Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP08/2018

Econometric Analysis of Productivity: Theory and Implementation in R

Robin C. Sickles, Wonho Song and Valentin Zelenyuk

Date: September 2018

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Forthcoming as Chapter in the 41 Handbook of Statistics co-edited by H D Vinod and C R Rao

Econometric Analysis of Productivity: Theory and Implementation in R

Robin C. Sickles * Wonho Song † Valentin Zelenyuk ‡§

September 3, 2018

Abstract

Our chapter details a wide variety of approaches used in estimating productivity and efficiency based on methods developed to estimate frontier production using Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA). The estimators utilize panel, single cross section, and time series data sets. The R programs include such approaches to estimate firm efficiency as the time invariant fixed effects, correlated random effects, and uncorrelated random effects panel stochastic frontier estimators, time varying fixed effects, correlated random effects, and uncorrelated random effects estimators, semi-parametric efficient panel frontier estimators, factor models for cross-sectional and time-varying efficiency, bootstrapping methods to develop confidence intervals for index number-based productivity estimates and their decompositions, DEA and Free Disposable Hull estimators. The chapter provides the professional researcher, analyst,

*Corresponding author. Professor of Economics, Department of Economics, Rice University, 6100 Main Street, Houston, Texas 77005, USA; phone: +1 713-348-3322; email: rsickles@rice.edu

†Professor of Economics, School of Economics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea, Phone: +82-2-820-5493, email: whsong@cau.ac.kr

‡Professor of Economics, School of Economics, University of Queensland, 530 Colin Clark Building (39), St Lucia, Brisbane, Qld 4072, Australia; Phone:+ 61 7 3346 7054 e-mail:v.zelenyuk@uq.edu.au

§The authors would like to thank Kerda Varaku for her assistance in preparing our manuscript. All remaining errors are the responsibility of the authors.

statistician, and regulator with the most up to date efficiency modeling methods in the easily accessible open source programming language R.

Keywords: Production (technical) efficiency; Stochastic frontier analysis; Data envelopment analysis; Panel data; Index numbers; Nonparametric analysis; Bootstrapping

1 Introduction

Our chapter provides a discussion of various statistical and mathematical procedures that firms, regulators, and academics and policy makers utilize in order to better understand performance and production (technical) efficiency of the entities they are benchmarking against other competitors or peers. As well we discuss, give examples of, and provide extensive links to R programs that implement these various methods and approaches. The various methods and approaches distinguish themselves by leveraging the estimation of relative performance and of production efficiency measures on regression-based methods and on linear programming methods, the former referred to as Stochastic Frontier Analysis (SFA) and the latter as Data Envelopment Analysis (DEA). We discuss the main approaches in turn, their relative strengths and weaknesses, and briefly touch on ways to aggregate the various methods using model averaging approaches.

Our chapter is organized in the following way. We first briefly discuss the motivation for using such methods that rest on the presence of production efficiency differences among units of production that are being compared. The various sets of rationales for such a discipline as efficiency and productivity analysis can be categorized into three main groups of motivating factors. They are based on varying management practices, which deliver heterogeneous outcomes, behavioral economics, and the presence of X-efficiency. We then discuss the two main classes of production efficiency estimators, the SFA and the DEA estimators. We next discuss examples of R programs that implement these various classical estimators and various extensions that have been introduced in the recent literature as well as protocols for accessing R programs on well-documented websites and other open source sites. We then provide concluding remarks.

2 Why Estimate Production (Technical) Efficiency?

One of the most compelling rationales for the study of production efficiency is substantial heterogeneity in management practices and resulting changes in the operating efficiency of a firm. A consensus in the empirical literature exists that indicates substantial productivity differences both within a firm[1] over time and among firms (Lieberman et al. (1990); Foster et al. (2008); Hsieh and Klenow (2009); Hall and Jones (1999)). Glaister (2014), among others, have noted that management practices are a key factor in explaining such productivity differences. Other factors, such as expenditures on R&D, utilization of capacity, and technology adoption, which are key decisions of management (Nallari and Bayraktar (2010)), are typically controlled by management practices. Bloom et al. (2012) regressed gross domestic product (GDP) per capita on a set of indicators of management practices among 17 countries. These indicators of management practices explained 87% of the variation in per capita GDP. These findings are corroborated in a more micro oriented study of Indian firms by Bloom et al. (2013). The engineering mechanism, blueprints, formal structural statistical model, or economic model that explains how a firm's productivity is linked to management skills and practices has not been developed in a way that lends itself to empirical analysis and to the generation of relative technical efficiency differences among peer firms. We thus tend to view such a factor as an unobservable latent factor and assume that management practices are one of the key factors in firm productivity. Another factor is innovation but innovation is oftentimes facilitated by decisions and practices of management. The literature on the effects of management practices on production efficiency, labor efficiency, and related measures of a firm's financial success is quite dense.[2] Behavioral economics provides another motivation for why firms may not operate at the frontier of production efficiency. The assumptions of efficient markets and rational decision makers have been leveraged with substantial success by neoclassical economists, but the footing on which these assumptions rest may be a bit loose and slippery.

[1]We will use the term "firm" to denote any generic unit whose production efficiency is being measured and estimated.

[2]Much of this literature is summarized and referenced in Grifell-Tatjé and Lovell (2013, 2015) and Grifell-Tatjé et al. (2018).

X-efficiency theory is a pragmatic paradigm that admits to the prevalence of various sources on inefficiency in economics. It was introduced by Leibenstein (1966, 1975, 1987), who pointed out that agency problems, asymmetric information, and monitoring by regulators were all factors that generated incentives to engage in sub-optimal decision making, absent these factors and constraints. Drawing from studies of the health care industry, telecommunications, airlines, and education, Frantz (1997, 2007) have documented the ubiquitous existence of levels of inefficiency with the predictions from X-efficiency theory and have also concluded that such production inefficiency is much more significant than inefficiencies due to incorrect output and input allocations (allocative inefficiency). Other studies that have documented such inefficiency levels in the banking system can be found in Kwan (2006); Jiang et al. (2009); Fu and Heffernan (2009); Yao et al. (2008); Rezvanian et al. (2011); Bauer and Hancock (1993); Mester (1993); DeYoung (1998). And as emphasized by Frantz (1997)

"*...what becomes of the word maximize if non-maximize is not possible? Is the concept of efficiency important if the possibility of inefficiency is ruled out a priori? The importance of efficiency remains as long as economics remains important...*"

X-efficiency theory and the methods that we present in our chapter is based on an interdisciplinary approach that combines psychology, management, statistics, applied mathematics, and engineering.

Finally, the usefulness and importance of SFA and DEA methods have passed the market test as they are required to be used in a wide variety of regulatory decision-making settings in Europe and elsewhere (Bogetoft (2013), Agrell et al. (2017)).

3 Regression-Based Methods to Estimate Production Efficiency

We begin our formal discussion of production efficiency measurement by first defining a few sets and functions that are necessary for our presentation of methods and how to implement them. To facilitate further discussion, let inputs and outputs be represented by $x = (x_1, ..., x_N)' \in \Re_+^N$ and $y = (y_1, ..., y_M)' \in \Re_+^M$, respectively and the production process characterized by a technology set T that satisfies standard regularity conditions (Sickles and Zelenyuk (2018)). The output set is defined as

$$P(x) \equiv \{y \in \Re_+^M : y \text{ is producible from } x \in \Re_+^N\}. \tag{1}$$

The production function $f : \Re_+^N \to \Re_+^1$ is defined as $f(x) \equiv \max\{y\colon y \in P(x)\}$. The maximum of the production function exists and is unique owing to the fact that *P(x)* is a compact set. Details on the regularity conditions that assure the maximum is unique and exists also can be found in Sickles and Zelenyuk (2018). When there is more than one output one can use the generalization of the production function, *output orientated* (Shephard (1970)) distance function $D_o : \Re_+^N \times \Re_+^M \to \Re_+^1 \cup \{+\infty\}$ as $D_o(x, y) \equiv \inf\{\theta > 0 : y/\theta \in P(x)\}$. For a technology with only one output and holding the inputs constant, D_o is the ratio of actual output to potential (maximal) output. Then frontier production is $f(x^o) = y^o/D_o(x^o, y^o)$. When there are multiple-outputs, the Shephard output distance function is the smallest scalar required to radially expand *all* outputs to the output set's boundary, again at a fixed level of inputs.

3.1 The Stochastic Frontier Paradigm

In cross-sectional Stochastic Frontier Analysis (SFA) there are at least two sources of error (panel extensions may extend the identifiable error components for four distinct sources of error). In addition to the error that is appended to the parametric or nonparametric production or distance function to account for standard statistical noise (assumed to have mean zero) an additional source of error is added to address the asymmetry in the errors displayed by most production or distance function estimates that is due to the boundary property of the function being estimated. Such a one-sided error is utilized additively as a placeholder for production or technical inefficiency, which diminishes the level of observed output from the frontier production or distance function specified parametrically or nonparametrically. An example of such an SFA model for a linear in logs (Cobb-Douglas) production function is,

$$\ln(y) = \beta_0 + \sum_{k=1}^{N} \beta_k \ln(x_k) + \varepsilon, \tag{2}$$

$$\varepsilon = v - u. \quad (3)$$

Here v is a random term with mean zero and variance σ_v^2. The error term u represents the latent production inefficiency term. The latent production inefficiency term has positive support with a mean $\mu > 0$ and a variance σ_u^2.[3] Although these assumptions have been modified to account for more realistic empirical settings, this forms the basis for the canonical model used in SFA. Most of the literature in SFA has worked to generalize the distributions of u and v, their moment properties, stochastic properties of the inputs (and for multi-output distance functions the outputs as well), and different dependency patterns among the disturbances and the inputs and outputs, as well as possibly distinguishing between variables that are used in developing the production boundary and environmental or confounding factors that may influence production efficiency, that is the deviations between the boundary and non-boundary observations, as opposed to influencing the boundary per se.

