

Stochastic Nonparametric Envelopment of Data: Frontier Estimation Subject to Shape Constraints

Timo Kuosmanen^{1,2} and Mika Kortelainen³

- 1) Helsinki School of Economics, 00101 Helsinki, Finland.
- 2) MTT Agrifood Research Finland, 00410 Helsinki, Finland. E-mail. Timo.Kuosmanen@mtt.fi.
- 3) University of Manchester, M13 9PL, UK; E-mail. Mika.Kortelainen@manchester.ac.uk.

Abstract

Literature of productive efficiency analysis is currently divided between two main paradigms: the parametric Stochastic Frontier Analysis (SFA) and the deterministic, nonparametric Data Envelopment Analysis (DEA). This paper develops a new encompassing framework that melds the SFA-style stochastic composite error term to the DEA-type nonparametric frontier that satisfies monotonicity and concavity. The new approach is referred to as Stochastic Nonparametric Envelopment of Data (StoNED). StoNED method utilizes convex nonparametric least squares (CNLS), which estimates the shape of the frontier without any assumptions about its functional form or smoothness. In cross-sectional settings, distinguishing inefficiency from noise requires distributional assumptions, which can be relaxed in the case of panel data. We estimate the conditional expectations of inefficiency based on the CNLS residuals, using the method of moments and pseudolikelihood techniques. Performance of the StoNED procedure is examined using Monte Carlo simulations.

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 productive efficiency analysis and frontier estimation is large and growing, consisting of several thousands of studies in the fields of applied economics, econometrics, operations research, and statistics (see e.g. Fried et al., 2008, for an up-to-date introduction and literature review). This field is currently 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 are frequently used. 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. 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.

Bridging the gap between SFA and DEA has been recognized as one of the most important research objectives in this field, and contributions to this end have accumulated since the early 1990s. The emerging literature on semi/nonparametric stochastic frontier estimation has thus far mainly departed from the SFA side, replacing the parametric frontier function by a nonparametric specification that can be estimated by kernel regression or local maximum likelihood (ML) techniques.

Fan et al. (1996) and Kneip and Simar (1996) were among the first to apply kernel regression to frontier estimation in the cross-sectional and panel data contexts, respectively. Fan et al. (1996) proposed a two-step method where the shape of the frontier is first estimated by kernel regression, and the conditional expected inefficiency is subsequently estimated based on the residuals, imposing the same distributional assumptions as in standard SFA. Kneip and Simar (1996) similarly use kernel regression for estimating the frontier, but they make use of panel data to avoid the distributional assumptions. Other semi/nonparametric panel data approaches include Park et al. (1998, 2003, 2006) and Henderson and Simar (2005).

Banker and Maindiratta (1992) were the first to consider ML estimation of the stochastic frontier model subject to nonparametric shape constraints regarding the frontier. Their monotonicity and concavity constraints are analogous to DEA, but solving the resulting complex ML problem has proved extremely difficult, if not impossible in practical applications. No reported applications of the Banker

and Maindiratta's constrained ML method are known in the literature. Recently, Kumbhakar et al. (2007) proposed a more flexible SFA method based on local polynomial ML estimation. While the model is parametrized similar to the standard SFA models, all model parameters are approximated by local polynomials. Simar and Zelenyuk (2008) have further extended the local polynomial ML method to multi-output technologies. The latter study also uses DEA to the fitted values of the Kumbhakar et al. (2007) method to impose monotonicity and concavity. However, the link to DEA remains instrumental.

While the earlier semi/nonparametric approaches come a long way in enhancing flexibility of the SFA frontier, a substantial gap to DEA remains. It is worth emphasizing that the nonparametric kernel and local ML methods build upon smoothing assumptions, which are incompatible with the global shape constraints typically imposed in DEA. Put conversely, lack of smoothing and adherence to global shape constraints make DEA incompatible with the nonparametric smoothing techniques. We do not claim that either DEA or nonparametric smoothing is superior, they are just different approaches. In any case, none of the earlier semi/nonparametric frontier methods contains DEA as a special case. Moreover, none of the earlier approaches 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.

This paper develops an encompassing framework that includes both SFA and DEA as its constrained special cases. More specifically, we introduce a stochastic SFA-style composite error term, consisting of noise and inefficiency components, to a nonparametric, DEA-style piecewise linear frontier. Such a unifying approach deserves a catchy name, so we will henceforth refer to this amalgam framework as *stochastic nonparametric envelopment of data* (StoNED).¹

We estimate the StoNED model by convex nonparametric least squares regression (CNLS: Hildreth, 1954; Hanson and Pledger, 1976; Groeneboom et al., 2001), which does not require any smoothing parameters. To our knowledge, CNLS has not been applied to frontier estimation before. In this respect, this study builds upon the prior work by Kuosmanen (2008), who was the first to point out the connection between DEA and CNLS (see also Kuosmanen and Johnson, 2009). Importantly, CNLS does not assume a priori any particular functional form for the regression function; it identifies the function that best fits the data from the family of continuous, monotonic increasing, concave functions that can be non-differentiable. Kuosmanen noted that the single-output DEA model can be regarded as a constrained variant of CNLS regression. In this paper we exploit this connection between DEA and CNLS further by introducing a stochastic noise term to a nonparametric regression model that is based on the same global shape constraints as the standard DEA.

While this paper focuses on the cross-sectional model, we will also briefly show how the approach can be extended to the panel data. In that setting, the time-invariant inefficiency components can be estimated in a fully nonparametric fashion by resorting the standard fixed effects or random

¹ By the term "nonparametric envelopment" we refer specifically to the nonparametric treatment of the frontier. Taken as a whole, the cross-sectional StoNED model is more appropriately described as "semiparametric" due to the distributional assumptions related to the stochastic components. Importantly, these parametric assumptions can be avoided in the panel data setting.

effects treatments. In the cross-sectional setting, some further distributional assumptions are necessary for identifying inefficiency from noise. Our cross-sectional StoNED method consists of two stages. In the first stage, we estimate the shape of the production function by CNLS without making any functional form, distributional or smoothness assumptions. CNLS provides an unbiased, consistent estimator for the shape of the production frontier, but inefficiency and noise terms remain indistinguishable. Therefore, in the second stage we follow Fan et al. (1996) and impose some standard distributional assumptions adopted from the SFA literature, and estimate the conditional expected value of the inefficiency term using the method of moments or pseudolikelihood techniques.

Our StoNED method differs from the parametric and semi/nonparametric SFA treatments in that we do not make any functional form or smoothness assumptions, but build upon the global shape constraints (monotonicity, concavity). Compared to DEA, the StoNED method differs in its robustness to outliers and extreme observations and in its probabilistic treatment of inefficiency and noise. While the DEA frontier is typically spanned by a small number of influential observations, StoNED method uses information contained in the entire sample of observations for estimating the frontier. The main appeal of StoNED does not necessarily lie in its attractive properties; the key motivation of this paper is to contribute to better understanding of the connections between SFA and DEA by developing an encompassing framework that contains both methods as its special cases. Such an amalgam framework offers exciting new prospects for cross-fertilization between DEA and SFA paradigms.

The remainder of the paper is organized as follows. Section 2 introduces the StoNED model as a generalization of DEA and SFA and motivates our two-step estimation strategy. Section 3 elaborates the first-step of nonparametric estimation of the production function by employing CNLS regression. Based on the CNLS residuals, we estimate the inefficiency and noise terms by means of method of moments and pseudolikelihood techniques in Section 4. Section 5 discusses some useful extensions to illustrate the potential of the proposed approach. Section 6 examines how the proposed techniques perform in a controlled environment of Monte Carlo simulations. Finally, Section 7 draws the concluding remarks. Proof of Theorem 3.2 and an illustrative example are presented in Appendices 1 and 2, respectively. Further supplementary material such as graphical illustrations, example applications, and computational codes are available in the working papers Kuosmanen (2006), Kuosmanen and Kortelainen (2007), and the StoNED homepage: <http://www.nomepre.net/stoned/>.

2. Stochastic nonparametric envelopment of data (StoNED)

This section formally introduces the StoNED model in the cross-sectional setting; an extension to panel data is presented in Section 5.1. To maintain direct contact with SFA, we describe the model for the single-output multiple input case.² The m -dimensional input vector is denoted by \mathbf{x} and the scalar output by y . The production technology is represented by the *production function* $y = f(\mathbf{x})$. We assume that

² Extensions to multi-output setting are possible by using distance functions; see the working paper Kuosmanen (2006) for further details.

function f belongs to the class of continuous, monotonic increasing and globally concave functions that can be nondifferentiable. In what follows, this class of functions will be 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,

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

Terms u_i and v_i ($i = 1, \dots, n$) 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 $u_i \sim \left| N(0, \sigma_u^2) \right|$ 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.

