Stochastic Nonparametric Envelopment of Panel Data:

Frontier Estimation with Fixed and Random Effects Approaches

Timo Kuosmanen

Economic Research Unit, MTT Agrifood Research Finland

Luutnantintie 13, 00410 Helsinki, Finland

E-mail. Timo.Kuosmanen@mtt.fi.

Abstract

Stochastic nonparametric envelopment of data (StoNED) combines the virtues of data envelopment analysis (DEA) and stochastic frontier analysis (SFA) into a unified framework of frontier estimation. StoNED melds the nonparametric piece-wise linear DEA-type frontier with stochastic SFA-type inefficiency and noise terms. We show that the StoNED model can be estimated in the panel data setting in a fully nonparametric fashion. Both fixed and random effects approaches are adapted to the StoNED framework. To disentangle changes in technology and efficiency, a dynamic semiparametric variant of the StoNED model is developed. An application to the wholesale and retail industry illustrates the approach.

Key Words: data envelopment analysis (DEA), nonparametric least squares, nonparametric regression, productive efficiency analysis, stochastic frontier analysis (SFA)

JEL Classification: C14, C51, D24

1. Introduction

The literature of productive efficiency analysis has been divided between two main 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 standard axioms of production theory: monotonicity, convexity, and homogeneity. The weakness of DEA is that it 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 a rigid ex ante specification of the functional form. Since the economic theory does not justify a particular functional form, the flexible functional forms, such as the translog or generalized McFadden, 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 axioms. On the other hand, imposing these axioms will sacrifice the flexibility (e.g., Diewert and Wales, 1987; Sauer, 2006). In summary, it is generally accepted that the virtues of DEA lie in its nonparametric treatment of the frontier, consistent with the axioms of production theory, 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., Simar, 1992; 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; and 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, there is no guarantee that these models satisfy the axioms of production theory. Therefore, there is an evident need for semi- and nonparametric stochastic frontier approaches that satisfy the standard axioms and thus combine the virtues of DEA and SFA in a unified framework of frontier estimation.

In the cross-sectional setting, Banker and Maindiratta (1992) were the first to propose an amalgam of DEA and SFA that combines a DEA-style nonparametric, convex, piecewise linear frontier with a SFA-style parametric composite error term consisting of noise and inefficiency components. However, their constrained maximum likelihood estimation procedure is extremely difficult to implement; no operational computational procedure or empirical applications have been reported. Recently, Kuosmanen (2006) and Kuosmanen and Kortelainen (2007) introduced the *stochastic nonparametric envelopment of data* (StoNED) model, which is an additive variant of Banker and Maindiratta's model. Kuosmanen and Kortelainen showed how the StoNED model can be estimated in practice by applying a two-stage procedure. In the first stage, the average practice frontier is estimated by means of nonparametric least squares subject to shape constraints (monotonicity, concavity, and/or homogeneity). In the second-stage, the standard deviations of the inefficiency and noise terms are estimated by the method of moments or pseudolikelihood techniques, and the conditional expected values of the inefficiency terms are computed. While the frontier is estimated nonparametrically, the procedure as a whole can be more precisely described as semiparametric.

Panel data offers generally better possibilities for disentangling inefficiency from stochastic noise than a cross-section: when each firm is observed many times, the effects of random noise can be averaged out. In the parametric SFA literature, the two main approaches for dealing with panel data are the fixed effects approach (Schmidt and Sickles 1984) and the random effects approach (Lee and Tyler

1978). In both these approaches, the distributional assumptions about the inefficiency and error terms can be relaxed. In the nonparametric literature, the opportunities provided by the panel data have been largely ignored, with a notable exception of Ruggiero (2004).

The purpose of this paper is to adapt the cross-sectional StoNED model to the panel data setting. The panel data enables us to average out the noise without any parametric assumptions about the distributions of the inefficiency and noise terms: we present the first fully nonparametric variant of the StoNED model. However, we also note that certain parametric assumptions about technical progress and the inefficiency term may prove useful for disentangling the frontier shifts from efficiency improvements over time.

Practical estimation of the StoNED model requires adapting the shape constrained nonparametric least squares estimator to the panel data setting. To our knowledge, this is the first paper to apply the fixed and random effects approaches to the shape constrained nonparametric least squares regression. We also present one of the first empirical applications of the shape constrained nonparametric least squares in the general multiple regression setting.

The remainder of the paper is organized as follows. Section 2 introduces the StoNED model in the cross-sectional and panel data settings. Section 3 discusses the fixed effects approach to estimating the StoNED model in the panel data setting. Section 4 describes the random effects approach. Section 5 illustrates the approach by means of a simulated example. Section 6 extends the StoNED model to account for intertemporary changes in the technology and efficiency, and derives a semiparametric estimator. Section 7 illustrates the semiparametric estimation by means of an application to industry-level panel data of wholesale and retail sectors in 14 OECD countries over the period 1975-2003. Section 8 draws the concluding remarks. For compactness, the proofs of the mathematical theorems are presented in Appendix A. Appendix B provides a GAMS code used for computing the CNLS problem of the application.

2. StoNED model

2.1 Cross sectional model

To gain intuition, we start from the cross-sectional StoNED model introduced by Kuosmanen (2006) and Kuosmanen and Kortelainen (2007). The *M*-dimensional input vector is denoted by **x** and the scalar output by *y*. The production technology is represented by the *production function* $y = f(\mathbf{x})$, where function *f* belongs to the class of continuous, 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; the production function is specified along the 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 $\mathbf{x}_i$. Kuosmanen and Kortelainen (2007) follow the standard SFA practice and assume $u_i \underset{i.i.d}{\sim} \left|N(0,\sigma_u^2)\right|$ and $v_i \underset{i.i.d}{\sim} N(0,\sigma_v^2)$. Banker and Maindiratta's (1992) model differs from (1) in that the composite errors are multiplicative (i.e., $y_i = f(\mathbf{x}_i)e^{v_i - u_i}$) and the distribution of the inefficiency term u_i is truncated normal.

Model (1) is referred to as the cross-sectional *stochastic nonparametric envelopment of data* (StoNED) model. It can be thought of 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. For estimation of model (1), a reader is referred to Kuosmanen and Kortelainen (2007).

2.2. Panel data model

We next adapt the cross-sectional StoNED model to the panel data setting where we assume a balanced data of *n* firms in *T* time periods. The panel variant of the StoNED model can be formally defined as

$$y_{it} = f(\mathbf{x}_{it}) - u_i + v_{it}, \;\; i = 1,...,n; \;\; t = 1,...,T \;, \tag{2}$$

where $u_i \geq 0$ is the inefficiency term of firm *i* and v_{it} is the idiosyncratic error of firm *i* in period *t*. Production function *f* is assumed to be monotonic increasing and concave as above; we assume no particular functional form for *f*. We assume that the idiosyncratic errors v_{it} are uncorrelated random variables with $E(v_{it}) = 0 \;\forall i,t$ and $Var(v_{it}) = \sigma_v^2 < \infty \;\forall i,t$ (i.e., the Gauss-Markov assumptions). Importantly, we impose no parametric assumptions about the distributions of u_i and v_{it}: the panel variant (2) is a fully nonparametric model.

It is worth to note that, similar to the standard panel data treatments in the SFA literature, we here assume that the production function *f* and the inefficiency terms u_i do not change over time. This allows us to estimate model (2) by means of the standard fixed effects and random effects techniques, to be considered next. A dynamic, semiparametric variant of the StoNED model that allows the production function and inefficiency terms change over time is introduced in Section 6 below.

3. Fixed effects estimation

In the fixed effects approach the inefficiency terms u_i are taken as unknown firm-specific constants. Since the idiosyncratic errors v_{it} are the only source of random variation, we can estimate the StoNED model (2) by nonparametric least squares subject to monotonicity and concavity constraints (Hildreth, 1954; Hanson and Pledger, 1976; Groeneboom et al., 2001) [here more shortly convex nonparametric least squares (CNLS)]. CNLS is particularly well 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 essential statistical properties of the CNLS estimators are nowadays well understood. The maximum likelihood property of the CNLS 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 $O_P(n^{-1/(2+m)})$.

All known treatments of CNLS focus on the cross-sectional estimation. To estimate model (2), we need to adapt the CNLS estimator to the panel data setting. Introducing fixed effects u_i, the panel variant of the nonparametric least squares problem can be formally stated as

$$\begin{aligned} &\min_{f,\mathbf{u}} \sum_{t=1}^{T}\sum_{i=1}^{n}(y_{it}-(f(\mathbf{x}_{it})-u_i))^2 \\ &s.t. \\ &f \in F_2 \end{aligned} \tag{3}$$

In words, the CNLS problem selects production function $f \in F_2$ and inefficiency terms u_i to minimize the L_2-norm of the residuals. Problem (3) does not restrict beforehand to any particular functional form of f, but searches the best-fit function from the family F_2, which includes an infinite number of functions. This makes problem (3) generally hard to solve. In statistics, efficient algorithms for solving CNLS problems in the single regressor (i.e., single input) cross-sectional 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 x. However, such a sorting trick is not possible in the general multiple regression (i.e., multi-input) setting where $\mathbf{x}$ is a vector rather than scalar.