3.1.1 Corrected OLS

One of the first approaches to develop measures of production efficiency is the corrected ordinary least squares. The steps employed in such an exercise involve first estimating the production (or distance) function by ordinary least squares to get estimates of the average production relationship. In the second step one simply shifts the intercept (the example here is for the single output production function) to ensure that the residuals are all non-positive (Olson et al. (1980)). With only mild assumptions, OLS will yield minimum variance unbiased linear estimators. The step with the Cobb-Douglas production function involves estimating

$$\ln(y_i) = \beta_0 + \beta_1 \ln(x_{i1}) + ... + \beta_N \ln(x_{iN}) + \varepsilon_i, \quad i = 1, ..., n. \quad (4)$$

The second step uses the intercept correction

$$\hat{\beta}_0^{cols} := \hat{\beta}_0 + \max_i \{\hat{\varepsilon}_i\}, \ i = 1, ..., n, \quad (5)$$

where $\hat{\varepsilon}_i$ are OLS residuals. The last step estimates production efficiency for a firm as

$$\text{Technical Inefficiency} = \hat{\varepsilon}_i^{cols} := \max_i \{\hat{\varepsilon}_i\} - \hat{\varepsilon}_i, \quad i = 1, ..., n. \quad (6)$$

[3] Cost inefficiency can be modeled with a one-sided error with only negative support.

Thus the corrected production function that estimates the production frontier is

$$\ln(y_i) = \hat{\beta}_0^{cols} + \hat{\beta}_1 \ln(x_{i1}) + ... + \hat{\beta}_N \ln(x_{iN}) - \hat{\varepsilon}_i^{cols}, \quad i = 1, ..., n. \tag{7}$$

3.1.2 Stochastic Frontier Model

Aigner et al. (1977) (hereafter ALS) and Meeusen and van den Broeck (1977) pursued a parametric model of the stochastic frontier based via maximum likelihood. Average inefficiency is defined in terms of the performance of a firm to the firm identified as having the best-practices, as measured by its level of efficiency. The original ALS model used a half-normal distribution for the efficiency term and a normal error for the idiosyncratic disturbance. Many other distributions have been considered, usually for the inefficiency term. These include the exponential, truncated normal, gamma, and doubly truncated normal (Stevenson (1980), Greene (1980a,b),Qian and Sickles (2008), Almanidis and Sickles (2012), and Almanidis et al. (2014)). In the canonical SFA the composite error terms are assumed to be independent. If the production function is linear in logs then a convenient parameterization is

$$y_i = f(x_i|\beta)\exp(\varepsilon_i), \quad i = 1, ..., n \tag{8}$$

and for $\varepsilon = v - u$ the inefficiency of firm i is measured as

$$\exp(-u_i) \equiv \frac{y_i}{f(x_i)\exp(v_i)}, \quad i = 1, ..., n. \tag{9}$$

The model is usually specified after log-transforming the production relationship

$$\ln y_i = \ln f(x_i|\beta) + v_i - u_i, \quad i = 1, ..., n \tag{10}$$

and assuming

$$v_i \sim \mathcal{N}(0, \sigma_v^2)$$

and

$$u_i \sim |\mathcal{N}(0, \sigma_u^2)|$$

and that u_i and v_i are independent and i.i.d., it follows that the density of ε_i is:

$$f_{\varepsilon_i}(\varepsilon) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\left[1 - \Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right], \quad -\infty \leq \varepsilon \leq +\infty \tag{11}$$

where $\Phi(\cdot)$ is the standard normal distribution function, $\sigma^2 = (\sigma_v^2 + \sigma_u^2)$, and $\lambda = \sigma_u/\sigma_v$. The composite errors, $\varepsilon_1, ..., \varepsilon_n$, are of course not observed, but from our model we know that

$$\varepsilon_i = \ln y_i - \ln f(x_i), \quad i = 1, ..., n. \tag{12}$$

The log-likelihood function is

$$\begin{aligned} \ell(y_1, ..., y_n|\beta, \lambda, \sigma^2) =& \frac{n}{2} \ln\big(\frac{2}{\pi}\big) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \big[\ln y_i - \ln f(x_i)\big]^2 \\ &+ \sum_{i=1}^{n} \ln\Big[1 - \Phi\big(\frac{[\ln y - \ln f(x_i)]\lambda}{\sigma}\big)\Big]. \end{aligned} \tag{13}$$

Inefficiency has a mean and variance given by

$$E(u_i) \equiv \mu = \frac{\sqrt{2}}{\sqrt{\pi}} \sigma_u \tag{14}$$

and

$$V(u_i) = \left(\frac{\pi - 2}{\pi}\right) \sigma_u^2 \tag{15}$$

and thus the composed error's mean and variance is

$$E(\varepsilon_i) = E(v_i - u_i) = E(-u_i) = -\mu = -\frac{\sqrt{2}}{\sqrt{\pi}} \sigma_u \tag{16}$$

$$V(\varepsilon_i) = V(v_i - u_i) = V(v) + V(u) = \sigma_v^2 + \left(\frac{\pi - 2}{\pi}\right) \sigma_u^2 \tag{17}$$

and

$$cov(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j. \tag{18}$$

Estimation of individual inefficiencies Cross-sectional SFA, unfortunately, does not provide a consistent estimate of the efficiency of each firm. A simple solution by Materov (1981) and further developed in Jondrow et al. (1982) uses $E(u_i|\varepsilon_i)$ as a point estimate of u_i. This estimator has its drawback as it is contaminated by statistical noise but it is appealing in that, from the law of iterated expectations,

$$E(u_i) = E_\varepsilon[E(u_i|\varepsilon_i)] \tag{19}$$

and thus to develop a consistent estimator of $E(u_i|\varepsilon_i)$ we can write

$$E(u_i|\varepsilon_i) \equiv \int_0^\infty u f_{u_i|\varepsilon_i}(u|\varepsilon) du \tag{20}$$

where, from the definition of the conditional density,

$$f_{u_i|\varepsilon_i}(u|\varepsilon) \equiv \frac{f_{u_i,\varepsilon_i}(u,\varepsilon)}{f_{\varepsilon_i}(\varepsilon)}. \tag{21}$$

Based on the normal/half normal composed error structure typically used in SFA we can write

$$E(u_i|\varepsilon_i) = \mu_* + \sigma_* \frac{1}{1-\Phi(-\mu_*/\sigma_*)} \phi\left(\frac{-\mu_*}{\sigma_*}\right) \tag{22}$$

or

$$E(u_i|\varepsilon_i) = \sigma_* \left[\frac{\mu_*}{\sigma_*} + \frac{1}{1-\Phi(-\mu_*/\sigma_*)} \phi\left(\frac{-\mu_*}{\sigma_*}\right)\right]. \tag{23}$$

Using the parameterization $-\frac{\mu_*}{\sigma_*} = \frac{\sigma_u^2 \varepsilon}{\sigma^2 \frac{\sigma_v \sigma_u}{\sigma}} = \frac{\sigma_u \varepsilon}{\sigma \sigma_v} = \frac{\varepsilon \lambda}{\sigma}$, the conditional expectation is:

$$E(u_i|\varepsilon_i) = \frac{\sigma_v \sigma_u}{\sigma}\left[-\frac{\varepsilon_i \lambda}{\sigma} + \frac{\phi\left(\varepsilon_i \lambda/\sigma\right)}{1-\Phi(\varepsilon_i \lambda/\sigma)}\right] \tag{24}$$

for which a consistent estimate can be obtained based on consistent estimates of the model parameters (e.g., obtained via MLE or COLS, or the Modified OLS procedures of Olson et al. (1980)). Thus a (conditional) consistent estimator of each firm's efficiency level can be based on this last equation.

3.1.3 Panel Stochastic Production Frontiers

The cross-sectional stochastic frontier model has a number of drawbacks that have been addressed over the forty years since it was introduced. As we pointed out, a consistent estimate for a firm's technical efficiency is not available, only a consistent estimate of the conditional mean of the firm's efficiency level. Of course the canonical ALS model is fully parametric and inputs are assumed to be uncorrelated with the regressors, which is particularly troubling when the regressors are inputs and the latent factor that typically is assumed to account for inefficiency is unobservable managerial expertise, an input whose independence from the levels of capital and labor used in production is a questionable assumption, and of course if a firm is aware of its level of technical efficiency this information, although unknown to the analyst, should not be independent of the firm's input choices.