In model (1), the deterministic part (i.e., production function f) is defined analogous to DEA, while the stochastic part (i.e., composite error term ε_i) is defined similar to SFA. As a result, model (1) encompasses the classic SFA and DEA models as its constrained special cases. 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 single-output DEA model with an additive output-inefficiency, first considered by Afriat (1972). In this sense, both SFA and DEA can be seen as constrained special cases of model (1).

It is easy to write a theoretical model like (1); the main challenge is its estimation. In this paper we propose a new strategy to estimating model (1), referred to as *stochastic nonparametric envelopment of data* (StoNED). Our objective is to estimate the deterministic part of the model in a fully nonparametric fashion imposing a minimal set of assumptions, in the spirit of DEA. Similar to DEA, we estimate the shape of the frontier by exploiting the regularity properties from the microeconomic theory (i.e., continuity, monotonicity, and concavity of f), free of any distributional assumptions or assumptions about the functional form of f or its smoothness. However, in the cross-sectional setting it is impossible to distinguish between inefficiency and noise without imposing some distributional assumptions (see Hall and Simar, 2002, for a detailed analysis). Having estimated the shape of function f , we make use of the standard distributional assumptions adopted from the SFA literature to estimate the expected location of the frontier f , and the firm-specific conditional expected values for the inefficiency term. In summary, the StoNED method consists of two-steps:

Step 1: Estimate the shape of function f by Convex Nonparametric Least Squares (CNLS) regression

Step 2: Using residuals of the CNLS regression, estimate the variance parameters σ_u^2, σ_v^2 by using the method of moments or pseudolikelihood techniques, and compute the conditional expected values of inefficiency

We elaborate the implementation of Steps 1 and 2 in Sections 3 and 4, respectively.

The main obstacle 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) have shown that

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

This implies that the StONED model (1) violates one of 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

$$y_i = [f(\mathbf{x}_i) - \mu] + [\varepsilon_i + \mu] = g(\mathbf{x}_i) + v_i, \quad i = 1, \dots, n, \quad (3)$$

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 $v_i \equiv \varepsilon_i + \mu$, $i = 1, \dots, n$, is a modified composite error term. It is easy to verify that function g inherits the monotonicity and concavity properties of f since μ is a constant, and that the modified errors v_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 (see Section 4).

Our two-step estimation strategy parallels the modified OLS (MOLS) approach to estimating parametric SFA models, originating from Aigner et al. (1977);³ our approach can be viewed as a nonparametric counterpart to MOLS. Although SFA models are commonly estimated by maximum likelihood (ML) techniques, MOLS provides a consistent method for estimating the SFA model. While the ML estimators are more efficient, provided that the heavy dose of functional form and distributional assumptions imposed in SFA are correct, the MOLS estimators are more robust to violations of the distributional assumptions about inefficiency terms u_i and noise v_i . Note that in MOLS the distributional assumptions about the composite error term do not influence the parameter estimates of f obtained in Step 1. We consider this robustness to distributional assumptions as a very attractive property, especially in the present nonparametric setting.⁴ As mentioned in the introduction, Fan et al. (1996) have earlier explored a parallel two-step approach in the context of kernel estimation.

³ MOLS should not be confused with the deterministic COLS (=corrected OLS) approach (Greene, 1980), where the frontier is shifted upward according to the largest OLS residual so as to envelop all observations.

⁴ Similar two-step estimation strategies are frequently employed in other areas of econometrics where efficient ML estimators are also available; consider e.g. Heckman’s (1979) celebrated model of sample selection (see also Ahn and Powell, 1993, and Das et al., 2003, for semi- and nonparametric variants of sample selection models).

3. CNLS estimators

This section elaborates the Step 1 of the two-step estimation procedure outlined in the previous section. Our purpose is to estimate the shape of the production function f by CNLS.

To estimate the average production function g , we propose to employ CNLS regression, which is particularly suited for the estimation of the StoNED model because it draws its power from the monotonicity and concavity conditions (which are the key maintained assumptions in DEA models) without any further assumptions about the functional form or its smoothness. This approach also circumvents the bias-variance tradeoff associated with other nonparametric regression techniques (such as kernel or spline techniques) (e.g., Yatchew 2003).

The CNLS problem can be formally stated as

$$\min_g \sum_{i=1}^n (y_i - g(\mathbf{x}_i))^2 \text{ s.t. } g \in F_2. \quad (4)$$

In words, the CNLS 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 by Hildreth (1954). Hanson and Pledger (1976) proved consistency of estimator (4) in the single input case. Nemirovskii et al. (1985), Mammen (1991) and Mammen and Thomas-Agnen (1999) have established the nonparametric convergence rate $O_p(n^{-1/(2+m)})$, and Groeneboom et al. (2001) derived the asymptotic distribution at a fixed point. If one imposes further smoothness assumptions, 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, imposing further smoothness assumptions would spoil the connection to DEA. Therefore, we here restrict to the non-smooth CNLS estimator.

The CNLS problem (4) 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 (4) generally hard to solve. The recent paper by Kuosmanen (2008) has shown that the infinite dimensional CNLS problem (4) has an equivalent finite dimensional representation, which can be stated as the following quadratic programming (QP) problem

$$\begin{aligned} \min_{\mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\beta}} \sum_{i=1}^n v_i^2 \\ y_i = \alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i + v_i \\ \alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i \leq \alpha_h + \boldsymbol{\beta}'_h \mathbf{x}_i \quad \forall h, i = 1, \dots, n \\ \boldsymbol{\beta}_i \geq 0 \quad \forall i = 1, \dots, n \end{aligned} \quad (5)$$

The first constraint of the least squares problem (5) can be interpreted as the regression equation. Note that coefficients $\alpha_i, \boldsymbol{\beta}_i$ are specific to each observation $i: i = 1, \dots, n$, which reveals a technical similarity to the random parameters SFA models (e.g., Greene, 2005). In the present setting, coefficients $\alpha_i, \boldsymbol{\beta}_i$ are not parameters of the estimated function g , but rather, they characterize tangent hyperplanes to an

unknown $g(\mathbf{x}_i)$. The inequality constraints in (5) can be interpreted as a system of Afriat inequalities (compare with Afriat, 1967, 1972; and Varian, 1984). When all inequalities of (5) are satisfied, we can employ the Afriat's Theorem to show that there exist a continuous, monotonic increasing and concave function \hat{g} that satisfies $y_i = \hat{g}(\mathbf{x}_i) + v_i$ for all $i = 1, \dots, n$. As Kuosmanen (2008) emphasizes, the Afriat inequalities are the key to modeling the concavity axiom in the general multiple regression setting where there is no unambiguous way of sorting input vectors \mathbf{x} . Interestingly, problems (4) and (5) are equivalent in the following sense (see Kuosmanen 2008, Appendix 1, for a formal proof).

Theorem 3.1: *Denote the optimal solution to the infinite dimensional CNLS problem (4) by s_{CNLS}^2 and the optimal solution to the finite quadratic programming problem (5) by s_{QP}^2 . Then for any real-valued data, $s_{CNLS}^2 = s_{QP}^2$.*

Given the estimated tangent coefficients $\hat{\alpha}_i, \hat{\beta}_i$ from (5), we construct the following piece-wise linear estimator for production function g

$$\hat{g}(\mathbf{x}) \equiv \min_{i \in \{1, \dots, n\}} (\hat{\alpha}_i + \hat{\beta}_i' \mathbf{x}). \quad (6)$$

This piece-wise linear estimator is legitimized by the following result.

Corollary to Theorem 3.1: *Denote the set of functions that minimize the CNLS problem (4) by $F_2^* : F_2^* \subset F_2$. For any real-valued data, the function \hat{g} characterized by (6) and (5) is one of the optimal solutions to problem (4), that is, $\hat{g} \in F_2^*$.*

The representor \hat{g} and its coefficients $(\hat{\alpha}_i, \hat{\beta}_i)$ have a compelling economic interpretation: vector $\hat{\beta}_i$ can be interpreted as the subgradient vector $\nabla g(\mathbf{x}_i)$, and thus it represents the vector of marginal products of inputs at point \mathbf{x}_i . Thus, coefficients $\hat{\beta}_i$ can be used for nonparametric estimation of substitution and scale elasticities. Equation $y = \hat{\alpha}_i + \hat{\beta}_i' \mathbf{x}$ can be interpreted as the tangent hyperplane to function g at point \mathbf{x}_i . Therefore, function \hat{g} provides a local first-order Taylor series approximation to any $f \in F_2^*$ in the neighborhood of the observed points \mathbf{x}_i . In contrast to the flexible functional forms that can be interpreted as second-order Taylor approximations around a single, unknown expansion point, CNLS uses all n observations as expansion points for the local linear approximation.