To solve the CNLS problem (3) in the general multi-input panel data setting, we utilize the insights from Afriat's Theorem, following Banker and Maindiratta (1992), Matzkin (1994), and

Kuosmanen (2006). We convert the infinite dimensional optimization problem (3) into the following finite quadratic programming (QP) problem:

$$
\begin{aligned}
&\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{u},\mathbf{v}} \sum_{t=1}^{T}\sum_{i=1}^{n} v_{it}^2 \\
&y_{it} = \alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{x}_{it} - u_i + v_{it} \;\; \forall i = 1,...,n; t = 1,...,T \\
&\alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{x}_{it} \le \alpha_{hs} + \boldsymbol{\beta}_{hs}'\mathbf{x}_{it} \;\; \forall h,i \in \{1,...,n\}; s,t \in \{1,...,T\} \\
&\boldsymbol{\beta}_{it} \ge \mathbf{0} \;\; \forall i = 1,...,n; t = 1,...,T
\end{aligned}
\tag{4}
$$

The first constraint of this problem can be interpreted as the regression equation, the second constraint enforces concavity analogous to the Afriat inequalities, and the third constraint ensures monotonicity. The analogy of model (4) with the conventional parametric regression models is useful for econometric model building. Note that (4) differs from the classic OLS problem in that the coefficients $\alpha_i, \boldsymbol{\beta}_i$ are here firm-specific. In this respect, model (4) 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; Tsionas, 2002), 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 CNLS regression (4) estimates *n* tangent hyper-planes to one unspecified production function. The slope coefficients $\boldsymbol{\beta}_{it}$ represent the marginal products of inputs (i.e., the sub-gradients $\nabla f(\mathbf{x}_{it})$). Problems (3) and (4) are equivalent in the following sense:

Proposition 1: Let s_F^2 be the minimum sum of squares of problem (3) and let s_A^2 be the minimum sum of squares of problem (4). Then for any real-valued data, $s_F^2 = s_A^2$.

The absolute inefficiency levels are unidentifiable from (3) and (4). Following Gabrielsen (1975) and Greene (1980) (see also Schmidt and Sickles 1984), we may use the best observed practice in the sample as a benchmark: given the CNLS estimates $\hat{u}_i$ from (4), we compute the relative inefficiency estimates as

$$\hat{\hat{u}}_i \equiv \hat{u}_i - \min_{h \in \{1,\ldots,n\}} (\hat{u}_h) . \tag{5}$$

Note that $\hat{u}_i$ can be negative. If the sampling procedure is such that fully efficient firms (i.e., firm *i* such that $u_i = 0$) are observed with a strictly positive probability, then $\hat{\hat{u}}_i$ is a consistent estimator of u_i.

Given the estimated coefficients $\hat{\alpha}_{it}, \hat{\boldsymbol{\beta}}_{it}$ from model (4), we may estimate the production function *f* by the following piece-wise linear function

$$\hat{f}(\mathbf{x}) \equiv \min_{i \in \{1,\ldots,n\}, t \in \{1,\ldots,T\}} (\hat{\alpha}_{it} + \hat{\boldsymbol{\beta}}_{it}' \mathbf{x}) . \tag{6}$$

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

Proposition 2: Denote the set of functions that minimize problem (3) by $F_2^* : F_2^* \subset F_2$. For any real-valued data, $\hat{f} \in F_2^*$.

The representor $\hat{f}$ and its coefficients $(\hat{\alpha}_{it}, \hat{\boldsymbol{\beta}}_{it})$ have a compelling economic interpretation. Vector $\hat{\boldsymbol{\beta}}_{it}$ represents the marginal products of firm *i* in period *t*. Moreover, function $\hat{f}$ provides a first order Taylor series approximation to any $f \in F_2^*$ in the neighborhood of the observed points. This justifies the use of the representor $\hat{f}$ for forecasting the output values in unobserved points within the observed range of input values.

Coefficients $\hat{\boldsymbol{\beta}}_{it}$ can also be used for nonparametric estimation of the marginal properties and elasticities. We can calculate the marginal rate of substitution between inputs *k* and *m* in point $\mathbf{x}_{it}$ as

$$\frac{\partial \hat{f}(\mathbf{x}_{it}) / \partial x_k}{\partial \hat{f}(\mathbf{x}_{it}) / \partial x_m} = \frac{\hat{\beta}_{kit}}{\hat{\beta}_{mit}} \quad , \tag{7}$$

and further, the elasticity of substitution as

$$e_{k,m}(\mathbf{x}_{it}) = \frac{\hat{\beta}_{kit}}{\hat{\beta}_{mit}} \cdot \frac{x_{mit}}{x_{kit}} \quad . \tag{8}$$

These substitution rates and elasticities are simple to compute given the estimated $\hat{\boldsymbol{\beta}}_{it}$ coefficients.

The piece-wise linear structure of the estimator (6) closely resembles that of the DEA frontier. Although problem (4) includes *n* different firm-specific coefficients $\alpha_i, \boldsymbol{\beta}_i$, the number of different hyperplane segments in $\hat{f}(\mathbf{x})$ is typically much lower than *n* as in DEA. One could easily impose further assumptions about returns to scale as in DEA: if function *f* exhibits constant returns to scale, we may simply set $\alpha_{it} = 0$ in problem (4). This will guarantee that function $\hat{f}$ passes through the origin. Another similarity is that the estimator $\hat{f}(\mathbf{x})$ and its coefficients $\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_i$ are not necessarily unique. To test for uniqueness, one could construct upper and lower bounds for function $\hat{f}(\mathbf{x})$ along the lines of Varian (1984). Finally, one can draw statistical inference about the inefficiency estimates or the coefficients $\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_i$ by applying the bootstrap approach by Efron (1979, 1982) (see also Simar, 1992; and Kuosmanen and Kortelainen, 2007)

According to Greene (1999, 2005), a problem of the fixed effects approach is that it attributes any time-invariant heterogeneity across firms to the inefficiency term. In our view, heterogeneity is not a problem as such if it arises endogenously from the decisions of the firm management (e.g. how much to invest in capital, or where to locate the firm). However, if firms are heterogeneous in attributes that are exogenously given and beyond their control, then labelling impacts of heterogeneity as inefficiency (or lack of it) is both unfair and misleading. Note that including these attributes as inputs (or as other explanatory variables) in the model does not help because the lack of variation in these attributes makes their coefficients unidentifiable from the fixed effects. The random effects approach can provide a remedy for this problem.

4. Random effects estimation

While in the fixed effects model the inefficiency terms were taken as firm-specific constants, the random effects approach views inefficiency terms as random variables with $E(u_i) = \mu > 0 \ \forall i,t$ and

$Var(u_i) = \sigma_u^2 < \infty \ \forall i,t$. Note that inefficiency terms u_i are still constant over time; random variation restricts to variation across firms. A critical assumption of the random effects approach is that the inefficiency terms must be uncorrelated with inputs $\mathbf{X}$ and idiosyncratic errors v_{it}. Parametric assumptions about the distribution of u_i are not necessary, but can be added to the model (e.g., Battese and Coelli, 1988).

The first step of the random effects estimation is to re-write model (2) as

$$y_{it} = [f(\mathbf{x}_{it}) - \mu] - [u_i - \mu] + v_{it} \tag{9}$$

$$= g(\mathbf{x}_{it}) + w_{it}, \tag{10}$$

where function $g(\mathbf{x}_{it}) \equiv f(\mathbf{x}_{it}) - \mu$ can be interpreted as the average production function (adjusted to the expected inefficiency) and $w_{it} \equiv -[u_i - \mu] + v_{it}$ is the composite error term. Note that g is a continuous, monotonic increasing and concave function that differs from frontier f only by a positive constant μ. Furthermore, note that the composite errors w_{it} are random variables with zero mean and constant variance $\sigma_u^2 + \sigma_v^2$. Thus, function g can be consistently estimated by CNLS.