These problems are potentially avoidable if one has panel data (Pitt and Lee (1981); Schmidt and Sickles (1984)), although the possible endogeneity of input choice must be addressed with care and its treatment using fixed effects type estimators may not be completely satisfactory. We will return to this issue shortly. If we have a panel of firms, then if the unobserved firm effects represent technical efficiency they can be estimated consistently for large T (assuming that our sample of firms n is large) after controlling for inputs and other environmental and observable factors that may impact production. To show how this is accomplished with panel data we use the general treatment discussed in Schmidt and Sickles (1984) (SS), which also considers the Pitt and Lee (1981) parametric random effects model. The model is

$$y_{it} = \alpha + x_{it}'\beta + v_{it} - u_i, \;\; i = 1, ..., n; t = 1, ..., T, \tag{25}$$

which can be rewritten as

$$y_{it} = \alpha^* + x_{it}'\beta + v_{it} - u_i^* \tag{26}$$

where $\alpha^* = \alpha - \mu$; $u_i^* = u_i - \mu$; $E(u_i) = \mu \geq 0$, and where x_{it} is a vector of N inputs. If we let $\alpha_i = \alpha^* - u_i^*$ then the model becomes the usual panel data model

$$y_{it} = \alpha_i + x_{it}'\beta + v_{it} \tag{27}$$

and cross-sectional effects that can be viewed as random, fixed, or simply ignored. Five estimators of the classical panel data model with time-invariant effects are discussed in SS. These are the pooled OLS model, the fixed effects within estimator, the random effects model, the Hausman-Taylor estimator, and the fully parametric random effects MLE model (Pitt and Lee (1981)). Schmidt and Sickles (1984) discuss the asymptotics of each of these estimators and the assumptions needed in order for the parameter estimates and estimates of the technical efficiency level to be consistently estimated.

Technical efficiency effects estimates with the SS estimators do not change over time and this is a strong and often unreasonable assumption that is not necessary. The suite of panel stochastic frontier estimators developed by

Cornwell et al. (1990) (CSS) had a parameterization that allowed for time-varying heterogeneity and is based on the model:

$$y_{it} = x_{it}'\beta + z_i'\gamma + w_{it}'\delta_i + v_{it} \quad i = 1, ..., n; \, t = 1, ..., T, \tag{28}$$

where x_{it}, z_i and w_{it} are $N \times 1$, $J \times 1$ and $L \times 1$ vectors, respectively, and the parameter vectors β, γ and δ_i are dimensioned conformably. With $\delta_0 = E[\delta_i]$, and $\delta_i = \delta_0 + u_i$ the model can be written as:

$$y_{it} = x_{it}'\beta + z_i'\gamma + w_{it}'\delta_0 + \varepsilon_{it},$$

where

$$\varepsilon_{it} = v_{it} + w_{it}'u_i \tag{29}$$

and where u_i is assumed to be $i.i.d.$, zero mean random variables with co-variance matrix Δ.The error term v_{it} is assumed to be $i.i.d.$, with zero mean and constant variance σ_v^2 . The basic model assumes that v_{it} is uncorrelated with z, x, and u_i. In order to allow for time-varying technical efficiency the stochastic frontier model of Schmidt and Sickles (1984)

$$y_{it} = \alpha + x_{it}'\beta + v_{it} - u_i = \alpha_i + x_{it}'\beta + v_{it}. \tag{30}$$

can be modified by replacing the α_i with, e.g.,

$$\alpha_{it} = \theta_{i1} + \theta_{i2}t + \theta_{i3}t^2, \tag{31}$$

and the model in matrix form becomes

$$y = X\beta + Z\gamma + W\delta_0 + \varepsilon, \tag{32}$$

$$\varepsilon = Qu + v. \tag{33}$$

Details of the estimator can be found in Cornwell et al. (1990). Fixed effects, random effects, and Hausman-Taylor estimators of the CSS model were developed to estimate productivity efficiency that is time varying and allow for consistent estimation under large n and T asymptotics of the time varying productivity efficiency for each firm. The parameter δ_i is estimated by regressing the residuals $(y_{it} - x_{it}'\beta)$ for firm i on w_{it}. This amounts to regressing the within residual on a constant term, time, and time-squared for the specification we introduced for α_{it} above and relative efficiencies can be approximated (in the linear in logs production function) by

$$\hat{\alpha}_t = \max_i(\hat{\alpha}_{it})\ i = 1, ..., n \tag{34}$$

$$\hat{u}_{it} = \hat{\alpha}_t - \hat{\alpha}_{it}. \tag{35}$$

Parametric MLE case also be used to estimate a random effects time varying technical efficiency models. Kumbhakar (1990) and Battese and Coelli (1992) present two such models. In the former inefficiency is modeled as

$$u_{it} = (1 + \exp(bt + ct^2))^{-1}\tau_i, \tag{36}$$

where a and b are parameters to be estimated and where τ_i's distribution is assumed to be $i.i.d$ $N^+(0, \sigma_\tau^2)$ and v_{it} is $i.i.d$ $\mathcal{N}(0, \sigma_v^2)$. In the latter specification inefficiency is modeled as

$$u_{it} = \eta_t\tau_i = \{\exp[-\eta(t-T)]\}\tau_i, \quad t = 1, \ldots, T;\ i = 1, \ldots, n, \tag{37}$$

where v_{it} is $i.i.d\,\mathcal{N}(0,\, \sigma_v^2)$ random variables, τ_i is assumed to be $i.i.d.$ and has a non-negative truncated distribution $\mathcal{N}(\mu,\, \sigma^2)$, and η is a scalar parameter.

3.1.4 Second-and Third-Generation Stochastic Frontier Models

The models we have just discussed provide the basic statistical intuition for SFA. Much has been done to extend and generalize the canonical models we have discussed and space prohibits us from providing comparable detail on these second and third generation extensions. However, the R codes that we will discuss shortly are equipped to deal with a number of these relatively new modeling scenarios and so we give a brief summary of what these second and third generation SFA models deliver and the general ideas behind how they are formulated and specified. We leave it to the reader to seek out the original sources and recent books, handbooks, and survey articles that we have referenced and that provide more detailed treatments.

Researchers have provided further generalizations of the CSS time-varying technical efficiency model that focus on two main aspects of the model. The first is to allow for a factor-type structure for the cross-section and time varying technical efficiency term (these are almost all exclusively panel data extensions). The second is to decompose the time invariant and time varying efficiency terms into a portion that is skewed, to represent technical inefficiency, and a portion that is symmetric, the latter somewhat courageously called the "true" effects model. There are four other generalizations that some may view as comparable to these but there is simply not enough space to address them in any substantive way. We will mention them and provide several references as well as links to R codes that deal with these four additional issues, which are: environmental factors, endogeneity, Bayesian methods for SFA, and nonparametric specifications of the technology and of the error structure.

We briefly discuss these sets of generalizations and available R code to implement them.

Factor Models and SFA

Lee (1991) and Lee and Schmidt (1993) were the first to propose a (one component) factor model to address time-varying and cross-sectional specific production efficiency. Their model is

$$y_{it} = \alpha_t + x'_{it}\beta + v_{it} - u_{it} \quad for \quad i = 1, \ldots, n; \, t = 1, \ldots, T, \tag{38}$$

or as

$$y_{it} = x'_{it}\beta + \alpha_{it} + v_{it}, \tag{39}$$

where $\alpha_{it} = \alpha_t - u_{it}$ is the time varying cross-sectional specific technical efficiency term, which are modeled as

$$\alpha_{it} = \eta_t \delta_i. \tag{40}$$

Here $\eta_t, t = 1, ..., T$, are the time-varying effects to be estimated and $\delta_i, i = 1, ..., n$, are the firm effects. Lee and Schmidt (1993) provide fixed-effects and random effects estimators for this one-factor model based one nonlinear regression estimators.

Ahn et al. (2007) generalized this model to allow for multiple factors that change over time and specify the production frontier as

$$\begin{aligned} y_{it} &= \delta_t + x'_{it}\beta + v_{it} - u_{it} && (41) \\ &= x'_{it}\beta + \eta_{it} + v_{it}, \; i = 1, 2, \ldots, n, \; t = 1, 2, \ldots, T, && (42) \end{aligned}$$

where, again, v_{it} is the usual disturbance term and $u_{it} \geq 0$ is the inefficiency term and where $\eta_{it} \equiv \delta_t - u_{it}$ is the time-varying and cross-sectional specific technical efficiency term that is expressed as a linear combination of p unrestricted components,

$$\eta_{it} = \theta_{1t}\alpha_{1i} + \theta_{2t}\alpha_{2i} + ... + \theta_{pt}\alpha_{pi} = \sum_{j=1}^{p} \theta_{jt}\alpha_{ji}. \tag{43}$$

The model is estimated using generalized methods of moments. Ahn et al. (2013) provide a focused study of consistency properties of their Ahn et al. (2007) model when different sorts of dependency relationships exist between the production efficiency effects and the regressors.

