



**Stochastic Nonparametric
Envelopment of Data:
Combining Virtues of SFA and DEA
in a Unified Framework**

Timo Kuosmanen

Stochastic Nonparametric Envelopment of Data: Combining Virtues of SFA and DEA in a Unified Framework

Timo Kuosmanen

Economic Research Unit, MTT Agrifood Research Finland, Luutnantintie 13, 00410 Helsinki, Finland.

Tel. +358 9 5608 6309. Fax: + 358 9 5608 6264. E-mail. Timo.Kuosmanen@mtt.fi.

Abstract

The literature of productive efficiency analysis is divided into two main branches: the parametric Stochastic Frontier Analysis (SFA) and nonparametric Data Envelopment Analysis (DEA). This paper attempts to combine the virtues of both approaches in a unified framework. We follow the SFA literature and introduce a stochastic component decomposed into idiosyncratic error and technical inefficiency components imposing the standard SFA assumptions. In contrast to the SFA, we do not make any prior assumptions about the functional form of the deterministic production function. In this respect, we follow the nonparametric route of DEA that only imposes free disposability, convexity, and some specification of returns to scale. From the postulated class of production functions, the proposed method identifies the production function with the best empirical fit to the data. The resulting function will always take a piece-wise linear form analogous to the DEA frontiers. We discuss the practical implementation of the method and illustrate its potential by means empirical examples.

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

JEL Classification: C14, C51, C61, D24

Copyright 2006 by Timo Kuosmanen. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

1. Introduction

The literature of production frontier estimation is voluminous and growing rapidly (see e.g. recent surveys by Cherchye and Post, 2003; Murillo-Zamorano, 2004; and Worthington, 2001, 2004). For decades, this literature has been dominated by two separate branches: the *nonparametric* data envelopment analysis (DEA) and the *parametric* stochastic frontier analysis (SFA). The origins of DEA date back to the seminal paper by Farrell (1957), but its current popularity is largely due to the influential work by Charnes et al. (1978). The late seventies also saw the birth of SFA in the works of Aigner et al. (1977) and Meeusen and van den Broeck (1977), among others. SFA builds directly on the classic econometric ordinary least squares (OLS) and corrected OLS (COLS: Aigner and Chu, 1968) approaches to production function estimation, which relies heavily on the ex ante specification of the functional form. The main attention has been in the decomposition of the residual into a non-negative inefficiency term and an idiosyncratic error. By contrast, DEA has focused on the 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, completely ignoring any stochastic noise in the data. In summary, it is generally accepted that the virtues of SFA lie in the stochastic, probabilistic treatment of inefficiency and noise, while the virtues of DEA lie in its general nonparametric frontier (see e.g. Bauer, 1990; and Seiford and Thrall, 1990).

The objective of this paper is to combine these virtues of SFA and DEA in a unified framework of frontier analysis. More specifically, we propose to combine the DEA-style nonparametric piece-wise linear frontier with the SFA-style decomposition of residuals into stochastic noise and inefficiency, and show that the nonparametric frontier and stochastic composite residual can be pursued simultaneously. The main advantage of our approach to the parametric SFA approach is the independence of the ad hoc parametric assumptions about the functional form of the production function (or cost/distance functions). In contrast to the flexible functional forms, one can impose monotonicity, concavity and

homogeneity constraints without sacrificing the flexibility of the regression function. On the other hand, the main advantage to the nonparametric DEA approach is the better 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, in our method all observations influence the shape of the frontier. Also many standard tools from parametric regression such as goodness of fit statistics and statistical tests are directly applicable in our approach. In summary, our method addresses the main points of critique that are usually presented against SFA and DEA, combining the advantages of them both. Such a powerful frontier framework warrants a catchy name, so we will henceforth refer to the proposed approach as *stochastic nonparametric envelopment of data* (StoNED).

Table 1 presents a classification of the frontier estimation literature into parametric vs. non-parametric approaches and deterministic vs. stochastic approaches (see e.g. Murillo-Zamorano, 2004), with two key references to the established approaches. Thus far, the area of stochastic non-parametric frontier estimation has been an uncharted terrain. The ambitious objective of this paper is to fill in this void corner.¹

Table 1: Classification of the frontier estimation literature

	parametric	non-parametric
deterministic	<p>COLS</p> <p>Aigner and Chu (1968), Timmer (1971)</p>	<p>DEA</p> <p>Farrell (1957) Charnes, Cooper, Rhodes (1978)</p>
stochastic	<p>SFA (MOLS)</p> <p>Aigner, Lovell, Schmidt (1977) Meeusen and van den Broeck (1977)</p>	<p>StoNED</p> <p>this paper</p>

¹ Two important sources of influence outside the realm of frontier estimation warrant acknowledgement: 1) Hildreth's (1954) nonparametric regression approach (see also Kuosmanen, 2006) and 2) Varian's (1985) nonparametric tests of optimizing behavior with measurement error.

This study is by no means the first attempt to combine features of DEA and SFA, but contributes to a long series of prior studies that have pursued similar aims. For example, Park and Simar (1994), Fan et al., (1996), and Park et al. (1998, 2003) have explored semiparametric estimation of SFA models in the context of panel data. On the other hand, the random parameters SFA models (e.g., Tsionas, 2002; Greene, 2005) allow for heterogeneity across firms by introducing firm-specific coefficients that are in common with DEA. On the purely nonparametric side, Cazals et al. (2002) and Aragon et al. (2002) have developed more robust versions of DEA-type estimators, but these approaches still do not allow for rigorous analysis of the stochastic noise. Two recent papers by Kumbhakar et al. (2004) and Henderson and Simar (2005) presented the first fully nonparametric SFA frameworks based on local maximum likelihood and kernel regression. While earlier papers come a long way of combining the virtues of DEA and SFA, they tend to limit to some specific settings and thus lack the generality of DEA and SFA approaches. More importantly, the conceptual link between the parametric and non-parametric branches is still missing: none of the recent nonparametric techniques relates to DEA like SFA relates to COLS.

This paper intends to show that DEA and SFA can be combined together without compromising their attractive features. The main advantage of the present approach to the existing alternatives is its heavy reliance on the established concepts and principles of SFA and DEA without introducing new concepts or tools (such as kernel regression). We build on the standard assumptions that practitioners of SFA and DEA are comfortable with. Thus, readers familiar with classic SFA and DEA approaches will be able to appreciate and apply the proposed approach relatively easily. The conceptual bridges between DEA and SFA are also of considerable instrumental value for the further integration of the field.

The remainder of the paper is organized as follows. Section 2 starts by presenting the StoNED model in the cross-sectional case assuming a single-output technology. Section 3 illustrates the approach by a simulated numerical example. Section 4 expands the analysis to the panel data settings. Section 5 presents some useful extensions that allow one to model alternative assumptions about

returns to scale and multiplicative error terms, introduce environmental variables and multiple outputs, and estimate cost functions. Section 6 presents an illustrative application to industry-level panel data of wholesale and retail sectors in 14 OECD countries. Section 7 presents the concluding remarks. Supplementary technical materials are presented in three appendices.

2. Cross-sectional model

This section presents the cross-sectional StoNED model in the multi-input single-output setting where input vector is denoted by \mathbf{x} and output by y . In this section we represent the production technology by the classic *production function* $f: y = f(\mathbf{x})$, which characterizes the boundary of the production possibility set. Production function is not known a priori, and one of our principal aims is to estimate it from data. Following the DEA literature, we assume that function f belongs to the class of monotonic increasing and concave functions (i.e., the sub-gradients of f satisfy $\nabla f(\mathbf{x}) \geq \mathbf{0}$, $|\nabla^2 f(\mathbf{x})| \leq 0$), denoted by F^2 . In contrast to the SFA literature, no specific functional form for f is assumed a priori. Rather, we will endogenously select function $f \in F^2$ that fits the data best. Thus, our specification of the deterministic production function proceeds along the nonparametric lines of the DEA literature.

We deviate from the DEA approach by introducing a stochastic component following the usual practice of the SFA literature. The observed output y_i for firm i may differ from the value of $f(\mathbf{x}_i)$ by residual $\varepsilon_i = v_i - u_i$, which consists of inefficiency term $u_i > 0$ and idiosyncratic error term v_i , formally,

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

The SFA literature typically assumes $u_i \underset{i.i.d.}{\sim} |N(0, \sigma_u^2)|$ and $v_i \underset{i.i.d.}{\sim} N(0, \sigma_v^2)$ (e.g. Aigner et al., 1977). Other distributions such as gamma or exponential are also sometimes used for the inefficiency term u_i (see e.g. Kumbhakar and Lovell, 2000), but in this paper we restrict to the standard half-normal specification.

Analogous to the modified ordinary least-squares (MOLS) approach of the SFA literature, we can estimate the StoNED model (1) by using the method of moments. The procedure consists of three

steps: 1) least-squares estimation of residuals $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1 \dots \hat{\varepsilon}_n)$, 2) estimation of variance parameters σ_u^2, σ_v^2 based on the second and third moments of the residual distribution, and 3) estimation of inefficiency term based on its conditional distribution.

Note first that the parameter estimates of the classic OLS model used in the MOLS procedure are obtained from the optimal solution to the quadratic programming problem:²

$$\begin{aligned} \min_{\alpha, \beta, \hat{\boldsymbol{\varepsilon}}} \sum_{i=1}^n \hat{\varepsilon}_i^2 \\ \text{s.t.} \\ y_i = \alpha + \beta' \mathbf{x}_i + \hat{\varepsilon}_i \quad \forall i = 1, \dots, n \end{aligned} \quad (2)$$

where α, β denote the intercept and slope coefficients of the linear regression function which are common to all observations (the input-output values y, \mathbf{x} are usually transformed by taking logarithms). The objective function (divided by n) provides an estimate for the error variance. Note that the classic deterministic COLS model by Aigner and Chu (1968) is obtained from the OLS model (2) by imposing the constraint $\hat{\boldsymbol{\varepsilon}} \leq \mathbf{0}$, and interpreting the residuals as inefficiency.

In the StoNED approach we replace the OLS regression by a nonparametric regression technique called *concave nonparametric least squares* (CNLS: see Kuosmanen, 2006, for details). The CNLS differs from the OLS in that it allows for a general nonparametric representation of a monotonic and concave regression function. Formally, the CNLS model can be written as the quadratic programming problem

$$\begin{aligned} \min_{\alpha, \beta, \hat{\boldsymbol{\varepsilon}}} \sum_{i=1}^n \hat{\varepsilon}_i^2 \\ \text{s.t.} \\ y_i = \alpha_i + \beta_i' \mathbf{x}_i + \hat{\varepsilon}_i \quad \forall i = 1, \dots, n \\ y_h \leq \alpha_i + \beta_i' \mathbf{x}_h + \hat{\varepsilon}_h \quad \forall h, i = 1, \dots, n \\ \beta_i' \geq \mathbf{0} \quad \forall i = 1, \dots, n \end{aligned} \quad (3)$$

² See Wang et al. (2004) for further discussion about the relationship between the OLS regression and quadratic programming.

In contrast to the OLS, CNLS allows the intercept and slope coefficients to vary from one firm to another. Similar to the random parameters models of the SFA literature (e.g., Tsionas, 2002; Greene, 2005), there are n different slope vectors $\beta'_i, i=1, \dots, n$. Whereas the random parameters models estimate n different production functions of the same a priori specified functional form, the CNLS regression (3) estimates n tangent hyper-planes to one unspecified production function. The slope coefficients β'_i represent the marginal products of inputs (i.e., the sub-gradients $\nabla f(\mathbf{x}_n)$). The second constraint imposes concavity by applying of a system of inequality constraints known as “Afriat inequalities” (Afriat 1967, 1972; see Kuosmanen, 2006, for further discussion). The third constraint imposes monotonicity.

The CNLS estimate of the production function is a piece-wise linear function that closely resembles the usual DEA frontiers (see Figure 1 below). Yet, the least-squares formulation (3) differs from the usual linear programming formulations of DEA (see e.g. Seiford and Thrall, 1990) in many notable respects. First, the usual DEA formulations solve a separate linear programming problem for each observed firm, whereas problem (3) solves the residuals $\hat{\epsilon}$ simultaneously for all firms (compare with Kuosmanen et al., 2006). Second, the standard DEA models resort to multiplicative efficiency measures defined on a relative (percentage) scale, whereas model (3) uses an additive, absolute-scale departure from the frontier in line with the SFA literature. Third, the usual DEA formulations provide only implicit representations of the production frontier, whereas model (3) explicitly characterizes the sub-gradients of the production function. Finally, the most important difference is that while DEA has one-sided inefficiency term, the CNLS model uses unrestricted residuals $\hat{\epsilon}$.

It is worth to note that if we impose in (3) an extra constraint $\hat{\epsilon} \leq \mathbf{0}$ and interpret the residuals as inefficiency terms (as in the COLS method), we obtain the standard output-oriented variable returns-to-scale (VRS) DEA model as a special case of (3) [see Appendix 1 for details]. Therefore, DEA is a deterministic special case of StoNED in the same way as COLS is a deterministic special case of MOLS in the SFA literature.

Returning to the StoNED model, the CNLS regression provides us with the composite residuals $\hat{\boldsymbol{\varepsilon}}$ which consist of error and inefficiency. To disentangle these two components, we next apply the method of moments and calculate the second and third central moments of residual distributions:

$$m_2 = \sum_{i=1}^n (\hat{\varepsilon}_i - \hat{E}(\varepsilon_i))^2 / n \quad (4)$$

$$m_3 = \sum_{i=1}^n (\hat{\varepsilon}_i - \hat{E}(\varepsilon_i))^3 / n. \quad (5)$$

These moments are consistent estimators of the true moments μ_2, μ_3 , which depend on the variance of the inefficiency term and the error terms in the following manner

$$\mu_2 = \left[\frac{\pi-2}{\pi} \right] \sigma_u^2 + \sigma_v^2 \quad (6)$$

$$\mu_3 = \left(\sqrt{\frac{2}{\pi}} \right) \left[1 - \frac{4}{\pi} \right] \sigma_u^3. \quad (7)$$

Thus, the variances σ_u^2, σ_v^2 are easily estimated based on the moments m_2 and m_3 . The estimation is based on the skewness of the distribution of the compound disturbance $\boldsymbol{\varepsilon}$, which is due to the inefficiency term. Thus, the third moment m_3 should be negative. In practice, it can occur that the CNLS residuals are skewed in the wrong direction (m_3 is positive), in which case the maximum likelihood estimate for the inefficiency term is $\hat{\mathbf{u}} = \mathbf{0}$. According to Greene (1999, p. 105): "One might view this as a built-in diagnostic, since the phenomenon is likely to arise in a badly specified model or in an inappropriate application". It can also occur that the skewness is so great that the $\hat{\sigma}_u$ estimate obtained from (7) based on (5) is greater than m_2 and thus $\hat{\sigma}_v$ would be negative. In that case, Kumbhakar and Lovell (2000) suggest to use $\hat{\sigma}_v = 0$ and attribute all variance to the inefficiency term.

Given the variance estimates, we can use the conditional estimator for the inefficiency term. Jondrow et al. (1982) showed that the conditional distribution of inefficiency u_i given $\hat{\varepsilon}_i$ is a zero-

truncated normal distribution with mean $\mu_* = -\hat{\varepsilon}_i \sigma_u^2 / (\sigma_u^2 + \sigma_v^2)$ 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

$$E(u_i | \hat{\varepsilon}_i) = \mu_* + \sigma_* \left[\frac{\phi(-\mu_* / \sigma_*)}{1 - \Phi(-\mu_* / \sigma_*)} \right], \quad (8)$$

where ϕ is the standard normal density function, and Φ is the standard normal cumulative distribution function. This conditional mean has an intuitive interpretation as the mode of the conditional distribution plus the standard deviation multiplied by the normal hazard function (the ratio in brackets). As an alternative point estimator one may use the mode of the conditional distribution, which is the minimum of μ_* and zero. We can also derive intervals for the inefficiency term based on the conditional distribution (Horrace and Schmidt, 1996): a $100(1-\alpha)$ percent confidence interval is given by $[-\mu_* - z_L \sigma_*, -\mu_* - z_U \sigma_*]$ where $z_L = \Phi^{-1}(1 - (\alpha/2)\Phi(\mu_* / \sigma_*))$ and $z_U = \Phi^{-1}(1 - (1 - \alpha/2)\Phi(\mu_* / \sigma_*))$.

Using the parametrization by Aigner et al. (1977) with $\sigma^2 \equiv \sigma_u^2 + \sigma_v^2$ and $\lambda \equiv \sigma_u / \sigma_v$, the log-likelihood function of the of the StoNED regression can be expressed as

$$\ln L(y | \sigma^2, \lambda) = \frac{n}{2} \ln(2/\pi) - n \ln \sigma + \sum_{i=1}^n \ln \Phi \left[\frac{-\hat{\varepsilon}_i \lambda}{\sigma} \right] - \frac{1}{2\sigma^2} \sum_{i=1}^n \hat{\varepsilon}_i^2. \quad (9)$$

The log-likelihood function is useful for testing additional constraints such as returns to scale (see Section 5 for details) using the likelihood ratio test. In principle, maximum likelihood estimation of the model parameters is possible, but the computational burden can be large compared to the method of moments. Statistical inference on other model parameters such as the marginal products β_i is also possible (see Kuosmanen, 2006, for details).