The CNLS problem (3) can be adapted to the estimation of the average production function g as

$$\begin{aligned}
&\min_g \sum_{t=1}^{T} \sum_{i=1}^{n} (y_{it} - g(\mathbf{x}_{it}))^2 \\
&s.t. \\
&g \in F_2
\end{aligned} \tag{11}$$

The key difference between (3) and (11) is that in (3) the inefficiency terms u_i are modeled explicitly (as dummy variables) while in (11) the inefficiency terms are attributed to the residuals. If the inefficiency terms are uncorrelated with inputs and errors, as assumed in this section, then g is consistently estimated by (11). Note that the random effects treatment decreases the number of unknowns compared to the CNLS problem (4) by n. Analogous to (4), we may convert this infinite dimensional optimization problem into a tractable quadratic programming problem:

$$
\begin{aligned}
&\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{w}} \sum_{t=1}^{T}\sum_{i=1}^{n} w_{it}^2 \\
&y_{it} = \alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{x}_{it} + w_{it} \;\; \forall i = 1,...,n; t = 1,...,T \\
&\alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{x}_{it} \le \alpha_{hs} + \boldsymbol{\beta}_{hs}'\mathbf{x}_{it} \;\; \forall h,i \in \{1,...,n\}; s,t \in \{1,...,T\} \\
&\boldsymbol{\beta}_{it} \ge \mathbf{0} \;\; \forall i = 1,...,n; t = 1,...,T
\end{aligned}
\tag{12}
$$

Proposition 3: Let s_G^2 be the minimum sum of squares of problem (11) and let s_B^2 be the minimum sum of squares of problem (12). Then for any real-valued data, $s_G^2 = s_B^2$.

To estimate the inefficiency component u_i, we note that

$$E\left(\sum_{t=1}^{T} w_{it} / T\right) = \mu - u_i . \tag{13}$$

Moreover, applying the Gabrielsen-Greene normalization to composite errors w_{it}, we have

$$E\left(\min_{h\in\{1,...,n\}} \sum_{t=1}^{T} w_{ht} / T - \sum_{t=1}^{T} w_{it} / T\right) = \left[\mu - \min_{h\in\{1,...,n\}}(u_h)\right] - [\mu - u_i] = u_i - \min_{h\in\{1,...,n\}}(u_h) . \tag{14}$$

If the sampling procedure is such that fully efficient firms (i.e., firm *i* such that $u_i = 0$) are observed with a strictly positive probability, then a consistent estimator of u_i is obtained from (14) by utilizing the residuals of the CNLS model (12) as

$$\hat{u}_i = \min_{h\in\{1,...,n\}}\left(\sum_{t=1}^{T} \hat{w}_{ht} / T\right) - \sum_{t=1}^{T} \hat{w}_{it} / T . \tag{15}$$

Expected inefficiency μ can be subsequently estimated by using the sample mean as $\hat{\mu} = \sum_{i=1}^{n} \hat{u}_i / n$, and variance σ_u^2 can be estimated by the sample variance $\hat{\sigma}_u^2 = \sum_{i=1}^{n} (\hat{u}_i - \hat{\mu})^2 / (n-1)$. The empirical distribution of inefficiency estimates $\hat{u}_i$ can be used for drawing nonparametric statistical inference.

It is worth to note that the estimated coefficients $\hat{\alpha}_{it}, \hat{\boldsymbol{\beta}}_{it}$ from (12) relate to the average production function *g*. Since $g(\mathbf{x}_{it}) \equiv f(\mathbf{x}_{it}) - \mu$, we may estimate the frontier production function *f* by

$$\hat{f}(\mathbf{x}) = \min_{i \in \{1,\ldots,n\}, t \in \{1,\ldots,T\}} (\hat{\alpha}_{it} + \hat{\boldsymbol{\beta}}_{it}'\mathbf{x}) + \hat{\mu} \ . \tag{16}$$

Estimator $\hat{f}$ and its coefficients $(\hat{\alpha}_{it}, \hat{\boldsymbol{\beta}}_{it})$ have the same economic interpretation as their corresponding fixed effects counter parts: vector $\hat{\boldsymbol{\beta}}_{it}$ represents the marginal products of firm *i* in period *t*, and function $\hat{f}$ provides a first order Taylor series approximation to any *f* in the neighborhood of the observed points.

Instead of applying the Gabrielsen-Greene normalization, the inefficiency terms could be estimated in a semiparametric fashion if one imposes further parametric assumptions about the distributions of the random inefficiency and error terms. Given the CNLS residuals from (11), the conditional expected value of the inefficiency term could be extracted as described by Battese and Coelli (1988). The benefits of the semiparametric estimation include the access to conventional tools of statistical inference on the inefficiency term and avoiding the use of a benchmark firm as in (15). The disadvantage of this alternative is the need to impose ad hoc distributional assumptions.

In conclusion, we have shown that the StoNED model can be estimated fully nonparametrically in the panel data setting by means of fixed effects or random effects estimation. The random effects approach allows time-invariant firm-specific attributes to enter the model as inputs (or as other explanatory variables) but assumes the inefficiency term to be uncorrelated with the input levels. By contrast, the fixed effects approach cannot distinguish inefficiency from time-invariant firm-specific attributes, but allows the inefficiency term to be correlated with the input levels. Thus far, both approaches have built upon the assumption that the inefficiency terms and the technology remain constant over time. Some alternative ways of modeling technical change and dynamic efficiency changes will be discussed in Section 6. But first, a simulated example illustrates the developments of Sections 2-4.

5. Simulated example

We next illustrate the estimation of the StoNED model by fixed and random effects approaches by means of a simulated numerical example. Consider a single-input setting with a balanced panel of five firms over 20 years. The input values were randomly drawn: Firm 1 from *Uni*[1,11], Firm 2 from *Uni*[3,13], Firm 3 from *Uni*[5,15], Firm 4 from *Uni*[7,17], and Firm 5 from *Uni*[9,19], respectively. Drawing input values of firms from different domains reflects the fact that firms typically differ in their scale size. The true $f(x)$ values were calculated using production function $f(x) = \ln(x) + 1$. From the true $f(x)$, the firm specific inefficiency term was subtracted: inefficiency levels were assumed as u_1=0, u_2=1.5, u_3=1, u_4=2 u_5=0.5. The resulting output values were next perturbed by adding a random error drawn independently from $N(0,0.25^2)$. This gave the observed output values y. Figure 1 illustrates the observed sample by the scatter plot; the true function f is also superimposed in the scatter.

<Figure 1 around here>

<Table 1 around here>

We next computed the fixed and random effects CNLS regressions (4) and (12) using the Minos NLP solver of the GAMS software. Table 1 reports the coefficient of determination (R^2) and the estimated firm-specific inefficiencies. For comparison, the fixed and random effects SFA models assuming a linear production function were computed using OLS. All four models identified Firm 1 correctly as efficient. Also the relative efficiency rankings were correct in all models, but the estimated efficiency levels showed considerable differences. All models underestimated inefficiency in this example, but this is by no means a structural property of the estimators. We note that the fixed effects regression gave much better empirical fit than the random effects treatment both in the CNLS and OLS cases. Consequently, the inefficiency estimates of the fixed effects models came closer to the true inefficiency levels than the corresponding random effects estimates. Both in the fixed and random effects cases, the CNLS estimators gave better empirical fit and inefficiency estimates than the corresponding OLS panel data models.

Figure 1 also illustrates the representor functions $\hat{f}$ obtained from the fixed and random effects models. In both cases, the piece-wise linear StoNED frontiers consist of three line segments (i.e., coefficients $\hat{\alpha}_{it}, \hat{\boldsymbol{\beta}}_{it}$ were clustered to three different values in the sample). Recall that the number and the location of segments are endogenously determined within the models. The fixed effects frontier provides a reasonably good approximation of the true *f* throughout the observed range of x values. The random effects frontier is more flat because the original least squares regression excludes the firm-specific dummy variables which influence the estimation of the $\hat{\alpha}_{it}, \hat{\boldsymbol{\beta}}_{it}$ coefficients.

<Figure 2 around here>

Figure 2 further illustrates the case by presenting the scatter plots of the true output values $f(x_{it}) - u_i$ (the points without filling) and the corresponding fitted values $\hat{y}_{it}$ (points with grey filling) from the fixed effects CNLS model. Each firm is represented by a clearly distinguishable cluster of points. Note that the scatter of Figure 1 is obtained by adding the random noise (v_{it}) to the true values of Figure 2. Given the rather chaotic scatter of Figure 1, it is remarkable how closely the CNLS fitted values approximate the true outputs in Figure 2.

6. Dynamic StoNED model

6.1 Model specification

Thus far we have assumed that the production technology and inefficiency terms do not change over time. In this section we allow the production function to exhibit technical progress (or regress) and include time *t* explicitly as an argument of *f* and write production function as $f(\mathbf{x},t)$, $t = 0,1,...,T$. Also efficiency can change over time; we replace the constant inefficiency terms u_i by functions $u_i(t)$, i=1,…,n. In addition, in this section we model the composite error term in multiplicative rather than additive fashion. These three departures from model (2) lead us to the following new specification of the StoNED model:

$$y_{it} = f(\mathbf{x}_{it}, t)/(1 + u_i(t) - v_{it}) . \tag{17}$$

We next discuss each of these three departures in more detail, and return to the estimation of model (17) in Section 6.5 below.