Note that the optimal solution to the CNLS problem (4) is not necessarily unique; there generally exists a family of alternate optima F_2^* . Similar to DEA models, the optimal solution to the QP

problem (5) need not be unique either. Therefore, Kuosmanen (2008) derived the following bounds for the alternate optima within F_2^* :

$$\hat{g}_{\min}(\mathbf{x}) = \min_{\alpha \in \mathbb{R}, \beta \in \mathbb{R}_+^m} \{ \alpha + \beta' \mathbf{x} \mid \alpha + \beta' \mathbf{x}_i \geq \hat{y}_i \quad \forall i = 1, \dots, n \}, \quad (7)$$

$$\hat{g}_{\max}(\mathbf{x}) = \max_{\phi \in \mathbb{R}, \alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^{m \times n}} \{ \phi \mid \phi \leq \alpha_i + \beta_i' \mathbf{x} \quad \forall i; \alpha_i + \beta_i' \mathbf{x}_i = \hat{y}_i \quad \forall i; \alpha_i + \beta_i' \mathbf{x}_h \geq \hat{y}_h \quad \forall h \neq i \}, \quad (8)$$

where $\hat{y}_i = \hat{g}(\mathbf{x}_i) = y_i - \hat{v}_i$, $i = 1, \dots, n$, denote the fitted values of the dependent variable. Kuosmanen (2008, Theorem 4.1) has shown that, for any finite real-valued data, function \hat{g}_{\min} is the tightest possible lower bound for the family of functions F_2^* (i.e., $\hat{g}_{\min}(\mathbf{x}) = \min_f f(\mathbf{x})$ s.t. $f \in F_2^*$), and \hat{g}_{\max} is the tightest possible upper bound (i.e., $\hat{g}_{\max}(\mathbf{x}) = \max_f f(\mathbf{x})$ s.t. $f \in F_2^*$). Note that for the observed points \mathbf{x}_i , the fitted values are always unique: $g(\mathbf{x}_i) = \hat{g}_{\min}(\mathbf{x}_i) = \hat{g}_{\max}(\mathbf{x}_i) \quad \forall i = 1, \dots, n$.

We can utilize the lower bound (7) for deriving an explicit connection between CNLS and the deterministic DEA method. The DEA frontier is defined as the tightest envelopment of observations subject to the conditions that the frontier must be monotonic increasing and globally concave (see e.g. Seiford and Thrall, 1990). Formally, the DEA frontier can be defined as

$$\hat{f}_{DEA}(\mathbf{x}) = \max_{z \in \mathbb{R}_+^n} \left\{ \sum_{h=1}^n z_h y_h \mid \mathbf{x} \geq \sum_{h=1}^n z_h \mathbf{x}_h; \sum_{h=1}^n z_h = 1 \right\}. \quad (9)$$

Interestingly, if we assume away the stochastic noise and interpret the CNLS residuals v_i as inefficiency, the DEA frontier (9) is obtained as a constrained special case of the CNLS lower bound estimator (7). More specifically, suppose we insert in the QP problem (5) an additional sign constraint for the residuals: $v_i \leq 0 \quad \forall i = 1, \dots, n$. Denote the fitted values of the resulting sign-constrained CNLS model by \hat{y}_i^{SC} . Then we obtain the following result (see Appendix for the proof).

Theorem 3.2: The DEA frontier estimator (9) is equivalent to the lower bound estimator (7) applied to the sign-constrained CNLS model (i.e., the QP problem (5) subject to $v_i \leq 0 \quad \forall i = 1, \dots, n$). Specifically,

$$\hat{f}_{DEA}(\mathbf{x}) = \hat{g}_{\min}^{SC}(\mathbf{x}) \equiv \min_{\alpha \in \mathbb{R}, \beta \in \mathbb{R}_+^m} \{ \alpha + \beta' \mathbf{x} \mid \alpha + \beta' \mathbf{x}_i \geq \hat{y}_i^{SC} \quad \forall i = 1, \dots, n \} \quad \forall \mathbf{x} \in \mathbb{R}^m. \quad (10)$$

This result complements Theorem 3.1 by Kuosmanen and Johnson (2009) in establishing a direct link between CNLS and DEA. Not only these two methods share the same shape constraints, the DEA frontier can be interpreted as a constrained special case of CNLS. This result offers a new compelling reinterpretation of DEA as sign-constrained nonparametric regression. This result also reveals an interesting connection between DEA and the classic parametric programming (PP) approach by Aigner and Chu (1968). Specifically, PP imposes the sign constraint $v_i \leq 0$ to the residuals of the linear regression model, whereas in our re-interpretation, DEA applies exactly the same sign-constraints

to the residuals of the shape-constrained nonparametric regression. Therefore, DEA can be seen as a nonparametric counterpart to Aigner and Chu's PP model. This result further implies that the StONED model proposed in this paper is a stochastic extension of DEA in the same way as SFA is a stochastic extension of Aigner and Chu's deterministic PP model.

Despite these compelling links between CNLS, DEA, and PP, the piece-wise linear estimator $\hat{g}(\mathbf{x})$ or the lower bound $\hat{g}_{\min}(\mathbf{x})$ do not estimate the frontier $f(\mathbf{x})$ but the average production function $g(\mathbf{x})$. However, the shape of the average production function $g(\mathbf{x})$ must be exactly the same as that of the frontier $f(\mathbf{x})$ because $g(\mathbf{x}) \equiv f(\mathbf{x}) - \mu$, where the expected inefficiency $\mu = \sigma_u \sqrt{2/\pi}$ is a constant by assumption. In the next section we show how the expected inefficiency μ and the unknown standard deviations σ_u, σ_v can be estimated from the CNLS residuals.

4. Efficiency estimation

Given the CNLS residuals $\hat{\mathbf{v}} \equiv (\hat{v}_1, \dots, \hat{v}_n)$, the next challenge is to disentangle inefficiency from noise. At least two possible approaches are known in the literature: the method of moments and pseudolikelihood estimation. We next briefly describe both these approaches and adapt them for our purposes.

4.1. Method of moments

Originating from the seminal paper by Aigner et al. (1977), the method of moments (MM) is commonly used in the MOLS estimation of SFA models (e.g. Greene, 2008). 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

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

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

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

$$\hat{M}_2 = \sum_{i=1}^n (\hat{v}_i - \hat{E}(v_i))^2 / n \quad (13)$$

$$\hat{M}_3 = \sum_{i=1}^n (\hat{v}_i - \hat{E}(v_i))^3 / n \quad (14)$$

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

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

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

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

These MM estimators are unbiased and consistent (Aigner et al., 1977; Greene, 2008), but not necessarily as efficient as the maximum likelihood estimators.

4.2. 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). 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 CNLS residuals $\hat{\mathbf{v}} \equiv (\hat{v}_1, \dots, \hat{v}_n)$. 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 expressed as a function of a single parameter (λ) as,

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

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

$$\hat{\sigma} = \left\{ \frac{1}{n} \sum_{j=1}^n \hat{v}_j^2 / \left[1 - \frac{2\lambda^2}{\pi(1+\lambda)} \right] \right\}^{1/2} . \quad (19)$$

Note that $\hat{\varepsilon}_i$ and $\hat{\sigma}$ cannot be computed from the CNLS 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 $\hat{\sigma}_v = \hat{\sigma}/(1+\hat{\lambda})$. Fan et al. (1996) show that estimators $\hat{\lambda}$ and $\hat{\sigma}$ converge to the true λ and σ at parametric rate $n^{-1/2}$.

4.3. 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

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

In practice, this means that frontier is obtained by shifting the CNLS 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

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

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

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

where $\hat{\varepsilon}_i = \hat{v}_i - \hat{\sigma}_u \sqrt{2/\pi}$ is the estimator of the composite error term (compare with (18)), not the CNLS 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.

4.4. Statistical inference

Even though the statistical distributions of the inefficiency and noise terms are known (by assumption), the conventional methods of statistical inference do not directly apply to 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, 2009).

In light of these complications, parametric bootstrap appears to be the best suited approach to statistical inference in the present context. Simar and Wilson (2009) have developed a bootstrap procedure for SFA. Their bootstrap algorithm can be adapted for the StONED model as follows:

[1] Given the sample data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, solve problem (5) to obtain estimates $\{(\hat{\alpha}_i, \hat{\beta}_i, \hat{v}_i)\}_{i=1}^n$. Use the

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

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

[3] Using the pseudo-data $\{(\mathbf{x}_i, \tilde{y}_i)\}_{i=1}^n$, solve problem (5) to obtain estimates $\{(\hat{\alpha}_i, \hat{\beta}_i, \hat{v}_i)\}_{i=1}^n$. Use the residuals $\{\hat{v}_i\}_{i=1}^n$ and the MM or PSL methods to obtain bootstrap estimates $\hat{\sigma}, \hat{\lambda}, \hat{\sigma}_u, \hat{\sigma}_v$.

