Keskustelualoitteita #46
Joensuun yliopisto, Taloustieteet

Stochastic Nonparametric Envelopment of Data: Cross-sectional frontier estimation subject to shape constraints

Timo Kuosmanen
and
Mika Kortelainen

ISBN 978-952-458-957-4
ISSN 1795-7885
no 46

Stochastic Nonparametric Envelopment of Data:

Cross-sectional frontier estimation subject to shape constraints

Timo Kuosmanen[1] and Mika Kortelainen[2]

1) Economic Research Unit, MTT Agrifood Research Finland, Luutnantintie 13, 00410 Helsinki, Finland. Tel. +358 9 5608 6309. Fax: + 358 9 5608 6264. E-mail. Timo.Kuosmanen@mtt.fi.

2) Department of Economics and Business Administration, University of Joensuu, P.O.B. 111, 80101 Joensuu, Finland; Tel. +358 13 251 4534; Fax. +358 13 251 3290; E-mail. mika.kortelainen@joensuu.fi

Abstract

The field of production frontier estimation is divided between the parametric Stochastic Frontier Analysis (SFA) and the deterministic, nonparametric Data Envelopment Analysis (DEA). This paper explores an amalgam of DEA and SFA that melds a nonparametric frontier with a stochastic composite error. Our model imposes the standard SFA assumptions for the inefficiency and noise terms. The frontier is estimated nonparametrically, imposing monotonicity and convexity as in DEA. For estimation, we propose two alternative methods based on shape constrained nonparametric least squares. The performance of the proposed estimation techniques is examined using Monte Carlo simulations and an illustrative application.

Key Words: nonparametric least squares, method of moments, productive efficiency, pseudolikelihood, stochastic frontier analysis (SFA), data envelopment analysis (DEA)

JEL Classification: C14, C51, D24

1. Introduction

The literature of production frontier estimation has been dominated by two approaches: the nonparametric data envelopment analysis (DEA: Farrell, 1957; Charnes et al., 1978) and the parametric stochastic frontier analysis (SFA: Aigner et al., 1977; Meeusen and van den Broeck, 1977). The main appeal of DEA lies in its nonparametric treatment of the frontier, which does not assume a particular functional form but relies on the general regularity properties such as monotonicity, convexity, and homogeneity. However, DEA attributes all deviations from the frontier to inefficiency, and completely ignores any stochastic noise in the data. The key advantage of SFA is its stochastic treatment of residuals, decomposed into a non-negative inefficiency term and an idiosyncratic error term that accounts for measurement errors and other random noise. However, SFA builds on the parametric regression techniques, which require an ex ante specification of the functional form. Since the economic theory rarely justifies a particular functional form, the flexible functional forms, such as the translog or generalized McFadden (Christensen et al., 1973; Diewert and Wales, 1987), are frequently used in the SFA literature. The problem with the flexible functional forms is that the estimated frontiers often violate the monotonicity, concavity/convexity and homogeneity conditions. On the other hand, imposing these regularity conditions will sacrifice the flexibility (see e.g. Christensen and Caves, 1980; Diewert and Wales, 1987; Sauer, 2006). In summary, it is generally accepted that the virtues of DEA lie in its general, nonparametric treatment of the frontier, while the virtues of SFA lie in its stochastic, probabilistic treatment of inefficiency and noise (e.g., Bauer, 1990; Seiford and Thrall, 1990).

To bridge the gap between SFA and DEA, a large and growing number of stochastic semi- or nonparametric frontier models have been developed (e.g., Park and Simar, 1994; Fan et al., 1996; Kneip and Simar, 1996; Park et al., 1998, 2003, 2006; Post et al., 2002; Griffin and Steel, 2004; Henderson and Simar, 2005; Kuosmanen et al., 2007; Kumbhakar et al., 2007). While these studies come a long way in combining some of the virtues of DEA and SFA, the conceptual link between the parametric and non-parametric branches is still missing: none of these techniques can be viewed as a stochastic extension of DEA in the same way as SFA extends the classic deterministic econometric frontier models by Aigner and Chu (1968), Timmer (1971), Richmond (1974), and others. Furthermore, while the assumptions required by the previous semi- and nonparametric SFA models are relatively weak (c.f., e.g., Henderson and Simar, 2005; Kumbhakar et al., 2007), there is no guarantee that these models satisfy the regularity conditions implied by the economic theory. Therefore, there is an evident need for semi- and nonparametric stochastic frontier approaches that could satisfy the regularity properties and thus combine the virtues of DEA and SFA in a unified framework of frontier estimation.

This paper develops an amalgam of DEA and SFA, which combines a DEA-style nonparametric, piecewise linear frontier with a SFA-style composite error term consisting of noise and inefficiency components. Distributions of the noise and inefficiency terms are assumed to be of a known form, similar to the traditional SFA, but no particular functional form for the production frontier is assumed. Rather, the frontier is only required to satisfy the global monotonicity and concavity properties, similar to DEA. In essence, this model combines the key characteristics of both DEA and SFA in the same framework, closing the gap between the parametric and nonparametric approaches. Such a unifying model deserves a name, so we will henceforth refer to this amalgam model as *stochastic nonparametric envelopment of data* (StoNED).[1]

Banker and Maindiratta (1992) (henceforth BM) were the first to propose such an amalgam model and explore its practical estimation. Unfortunately, this elegant paper has not attracted deserved attention. Presumably, this is largely due to the lack of an operational estimation procedure. BM pursued constrained maximum likelihood (ML) estimation of their model, but in practice, the resulting ML problem is extremely difficult (if not impossible) to solve. Therefore, one of our main contributions is to develop an operational estimation method for the StoNED model.

Our estimation method consists of two stages. In the first stage, we estimate the average production function by nonparametric least squares (NLS) subject to monotonicity and concavity constraints (Hildreth, 1954; Hanson and Pledger, 1976; Mammen, 1991; Groeneboom et al., 2001). NLS provides an unbiased, consistent estimator for the shape of the production frontier. However, it underestimates the true frontier due to the inefficiency term. Therefore, in the second stage we estimate the conditional expected value of the inefficiency term, and correct the NLS estimates. We show how the conditional expected values can be estimated by using the method of moments or pseudolikelihood techniques. We examine the performance of both these methods by means of Monte Carlo simulations.

The remainder of the paper is organized as follows. Section 2 introduces the StoNED model as a generalization of DEA and SFA. Section 3 discusses its estimation by means of constrained maximum likelihood along the lines of BM. Section 4 describes the estimation of the average production function by means of NLS. Based on the NLS residuals, we estimate the inefficiency and noise terms by means of method of moments or pseudolikelihood techniques in Section 5. Section 6 discusses some useful extensions to estimation of cost functions and modeling alternative assumptions about the returns to scale. Section 7 examines how the proposed techniques perform in a controlled environment of Monte Carlo simulations. Section 8 presents an illustrative application using data on tax collection offices. Section 9 draws the concluding remarks.

[1] The term "nonparametric" refers here specifically to the fully nonparametric treatment of the frontier. As a whole, the cross-sectional model discussed in this paper could be more appropriately described as "semiparametric" due to the distributional assumptions about the stochastic components. These parametric assumptions can be avoided in the panel data setting (see the working paper Kuosmanen, 2006, for discussion).

2. Stochastic nonparametric envelopment of data (StoNED)

This section formally introduces the StoNED model in the cross-sectional, multi-input single-output setting. The *m*-dimensional input vector is denoted by **x** and the scalar output by *y*. The production technology is represented by the *production function* $y = f(\mathbf{x})$. We assume that function *f* belongs to the class of monotonic increasing and concave functions, denoted by F_2. In contrast to the SFA literature, no specific functional form for *f* is assumed a priori; our specification of the production function proceeds along the nonparametric lines of the DEA literature.

The observed output y_i of firm *i* may differ from $f(\mathbf{x}_i)$ due to inefficiency and noise. We follow the SFA literature and introduce a composite error term $\varepsilon_i = v_i - u_i$, which consists of the inefficiency term $u_i > 0$ and the idiosyncratic error term v_i, formally,

$$y_i = f(\mathbf{x}_i) + \varepsilon_i = f(\mathbf{x}_i) - u_i + v_i,\ \ i = 1,...,n. \tag{1}$$

Terms u_i and v_i ($i = 1,...,n$) are assumed to be statistically independent of each other as well as of inputs $\mathbf{x}_i$. Furthermore, we follow the standard SFA practice and assume $u_i \underset{i.i.d}{\sim} |N(0,\sigma_u^2)|$ and $v_i \underset{i.i.d}{\sim} N(0,\sigma_v^2)$. Other distributions such as gamma or exponential are also used for the inefficiency term u_i (e.g. Kumbhakar and Lovell, 2000), but this paper focuses on the standard half-normal specification.

Model (1) is referred to as the *stochastic nonparametric envelopment of data* (StoNED) model. It can be thought as a generalization of the classic SFA and DEA. Specifically, if *f* is restricted to some specific functional form (instead of the class F_2), model (1) boils down to the SFA model by Aigner et al. (1977). On the other hand, if we impose the restriction $\sigma_v^2 = 0$ and relax the distributional assumption concerning the inefficiency term, we obtain the DEA model by Banker et al. (1984). In this sense, both SFA and DEA can be seen as special cases of the more general StoNED framework.