6.2 Modeling technical change

There are many ways to model technical progress by means of production function $f(\mathbf{x},t)$. One convenient approach is to assume that technical progress is of *output augmenting, input additive* type (see Beckmann and Sato, 1969, for a detailed classification), and specify function $f(\mathbf{x},t)$ as

$$f(\mathbf{x},t) = f(\mathbf{x},0) + \sum_{m=1}^{M} A_m(t) x_m , \quad i = 1,...,n; \quad t = 0,1,...,T . \tag{18}$$

Function $f(\mathbf{x},0)$ is referred to as the base period production function, and functions $A_m : \mathbb{R}_+ \to \mathbb{R}$ represent *input embodied* technical change (e.g. improvement of capital inputs due to better designs, or improvement of labor quality due to learning by doing). The properties of the base period production function $f(\mathbf{x},0)$ carry over to other periods according to the following theorem:

Proposition 4: If the base period production function $f(\mathbf{x},0)$ is monotonic increasing for all $\mathbf{x} \in \mathbb{R}_+^M$ and $A_m(t) \geq 0 \quad \forall m = 1,...,M; \; t = 0,1,...,T$, then functions $f(\mathbf{x},t)$ that satisfy (17) are monotonic increasing for all $\mathbf{x} \in \mathbb{R}_+^M$ and $t \in \{0,1,...,T\}$. Furthermore, if $f(\mathbf{x},0)$ is globally concave for all $\mathbf{x} \in \mathbb{R}_+^M$, then functions $f(\mathbf{x},t)$ that satisfy (17) are globally concave for all $\mathbf{x} \in \mathbb{R}_+^M$ and $t \in \{0,1,...,T\}$.

This result shows the generality of the specification (18) and proves convenient for modeling technical change in the StoNED framework. We can estimate the base period production function $f(\mathbf{x},0)$ in the nonparametric fashion as in Sections 2 and 3, and model the technical change component

A in a nonparametric or parametric fashion. Importantly, the concavity properties of $f(\mathbf{x},0)$ carry over to all future periods regardless of how functions A_m are modelled.

Purely nonparametric specification of functions A_m is possible by writing

$$A_m(t) = \gamma_{mt} , \tag{19}$$

where γ_{mt} are input and time-period specific constants. However, this introduces a very large number (*MT*) of new unknown parameters to the model. Alternatively, one could resort to a parametric approximation, for example, a quadratic function

$$A_m(t) = \theta_m t + \psi_m t^2 . \tag{20}$$

This quadratic specification adds 2*M* new unknown parameters to the model.

Further restrictions on functions A_m are useful, especially in the purely nonparametric specification (19). If one assumes that technical regress does not occur, then one can impose further constraints

$$A_m(t+1) \geq A_m(t) \;\; \forall t = 0,...,T-1. \tag{21}$$

One might assume that technical progress is embodied to certain inputs; consider for example capital embodied or labor embodied technical progress. In these cases, the A_m functions for other inputs can be eliminated from the model. One could also assume that production function *f* is a homothetic function of time, such that technical change does not influence the shape of the isoquants but shifts them parallel towards the origin. This assumption can be simply imposed as

$$A_1(t) = A_2(t) = ... = A_M(t) \;\; \forall t = 0,...,T-1. \tag{22}$$

This could be seen as an additive type of Hicks neutrality (compare with Blackorby et al. 1976). However, the homotheticity property (22) differs from the standard Hicks neutrality in that the marginal rates of substitution can change over time.

6.3 Modeling efficiency change

Improvement of technical efficiency over time is commonly referred to as catching up and the decline of efficiency as falling behind. Distinguishing the effects of technical progress from efficiency change is generally very challenging. In this respect, the previous specification of input embodied technical progress is attractive because it allows the rate of technical progress to vary across firms depending on their input use, but it is neutral to the identity of the firm: two firms with identical input mix always exhibit the same rate of technical progress. Thus, any systematic firm specific deviations from the frontier remain identifiable as inefficiency.

Modeling efficiency changes over time in purely nonparametric fashion is not very practical in the CNLS framework, unless one imposes some rather restrictive assumptions (e.g., that efficiency increases/declines at increasing/decreasing rate). Without sufficiently stringent structure, the efficiency changes remain indistinguishable from the noise. Therefore, if the model of constant efficiency of Section 3 seems inappropriate, we suggest to resort to a semiparametric approach where the production function is estimated in a nonparametric fashion and the inefficiency function $u_i(t)$ is approximated by some parametric function. Following Cornwell et al. (1990), a second-order polynomial specification

$$u_i(t) = a_i + b_i t + c_i t^2 \tag{23}$$

seems useful because it allows the level of technical change to be constant ($b_i = c_i = 0$), or increase in a linear ($b_i > 0, c_i = 0$) or nonlinear ($c_i \neq 0$) fashion. Constant a_i can be interpreted as the initial inefficiency level of firm *i* (in period 0). Alternatively, a_i can be seen as a proxy of heterogeneity across firms (cf., e.g., Greene, 2005). It is worth to emphasize that if one is primarily interested in the efficiency changes over time rather than their absolute levels (consider e.g. the Malmquist productivity index and its decomposition), then the interpretation of constants a_i does not influence the results.

6.4 Multiplicative inefficiency and error terms

While the additive error model introduced in Section 2 is standard in econometrics, the SFA models typically assume a multiplicative error structure of form

$$y_{it} = f(\mathbf{x}_{it}) \cdot \exp(v_{it} - u_i) . \tag{24}$$

Model (24) can be transformed to the additive form by taking logarithms of both sides of the equation. The multiplicative error structure is convenient at least for three reasons. First, the multiplicative error specification can alleviate heteroskedasticity across different sized firms. Second, the multiplicative model of errors is useful for modeling constant returns to scale. Third, the multiplicative structure of the inefficiency term is consistent with the Farrell output efficiency index and the Shephard distance function.

In the CNLS framework, applying logarithmic transformations to data would violate the concavity assumption. For our purposes, the multiplicative specification of (17) is convenient, because it can be rewritten in additive form without any data transformations as

$$y_{it} = f(\mathbf{x}_{it}, t) - u_i(t) y_{it} + v_{it} y_{it} . \tag{25}$$

Note that both inefficiency $u_i(t)$ and errors v_{it} are now proportionate to the output y_{it} .

6.5 Estimation

Combining specifications (18), (20), (23) and (25) into the StoNED model (17), we obtain the regression equation to be estimated:

$$y_{it} = f(\mathbf{x}_{it}, 0) + (t\boldsymbol{\theta} + t^2\boldsymbol{\psi})'\mathbf{x}_{it} - (a_i + b_i t + c_i t^2) y_{it} + v_{it} y_{it} , \tag{26}$$

where $\boldsymbol{\theta} \equiv (\theta_1 ... \theta_M)'$ and $\boldsymbol{\psi} \equiv (\psi_1 ... \psi_M)'$. Note that coefficients $\boldsymbol{\theta}, \boldsymbol{\psi}$ vary across inputs but are constant across firms, whereas coefficients a_i, b_i, c_i vary across firms but are constant across inputs. This distinction allows us to identify the technical change from changes in efficiency.

To estimate the unknown *f* and the parameters $\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{a}, \mathbf{b}, \mathbf{c}$ from (26), we apply insights from Propositions 1-3 and construct the following CNLS estimator:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\varphi,\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{v}} \sum_{t=1}^{T}\sum_{i=1}^{n} v_{it}^2$$
$$y_{it} = \alpha_{it} + \boldsymbol{\beta}'_{it}\mathbf{x}_{it} + (t\boldsymbol{\theta} + t^2\boldsymbol{\psi})'\mathbf{x}_{it} - (a_i + b_i t + c_i t^2)y_{it} + v_{it}y_{it} \;\; \forall i = 1,...,n; t = 1,...,T$$
$$\alpha_{it} + \boldsymbol{\beta}'_{it}\mathbf{x}_{it} \le \alpha_{hs} + \boldsymbol{\beta}'_{hs}\mathbf{x}_{it} \;\; \forall h,i \in \{1,...,n\}; s,t \in \{1,...,T\} \quad (27)$$
$$\boldsymbol{\beta}_{it} \ge \mathbf{0} \;\; \forall i = 1,...,n; t = 1,...,T$$
$$t\boldsymbol{\theta} + t^2\boldsymbol{\psi} \ge \mathbf{0} \;\; \forall t = 1,...,T$$

The objective function minimizes the L_2 norm of the error terms v_{it}. Note that the error terms are not multiplied by y_{it} in the objective function; the multiplication in (25) and (26) is used for transforming the multiplicative error term into additive form. The first constraint is the regression equation (26) where the unknown $f(\mathbf{x}_{it},0)$ is represented by the tangent hyperplane $\alpha_{it} + \boldsymbol{\beta}'_{it}\mathbf{x}_{it}$. The second constraint imposes the concavity constraints to the tangent hyperplanes by applying the Afriat inequalities. The third constraint imposes monotonicity. The fourth constraint is a technical restriction: by Proposition 4, this condition together with the monotonicity and concavity of the base period production function $f(\mathbf{x},0)$ will ensure that $f(\mathbf{x},t)$ is monotonic and concave for all *t*. Note that the fourth constraint does not rule out the possibility of technical regress.