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

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

Practical usefulness of the bootstrap procedure is further highlighted by the fact that the least-squares residuals are often skewed in the wrong direction ($\hat{M}_3 > 0$) in empirical applications (e.g., Carree, 2002). 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, 2008). However, several Monte Carlo simulations show that wrongly skewed residuals can frequently arise even in correctly specified SFA models (e.g., Fan et al., 1996; Simar and Wilson, 2009; see also Carree, 2002). Thus, wrongly skewed residuals are also likely to occur in correctly specified StoNED models. This is not only a problem for the method of moments, it equally affects the pseudolikelihood method. In this respect, it is comforting to note that the bootstrap approach described above can provide useful information about the inefficiency levels even when the residuals are wrongly skewed (see Simar and Wilson, 2009).

5. Extensions

5.1 Panel data model

Panel data enables us to relax the distributional assumptions, and estimate the model in a fully nonparametric fashion. In the following we describe the fixed effects approach to estimating time-invariant inefficiency. Alternative panel data approaches such as random effects modeling, time-varying inefficiency, and modeling technical progress are left as interesting topics for future research.

Assuming a balanced panel where each firm is observed over time periods $t = 1, \dots, T$, the StoNED model with time-invariant inefficiency can be described as

$$y_{it} = f(\mathbf{x}_{it}) - u_i + v_{it}, \quad i=1, \dots, n; \quad t=1, \dots, T, \quad (23)$$

where $u_i \geq 0$ is a time-invariant inefficiency term of firm i and v_{it} is the stochastic noise term for firm i in period t . Production function f is assumed to be monotonic increasing and concave as before. We assume that the noise components v_{it} are uncorrelated random variables with $E(v_{it})=0 \forall i, t$, $E(v_{it}^2) = \sigma_v^2 < \infty \forall i, t$, and $E(v_{js}v_{it})=0 \forall j \neq i, s \neq t$. Importantly, no distributional assumptions are imposed: the panel model (23) is fully nonparametric.

To estimate model (23) by CNLS, we first rephrase it as

$$y_{it} = [f(\mathbf{x}_{it}) - \mu] + [v_{it} - u_i + \mu] = g(\mathbf{x}_{it}) + v_{it}, \quad i=1, \dots, n; \quad t=1, \dots, T, \quad (24)$$

where $\mu \equiv \sum u_i / n$ is the average inefficiency, $g(\mathbf{x}) \equiv f(\mathbf{x}) - \mu$ is the average production function, and $v_{it} \equiv v_{it} - u_i + \mu$ is the modified composite error term. Note that function g inherits the monotonicity and concavity properties of f , and that the modified errors v_{it} satisfy the Gauss-Markov conditions. Thus, the average production function g can be consistently estimated by CNLS. Adapting the cross-sectional CNLS estimator to the panel data, we rewrite the CNLS problem (5) as

$$\begin{aligned} \min_{\alpha, \beta, u, v} \quad & \sum_{t=1}^T \sum_{i=1}^n v_{it}^2 \\ y_{it} = & \alpha_{it} + \beta'_{it} \mathbf{x}_{it} + v_{it} \quad \forall i=1, \dots, n; t=1, \dots, T \\ \alpha_{it} + \beta'_{it} \mathbf{x}_{it} \leq & \alpha_{hs} + \beta'_{hs} \mathbf{x}_{it} \quad \forall h, i \in \{1, \dots, n\}; s, t \in \{1, \dots, T\} \\ \beta_{it} \geq 0 \quad & \forall i=1, \dots, n; t=1, \dots, T \end{aligned} \quad (25)$$

The piece-wise linear CNLS estimator of g is obtained analogous to (6) as

$$\hat{g}(\mathbf{x}) \equiv \min_{i \in \{1, \dots, n\}, t \in \{1, \dots, T\}} (\hat{\alpha}_{it} + \hat{\beta}'_{it} \mathbf{x}). \quad (26)$$

Given the optimal solution to (25), we can compute the average residual of firm i as

$$r_i \equiv \sum_{t=1}^T v_{it} / T. \quad (27)$$

Applying the nonparametric inefficiency estimator of Schmidt and Sickles (1984) (see also Greene, 1980), we can estimate the firm-specific inefficiency terms u_i by

$$\hat{u}_i = \max_{h \in \{1, \dots, n\}} r_h - r_i. \quad (27)$$

If the sampling procedure is such that efficient firms with $u_i = 0$ are observed with a strictly positive probability, then \hat{u}_i provides a consistent estimator of u_i .

5.2. 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. Imposing RTS is straightforward in the QP problems (5) and (25). In (5) we can simply add the following constraints:

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

Rationale of these constraints is directly analogous to the standard multiplier-side DEA formulations where parallel constraints are employed for enforcing RTS assumptions.

While the CNLS 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 estimated 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.

5.3. Multiplicative model

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 CRS assumption requires a multiplicative error structure. Moreover, multiplicative error specification might be a useful remedy to heteroskedasticity from different scale sizes.

Adhering to the standard multiplicative formulation from SFA, we can rephrase model (1) as

$$y_i = f(\mathbf{x}_i) \cdot \exp(\varepsilon_i) = f(\mathbf{x}_i) \cdot \exp(v_i - u_i) \quad , i = 1, \dots, n. \quad (28)$$

We maintain the same assumptions on production function f and the composite error term as in model (1). Applying the log-transformation to equation (28), we obtain

$$\ln y_i = \ln f(\mathbf{x}_i) + \varepsilon_i \quad , i = 1, \dots, n. \quad (29)$$

Note that the log-transformation is applied to function f , not directly to inputs \mathbf{x} . Next, we may apply the decomposition presented in (3) to restore the Gauss-Markov conditions, rephrasing model (29) as

$$\ln y_i = [\ln f(\mathbf{x}_i) - \mu] + [\varepsilon_i + \mu] = g(\mathbf{x}_i) + v_i, \quad i = 1, \dots, n, \quad (30)$$

where μ is the expected inefficiency and g is the average production function as before. To estimate g by CNLS, we may rephrase the QP problem (5) as

$$\begin{aligned} \min_{\hat{y}, \alpha, \beta} & \sum_{i=1}^n (\ln y_i - \ln \hat{y}_i)^2 \\ \hat{y}_i &= \alpha_i + \beta'_i \mathbf{x}_i \\ \alpha_i + \beta'_i \mathbf{x}_i &\leq \alpha_h + \beta'_h \mathbf{x}_i \quad \forall h, i = 1, \dots, n \\ \beta_i &\geq 0 \quad \forall i = 1, \dots, n \end{aligned} \quad (31)$$

This yields a convex programming problem with a convex objective function and a system of linear inequality constraints. Note that the fitted values \hat{y}_i are model variables in (31). Although the objective function involves logarithms of model variables, global convexity of the objective function of problem

(31) presents an important advantage compared to the constrained ML problem suggested by Banker and Maindiratta (1992). With today's computational capacity, convex programming problems are considered not less tractable than linear programming.

Given the composite residuals from model (31) (i.e., $v_i = \ln y_i - \ln \hat{y}_i$), the standard MM or PSL procedures can be applied, as described in Section 4. The log-transformation only concerns Step 1, and makes no difference in the estimation of Step 2. However, the interpretation of inefficiency term u_i changes: $\exp(u_i)$ provides the Farrell output efficiency measure.

5.4 Cost functions

We can make use of the previous developments in the estimation of cost functions. Duality theory has established that the production technology can be equivalently modeled by means of monetary representations, such as the cost function, which is formally defined as

$$C(y, \mathbf{w}) = \min_{\mathbf{x}} \{ \mathbf{w}'\mathbf{x} \mid f(\mathbf{x}) = y \}. \quad (32)$$

Vector \mathbf{w} represents the exogenously given input prices. The cost function indicates the minimum cost of producing a given target output at given input prices. According to the microeconomic theory, it must be non-negative, non-decreasing, homogenous of degree one, concave and continuous in prices \mathbf{w} . These known properties provide a sound rationale for the nonparametric estimation.

In the stochastic cost frontier model, the observed costs C_i ($i = 1, \dots, n$) 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). To ensure homogeneity of degree one, we postulate a multiplicative error term as in Section 5.3, that is,

$$C_i = C(y_i, \mathbf{w}_i) \cdot \exp(\varepsilon_i) = C(y_i, \mathbf{w}_i) \cdot \exp(v_i + u_i). \quad (33)$$

Maintaining the earlier assumptions of u_i and v_i (note the changed sign of the inefficiency term), the cost frontier can be estimated analogous to the production function using the StoNED method.