3. Constrained maximum likelihood estimation

Banker and Maindiratta (1992) (henceforth BM) considered a multiplicative variant of model (1), and proposed to estimate it by constrained maximum likelihood (ML) method. While BM assumed the distribution of the inefficiency term to be truncated normal, we here rephrase their approach in terms of the more standard half-normal specification.[2] Using the parametrization by Aigner et al. (1977) with $\sigma^2 \equiv \sigma_u^2 + \sigma_v^2$ and $\lambda \equiv \sigma_u / \sigma_v$, the log-likelihood function can we written in terms of the unobserved

[2] Truncated normal specification (or any other distributional assumption) for u_i only influences the log-likelihood function. It is straightforward to adapt the model to these situations.

frontier outputs $y_i^f \equiv f(\mathbf{x}_i)$ as

$$\ln L(\mathbf{y}^f, \sigma^2, \lambda) = \frac{n}{2}\ln(2/\pi) - n\ln\sigma + \sum_{i=1}^{n}\ln\Phi\left[\frac{-\left(y_i - y_i^f\right)\lambda}{\sigma}\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - y_i^f\right)^2, \qquad (2)$$

where Φ denotes the cumulative distribution function of the standard normal distribution. Given the regularity condition $f \in F_2$, the ML estimates of the frontier outputs $\hat{y}_i^f$ and the density function parameters $\hat{\sigma}$ and $\hat{\lambda}$ are obtained as the optimal solutions to the following constrained ML problem

$$\max_{\substack{y_1^f,\ldots,y_n^f \\ \sigma,\lambda}} \left\{ \ln L(\mathbf{y}^f, \sigma^2, \lambda) \;\middle|\; y_i^f = f(\mathbf{x}_i)\ \forall i;\ f \in F_2;\ \mathbf{y}^f, \sigma, \lambda \geq 0 \right\}. \qquad (3)$$

This is a complex, infinite dimensional optimization problem. An important contribution of BM was to transform (3) into a finite dimensional optimization problem by applying insights from Afriat's Theorem (see Afriat, 1967, 1972; Hanoch and Rotchild, 1972; and Diewert and Parkan, 1983). Their main result was to show that the constraints of (3) can be linearized, and the constrained ML problem can be equivalently written as

$$\begin{aligned} &\max_{\mathbf{y}^f, \boldsymbol{\beta}, \sigma, \lambda} \ \frac{n}{2}\ln(2/\pi) - n\ln\sigma + \sum_{i=1}^{n}\ln\Phi\left[\frac{-\left(y_i - y_i^f\right)\lambda}{\sigma}\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - y_i^f\right)^2 \\ &y_i^f - \boldsymbol{\beta}_i'\mathbf{x}_i \geq y_j^f - \boldsymbol{\beta}_i'\mathbf{x}_j \ \ \forall i, j = 1,\ldots,n \\ &\boldsymbol{\beta}_i \geq \mathbf{0} \ \forall i = 1,\ldots,n, \\ &y_i^f \geq 0 \ \forall i = 1,\ldots,n; \ \ \sigma, \lambda \geq 0. \end{aligned} \qquad (4)$$

BM did not present any asymptotic results for their method. Later, Sarath and Maindiratta (1997) proved consistency of the estimators for frontier outputs and the composed error density function parameters under relatively weak assumptions.

Unfortunately, the objective function of (4) remains a complex, nonlinear, non-convex function. In particular, the sum of the logarithms of the cumulative normal density in *n* different points is hard to compute because there is no closed form solution to the definite integrals of Φ. BM suggested various solution strategies including grid search and cutting plane algorithms, but no practical computational procedure is available (see also Banker et al., 2002; and Allon et al., 2007, for critical discussion). To our knowledge, there are no reported empirical applications of this ML method.

4. Constrained nonparametric least squares estimation

In this section we develop a least squares estimator for the shape of the production frontier. The challenge in the least squares estimation of model (1) is that the expected value of the composite error term is greater than zero. Given the half-normal specification of the inefficiency term, Aigner et al. (1977)

showed that

$$E(\varepsilon_i) = E(u_i) = \sigma_u \sqrt{2/\pi} > 0 \,. \quad (5)$$

This implies that the StoNED model (1) violates the Gauss-Markov assumptions and hence the least squares estimators are biased and inconsistent. However, the Gauss-Markov properties can be restored by rephrasing the model as

$$y_i = \left[f(\mathbf{x}_i) - \mu\right] + \left[\varepsilon_i + \mu\right] = g(\mathbf{x}_i) + \upsilon_i, \;\; i = 1,...,n \,, \quad (6)$$

where $\mu \equiv E(u_i)$ is the expected inefficiency and $g(\mathbf{x}) \equiv f(\mathbf{x}) - \mu$ can be interpreted as an "average" production function (in contrast to the "frontier" production function *f*), and $\upsilon_i \equiv \varepsilon_i + \mu, \;\; i = 1,...,n$, is a modified composite error term. It is easy to verify that the modified errors υ_i satisfy the Gauss-Markov conditions under the maintained assumptions of the StoNED model. Thus, the average production function *g* can be consistently estimated by nonparametric regression techniques. Subsequently, the expected value μ and the parameters of the inefficiency and noise distributions can be estimated based on the regression residuals by the method of moments or pseudolikelihood techniques (as discussed in Section 5).

To estimate the average production function *g*, we employ nonparametric least squares (NLS) subject to monotonicity and concavity restrictions (Hildreth, 1954; Hanson and Pledger, 1976; Mammen, 1991; Groeneboom et al., 2001). The constrained NLS estimator is particularly suited for the estimation of the StoNED model because it draws its power from the monotonicity and concavity conditions (which are the maintained assumptions of both StoNED and DEA models) without any further assumptions about the functional form or its smoothness. This approach avoids the bias-variance tradeoff associated with other nonparametric regression techniques (such as kernel or spline techniques) (e.g., Yatchew 2003).

The shape constrained NLS problem can be stated as

$$\min_g \sum_{i=1}^{n} (y_i - g(\mathbf{x}_i))^2 \;\; s.t. \; g \in F_2 \,. \quad (7)$$

In words, the NLS estimator of *g* is a monotonic increasing and concave function that minimizes the L_2-norm of the residuals. The maximum likelihood property of this estimator was noted already by Hildreth (1954). Hanson and Pledger (1976) proved consistency of estimator (7) in the single regression case. Nemirovskii et al. (1985), Mammen (1991) and Mammen and Thomas-Agnen (1999) have established convergence rates and Groeneboom et al. (2001) derived the asymptotic distribution at a fixed point. In the case of *m* inputs, the NLS estimator (7) achieves the standard nonparametric rate of convergence $O_P(n^{-1/(2+m)})$. If one imposes further smoothness assumption by postulating that *g* belongs to some constrained Sobolev functional class, the optimal convergence rate of nonparametric estimator in the sense of Stone (1980, 1982) can be achieved (see e.g. Mammen and Thomas-Agnen, 1999; and

Yatchew, 2003). While introducing more stringent smoothness assumptions can improve the rate of convergence, the economic theory does not provide guidance regarding the appropriate degree of smoothness (or bounds of derivatives). More importantly, imposing further smoothness assumptions spoils the connection to DEA. Therefore, we here restrict to the non-smooth NLS.

The NLS problem (7) does not restrict beforehand to any particular functional form of *g*, but searches the best-fit function from the family F_2, which includes an infinite number of functions. This makes problem (7) generally hard to solve. In statistics, efficient algorithms for solving problem (3) in the single regressor (i.e., single input) case have been developed (e.g., Fraser and Massam, 1989; Meyer, 1999). These algorithms require that the data is sorted in ascending order according to the regressor. However, such a sorting trick is not possible in the general multiple regression (i.e., multi-input) setting where **x** is a vector rather than scalar.

To estimate the NLS problem (7) in the general multi-input setting, we utilize the insights from Afriat's Theorem, in line with BM and Matzkin (1994). Specifically, we take the constraints of the ML problem (3) by BM and form the following least-squares problem

$$\begin{aligned} &\min_{\mathbf{y}^g,\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - y_i^g \right)^2 \\ &y_i^g - \boldsymbol{\beta}_i'\mathbf{x}_i \geq y_j^g - \boldsymbol{\beta}_i'\mathbf{x}_j \;\; \forall i,j = 1,...,n \\ &\boldsymbol{\beta}_i \geq \mathbf{0} \;\; \forall i = 1,...,n. \end{aligned} \tag{8}$$

This gives a quadratic programming (QP) problem with $n(m+1)$ unknowns and n^2+n linear inequalities, which is relatively easy to solve by standard QP algorithms and solver software.[3] Interestingly, this finite QP problem is equivalent to the infinite dimensional NLS problem (7) in the following sense:

Proposition 1: Let s_{NLS}^2 be the minimum sum of squares of problem (7) and let s_{Afriat}^2 be the minimum sum of squares of problem (8). Then for any real-valued data, $s_{NLS}^2 = s_{Afriat}^2$.

This equivalence result is an important step towards operationalizing NLS in the multiple regression setting. While the use of Afriat inequalities to model concavity constraints in NLS has been briefly suggested earlier (e.g., Matzkin, 1994; Yatchew, 1998), to our knowledge, the equivalence of the infinite dimensional NLS problem and a finite QP problem has not been formally proven before.

[3] QP is a standard class of problems within nonlinear programming (NLP). A variety of commercial and shareware solver software are available for solving QP problems. High-performance QP solvers include, e.g., CPLEX, LINDO, MOSEK, and QPOPT, but also general NLP solvers such as MINOS and BQPD can handle QP problems. Most solvers can be integrated with standard mathematical modeling systems/languages such as GAMS, Gauss, Mathematica, and Matlab.

Proposition 1 implies that the QP problem (8) yields unbiased and consistent fits $\hat{y}_i^g$ in the observed points $\mathbf{x}_i, i = 1,...,n$. To estimate a full-fledged production function throughout the observed range of input values, including the unobserved points, the following model proves convenient:

$$\begin{aligned}
&\min_{\boldsymbol{\upsilon},\boldsymbol{\alpha},\boldsymbol{\beta}} \sum_{i=1}^{n} \upsilon_i^2 \\
&y_i = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i + \upsilon_i \\
&\alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i \le \alpha_h + \boldsymbol{\beta}_h'\mathbf{x}_i \;\; \forall h,i = 1,...,n \\
&\boldsymbol{\beta}_i \ge \mathbf{0} \;\; \forall i = 1,...,n.
\end{aligned} \tag{9}$$

The first constraint of this problem can be interpreted as a standard regression equation, the second constraint enforces concavity analogous to the Afriat inequalities, and the third constraint ensures monotonicity. The analogy of model (9) with the conventional parametric regression models is useful for econometric model building (e.g., we exploit it in Section 6.3 for estimating cost functions). Note that (9) differs from the classic OLS problem in that the coefficients $\alpha_i, \boldsymbol{\beta}_i$ are here firm-specific. In this respect, model (9) is structurally similar to the varying coefficient (VC) regression models (also referred to as random parameters models) (e.g., Fan and Zhang, 1999; Greene, 2005), which typically assume a conditional linear structure. However, while the random parameters models estimate *n* different production functions of the same a priori specified functional form, the NLS regression (9) estimates *n* tangent hyper-planes to one unspecified production function. The slope coefficients $\boldsymbol{\beta}_i$ represent the marginal products of inputs (i.e., the sub-gradients $\nabla \hat{g}_i(\mathbf{x})$). Interestingly, problems (8) and (9) are equivalent in the following sense:

Proposition 2: Let s_{VC}^2 be the minimum sum of squares of problem (9). For any real-valued data, $s_{Afriat}^2 = s_{VC}^2$.

Given the estimated coefficients $\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_i$ from model (9), we may estimate the average production function *g* by the following piece-wise linear function

$$\hat{g}(\mathbf{x}) \equiv \min_{i \in \{1,...,n\}} (\hat{\alpha}_i + \hat{\boldsymbol{\beta}}_i'\mathbf{x}). \tag{10}$$

This function provides estimates for function *g* in unobserved points **x**. In addition, one can use this function for computing substitution and scale elasticities. This piece-wise linear estimator is legitimized by the following result.