We conclude this section by noting two critical remarks to the cross-sectional model. First, it is well known in the SFA literature that the conditional mean estimator for the inefficiency term u_i is not statistically consistent (e.g., Greene, 1999). Second, the absolute levels of the inefficiency terms are driven by rather ad hoc assumptions about the functional form of the inefficiency distribution. As Ondrich

and Ruggiero (2001) point out, decomposition of the residual into inefficiency and error terms does not influence the relative efficiency rankings of the firms - whichever distribution is assumed. Thus, if we are interested in the relative efficiency of the firms, the CNLS residuals would be equally good indicators as the decomposed inefficiency terms. Both these problems arise from the fact that we try to disentangle inefficiency from the noise based on merely a single observation per firm. Therefore, resorting to a panel data can help to solve both these problems. We return to the panel data models in more detail in Section 4. But first, consider a simple numerical example.

3. Simulation example

This section illustrates the cross-sectional StoNED model by a simulated numerical example. We assume a sample of 100 firms who employ a concave single-input single-output technology. The input values of the firms were randomly sampled from $Uni[1,11]$, and the efficient output levels were calculated using the production function $y = \ln(x) + 2$. Subsequently, we subtracted from the efficient output level a random inefficiency term $u_i \sim |N(0,0.6^2)|$ and a random error $v_i \sim N(0,0.3^2)$, which results as the observed outputs. Figure 1 illustrates the observed sample by means of the scatter plot diagram. Figure 1 also shows the true production function assumed in the simulations (the broken line).

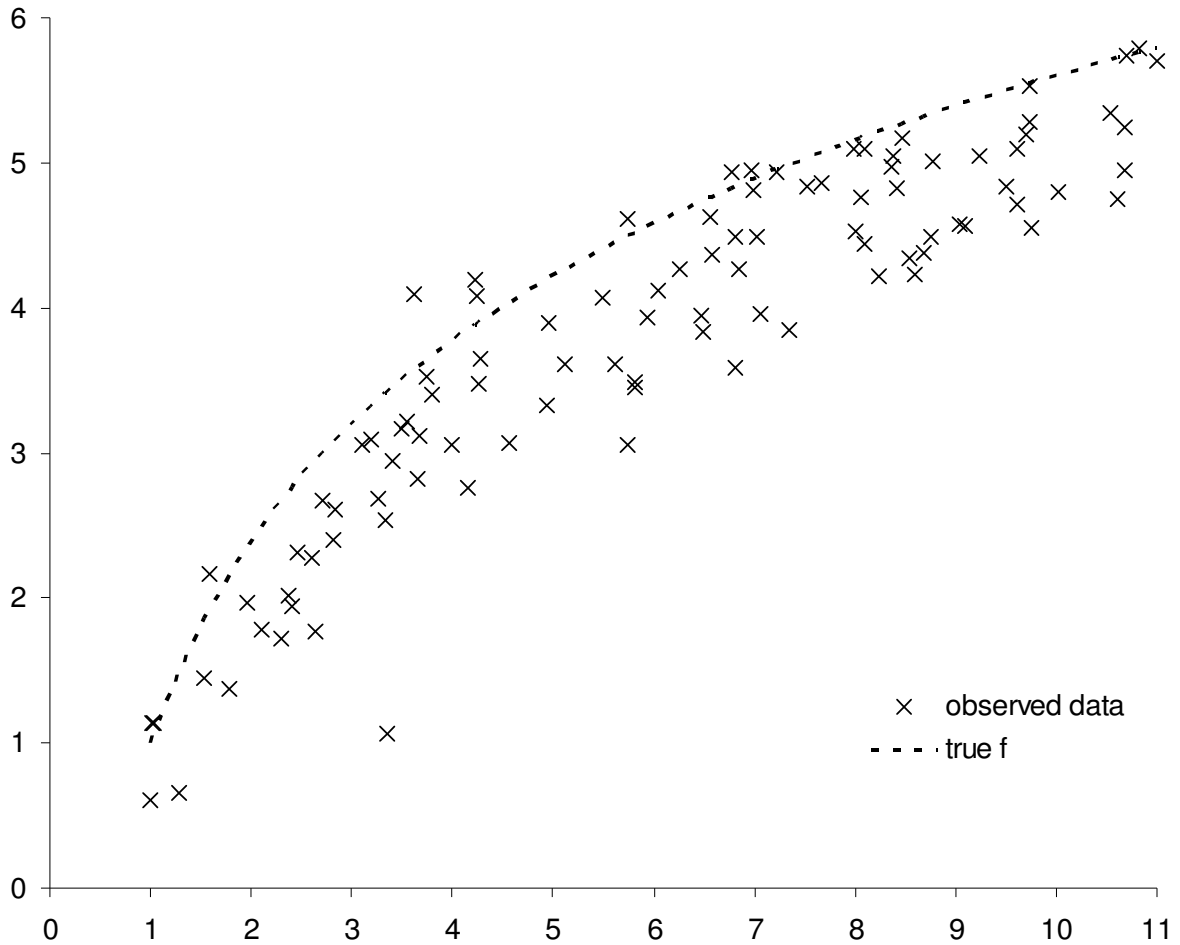


Figure 1: Scatter plot of the simulated data and the true production function (broken line)

We next ran the CNLS regression (3), which gave the R^2 value 0.887. From the distribution of the CNLS residuals we calculated the second and third moments m_2 and m_3 (see (4) and (5)) and estimated the standard deviations σ_u, σ_v (see (6) and (7)). The estimated values $\hat{\sigma}_u = 0.574$, $\hat{\sigma}_v = 0.250$ come rather close to their true values (0.6 and 0.3). Based on these estimates, we calculated the conditional expected value for the firm-specific inefficiency using the Jondrow et al. (1982) result. The correlation coefficient between the estimated (\hat{u}_i) and true inefficiencies (u_i) was 0.746 (rank correlation 0.683). We also used the expected value of the inefficiency distribution (derived by Aigner et al., 1977) $E[u] = 0.458$ to correct for the nonparametric CNLS estimate to obtain the StoNED production function estimate.

Figure 2 illustrates graphically the CNLS regression curve and the StoNED frontier. This figure shows how the CNLS regression estimates the central tendency by means of a piecewise linear function that consists of five different line segments; the number of line segments is endogenously determined in the model to maximize the empirical fit. The StoNED frontier function that takes into account the expected value of inefficiency comes very close to the true production function used in the simulation.

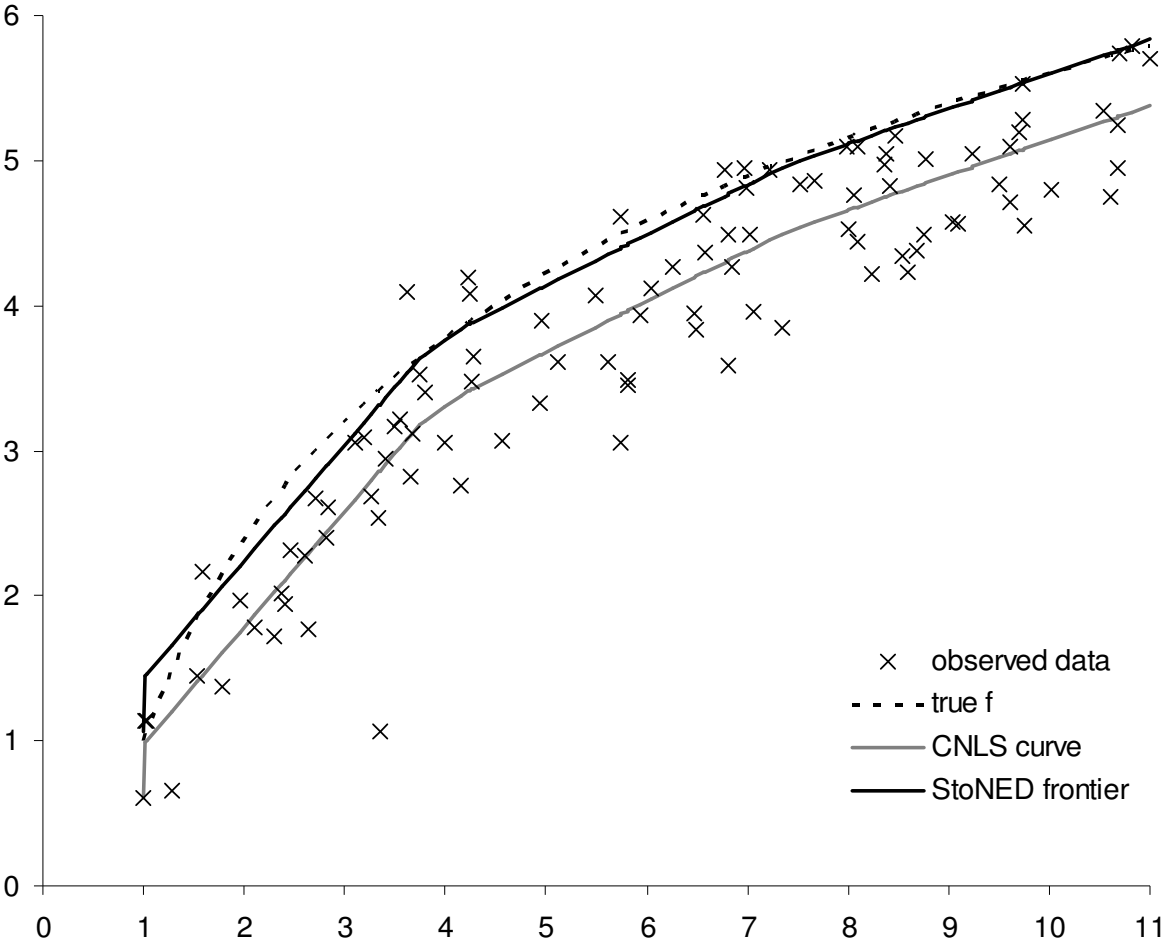


Figure 2: The CNLS regression curve and the StoNED frontier

We next computed the 95 percent confidence intervals for the StoNED frontier applying the results derived by Horrace and Schmidt (1996). Figure 3 plots the confidence intervals in the scatterplot of observation, together with the true frontier and the mean StoNED estimate. As can be seen from the

figure, the true frontier coincides within the confidence interval for the most parts of the frontier. The true output value fits within the 95 percent confidence interval for 86 percent of the observations.

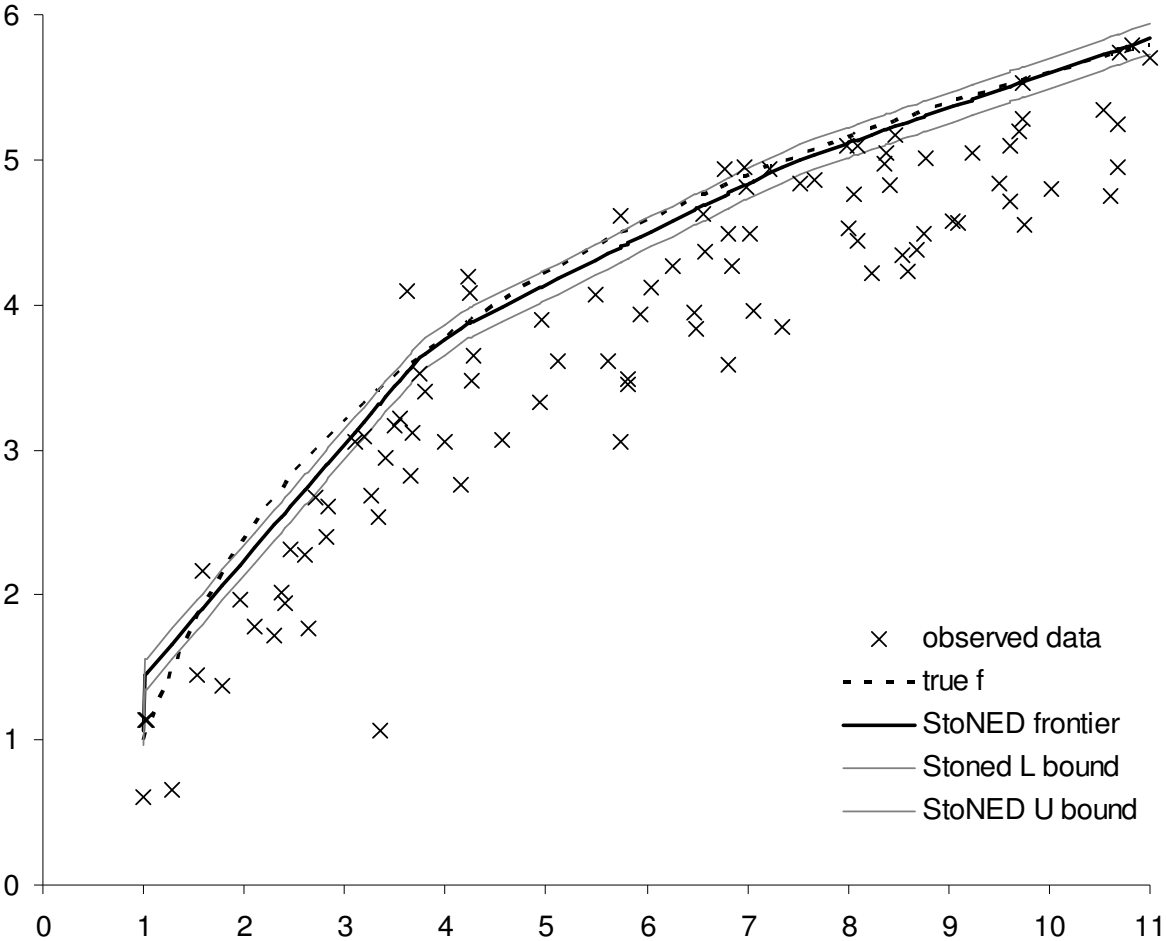


Figure 3: 95 percent confidence intervals of the StoNED frontier.

For comparison, we also considered the standard parametric MOLS procedure assuming the Cobb-Douglas production function. Running the OLS model (2) with the log-transformed data gave an R^2 value of 0.917.³ The slope estimate 0.700 was statistically highly significant (p-value 0.000) and the intercept estimate 0.097 was almost significant (p-value 0.076). However, the residuals were too much skewed: the estimated $\hat{\sigma}_u^2 (= 0.131)$ was much larger than the second moment of the residual

³ This statistic is based on the log-transformed data and is hence not directly comparable with the R^2 of the CNLS regression.

distribution m_2 ($=0.033$), which leaves a negative value for $\hat{\sigma}_v$. The usual approach to resolve this inconsistency is to assign a very small value for $\hat{\sigma}_v$ (here 0.001) and estimate $\hat{\sigma}_u^2$ based on the second moment m_2 . This gives $\hat{\sigma}_u^2 = 0.008$.

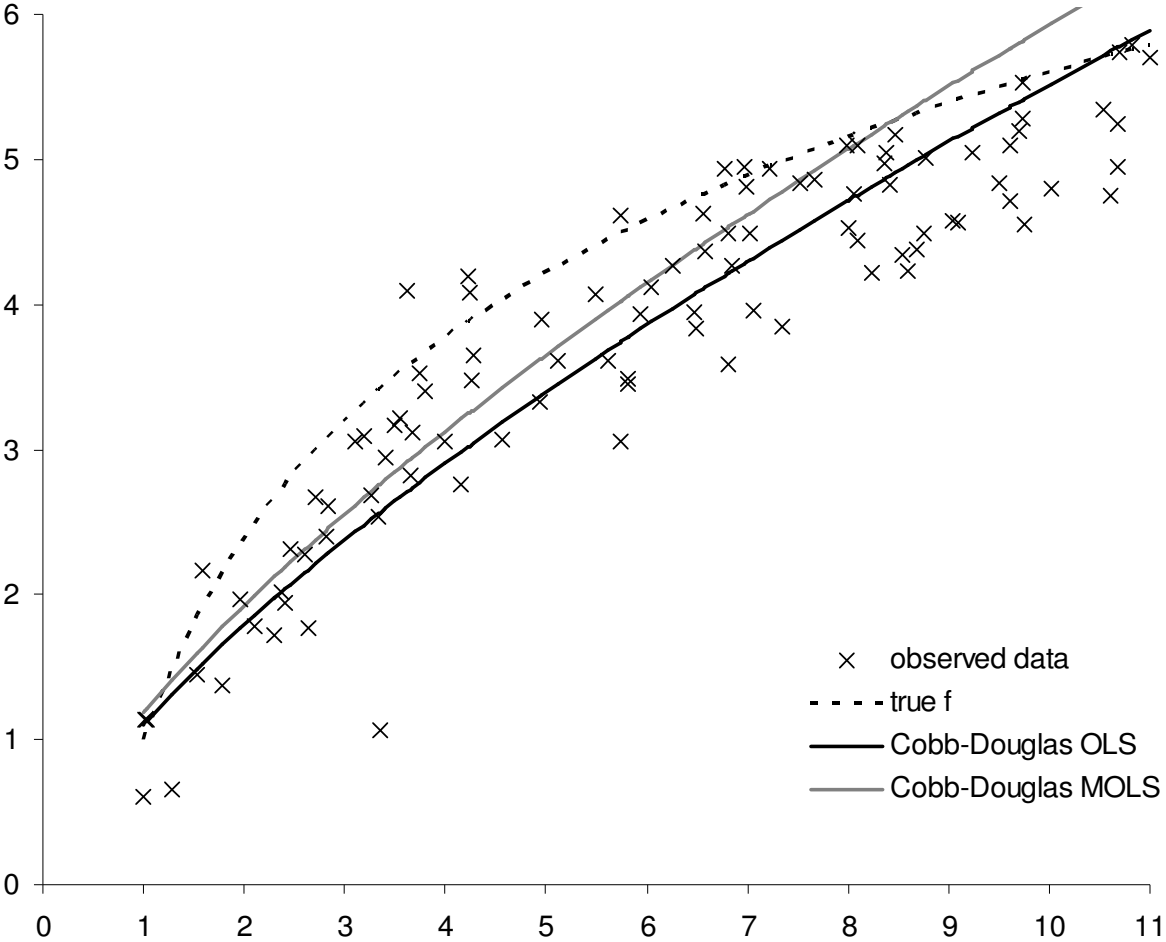


Figure 4: Illustration of the Cobb-Douglas production function estimated with the OLS and MOLS.