In the Ahn et al. (2007) model the effects are multiplicative. Kneip et al. (2012) (KSS) provided a more general model than Ahn et al. (2007) by allowing for a general nonparametric time-varying and cross-section specific productivity efficiency. The KSS model is:

$$y_{it} = \beta_0(t) + \sum_{j=1}^{N} \beta_j x_{itj} + u_i(t) + v_{it}, \quad i = 1, \dots, n; \, t = 1, \dots, T. \tag{44}$$

The $u_i(t)$'s are assumed to be smooth time-varying individual effects that satisfy a normalization that $\sum_i u_i(t) = 0$ and are a linear combination of $L < T$ basis functions (common factors) $g_1, ..., g_L$:

$$u_i(t) = \sum_{r=1}^{L} \theta_{ir} g_r(t). \tag{45}$$

The term $\beta_0(t)$ is an average function that is eliminated by centering the model yielding

$$y_{it} - \bar{y}_t = \sum_{j=1}^{N} \beta_j(x_{itj} - \bar{x}_{tj}) + u_i(t) + v_{it} - \bar{v}_i, \quad i = 1, \dots, n; \, t = 1, \dots, T \tag{46}$$

where $\bar{y}_t = \frac{1}{n}\sum_i y_{it}$, $\bar{x}_{tj} = \frac{1}{n}\sum_i x_{itj}$ and $\bar{v}_i = \frac{1}{n}\sum_i v_{it}$. As with the CSS and many other panel stochastic frontier estimators, technical efficiency is calculated as $TE_i(t) = \exp\{u_i(t) - \max_{j=1,\dots,n}(u_j(t))\}$, just as it is calculated with the CSS estimator. The estimator used is a three step one that first estimates β using penalized least squares and smoothing splines to approximate the factors, then second generates estimates of the covariance structure of the $u_i(t)'s$, and the third step estimates the basis functions $\hat{g}_r$ and then updates estimates of u_i by $\sum_{r=1}^{L}\hat{\theta}_{ir}\hat{g}_r$. The KSS estimator and related estimators such as those introduced by Bai and Ng (2002) and Bai (2009), as well as tests for the number of factors, have been programmed in R and are available not only in the suite of programs we discuss at the end of this chapter but also can be found on the website referenced in Bada and Liebl (2014).

True Fixed Effects and SFA

Greene (2005a,b) proposed a stochastic panel frontier model in which the intercept fixed effects were not measures of persistent inefficiencies, as had been assumed by SS, but rather was simply firm specific heterogeneity. This has become known as the "true" fixed effects model, although a less ambitious label may be the panel frontier with only transitory inefficiency. The model is

$$y_{it} = \alpha_i + x'_{it}\beta + v_{it} - u_{it} \tag{47}$$

$$\varepsilon_{it} = v_{it} - u_{it}, \tag{48}$$

or

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it} \tag{49}$$

and thus the source of inefficiency is u_{it}, which is distinguished from the intercept terms α_i. Greene outlines estimators for both the fixed effects and random effects panel stochastic frontier model with only transitory inefficiency based on simulated MLE and technical efficiency estimates are based on the familiar expression

$$\widehat{TE}_{it} = \exp[-\{max(\widehat{u}_{it}) - \widehat{u}_{it}\}]. \tag{50}$$

Colombi et al. (2011), Colombi et al. (2014), and Tsionas and Kumbhakar (2014), provided further generalizations of this model and alternative estimators based on Bayesian methods, full information maximum likelihood, quasi-MLE, and method of moments. One important generalization proposed by these authors specifies the persistent and transitory efficiency terms as separate skew-normal errors based on the following parameterization ofColombi et al. (2014)

$$y_{it} = x'_{it}\beta + \alpha_i + v_{it} - u_{it} - \eta_i, \tag{51}$$

where $\varepsilon_{it} = \alpha_i + v_{it} - u_{it} - \eta_i$ is the error structure and where u_{it} and η_i are non-negative random variables that capture firm-specific random effects, random noise, short-run technical inefficiency, and long-run technical inefficiency.

Environmental factors can be introduced into the SFA paradigm in a natural way by including them in a vector of variables that is assumed to impact the mean and the variance of production efficiency. For an excellent treatment of the problem see Wang and Schmidt (2002), Simar and Wilson (2008) and Kim and Schmidt (2008). The marginal effects of the environmental factors, based on Battese and Coelli (1995), can be estimated using the R package Frontier (http://cran.r-project.org/web/packages/frontier/) developed by Arne Henningsen.

Various dependency structures have been assessed within the SFA paradigm. An excellent source of recent work on this topic can be found in Kumbhakar and Schmidt (2016).

No one has had more of an impact on the introduction of Bayesian methods into the SFA and DEA paradigms than Mark Steel and Mike Tsionas (Griffin and Steel (2007), Tsionas and Papadakis (2010), and Liu et al. (2017)). The WinBUGS package of Steele is Fortran and Matlab-based open source code that can be modified to run in R. Nonparametric models in SFA have been studied in a variety of settings by Park et al. (1998), Adams et al. (1999), Park et al. (2003), Park et al. (2007), among others, and the R codes for the latter models are discussed in more detail in the last sections of this chapter on specific implementation of open source R codes for SFA and DEA. Experimental R codes that can estimate both parametric and non-parametric stochastic frontiers can also be found in the experimental R-Forge package maintained by Arne Henningsen.

4 Envelopment Estimators

In this section we will briefly review another very popular approach in the measurement and empirical estimation of the efficiency of economic systems (firms, industries, etc.), called Data Envelopment Analysis (DEA).

4.1 The Origins of DEA

In a nutshell, DEA is rooted in and coherent with theoretical economic modeling using *Activity Analysis Models* (AAM) and is estimated via the powerful linear programming approach. Its name was branded in the seminal work of Charnes et al. (1978), who refined and generalized the approach of Farrell (1957) to estimate production efficiency, which in turn was influenced by seminal works of Debreu (1951), Koopmans (1951a,b) and Shephard (1953, 1970).

4.2 The Basic DEA Model

In his seminal work, Farrell (1957) focused on the constant returns to scale model, with multiple inputs and a single output. About two decades later, Charnes et al. (1978) generalized Farrell's approach to the multi-output case but started with a very different formulation–a fractional programming problem formulation with the objective being to optimize the ratio of a weighted aggregate of outputs to a weighted aggregate of inputs (i.e., a kind of productivity index). They then transformed this problem into a linear programming (LP) problem and derived its dual, which turned out to be the (generalized version of) AAM proposed by Farrell. Specifically, the formulation of Charnes et al. (1978) for estimating the efficiency score of a firm or decision-making unit (DMU) $j \in (1, ..., n)$ with an allocation (x^j, y^j), states

$$E^j_{i,CCR} = \max_{\substack{v_1,...,v_N; \\ u_1,...,u_M}} \left\{ \begin{array}{c} \frac{\sum_{m=1}^M u_m y_m^j}{\sum_{l=1}^N v_l x_l^j} : \frac{\sum_{m=1}^M u_m y_m^k}{\sum_{l=1}^N v_l x_l^k} \leq 1, k = 1, ..., n, \\ u_m \geq 0, v_l \geq 0, l = 1, ..., N; m = 1, ..., M \end{array} \right\} \tag{52}$$

where $u' = (u_1, ..., u_M)$ and $v' = (v_1, ..., v_N)$ are optimization variables (also called here 'multipliers'). This DEA formulation is more popular in the operations research and management science literature and is often referred to as the 'CCR model' or the 'multiplier form of DEA' under CRS, additivity and free disposability.

Importantly, Charnes et al. (1978) then showed that (52) is equivalent to the AAM version of the DEA-*estimator of the Farrell input oriented technical efficiency* score of a DMU with an allocation (x^j, y^j) under the assumption of CRS (also assuming additivity and free disposability of all inputs and all outputs), formulated as

$$
\begin{aligned}
\widehat{E}_i(x^j, y^j) &\equiv \min_{\lambda, z^1, ..., z^n} \lambda && (53)\\
& s.t. \\
\sum_{k=1}^{n} z^k y_m^k &\geq y_m^j, m = 1, ..., M, && (54)\\
\sum_{k=1}^{n} z^k x_l^k &\leq \lambda x_l^j, l = 1, ..., N, && (55)\\
\lambda &\geq 0, z^k \geq 0, k = 1, ..., n, && (56)
\end{aligned}
$$

which is an LP problem that can be solved via any standard LP solver. This formulation is more common in the economics literature (largely due to its connection to works of Debreu (1951) and Koopmans (1951a,b)), and is often referred to as the *envelopment form of DEA* under CRS, additivity and free disposability. We will use this envelopment formulation to describe other variants of DEA, though it is useful to keep in mind that there is also a multiplier form that optimizes what can be viewed as a normalized productivity index.

The previous formulation looks at minimization of all inputs, while keeping outputs fixed, and hence is called the input orientation. Similarly, the DEA-*estimator of the Farrell output oriented technical efficiency* score of any (x^j, y^j) allocation, under the assumptions of CRS, and additivity and free disposability of all outputs and all inputs is formulated as

$$
\begin{aligned}
\widehat{E}_o(x^j, y^j) &\equiv \max_{\lambda, z^1, ..., z^n} \lambda && (57)\\
& s.t. \\
\sum_{k=1}^{n} z^k y_m^k &\geq \lambda y_m^j, m = 1, ..., M, && (58)\\
\sum_{k=1}^{n} z^k x_l^k &\leq x_l^j, l = 1, ..., N, && (59)\\
\lambda &\geq 0, z^k \geq 0, k = 1, ..., n. && (60)
\end{aligned}
$$

Immediately note that for these formulations we have $\widehat{E}_i(x^j, y^j) = 1/\widehat{E}_o(x^j, y^j)$ for any (x, y), as is required theoretically due to CRS.