Proposition 3: Denote the set of functions that minimize problem (7) by $G_2^* : G_2^* \subset F_2$. For any real-valued data, $\hat{g} \in G_2^*$.

The piece-wise linear structure of the estimator (10) closely resembles that of the DEA frontier. Although problem (9) includes *n* different firm-specific coefficients α_i ,$\boldsymbol{\beta}_i$, the number of different hyperplane segments in $\hat{g}(\mathbf{x})$ is typically much lower than *n* (for graphical illustration, see Figures 1-3 in Section 7). Second similarity with DEA is that the estimator $\hat{g}(\mathbf{x})$ and its coefficients $\hat{\alpha}_i$,$\hat{\boldsymbol{\beta}}_i$ are not necessarily unique. This is because $\hat{g}(\mathbf{x})$ depends on the particular choice of subgradients $\nabla\hat{g}(\mathbf{x})$ that are represented by the slope coefficients $\hat{\boldsymbol{\beta}}_i$. To test for uniqueness, one could construct upper and lower bounds for function $\hat{g}(\mathbf{x})$ along the lines of Varian (1984).

Despite these links to DEA, the piece-wise linear function $\hat{g}(\mathbf{x})$ does not estimate the frontier, but the average production function *g*(**x**). Estimation of *g* is a common approach both in parametric and semiparametric panel data stochastic frontier models (e.g., Schmidt and Sickles, 1984; Cornwell et al., 1990; Kneip and Simar, 1996). However, some authors claim that the technology in use at the average production frontier can differ from that of the best practice frontier (see, e.g., Greene, 1997, for discussion). Yet, no convincing theory supports such an argument. In the present framework (which is standard in the literature), the shape of the frontier *f*(**x**) must be exactly the same as that of the average production function *g*(**x**) because $g(\mathbf{x}) \equiv f(\mathbf{x}) - \mu$, where the expected inefficiency μ is a constant. To estimate $f(\mathbf{x})$ consistently, we next need to estimate μ , and then shift the estimated $\hat{g}(\mathbf{x})$ upward, similarly to the modified OLS (MOLS) approach in the SFA literature. In next section we show how the expected inefficiency μ and the unknown standard deviations σ_u, σ_v can be estimated from the NLS residuals by the method of moments or pseudolikelihood techniques.

5. Efficiency estimation

Given the NLS residuals $\hat{\mathbf{\upsilon}} \equiv (\hat{\upsilon}_1,...,\hat{\upsilon}_n)$, the next challenge is to disentangle inefficiency from noise. This can be done by method of moments or pseudolikelihood techniques.

5.1. Method of moments

The method of moments (MM) is widely used in the SFA literature (referred to as MOLS, e.g. Greene 1997). Under the maintained assumptions of half-normal inefficiency and normal noise, the second and third central moments of the composite error distribution are given by

$$M_2 = \left[\frac{\pi - 2}{\pi}\right]\sigma_u^2 + \sigma_v^2 \tag{11}$$

$$M_3 = \left(\sqrt{\frac{2}{\pi}}\right)\left[1 - \frac{4}{\pi}\right]\sigma_u^3 . \tag{12}$$

These can be estimated based on the distribution of the NLS residuals as

$$\hat{M}_2 = \sum_{i=1}^{n} (\hat{\upsilon}_i - \hat{E}(\upsilon_i))^2 / n \tag{13}$$

$$\hat{M}_3 = \sum_{i=1}^{n} (\hat{\upsilon}_i - \hat{E}(\upsilon_i))^3 / n . \tag{14}$$

Note that the third moment (which represents the skewness of the distribution) only depends on the standard deviation parameter σ_u of the inefficiency distribution. Thus, given the estimated $\hat{M}_3$ (which should be negative), we can estimate σ_u parameter by

$$\hat{\sigma}_u = \sqrt[3]{\frac{\hat{M}_3}{\left(\sqrt{\frac{2}{\pi}}\right)\left[1 - \frac{4}{\pi}\right]}} . \tag{15}$$

Subsequently, the standard deviation of the error term σ_v is estimated based on (11) as

$$\hat{\sigma}_v = \sqrt{\hat{M}_2 - \left[\frac{\pi - 2}{\pi}\right]\hat{\sigma}_u^2} . \tag{16}$$

These MM estimators are unbiased and consistent (Aigner et al., 1977; Greene, 1997), but not necessarily as efficient as the maximum likelihood estimators.

5.2. Pseudolikelihood estimation

An alternative way to estimate the standard deviations σ_u, σ_v is to apply the pseudolikelihood (PSL) method suggested by Fan et al. (1996).[4] Compared to the MM, PSL is potentially more efficient, but is computationally and conceptually somewhat more demanding.

Like in the MM approach, our starting point is the NLS residuals $\hat{\mathbf{U}} \equiv (\hat{\upsilon}_1,...,\hat{\upsilon}_n)$. In the PSL approach we set parameters $\sigma \equiv \sigma_u + \sigma_v$ and $\lambda \equiv \sigma_u / \sigma_v$ to maximize the concentrated log-likelihood function. One of the main contributions of Fan et al. (1996) was to show that the log-likelihood can be

[4] Fan et al. (1996) estimated a variant of the StoNED using a two-stage method: in stage 1) the average production function *g* is estimated using the kernel regression, and in stage 2) the standard deviations σ_u, σ_v are estimated by PSL. While the kernel regression yields a consistent estimator of *g*, the estimated frontier may violate monotonicity and concavity. Our approach deviates from Fan et al. in the first stage where we estimate *g* by using shape restricted NLS instead of kernel techniques.

expressed as a function of a single parameter (λ) as,

$$\ln L(\lambda) = -n \ln \hat{\sigma} + \sum_{i=1}^{n} \ln \Phi \left[\frac{-\hat{\varepsilon}_i \lambda}{\hat{\sigma}} \right] - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 , \tag{17}$$

where

$$\hat{\varepsilon}_i = \hat{\upsilon}_i - \left(\sqrt{2}\lambda\hat{\sigma}\right) \Big/ \left[\pi\left(1+\lambda^2\right) \right]^{1/2} , \tag{18}$$

and

$$\hat{\sigma} = \left\{ \frac{1}{n} \sum_{j=1}^{n} \hat{\upsilon}_i^2 \Big/ \left[1 - \frac{2\lambda^2}{\pi\left(1+\lambda\right)} \right] \right\}^{1/2} . \tag{19}$$

Note that $\hat{\varepsilon}_i$ and $\hat{\sigma}$ cannot be computed from the NLS residuals as they depend on the unknown parameter λ. In practice, we maximize the log-likelihood function (17) by enumerating over λ values, using a simple grid search or more sophisticated search algorithms. After the pseudolikelihood estimate $\hat{\lambda}$ that maximizes (17) is found, estimates for ε_i and σ are obtained from (18) and (19). Subsequently, we obtain $\hat{\sigma}_u = \hat{\sigma}\hat{\lambda}/(1+\hat{\lambda})$ and $\hat{\sigma}_v = \hat{\sigma}/(1+\hat{\lambda})$. Regarding convergence, Fan et al. (1996) note that estimators $\hat{\lambda}$ and $\hat{\sigma}$ converge to the true λ and σ at parametric rate $n^{-1/2}$.

5.3. Estimation of the inefficiency term

Given a consistent estimator $\hat{\sigma}_u$ (obtained by either MM or PSL), the frontier production function *f* can be consistently estimated by

$$\hat{f}(\mathbf{x}_i) = \hat{g}(\mathbf{x}_i) + \hat{\sigma}_u \sqrt{2/\pi} . \tag{20}$$

In practice, this means that frontier is obtained by shifting the NLS estimate of the average production function upwards by the expected value of the inefficiency term, analogous to the MOLS approach.

Regardless of how σ_u, σ_v are estimated, the firm-specific inefficiency component u_i must be inferred indirectly in the cross-sectional setting. Jondrow et al. (1982) have shown that the conditional distribution of inefficiency u_i given ε_i is a zero-truncated normal distribution with mean $\mu_* = -\varepsilon_i \sigma_u^2 / (\sigma_u^2 + \sigma_v^2)$ and variance $\sigma_*^2 = \sigma_u^2 \sigma_v^2 / (\sigma_u^2 + \sigma_v^2)$. As a point estimator for u_i, one can use the conditional mean

$$E(u_i | \varepsilon_i) = \mu_* + \sigma_* \left[\frac{\phi(-\mu_* / \sigma_*)}{1 - \Phi(-\mu_* / \sigma_*)} \right], \tag{21}$$

where ϕ is the standard normal density function, and Φ is the standard normal cumulative distribution function. Given the estimated $\hat{\sigma}_u, \hat{\sigma}_v$ parameters, the conditional expected value of inefficiency can be computed as

$$\hat{E}(u_i | \hat{\varepsilon}_i) = -\frac{\hat{\varepsilon}_i \hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_v^2} + \frac{\hat{\sigma}_u^2 \hat{\sigma}_v^2}{\hat{\sigma}_u^2 + \hat{\sigma}_v^2} \left[\frac{\phi(\hat{\varepsilon}_i / \hat{\sigma}_v^2)}{1 - \Phi(\hat{\varepsilon}_i / \hat{\sigma}_v^2)} \right], \tag{22}$$

where $\hat{\varepsilon}_i = \hat{\upsilon}_i - \hat{\sigma}_u \sqrt{2/\pi}$ is the estimator of the composite error term (compare with (18)), not the NLS residual. The conditional expected value (22) is an unbiased but inconsistent estimator of u_i: irrespective of the sample size n, the variance of the estimator does not converge to zero.

5.4. Statistical inference

Even though the log-likelihood function and the statistical distributions of the inefficiency and noise terms are known (by assumption), the conventional methods of statistical inference do not directly apply in the present setting. For example, one might apply the likelihood ratio test for testing significance of two alternative hierarchically nested StoNED models, but the degrees of freedom are difficult to specify (see Meyer, 2003, 2006, for discussion). One could also construct confidence intervals based on the known conditional distribution of the inefficiency term (see Horrace and Schmidt, 1996, for details). However, such confidence intervals do not take into account the sampling distribution of the inefficiency estimators, and consequently, have poor coverage properties (Simar and Wilson, 2005).

In light of these complications, parametric bootstrap appears to be the best suited approach to statistical inference in the present context. Simar and Wilson (2005) have recently developed a bootstrap procedure for SFA, which can be adapted to the context of the StoNED model as follows:

[1] Given the sample data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, solve the NLS problem (9) to obtain estimates $\{(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_i, \hat{\upsilon}_i)\}_{i=1}^n$.