Figure 4 illustrates the OLS and MOLS estimates of the Cobb-Douglas production function. The Cobb-Douglas function appears too inflexible to model the curvature of the true production function in this case: the log-transformation of input and output still leaves a non-linear relationship between the variables. This results as heteroskedasticity in residuals, and may also contribute to the too high skewness that contradicts method of moments variance estimators. Note that the MOLS correction that uses a multiplicative inefficiency term makes the MOLS frontier steeper than the OLS curve.

Finally, we tried the MOLS estimation with the translog specification for the production function. This gave the R^2 value of 0.917 which is exactly the same as that of the Cobb-Douglas model. All parameter estimates (intercept -0.177 (p-value 0.027); slope of $\ln x$ 1.210 (p-value 0.000); slope of $(\ln x)^2$ -0.185 (p-value 0.000)) were statistically significant. However, also in the translog case the residuals were too skewed: the estimated $\hat{\sigma}_u^2$ (= 0.120) was much larger than the second moment of the residual distribution m_2 (=0.027), which leaves a negative value for $\hat{\sigma}_v$. Again, we resolve this inconsistency by assigning a very small value for $\hat{\sigma}_v$ (0.001) and estimating $\hat{\sigma}_u^2$ based on the second moment m_2 . This gives $\hat{\sigma}_u^2 = 0.006$.

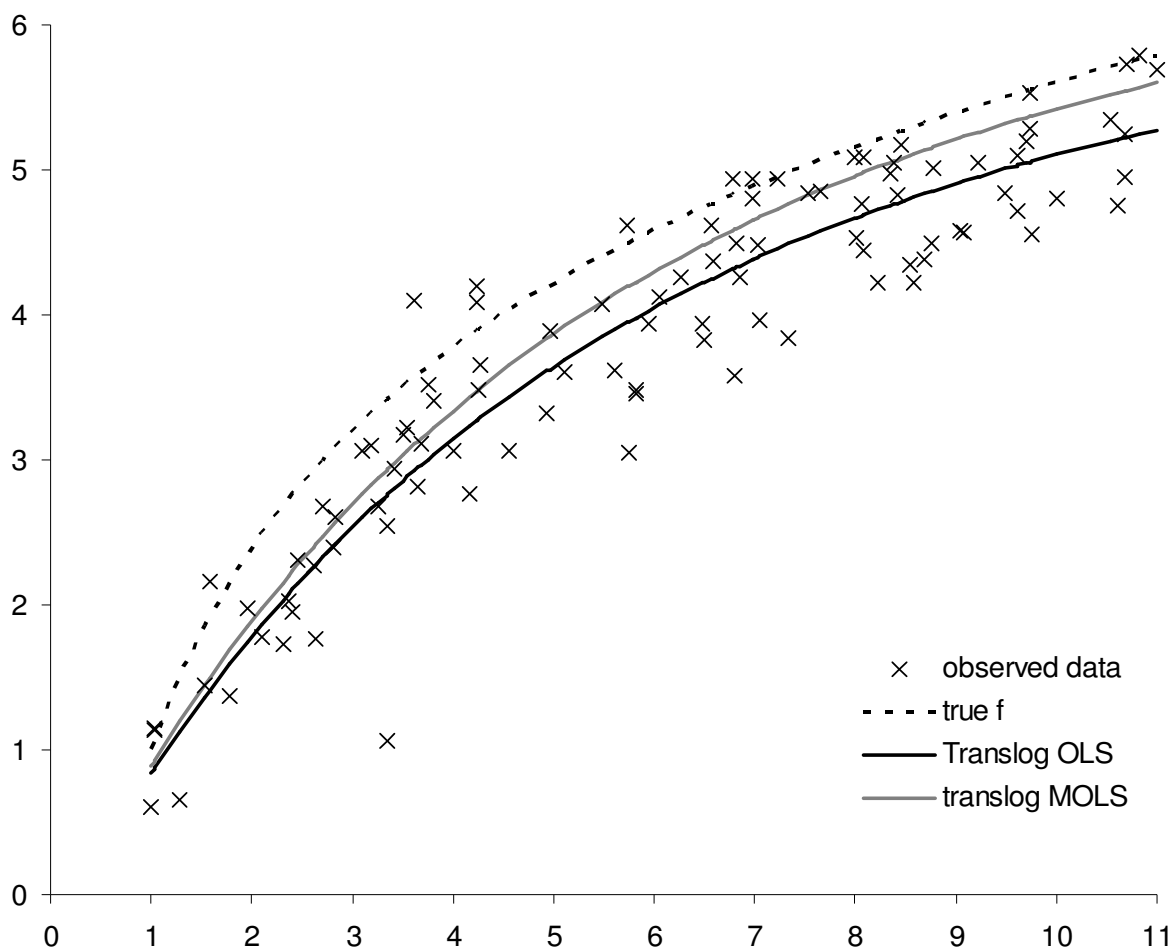


Figure 5: Illustration of the translog production function estimated with the OLS.

Figure 5 illustrates the OLS and MOLS estimates of the translog production function. Translog proves more flexible than the Cobb-Douglas in following the curvature of the true production function. However, the method of moments estimator $\hat{\sigma}_v^2$ underestimates the magnitude of inefficiency even though we assign all of the residual variance to the inefficiency term. Consequently, the expected MOLS frontier falls considerably below the true production function in particular at low input levels.

In conclusion, the StoNED model worked remarkably well in this simple example. This is in line with the more extensive simulation results by Kuosmanen (2006), which compared the performance of the CNLS regression with that of the OLS (assuming Cobb-Douglas and translog functions) in nine alternative scenarios, and found CNLS to be more robust than the parametric approach. While these simulation results are promising and suggest a great potential for the approach, they do not guarantee good performance in all settings. In particular, as the number of input variables increases, ever larger sample sizes are necessary to counter the curse of dimensionality associated with the nonparametric approaches.

4. Panel data models

In this section we assume a balanced panel data covering T time periods indexed by subscript t . Extensions to unbalanced data are straightforward and will not be discussed herein. Observing each firm T times provides additional information that can be helpful for efficiency estimation (see e.g. Kumbhakar and Lovell, 2000). The possibilities to exploit this information in the StoNED panel model will be explored next.

4.1 Freely varying efficiency and technology

The DEA treatments typically choose to ignore the panel structure even if it is available (a notable exception is Ruggiero, 2004). Indeed, one could treat each time period as a unique and independent cross-section, and proceed with estimation as described in Section 2. This approach does not impose

any structure on the inter-temporal changes in the production function (i.e., technical progress or regress) or efficiency – they both can vary freely from one period to another. This might be a desirable feature when the model results are used for estimating and decomposing total factor productivity growth using the Malmquist productivity indices. However, the problems of the cross-sectional model noted above (i.e., inconsistency of inefficiency estimator, dependence on the parametric assumptions) remain. To circumvent these problems, one might impose additional structure on the dynamic patterns of technical progress or efficiency changes.

4.2 Constant efficiency and technology

The opposite extreme is to assume that both the production technology and the efficiency levels remain constant over time; these assumptions give rise to the basic fixed effects treatment in the SFA literature. In the StoNED framework, constant efficiency and technology facilitate a fully nonparametric estimation of efficiency: we can relax the assumption about half-normally distributed inefficiency. We can write the fixed effects panel data variant of the CNLS model (3) as follows:

$$\begin{aligned}
& \min_{\alpha_i, \beta_j, u_i, v} \sum_{t=1}^T \sum_{j=1}^n v_{it}^2 \\
& \text{s.t.} \\
& y_{it} = \alpha_{it} + \beta'_j \mathbf{x}_{it} - u_i + v_{it} \quad \forall i = 1, \dots, n; \quad \forall t = 1, \dots, T \\
& y_{hs} \leq \alpha_{it} + \beta'_j \mathbf{x}_{hs} - u_i + v_{hs} \quad \forall h, j = 1, \dots, n; \quad \forall s, t = 1, \dots, T \\
& \beta'_j \geq \mathbf{0} \quad \forall j = 1, \dots, n
\end{aligned} \tag{10}$$

This least-squares problem minimizes the sum of squares of the error terms v_{it} , and treats the one-sided inefficiency term u_i as a firm-specific constant. Note that the coefficients α_{it}, β_{it} are specific to each firm and period; the variable parameter models in the SFA literature impose firm-specific coefficients that do not change over time. In contrast to the fixed effects OLS model, we cannot incorporate the inefficiency term to the varying intercept. Nevertheless, the absolute levels of inefficiency terms are not identifiable: any increase in u_i can be offset by a parallel increase of α_{it} intercepts for all $t = 1, \dots, T$.

The standard Gabrielsen-Greene approach (Gabrielsen, 1975; Greene, 1980) uses the benchmark value of zero for the most efficient firm in the sample and normalize the inefficiency terms as

$$\hat{u}_i = u_i - \min_h u_h. \quad (11)$$

A weakness of this additive, absolute-scale normalization is that it does not take into account the magnitude of inefficiency relative to the output size. As an alternative approach we propose to normalize the Shephard output distance function (or the inverse of the Farrell efficiency measure) to be less than or equal to one using

$$D_t^o(\mathbf{x}_{it}, y_{it}) = \frac{\hat{y}_{it}/(\hat{y}_{it} + u_i)}{\max_h [\hat{y}_{ht}/(\hat{y}_{ht} + u_h)]}, \quad (12)$$

where $\hat{y}_{it} = y_{it} - v_{it}$ denotes the fitted output values. Note that the two normalizations can yield different results.⁴ We prefer the latter approach that proportions the inefficiency term to the fitted output level, and thus adjusts for possible differences in the firm size.

The main advantages of the constant efficiency approach include the fully nonparametric treatment of inefficiency (i.e., no assumption about its distribution is needed) and the statistical consistency of the inefficiency estimator. The main disadvantage of the approach is that it does not allow for inter-temporal changes in efficiency levels or production technology. The approach can be very useful for applications involving high-frequency (i.e., weekly, monthly) data over a relatively short time span, but in long panels the assumptions of constant efficiency and technology seem untenable.

4.3 Polynomial model of efficiency or technical change

Instead of a constant inefficiency term u_i , one could specify the inefficiency term as a polynomial function of time ($u_i(t)$). In the SFA literature, the most commonly used functional forms are the linear

⁴ The following numerical example illustrates. Assume three firms with outputs $y_1=1$, $y_2=10$, and $y_3=100$, and let $u_1=0.5$, $u_2=2$, and $u_3=5$. Thus, using additive normalization we have $\hat{u}_1=0$, $\hat{u}_2=1.5$, and $\hat{u}_3=4.5$. By contrast, using the proportional normalization (assuming errors v are zero), we obtain $y_1/(y_1+u_1)=2/3$, $y_2/(y_2+u_2)=5/6$, and $y_3/(y_3+u_3)=100/105$, and thus $D_1=0.7$, $D_2=0.875$, and $D_3=1$.

($u_i(t) = a_i + b_i t$) and quadratic ($u_i(t) = a_i + b_i t + c_i t^2$) [e.g., Cornwell et al., 1990]. Technically, these variants can be directly inserted to the CNLS problem (10) in the place of constant u_i . The output distance function estimates relative to the benchmark technology of period s can be subsequently calculated similar to (12) as

$$D_s^o(\mathbf{x}_{it}, y_{it}) = \frac{\hat{y}_{it}/(\hat{y}_{it} + u_i(t))}{\max_h [\hat{y}_{hs}/(\hat{y}_{hs} + u_h(s))]} \quad (13)$$

Provided that the semiparametric specification of efficiency term is appropriate, the approach provides consistent estimates and does allow for inter-temporal changes in efficiency. Note that the case of constant efficiency and technology described in Section 3.2 is a special case of this model.

This semiparametric specification also features Hicks neutral technical progress (or regress) which preserves the shape of the input isoquants. One can use the output distance function (13) for calculating the Malmquist productivity index and its components using the standard formulas (see e.g. Färe et al., 1994). Hicks neutrality implies that the rate of technical change is uniform across all firms but varies over time. Efficiency changes can vary both across firms and across time periods.

4.4 Rolling window analysis

Rolling window analysis is one of the few panel data techniques known in the DEA literature (e.g., Asmild et al., 2004). The main idea is simple: the technology of period t is estimated using a cross-sectional model with data that includes observations from periods $[t - w/2, t + w/2]$ where w is the window length. The main appeal of the rolling window approach is that it does not impose any limitations on technical progress (regress) or changes in efficiency over time; it can also capture biases in technological progress. Its main drawback is its ad hoc nature. If the rate of technical progress is speedy, then including observations from multiple time periods in the same sample can bias the estimates. For example, the assumptions of the stochastic component such as independence and identical distribution of the noise term are likely violated. While in DEA the inefficient observations do not

influence the shape of the frontier, the StoNED model is similar to SFA in that all observations count. The rolling window approach may be a useful approach for smoothing random temporary fluctuations in applications where technical progress occurs relatively slowly (consider e.g. agriculture).

5. Extensions

5.1 Returns to scale

All models presented thus far exhibit variable returns to scale. In many applications it is meaningful to impose constant or non-increasing returns. Imposing further structure on returns to scale is relatively straightforward in the StoNED framework. We only need to add the following constraints to the intercept term of the relevant CNLS regression:

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

In the CRS case one can delete the intercept term (α_i) altogether.

5.2 Multiple outputs

Thus far we have restricted to the single-output setting. One important strength of the DEA is its ability to handle joint production of multiple outputs. In the SFA and StoNED frameworks, multiple outputs can be handled by means of distance functions. For generality, consider the directional distance function (Chambers et al., 1996, 1998) that encompasses almost all known distance metrics as its special cases.

The directional distance function is defined as

$$DD(\mathbf{x}, \mathbf{y}; \mathbf{g}^x, \mathbf{g}^y) = \sup \left\{ \delta \mid (\mathbf{x} - \delta \mathbf{g}^x, \mathbf{y} + \delta \mathbf{g}^y) \in T \right\}. \quad (14)$$

where T denotes a closed and compact production possibility set and $(\mathbf{g}^x, \mathbf{g}^y) \geq \mathbf{0}$ is an ex ante specified direction vector that defines the projection path to the frontier (e.g., setting $(\mathbf{g}^x, \mathbf{g}^y) = (\mathbf{0}, \mathbf{y})$,

we obtain the Shephard output distance function as $1+DD(\mathbf{x},\mathbf{y};\mathbf{0},\mathbf{y})$. As Chambers et al. (1998) have shown, the directional distance function has an equivalent dual formulation as

$$DD(\mathbf{x},\mathbf{y};\mathbf{g}^x,\mathbf{g}^y) = - \inf_{\substack{(\mathbf{x}^*,\mathbf{y}^*) \\ (\boldsymbol{\omega},\boldsymbol{\rho}) \geq \mathbf{0}}} \{(\boldsymbol{\rho}'\mathbf{y} - \boldsymbol{\omega}'\mathbf{x}) - (\boldsymbol{\rho}'\mathbf{y}^* - \boldsymbol{\omega}'\mathbf{x}^*) \mid (\mathbf{x}^*,\mathbf{y}^*) \in T; \boldsymbol{\rho}'\mathbf{g}^x + \boldsymbol{\omega}'\mathbf{g}^y = 1\}, \quad (15)$$

where vectors $(\boldsymbol{\omega},\boldsymbol{\rho})$ represent the shadow prices of inputs and outputs. This dual formulation proves useful for estimation within the StoNED framework.