Efficiency measures must be inferred indirectly from model (27). Given the parameter estimates $\hat{a}_i, \hat{b}_i, \hat{c}_i$ from (27), the firm specific inefficiency estimates for period *t* are calculated relative to the best-practice benchmark using the Gabrielsen-Greene normalization (see Cornwell et al., 1990) as

$$\hat{u}_i(t) = (\hat{a}_i + \hat{b}_i t + \hat{c}_i t^2) - \min_{h\in\{1,...,n\}}(\hat{a}_h + \hat{b}_h t + \hat{c}_h t^2) \quad (28)$$

The estimator of the base period production function $f(\mathbf{x},0)$ is obtained analogous to (6). For periods *t*>0, the production function is estimated by

$$\hat{f}(\mathbf{x},t) = \min_{i\in\{1,...,n\},t\in\{1,...,T\}}(\hat{\alpha}_{it} + \hat{\boldsymbol{\beta}}'_{it}\mathbf{x}) + (t\boldsymbol{\theta} + t^2\boldsymbol{\psi})'\mathbf{x} \quad (29)$$

Note that the technical change influences the marginal products of inputs and the substitution properties: for function (29), the marginal rate of substitution between inputs *k* and *m* in point $\mathbf{x}_{it}$ is

$$\frac{\partial \hat{f}(\mathbf{x}_{it})/\partial x_k}{\partial \hat{f}(\mathbf{x}_{it})/\partial x_m} = \frac{\hat{\beta}_{kit} + t\theta_k + t^2\psi_k}{\hat{\beta}_{mit} + t\theta_m + t^2\psi_m} \quad . \tag{30}$$

Thus, the shape of the production function is determined partly by the nonparametric part $f(\mathbf{x},0)$ and the parametric part $(t\boldsymbol{\theta} + t^2\boldsymbol{\psi})'\mathbf{x}$ that represents technical change.

7. Application: productive efficiency in the wholesale and retail trade

7.1 Motivation and setup

In this section we estimate the dynamic, semiparametric, multiplicative StoNED model (17) from the industry-level panel data of the wholesale and retail trade sectors in fourteen OECD countries for the period 1975-2003, obtained from the OECD's Structural Analysis database STAN (http://www.oecd.org). The industry classification covers the ISIC codes 51-52; data for Canada, Sweden, and New Zeeland also include ISIC code 50 (the sale and repair of motor vehicles and the retail sale of fuels). The list of countries (abbreviation) is the following: Austria (AUT), Canada (CAN), Denmark (DEN), Finland (FIN), France (FRA), Great Britain (GBR), Germany (GER, 1975-1990 West Germany), Italy (ITA), Netherlands (NLD), New Zealand (NZL), Norway (NOR), Portugal (POR), Sweden (SWE), and United States (USA). We will henceforth refer to the wholesale and retail trade sectors of countries by these three-letter abbreviations.

The input and output variables are the following:

Output: value added (*y*) (in Bill. €, prices of 2000)

Inputs: 1) gross capital stock (*K*) (in Bill. €, prices of 2000),

2) labor (*L*) (Bill. hours),

3) intermediate inputs (excluding items for sale) (*M*) (Bill. €, prices of 2000).

The data were obtained directly from the OECD STAN database, except in the following cases of missing data. For AUT, GER, ITA, NED, NOR, POR, and SWE, the capital stock was estimated based on the data of gross fixed capital formation using the perpetual inventory method assuming 8%

depreciation rate. The labor hours were estimated based on the employment data (full-time equivalents) for AUT, GER, ITA, POR, and GBR.

7.2 Model specification

The time horizon of the study spans 29 years so it is necessary to account for technical progress and changes in efficiency levels over time. Therefore, we estimate the dynamic, semi-parametric, multiplicative StoNED model (26) where the base period production technology is estimated in nonparametric fashion using CNLS, and technical progress and efficiency changes are modeled by means of parametric approximations. Given the vast differences in the sizes of the sectors, a multiplicative error structure is preferred. Constant returns to scale were postulated to assess small and large sectors against a common global benchmark. To exclude the possibility of technical regress, we added constraints of form (21). For clarity, we write out the estimated StoNED model in full as

$$
\begin{aligned}
&\min_{\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\psi},\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{v}} \sum_{t=1}^{T}\sum_{i=1}^{n} v_{it}^2 \\
&s.t. \\
&y_{it} = [\beta_{Lit}L_{it} + \beta_{Kit}K_{it} + \beta_{Mit}M_{it}] + \left[(\theta_L t + \psi_L t^2)L_{it} + (\theta_K t + \psi_K t^2)K_{it} + \right. \\
&\left.(\theta_M t + \psi_M t^2)M_{it}\right] - \left[a_i + b_i t + c_i t^2\right] y_{it} + v_{it} y_{it} \;\; \forall i = 1,...,n;\; \forall t = 1,...,T \\
&\beta_{Lit}L_{it} + \beta_{Kit}K_{it} + \beta_{Mit}M_{it} \le \beta_{Lhs}L_{it} + \beta_{Khs}K_{it} + \beta_{Mhs}M_{it} \;\; \forall h,i = 1,...,n;\; \forall s,t = 1,...,T \\
&\beta_{Lit}, \beta_{Kit}, \beta_{Mit} \ge 0 \;\; \forall i = 1,...,n;\; \forall t = 1,...,T \\
&\theta_L t + \psi_L t^2 \ge 0 \;\; \forall t = 1,...,T \\
&\theta_K t + \psi_K t^2 \ge 0 \;\; \forall t = 1,...,T \\
&\theta_M t + \psi_M t^2 \ge 0 \;\; \forall t = 1,...,T
\end{aligned}
\tag{31}
$$

This QP problem involves 1266 unknowns and 166,547 linear inequality constraints. The GAMS code for solving problem (31) is provided in Appendix B.

7.3 Results

Model (31) gave a good empirical fit with the coefficient of determination as high as R^2=0.98. The large R^2 value is due to the fact that there are large cross-country differences in the industry sizes, and the input use and the time trends capture a very large proportion of the variation in the output.

To shed light on the estimated semiparametric production function, Table 2 reports the marginal product (MP) estimates for all countries in selected years. Recall that the MP estimates (i.e, coefficients $\beta_L, \beta_K, \beta_M$) of the StoNED model are specific to each sector and year. The MP of labor has increased over time in almost all countries. This is an expected result as the capital stock has increased considerably. Compared with the hourly wage rates, the MPs are generally of the right magnitude, although the values for FRA appear unrealistically high and values of NZL and POR too low. Recall that the shadow prices of the nonparametric part of the production function need not be unique; the same applies to the input-output multipliers in the standard DEA models. The estimated MPs of capital are low for most observations; even close to zero. Since the MP=1 is consistent with the long-run profit maximization in competitive markets, a low MP suggests that the sector has over-invested in capital. Only NZL had MPs significantly greater than one, which suggests under-investment.

<Table 2 around here>

The MP estimates reported in Table 2 include both the nonparametric estimate of $f(\mathbf{x},0)$ and the parametric estimates of the technical change components A_m. Regarding the nonparametric part $f(\mathbf{x},0)$, the estimated frontier consists of 72 different hyper-plane segments. Majority of these segments are small: total of 33 segments contain only one observation, and 12 segments contained two or three observations. On the other hand, there were six large segments with more than 20 observations each. The largest hyper-plane segment contains 61 observations. Two thirds of the observations are projected to thirteen largest segments with at least ten observations each. Thus, the frontier could be reasonably approximated by a relatively small number of hyperplane segments.

Regarding the technical change functions A_m, the labor embodied technical change A_L was equal to zero (i.e., $\theta_L = \psi_L = 0$), and capital embodied technical change A_K was also very small ($\theta_K = -9.5 \cdot 10^{-5}, \psi_K = 9.5 \cdot 10^{-5}$). This suggests that no significant capital or labor saving technical progress occurred during the period. Our results suggest that technical change has been embodied in the intermediate inputs: the A_M function has grown by factor 12 during the 29-year period ($\theta_M = 0.0683, \psi_M = -0.0014$). A natural explanation for this finding is that the price margins of the wholesale and retail trade have increased over time.

Frontier analysis is typically geared towards efficiency estimation. Table 3 reports the industry-specific efficiency levels in all countries in the selected years. For clarity, the inefficiency functions are reported in a normalized form as $1/(1+u_i(t))$; the efficiency estimates reported in Table 3 lie within interval [0,1] with the value of one indicating full efficiency. The results suggest that CAN and GER were the most efficient sectors in 1975, but their efficiency has been in relative decline. USA reached the global efficient frontier in the mid-1980s, but its efficiency level declined in the 1990s. FRA took over the leading position in the early 1990s, but its efficiency has also declined since the late 1990s. In early 2000s, NOR has been closest to the global frontier. The least efficient sectors were found in NZL and POR.