Use the NLS residuals $\{\hat{\upsilon}_i\}_{i=1}^n$ and the MM or PSL methods to obtain estimates $\hat{\sigma}, \hat{\lambda}, \hat{\sigma}_u, \hat{\sigma}_v$.

[2] For $i = 1,\ldots,n$, draw $\tilde{u}_i \sim |N(0, \hat{\sigma}_u^2)|$ and $\tilde{v}_i \sim N(0, \hat{\sigma}_v^2)$, and compute $\tilde{y}_i = \hat{\alpha}_i + \hat{\boldsymbol{\beta}}_i' \mathbf{x}_i - \tilde{u}_i + \tilde{v}_i$.

[3] Using the pseudo-data $\{(\mathbf{x}_i, \tilde{y}_i)\}_{i=1}^n$, solve the NLS problem (9) to obtain estimates $\{(\hat{\hat{\alpha}}_i, \hat{\hat{\boldsymbol{\beta}}}_i, \hat{\hat{\upsilon}}_i)\}_{i=1}^n$.

Use the residuals $\{\hat{\hat{\upsilon}}_i\}_{i=1}^n$ and the MM or PSL methods to obtain bootstrap estimates $\hat{\hat{\sigma}}, \hat{\hat{\lambda}}, \hat{\hat{\sigma}}_u, \hat{\hat{\sigma}}_v$.

[4] Repeat steps [2]-[3] to obtain bootstrap estimates $\left\{\left\{(\hat{\hat{\alpha}}_{ib}, \hat{\hat{\boldsymbol{\beta}}}_{ib}, \hat{\hat{\upsilon}}_{ib})\right\}_{i=1}^n\right\}_{b=1}^B$ and $\left\{\hat{\hat{\sigma}}_b, \hat{\hat{\lambda}}_b, \hat{\hat{\sigma}}_{ub}, \hat{\hat{\sigma}}_{vb}\right\}_{b=1}^B$.

The resulting bootstrap estimates can be used for statistical inference in many ways (see, e.g., Efron, 1979, 1982; and Efron and Tibshirani, 1993, for discussion). For example, the $100 \cdot (1-\alpha)\%$

confidence interval for parameter σ_u is constructed as $\left[\hat{\hat{\sigma}}_u^{[\alpha/2]}, \hat{\hat{\sigma}}_u^{[(1-\alpha)/2]}\right]$ where $\hat{\hat{\sigma}}_u^{[p]}$ denotes the $p \cdot 100$-percentile of the elements of $\left\{\hat{\hat{\sigma}}_{ub}\right\}_{b=1}^{B}$. The confidence interval for the expected inefficiency μ is hence $\left[\hat{\hat{\sigma}}_u^{[\alpha/2]}\sqrt{2/\pi}, \hat{\hat{\sigma}}_u^{[(1-\alpha)/2]}\sqrt{2/\pi}\right]$. Thus, the confidence interval for the production function *f* in point **x** is $\left[\hat{g}(\mathbf{x}) + \hat{\hat{\sigma}}_u^{[\alpha/2]}\sqrt{2/\pi},\ \hat{g}(\mathbf{x}) + \hat{\hat{\sigma}}_u^{[(1-\alpha)/2]}\sqrt{2/\pi}\right]$.

Finally, we note that in empirical applications the least-squares residuals are often skewed in the wrong direction ($\hat{M}_3 > 0$). In the SFA literature, the usual approach is to set $\hat{\sigma}_u = 0$, which means that all firms are diagnosed as efficient. It may also occur that the skewness is so great that $\hat{\sigma}_u > \hat{\sigma}$, and thus $\hat{\sigma}_v$ becomes negative. In that case, the typical approach is to set $\hat{\sigma}_v$ = 0 and attribute all observed variation to inefficiency (as in DEA). The "wrong skewness" is conventionally seen as a built-in diagnostic, which signals model misspecification or inappropriate data (Greene, 1997). However, Simar and Wilson (2005) have shown by means of Monte Carlo simulations that the wrongly skewed residuals can frequently arise even in correctly specified SFA models. Wrongly skewed residuals can also occur in correctly specified StoNED models. This is not only a problem for the method of moments, it equally affects the pseudolikelihood method. Thus, it is comforting to note that Simar and Wilson (2005) have shown that the parametric bootstrap method described above can provide useful information about the inefficiency levels even in such situations where the residuals are wrongly skewed.

6. Extensions

6.1. Returns to scale

We have thus far left returns to scale (RTS) unrestricted. In many applications, it is meaningful to impose further structure on RTS or it is interesting to test for alternative RTS assumptions statistically. Imposing RTS is straightforward in the NLS regression (9). We simply add the following constraints:

- *constant returns to scale* (CRS): $\alpha_i = 0\ \forall i = 1,...,n$
- *non-increasing returns to scale* (NIRS): $\alpha_i \geq 0\ \forall i = 1,...,n$
- *non-decreasing returns to scale* (NDRS): $\alpha_i \leq 0\ \forall i = 1,...,n$

While the NLS regression is easily adapted to alternative RTS assumptions, the implications to the efficiency estimation are somewhat trickier. Specifically, if one estimates the average technology *g* subject to CRS, and subsequently shifts the frontier upward by the expected inefficiency, the resulting best-practice frontier does not generally satisfy CRS. This is due to the mismatch of the additive structure of the inefficiency and noise terms assumed in (1) and the multiplicative nature of the scale properties. If one imposes CRS, NIRS, or NDRS assumptions, it is logically consistent to employ the multiplicative specification of inefficiency and noise, to be discussed next.

6.2. Multiplicative errors

Although the additive model of errors is standard in the econometric theory, in practice, most SFA studies employ a multiplicative error model due to the log-transformations applied to the data (e.g., when the popular Cobb-Douglas or translog functional forms are used). As noted above, the RTS assumptions require a multiplicative specification of errors. Moreover, multiplicative error specification might alleviate heteroskedasticity across different sized firms.

In the present setting, applying log-transformations would violate the concavity constraints of the NLS regression. Therefore, we need an alternative way of modeling noise and inefficiency in a multiplicative fashion. The following multiplicative specification proves convenient for our purposes:

$$y_i = f(\mathbf{x}_i)\cdot(1-\varepsilon_i)^{-1} = f(\mathbf{x}_i)/(1+u_i-v_i) \quad ,i=1,...,n . \tag{23}$$

The composite error term ε_i and its components u_i and v_i are assumed to satisfy the standard assumptions imposed in Section 2.

For the purposes of estimation, we decompose the multiplicative model (23) as

$$y_i = \left[f(\mathbf{x}_i)/(1+\mu)\right]\cdot\left[(1+\mu)/(1+u_i-v_i)\right] = g(\mathbf{x}_i)/(1-\upsilon_i) \ ,i=1,...,n , \tag{24}$$

where μ is the expected inefficiency as in (5), $g(\mathbf{x}_i) \equiv f(\mathbf{x}_i)/(1+\mu)$ is the average production function, and $\upsilon_i = (\mu - u_i + v_i)/(1+\mu)$ is the modified composite error term with $E(\upsilon_i)=0$. Since $g(\mathbf{x}_i) = (1-\upsilon_i)y_i$, the average production function can be consistently estimated by the following NLS regression:

$$\begin{aligned}
&\min_{\boldsymbol{\alpha},\boldsymbol{\beta}_i,\boldsymbol{\upsilon}} \sum_{i=1}^{n} \upsilon_i^2 \\
&s.t. \\
&(1-\upsilon_i)y_i = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i \ \ \forall i=1,...,n \\
&\alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i \le \alpha_h + \boldsymbol{\beta}_h'\mathbf{x}_i \ \ \forall h,i=1,...,n \\
&\boldsymbol{\beta}_i \ge \mathbf{0} \ \ \forall i=1,...,n
\end{aligned} \tag{25}$$

Subsequently, the method of moments and pseudolikelihood techniques can be used for filtering out the noise from inefficiency.[5] Given the NLS production function $\hat{g}$ obtained from (25) and the parameter estimate $\hat{\sigma}_u$, the production function is estimated by

$$\hat{f}(\mathbf{x}_i) = \hat{g}(\mathbf{x}_i)(1+\hat{\sigma}_u\sqrt{2/\pi}) . \tag{26}$$

Firm efficiency can be gauged using the multiplicative Farrell output efficiency measure (the reciprocal of the Shephard output distance function), which can be estimated by $1+\hat{E}(u_i|\hat{\varepsilon}_i)$, where $\hat{E}(u_i|\hat{\varepsilon}_i)$ is calculated according to equation (22) using $\hat{\varepsilon}_i = \hat{\upsilon}_i + \hat{\sigma}_u\sqrt{2/\pi}(\hat{\upsilon}_i - 1)$.

6.3 Cost functions

Duality theory has established that the production technology can be equivalently modeled by means of monetary representations. The most popular dual representation is the cost function, formally defined as

$$C(y,\mathbf{w}) = \min_{\mathbf{x}}\left\{\mathbf{w}'\mathbf{x} \middle| f(\mathbf{x}) = y\right\}, \tag{27}$$

where **w** denotes the vector of exogenously given input prices. The cost function indicates the minimum cost of producing a given target output at given input prices. It is non-negative, non-decreasing, homogenous of degree one, concave and continuous in prices **w** (Kuosmanen, 2003; Theorem 3.3). These known properties provide a sound rationale for the nonparametric estimation.

In the stochastic cost frontier model, the observed costs C_i (i = 1,…,n) are assumed to differ from the cost function due to a composite error term (ε_i) which is the sum of a non-negative inefficiency term (u_i) and a noise term (v_i), that is,

$$C_i = C(y_i,\mathbf{w}_i) + \varepsilon_i = C(y_i,\mathbf{w}_i) + u_i + v_i. \tag{28}$$

Maintaining the assumptions of u_i and v_i stated in Section 2, the cost frontier can be estimated analogous to the production function procedure using the constrained NLS regression together with the method of moments or pseudolikelihood techniques.