Since the distance to the frontier depends on both inefficiency and noise, a common approach is to specify the value of the distance function as a stochastic composite residual $\delta_i = v_i - u_i$ that is subsequently decomposed to inefficiency term $u_i > 0$ and idiosyncratic error v_i . Thus, we can apply the method of moments estimation described in Section 2. In the case of the directional distance function, the CNLS regression can be written as

$$\begin{aligned} & \min_{\boldsymbol{\rho}, \boldsymbol{\omega}, \boldsymbol{\varepsilon}} \sum_{i=1}^n \delta_i^2 \\ & \text{s.t.} \\ & \delta_i - \delta_h \geq (\boldsymbol{\rho}'_i \mathbf{y}_h - \boldsymbol{\omega}'_i \mathbf{x}_h) - (\boldsymbol{\rho}'_i \mathbf{y}_i - \boldsymbol{\omega}'_i \mathbf{x}_i) \quad \forall h, i = 1, \dots, n \\ & \boldsymbol{\rho}'_i \mathbf{g}_y + \boldsymbol{\omega}'_i \mathbf{g}_x = 1 \quad \forall i = 1, \dots, n \\ & (\boldsymbol{\rho}'_i, \boldsymbol{\omega}'_i) \geq \mathbf{0} \quad \forall i = 1, \dots, n \end{aligned} \quad (16)$$

This quadratic programming problem minimizes the square-sum of the composite residuals, subject to the constraints derived from the dual expression (15). Further details of the derivation are presented in Appendix 2. The decomposition of the composite residual into the inefficiency and error terms is subsequently conducted as described in Section 2.

5.3 Multiplicative errors

Thus far we have applied an additive specification of the composite error following the standard model of errors in variables. This specification implicitly assumes that the error variance is independent on the firm size. Violations of this assumption may show up as heteroskedasticity across different sized firms.

Such heteroskedasticity may be alleviated by assuming a multiplicative error structure. Note that the SFA literature typically postulates a multiplicative error structure of form

$$y_i = f(\mathbf{x}_i) \cdot \exp^{\varepsilon_i}, \quad i = 1, \dots, n, \quad (17)$$

which is converted to the additive form by taking logarithms of both sides of the equation. For our purposes, it is more convenient to use the following multiplicative specification

$$y_i = f(\mathbf{x}_i) \cdot (1 - \varepsilon_i)^{-1}, \quad i = 1, \dots, n, \quad (18)$$

which can be equivalently expressed as $y_i(1 - \varepsilon_i) = f(\mathbf{x}_i)$, and further $y_i = f(\mathbf{x}_i) + y_i \varepsilon_i$. Thus, the cross sectional model can be estimated with the multiplicative error specification by running the following CNLS regression

$$\begin{aligned} & \min_{\alpha, \beta, \hat{\varepsilon}} \sum_{i=1}^n \hat{\varepsilon}_i^2 \\ & \text{s.t.} \\ & y_i = \alpha_i + \beta_i' \mathbf{x}_i + y_i \hat{\varepsilon}_i \quad \forall i = 1, \dots, n \\ & y_h \leq \alpha_i + \beta_i' \mathbf{x}_h + y_h \hat{\varepsilon}_h \quad \forall h, i = 1, \dots, n \\ & \beta_i' \geq \mathbf{0} \quad \forall i = 1, \dots, n \end{aligned} \quad (19)$$

and applying the method of moments to decompose the composite residuals.

5.4 Firm characteristics and environmental variables

It is popular to estimate the impact of the firm characteristics and environmental factors on the firm efficiency in a parametric or semi-parametric fashion (e.g., Wang and Schmidt, 2002; and Simar and Wilson, 2006). Such estimation can be easily incorporated into the present StoNED framework. Denote the vector of environmental variables by \mathbf{z} , which may include binary dummy variables as its elements. Suppose for simplicity that the environmental variables have a linear effect on efficiency as

$$u_i(\mathbf{z}_i) = u_i + \mathbf{\Upsilon}' \mathbf{z}_i \quad (20)$$

where u_i is a firm-specific random variable and multipliers $\boldsymbol{\gamma}$ describe the efficiency impacts of the environmental variables, which are assumed to be uniform across all firms. Thus, the model to be estimated can be written as

$$y_i = f(\mathbf{x}_i) - (u_i + \boldsymbol{\gamma}'\mathbf{z}_i) + v_i = f(\mathbf{x}_i) + \boldsymbol{\gamma}'\mathbf{z}_i + \varepsilon_i, \quad i = 1, \dots, n. \quad (21)$$

Since the functional form of the production function f is unknown but the effects of the environmental variables are assumed to linear and common across all firms, we can resort to a semi-parametric version of the cross-sectional CNLS regression. This regression model can be written as

$$\begin{aligned} & \min_{\alpha_n, \boldsymbol{\beta}_n, \hat{\boldsymbol{\varepsilon}}} \sum_{i=1}^n \hat{\boldsymbol{\varepsilon}}_i^2 \\ & \text{s.t.} \\ & y_i = \alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i + \boldsymbol{\gamma}'\mathbf{z}_i + \hat{\boldsymbol{\varepsilon}}_i \quad \text{for all } i = 1, \dots, n \\ & y_h \leq \alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_h + \boldsymbol{\gamma}'\mathbf{z}_h + \hat{\boldsymbol{\varepsilon}}_h \quad \text{for all } h, i = 1, \dots, n \\ & \boldsymbol{\beta}'_i \geq \mathbf{0} \quad \text{for all } i = 1, \dots, n \end{aligned} \quad (22)$$

In this model, the parameters $\alpha_i, \boldsymbol{\beta}_i$ describing the production function are specific to each firm i , whereas the coefficients $\boldsymbol{\gamma}$ describing the effects of environmental variables are uniform across firms. Given the optimal solution to problem (22), the composite residuals can be decomposed as described in Section 2.

The linear specification of the environmental variables was presented here for sake of example; nonlinear and multiplicative effects of environmental variables can also be modeled in the semi-parametric fashion in the StoNED framework. Nonlinear effects of environmental variables could be modeled directly analogous to the treatments of nonlinear polynomial time trends (Section 4.3). Multiplicative effects of the environmental variables could be treated similar to the multiplicative errors (Section 5.3).

5.5 Cost functions

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

$$C(y, \mathbf{w}) = \min_{\mathbf{x}} \{ \mathbf{w}'\mathbf{x} \mid f(\mathbf{x}) = y \}, \quad (23)$$

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

The observed costs C_i ($i = 1, \dots, n$) differ from the cost function due to stochastic term (ε_i) which is the sum of the inefficiency term (u_i) and the noise term (v_i), that is,

$$C_i = C(y_i, \mathbf{w}_i) + \varepsilon_i = C(y_i, \mathbf{w}_i) + u_i + v_i. \quad (24)$$

Maintaining the stochastic assumptions of Section 2, the cross-sectional cost frontier can be estimated analogous to the production function procedure. The cost based CNLS problem can be written as

$$\begin{aligned} & \min_{\alpha_n, \beta_n, \chi} \sum_{i=1}^n \varepsilon_i^2 \\ & \text{s.t.} \\ & C_i = \alpha_i + \beta_i' \mathbf{w}_i + \chi_i y_i + \varepsilon_i \text{ for all } i = 1, \dots, n \\ & C_h \geq \alpha_i + \beta_i' \mathbf{w}_h + \chi_i y_h + \varepsilon_h \text{ for all } h, i = 1, \dots, n \\ & \beta_i' \geq \mathbf{0}, \chi_i \geq 0 \text{ for all } i = 1, \dots, n \end{aligned} \quad (25)$$

where coefficients χ_i represent the marginal cost of output, coefficients β_i indicate the marginal cost of input prices (which depends on the input substitution possibilities), and intercepts α_i have an interpretation as the fixed cost. The composite residual can be decomposed along the lines described in Section 2. Note the changed sign of the inefficiency component and the direction of skewness. The interpretation of the inefficiency term also changes: it here represents (overall) cost efficiency that captures both technical and allocative efficiency. Extending the cost function estimation to multi-output settings is straightforward; the cost function provides another useful way to handle multi-output technologies.

6. Application: productive efficiency in the wholesale and retail trade sector

6.1 Motivation and setup

The purpose of this section is to illustrate the application of the StoNED model in a real-world data and show how the various tools and extensions discussed in Sections 4 and 5 can be combined together. The present application uses the industry-level data of the wholesale and retail trade sectors in thirteen OECD countries.⁵ A balanced panel of industry level data for the period 1975-2003 was collected from the OECD's Structural Analysis database STAN (<http://www.oecd.org>). The list of countries (abbreviation) is the following: Austria (AUT), Canada (CAN), Denmark (DEN), Finland (FIN), France (FRA), Germany (GER, 1975-1990 West Germany), Italy (ITA), Netherlands (NLD), New Zealand (NZL), Norway (NOR), Portugal, Sweden (SWE), and United Kingdom (GBR), and United States (USA). We will henceforth refer to the trade sector of a specific country by these three-letter abbreviations.

The input and output variables are the following

Output: gross sales (y) (in Bill. €, prices of 2000)

Inputs: 1) gross capital stock (K) (in Bill. €, prices of 2000),⁶

2) labor (L) (Bill. hours),⁷

3) intermediate inputs (M) (Bill. €, prices of 2000).

Table 2 illustrates the data by providing the output and input values for all countries in years 1975, 1990, and 2003. The data shows that the trade sector has grown fast in all countries during this 29 year period. USA is the largest sector, while GBR overtook GER as the second largest.

⁵ This industry classification covers the ISIC codes 51-52. Data for Canada, Sweden, and New Zealand also include ISIC code 50 (the sale and repair of motor vehicles and the retail sale of fuels).

⁶ For Austria, West Germany, Italy, Netherlands, Norway, Portugal, and Sweden, the capital stock was estimated based on the data of gross fixed capital formation using the perpetual inventory method assuming 8% depreciation rate.

⁷ The labor hours were estimated based on the employment data (full-time equivalents) for the following countries: Austria, Germany, Italy, Portugal, United Kingdom.

Table 2: Output and input data for 1975, 1990, and 2003

	output			labor (L)			capital (K)			interm. inputs (M)		
	1975	1990	2003	1975	1990	2003	1975	1990	2003	1975	1990	2003
AUT	8.5	23.3	38.8	690.7	754.5	812.2	8.8	14.2	22.7	2.6	8.8	16.8
CAN	19.0	72.7	143.4	2996.5	4231.5	5024.5	22.3	33.2	67.6	3.9	19.7	52.3
DEN	6.3	20.5	33.5	596.6	522.6	545.8	24.0	31.2	41.3	2.0	8.1	15.0
FIN	2.6	11.4	20.0	519.6	487.3	427.9	13.3	21.1	21.0	1.0	4.1	8.9
FRA	32.9	123.3	171.1	1292.2	1423.3	1729.3	74.9	151.2	225.0	11.5	37.4	57.6
GBR	47.7	133.5	314.1	2430.0	2763.6	3225.2	75.9	149.6	270.5	16.9	57.0	152.5
GER	74.2	165.4	249.5	3515.2	4399.2	5642.0	50.7	89.3	249.5	25.6	62.9	105.4
ITA	10.9	112.1	217.8	3417.4	4320.2	4415.3	46.1	110.9	183.1	3.6	37.3	88.0
NLD	16.1	42.2	74.9	921.7	1140.1	1412.0	13.3	25.2	45.5	5.4	14.9	30.1
NZL	2.5	12.7	21.3	330.8	1680.8	2532.0	5.0	7.1	8.2	1.3	7.0	10.7
NOR	4.1	14.3	25.6	356.3	364.7	370.8	4.6	14.3	13.2	1.4	5.8	11.9
POR	0.6	10.0	21.7	2576.9	3090.7	3680.2	1.7	6.0	18.3	0.2	4.0	10.4
SWE	5.8	25.1	37.3	982.8	996.3	911.5	4.1	13.5	26.3	2.7	9.6	12.7
USA	375.2	1051.5	1968.1	18735.8	28340.4	33930.5	243.8	701.8	3575.7	128.5	332.2	603.7

6.2 Model specification

Our preliminary data analysis showed the sample size to be too small for a meaningful cross-sectional estimation. For example, the standard DEA models identify 10-13 efficient sectors out of 14 depending on year and the model assumptions. On the other hand, the time horizon spans 29 years so it is necessary to account for technical progress. For these reasons, we resort to a semi-parametric model where the production technology is modeled in nonparametric fashion and Hicks neutral technical change is modeled by including a quadratic trend as discussed in Section 3.3. Constant returns to scale were postulated to assess small and large sectors against the same benchmark. We first estimated the model using the additive error specification, but this resulted in major heteroskedasticity across sectors. Given the vast differences in the sizes of the sectors (as seen in Table 1), a multiplicative error structure (discussed in Section 4.3) was found to be more appropriate. Thus, our preferred model is described by the following least squares problem

$$\begin{aligned}
 & \min_{\alpha_i, \beta_i, u, v} \sum_{t=1}^T \sum_{l=1}^n v_{it}^2 \\
 & \text{s.t.} \\
 & y_{it} = [\beta_{Lit}L_{it} + \beta_{Kit}K_{it} + \beta_{Mit}M_{it}] + [a_i + b_it + c_it^2] + y_{it}v_{it} \quad \forall i = 1, \dots, n; \quad \forall t = 1, \dots, T \\
 & y_{hs} \leq [\beta_{Lit}L_{hs} + \beta_{Kit}K_{hs} + \beta_{Mit}M_{hs}] + [a_h + b_h t + c_h t^2] + y_{hs}v_{hs} \quad \forall h, i = 1, \dots, n; \quad \forall s, t = 1, \dots, T \\
 & \beta_{Lit}, \beta_{Kit}, \beta_{Mit} \geq 0 \quad \forall i = 1, \dots, n; \quad \forall t = 1, \dots, T
 \end{aligned} \tag{25}$$

The coefficients β characterize a common CRS production function. The quadratic sector-specific time trend within the brackets describes the Hicks neutral technical progress; exceeding the growth rate of the frontier countries is interpreted as catching up. Finally, the error terms v are multiplied by the corresponding output values to model errors proportional to the output level. The GAMS code used for solving problem (25) is presented in the Appendix 3.

6.3 Results

The StoNED model (25) gave a good empirical fit: the R^2 statistic was as high as 0.999 and log-likelihood value -2772.51. The large value of the R^2 statistic (almost 1) is due to the fact that there are large differences in the output sizes of the sectors (e.g., output of FIN is only about 1% of that of USA), and the input use and the time trend explain most of the variation in the output. The total monetary value of the residuals amounts to 6.7 Mill. Euro, which is a considerable amount of money in absolute sense, but of minuscule scale compared to the variation in the output size.

Note that the marginal product (MP) estimates ($\beta_L, \beta_K, \beta_M$ coefficients) of the StoNED model are specific to each country and year. To provide further insight to the working of the model, we report the complete set of MP coefficients in Tables 3-5 for all observations.

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 and GBR appear unrealistically high and values of NZL and POR too low. Note however that these shadow prices need not be unique; the same applies to the input-output multipliers in the standard DEA models.