<Table 3 around here>

It is worth to note that the relative efficiency measures tend to capture any cross-country heterogeneity in the geographic, logistic, environmental or institutional factors. This is a known problem in all fixed effects treatments (see e.g. Greene 1999, 2006 for further discussion). However, if we are interested in the changes in efficiency over time, then any time-invariant heterogeneity across countries will not disturb the results. The bottom row of Table 3 presents the (geometric) average annual efficiency change rates for the sectors. We note that especially SWE, FIN, NOR, and ITA showed relatively high efficiency improvements. On the other hand, the rates of efficiency changes were rather modest in all countries (less than one percent per year).

8. Concluding remarks

This paper has proposed a new way of combining the nonparametric DEA-type frontier with the stochastic SFA-type treatment of inefficiency and noise: we have adapted the Stochastic Nonparametric

Envelopment of Data (StoNED) framework, developed in the cross-sectional setting by Banker and Maindiratta (1992), Kuosmanen (2006), and Kuosmanen and Kortelainen (2007), to the panel data setting. Importantly, the panel data enabled us to drop the parametric distributional assumptions of the cross-sectional model and estimate both the frontier and the inefficiency terms in a fully nonparametric fashion. To account for intertemporal changes in production technology and efficiency, a semiparametric variant of the StoNED model was developed.

Compared to the parametric SFA, the main advantage of the StoNED approach is its avoidance of ad hoc assumptions about the functional form of the production function (or alternatively, cost or distance functions). In contrast to the flexible functional forms often used in SFA, one can easily impose monotonicity, concavity and homogeneity axioms without scarifying the local flexibility of the regression function. On the other hand, the main advantage of StoNED to the nonparametric DEA is its robustness to outliers, data errors, and other stochastic noise in the data. While in DEA the frontier is spanned by a relatively small number of efficient firms, all observations have equal influence on the shape of the StoNED frontier. In essence, StoNED addresses the main points of critique presented against SFA or DEA, and melds the advantages of both approaches into a unified framework.

The main advantage of StoNED to the alternative semiparametric estimation approaches (e.g., kernel or spline regression techniques) is its utilization of the established concepts and principles of SFA and DEA: the approach is based on the standard assumptions that practitioners of SFA and DEA are comfortable with. Thus, readers familiar with classic SFA and DEA approaches are expected to easily grasp the essential features of the proposed approach. The conceptual bridges between DEA and SFA are also important for the further integration of the parametric and nonparametric fields of productive efficiency analysis.

Since the StoNED model is a genuine hybrid of SFA and DEA, many existing tools and techniques from SFA and DEA can be easily incorporated into the proposed framework. However, the hybrid nature of StoNED also means that there are many important differences to both SFA and DEA which should be kept in mind. For example, the interpretation of the StoNED input coefficients differs considerably from those of the SFA coefficients. In contrast to DEA, all observations influence the shape of the StoNED frontier. In this respect, further research is needed for a better understanding of the similarities and differences.

While the StoNED approach combines the most attractive features of DEA and SFA, it also shares some of their limitations. Like DEA, also StoNED model is vulnerable to the curse of dimensionality; when the number of input variables increases, the sample size should expand exponentially. In this respect, utilization of panel data can help to alleviate the problem. On the other hand, stochastic noise does not necessarily restrict to the output data, also input data may be perturbed by measurement errors and other noise. Noisy input data remains somewhat problematic for SFA, and the same applies to the StoNED model. Despite these shared limitations, the benefits of the unified amalgam model clearly outweigh the costs.

Appendix A: Proofs of propositions

Proposition 1: Assuming $f(\mathbf{x}_{it}) = \alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{x}_{it} \quad \forall \mathbf{x}_{it}, i = 1,...,n; t = 1,...,T$ and writing $v_{it} = y_{it} - (\alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{x}_{it} - u_i)$, we see that the objective functions of (3) and (4) are equivalent. By Afriat's Theorem (Afriat 1967, 1972; Varian 1982), the following conditions are equivalent:

(1) there exist a set of coefficients α_{it} and $\boldsymbol{\beta}_{it} \geq \mathbf{0}$ that satisfy the Afriat inequalities $\alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{x}_{it} \leq \alpha_{hs} + \boldsymbol{\beta}_{hs}'\mathbf{x}_{it} \ \forall h,i \in \{1,...,n\}; s,t \in \{1,...,T\}$.

(2) there exists a continuous, monotonic increasing, concave function $f \subset F_2$ such that $f(\mathbf{x}_{it}) = \alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{x}_{it} \ \forall \mathbf{x}_{it}, i = 1,...,n; t = 1,...,T$.

The objective function of (3) depends on the value of f only in a finite set of points $\mathbf{x}_{it}, i = 1,...,n; t = 1,...,T$. Thus, the assumption $f(\mathbf{x}_{it}) = \alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{x}_{it} \quad \forall \mathbf{x}_{it}, i = 1,...,n; t = 1,...,T$ does not involve a loss of generality. Therefore, the equality $s_F^2 = s_A^2$ holds for any real-valued data set (**X**,**y**). □

Proposition 2:

It is straightforward to verify that $\hat{f} \in F_2$. Proposition 1 directly implies that $\hat{f} \in F_2^*$.

Proposition 3:

Directly analogous to Proposition 1 and hence omitted.

Proposition 4:

Monotonicity property is obvious. Regarding concavity, we note that $\sum_{m=1}^{M} A_m(t)x_m$ is an affine function of $\mathbf{x}$, and thus $f(\mathbf{x},t)$ is essentially a sum of two proper concave functions. It is known that the sum of two proper concave functions must itself be a concave function (see e.g. Rockafellar 1970).

Appendix B: THE GAMS code for computing the StoNED model (31) in Section 7

```
*14 industries, 29 time periods, 3 inputs, 1 output

SETS
i           index of industries /i1*i14/
t           time periods /t1*t29/;
alias(i,h)
alias(t,s);

PARAMETERS
Y(i,t)          output
L(i,t)          labor
K(i,t)          capital
M(i,t)          intermediate input
Tr(t)           time trend
Tr2(t)          time trend squared;

VARIABLES
VARIABLES
SS              sum of squares of errors
V(i,t)          residual
ThetaL          labor trend coefficient
PsiL            labor trend coefficient 2
```

```
ThetaK          capital trend coefficient
PsiK            capital trend coefficient 2
ThetaM          material trend coefficient
PsiM            material trend coefficient 2
A(i)            inefficiency constant
B(i)            inefficiency slope
C(i)            inefficiency quadratic;

POSITIVE VARIABLES
BL(i,t)         marginal product of labor
BK(i,t)         marginal product of capital
BM(i,t)         marginal product of intermediate inputs;

EQUATIONS
QSSE                    objective=sum of squares of errors
QREGRESSION(i,t)        regression equation
QCONCAVITY(i,j,t,s)     concavity constraint
QLCON(t)                nonnegative labor embodied technical progress
QKCON(t)                nonnegative capital embodied technical progress
QMCON(t)                nonnegative material embodied technical progress
QEFF(i,t)               nonnegative inefficiency;

QSSE..                  SS=e=sum(t,sum(i,V(i,t)*V(i,t)));
QREGRESSION(i,t)..      Y(i,t)=e=BL(i,t)*L(i,t)+BK(i,t)*K(i,t)+BM(i,t)*M(i,t)+
                        (ThetaL*Tr(t)+PsiL*Tr2(t))*L(i,t)+(ThetaK*Tr(t)+PsiK*Tr2(t))*K(i,t)+
                        (ThetaM*Tr(t)+PsiM*Tr2(t))*M(i,t) - (A(i)+B(i)*Tr(t)+C(i)*Tr2(t))*Y(i,t)
                        +Y(i,t)*V(i,t);
QCONCAVITY(i,j,t,s)..   BL(i,t)*L(i,t)+BK(i,t)*K(i,t)+BM(i,t)*M(i,t)=l=
                        BL(j,s)*L(i,t)+BK(j,s)*K(i,t)+BM(j,s)*M(i,t);
QLCON(t)..              ThetaL*Tr(t)+PsiL*Tr2(t)=g=0;
QKCON(t)..              ThetaK*Tr(t)+PsiK*Tr2(t)=g=0;
QMCON(t)..              ThetaM*Tr(t)+PsiM*Tr2(t)=g=0;

MODEL StoNED /all/

SOLVE StoNED using NLP Minimizing SS;
```


Acknowledgements

We would like to thank Mika Kortelainen, Timo Sipiläinen, Tarmo Räty and Guan Zhengfei for helpful comments and discussions. The usual disclaimer applies.

References

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

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

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

Aigner DJ, Lovell CAK, and Schmidt P. 1977. Formulation and Estimation of Stochastic Frontier Models *Journal of Econometrics* **6**: 21-37.

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

Banker RD, Charnes A, and Cooper WW. 1984. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* **309**: 1078–1092.

Battese G, and Coelli T. 1988. Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data. *Journal of Econometrics* **38**: 387-399.

Bauer PW. 1990. Recent Developments in the Econometric Estimation of Frontiers. *Journal of Econometrics* **46**: 39-56.