The main challenge of the cost function estimation concerns the specification of the CNLS model to estimate the conditional expected values $E(C_i \mid \mathbf{w}_i, y_i)$. 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 \mathbf{w} . Thus, we need to transform the cost function as a concave (or convex) function of all its arguments. Note that if the cost function is a convex function of output, then it is a concave function of its additive inverse. To this end, we may rephrase equation (28) as

$$\ln C_i = [\ln C(y_i, \mathbf{w}_i) + \mu] + [v_i + u_i - \mu] = AC(-y_i, \mathbf{w}_i) + v_i, \quad (34)$$

where $AC(-y_i, \mathbf{w}_i) = \ln C(y_i, \mathbf{w}_i) + \mu$ is a concave function of all its arguments, and v_i is a modified error term that satisfies the Gauss-Markov assumptions. The average cost curve AC can be estimated by solving the CNLS problem:

$$\begin{aligned}
& \min_{\beta, \delta, \hat{c}} \sum_{i=1}^N (\ln C_i - \ln \hat{c}_i)^2 \\
& s.t. \\
& \hat{c}_i = \beta'_i \mathbf{w}_i + \delta_i(-y_i) \quad \forall i=1, \dots, n \\
& \beta'_i \mathbf{w}_i + \delta_i(-y_i) \leq \beta'_h \mathbf{w}_i + \delta_h(-y_i) \quad \forall h, i=1, \dots, n \\
& \beta'_i \geq \mathbf{0}, \delta_i \leq 0 \quad \forall i=1, \dots, n
\end{aligned} \tag{35}$$

Fitted values \hat{c}_i are model variables similar to problem (31), so problem (35) is a convex programming problem with linear constraints. Coefficients δ_i represent (the additive inverse of) the marginal cost of output, and are postulated to be non-positive. Coefficients β_i indicate the marginal cost of input prices (which depends on the input substitution possibilities). Note that model (35) excludes the intercept coefficients (compare with Section 5.1); by excluding the intercept terms and applying the multiplicative error term, we effectively force the estimated cost function to be homogenous of degree one, as required by the microeconomic theory.

Given the CNLS residuals, the conditional expected values of the inefficiency terms can be estimated along the lines described in Section 4. Note the changed sign of the inefficiency component and the direction of skewness. 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.

5.5 Heteroskedasticity

We have thus far assumed that standard deviations σ_u, σ_v are the same across all firms. This assumption is referred to as homoskedasticity, and it forms one of the maintained assumptions of the original SFA model by Aigner et al. (1977). As Florens and Simar (2005) have shown, violation of the homoskedasticity assumption leads to potentially serious problems in the context of parametric frontier estimation, and similar problems are likely to carry over to the present framework. Thus, a brief discussion about robustness of the proposed method to heteroskedasticity is necessary, even though more systematic treatment of the topic deserves a separate paper.

Firstly, we must distinguish between 1) heteroskedasticity of the noise term (i.e., parameter σ_v varies across firms) and 2) heteroskedasticity of the inefficiency term (i.e., σ_u varies across firms). Let us first consider heteroskedasticity of type 1). Of course, both types of heteroskedasticity may be present at the same time. However, their impacts on the StoNED estimators differ.

Note first that the expected inefficiency $\mu = \sigma_u \sqrt{2/\pi}$ does not depend on σ_v . Therefore, the shape of the average production function g remains identical to that of the frontier f even if the noise terms are heteroskedastic. Hence, the proposed approach is not particularly sensitive to heteroskedasticity of type 1). Least squares estimators (incl. CNLS) are known to be unbiased and consistent under symmetric heteroskedasticity, even though more efficient estimators are possible if

heteroskedasticity is modeled correctly. Given unbiased CNLS residuals, heteroskedastic σ_v will likely increase variance of the parameter estimators $\hat{\sigma}_u, \hat{\sigma}_v$. However, since σ_u is estimated based on the skewness of the residual distribution, and heteroskedasticity in the symmetric noise component does not affect skewness, the estimator $\hat{\sigma}_u$ remains consistent. Thus, frontier f and expected inefficiency μ can be consistently estimated even under heteroskedasticity of type 1). The only problem is that the conditional expected value of inefficiency $\hat{E}(u_i | \hat{\varepsilon}_i)$ is a function of heteroskedastic $\hat{\sigma}_v$. Thus, firm-specific efficiency scores and rankings can be affected by heteroskedasticity of type 1).

Heteroskedasticity of type 2) is a much more serious problem because σ_u does directly influence the expected inefficiency $E(u_i)$. When σ_u is heteroskedastic, the expected inefficiency $E(u_i)$ differs across firms, and thus the shape of the average production function g is no longer identical to that of the frontier f . We stress that this problem arises only in case 2), not in case 1). Since the proposed StoNED method relies on consistent estimation of the average production g in the step 1), the estimates can be sensitive to the violation of the homoskedasticity assumption for σ_u (see the next section for some evidence from Monte Carlo simulations). Therefore, it is critically important to develop statistical tests of the homoskedasticity assumption and more general estimation methods that can deal with heteroskedastic inefficiency. Fortunately, such tests and methods have been developed for the least squares estimation in the context of the linear regression model (consider, e.g., the generalized least squares (GLS) method). The main challenge is to adapt and extend existing techniques from the linear regression analysis to the CNLS framework. This forms a fascinating topic for future research.

6. Monte Carlo simulations

In this section we examine performance of the StoNED method in the controlled environment of Monte Carlo simulations. Our objective is to compare performance of the StoNED method with the standard DEA and SFA under alternative conditions where the distributional assumptions of the StoNED model are violated.⁵ The data generating processes used in the simulations has been adopted from Simar and Zelenyuk (2008). Systematic performance comparisons with other semi- and nonparametric methods is left as a topic for future research.⁶

We consider performance in terms of the standard mean squared error (MSE) criterion, applying it to estimates of the frontier f and the inefficiency term u . For the frontier estimates, the MSE statistic is defined as

⁵ For an illustrative example of the functioning and performance of the method with simulated data under ideal conditions, see Appendix 2. Further examples are available in the working papers Kuosmanen (2006) and Kuosmanen and Kortelainen (2007).

⁶ Since we replicate some of the simulations conducted by Simar and Zelenyuk (2008), an interested reader may compare our results with those reported by Simar and Zelenyuk for their local maximum likelihood estimator. However, it is worth to note that the synthetic data sets used in the different simulations are not exactly identical, but each random draw from the DGP yields unique data, which may have effect on the performance of estimators. The results reported here are averages over 50 replications of each scenario, whereas Simar and Zelenyuk (2008) report results of a single simulation run for each scenario.

$$MSE_f = \frac{1}{nR} \sum_{r=1}^R \sum_{i=1}^n (\hat{f}_r(\mathbf{x}_i) - f(\mathbf{x}_i))^2 ,$$

where \hat{f} denotes the estimated frontier function (estimated by DEA, SFA, or StoNED), and $r = 1, \dots, R$ is the index of replications of a given scenario. Analogously, the MSE of the inefficiency estimates is defined as

$$MSE_u = \frac{1}{nR} \sum_{r=1}^R \sum_{i=1}^n (\hat{u}_{i,r} - u_i)^2 .$$

For DEA, the standard output-oriented variable returns to scale (VRS) specification is used. Given the DEA efficiency score $\theta = \hat{f}^{DEA}(\mathbf{x}_i) / y_i$, the DEA inefficiency estimator is obtained as $\hat{u}_i^{DEA} = (\theta - 1)y_i$. For SFA, we use the Cobb-Douglas production function with the half-normal inefficiency term. The MOLS estimator is used to ensure the best comparability with the StoNED method. For the StoNED method, we assume the multiplicative specification (28) and the half-normal inefficiency distribution. Since the MC simulations are computationally intensive, we restrict to the simpler method of moment (MM) estimator in this section. In the MM estimation of SFA and StoNED models, we have dealt with the wrong skewness problem as follows. If \hat{M}_3 is non-negative, we set $\hat{M}_3 = -0.0001$. On the other hand, if $\hat{\sigma}_v$ is negative, we set $\hat{\sigma}_v = 0.0001$. These settings ensure that the algorithm runs smoothly even in those scenarios where the DGP is inconsistent with the model assumptions (e.g., there are outliers or no inefficiency). Of course, the wrong skewness can be a signal of model misspecification (e.g., in scenarios involving outliers), but in these MC simulations we disregard this potentially useful information and force the postulated skewness to the estimated distributions of the composite error term.

6.1 Univariate Cobb-Douglas frontier

We start by replicating the first six scenarios of Simar and Zelenyuk (2008) as reported in their Section 3.1.1. The DGP is characterized by the univariate Cobb-Douglas model

$$y_i = x_i^{0.5} \cdot \exp(-u_i) \cdot \exp(v_i) ,$$

where $x_i \sim Uni[0,1]$, $u_i \sim Exp[\mu = 1/6]$ with parameter μ representing the mean inefficiency, and $v_i \sim N(0, \sigma_v^2)$ where $\sigma_v = \rho_{ns} \cdot \mu$ and parameter ρ_{ns} represents the noise-to-signal ratio. Using this DGP, Simar and Zelenyuk construct six alternative scenarios corresponding to different values of sample size n and parameter ρ_{ns} . Before proceeding to the results, we note that the SFA model assumes the correct functional form for the frontier. However, both SFA and StoNED models assume a wrong distribution for the inefficiency term.