Table 3: Marginal product of labor (β_L , €/hour), StoNED regression (25)

year	AUT	CAN	DEN	FIN	FRA	GBR	GER	ITA	NED	NOR	NZL	POR	SWE	USA
1975	6,8	2,0	6,8	6,8	11,5	8,4	8,5	2,0	7,3	6,8	6,8	0,0	2,8	7,6
1976	6,8	2,0	6,8	6,8	13,1	8,4	8,5	6,8	7,6	6,8	6,9	0,0	2,8	7,6
1977	6,9	4,0	11,5	6,8	13,1	8,4	8,5	6,8	7,7	6,9	6,8	0,0	2,8	7,7
1978	6,9	4,0	11,5	6,8	13,1	8,8	8,5	6,8	8,5	6,9	6,8	0,0	2,3	8,5
1979	7,1	4,0	11,5	6,8	13,1	11,6	8,5	6,8	8,5	6,9	7,1	0,0	2,3	8,5
1980	7,3	4,0	11,5	6,8	13,1	12,9	8,6	6,8	8,5	6,9	6,9	0,0	2,3	8,5
1981	7,3	4,0	11,5	6,8	13,1	13,1	8,6	6,8	8,5	6,9	2,8	0,0	2,3	8,5
1982	7,7	4,0	11,5	6,8	13,1	13,1	8,6	6,8	8,5	6,9	2,8	0,0	2,3	7,7
1983	7,7	4,0	12,9	6,8	13,1	13,1	8,6	6,8	8,5	8,8	2,8	0,1	2,3	7,7
1984	8,5	4,0	13,1	6,8	13,1	16,8	8,6	6,8	8,6	12,4	2,3	0,0	2,3	8,5
1985	8,5	4,0	13,1	6,8	13,1	16,8	10,4	6,9	10,4	12,4	2,3	0,0	7,7	8,5
1986	8,5	4,0	13,1	10,2	13,1	16,8	8,5	6,9	13,8	12,4	2,3	0,0	7,7	10,5
1987	8,5	4,0	13,1	11,5	13,1	19,7	8,6	6,9	13,8	12,4	2,3	0,1	7,7	10,1
1988	8,5	4,0	13,1	11,5	13,1	19,9	10,4	6,9	13,8	12,5	2,3	0,2	8,5	10,1
1989	8,6	7,1	13,1	11,5	13,1	19,9	10,4	8,8	13,8	16,8	2,3	0,0	8,5	11,8
1990	8,6	7,7	13,1	11,5	13,1	19,9	10,4	11,3	13,8	16,8	2,3	0,0	8,6	11,7
1991	8,6	7,7	13,1	13,1	19,9	19,9	17,4	12,4	13,8	19,9	2,3	0,0	8,6	11,9
1992	13,8	7,7	13,1	13,1	19,3	19,9	17,4	12,4	13,8	19,9	2,3	1,1	8,5	11,9
1993	13,8	7,7	13,1	13,1	59,7	19,9	17,4	12,5	13,8	19,9	2,3	0,3	8,5	11,9
1994	13,8	7,7	14,9	13,1	18,5	19,9	17,4	12,5	13,8	19,9	1,1	1,1	8,5	16,0
1995	13,8	7,7	16,8	13,1	59,7	19,9	19,9	17,4	13,8	19,9	1,1	2,3	8,5	13,8
1996	13,8	7,7	16,8	13,1	59,7	19,9	17,7	16,8	13,8	19,9	1,1	2,8	8,5	17,4
1997	13,8	7,7	19,9	13,1	59,7	19,9	19,9	19,9	13,8	19,9	1,1	2,8	8,5	16,8
1998	13,8	7,7	19,9	16,8	59,7	19,9	19,9	19,9	13,8	19,9	1,1	2,3	11,7	13,1
1999	13,8	7,7	19,9	16,8	59,7	19,9	19,9	19,9	13,8	22,3	1,1	2,8	11,7	13,1
2000	13,8	8,5	19,9	19,9	59,7	19,9	19,9	19,9	13,8	13,0	3,8	2,8	11,8	13,1
2001	11,7	8,5	19,9	19,9	59,7	59,7	19,9	19,9	17,4	16,2	1,1	2,8	11,9	13,1
2002	16,5	9,3	19,9	19,9	59,7	59,7	19,9	19,9	19,9	30,3	1,1	2,8	13,8	13,1
2003	16,5	10,4	19,9	19,9	59,7	59,7	19,9	19,9	19,9	30,3	1,1	2,4	13,8	13,1

The estimated MPs of capital are low for most observations; even zero values are common. Since the MP of one would be consistent with long-run profit maximization, a low MP suggests that the sector has over-invested in capital inputs. Only NZL and POR had MPs significantly greater than one, which indicates under-investment. Note the strong negative correlation between the MPs of labor and capital: the Pearson correlation coefficient between the two MP series is -0.42.

Table 4: Marginal product of capital (β_K), StoNED regression (25)

year	AUT	CAN	DEN	FIN	FRA	GBR	GER	ITA	NED	NOR	NZL	POR	SWE	USA
1975	0,10	0,37	0,10	0,10	0,00	0,07	0,41	0,37	0,34	0,10	0,10	0,42	0,78	0,40
1976	0,10	0,37	0,10	0,10	0,00	0,07	0,41	0,10	0,38	0,10	0,10	0,47	0,78	0,40
1977	0,10	0,30	0,00	0,10	0,00	0,07	0,41	0,10	0,37	0,10	0,10	0,47	0,78	0,37
1978	0,10	0,30	0,00	0,10	0,00	0,08	0,41	0,10	0,41	0,10	0,10	0,47	1,48	0,41
1979	0,22	0,30	0,00	0,10	0,00	0,08	0,41	0,10	0,41	0,10	0,33	0,51	1,48	0,41
1980	0,34	0,30	0,00	0,10	0,00	0,00	0,42	0,10	0,41	0,10	0,33	0,55	1,48	0,41
1981	0,34	0,30	0,00	0,10	0,00	0,00	0,42	0,10	0,41	0,10	0,78	0,55	1,48	0,41
1982	0,37	0,30	0,00	0,10	0,00	0,00	0,42	0,10	0,41	0,10	0,78	0,53	1,48	0,37
1983	0,37	0,30	0,00	0,10	0,00	0,00	0,42	0,10	0,41	0,08	0,78	0,57	1,48	0,21
1984	0,41	0,30	0,00	0,10	0,00	0,16	0,42	0,10	0,42	0,15	1,48	0,66	1,48	0,41
1985	0,41	0,30	0,00	0,10	0,00	0,16	0,61	0,10	0,61	0,15	1,48	0,66	0,45	0,41
1986	0,41	0,30	0,00	0,03	0,00	0,16	0,41	0,10	0,36	0,15	1,48	0,66	0,45	0,29
1987	0,41	0,30	0,00	0,00	0,00	0,20	0,42	0,10	0,36	0,15	1,48	0,57	0,45	0,32
1988	0,41	0,30	0,00	0,00	0,00	0,20	0,61	0,10	0,36	0,16	1,48	0,65	0,48	0,32
1989	0,42	0,33	0,00	0,00	0,00	0,20	0,61	0,08	0,36	0,16	1,48	0,97	0,48	0,24
1990	0,42	0,45	0,00	0,00	0,00	0,20	0,61	0,08	0,36	0,16	1,48	1,62	0,42	0,29
1991	0,41	0,45	0,00	0,00	0,20	0,20	0,25	0,15	0,36	0,20	1,48	2,70	0,42	0,24
1992	0,36	0,45	0,00	0,00	0,01	0,20	0,25	0,15	0,36	0,20	1,48	2,02	0,41	0,24
1993	0,36	0,45	0,00	0,00	0,00	0,20	0,25	0,17	0,36	0,20	1,48	0,58	0,41	0,24
1994	0,36	0,45	0,08	0,00	0,00	0,20	0,25	0,17	0,36	0,20	2,02	2,02	0,41	0,18
1995	0,36	0,45	0,16	0,00	0,00	0,20	0,20	0,25	0,36	0,20	2,02	0,51	0,41	0,36
1996	0,36	0,45	0,16	0,00	0,00	0,20	0,17	0,16	0,36	0,20	2,02	0,78	0,41	0,25
1997	0,36	0,45	0,20	0,00	0,00	0,20	0,20	0,20	0,36	0,20	2,02	0,78	0,41	0,16
1998	0,36	0,45	0,20	0,16	0,00	0,20	0,20	0,20	0,36	0,20	2,02	1,48	0,29	0,00
1999	0,36	0,45	0,20	0,16	0,00	0,20	0,20	0,20	0,36	0,31	2,02	0,78	0,29	0,00
2000	0,36	0,41	0,20	0,20	0,00	0,20	0,20	0,20	0,36	1,00	2,14	0,78	0,29	0,00
2001	1,05	0,41	0,20	0,20	0,00	0,00	0,20	0,20	0,25	0,90	2,02	0,78	0,30	0,00
2002	0,34	0,59	0,20	0,20	0,00	0,00	0,20	0,20	0,20	0,50	2,02	0,78	0,36	0,00
2003	0,34	0,61	0,20	0,20	0,00	0,00	0,20	0,20	0,20	0,50	2,02	0,51	0,36	0,00

The MPs of intermediate inputs have decreased over time in most sectors, USA being a notable exception. Again, the value of one would be consistent with the long-run profit maximization. In most countries, the under-utilization of intermediate inputs in the 1970s and 1980s has changed to over-use of intermediate inputs in the 1990s.

Table 5: Marginal product of intermediate inputs (β_M), StoNED regression (25)

year	AUT	CAN	DEN	FIN	FRA	GBR	GER	ITA	NED	NOR	NZL	POR	SWE	USA
1975	2,20	3,38	2,20	2,20	2,00	2,06	1,17	3,38	1,48	2,20	2,20	4,61	1,78	1,30
1976	2,20	3,38	2,20	2,20	1,81	2,06	1,17	2,20	1,37	2,20	2,18	4,32	1,78	1,30
1977	2,18	2,38	2,00	2,20	1,81	2,06	1,17	2,20	1,37	2,18	2,20	4,32	1,78	1,37
1978	2,18	2,38	2,00	2,20	1,81	2,00	1,17	2,20	1,17	2,18	2,20	4,32	0,92	1,17
1979	1,83	2,38	2,00	2,20	1,81	1,66	1,17	2,20	1,17	2,18	1,55	4,07	0,92	1,17
1980	1,48	2,38	2,00	2,20	1,81	1,83	1,14	2,20	1,17	2,18	1,60	3,93	0,92	1,17
1981	1,48	2,38	2,00	2,20	1,81	1,81	1,14	2,20	1,17	2,18	1,78	3,93	0,92	1,17
1982	1,37	2,38	2,00	2,20	1,81	1,81	1,14	2,20	1,17	2,18	1,78	3,95	0,92	1,37
1983	1,37	2,38	1,83	2,20	1,81	1,81	1,14	2,20	1,17	2,00	1,78	3,70	0,92	1,78
1984	1,17	2,38	1,81	2,20	1,81	1,03	1,14	2,20	1,14	1,37	0,92	3,59	0,92	1,17
1985	1,17	2,38	1,81	2,20	1,81	1,03	0,69	2,19	0,69	1,37	0,92	3,59	1,21	1,17
1986	1,17	2,38	1,81	2,03	1,81	1,03	1,17	2,18	0,81	1,37	0,92	3,59	1,21	1,22
1987	1,17	2,38	1,81	2,00	1,81	0,76	1,14	2,19	0,81	1,37	0,92	3,70	1,21	1,19
1988	1,17	2,38	1,81	2,00	1,81	0,74	0,69	2,18	0,81	1,36	0,92	3,40	1,06	1,19
1989	1,14	1,55	1,81	2,00	1,81	0,74	0,69	2,00	0,81	1,03	0,92	3,04	1,06	1,21
1990	1,14	1,21	1,81	2,00	1,81	0,74	0,69	1,70	0,81	1,03	0,92	2,03	1,14	1,10
1991	1,15	1,21	1,81	1,81	0,74	0,74	0,79	1,37	0,81	0,74	0,92	0,47	1,14	1,21
1992	0,81	1,21	1,81	1,81	1,55	0,74	0,79	1,37	0,81	0,74	0,92	0,73	1,17	1,21
1993	0,81	1,21	1,81	1,81	0,18	0,74	0,79	1,32	0,81	0,74	0,92	3,48	1,17	1,21
1994	0,81	1,21	1,44	1,81	1,61	0,74	0,79	1,32	0,81	0,74	0,73	0,73	1,17	1,04
1995	0,81	1,21	1,03	1,81	0,18	0,74	0,74	0,79	0,81	0,74	0,73	2,45	1,17	0,81
1996	0,81	1,21	1,03	1,81	0,18	0,74	0,94	1,03	0,81	0,74	0,73	1,78	1,17	0,79
1997	0,81	1,21	0,74	1,81	0,18	0,74	0,74	0,74	0,81	0,74	0,73	1,78	1,17	1,03
1998	0,81	1,21	0,74	1,03	0,18	0,74	0,74	0,74	0,81	0,74	0,73	0,92	1,10	1,81
1999	0,81	1,21	0,74	1,03	0,18	0,74	0,74	0,74	0,81	0,50	0,73	1,78	1,10	1,81
2000	0,81	1,17	0,74	0,74	0,18	0,74	0,74	0,74	0,81	0,00	0,00	1,78	1,11	1,81
2001	0,00	1,17	0,74	0,74	0,18	0,18	0,74	0,74	0,79	0,00	0,73	1,78	1,08	1,81
2002	0,69	0,82	0,74	0,74	0,18	0,18	0,74	0,74	0,74	0,00	0,73	1,78	0,81	1,81
2003	0,69	0,69	0,74	0,74	0,18	0,18	0,74	0,74	0,74	0,00	0,73	2,38	0,81	1,81

As a general observation, we see that the shadow prices reported in Tables 3-5 are concentrated to certain levels. This reflects the piece-wise linear nature of the StoNED frontier. The estimated frontier consists of 77 linear hyper-plane segments in total. Majority of these segments were small: only one observation was projected to 39 segments, and 20 segments contained two or three projections. On the other hand, there were five super-segments with more than 25 observations each. The largest hyper-plane segment contains 50 observations. More than 71 percent of the observations are projected to segments with at least ten other observations.

Table 5 presents the sector-specific coefficients of the quadratic time trend function; all coefficient are statistically highly significant. The constants a_i are negative for all sectors, whereas both b_i and c_i are positive for most sectors, signaling productivity growth over the time period. Only POR had

both b_i and c_i , which implies productivity decline. The time trend is graphically illustrated in Figure 6 for the five largest sectors. The diagram shows that the USA experienced the greatest output growth that can not explained by input use. All smaller sectors not depicted in the figure would lie somewhere between GER and CAN series.

Table 6: Time trend ($a_i + b_i t + c_i t^2$), StoNED regression (25)

	a	b	c
AUT	-3063,9	118,1	7,4
CAN	-8266,2	81,6	23,3
DEN	-4858,9	165,1	5,3
FIN	-4540,9	93,1	2,4
FRA	-9478,1	1938,3	30,7
GBR	-10583,7	-1474,4	169,4
GER	-7088,8	1761,1	-16,8
ITA	-24838,9	-248,6	79,0
NED	-3432,7	410,0	8,3
NOR	-1600,2	-190,4	18,2
NZL	-2467,6	-630,6	16,5
POR	-998,0	-77,7	-21,8
SWE	-4987,0	226,4	4,0
USA	-72073,4	10874,6	209,8

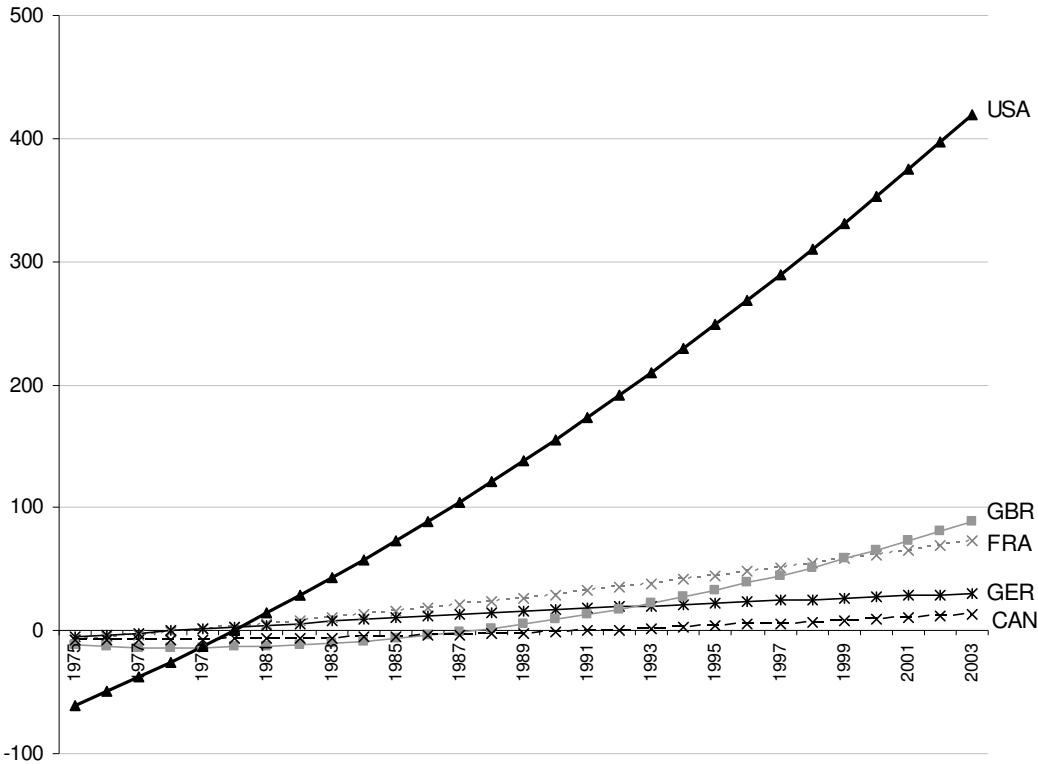


Figure 6: Time-trend functions $u(t)$ for five largest sectors.

Having discussed the non-parametric production frontier and the parametric trend function, it is worth to pay attention on the residuals. Figure 7 plots the residuals by sector over the time period (horizontal axis). The residuals appear to be randomly distributed, and no systematic sector specific or inter-temporal patterns are visible in the residual plot. In other words, no signs of heteroskedasticity or autocorrelation are present.