Beckmann MJ, and Sato R. 1969. Aggregate production Functions and Types of Technical Progress: A Statistical Analysis. *American Economic Review* **59**: 88-101.

Blackorby C., Lovell CAK, and Thursby MC. 1976. Extended Hicks Neutral Technical Change. *Economic Journal* **86**: 845-852.

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

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

Diewert WE, and Wales TJ. 1987. Flexible Functional Forms and Global Curvature Conditions. *Econometrica* **55**: 43-68.

Efron B. 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**: 1-26.

Efron B. 1982. *The jackknife, the bootstrap and other resampling plans*. CBMS-NSF regional conference series in applied mathematics 38, SIAM: Philadelphia, PA.

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

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

Farrell MJ. 1957. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society Series A* **120**: 253-282.

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

Gabrielsen A. 1975. *On Estimating Efficient Production Functions*. Working Paper A-85: Dept. of Humanities and Social Sciences, Chr. Michelsen Institute, Bergen.

Greene WH. 1980. Maximum Likelihood Estimation of Econometric Frontier Functions. *Journal of Econometrics* **13**: 27-56.

Greene WH. 1999. Frontier Production Functions. In *Handbook of applied econometrics, Vol II: Microeconomics*, Pesaran M, and Schmidt P (eds). Oxford: Blackwell.

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

Griffin JF, and Steel MJF. 2004. Semiparametric Bayesian Inference for Stochastic Frontier Models. *Journal of Econometrics* **123**: 121-152.

Groeneboom P, Jongbloed G, and Wellner JA. 2001. Estimation of a Convex Function:

Characterizations and Asymptotic Theory. *Annals of Statistics* **29**: 1653-1698.

Hanson DL, and Pledger G. 1976. Consistency in concave regression. *Annals of Statistics* **4**: 1038-1050.

Henderson DJ, and Simar L. 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**, 598-619.

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

Kumbhakar SC, Park BU, Simar L, and Tsionas EG. 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, and Kortelainen M. 2007. *Stochastic Nonparametric Envelopment of Data: Cross-sectional Frontier Estimation Subject to Shape Constraints,* University of Joensuu, Economics Discussion Paper 46.

Kuosmanen T, Post GT, and Scholtes S. 2007. Testing for Productive Efficiency in Case of Errors-in-Variables. *Journal of Econometrics* **136**: 131-162.

Lee L, and Tyler M. 1978. The Stochastic Frontier Production Functions and Average Efficiency. *Journal of Econometrics* **7**: 385-390.

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

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

Matzkin RL. 1994. Restrictions of Economic Theory in Nonparametric Methods. Chapter 42 in

Handbook of Econometrics Vol IV, Engle RF, and McFadden, DL (eds). Elsevier.

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

Meyer MC. 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.

Nemirovskii AS, Polyak BT, and Tsybakov AB. 1985. Rates of Convergence of Nonparametric Estimates of Maximum Likelihood Type. *Problems of Information Transmission* **21**, 258-271.

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

Park B, Sickles RC, and Simar L. 2003. Semiparametric Efficient Estimation of AR1 Panel Data Models. *Journal of Econometrics* **117**: 279-309.

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

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

Post GT, Cherchye L, and Kuosmanen T. 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.

Rockafellar RT. 1970. *Convex Analysis*. Princeton University Press.

Ruggiero J. 2004. Data Envelopment Analysis with Stochastic Data. *Journal of the Operational Research Society* **55**: 1008-1012.

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

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

Seiford LM, and Thrall RM. 1990. Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis. *Journal of Econometrics* **46**: 7-38.

Simar L. 1992. Estimating efficiencies from frontier models with panel data: A comparison of parametric, non-parametric and semi-parametric methods with bootstrapping. *Journal of Productivity Analysis* **3**: 171-203.

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

Tsionas M. 2002. Stochastic frontier models with random coefficients. *Journal of Applied Econometrics* **17**: 127–147.

Varian HR. 1984. The Nonparametric Approach to Production Analysis. *Econometrica* **52**: 579-598.

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

Figure 1. Simulated example with 5 firms observed over 20 periods. The scatter is an artificial sample of data points from a concave production function with firm-specific inefficiency and Gaussian noise. Superimposed on the scatter are the true production function *f* and the estimated production frontiers from the fixed and random effects StoNED models.

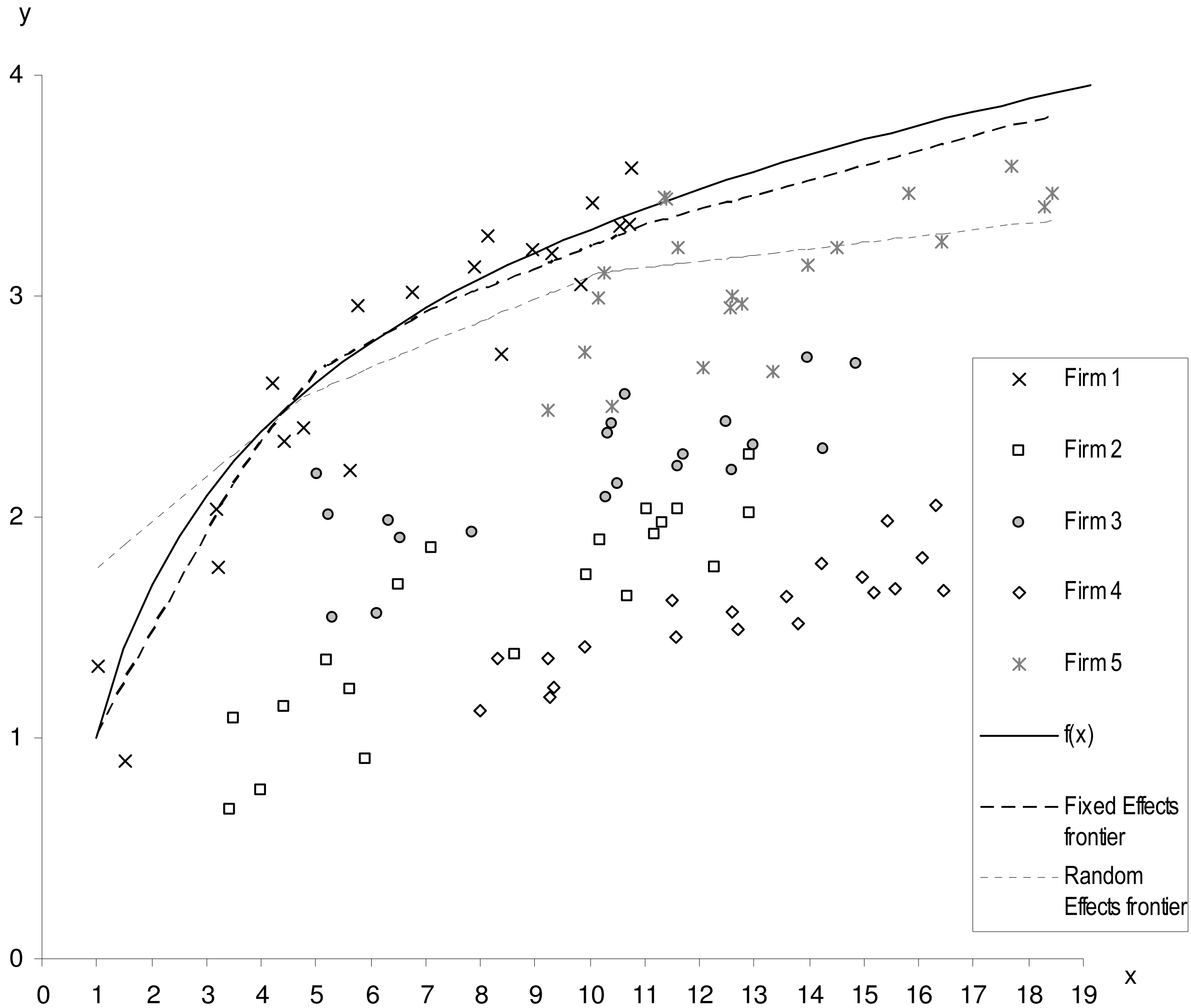

Figure 2. True data and the fitted values. The scatter illustrates the true and estimated input-output values of the five firms when the noise has been filtered out. The data points without filling indicate the true outputs ($f(\mathbf{x}_i)$-u_i) at the firm specific frontier. The data points with grey filling indicate the fitted values from the StoNED model estimated by the fixed effects CNLS regression.

