940	6.0	11.9694	12.0506	0.0938	0.9105
18	3.1739	2.4849	0.4055	11.2020	6.0	12.1741	12.0862	0.0401	0.9607
333	2.7600	2.0541	0.4055	9.7962	3.0	11.0250	11.2389	0.1918	0.8255
347	3.1311	2.7726	0.0000	10.7722	5.0	11.8641	11.8656	0.0617	0.9402
54	1.8405	2.3026	0.4055	9.4969	6.0	10.0470	10.3164	0.2231	0.8000
355	2.8273	2.8034	0.4055	9.9478	1.0	11.2372	11.3386	0.1039	0.9013
323	2.6101	1.9459	0.0000	9.9372	5.0	11.0377	11.1660	0.1193	0.8876
288	2.4248	2.5257	0.0000	9.8069	6.0	11.0120	11.0208	0.0562	0.9453
94	3.1398	2.1972	0.4055	11.3204	4.0	11.9653	12.0919	0.1191	0.8877
222	2.6878	2.0794	0.4055	10.0044	6.0	11.1571	11.2890	0.1225	0.8847
346	3.1135	2.7726	0.0000	10.7179	4.0	11.7776	11.8296	0.0798	0.9233
327	2.7279	2.0794	0.6931	9.9072	3.0	11.1376	11.2458	0.1041	0.9012
203	3.0445	1.8871	0.4055	10.9028	5.0	11.7769	11.8274	0.0678	0.9344
22	3.2149	2.1748	0.4055	11.1983	4.0	11.8308	12.0941	0.2369	0.7890
257	2.8848	2.4849	0.4055	9.9955	5.0	11.3363	11.4581	0.1159	0.8906
178	3.2696	2.3026	0.4055	11.3728	4.0	12.0477	12.2067	0.1436	0.8662
134	3.4965	2.6101	0.4055	11.6542	2.0	12.4795	12.4608	0.0563	0.9452
105	2.7408	2.1401	0.4055	10.4996	3.0	11.3687	11.4810	0.1082	0.8974
424	2.9285	2.9957	0.4055	9.3154	4.0	11.2699	11.3178	0.0762	0.9266
251	3.2108	2.9178	0.4055	11.1984	5.0	12.2027	12.0863	0.0476	0.9535
469	2.0669	2.2513	0.0000	9.8576	1.0	10.5367	10.6983	0.1408	0.8687
4	3.1739	2.1748	0.4055	11.5400	4.0	12.1517	12.2001	0.0691	0.9332
382	3.4372	2.7081	0.6931	11.2708	4.0	12.1887	12.3008	0.1079	0.8977
238	3.8067	2.7726	0.6931	11.8309	4.0	12.5864	12.7306	0.1228	0.8845
395	3.7062	3.4012	0.6931	11.3640	5.0	12.3323	12.4333	0.0890	0.9148

Table 3: *Dairy data: Table of some fitted values*