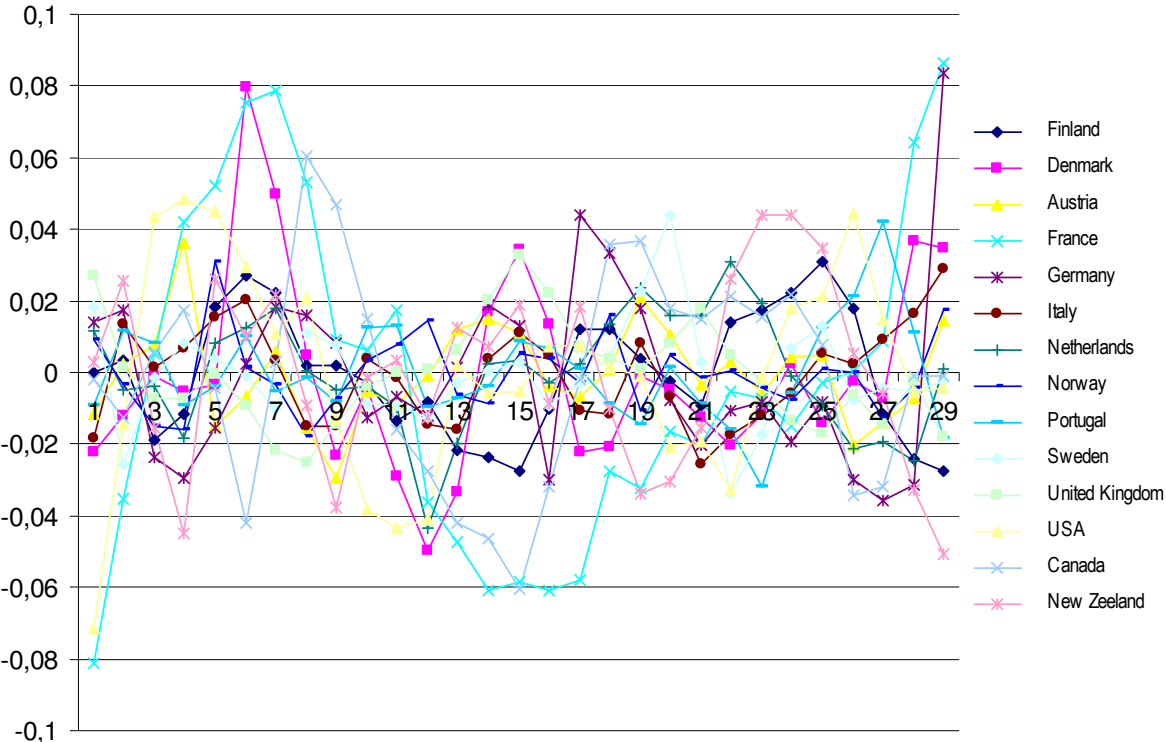


Figure 7: Residual plot.

6.4 Absolute vs. relative normalization

Returning back to Figure 6, we saw that the trend function for USA showed most rapid growth. This pattern is largely due to the dominant size of USA among other sectors. Resorting to the standard Gabrielsen-Greene standardization would use USA as a benchmark and compare the performance of other countries to the vast growth of USA in absolute terms without taking into account the difference in

the industry size. Consequently, all other countries appear utterly inefficient compared to USA. In fact, in 2003 all other countries yield larger inefficiency estimates than their total output. This implies that the output distance functions get negative values. Figure 8 illustrates the development of the output distance function values (inverse of the Farrell output efficiency measure) over the study period for the five largest sectors. This figure aptly illustrates the problem of the additive normalization in the present setting.

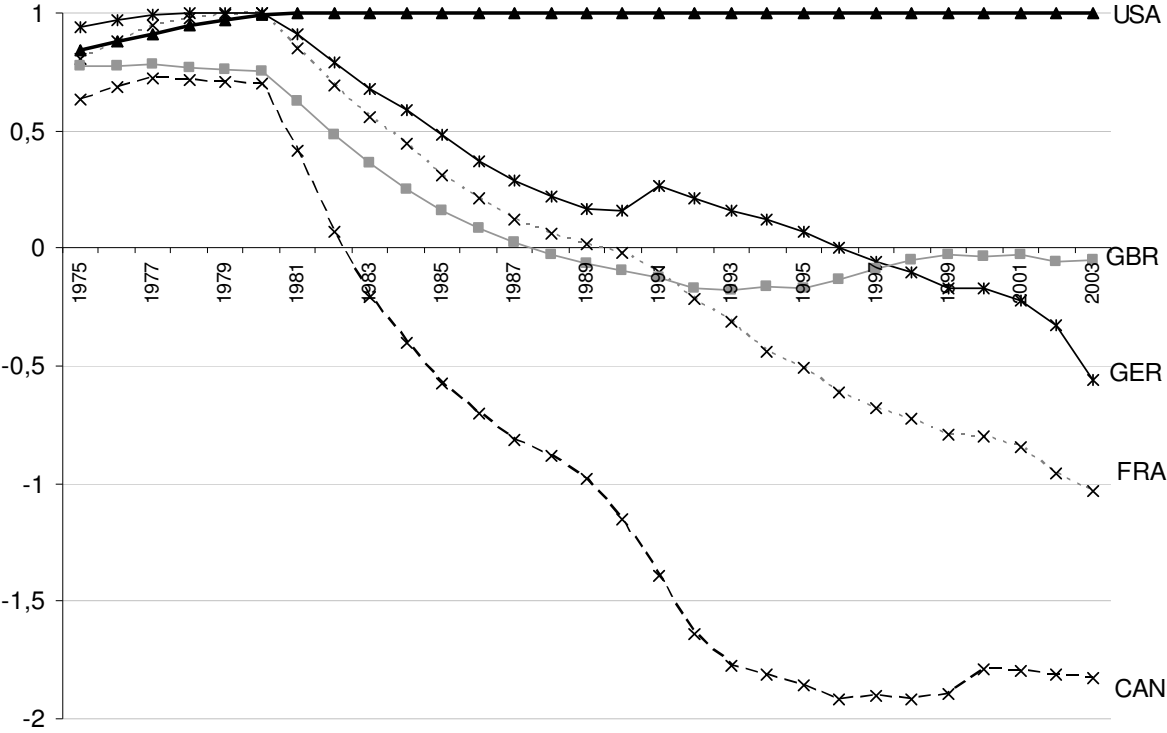


Figure 8: Relative output efficiency: additive Gabrielse-Greene normalization.

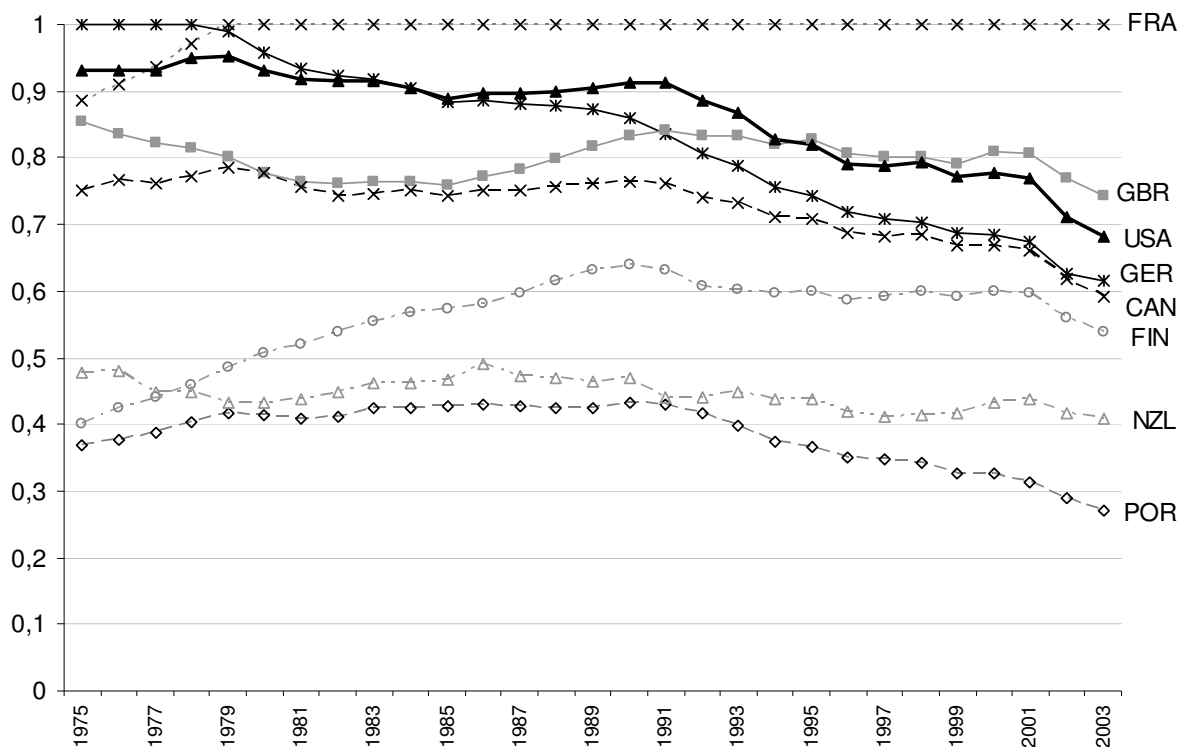


Figure 9: Relative output efficiency: direct multiplicative normalization.

More meaningful efficiency estimates are obtained by applying the multiplicative normalization presented in equations (12) and (13). Figure 9 describes the development of the output distance function values over the time period for the five largest and the three smallest sectors. Although USA showed the largest growth in absolute sense, the largest percentage growth occurs in FRA, which overtook GER as the most productive sector in 1979 and kept its leading position till the end of the study period. In fact, the relative normalization indicates that the growth of USA has been slow in relative terms. Regarding the small sectors, we note that FIN has shown considerable catching up while NZL and POR have fallen somewhat behind in relative efficiency terms.

6.5 Comparison with Parametric SFA

For comparison, we ran analogous fixed-effects SFA models with quadratic trend using the Cobb-Douglas and translog specifications for production function. The Cobb-Douglas model can be presented as the following least-squares problem:

$$\begin{aligned}
& \min_{\alpha, \beta, u, v} \sum_{t=1}^T \sum_{i=1}^n v_{it}^2 \\
& \text{s.t.} \\
& \ln y_{it} = [\alpha + \beta_L \ln L_{it} + \beta_K \ln K_{it} + \beta_M \ln M_{it}] + [a_i + b_i t + c_i t^2] + y_{it} v_{it}
\end{aligned} \tag{26}$$

In contrast to the StoNED model (25), the coefficients β are here the same for all observations. Further, the translog model can be written as

$$\begin{aligned}
& \min_{\alpha, \beta, u, v} \sum_{t=1}^T \sum_{i=1}^n v_{it}^2 \\
& \text{s.t.} \\
& \ln y_{it} = [\alpha + \beta_L \ln L_{it} + \beta_K \ln K_{it} + \beta_M \ln M_{it} + \gamma_{LL} (\ln L_{it})^2 + \gamma_{KK} (\ln K_{it})^2 + \gamma_{MM} (\ln M_{it})^2 \\
& + \gamma_{LK} \ln L_{it} \ln K_{it} + \gamma_{KM} \ln K_{it} \ln M_{it} + \gamma_{LM} \ln L_{it} \ln M_{it}] + [a_i + b_i t + c_i t^2] + y_{it} v_{it}
\end{aligned} \tag{27}$$

The Cobb-Douglas model is a special case of (27) where all second-order coefficients γ are set equal to zero.

Table 7 presents the results of the two SFA models. For comparison, we also include the average values of the StoNED model. We note that the translog model achieved the highest R² and log-likelihood values. In theory, the CNLS regression yields a better fit than any non-linear OLS regression; the higher values of the translog regression are here due to the fact that we impose both concavity and CRS restrictions in the StoNED model while the SFA model is unrestricted in these respects.

It is worth to emphasize that the coefficients of the trend function are country-specific, as are all coefficients of the StoNED model. For clarity, the average values and their standard deviations are typeset in *italics*. These should not be confused with the estimated regression coefficients and standard errors typeset in normal font.

Table 7: Comparison of the regression results

	StoNED	SFA Cobb-Douglass	SFA translog
R ²	0.9994	0.9992	0.9996
ln L	-2772.51	-2837.97	-2670.51
Variable	Average (St.Dev)	Coefficient (st. err.)	Coefficient (st. err.)
β_L	0.0123 (0.0105)	0.4748 (0.0121)	0.9972 (0.0099)
β_K	0.4487 (0.4619)	0.0512 (0.0094)	-0.4596 (0.0078)
β_M	1.4317 (0.7786)	0.5532 (0.0095)	1.6840 (0.0078)
γ_{LL}			-0.0806 (0.000)
γ_{KK}			0.0476 (0.000)
γ_{MM}			0.0415 (0.000)
γ_{LK}			0.1632 (0.000)
γ_{KM}			-0.2326 (0.000)
γ_{LM}			-0.0461 (0.000)
Variable	Average (St.Dev)	Average (St.Dev)	Average (St.Dev)
a	-11305.5943 (18485.8823)	0.0892 (0.0309)	0.0001 (0.0000)
b	931.8968 (2985.9747)	1.0445 (0.0356)	1.0361 (0.0340)
c	38.2488 (68.7749)	0.9992 (0.0007)	0.9994 (0.0007)

Figures 10 and 11 describe the development of the efficiency scores for the five largest and three smallest sectors according to the SFA models, using the multiplicative normalization (13). All models agree that FRA was the benchmark country, but for other countries there are considerable differences in the results. Consider USA (the thickest line), for example. The Cobb-Douglas model suggests at least 10 percent points lower efficiency level for USA compared to the StoNED model, and also the shape of the trend is different. By contrast, the translog model suggests that USA reached the world frontier in years 1980 and 2003 with a temporary decline between those years. The good performance of USA in the translog model is explained by the fact that the flexible production function allows for locally varying returns to scale, and allowing for decreasing returns to scale makes USA look

more efficient in the translog model. By contrast, the Cobb-Douglas regression exhibited globally increasing returns to scale (scale elasticity 1.08) which explains the lower efficiency of USA. In conclusion, the efficiency estimates proved rather sensitive to the specification of the functional form of the production function even though the differences in the empirical fit were negligible.

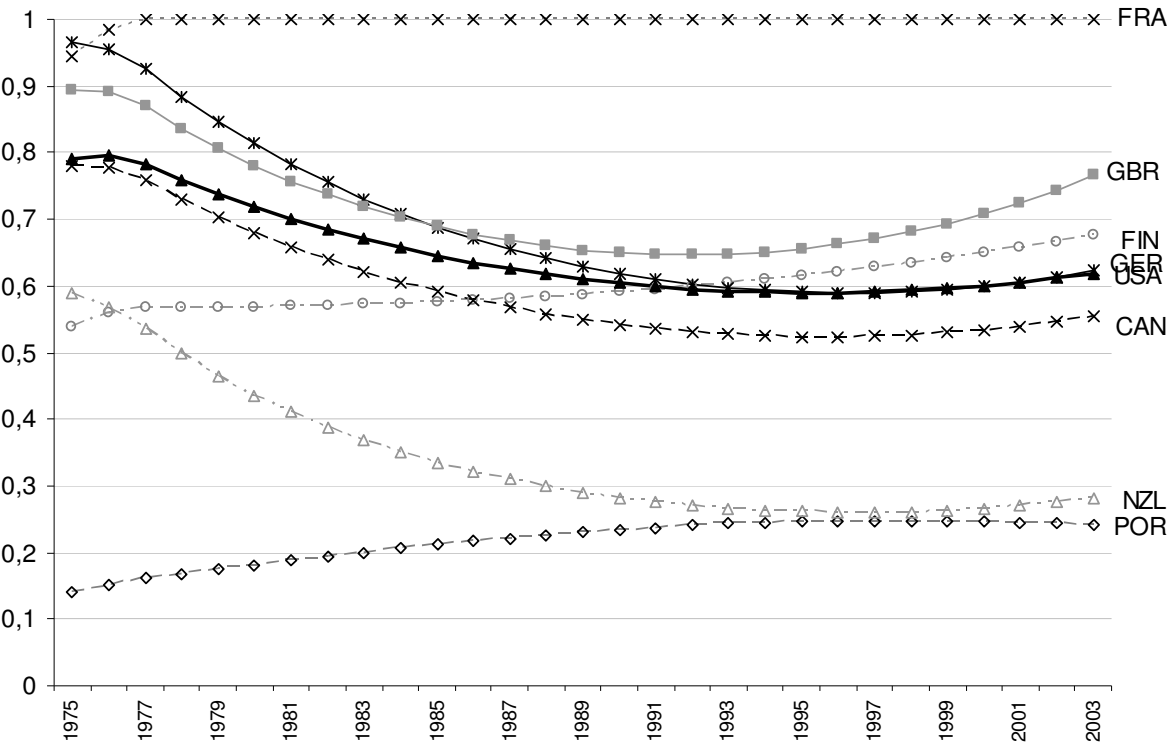


Figure 10: Relative output efficiency – the SFA model with Cobb-Douglas production function.