Also note that the reciprocals of these estimated Farrell efficiency measures also give estimates of the input and output oriented Shephard's distance functions (Shephard (1953, 1970)), under the same assumptions on technology, i.e., CRS, additivity and free disposability.

Sometimes a researcher may not be interested in fixing only the levels of inputs or outputs, but may want to simultaneously expand outputs and contract inputs, thus requiring other orientations. The DEA-estimator of a general efficiency measure of any (x^j, y^j) allocation in such a case (also here under the assumptions of CRS and additivity and free disposability of all outputs and all inputs), can be obtained as follows

$$\begin{aligned} \widehat{GE}_o(x^j, y^j) &\equiv \max_{\lambda_1,...,\lambda_N,,\theta_1,...,\theta_M,z^1,...,z^n} f(\lambda_1, ..., \lambda_N, , \theta_1, ..., \theta_M) && (61) \\ & \quad s.t. \\ \sum_{k=1}^{n} z^k y_m^k &\geq \theta_m y_m^j, m = 1, ..., M, && (62) \\ \sum_{k=1}^{n} z^k x_l^k &\leq x_l^j / \lambda_l, l = 1, ..., N, && (63) \\ \lambda &\geq 0, z^k \geq 0, k = 1, ..., n, && (64) \end{aligned}$$

where certain restrictions can be imposed on the objective function $f(\lambda_1, ..., \lambda_N, , \theta_1, ..., \theta_M)$ and its arguments, depending on the interest of the researcher. For example, restricting f to be additive (but summing only positive arguments) will give the DEA estimate of the general Russell efficiency measure and an additional restriction of $\theta_1 = ... = \theta_M = 1$ will turn it into the input oriented Russell efficiency measure (as was originally introduced by Färe and Lovell (1978)). If one instead restricts $\lambda_1 = ... = \lambda_N = 1$ then one obtains the output oriented Russell efficiency measure. Meanwhile, restricting $f(\lambda_1, ..., \lambda_N, \theta_1, ..., \theta_M)$ to be multiplicative (a geometric mean) will generate the multiplicative-Russell efficiency measure introduced by Färe et al. (2007). Furthermore, if instead one imposes $\theta_1 = ... = \theta_M = \theta$ and $\lambda_1 = ... = \lambda_N = \lambda$ then the DEA estimate of the what has been termed the general hyperbolic efficiency measure is obtained. This latter measure sometimes also appears with the additional restriction that $\theta = \lambda$, which imposes the properties of equiproportional expansion (contraction) of all output (inputs). Note that in general, this last formulation is no longer an LP problem and therefore, it is typically more challenging to estimate.

Another very general measure, as well as a primal characterization of technology, is the directional distance function,[4], which can be estimated via the following DEA formulation (here also under the assumptions of CRS and additivity and free disposability of all outputs and all inputs):

$$\begin{aligned}
\widehat{D}_d(x^j, y^j|d_x, d_y) &\equiv \max_{\lambda, z^1,...,z^n} \lambda, &(65)\\
&\quad s.t. \\
\sum_{k=1}^{n} z^k y_m^k &\geq y_m^j + \lambda d_{y_m}, m = 1, ..., M, &(66)\\
\sum_{k=1}^{n} z^k x_l^k &\leq x_l^j - \lambda d_{x_l}, l = 1, ..., N, &(67)\\
\lambda &\geq 0, z_k \geq 0, k = 1, ..., n. &(68)
\end{aligned}$$

In a similar fashion, DEA can be used to model and estimate cost, revenue and profit functions and associated efficiency measures. More details on this can be found in Sickles and Zelenyuk (2018).

4.3 The Myriad of DEA models

The approach briefly described in the previous section in the context of different efficiency measures constitutes the canonical forms of the DEA paradigm. Essentially, all the other versions are modifications or extensions of the models we have outlined. In this section we briefly discuss a few of these modifications and extensions.

Oftentimes, extensions are obtained by imposing various additional constraints onto either the envelopment form or the multiplier form of DEA, with an aim to better mimic the particular actual production process under study. Such additional restrictions should be imposed with care, since they may (and often do) affect other desirable properties related to previously added constraints, such as CRS, convexity, free disposability, additivity, etc.

Additional constraints also may create computational complications, e.g., turning the problem from a linear to a non-linear one or a hybrid problem

[4]The origins of ideas for this function go back to at least Allais (1943), Diewert (1983), and Luenberger (1992), and were later revived and thoroughly developed by Chambers et al. (1996)

that may require integer-programming problems. This may result in possibly making the problem much harder to compute, possibly infeasible, or may give inferior local optima or degenerate solutions.

4.3.1 Relaxing Constant Returns to Scale and Convexity

The first wave of extensions of the DEA-CRS model mainly focused on relaxing assumptions such as CRS and convexity. This line of research was pursued by a number of researchers who extended and enriched the DEA paradigm, among them Afriat (1972); Färe et al. (1983); Banker et al. (1984); Deprins et al. (1984) and Petersen (1990); Bogetoft (1996), to mention a few.

Out of the many modifications and extensions of the canonical DEA problem the several that have sustained the test of time and popularity are the DEA-VRS (variable returns to scale) and the DEA-NIRS (non-increasing returns to scale) models, which simply amount to adding additional constraints in the form of $\sum_{k=1}^{n} z_k = 1$ or $\sum_{k=1}^{n} z_k \leq 1$, respectively, to the DEA-CRS formulations as described above.

Meanwhile, the Free Disposal Hull (FDH) approach, can be implemented via a hybrid of the linear programming and the integer programming problems, which is formulated in exactly the same way as the DEA-VRS programming problem with the exception that the constraints "$z_k \geq 0, k = 1, ..., n$" are replaced with "$z_k \in \{0, 1\}, k = 1, ..., n$". For example, using the output oriented Farrell efficiency with an allocation (x^j, y^j) the FDH formulation is given by

$$\begin{aligned}
\widehat{E}_o(x^j, y^j) &\equiv \max_{\theta} \theta && (69)\\
& \quad s.t. \\
\sum_{k=1}^{n} z^k y_m^k &\geq \theta y_m^j, m = 1, ..., M, && (70)\\
\sum_{k=1}^{n} z^k x_l^k &\leq x_l^j, l = 1, ..., N, && (71)\\
\sum_{k=1}^{n} z^k &= 1, && (72)\\
\theta &\geq 0, z_k \in \{0, 1\}, k = 1, ..., n. && (73)
\end{aligned}$$

The value of such formulation is that it hints at the relationship between FDH and DEA-VRS: indeed, the DEA-VRS estimated technology set is simply the *convex closure* of the FDH-estimated technology set. Thus, the FDH estimator can be viewed as a special case of the data envelopment analysis approach since it also envelopes the data but does so without imposing convexity. For historical reasons these names are kept separate to avoid confusion. An alternative yet equivalent form of the FDH estimator can be given via the min-max problem, which is faster to compute (e.g., see Simar and Wilson (2013) for more details).

Finally, stochastic versions of DEA and FDH are also available (e.g., Simar (2007); Simar and Zelenyuk (2011)).

4.3.2 Modeling with Undesirable Outputs or with Congesting Inputs

Another important stream of DEA literature has focused on estimating technologies with weak disposability of inputs or (and especially) outputs, to account for the fact that some outputs are undesirable ('bad') and some inputs can cause congestion.

The ideas for such modeling approaches go back to at least Shephard (1974), and then was elaborated on in Färe and Svensson (1980); Färe and Grosskopf (1983); Grosskopf (1986); Tyteca (1996); Chung et al. (1997), which defined the mainstream approach on this matter. More recently this mainstream approach was re-evaluated in several important works, including Seiford and Zhu (2002), Färe and Grosskopf (2003, 2004, 2009), Førsund (2009), Podinovski and Kuosmanen (2011), Pham and Zelenyuk (2018, 2017).[5]

While many proposals have been made, the most popular approach in this context so far continues to be the mainstream one. In this approach, for example, if one is interested in measuring the radial expansion of the good outputs (g) while having no more inputs (x) and no more of bad outputs (b), then the DEA estimate of *the good-output-oriented Farrell technical efficiency* (under VRS) is given by

[5] Also see Dakpo et al. (2017) and Sueyoshi et al. (2017) for reviews of this research stream.

$$
\begin{aligned}
\widehat{TE}_g(x, g, b) &\equiv \max_{\gamma, z^1, \ldots, z^n, \delta} \gamma && (74)\\
x &\geq \textstyle\sum_{k=1}^n z^k x^k && (75)\\
g\gamma &\leq \delta \textstyle\sum_{k=1}^n z^k g^k, \quad \gamma \geq 1 && (76)\\
b &= \delta \textstyle\sum_{k=1}^n z^k b^k, \quad 0 \leq \delta \leq 1 && (77)\\
\sum_{k=1}^n z^k &= 1, \quad z^k \geq 0, \quad k = 1, ..., n. && (78)
\end{aligned}
$$

For more theoretical and practical (computational) discussions on this topic see Pham and Zelenyuk (2017) and references therein.