The main challenge of the cost function estimation concerns the specification of the NLS model to estimate the conditional expected values $E(C_i|\mathbf{w}_i, y_i)$. If the production function f is concave, then the cost function is a convex function of output y. However, the cost function must be a concave function of input prices **w**. To estimate a cost frontier by NLS, we need to transform the cost function as a concave (or convex) function of all its arguments. To this end, we note that if the cost function is a convex function of output, then it is a concave function of its additive inverse. Introducing the expected inefficiency, we may rephrase equation (28) as

$$C_i = \left[C(y_i,\mathbf{w}_i) + \mu\right] + \left[u_i + v_i - \mu\right] = AC(-y_i,\mathbf{w}_i) + \upsilon_i, \tag{29}$$

where $AC(-y_i,\mathbf{w}_i) \equiv C(y_i,\mathbf{w}_i) + \mu$ is a concave function of all its arguments, and υ_i is a modified error term that satisfies the Gauss-Markov assumptions. The average cost curve *AC* can be consistently estimated by the NLS model:

---

[5] The multiplicative error structure should be taken into account when forming the log-likelihood function for PSL; equations (17)-(19) need to be adjusted to the multiplicative error structure.

$$
\begin{aligned}
&\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\upsilon}} \sum_{I=1}^{N} \upsilon_i^2 \\
&s.t. \\
&C_i = \alpha_i + \boldsymbol{\beta}_i'\mathbf{w}_i + \delta_i(-y_i) + \upsilon_i \quad \forall i = 1,...,n \\
&\alpha_i + \boldsymbol{\beta}_i'\mathbf{w}_i + \delta_i(-y_i) \le \alpha_h + \boldsymbol{\beta}_h'\mathbf{w}_i + \delta_h(-y_i) \quad \forall h,i = 1,...,n \\
&\boldsymbol{\beta}_i' \ge \mathbf{0},\ \delta_i \le 0 \quad \forall i = 1,...,n
\end{aligned}
\tag{30}
$$

Coefficients δ_i represent (the additive inverse of) the marginal cost of output, and are postulated to be non-positive. Coefficients $\boldsymbol{\beta}_i$ indicate the marginal cost of input prices (which depends on the input substitution possibilities). Intercepts α_i have an interpretation as the fixed cost.

Given the NLS residuals, the conditional expected values of the inefficiency terms can be estimated along the lines described in Section 5. Note the changed sign of the inefficiency component and the direction of skewness (compare, e.g., with Greene, 1997, and Kumbhakar and Lovell, 2000). The interpretation of the inefficiency term also changes: u_i here represents (overall) cost inefficiency that captures both technical and allocative aspects of inefficiency. Extending the cost frontier estimation to multi-output settings is straightforward.

7. Monte Carlo Simulations

The purpose of this section is three-fold. First, we present four simulated examples to illustrate how the proposed approach works in practice and to visualize the estimated StoNED frontiers and isoquants. Second, the simulations demonstrate that the proposed estimation methods can perform better than the existing parametric, nonparametric and semiparametric techniques at least in some nontrivial settings. Third, we compare systematically the performance of the method of moments and the pseudolikelihood estimators suggested in Section 5.

7.1 Illustration

We first estimate the StoNED frontiers in four simulated scenarios described by Table 1. Scenario A represents a single-input case, with a small sample size and relatively large noise. Scenario B also has a single input, but has a larger sample size and lower noise. Scenario C involves two inputs, constant returns to scale, and a multiplicative error structure. Scenario D is the most difficult to estimate, because it involves three inputs, and an exponential error structure that contradicts our assumptions.

Table 1: Description of the four scenarios

Scenario	inputs	true production function	n	σ_u	σ_v
A)	x	$y_i = x_i^{1/2} + \ln(x_i) + v_i - u_i$	50	0.6	0.4
B)	x	$y_i = \ln(x_i) + 2 + v_i - u_i$	100	0.6	0.3
C)	x_1,x_2	$y_i = (0.1x_{1i} + 0.1x_{2i} + 0.3(x_{1i}x_{2i})^{1/2})/(1 - v_i + u_i)$	100	0.4	0.2
D)	x_1,x_2,x_3	$y_i = (0.1x_{1i} + 0.1x_{2i} + 0.1x_{3i} + 0.3(x_{1i}x_{2i}x_{3i})^{1/3})\exp(v_i - u_i)$	100	0.4	0.2

In all scenarios, the input data were randomly sampled from *Uni*[1,11], independently for each input and firm. The efficient output levels were calculated using the production function described by Table 1. From the efficient output level, we subtracted a random inefficiency term $u_i \underset{i.i.d}{\sim} \left|N(0,\sigma_u^2)\right|$ and added a random error $v_i \underset{i.i.d}{\sim} N(0,\sigma_v^2)$, to obtain the "observed" output data used in estimation. The standard deviations σ_u^2,σ_v^2 used in each scenario are described in Table 1.

Given the four randomly generated data sets, we computed the shape constrained NLS regression and subsequently the MM and PSL estimators for each data set separately. In Scenarios A and B we employed unrestricted RTS and an additive error structure, while in Scenarios C and D we imposed CRS and a multiplicative error structure.

Figure 1 illustrates the results of Scenario A by plotting a scatter of the sample data (points ×), the true frontier (thick black curve), the NLS estimate of the average production function (thick, grey, piece-wise linear curve), and the StoNED frontiers estimated by the MM (solid, thin, piece-wise linear curve) and PSL (broken, piece-wise linear curve), respectively. The NLS estimator for the average production function consists of four different line segments (segments 1 and 2 have almost identical slopes and are indistinguishable in Figure 1). It is worth emphasizing that, in contrast to linear splines, the number of segments and the location of the vertices (or knots) are not specified a priori but are endogenously determined in the NLS problem (9) to minimize the sum of squares. Despite the small sample size, the NLS approximates the shape of the true production function reasonably well. As a result, both the MM and PSL estimators come rather close to the true production function, the PSL curve slightly exceeding the MM curve.

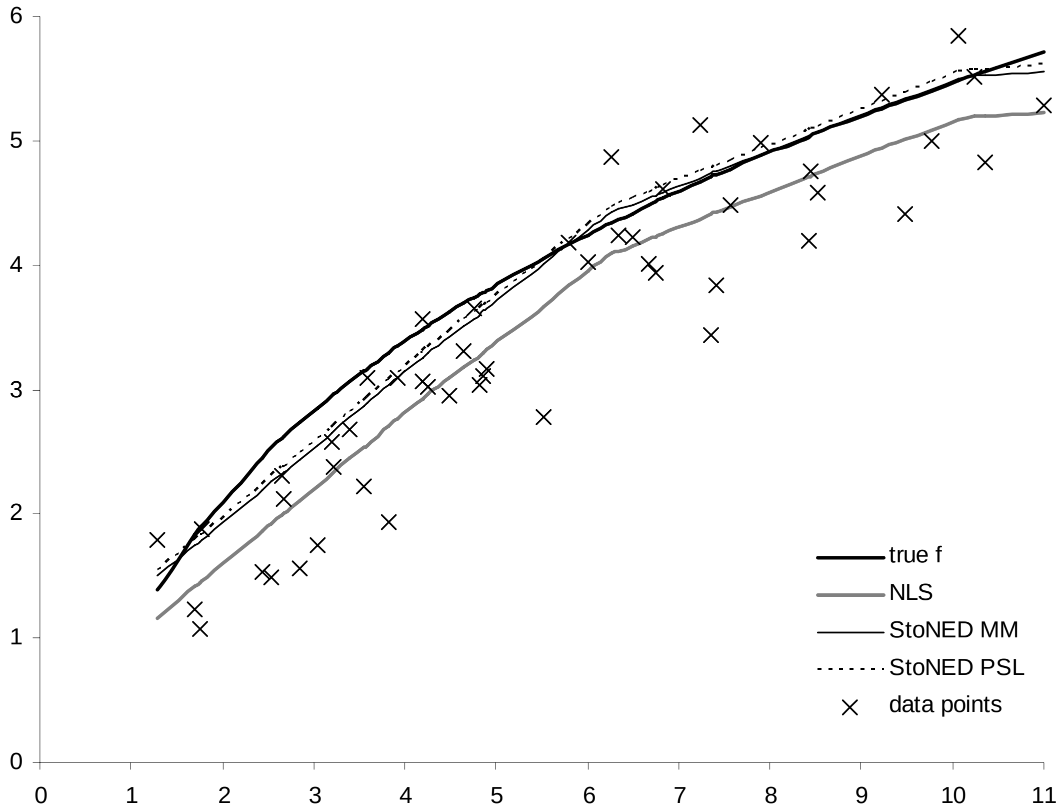

Figure 1: Scenario A: scatter of data, the NLS regression curve, and the StoNED frontiers

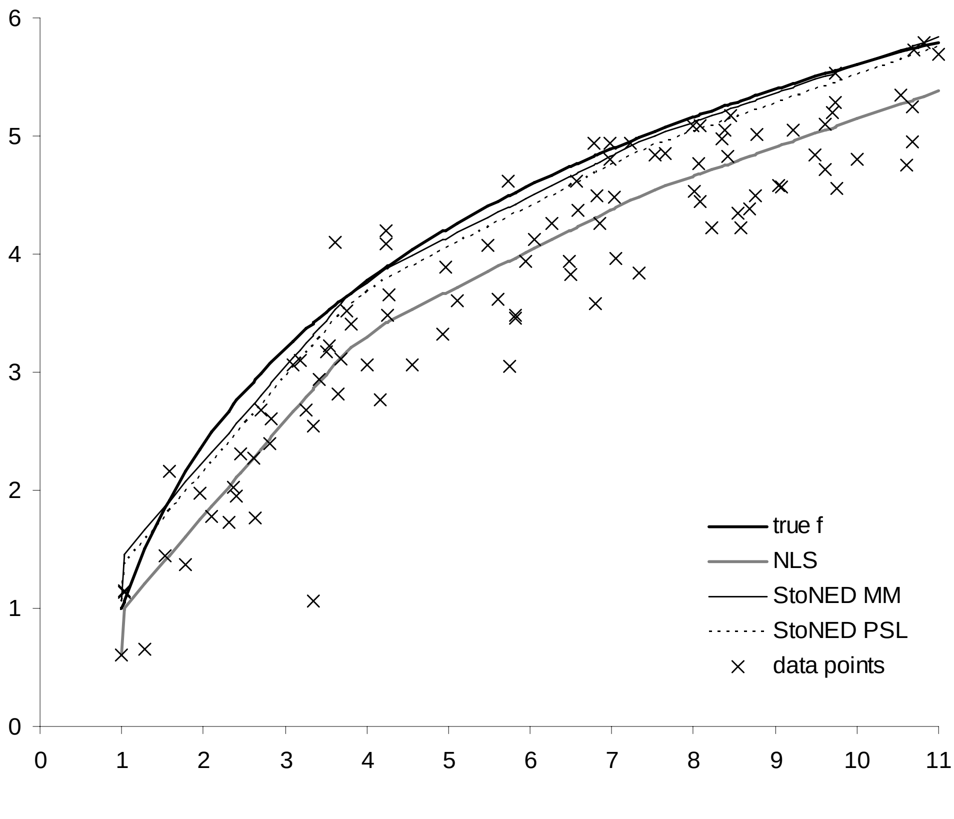

Figure 2: Scenario B: scatter of data, the NLS regression curve, and the StoNED frontiers.