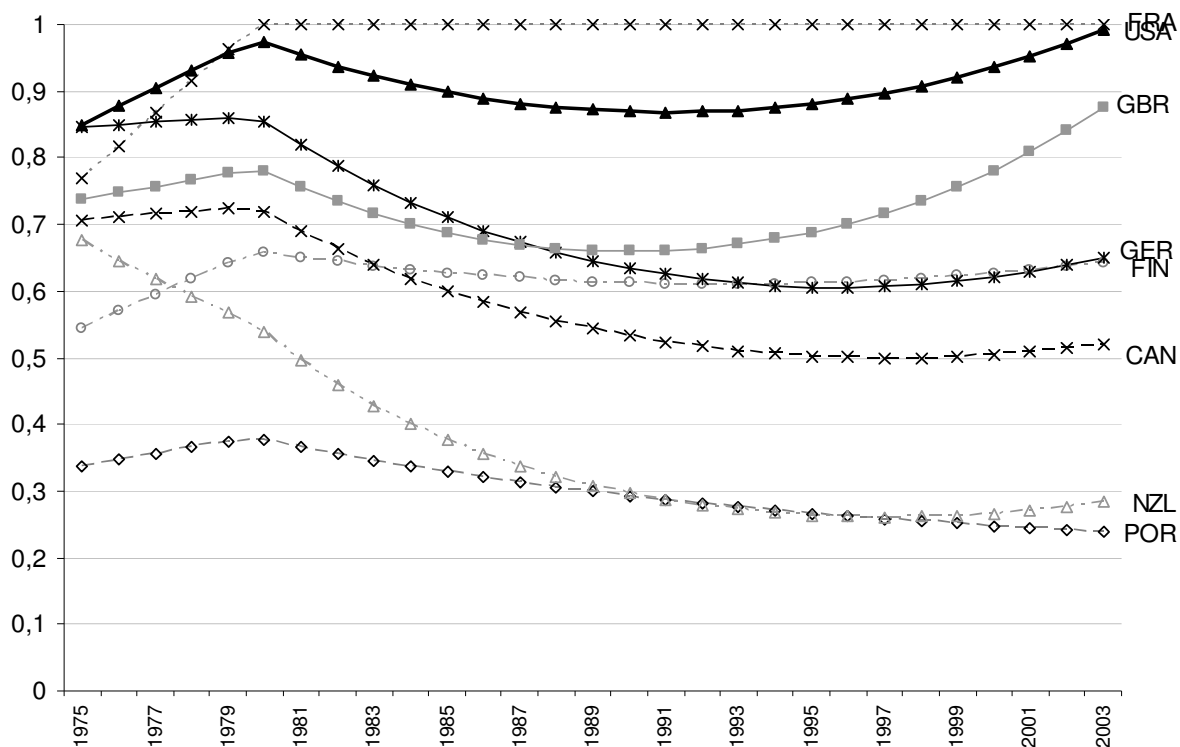


Figure 11: Relative output efficiency – the SFA model with translog production function.

6.6 Malmquist productivity index

We next used the output distance function estimates from the StoNED model to calculate the Malmquist productivity indices and decompose them to technical change and efficiency change components. Figure 12 illustrates the cumulative Malmquist indices for the five largest sectors, together with ITA and FIN which showed the greatest productivity growth during the period. Scandinavian sectors SWE, DEN and NOR not depicted in Figure 12 followed closely the productivity development of FRA. All other sectors followed the paths of GBR and USA. GER exhibited notably the lowest productivity growth especially since 1987.

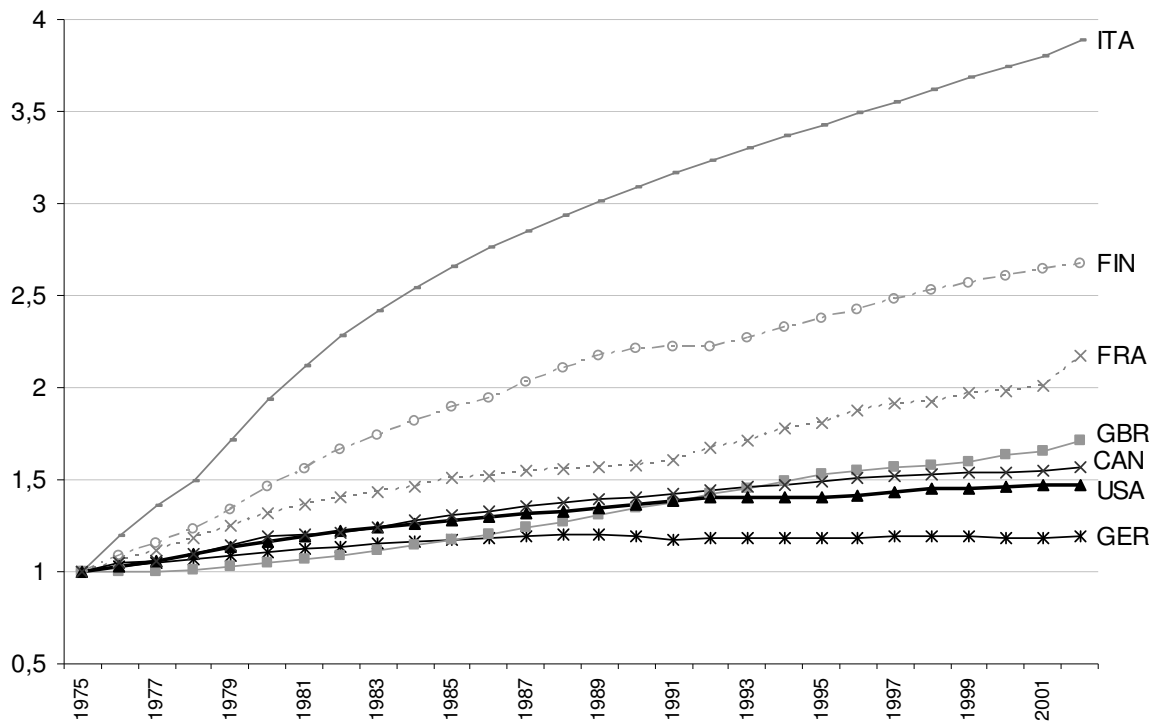


Figure 12: Malmquist productivity indices (cumulative, 1975=1)

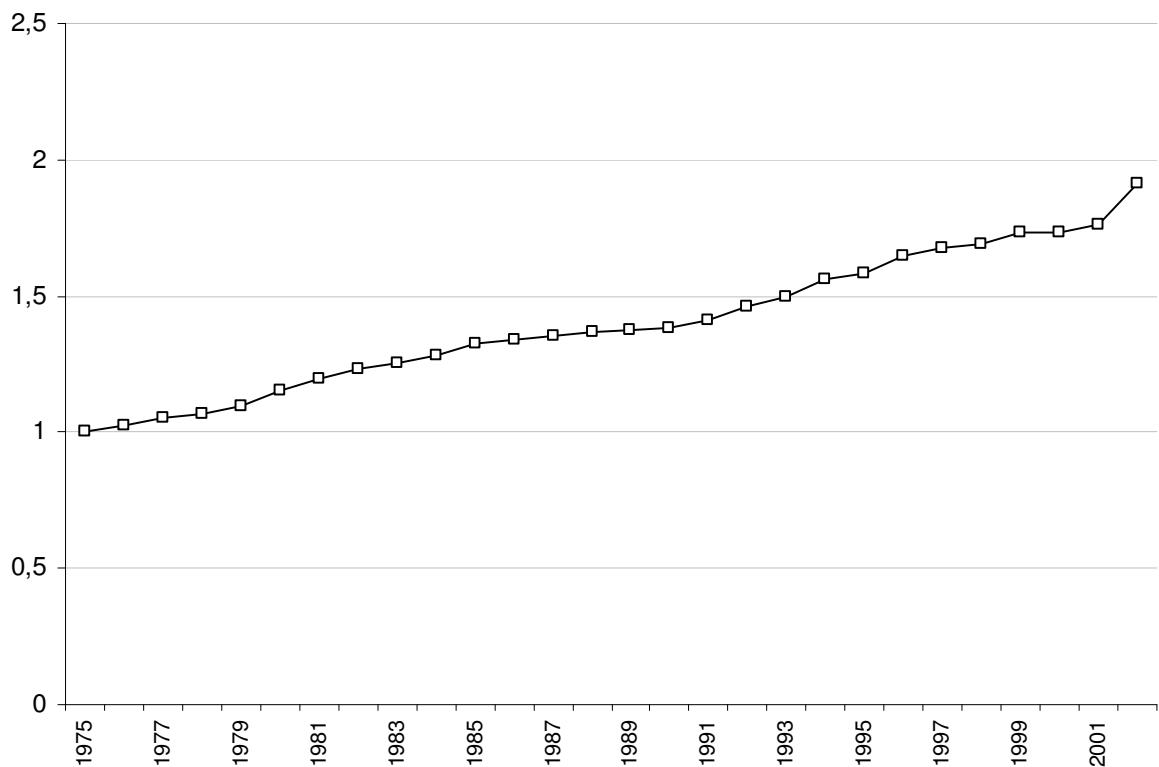


Figure 13: Technical change component of the Malmquist index (cumulative, 1975=1)

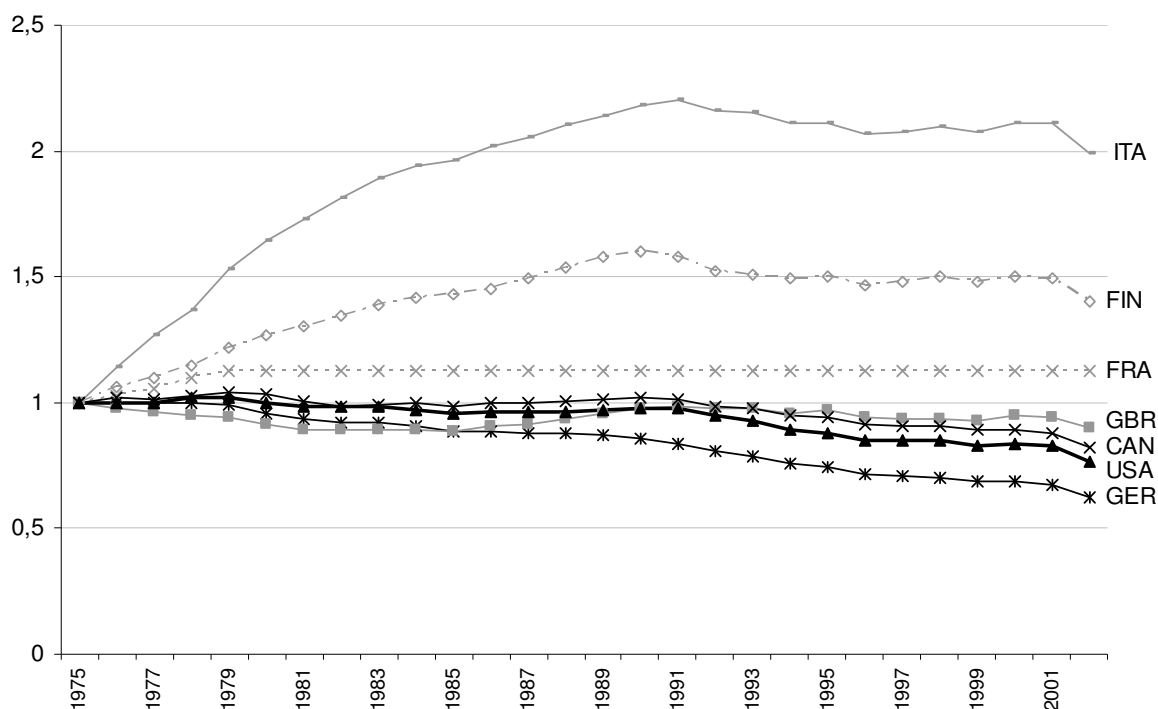


Figure 14: Efficiency change component of the Malmquist index (cumulative, 1975=1)

The technical change and efficiency components of the Malmquist index are depicted in Figures 13 and 14 for these seven countries. The technical change component of Figure 13 is the same for all sectors due to the Hicks neutral specification of technical progress. Comparing Figures 13 and 14 we see that the overall technical change follows very closely the productivity development of FRA, which was the frontier shifting sector throughout 1979-2003. From Figure 14 we see that the rapid productivity growth of ITA and FIN has been largely due to efficiency increase, in other words catching up. Also in Figure 14, the efficiency change of FRA also approximates that of Scandinavian sectors SWE, DEN, and NOR, while all other sectors lie somewhere between GBR and USA.

Comparing the results of the StoNED model with the recent report by the Nordic Competition Authorities (2005) reveals some interesting insights. It is not surprising that FRA distinguishes itself as a frontier shifting sector: France has a modern hypermarket sector and today more than half of the retail sales are done in markets with floor area of 10,000 m² or greater. Also the trading hours have become very flexible in France, the main limitation being the maximum number of hours that employees are

permitted to work per week. Development of large-scale super and hypermarkets has also been rapid in Italy and Finland, which explains their catching up. By contrast, in Germany the trading hours and the construction of large-scale retail schemes are more restrictive than anywhere in else in Europe. Consequently, Germany has relatively few large hypermarkets and the small discount shops and hard discounters have currently almost 40 percent market share. The United Kingdom had an early start in the development of shopping centers with hypermarkets, but since the 1980s the building of large out-of-town shopping centers has been restricted in favor of town center development. The USA and Canada have large-scale shopping malls and hypermarkets, and the trading hours are very liberal. However, the relatively low wage rates and the competition with opening hours allow the North American retail markets to use more inputs per sale than their European counterparts, which shows in the productivity and efficiency comparisons.

7. Concluding discussion

We have shown how a stochastic noise term can be introduced to the nonparametric frontier estimation, to combine the key merits of SFA and DEA in a uniform framework. The new approach was named *Stochastic Nonparametric Envelopment of Data* (StoNED). We started from the estimation of a single-output production function in the cross-sectional setting, and then extended the framework towards panel data settings, and allowed for multiple outputs by using distance or cost functions. We also showed how alternative assumptions about returns to scale or multiplicative errors can be implemented in the StoNED framework, and discussed the semi-parametric treatment of environmental factors.

The approach was illustrated by means of a simulated numerical example and an industry-level application to wholesale and retail trade sector in 14 OECD countries. The simulation showed the potential of the StoNED approach, and the application demonstrated its usefulness in the panel data settings where the standard DEA does not have sufficient discriminatory power. The application also demonstrated the sensitivity of the results to the parametric specification of the functional form even

when the empirical fit is very high. In this respect, the StoNED approach uses a nonparametric function that has its foundation in optimization. The application also demonstrated how one can apply the StoNED method for estimating and decomposing total factor productivity growth using the Malmquist productivity indices.

The proposed StoNED approach shares many similarities with SFA and DEA, being a genuine hybrid of the two. Thus, many existing tools and techniques can be incorporated into the StoNED framework. However, the hybrid nature of StoNED also means that there are many important differences to both SFA and DEA which must be kept in mind in application and interpretation of the StoNED models. For example, the interpretation of the StoNED input coefficients differs considerably from those of the SFA coefficients. Moreover, in contrast to DEA, all observations influence the shape of the frontier. In this respect, further research is needed for a better understanding of these similarities and differences. This paper has briefly noted a number of possibilities to apply the StoNED approach: each of them would merit more detailed discussion than the short sketches provided here. We hope that this paper can provide inspiration for further work and thus contribute the unification of the productive efficiency analysis and the integration and cross-fertilization of the parametric and nonparametric streams of the literature.

While the StoNED approach combines the best features of DEA and SFA, it also shares some of their limitations. Similar to DEA, the nonparametric approach makes also StoNED vulnerable to the curse of dimensionality, which means that the sample size must be very large when the number of input variables is high. On the other hand, the maintained SFA assumptions regarding the error distribution (e.g., normality, independence) may be violated. Moreover, stochastic noise does not restrict to the output, but also input data may be perturbed by measurement errors and other noise. Treatment of the noise in input data remains somewhat problematic in the SFA framework, and hence also in the StoNED approach. Despite these shared limitations, the benefits of the unified framework clearly outweigh the costs.