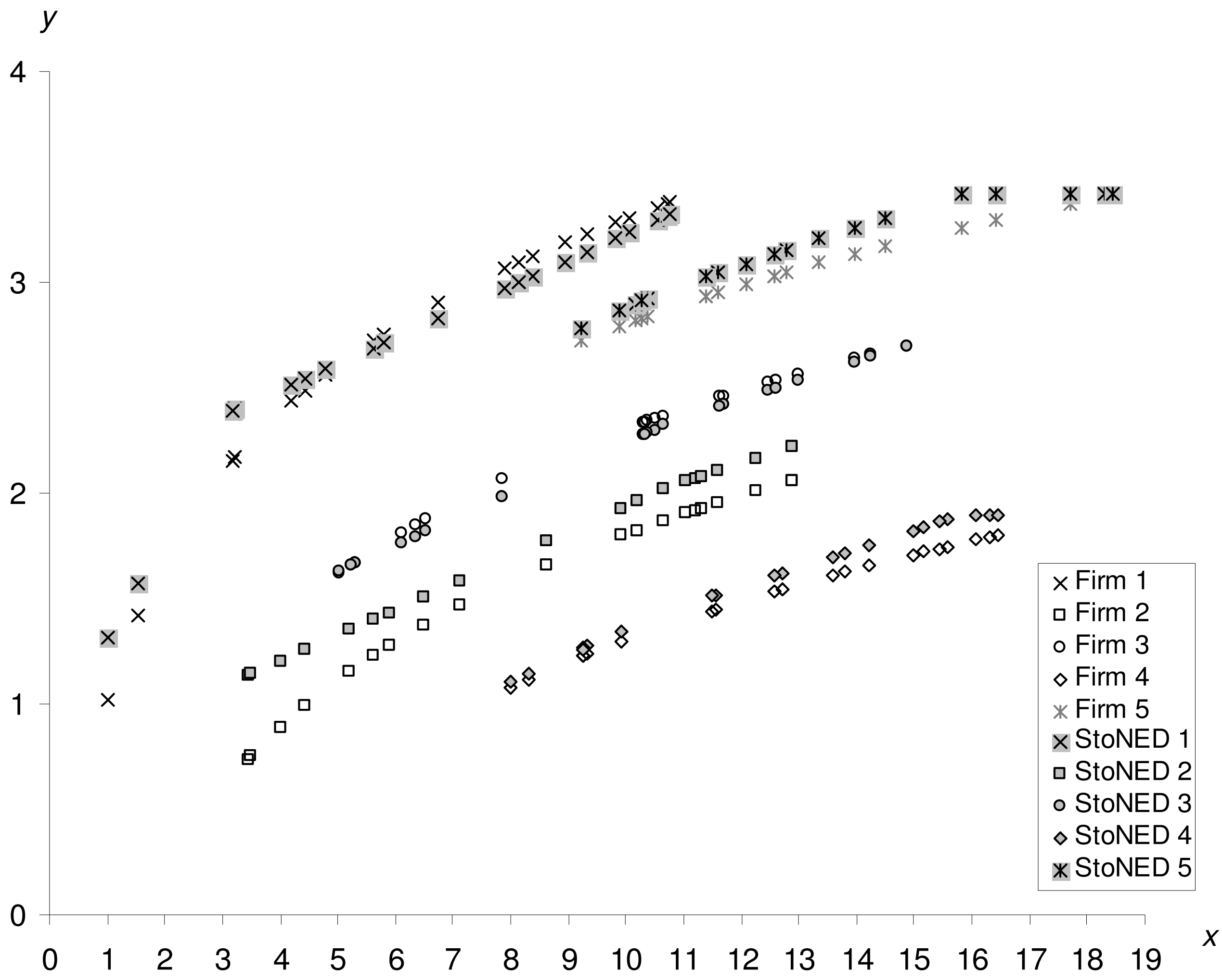

Table 1: Numerical example: true inefficiency levels and the StoNED and SFA estimates in the fixed and random effects models

	R^2	*Firm 1*	*Firm 2*	*Firm 3*	*Firm 4*	*Firm 5*
true inefficiency		0	1.50	1.00	2.00	0.50
StoNED fixed effects	0.925	0	1.41	0.98	1.85	0.37
StoNED random effects	0.183	0	1.29	0.80	1.59	0.10
SFA fixed effects	0.868	0	1.32	0.87	1.82	0.35
SFA random effects	0.156	0	1.25	0.74	1.58	0.09

Table 2: Estimated marginal products ($\hat{\beta}_{mit} + t\hat{\theta}_m + t^2\hat{\psi}_m$) of inputs in selected years

	year	AUT	CAN	DEN	FIN	FRA	GBR	GER	ITA	NED	NOR	NZL	POR	SWE	USA
labor	1975	2.22	0.03	1.70	0.75	9.75	7.71	6.38	0.75	4.26	2.22	2.22	0.01	0.76	6.02
	1980	3.54	2.22	9.75	1.60	9.75	7.98	8.49	1.60	8.49	4.26	3.54	0.01	0.76	8.49
	1985	6.38	2.22	9.52	7.71	9.67	12.3	10.8	3.54	12.2	9.13	3.75	0.01	10.2	8.49
	1990	11.1	3.54	9.67	9.75	9.67	12.3	11.1	7.98	12.2	12.3	0.49	0.01	9.18	9.13
	1995	12.2	6.02	12.3	9.52	22.8	21.0	12.2	12.2	12.2	12.2	0.00	0.76	8.49	12.2
	2000	12.2	6.38	12.3	12.3	48.2	48.2	12.2	12.2	12.2	27.4	2.99	0.76	9.13	9.67
	2003	12.2	11.3	24.4	12.2	48.2	48.2	12.2	12.2	12.2	27.4	0.00	0.76	12.2	9.75
capital	1975	0.16	0.04	0.00	0.00	0.00	0.09	0.29	0.00	0.17	0.16	0.16	0.00	0.84	0.28
	1980	0.17	0.16	0.00	0.06	0.00	0.09	0.29	0.06	0.29	0.18	0.17	0.25	0.84	0.29
	1985	0.30	0.17	0.04	0.10	0.07	0.19	0.52	0.17	0.35	0.28	1.31	0.26	1.09	0.30
	1990	0.39	0.19	0.08	0.02	0.08	0.20	0.60	0.11	0.36	0.20	1.21	0.59	0.45	0.29
	1995	0.38	0.32	0.22	0.07	0.09	0.34	0.38	0.38	0.38	0.38	1.67	0.88	0.33	0.38
	2000	0.40	0.35	0.24	0.24	0.23	0.23	0.40	0.40	0.40	0.52	2.32	0.90	0.33	0.12
	2003	0.42	1.03	0.31	0.42	0.24	0.24	0.42	0.42	0.42	0.53	3.26	0.92	0.42	0.08
materials	1975	2.58	4.56	3.89	4.38	2.05	1.86	1.50	4.38	2.07	2.58	2.58	5.09	1.90	1.55
	1980	2.83	3.18	2.65	3.89	2.65	2.43	1.84	3.89	1.84	2.68	2.83	3.96	2.50	1.84
	1985	2.56	3.64	3.00	2.92	2.84	2.28	1.74	3.28	1.89	2.27	1.87	4.41	1.13	2.30
	1990	2.26	3.61	3.17	3.43	3.17	2.61	1.96	3.21	2.22	2.61	2.97	4.18	2.35	2.60
	1995	2.41	3.13	2.80	3.52	2.96	2.09	2.41	2.41	2.41	2.41	2.84	3.47	2.82	2.41
	2000	2.45	3.12	2.84	2.84	1.69	1.69	2.45	2.45	2.45	1.69	1.69	3.52	2.84	3.40
	2003	2.41	1.65	2.12	2.41	1.65	1.65	2.41	2.41	2.41	1.65	1.65	3.48	2.41	3.63

Table 3: Relative output efficiency measures $[1/(1+u_i(t))]$ in selected years, the arithmetic average averages of 1975-2003, and the average annual efficiency change rate [ΔEff] in percent

	AUT	CAN	DEN	FIN	FRA	GBR	GER	ITA	NED	NOR	NZL	POR	SWE	USA
1975	0.88	1.00	0.69	0.54	0.87	0.85	1.00	0.60	0.93	0.80	0.51	0.53	0.65	0.92
1980	0.86	0.96	0.72	0.60	0.93	0.81	0.97	0.68	0.93	0.77	0.49	0.57	0.71	0.97
1985	0.85	0.93	0.75	0.64	0.98	0.79	0.93	0.75	0.93	0.77	0.49	0.58	0.76	1.00
1990	0.85	0.90	0.76	0.67	1.00	0.79	0.89	0.80	0.92	0.79	0.49	0.57	0.81	0.99
1995	0.84	0.87	0.76	0.69	0.99	0.81	0.85	0.82	0.91	0.83	0.51	0.54	0.85	0.95
2000	0.84	0.85	0.75	0.69	0.96	0.86	0.81	0.81	0.90	0.90	0.54	0.49	0.87	0.88
2003	0.85	0.84	0.73	0.68	0.93	0.91	0.78	0.78	0.88	0.97	0.57	0.45	0.88	0.83
Average 1975-2003	0.85	0.91	0.74	0.65	0.96	0.82	0.89	0.75	0.92	0.82	0.51	0.54	0.79	0.95
ΔEff (%)	-0.14	-0.61	0.22	0.78	0.24	0.23	-0.88	0.95	-0.19	0.70	0.39	-0.59	1.06	-0.34