4.3.3 Other Streams of DEA

Another stream of DEA focuses on accounting for the network structure of production technologies whether static or dynamic. This stream originated in the seminal works of Färe and Grosskopf (1996); Färe et al. (1996) and was taken further in many other works, e.g., see Kao (2009a,b, 2014) and references therein.

Another important stream of DEA literature is on the topic of weight restrictions in multiplier form of DEA and the classical works here are by Dyson and Thanassoulis (1988); Charnes et al. (1990); Thompson et al. (1990), with more recent and fundamental contributions including new interpretations (as 'technological trade-offs') of various weight restrictions in DEA from Podinovski and Bouzdine-Chameeva (2013), to mention just a few. Also see reviews on this topic by Allen et al. (1997); Podinovski (2015).

Yet another interesting research stream of DEA overlaps with game theory, which also has its roots in the seminal work of von Neumann (1945) and more explicit treatment, for example, in Hao et al. (2000); Nakabayashi and Tone (2006); Liang et al. (2008); Lozano (2012).

4.4 Statistical analysis of DEA and FDH

Another key research wave that brought DEA and FDH to a totally different level is related to their statistical aspects—this wave was mainly influenced by the seminal works of Léopold Simar and many of his co-authors.

The first breakthrough in this area was made by Banker (1993), where the first proof of consistency of the DEA estimator was sketched for the single-output case (in output oriented context), and was pointed out that it belongs to the class of maximum likelihood estimators. This important discovery was then substantially enriched by Korostelev et al. (1995a,b) who proved convergence of the estimated technology to the true technology for both DEA and FDH estimators, and derived convergence rates of these estimators, clarifying that they depend on the dimension of the production model, yet also have some optimality properties under certain conditions.

The convergence properties for the multi-input-multi-output case were first presented in the seminal work of Kneip et al. (1998). Meanwhile, the discovery of the limiting distribution of the DEA estimator was done by Gijbels et al. (1999), only for the 1-input-1-output case and a decade later, Kneip et al. (2008) derived it for the fully multivariate case for DEA with VRS and also proved consistency of various bootstrap procedures. The limiting distribution for the case of DEA with CRS was established by Park et al. (2010) and the limiting distribution of the FDH estimator for the fully multivariate case was also derived by Park et al. (2000), while consistency of the bootstrap for FDH estimator was first presented in Jeong and Simar (2006).

This stream also includes the approach of analyzing the DEA (or FDH) estimated efficiency scores. Perhaps the most popular of these is the so-called 'two-stage DEA', which involves regression analysis of efficiency scores on some factors. The state of the art here is the approach proposed by Simar and Wilson (2007), which is based on truncated regression where the inference is done with the help of a double bootstrap.[6] A non-parametric version of this approach (based on non-parametric truncated regression) was proposed by Park et al. (2008). Furthermore, methods to analyze the distributions of DEA and FDH efficiency scores were explored in Simar and Zelenyuk (2006), while methods to analyze industry efficiency were explored by Simar and Zelenyuk (2007). These approaches were applied in various contexts and industries, e.g., Zelenyuk and Zheka (2006), Demchuk and Zelenyuk (2009), Curi et al. (2015), Chowdhury and Zelenyuk (2016), Du et al. (2018), to mention a few.

A particularly notorious drawback of DEA and FDH—not allowing for noise and sensitivity to 'super-efficient' outliers—was addressed by Simar

[6] Also see Simar and Wilson (2011) for the discussion on caveats and limitations.

(2007); Simar and Zelenyuk (2011), who proposed their version of Stochastic DEA and Stochastic FDH, which consists of two stages: (i) filter the data from the noise using a non-parametric stochastic frontier method[7] and then (ii) use DEA or FDH on the filtered data.

More recently, Kneip et al. (2015) derived new central limit theorems for the context where DEA or FDH estimates are used in place of the true efficiency and thus provided the foundation for many useful statistical tests involving DEA or FDH estimators, including the two-stage 'DEA+regression' context. This foundation was then used by Kneip et al. (2016); Daraio et al. (2017b) to develop various statistical tests and by Simar and Zelenyuk (2017) to develop two new central limit theorems for the aggregate efficiency scores (industry efficiency, etc.) of the type described above and more work continues in this area.

5 SFA Efficiency Software in R

Reproducing the R code needed to implement the methods we have discussed in our chapter is not feasible and thus we provide a short tutorial on how to use a suite of estimators that can easily be accessed via the website "Productivity in R" that can be found as https://sites.google.com/site/productivityinr/. We use standard notations in the codes and let n be the number of firms, T the number of time series and nT the total number of observations, i.e., $nT = n \times T$. The data input files configured with the first column a $nT \times 1$ column vector of the dependent variable y, and the next k containing $nT \times 1$ vectors of the independent variables contained in x'. The convention used is for the first T observations to be for the first cross-section, the second T observations for the second cross-section and so on. The default program output ('results.out' is the default output filename) contains parameter estimates, standard errors, t-values, average technical efficiency, correlation of effects and efficiencies, Spearman rank order correlation of effects and efficiencies, R-squared and adjusted R-squared.

[7]While originally Simar and Zelenyuk (2011) considered the approach of Kumbhakar et al. (2007) for the first stage, one could use other non-parametric SFA approaches, e.g. a more general version proposed by Park et al. (2015) or Simar et al. (2017).

5.1 Basic Model Setup

The default model is the panel data model we introduced in section 3, which we rewrite here:

$$y_{it} = \alpha + x'_{it}\beta - u_{it} + v_{it}, \tag{79}$$

where the global mean (α) is subtracted out in the regression and where this demeaning method is applied to the suite of estimators discussed in SS and CSS as well as for the Battese and Coelli (1992) (BC) estimator. Time trends can be added as well to account for disembodied technical change that is available for adoption by all firms and various right-hand-side (rhs) variables that may be correlated with the effects can be identified in the main R file. Efficiencies can also be averaged over different estimators using model averaging weights based on a simple average (`AVE=0`), AIC weights (`AVE=1`), or BIC weights (`AVE=2`). Outliers can be addressed by trimming (this does not apply to the `BC` and the `DEA` estimators).

5.2 Figures and Tables

Figures and tables are selected by setting values of 1 to print and 0 to skip the results. There are a number of figures that are already set up to print. For example, when `fig1=1` a figure for the average of efficiencies of the time-variant estimators is printed. When `fig2=1` one for the efficiencies from the time-variant estimators is printed. When `fig3=1` a figure for the average of efficiencies from all estimators is printed, while `fig4=1` the weighted average of the efficiencies is printed. `tab1=1` prints a table for the average of efficiencies from the time-variant estimators, `tab2=1` prints one for the efficiencies from the time-variant estimators are printed and `tab3=1` prints a table of the individual effects from all estimators. Also, setting `firmeff=1` saves the efficiencies of each individual firm for each estimator utilized.

5.3 Different Estimators

The R codes for the SFA models include the fixed effect, random effects, and the Hausman-Taylor (`HT`) version of the stochastic panel frontier models of Schmidt and Sickles (1984). The `FR` and the `HT` global parameters are set to determine which model is estimated. Since the `HT` estimator of the panel stochastic frontier allows for selected regressors to be correlated with the efficiency effects term the global option `k1` is set at the number of variables in X not correlated with the efficiency effects and are loaded into the first k1 columns for the regressors. These different models are referred to as `FIX`, `RND`, and `HT`. The global options `PSS1`, `PSS2`, `PSS3` designate the estimators for the Park, Sickles and Simar (1998, 2003, 2007) models. The value of 1 prints the results and 0 skips the results. Bandwidth selection is based on leave-one-out least squares cross-validation. The `CSS` global option allows for the estimates of the Cornwell et al. (1990) models to be generated. A global option of `CSS=1` prints results of the Fixed Effects/Within estimator (`CSSW`) estimator, while `CSS=2` generated the GLS random effects estimator. The efficient IV estimator (analogous to the HT estimator for the SS Model) is engaged with `CSS=3`. Finally, `CSS=4` prints out all four model results. Time invariant variables should be placed at the end of X. The parameter `zp` is set equal to the number of time invariant variables. The Kneip, Sickles and Song (2012) factor-model is called by setting the global parameter `KSS=1.` The `KSS` estimator also finds the number of factors and the code varies the number of factors from L_{max} to L_{min} and finds the first highest number at which the dimensionality test is not rejected. Cross-validation is used to estimate the optimal smoothing parameter. `gr_st` is the starting point of the grid search, `gr_in` is the increment in the grid search, and `gr_en` is the end point of the grid search. The Battese and Coelli (1992) model is called by setting the `BC` option=1. The bounded inefficiency model of Almanidis, Qian, and Sickles (2014) is called by setting the `BIE` option=1. Different distributions for the lower bound on inefficiency can also be specified using the `bie_dist` parameter. For a `bie_dist=0` the truncated is specified, while `bie_dist=1` uses the truncated half normal, and `bie_dist=2` the doubly truncated normal distribution. Finally, the Kutlu (2018) endogeneity correction for selected regressors (e.g., input levels in the production frontier) based on the CSS estimator is also available on this website along with instructions for its use.