Table 1 describes the six scenarios and reports the average MSEs over 50 replications for the frontier estimates. We note first that the results for the DEA frontier estimator come reasonably close to

those reported by Simar and Zelenyuk (2008). We see that the SFA estimator has a larger MSE than DEA in scenario a) that does not involve any noise whatsoever, but it performs considerably better than DEA in other scenarios involving outliers or noise. Interestingly, the StoNED estimator has a lower MSE than the SFA estimator in all scenarios, even though the functional form of SFA is correct.

Table 1: Performance in estimating frontier f ; univariate C-D frontier

Scenario	Description	MSE _{DEA}	MSE _{SFA}	MSE _{StoNED}
a)	$n = 100, \rho_{nts} = 0$	0.0002	0.0060	0.0052
b)	$n = 103, 3$ outliers	0.0999	0.0068	0.0064
c)	$n = 100, \rho_{nts} = 1$	0.0398	0.0070	0.0067
d)	$n = 200, \rho_{nts} = 1$	0.0640	0.0068	0.0067
e)	$n = 500, \rho_{nts} = 1$	0.0966	0.0058	0.0057
f)	$n = 500, \rho_{nts} = 2$	0.7053	0.0077	0.0075

Table 2 reports the corresponding statistics for the inefficiency estimates. Interestingly, while the DEA estimator captures the frontier better than SFA or StoNED in scenario a) that involves no noise, the DEA inefficiency estimator has a higher MSE than the two stochastic alternatives. While the SFA and StoNED estimators over-estimate the frontier when the true DGP has no noise, in the case of efficiency estimation, attributing a part of the total variance to the noise term will tend to offset the upward bias in the frontier estimation. This explains the better performance of SFA and StoNED in efficiency estimation in scenario a). On the other hand, in the noisy scenarios, the advantages of SFA and StoNED are not so great in terms of inefficiency estimates as they are in the case of frontier estimation. Estimating inefficiency at the firm level in a cross-sectional setting is a notoriously challenging task when both the frontier and the evaluated input-output vector are subject to noise.

Table 2: Performance in estimating inefficiency term u ; univariate C-D frontier

Scenario	Description	MSE _{DEA}	MSE _{SFA}	MSE _{StoNED}
a)	$n = 100, \rho_{nts} = 0$	0.0161	0.0109	0.0097
b)	$n = 103, 3$ outliers	0.0854	0.0322	0.0317
c)	$n = 100, \rho_{nts} = 1$	0.0424	0.0294	0.0282
d)	$n = 200, \rho_{nts} = 1$	0.0600	0.0301	0.0288
e)	$n = 500, \rho_{nts} = 1$	0.0829	0.0265	0.0258
f)	$n = 500, \rho_{nts} = 2$	0.6236	0.0377	0.0362

6.2 Trivariate Cobb-Douglas frontier

We next extend the previous six scenarios to the three-input case, characterized by the Cobb-Douglas model

$$y_i = x_{1,i}^{0.4} \cdot x_{2,i}^{0.3} \cdot x_{3,i}^{0.2} \cdot \exp(-u_i) \cdot \exp(v_i),$$

where $x_{j,i} \sim \text{Uni}[0,1]$, $j = 1, 2$. The inefficiency and the noise terms are drawn in the identical manner to Section 6.1. The purpose of these scenarios is to examine how the curse of dimensionality might affect performances of alternative estimators.

Table 3 describes the six scenarios and reports the average MSEs over 50 replications for the frontier estimates. Firstly, we must emphasize that the MSEs reported in Table 3 are not directly comparable with those of Table 1 because the scale of output values is somewhat different. As expected, the DEA estimator performs best in scenarios a) and b) involving little or no noise. Its precision deteriorates dramatically when the noise to signal ratio increases. The MSEs of SFA and StoNED estimators are more stable across scenarios. The StoNED performs better than SFA in most scenarios, except for c) and f) that involve the largest noise to signal ratios at given sample sizes.

Table 3: Performance in estimating frontier f ; trivariate C-D frontier

Scenario	Description	MSE _{DEA}	MSE _{SFA}	MSE _{StoNED}
a)	$n = 100, \rho_{nts} = 0$	0.0014	0.0028	0.0020
b)	$n = 100, \rho_{nts} = 0.5$	0.0013	0.0028	0.0021
c)	$n = 100, \rho_{nts} = 1$	0.0063	0.0028	0.0029
d)	$n = 200, \rho_{nts} = 1$	0.0084	0.0037	0.0036
e)	$n = 300, \rho_{nts} = 1$	0.0137	0.0031	0.0028
f)	$n = 300, \rho_{nts} = 2$	0.1583	0.0073	0.0080

Table 4 presents the corresponding MSE statistics for the inefficiency estimates. Interestingly, the DEA estimator has large MSE in scenarios a) and b) where it estimates the frontier most accurately. Recall that the DEA estimator is by construction downward biased. The SFA and StoNED estimators overestimate the frontier in scenarios a) and b) involving little or no noise. However, this overshooting is offset in the efficiency estimation as these two methods attribute a part of the regression residual to the noise term. This explains why SFA and StoNED perform better than DEA in estimating the inefficiency term u than the frontier f in scenarios a) and b). Due to its better empirical fit, the StoNED estimator yields lower MSE than SFA in all scenarios, including scenario f).

Table 4: Performance in estimating inefficiency term u ; trivariate C-D frontier

Scenario	Description	MSE _{DEA}	MSE _{SFA}	MSE _{StoNED}
a)	$n = 100, \rho_{nts} = 0$	0.0334	0.0011	0.0010
b)	$n = 100, \rho_{nts} = 0.5$	0.0295	0.0163	0.0135
c)	$n = 100, \rho_{nts} = 1$	0.0283	0.0267	0.0250
d)	$n = 200, \rho_{nts} = 1$	0.0268	0.0309	0.0297
e)	$n = 300, \rho_{nts} = 1$	0.0284	0.0265	0.0262
f)	$n = 300, \rho_{nts} = 2$	0.1288	0.0512	0.0511

6.3 Trivariate Cobb-Douglas frontier with heteroskedastic inefficiency

We next adapt the DGP of the previous section by introducing heteroskedasticity in the inefficiency term u . Following Simar and Zelenyuk (2008) Section 3.1.4, we draw inefficiency terms from the half-normal distribution as $u_i | \mathbf{x}_i \sim \left| N(0, (\sigma_u (x_{1,i} + x_{2,i}) / 2)^2) \right|$, where $\sigma_u = 0.3$. Note that variance of inefficiency distribution depends on inputs 1 and 2, which results as heteroskedasticity. The noise term

is homoskedastic normal, $v_i \sim N(0, \sigma_v^2)$, where $\sigma_v = \rho_{nts} \cdot \sigma_u \cdot \sqrt{(\pi - 2)/\pi}$. Parameter ρ_{nts} can be interpreted as the average noise to signal ratio, and it is varied across scenarios.

Table 5 reports the average MSEs over 50 replications for the frontier estimates. The MSEs reported in Tables 3 and 5 are comparable as we have used the same production function, the same sample sizes, and the same noise to signal ratios; the only difference is the heteroskedastic inefficiency term. Interestingly, although DEA is a distribution-free method, MSEs of the DEA estimator increase notably. This is because observations with large values of inputs 1 and 2 are likely to have larger inefficiencies. This will directly affect the local DEA approximation of the frontier in the region where x_1 and x_2 are greater than 0.5. By contrast, the MSEs of the SFA estimator decrease in all scenarios. The SFA frontier is more rigid by construction, and hence less sensitive to local heteroskedasticity. Moreover, the SFA benefits from the correct half-normal distributional function of the inefficiency term in this case. Performance of the StoNED estimator deteriorates for the same reason as for DEA. While the StoNED estimator is more sensitive to local heteroskedasticity than SFA, its MSE remains lower than that of DEA in all noisy scenarios where the average noise to signal ratio is equal to one or higher.

Table 5: Performance in estimating frontier f ; trivariate C-D frontier with heteroskedastic inefficiency

Scenario	Description	MSE _{DEA}	MSE _{SFA}	MSE _{StoNED}
a)	$n = 100, \rho_{nts} = 0$	0.0036	0.0016	0.0042
b)	$n = 100, \rho_{nts} = 0.5$	0.0024	0.0015	0.0038
c)	$n = 100, \rho_{nts} = 1$	0.0051	0.0030	0.0051
d)	$n = 200, \rho_{nts} = 1$	0.0071	0.0017	0.0038
e)	$n = 300, \rho_{nts} = 1$	0.0067	0.0011	0.0023
f)	$n = 300, \rho_{nts} = 2$	0.0895	0.0036	0.0041

Table 6: Performance in estimating inefficiency term u ; trivariate C-D frontier with heteroskedastic inefficiency

Scenario	Description	MSE _{DEA}	MSE _{SFA}	MSE _{StoNED}
a)	$n = 100, \rho_{nts} = 0$	0.0574	0.0108	0.0192
b)	$n = 100, \rho_{nts} = 0.5$	0.0498	0.0191	0.0210
c)	$n = 100, \rho_{nts} = 1$	0.0439	0.0401	0.0363
d)	$n = 200, \rho_{nts} = 1$	0.0371	0.0370	0.0377
e)	$n = 300, \rho_{nts} = 1$	0.0358	0.0346	0.0335
f)	$n = 300, \rho_{nts} = 2$	0.0651	0.0629	0.0613

For completeness, Table 6 presents the corresponding MSEs of the inefficiency estimates. Compared to Table 4, the MSEs of all three methods increase. In particular, performances of SFA and StoNED deteriorate notably in all scenarios, but especially in a) and b) involving little or no noise. Still, SFA and StoNED outperform DEA in those two scenarios. As the sample size and the noise to signal ratio increase, the StoNED estimator becomes more competitive in comparison to SFA.