Figure 2 illustrates the results of Scenario B in an analogous fashion. The true frontier is now more steeply curved, and the sample size is larger. The NLS curve consists of five different line segments (segments 3 and 4 are difficult to distinguish in Figure 2). In this Scenario, the MM curve indicates slightly higher output levels than the PSL curve. Nevertheless, both curves closely approximate the true frontier.

The multi-input scenarios are more difficult to visualize. Figure 3 illustrates Scenario C by means of an isoquant map. Since the production function exhibits CRS, we plot the scatter of data on normalized axes (x_1/y and x_2/y). The isoquant of the true production function is illustrated by the thin solid curve. The isoquant of the average production function estimated by NLS is represented by the thick, grey, piece-wise linear curve. It consists of four different line segments, and captures the shape of the true isoquant relatively well. The MM and PSL estimators of the frontier proved indistinguishable; they both are represented by the thin, broken, piece-wise linear curve. In this scenario, the estimated frontier falls somewhat short from the true frontier; the skewness of NLS residuals underestimated the magnitude of the true σ_u^2, which shows up both the MM and PSL estimates.

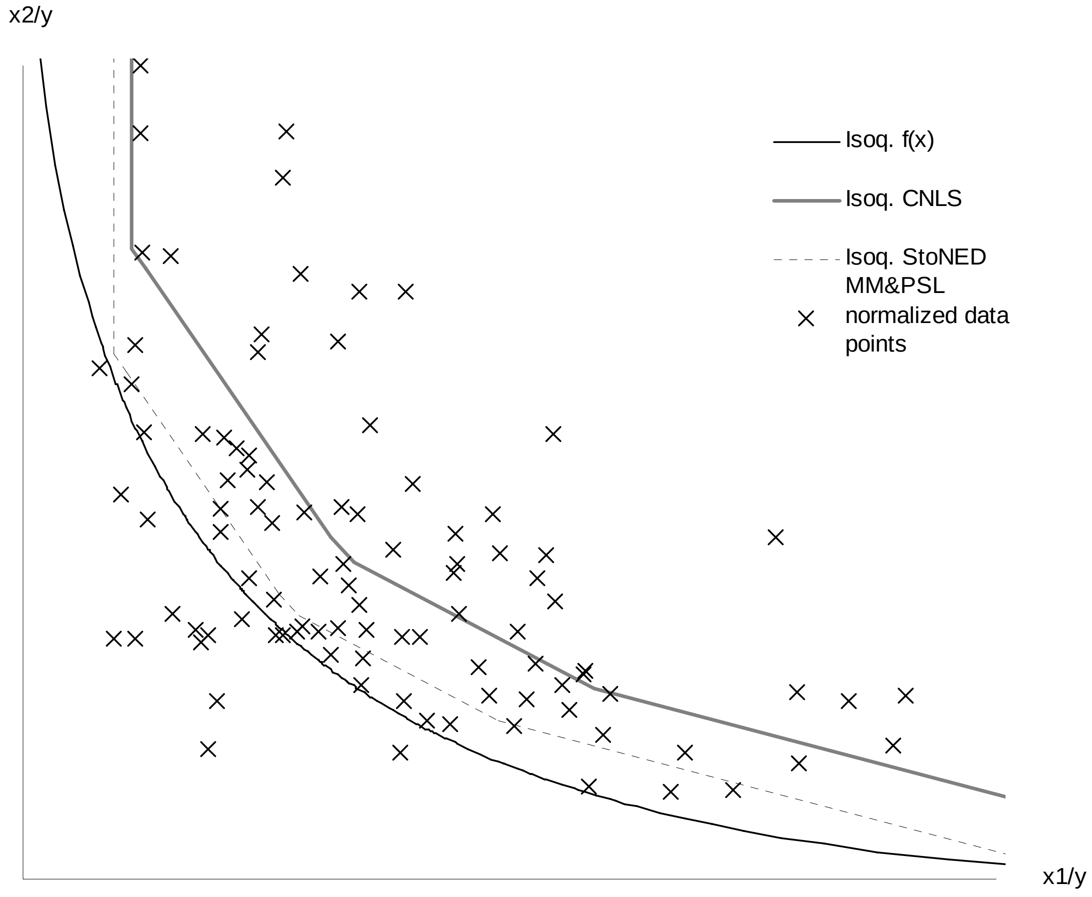

Figure 3: Scenario C: scatter of data and input isoquants of the NLS regression curve and the StoNED frontier

7.2 Comparison to other frontier estimation techniques

We next compared the performance of the two StoNED frontier estimation approaches to the conventional SFA and DEA approaches in the previous four scenarios. The comparison also includes the semi-parametric kernel regression approach by Fan et al. (1996).[6] We restrict attention on the estimation of the frontier production function in *n* observed points (for the stochastic approaches, the conditional expected value is used as an estimator). Table 2 reports the mean squared error (MSE) and the bias of the alternative estimators.

Table 2: Performance of alternative estimation techniques in the frontier estimation

		scenario A		scenario B		scenario C		scenario D	
		MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS
StoNED	MM	0,030	-0,105	0,010	-0,049	0,081	-0,266	0,296	-0,484
	PSL	0,022	-0,055	0,025	-0,132	0,085	-0,272	0,417	-0,594
SFA	CD	0,102	-0,023	0,189	-0,243	0,361	-0,544	0,092	0,190
	translog	0,101	-0,023	0,110	-0,307	0,371	-0,547	0,145	0,047
DEA	CRS	21,594	3,868	23,490	3,980	2,571	1,313	1,078	0,830
	VRS	0,097	0,260	0,151	0,364	1,065	0,802	0,581	0,427
semiparam. kernel		0,297	0,508	0,075	-0,093	0,103	0,272	0,943	-0,850

Overall, the performance of the proposed estimators was strong. In Scenarios B and C, the StoNED estimators achieved both the lowest MSE and bias. In Scenario A, the StoNED estimators achieved the lowest MSE, but the SFA estimators had the lowest bias. In Scenario D, the SFA estimators achieved the lowest MSE and bias, followed by the StoNED estimators. In general, SFA achieved lower MSE and bias than DEA, with the exception of Scenario A. Note that the MM estimator of the StoNED model outperformed the kernel estimator by both criteria in all four scenarios. The MM performed better than the PSL method in Scenarios B-D, but the PSL was superior in Scenario A.

7.3 MM vs. PSL estimators

We next evaluate the performance of the MM and PSL estimators for the model parameters σ^2, λ and the standard deviations σ_v, σ_u in a more systematic manner. To facilitate comparisons, we replicate the setting by Aigner et al. (1977) and Fan et al. (1996). We focus on the model $y_i = 1 + \varepsilon_i$, where $\varepsilon_i = v_i - u_i$, $v_i \underset{iid}{\sim} N(0, \sigma_v^2)$, and $u_i \underset{iid}{\sim} \left| N(0, \sigma_u^2) \right|$ for all $i = 1, \ldots, n$. As estimator for *g*, we use the

[6] The Gaussian kernel function was used. Following Fan et al. (1996), the smoothing parameter was set at $x_{sd} n^{-1/(m+4)}$ where x_{sd} is the sample standard deviation of input *x*.

arithmetic average of the observed output values, and hence the estimated $\hat{g}$ may differ from the true value of g (i.e., $1-\mu$). Hence, this simple model suffices to capture the essential elements of frontier estimation for the comparison of the MM and PSL estimators.

Following Aigner et al. (1997) and Fan et al. (1996), three different levels of the true (σ^2, λ) parameters are considered: (σ^2, λ) = (1.88, 1.66), (1.63, 1.24), (1.35, 0.83). These correspond to standard deviations (σ_v, σ_u) = (0.71, 1.17), (0.80, 0.99), (0.89, 0.74), respectively. All three models are replicated 1,000 times and the MM and PSL estimates are computed for $\sigma^2, \lambda, \sigma_v$, and σ_u. This allows us to assess the performance of the estimators both within and across scenarios.

Table 3 reports the MSE and bias of the MM and PSL estimators in the three scenarios. In general, our results are in line with those reported by Aigner et al. (1977) and Fan et al. (1996). Although there can be large differences between the MM and PSL estimators in a single model run, neither MM nor PSL approach dominates in light of these results. An interesting finding in the simulations was that the MM estimator of parameter λ overshoots the true value more often than the PSL estimator. This explains the much better performance of PSL in the estimation of λ. On the other hand, the PSL estimator yields zero estimates of parameter λ more often than the MM estimator (the negative values of λ are truncated to zero). This could partly explain the good performance of MM in the estimation of σ_u^2 and σ_v^2 parameters.

Table 3: Mean squared error (MSE) and bias statistics of the MM and PSL estimators

	Scenario 1		Scenario 2		Scenario 3	
	σ^2	λ	σ^2	λ	σ^2	λ
	1,88	1,66	1,63	1,24	1,35	0,83
PSL	MSE	BIAS	MSE	BIAS	MSE	BIAS
σ^2	0,297	-0,075	0,285	-0,042	0,150	-0,149
λ	1,107	0,043	0,578	-0,139	0,466	-0,400
σ_u^2	0,524	-0,097	0,541	0,262	0,450	-0,483
σ_v^2	0,054	0,022	0,141	-0,304	0,147	0,334
MM	MSE	BIAS	MSE	BIAS	MSE	BIAS
σ^2	0,282	-0,079	0,294	-0,012	0,210	0,038
λ	2,438	0,117	0,920	-0,030	0,587	-0,019
σ_u^2	0,500	-0,103	0,600	0,310	0,435	-0,189
σ_v^2	0,054	0,024	0,157	-0,322	0,108	0,227

Table 4: Head-to-head comparison of the MM and PSL estimators

	Scenario 1		Scenario 2		Scenario 3	
	σ^2	λ	σ^2	λ	σ^2	λ
	1,88	1,66	1,63	1,24	1,35	0,83
PSL wins	47,0 %	49,6 %	40,5 %	39,0 %	35,0 %	37,5 %
MM wins	45,8 %	43,2 %	40,5 %	42,0 %	33,8 %	31,3 %
equally good	7,2 %	7,2 %	19,0 %	19,0 %	31,3 %	31,3 %

In addition to the MSE and bias statistics, we also conducted a head-to-head comparison of the two estimators for parameters σ^2, λ in the same model runs. Table 4 reports the percentages of model runs where the PSL yields more accurate estimates than the MM, and vice versa. These results suggest that the PSL approach performs generally somewhat better than MM. Still, it is fair to note that the MM wins the PSL in a significant proportion of the cases. The cases where the two approaches yield exactly the same result occur when the estimated λ parameter yields the zero value. The percentage of these cases increases as the true λ decreases. Parameter λ can be interpreted as a signal-to-noise ratio; obviously, detecting the signal (inefficiency) is more difficult in a noisier environment.