6 DEA Efficiency Software in R

The Jeon and Sickles (2004) model is the directional distance function method outlined in the section on DEA estimators. This estimator also allows us to address the presence of undesirable outputs. Jeon and Sickles (JS) used this DEA estimator to examine the productivity effects of controlling for carbon dioxide emissions on productivity growth using Mamlquist indexes.

The JS estimator uses OECD data while Efficiency software uses UNIDO and Bank data. The JS R codes are in a separate Jeon_Sickles_2004.zip file on the website. The default results that are printed by the R codes are productivity growth, efficiency change, and technology change and confidence intervals for the growth decompositions are based on the Simar and Wilson (2007) bootstrapping method. Next, we have the Simar and Zelenyuk (2006) model, which implements the Li (1996) test in the context of comparing distributions of efficiency estimated via DEA and the Simar and Zelenyuk (2007) model that constructs confidence intervals and bias corrections for DEA-estimated aggregate efficiencies of a set of firms. It also provides a test for the comparison of these group efficiencies.

There are other open source R code platforms other than the one that we have focused on in our discussions so far. The *Benchmarking* package in R prepared by Peter Bogetoft and Lars Otto is one such program and the manual and other directions for its use can be found at https://cran.r-project.org/web/packages/Benchmarking/Benchmarking.pdf. This can be used to estimate both DEA and FDH models. There have been many other programs developed by Simar and his colleagues in Matlab and these can be compiled into R using Matlab-to-R converters. One such package, the 'matconv' package by Siddarta Jairam is a useful conversion tool, although some testing for accuracy is always advised for such automatic translations.[8] Many of these Matlab programs can be found at https://sites.google.com/site/productivityefficiency.[9] Finally, Daraio et al. (2017a) provide a survey on a variety of software platforms that can estimate SFA and DEA models, including a number with R coding.

7 Summary and Final Remarks

This chapter has discussed a number of statistical and programming techniques used to evaluate the production efficiency of economic units, whether they be firms, sectors, or countries. We have discussed the methods most widely used for these sorts of evaluations and have provided references and

[8] See https://cran.r-project.org/web/packages/matconv/matconv.pdf for more details.

[9] Arne Henningsen has provided the R conversion to a number of Fortran-based programs (FRONTIER) developed by Tim Coelli as well as a set of new programs in his *A Package for Stochastic Frontier Analysis (SFA) in R*. This platform is available at https://cran.r-project.org/.

url's to the most up to date websites that provide the R code to implement these methods. We trust that the interested reader will find our discussions and the software helpful for the purposes of practical evaluations of the performance of business entities as well as in their academic research. We have also provided an up to date set of references that productivity and efficiency researchers can use for extended and deeper readings on these widely used and adopted performance benchmarkingu; methods.

References

Adams, R. M., Berger, A. N., Sickles, R. C., 1999. Semiparametric approaches to stochastic panel frontiers with applications in the banking industry. Journal of Business and Economic Statistics 17 (3), 349–358.

Afriat, S. N., 1972. Efficiency estimation of production functions. International Economic Review 13 (3), 568–598.

Agrell, P. J., Bogetoft, P., et al., 2017. Regulatory benchmarking: Models, analyses and applications. Data Envelopment Analysis Journal 3 (1–2), 49–91.

Ahn, S. C., Lee, Y. H., Schmidt, P., 2007. Stochastic frontier models with multiple time-varying individual effects. Journal of Productivity Analysis 27 (1), 1–12.

Ahn, S. C., Lee, Y. H., Schmidt, P., 2013. Panel data models with multiple time-varying individual effects. Journal of Econometrics 174 (1), 1–14.

Aigner, D., Lovell, C., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6 (1), 21–37.

Allais, M., 1943. Traitè D'Èconomie Pure. Vol. 3. Paris, FR: Imprimerie Nationale.

Allen, R., Athanassopoulos, A., Dyson, R. G., Thanassoulis, E., 1997. Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. Annals of Operations Research 73, 13–34.

Almanidis, P., Qian, J., Sickles, R. C., 2014. Stochastic frontier with bounded inefficiency. In: Sickles, R. C., Horrace, W. C. (Eds.), Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications. New York, NY: Springer, pp. 47–82.

Almanidis, P., Sickles, R. C., 2012. The skewness problem in stochastic frontier models: Fact or fiction? In: Van Keilegom, I., Wilson, P. W. (Eds.), Exploring Research Frontiers in Contemporary Statistics and Econometrics: A Festchrift in Honor of Léopold Simar. New York, NY: Springer, pp. 201–227.

Bada, O., Liebl, D., 2014. The r-package phtt: Panel data analysis with heterogeneous time trends. Journal of Statistical Software 59 (6), 1–33.

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77 (4), 1229–1279.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70(1), 191–221.

Banker, R. D., 1993. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. Management Science 39 (10), 1265–1273.

Banker, R. D., Charnes, A., Cooper, W. W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30 (9), 1078–1092.

Battese, G., Coelli, T., 1992. Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. Journal of Productivity Analysis 3 (1–2), 153–169.

Battese, G. E., Coelli, T. J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical economics 20 (2), 325–332.

Bauer, P. W., Hancock, D., 1993. The efficiency of the Federal Reserve in providing check processing services. Journal of Banking & Finance 17 (2-3), 287–311.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., Roberts, J., 2013. Does management matter? Evidence from India. The Quarterly Journal of Economics 128 (1), 1–51.

Bloom, N., Genakos, C., Sadun, R., Van Reenen, J., 2012. Management practices across firms and countries. The Academy of Management Perspectives 26 (1), 12–33.

Bogetoft, P., 1996. DEA on relaxed convexity assumptions. Management Science 42 (3), 457–465.

Bogetoft, P., 2013. Performance benchmarking: Measuring and managing performance. Springer Science & Business Media.

Chambers, R., Färe, R., Grosskopf, S., 1996. Productivity growth in APEC countries. Pacific Economic Review 1 (3), 181–190.

Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2 (6), 429–444.

Charnes, A., Cooper, W. W., Huang, Z. M., Sun, D. B., 1990. Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. Journal of econometrics 46 (1-2), 73–91.

Chowdhury, H., Zelenyuk, V., 2016. Performance of hospital services in Ontario: DEA with truncated regression approach. Omega 63, 111–122.

Chung, Y. H., Färe, R., Grosskopf, S., 1997. Productivity and undesirable outputs: A directional distance function approach. Journal of Environmental Management 51 (3), 229–240.

Colombi, R., Kumbhakar, S. C., Martini, G., Vittadini, G., 2014. Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. Journal of Productivity Analysis 42 (2), 123–136.

Colombi, R., Martini, G., Vittadini, G., 2011. A stochastic frontier model with short-run and long-run inefficiency random effects, university of Bergamo Department of Economics and Technology Management Working Paper No. 012011.

Cornwell, C., Schmidt, P., Sickles, R. C., 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of Econometrics 46 (1), 185–200.

Curi, C., Lozano-Vivas, A., Zelenyuk, V., 2015. Foreign bank diversification and efficiency prior to and during the financial crisis: Does one business model fit all? Journal of Banking and Finance 61 (S1), S22–S35.

Dakpo, K. H., Jeanneaux, P., Latruffe, L., 2017. Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits and future prospects in the nonparametric framework. European Journal of Operational Research 250 (2), 347–359.

Daraio, C., Kerstens, K. H., Nepomuceno, T. C. C., Sickles, R., 2017a. Productivity and efficiency analysis software: An exploratory bibliographical survey of the options. Journal of Economic Surveys.

Daraio, C., Simar, L., Wilson, P. W., 2017b. Central limit theorems for conditional efficiency measures and tests of the "separability" condition in nonparametric, two-stage models of production. The Econometrics Journal.

Debreu, G., 1951. The coefficient of resource utilization. Econometrica 19 (3), 273–292.

Demchuk, P., Zelenyuk, V., 2009. Testing differences in efficiency of regions within a country: The case of Ukraine. Journal of Productivity Analysis 32 (2), 81–102.

Deprins, D., Simar, L., Tulkens, H., 1984. Measuring labour efficiency in post offices. In: Marchand, M., Pestieau, P., Tulkens, H. (Eds.), The Performance of Public Enterprises: Concepts and Measurement. Amsterdam, NL: Springer, pp. 243–267.

DeYoung, R., 1998. Management quality and x-inefficiency in national banks. Journal of Financial Services Research 13 (1), 5–22.

Diewert, W. E., 1983. The measurement of waste within the production sector of an open economy. The Scandinavian Journal of Economics 85 (2), 159–179.

Du, K., Worthington, A. C., Zelenyuk, V., 2018. Data envelopment analysis, truncated regression and double-bootstrap for panel data with application to Chinese banking. European Journal of Operational Research 265 (2), 748 – 764.

Dyson, R. G., Thanassoulis, E., 1988. Reducing weight flexibility in data envelopment analysis. Journal of the Operational Research Society 39 (6), 563–576.

Färe, R., Grosskopf, S., 1983. Measuring congestion in production. Zeitschrift für Nationalökonomie / Journal of Economics 43 (3), 257–271.

Färe, R., Grosskopf, S., 1996. Intertemporal Production Frontiers: With Dynamic DEA. Norwell, MA: Kluwer Academic Publishers.

Färe, R., Grosskopf, S., 2003. Nonparametric productivity analysis with undesirable outputs: Comment. American Journal of Agricultural Economics 85 (4), 1070–1074.