In conclusion, the proposed StoNED estimator proved a competitive alternative to the conventional DEA and SFA estimators in the simulations adopted from Simar and Zelenyuk (2008). We

emphasize that the distributional assumptions for the inefficiency term were incorrect in all scenarios that were considered. Despite this specification error, the StoNED estimator outperformed the distribution-free DEA estimator in 29 scenarios out of 36. This suggests that it is often preferable to model noise even at the risk of making specification error in the distributional assumptions of the inefficiency term than assume away noise completely. The StoNED estimator achieved lower MSE than the corresponding SFA estimator in 25 cases out of 36 even though the SFA estimator assumed the correct functional form for the frontier (the inefficiency term was wrongly specified exactly the same way as for the StoNED estimator). It appears that the better empirical fit in the estimation of the frontier can also partly offset the possible specification errors in the estimation of the inefficiency distribution. Of course, evidence from any Monte Carlo study is limited, and the present comparison only limits to the most basic variants of DEA and SFA. We recognize the need to compare the performance of the proposed method with other recently developed semiparametric and nonparametric approaches that were briefly reviewed in the Introduction, but we also realize that designing and implementing such a comparison of many computationally intensive methods in a fair and objective way is a daunting task that deserves a thorough investigation of its own.

7. Conclusions and discussion

We have developed a new encompassing framework for productive efficiency analysis, referred to as *Stochastic Nonparametric Envelopment of Data* (StoNED). The StoNED model combines a nonparametric DEA-like frontier with a stochastic SFA-like inefficiency and noise terms, melding the main virtues of both DEA and SFA into a unified model that can be estimated in practice. Importantly, both DEA and SFA can be viewed as special cases of StoNED, obtainable by imposing some more restrictive assumptions to the StoNED model.

To estimate the StoNED model, we employed a two-stage estimation strategy that is commonly used in many areas of econometrics, particularly in semi- and nonparametric estimation. In the first stage, the shape of the frontier is consistently estimated by using convex nonparametric least squares (CNLS), which does not assume any smoothing parameters, building upon the same shape constraints as DEA. In the second stage, we apply method of moments or pseudolikelihood techniques, adopted from the SFA literature, to disentangle the inefficiency and noise components from the CNLS residuals. Although this stepwise estimation strategy may not be as efficient as the constrained maximum likelihood, it has some important advantages, including the robustness to distributional assumptions about the inefficiency and noise terms, and substantially lower computational barriers (i.e., the constrained ML estimators are often computationally infeasible in the present setting). To our knowledge, the connection between CNLS regression and DEA has not been examined before. We have here established formal connections between CNLS, DEA, and parametric programming approaches to frontier estimation, proving that the DEA frontier is obtainable as a constrained special case of the CNLS regression. In light of these results, we consider CNLS regression (first proposed by Hildreth,

1954) as a long-sought missing link between DEA and SFA approaches in the evolution of productive efficiency analysis.

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 for alternative specifications of returns to scale. Moreover, performance of the StoNED approach was examined in the controlled environment of Monte Carlo simulations. The evidence from the simulations demonstrates that the proposed method is a competitive alternative to standard DEA and SFA methods even when the distribution of the inefficiency term is wrongly specified.

The proposed StoNED approach shares many common features with SFA and DEA, being an amalgam of the two. Thus, many of the existing tools and techniques for SFA and DEA can be readily incorporated into the StoNED framework. Further exploiting of the connections established in this paper provide many interesting opportunities for future research. The hybrid nature of StoNED also means that there are many important differences to both SFA and DEA, which must be kept in mind when applying the StoNED approach to empirical data. 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. We hope that this paper could inspire further theoretical and empirical work in this direction, and thus contribute to cross-fertilization and unification of the parametric and nonparametric streams of productive efficiency analysis in the future.

While the StoNED approach combines the appealing features of DEA and SFA, it also shares 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 distributional assumptions of SFA are rather *ad hoc*, and might often be inappropriate. 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. Other important extensions include relaxations of homoskedasticity and concavity assumptions that are used frequently in SFA and DEA applications. Addressing these shared limitations of DEA and SFA presents important challenges for future research. In this respect, we emphasize again that the focus of this paper has been on the development of an amalgam model that encompasses the classic DEA and SFA models as its special cases. Improving upon DEA and/or SFA aspects of the model is another issue, which falls beyond the scope of the present paper.

The present paper restricts to the estimation of a single-output technology. The ability of DEA to deal with multi-input multi-output technologies is often an important rationale for using that method. Extending the StoNED approach to the general multi-output technologies is another important avenue for future research.

Finally, panel data can offer a richer set of information, which is often utilized in SFA. In this

respect, applying the standard fixed effects approach to CNLS regression is straightforward, as discussed in the paper. Interpreting the fixed effects as firm specific inefficiency terms enables us to abolish the distributional assumptions about the inefficiency and noise term, resulting in a fully nonparametric approach to frontier estimation. However, appropriate modeling of technical progress, intertemporal efficiency changes, and heterogeneity across firms remain as major challenges. Thus, a more detailed examination of the panel data approach presents yet another fascinating topic for future research.

Acknowledgements

We thank Associate Editor and two anonymous reviewers of this journal for helpful comments and suggestions for improving the paper. An earlier version of this paper has been presented at EWEPA X 2007; 4th Nordic Econometric Meeting 2007; XXIX Annual Meeting of the Finnish Society for Economic Research 2007; Nordic Efficiency and Productivity Workshop 2006; and NAPW IV 2006. We are grateful to M. Browning, W.H. Greene, D. Henderson, D. Kristensen, S. Kumbhakar, R. Sickles, T. Sipiläinen, R. Svento, M. Vardanyan, and other participants to these workshops and seminars for commenting earlier versions of this paper. Special thanks to P. Agrell, K. Kerstens, and P. Vanden Eeckaut for the originality prize at EWEPA X, and R. Conditions for computational assistance.

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.
- Ahn, H., and J.L. Powell, 1993, Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism, *Journal of Econometrics* 58, 3-29.
- 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.
- Banker, R.D., and A. Maindiratta, 1992, Maximum Likelihood Estimation of Monotone and Concave Production Frontiers, *Journal of Productivity Analysis* 3, 401-415.
- Carree, M.A., 2002, Technological Inefficiency and the Skewness of the Error Component in Stochastic Frontier Analysis, *Economics Letters* 77, 101-107.
- 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.
- Das, M., W.K. Newey, and F. Vella, 2003, Nonparametric Estimation of Sample Selection Models, *Review of Economic Studies* 70, 3-58.

- 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, 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.
- Florens, J.P., and L. Simar, 2005, Parametric Approximations of Nonparametric Frontier. *Journal of Econometrics* 124(1), 91-116.
- Fried, H., C.A.K. Lovell, and S. Schmidt, 2008, *The Measurement of Productive Efficiency and Productivity Change*, Oxford University Press, New York.
- Greene, W., 1980, Maximum likelihood estimation of econometric frontier functions, *Journal of Econometrics* 13, 26-57.
- Greene, W.H., 2005, Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model, *Journal of Econometrics* 126, 269-303.
- Greene, W.H., 2008, The Econometric Approach to Efficiency Analysis, Chapter 2 in H Fried, K. Lovell and S. Schmidt, eds., *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press.
- Groeneboom, P., G. Jongbloed, and J.A. Wellner, 2001, Estimation of a Convex Function: Characterizations and Asymptotic Theory, *Annals of Statistics* 29, 1653-1698.
- Hall, P., and L. Simar, 2002, Estimating a Change-point, Boundary, or Frontier in the Presence of Observation Error, *Journal of the American Statistical Association* 97(458), 523-534.
- Hanson, D.L., and G. Pledger, 1976, Consistency in Concave Regression, *Annals of Statistics* 4(6), 1038-1050.
- Heckman, J., 1979, Sample Selection Bias as a Specification Error, *Econometrica* 47, 153-161.
- 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., B.U. Park, L. Simar, and E.G. Tsionas, 2007, Nonparametric Stochastic Frontiers: A Local Maximum Likelihood Approach, *Journal of Econometrics*, 137, 1-27.
- 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., 2008, Representation Theorem for Convex Nonparametric Least Squares, *Econometrics Journal* 11, 308-325.
- Kuosmanen, T. and A. Johnson, 2009, Data Envelopment Analysis as Nonparametric Least Squares Regression, *Operations Research*, to appear.
- Kuosmanen, T., and M. Kortelainen, 2007, Stochastic Nonparametric Envelopment of Data: Cross-sectional Frontier Estimation Subject to Shape Constraints, Univ. of Joensuu, Economics Discussion Paper No. 46.
- 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.
- 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. , 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.
- 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, 2009, Estimation and Inference in Cross-Sectional Stochastic Frontier Models, *Econometric Reviews*, in press.
- Simar, L., and V. Zelenyuk 2008, Stochastic FDH/DEA Estimators for Frontier Analysis. Discussion paper 0820, 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.
- 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., 2003, *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press.