8. Application to tax collection offices

We next apply the proposed estimation techniques to empirical data to compare the efficiency estimates with those obtained by standard DEA and SFA methods. We re-examine a cross-sectional data of 62 local property tax collection offices (called rates departments) in the London Boroughs and Metropolitan Districts; the data has been documented and used for relative performance assessment by Thanassoulis et al. (1987) and Dyson and Thanassoulis (1988).[7]

The data set includes the total annual cost (considered as an input by Thanassoulis et al.) and four output variables. Thus, we resort to the cost frontier approach (see Section 6.3), assuming that the cost function is increasing and convex in outputs and that the input prices are the same across all offices. The output variables are: (y_1) non-council hereditaments, (y_2) rates rebates granted, (y_3) summonses issued and distress warrants obtained, and (y_4) NPV of non-council rates collected.

We first estimated the average cost function using the shape constrained NLS model. This gave a good empirical fit, with the coefficient of determination R^2=0.974. The piece-wise linear NLS cost function consists of 29 different segments, the largest one containing 11 observations. Based on the NLS residuals, the standard deviations of the inefficiency and error terms were estimated using the MM and PSL techniques. Based on the NLS residuals and the estimated standard deviations $\hat{\sigma}_u^2, \hat{\sigma}_v^2$, the

[7] The data set is available online at: http://www.etm.pdx.edu/dea/dataset/

conditional expected values of the inefficiency terms were computed using the results by Jondrow et al.

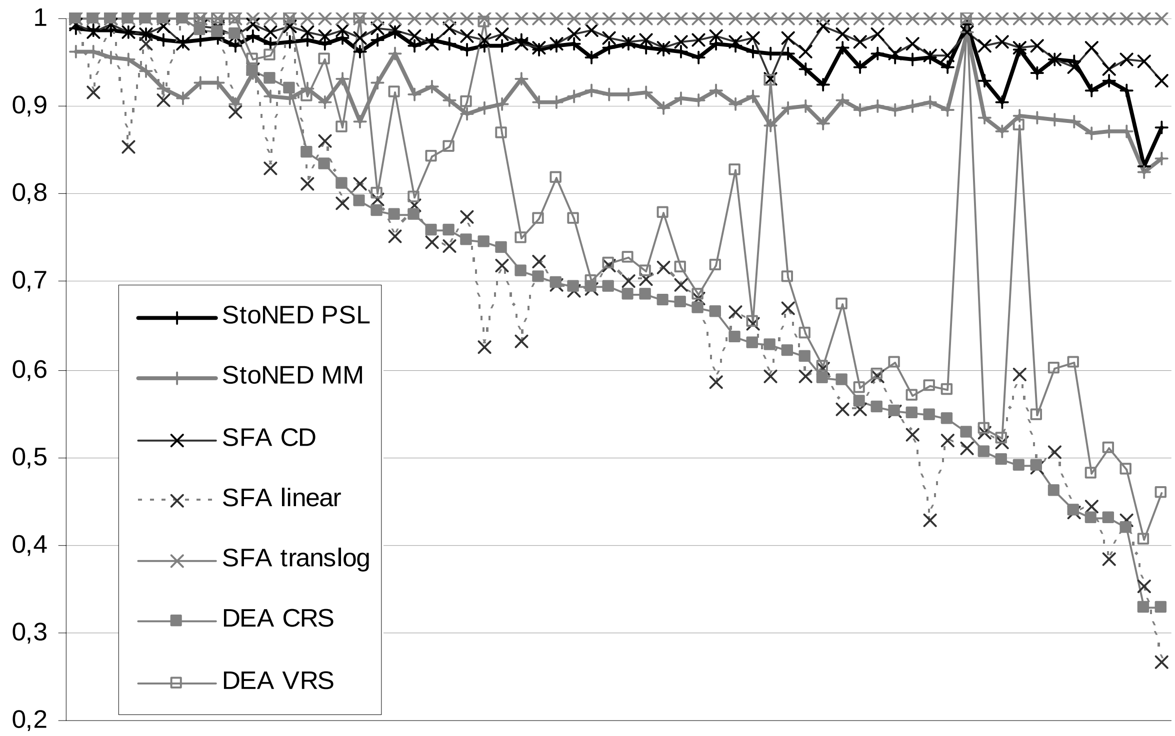

Figure 4: Cost efficiency indices of the 62 authorities by different methods (minimum / observed cost)

Figure 4 presents the distribution of the cost efficiency estimates (i.e., the ration of minimum to observed cost) computed with the StoNED, SFA, and DEA methods. Each point in the figure represents a cost efficiency index of an observed unit: the 62 units run on the horizontal axis, ranked in decreasing order according to the input-oriented DEA CRS efficiency scores. Regarding the StoNED estimates, the PSL technique yields systematically higher cost efficiency estimates than the MM in this application. Still, the efficiency rankings are very similar: the correlation coefficient between the PSL and MM efficiency estimates is 0.845. For comparison, we estimated three SFA models using the linear, Cobb-Douglas and translog functional forms, assuming the half-normal specification of the error term. In the translog specification, the OLS residuals were skewed in the wrong direction, and thus all units were diagnosed as efficient. Also the Cobb-Douglas specification yields higher efficiency levels than the StoNED models. By contrast, the linear specification resulted with efficiency estimates that come very close to the DEA CRS efficiency scores (the correlation coefficient 0.967). In light of these observations, we conclude that the SFA estimates are sensitive to chosen functional form in this application. The results obtained with the most standard DEA and SFA methods diverge extensively, while the StoNED estimates fell somewhere between the two. Of course, we cannot say which of the methods performs best in this application.

Nevertheless, this application demonstrates that the StoNED approach can provide empirical results that are not obtainable by the established SFA and DEA methods.

9. Conclusions and discussion

We have shown that the frontier estimation based on a nonparametric DEA-like production function and stochastic SFA-like inefficiency and noise terms is possible in practice. The new approach, referred to as *Stochastic Nonparametric Envelopment of Data* (StoNED), melds the virtues of both DEA and SFA into a uniform framework of frontier estimation. Indeed, both DEA and SFA can be viewed as special cases of StoNED under some more restrictive assumptions. While we mainly focused on the estimation of production functions under variable returns to scale, we also demonstrated how the method extends to the estimation of cost functions and other representations of technology and allows one to postulate or test alternative specifications of returns to scale.

The potential of the StoNED approach was illustrated by means of numerical examples, Monte Carlo simulations, and an empirical application. The simulated examples demonstrated that the proposed method can outperform the existing parametric, nonparametric and semiparametric frontier estimation techniques at least in some circumstances. Both the method of moments and the pseudo-likelihood estimators proved competitive in the Monte Carlo simulations. Finally, the example application proves that the StoNED approach can provide empirical results that cannot be obtained by the standard SFA or DEA models. This information can be particularly valuable when the results of SFA and DEA diverge and the SFA results are sensitive to the functional form specification.

The proposed StoNED approach shares many common features with SFA and DEA, being a genuine amalgam of the two. Thus, many of the existing tools and techniques for SFA and DEA can be readily incorporated into the StoNED framework. However, the hybrid nature of StoNED also means that there are many important differences to both SFA and DEA, which must be kept in mind in the application and interpretation of StoNED models. For example, the interpretation of the StoNED input coefficients differs considerably from those of the SFA coefficients. Moreover, in contrast to DEA, all observations influence the shape of the frontier. Further research is needed for a better understanding of these similarities and differences. We hope that this paper could inspire further theoretical and empirical work in this direction, and thus contribute to the unification of the parametric and nonparametric streams of productive efficiency analysis.

While the StoNED approach combines the appealing features of DEA and SFA, it also shares some of their limitations. Similar to DEA, the nonparametric orientation of StoNED can make it vulnerable to the curse of dimensionality, which means that the sample size must be very large when the number of input variables is high. On the other hand, the maintained SFA assumptions regarding the composite error distribution may be violated. Moreover, stochastic noise does not necessarily restrict to the output, but also input data may be perturbed by measurement errors and other noise. Treatment of noise in the input data remains somewhat problematic in the SFA framework, and hence also in the StoNED approach. Addressing these shared limitations presents interesting challenges for the future research.

Acknowledgements

We are grateful to Bill Greene, Daniel Henderson, Subal Kumbhakar, Timo Sipiläinen, Rauli Svento, and Michael Vardanyan for valuable comments and suggestions. A part of this paper has been written while the second author was visiting Leonard N. Stern School of Business at New York University. He wants to thank his hosts for hospitality and the Finnish Doctoral Programme in Economics (FDPE) and the Northern Karelia Fund of the Finnish Cultural Foundation for financial support.

References

Afriat, S.N., 1967, The Construction of a Utility Function from Expenditure Data, *International Economic Review* 8, 67-77.

Afriat, S., 1972, Efficiency Estimation of Production Functions, *International Economic Review* 13, 568-598.

Aigner, D.J., and S. Chu, 1968, On Estimating the Industry Production Function, *American Economic Review* 58, 826-839.

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

Allon, G., M. Beenstock, S. Hackman, U. Passy, and A. Shapiro, 2007, Nonparametric Estimation of Concave Production Technologies by Entropic Methods, *Journal of Applied Econometrics*, in press.

Bauer, P.W., 1990, Recent Developments in the Econometric Estimation of Frontiers, *Journal of Econometrics* 46 (1-2), 39-56.

Banker, R.D., A. Charnes, W. W. Cooper, 1984, Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science* 30(9), 1078 – 1092.

Banker, R.D., and A. Maindiratta, 1992, Maximum Likelihood Estimation of Monotone and Concave Production Frontiers, *Journal of Productivity Analysis* 3, 401-415.

Banker, R.D., S. Janakiraman, and R. Natarajan, 2002, Evaluating the Adequacy of Parametric Functional Forms in Estimating Monotone and Concave Production Functions, *Journal of Productivity Analysis* 17, 111-132.

Charnes, A., W.W. Cooper, and E. Rhodes, 1978, Measuring the Inefficiency of Decision Making Units, *European Journal of Operational Research* 2(6), 429-444.

Christensen, L.R., and D.W. Caves, 1980, Global Properties of Flexible Functional Forms, *American Economic Review* 70, 322-332.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau, 1973, Transcendential Logarithmic Production Functions, *Review of Economics and Statistics* 55(1), 28-45.

Cornwell, C., P. Schmidt, and R. Sickles, 1990, Production Frontiers with Cross Sectional and Time Series Variation in Efficiency Levels, *Journal of Econometrics* 46, 185-200.