References

- Afriat, S.N. (1967): The Construction of a Utility Function from Expenditure Data, *International Economic Review* 8, 67-77.
- Afriat, S. (1972): Efficiency Estimation of Production Functions, *International Economic Review* 13, 568-598.
- Aigner, D.J., and S. Chu (1968): On Estimating the Industry Production Function, *American Economic Review* 58, 826-839.
- Aigner, D.J., C.A.K. Lovell, and P. Schmidt (1977): Formulation and Estimation of Stochastic Frontier Models, *Journal of Econometrics* 6, 21-37.
- Asmild, M., J.C. Paradi, V. Aggarwall, and C. Schaffnit (2004): Combining DEA Window Analysis with the Malmquist Index Approach in a Study of the Canadian Banking Industry, *Journal of Productivity Analysis* 21(1), 67-89.
- Aragon, Y., A. Daouia, and C. Thomas-Agnan (2002): *Nonparametric Frontier Estimation: A Conditional Quantile-based Approach*, Discussion paper, GREMAQ et LSP, Université de Toulouse.
- Bauer, P.W. (1990): Recent Developments in the Econometric Estimation of Frontiers, *Journal of Econometrics* 46 (1-2), 39-56.
- Cazals, C., J.P. Florens, and L. Simar (2002): Nonparametric Frontier Estimation: A Robust Approach, *Journal of Econometrics* 106, 1-25.
- Chambers, R.G., Y. Chung and R. Färe (1996): Benefit and Distance Functions, *Journal of Economic Theory* 70(2), 407-419.
- Chambers, R.G., Y. Chung and R. Färe (1998): Profit, Directional Distance Functions, and Nerlovian Efficiency, *Journal of Optimization Theory and Applications* 98(2), 351-364.
- Cherchye, L., and T. Post (2003): Methodological Advances in DEA: A Survey and an Application for the Dutch Electricity Sector, *Statistica Neerlandica* 57(4), 410-438.
- 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.

- Cornwell, C., P. Schmidt, and R. Sickles (1990): Production Frontiers with Cross Sectional and Time Series Variation in Efficiency Levels, *Journal of Econometrics* 46, 185-200.
- 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.
- Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994): Productivity Growth, Technical Progress and Efficiency Change in Industrialized Countries. *American Economic Review* 84(1): 66-83.
- Farrell, M.J. (1957): The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society Series A. General* 120(3): 253-282.
- Gabrielsen, A. (1975): On Estimating Efficient Production Functions, Working Paper A-85, Dept. of Humanities and Social Sciences, Chr. Michelsen Institute, Bergen.
- Greene, W.H. (1980): Maximum Likelihood Estimation of Econometric Frontier Functions, *Journal of Econometrics* 13, 27-56.
- Greene, W.H. (1999): Frontier Production Functions, Ch. 3 (pp. 81–166) in: M. Pesaran and P. Schmidt (Eds.): *Handbook of applied econometrics, Vol II: Microeconomics*, Blackwell Publishers, Oxford.
- Greene, W.H. (2005): Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model, *Journal of Econometrics* 126, 269-303.
- 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.

- 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 (2004): Nonparametric Stochastic Frontiers: A Local Maximum Likelihood Approach, *Journal of Econometrics*, to appear.
- Kuosmanen, T., (2003): Duality Theory of Non-convex Technologies, *Journal of Productivity Analysis* 20, 273–304.
- Kuosmanen, T., (2006): *Convex Nonparametric Least Squares: Multiple Regression, Interpolation and Statistical Inference*, unpublished manuscript, available from the author by request.
- Kuosmanen, T., L. Cherchye, and T. Sipiläinen (2006): The Law of One Price in Data Envelopment Analysis: Restricting Weight Flexibility across Firms, *European Journal of Operational Research* 170(3), 735-757.
- Meeusen, W., and J. van den Broeck (1977): Efficiency Estimation from Cobb-Douglas Production Function with Composed Error, *International Economic Review* 8, 435-444.
- Murillo-Zamorano, L.R. (2004): Economic Efficiency and Frontier Techniques, *Journal of Economic Surveys* 18(1), 33-77.
- Nordic Competition Authorities (2005): *Nordic Food Markets – A Taste for Competition*, report from the Nordic Competition Authorities No 1/2005.
- Ondrich, J., and J. Ruggiero (2001): Efficiency Measurement in the Stochastic Frontier Model, *European Journal of Operational Research* 129(2), 434-442.
- 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., and L. Simar (1994): Efficient Semiparametric Estimation in a Stochastic Frontier Model, *Journal of the American Statistical Association* 89(427), 929-936.

- Ruggiero, J. (2004): Data Envelopment Analysis with Stochastic Data, *Journal of the Operational Research Society* 55(9), 1008-1012.
- 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. Wilson (2006): Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes, *Journal of Econometrics* (in press).
- 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, H. (1985): Nonparametric Analysis of Optimizing Behavior with Measurement Error, *Journal of Econometrics* 30(1/2), 445-458.
- Wang, D.Q., S. Chukova, and C.D. Lai (2004): On the Relationship between Regression Analysis and Mathematical Programming, *Journal of Applied Mathematics and Decision Sciences* 8(2), 131–140.
- Wang, H.-J., and P. Schmidt (2002): One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels, *Journal of Productivity Analysis* 18, 129-144.
- Worthington, A.C. (2001): An Empirical Survey of Frontier Efficiency Measurement Techniques in Education, *Education Economics* 9(3), 245-268
- Worthington, A.C. (2004): An Empirical Survey of Frontier Efficiency Measurement Techniques in Healthcare Services, *Medical Care Research and Review* 61(2), 1-36.

Appendix 1: DEA as least-squares regression

The standard output-oriented variable returns to scale (VRS) DEA envelopment model can be written as

$$\begin{aligned}
 & \max_{\theta, \lambda} \theta \\
 & \text{s.t.} \\
 & y_0 \theta \leq \sum_{i=1}^n y_i \lambda_i \\
 & \mathbf{x}_i \geq \sum_{i=1}^n \mathbf{x}_i \lambda_i \\
 & \sum_{i=1}^n \lambda_i = 1 \\
 & \lambda \geq \mathbf{0}
 \end{aligned} \tag{A1.1}$$

In the single-output case, we can harmlessly write the problem in an additive form as

$$\begin{aligned}
 & \max_{\phi, \lambda} \phi \\
 & \text{s.t.} \\
 & y_0 + \phi \leq \sum_{i=1}^n y_i \lambda_i \\
 & \mathbf{x}_i \geq \sum_{i=1}^n \mathbf{x}_i \lambda_i \\
 & \sum_{i=1}^n \lambda_i = 1 \\
 & \lambda \geq \mathbf{0}
 \end{aligned} \tag{A1.2}$$

Note that the multiplicative efficiency score can be recovered as $\theta^* = 1 + \phi^* / y_0$.

The dual problem of (A1.2) can be written as

$$\begin{aligned}
 & \min_{\beta, \alpha, \gamma} (\beta' \mathbf{x}_0 + \alpha) - \gamma y_0 \\
 & \text{s.t.} \\
 & \gamma = 1 \\
 & \beta' \mathbf{x}_i + \alpha - \gamma y_i \geq 0 \quad \forall i = 1, \dots, n \\
 & \beta \geq \mathbf{0}
 \end{aligned} \tag{A1.3}$$

where vector β represents the multiplier weights for inputs, and α is the shadow price of the VRS constraint. Note that the additive formulation of the efficiency score implies that the output y_0 appears in the objective function and not in the first constraint (as it would be in the dual of (A1.1)). Thus, we can remove the redundant multiplier γ and write the dual problem as

$$\begin{aligned}
& \min_{\beta, \alpha} (\alpha + \beta' \mathbf{x}_0) - y_0 \\
& \text{s.t.} \\
& y_i \leq \alpha + \beta' \mathbf{x}_i \quad \forall i = 1, \dots, n \\
& \beta \geq \mathbf{0}
\end{aligned} \tag{A1.4}$$

Next, we introduce $\tilde{\varepsilon}_0 = \alpha + \beta' \mathbf{x}_0 - y_0$ as an auxiliary variable for the objective function, and write the problem as

$$\begin{aligned}
& \min_{\beta, \alpha} \tilde{\varepsilon}_0 \\
& \text{s.t.} \\
& y_0 = \alpha + \beta' \mathbf{x}_0 - \tilde{\varepsilon}_0 \\
& y_i \leq \alpha + \beta' \mathbf{x}_i \quad \forall i = 1, \dots, n \\
& \beta \geq \mathbf{0}
\end{aligned} \tag{A1.5}$$

Instead of solving problem (A1.5) for each firm separately, we can solve the efficiency scores simultaneously for all firms. As in Kuosmanen et al. (2006), we can minimize the sum of the efficiency scores as

$$\begin{aligned}
& \min_{\beta, \alpha} \sum_{i=1}^n \tilde{\varepsilon}_i \\
& \text{s.t.} \\
& y_i = \alpha_i + \beta_i' \mathbf{x}_i - \tilde{\varepsilon}_i \quad \forall i = 1, \dots, n \\
& y_j \leq \alpha_j + \beta_j' \mathbf{x}_j \quad \forall i, j = 1, \dots, n \\
& \beta_i \geq \mathbf{0} \quad \forall i = 1, \dots, n
\end{aligned} \tag{A1.6}$$

where (α_i, β_i) are the multiplier weights used for evaluating firm i . The optimal $\tilde{\varepsilon}_i^*$ from (A1.6) gives the efficiency score of firm i . Since the efficiency score is invariant to monotonic transformation, we can equally well minimize the sum of squares as

$$\begin{aligned}
& \min_{\beta, \alpha} \sum_{i=1}^n \tilde{\varepsilon}_i^2 \\
& \text{s.t.} \\
& y_i = \alpha_i + \beta_i' \mathbf{x}_i - \tilde{\varepsilon}_i \quad \forall i = 1, \dots, n \\
& y_j \leq \alpha_j + \beta_j' \mathbf{x}_j \quad \forall i, j = 1, \dots, n \\
& \beta_i \geq \mathbf{0} \quad \forall i = 1, \dots, n
\end{aligned} \tag{A1.7}$$

Since $\tilde{\varepsilon}_i^2 = (-\tilde{\varepsilon}_i)^2$, we can substitute $\tilde{\varepsilon}_i$ by $-\hat{\varepsilon}_i$ and write the least squares problem as

$$\begin{aligned}
& \min_{\beta, \alpha} \sum_{i=1}^n \hat{\varepsilon}_i^2 \\
& \text{s.t.} \\
& y_i = \alpha_i + \beta'_i \mathbf{x}_i + \hat{\varepsilon}_i \quad \forall i = 1, \dots, n \\
& y_j \leq \alpha_i + \beta'_i \mathbf{x}_j \quad \forall i, j = 1, \dots, n \\
& \beta_i \geq \mathbf{0} \quad \forall i = 1, \dots, n \\
& \hat{\varepsilon} \leq 0
\end{aligned} \tag{A1.8}$$

Finally, since the inefficient firms (for which $\hat{\varepsilon}_i < 0$) do not influence the shape of the DEA frontier, we can harmlessly add the inefficiency component to the constraints of the reference firms (the second constraint) as

$$\begin{aligned}
& \min_{\alpha, \beta, \hat{\varepsilon}} \sum_{i=1}^n \hat{\varepsilon}_i^2 \\
& \text{s.t.} \\
& y_i = \alpha_i + \beta'_i \mathbf{x}_i + \hat{\varepsilon}_i \quad \forall i = 1, \dots, n \\
& y_h \leq \alpha_i + \beta'_i \mathbf{x}_h + \hat{\varepsilon}_h \quad \forall h, i = 1, \dots, n \\
& \beta'_i \geq \mathbf{0} \quad \forall i = 1, \dots, n \\
& \hat{\varepsilon} \leq 0
\end{aligned} \tag{A1.9}$$

This gives the CNLS problem (3) augmented by the non-positivity constraint $\hat{\varepsilon} \leq 0$ for the residuals.

Similar to the deterministic COLS model by Aigner and Chu (1968) that inserts constraint $\hat{\varepsilon} \leq 0$ for the residuals of the OLS regression, we have here shown that DEA model can be viewed as a “corrected CNLS” regression.

Appendix 2: Derivation of the directional distance function StoNED model of Section 5.2

Recall from (15) that the dual expression of the directional distance function can be written as

$$DD(\mathbf{x}, \mathbf{y}; \mathbf{g}^x, \mathbf{g}^y) = - \inf_{\substack{(\mathbf{x}^*, \mathbf{y}^*) \\ (\boldsymbol{\omega}, \boldsymbol{\rho}) \geq 0}} \{ (\boldsymbol{\rho}'\mathbf{y} - \boldsymbol{\omega}'\mathbf{x}) - (\boldsymbol{\rho}'\mathbf{y}^* - \boldsymbol{\omega}'\mathbf{x}^*) \mid (\mathbf{x}^*, \mathbf{y}^*) \in T; \boldsymbol{\rho}'\mathbf{g}^x + \boldsymbol{\omega}'\mathbf{g}^y = 1 \}. \quad (\text{A2.1})$$

Provided that the direction vector $(\mathbf{g}^x, \mathbf{g}^y)$ is the same for all firms, the DEA estimates for the directional distance functions can be obtained from the optimal solutions to the following linear programming problem

$$\begin{aligned} \min_{\boldsymbol{\rho}, \boldsymbol{\omega}, \boldsymbol{\varepsilon}} \quad & \sum_{i=1}^N \delta_i \\ \text{s.t.} \quad & \\ & \delta_i \geq (\boldsymbol{\rho}'_i \mathbf{y}_h - \boldsymbol{\omega}'_i \mathbf{x}_h) - (\boldsymbol{\rho}'_i \mathbf{y}_i - \boldsymbol{\omega}'_i \mathbf{x}_i) \quad \forall h, i = 1, \dots, n \\ & \boldsymbol{\rho}'_i \mathbf{g}_y + \boldsymbol{\omega}'_i \mathbf{g}_x = 1 \quad \forall i = 1, \dots, n \\ & (\boldsymbol{\rho}'_i, \boldsymbol{\omega}'_i) \geq \mathbf{0} \quad \forall i = 1, \dots, n \end{aligned} \quad (\text{A2.2})$$

as $\widehat{DD}(\mathbf{x}_i, \mathbf{y}_i; \mathbf{g}^x, \mathbf{g}^y) = \delta_i^*$. Now, adopting a stochastic specification of the distance measure δ_i , we will minimize the square sum of δ_i distances. We also need to adjust the “shadow profit” of the reference firm h by adding its distance to the frontier δ_h as $\boldsymbol{\rho}'_i \mathbf{y}_h - \boldsymbol{\omega}'_i \mathbf{x}_h + \delta_h$. This provides the least-squares problem:

$$\begin{aligned} \min_{\boldsymbol{\rho}, \boldsymbol{\omega}, \boldsymbol{\varepsilon}} \quad & \sum_{i=1}^N \delta_i^2 \\ \text{s.t.} \quad & \\ & \delta_i \geq (\boldsymbol{\rho}'_i \mathbf{y}_h - \boldsymbol{\omega}'_i \mathbf{x}_h + \delta_h) - (\boldsymbol{\rho}'_i \mathbf{y}_i - \boldsymbol{\omega}'_i \mathbf{x}_i) \quad \forall h, i = 1, \dots, n \\ & \boldsymbol{\rho}'_i \mathbf{g}_y + \boldsymbol{\omega}'_i \mathbf{g}_x = 1 \quad \forall i = 1, \dots, n \\ & (\boldsymbol{\rho}'_i, \boldsymbol{\omega}'_i) \geq \mathbf{0} \quad \forall i = 1, \dots, n \end{aligned} \quad (\text{A2.3})$$

Reorganizing the first constraint gives

$$\begin{aligned} \min_{\boldsymbol{\rho}, \boldsymbol{\omega}, \boldsymbol{\varepsilon}} \quad & \sum_{i=1}^N \delta_i^2 \\ \text{s.t.} \quad & \\ & \delta_i - \delta_h \geq (\boldsymbol{\rho}'_i \mathbf{y}_h - \boldsymbol{\omega}'_i \mathbf{x}_h) - (\boldsymbol{\rho}'_i \mathbf{y}_i - \boldsymbol{\omega}'_i \mathbf{x}_i) \quad \forall h, i = 1, \dots, n \\ & \boldsymbol{\rho}'_i \mathbf{g}_y + \boldsymbol{\omega}'_i \mathbf{g}_x = 1 \quad \forall i = 1, \dots, n \\ & (\boldsymbol{\rho}'_i, \boldsymbol{\omega}'_i) \geq \mathbf{0} \quad \forall i = 1, \dots, n \end{aligned} \quad (\text{A2.4})$$

**Appendix 3: The GAMS code for the panel data StoNED model used in Section 6 (14 industries,
29 time periods, 3 inputs, 1 output)**

SETS

i observations /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) trend
Tr2(t) trend-squared;

VARIABLES

E(i,t) error term
SS sum of square of errors
A(i) trend coefficient
B(i) trend coefficient
C(i) trend coefficient;

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
QREGR(i,t) regression equation
QCONC (i,h,t,s) concavity constraint;

QSSE.. SS =e= sum(t, sum(i, E(i,t)*E(i,t)));
QREGR (i,t).. Y(i,t) =e= BL(i,t)*L(i,t) + BK(i,t)*K(i,t) + BM(i,t)*M(i,t) - (A(i) + B(i)*Tr(t) + C(i)*Tr2(t))
+ Y(i,t)*E(i,t);
QCONC(i,h,t,s).. Y(h,s) =l= BL(i,t)*L(h,s) + BK(i,t)*K(h,s) + BM(i,t)*M(h,s) - (A(h) + B(h)*Tr(s)
+ C(h)*Tr2(s)) + Y(h,s)*E(h,s);

MODEL StoNED /all/

SOLVE StoNED using NLP Minimizing SS;