Appendix 1: Proof of Theorem 3.2

Consider the following sign-constrained QP problem (i.e., add sign-constraint $v_i \leq 0$ to (5))

$$\begin{aligned}
 & \min_{\mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\beta}} \sum_{i=1}^n v_i^2 \\
 & y_i = \alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i + v_i \\
 & \alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i \leq \alpha_h + \boldsymbol{\beta}'_h \mathbf{x}_i \quad \forall h, i = 1, \dots, n \\
 & \boldsymbol{\beta}_i \geq 0 \quad \forall i = 1, \dots, n \\
 & v_i \leq 0 \quad \forall i = 1, \dots, n
 \end{aligned} \tag{A.1}$$

The optimal solution to (A.1) yields a unique set of fitted values $\hat{y}_i^{SC} = \alpha_i^* + \boldsymbol{\beta}'_i^* \mathbf{x}_i$, which can be used for constructing the lower bound function $\hat{g}_{\min}^{SC}(\mathbf{x}) \equiv \min_{\alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}_+^m} \{ \alpha + \boldsymbol{\beta}' \mathbf{x} \mid \alpha + \boldsymbol{\beta}' \mathbf{x}_i \geq \hat{y}_i^{SC} \quad \forall i = 1, \dots, n \}$.

Kuosmanen and Johnson (2009, Theorem 3.1) have shown that the fitted values \hat{y}_i^{SC} obtained from (A.1) are equivalent to those of the DEA estimator (9) for all observations $i = 1, \dots, n$, specifically,

$$\hat{f}_{DEA}(\mathbf{x}_i) = \hat{y}_i^{SC} = \hat{g}_{\min}^{SC}(\mathbf{x}_i) \quad \forall i = 1, \dots, n. \tag{A.2}$$

The proof of Theorem 3.2 requires that we extend this result to unobserved vectors $\mathbf{x} \in \mathbb{R}_+^m$.

Using duality theory of linear programming, one can verify that the lower-bound function $\hat{g}_{\min}^{SC}(\mathbf{x})$ has the following equivalent dual formulation

$$\hat{g}_{\min}^{SC}(\mathbf{x}) = \max_{\mathbf{z} \in \mathbb{R}_+^n} \left\{ \sum_{i=1}^n z_i \hat{y}_i \mid \mathbf{x} \geq \sum_{i=1}^n z_i \mathbf{x}_i; \sum_{i=1}^n z_i = 1 \right\}. \tag{A.3}$$

Comparing this dual formulation with (9), we note that the only difference between $\hat{g}_{\min}^{SC}(\mathbf{x})$ and the DEA frontier $\hat{f}_{DEA}(\mathbf{x})$ is that our lower bound function uses the fitted values \hat{y}_i^{SC} whereas the DEA frontier uses the observed y_i , $i = 1, \dots, n$. Note that for the efficient subset $Eff_{DEA} = \{i \mid y_i = \hat{f}_{DEA}(\mathbf{x}_i)\}$, equality $\hat{y}_i^{SC} = y_i$ holds. Moreover, the inefficient subset $Ineff_{DEA} = \{i \mid y_i < \hat{f}_{DEA}(\mathbf{x}_i)\}$ does not influence the DEA frontier at all. Therefore,

$$\hat{f}_{DEA}(\mathbf{x}) = \hat{g}_{\min}^{SC}(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbb{R}_+^m. \tag{A.4}$$

Appendix 2: Illustrative example

The purpose of this appendix is to illustrate the estimated StoNED frontiers graphically in a single-input single-output setting. Further examples and illustrations can be found in working papers Kuosmanen (2006) and Kuosmanen and Kortelainen (2007). Some computational codes for the GAMS and Matlab software are available at the homepage: <http://www.nomepre.net/stoned/>.

In the present example, the input data were randomly sampled from $Uni[1,11]$ for a random sample of 100 firms, independently for each input and firm. The efficient output levels were calculated using the production function $f(x_i) = \ln(x_i) + 2$. From the efficient output level, we subtracted a random inefficiency term $u_i \sim \left| N(0, \sigma_u^2) \right|$ and added a random error $v_i \sim N(0, \sigma_v^2)$, to obtain the “observed” output data used in estimation as $y_i = \ln(x_i) + 2 + v_i - u_i$. The standard deviations of the inefficiency and noise terms are $\sigma_u = 0.6$ and $\sigma_v = 0.3$.

We applied the shape constrained CNLS method with additive error structure without restrictions on RTS to this simulated data, and subsequently computed the MM and PSL estimators using the CNLS residuals. Figure 1 illustrates the results by plotting a scatter of the sample data (points \times), the true frontier (thick black curve), the CNLS 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 CNLS 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.

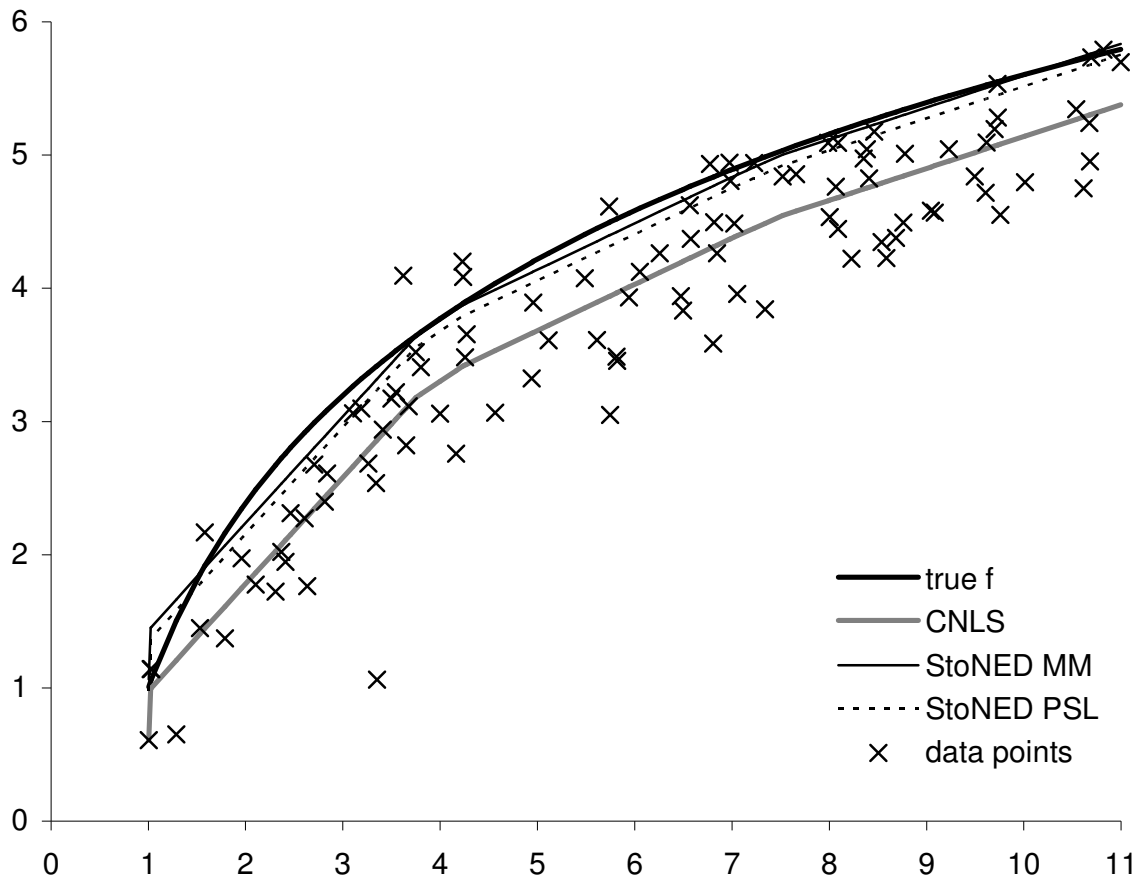


Figure 1: Graphical illustration of the CNLS regression curve and the StoNED frontiers.