Diewert, W.E., and C. Parkan, 1983, Linear Programming Tests of Regularity Conditions for Production Frontiers, in: Eichhorn, W., Henn, R., Neumann, K., Shephard, R.W., (Eds.), *Quantitative studies on Production and Prices*. Physica-Verlag, Würzburg.

Diewert, W.E., and T.J. Wales, 1987, Flexible Functional Forms and Global Curvature Conditions, *Econometrica* 55(1), 43-68.

Dyson, R.G., and E. Thanassoulis, 1988, Reducing Weight Flexibility in Data Envelopment Analysis, *Journal of the Operational Research Society* 39, 563-576.

Efron, B., 1979, Bootstrap Methods: Another Look at the Jackknife, *Annals of Statistics* 7, 1-16.

Efron, B., 1982, The Jackknife, the Bootstrap and Other Resampling Plans, *CBMS-NSF Regional Conference Series in Applied Mathematics* #38. Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B., and R.J. Tibshirani, 1993, *An Introduction to the Bootstrap*, London: Chapman and Hall.

Fan, J., and W. Zhang, 1999, Statistical Estimation in Varying Coefficients Models, *Annals of Statistics* 27, 1491-1518.

Fan, Y., Q. Li, and A. Weersink, 1996, Semiparametric Estimation of Stochastic Production Frontier Models, *Journal of Business and Economic Statistics* 14(4), 460-468.

Farrell, M.J., 1957, The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society Series A. General* 120(3), 253-282.

Fraser, D.A.S., and H. Massam, 1989, A Mixed Primal-Dual Bases Algorithm for Regression Under Inequality Constraints: Application to Convex Regression, *Scandinavian Journal of Statistics* 16, 65-74.

Greene, W.H., 1997, Frontier Production Functions. in: M. Pesaran and P. Schmidt, (Eds.), *Handbook of Applied Econometrics, Vol II: Microeconomics*. Blackwell Publishers, Oxford, pp. 81–166.

Greene, W.H., 2005, Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model, *Journal of Econometrics* 126, 269-303.

Griffin, J.F., and M.J.F. Steel, 2004, Semiparametric Bayesian Inference for Stochastic Frontier Models, *Journal of Econometrics* 123, 121-152.

Groeneboom, P., G. Jongbloed, and J.A. Wellner, 2001, Estimation of a Convex Function: Characterizations and Asymptotic Theory, *Annals of Statistics* 29, 1653-1698.

Hanson, D.L., and G. Pledger, 1976, Consistency in Concave Regression, *Annals of Statistics* 4(6), 1038-1050.

Hanoch, G. and M. Rotschild, 1972, Testing the Assumptions of Production Theory: A Nonparametric Approach, *Journal of Political Economy* 80, 256-275.

Henderson, D.J., and L. Simar, 2005, A Fully Nonparametric Stochastic Frontier Model for Panel Data, Discussion Paper 0417, Institut de Statistique, Universite Catholique de Louvain.

Hildreth, C., 1954, Point Estimates of Ordinates of Concave Functions, *Journal of the American Statistical Association* 49(267), 598-619.

Horrace, W.C., and P. Schmidt, 1996, Confidence Statements for Efficiency Estimates from Stochastic Frontier Models, *Journal of Productivity Analysis* 7, 257-282.

Jondrow, J., C.A.K. Lovell, I.S. Materov, and P. Schmidt, 1982, On Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model, *Journal of Econometrics* 19, 233-238.

Kneip, A., and L. Simar, 1996, A General Framework for Frontier Estimation with Panel Data, *Journal of Productivity Analysis* 7, 187-212.

Kumbhakar, S.C., and C.A.K. Lovell, 2000, *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge.

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

Kuosmanen, T., 2003, Duality Theory of Non-Convex Technologies, *Journal of Productivity Analysis* 20, 273–304.

Kuosmanen, T., 2006, Stochastic Nonparametric Envelopment of Data: Combining Virtues of SFA and DEA in a Unified Framework, MTT Discussion Paper 3/2006.

Kuosmanen, T., G.T. Post, and S. Scholtes, 2006, Nonparametric Tests of Productive Efficiency with Errors-in-Variables, *Journal of Econometrics* 136, 131-162.

Mammen, E., 1991, Nonparametric Regression under Qualitative Smoothness Assumptions, *Annals of Statistics* 19, 741-759.

Mammen, E., and C. Thomas-Agnan, 1999, Smoothing Splines and Shape Restrictions, *Scandinavian Journal of Statistics* 26, 239-252.

Matzkin, R.L., 1994, Restrictions of Economic Theory in Nonparametric Methods, Chapter 42 in: R.F. Engle and D.L McFadden,Eds., *Handbook of Econometrics Vol IV,* Elsevier.

Meeusen, W., and J. van den Broeck, 1977, Efficiency Estimation from Cobb-Douglas Production Function with Composed Error, *International Economic Review* 8, 435-444.

Meyer, M.C., 1999, An Extension of the Mixed Primal-Dual Bases Algorithm to the Case of More Constraints than Dimensions, *Journal of Statistical Planning and Inference* 81, 13-31.

Meyer, M.C. , 2003, A Test for Linear vs. Convex Regression Function using Shape-Restricted Regression, *Biometrika* 90(1), 223–232.

Meyer, M.C., 2006, Consistency and Power in Tests with Shape-Restricted Alternatives, *Journal of Statistical Planning and Inference* 136, 3931-3947.

Nemirovskii, A.S., B.T. Polyak, and A.B. Tsybakov, 1985, *Rates of Convergence of Nonparametric Estimates of Maximum Likelihood Type*, Problems of Information Transmission 21, 258-271.

Park, B., and L. Simar, 1994, Efficient Semiparametric Estimation in Stochastic Frontier Models, *Journal of the American Statistical Association* 89, 929-936.

Park, B., R.C. Sickles, and L. Simar, 1998, Stochastic Panel Frontiers: A Semiparametric Approach, *Journal of Econometrics* 84, 273-301.

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

Park, B., R.C. Sickles, and L. Simar, 2006, Semiparametric Efficient Estimation of Dynamic Panel Data Models, *Journal of Econometrics* 136, 281-301.

Post, G.T., L. Cherchye, and T. Kuosmanen, 2002, Nonparametric Efficiency Estimation in Stochastic Environments, *Operations Research* 50, 645-655.

Richmond, J., 1974, Estimating the Efficiency of Production, *International Economic Review* 15, 515-521.

Sarath, B., and A. Maindiratta, 1997, On the Consistency of Maximum Likelihood Estimation of Monotone and Concave Production Frontiers, *Journal of Productivity Analysis* 8, 239–246.

Sauer, J., 2006, Economic Theory and Econometric Practice: Parametric Efficiency Analysis, *Empirical Economics* 31, 1061-1087.

Schmidt, P., and R. Sickles, 1984, Production Frontiers and Panel Data, *Journal of Business and Economic Statistics* 2, 367–374.

Seiford, L. M., and R.M. Thrall, 1990, Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis, *Journal of Econometrics* 46,1-2), 7-38.

Simar, L., and P.W. Wilson, 2005, Estimation and Inference in Cross-Sectional Stochastic Frontier Models, Discussion paper 0525, Institut de Statistique, UCL.

Stone, C.J., 1980, Optimal Rates of Convergence for Nonparametric Estimators, *Annals of Statistics* 8(6), 1348-1360.

Stone, C.J., 1982, Optimal Global Rates of Convergence for Nonparametric Regression, *Annals of Statistics* 10(4), 1040-1053.

Thanassoulis, E., R.G. Dyson, and M.J. Foster, 1987, Relative Efficiency Assessments Using Data Envelopment Analysis: An Application to Data on Rates Departments, *Journal of the Operational Research Society* 38, 397-411.

Timmer, P., 1971, Using a Probabilistic Frontier Production Function to Measure Technical Efficiency, *Journal of Political Economy* 79, 776-794.

Varian, H.R., 1984, The Nonparametric Approach to Production Analysis, *Econometrica* 52, 579-598.

Yatchew, A.J., 1998, Nonparametric Regression Techniques in Economics, *Journal of Economic Literature* 36, 669-721.

Yatchew, A., 2003, *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press.

.**Appendix: Proofs of the propositions**

Proposition 1: By Afriat's Theorem, if the fitted output values y_i^g (i=1,…,n) satisfy the constraints of (8), then there exists a continuous, monotonic increasing, concave function $g \subset F_2$ such that $y_i^g = g(\mathbf{x}_i)\ \forall i = 1,...,n$ (see, e.g., Banker and Maindiratta, 1992; Proposition 1). Since the objective function of (7) depends on the value of *g* only in a finite set of points $\mathbf{x}_i, i = 1,...,n$, representing function *g* with the fitted output values y_i^g (i=1,…,n) does not involve a loss of generality. Therefore, the equality $s_{NLS}^2 = s_{Afriat}^2$ holds for any real-valued data set (**X**,**y**). □

Proposition 2: First, introduce intercepts $\alpha_i \equiv y_i^g - \boldsymbol{\beta}_i'\mathbf{x}_i$ and modified composite errors $\upsilon_i \equiv y_i - y_i^g$, $i = 1,...,n$. Thus, $y_i = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i + \upsilon_i$. We now see that the objective functions of (8) and (9) are equivalent. The proof is completed by transforming the first constraint of (8) into the second constraint of (9). To see this, we start from the first inequality of (8):

$$y_i^g - \boldsymbol{\beta}_i'\mathbf{x}_i \geq y_j^g - \boldsymbol{\beta}_i'\mathbf{x}_j \ \ \forall i, j = 1,...,n. \qquad \text{(A1)}$$

Substituting $y_i^g = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i$ and $y_j^g = \alpha_j + \boldsymbol{\beta}_j'\mathbf{x}_j$ into (A1) and reorganizing the terms, we obtain

$$\alpha_j + \boldsymbol{\beta}_j'\mathbf{x}_j \leq \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_j \ \ \forall i, j = 1,...,n. \qquad \text{(A2)}$$

Indices *i*,*j* run through all firms in the sample, so we can harmlessly substitute the pair (*j*,*i*) by (*i*,*h*) and rewrite inequalities (A2) identical to the second constraint of (9)as

$$\alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i \leq \alpha_h + \boldsymbol{\beta}_h'\mathbf{x}_i \ \ \forall i, h = 1,...,n. \quad □ \qquad \text{(A3)}$$

Proposition 3:

It is straightforward to verify that $\hat{g} \in F_2$. Combining propositions 1 and 2, we have $s_{NLS}^2 = s_{VC}^2$. This directly implies that $\hat{g} \in G_2^*$.