Färe, R., Grosskopf, S., 2004. Modeling undesirable factors in efficiency evaluation: Comment. European Journal of Operational Research 157 (1), 242–245.

Färe, R., Grosskopf, S., 2009. A comment on weak disposability in nonparametric production analysis. American Journal of Agricultural Economics 91 (2), 535–538.

Färe, R., Grosskopf, S., Logan, J., 1983. The relative efficiency of Illinois electric utilities. Resources and Energy 5 (4), 349–367.

Färe, R., Grosskopf, S., Roos, P., 1996. On two definitions of productivity. Economics Letters 53 (3), 269–274.

Färe, R., Grosskopf, S., Zelenyuk, V., 2007. Finding common ground: Efficiency indices. In: Färe, R., Grosskopf, S., Primont, D. (Eds.), Aggregation, Efficiency and Measurement. Boston, MA: Springer, pp. 83–95.

Färe, R., Lovell, C. K., 1978. Measuring the technical efficiency of production. Journal of Economic Theory 19 (1), 150–162.

Färe, R., Svensson, L., 1980. Congestion of production factors. Econometrica: Journal of the Econometric Society 48 (7), 1745–1753.

Farrell, M. J., 1957. The measurement of productive efficiency. Jornal of the Royal Statistical Society. Series A (General) 120 (3), 253–290.

Førsund, F. R., 2009. Good modelling of bad outputs: Pollution and multiple-output production. International Review of Environmental and Resource Economics 3 (1), 1–38.

Foster, L., Haltiwanger, J., Syverson, C., 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? American Economic Review 98 (1), 394–425.

Frantz, R. S., 1997. X-Efficiency: Theory, Evidence and Applications, 2nd Edition. Norwell, MA: Kluwer Academic Publishers.

Frantz, R. S., 2007. Empirical evidence on X-efficiency, 1967-2004. In: Frantz, R. (Ed.), Renaissance in Behavioral Economics: Essays in Honour of Harvey Leibenstein. New York, NY: Routledge, pp. 221–227.

Fu, X. M., Heffernan, S., 2009. The effects of reform on China's bank structure and performance. Journal of Banking & Finance 33 (1), 39–52.

Gijbels, I., Mammen, E., Park, B. U., Simar, L., 1999. On estimation of monotone and concave frontier functions. Journal of the American Statistical Association 94 (445), 220–228.

Glaister, K. W., 2014. The contribution of management to economic growth: A review. Prometheus 32 (3), 227–244.

Greene, W. H., 1980a. Maximum likelihood estimation of econometric frontier functions. Journal of Econometrics 13 (1), 27–56.

Greene, W. H., 1980b. On the estimation of a flexible frontier production model. Journal of Econometrics 13 (1), 101–115.

Greene, W. H., 2005a. Fixed and random effects in stochastic frontier models. Journal of Productivity Analysis 23 (1), 7–32.

Greene, W. H., 2005b. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126 (2), 269–303.

Grifell-Tatjé, E., Lovell, C. A. K., 2013. Advances in cost frontier analysis of the firm. In: Thomas, C. R., Shughart, W. (Eds.), The Oxford Handbook of Managerial Economics. London, UK: Oxford University Press, pp. 1–18.

Grifell-Tatjé, E., Lovell, C. A. K., 2015. Productivity Accounting. New York, NY: Cambridge University Press.

Grifell-Tatjé, E., Lovell, C. K., Sickles, R. C., 2018. The Oxford Handbook of Productivity Analysis. Oxford University Press.

Griffin, J. E., Steel, M. F., 2007. Bayesian stochastic frontier analysis using winbugs. Journal of Productivity Analysis 27 (3), 163–176.

Grosskopf, S., 1986. The role of the reference technology in measuring productive efficiency. The Economic Journal 96 (382), 499–513.

Hall, R. E., Jones, C. I., 1999. Why do some countries produce so much more output per worker than others? The Quarterly Journal of Economics 114 (1), 83–116.

Hao, G., Wei, Q. L., Yan, H., 2000. A game theoretical model of DEA efficiency. The Journal of the Operational Research Society 51 (11), 1319–1329.

Hsieh, C.-T., Klenow, P. J., 2009. Misallocation and manufacturing TFP in China and India. The Quarterly Journal of Economics 124 (4), 1403–1448.

Jeon, B. M., Sickles, R. C., 2004. The role of environmental factors in growth accounting. Journal of Applied Econometrics 19 (5), 567–591.

Jeong, S.-O., Simar, L., 2006. Linearly interpolated FDH efficiency score for nonconvex frontiers. Journal of Multivariate Analysis 97 (10), 2141–2161.

Jiang, C., Yao, S., Zhang, Z., 2009. The effects of governance changes on bank efficiency in China: A stochastic distance function approach. China Economic Review 20 (4), 717–731.

Jondrow, J., Lovell, C. A. K., Materov, I. S., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19 (2–3), 233–238.

Kao, C., 2009a. Efficiency decomposition in network data envelopment analysis: A relational model. European Journal of Operational Research 192 (3), 949–962.

Kao, C., 2009b. Efficiency measurement for parallel production systems. European Journal of Operational Research 196 (3), 1107–1112.

Kao, C., 2014. Network data envelopment analysis: A review. European Journal of Operational Research 239 (1), 1–16.

Kim, M., Schmidt, P., 2008. Valid tests of whether technical inefficiency depends on firm characteristics. Journal of Econometrics 144 (2), 409–427.

Kneip, A., Park, B. U., Simar, L., 1998. A note on the convergence of nonparametric DEA estimators for production efficiency scores. Econometric Theory 14 (6), 783–793.

Kneip, A., Sickles, R. C., Song, W., 2012. A new panel data treatment for heterogeneity in time trends. Econometric Theory 28 (3), 590–628.

Kneip, A., Simar, L., Wilson, P. W., 2008. Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. Econometric Theory 24 (6), 1663–1697.

Kneip, A., Simar, L., Wilson, P. W., 2015. When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores. Econometric Theory 31 (2), 394–422.

Kneip, A., Simar, L., Wilson, P. W., 2016. Testing hypotheses in nonparametric models of production. Journal of Business & Economic Statistics 34 (3), 435–456.

Koopmans, T., 1951a. Activity Analysis of Production and Allocation. New York, NY: Wiley.

Koopmans, T. C., 1951b. Analysis of production as an efficienct combination of activities. Activity Analysis of Production and Allocation 13, 33–37.

Korostelev, A., Simar, L., Tsybakov, A. B., 1995a. Efficient estimation of monotone boundaries. Annals of Statistics 23 (2), 476–489.

Korostelev, A., Simar, L., Tsybakov, A. B., 1995b. On estimation of monotone and convex boundaries. Publications de l'Institut de Statistique des Universités de Paris 39 (1), 3–18.

Kumbhakar, S., Schmidt, P., 2016. Endogeneity Problems in Econometrics, Special Issue of the Journal of Econometrics. Vol. 190. Amsterdam, Netherlands: Elsevier.

Kumbhakar, S. C., 1990. Production frontiers, panel data, and time-varying technical inefficiency. Journal of Econometrics 46 (1), 201–211.

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

Kutlu, L., 2018. A distribution-free stochastic frontier model with endogenous regressors. Economics Letters 163, 152–154.

Kwan, S. H., 2006. The X-efficiency of commercial banks in Hong Kong. Journal of Banking & Finance 30 (4), 1127–1147.

Lee, Y. H., 1991. Panel data models with multiplicative individual and time effects: Applications to compensation and frontier production functions. Ph.D. thesis, Michigan State University, East Lansing, MI, USA.

Lee, Y. H., Schmidt, P., 1993. A production frontier model with flexible temporal variation in technical efficiency. In: Fried, H. O., Lovell, C. A. K., Schmidt, S. S. (Eds.), The Measurement of Productive Efficiency: Techniques and Applications. New York, NY: Oxford University Press, pp. 237–255.

Leibenstein, H., 1966. Allocative efficiency vs. "X-efficiency". American Economic Review 56 (3), 392–415.

Leibenstein, H., 1975. Aspects of the X-efficiency theory of the firm. The Bell Journal of Economics 6 (2), 580–606.

Leibenstein, H., 1987. Inside the Firm: The Inefficiencies of Hierarchy. Cambridge, MA: Harvard University Press.

Li, Q., 1996. Nonparametric testing of closeness between two unknown distribution functions. Econometric Reviews 15 (3), 261–274.

Liang, L., Wu, J., Cook, W. D., Zhu, J., 2008. The DEA game cross-efficiency model and its Nash equilibrium. Operations Research 56 (5), 1278–1288.

Lieberman, M. B., Lau, L. J., Williams, M. D., 1990. Firm-level productivity and management influence: A comparison of US and Japanese automobile producers. Management Science 36 (10), 1193–1215.

Liu, J., Sickles, R. C., Tsionas, E. G., 2017. Bayesian treatments to panel data models with time-varying heterogeneity. Econometrics 5 (33), 1–21.

Lozano, S., 2012. Information sharing in DEA: A cooperative game theory approach. European Journal of Operational Research 222 (3), 558–565.

Luenberger, D. G., 1992. Benefit functions and duality. Journal of Mathematical Economics 21 (5), 461–481.

Materov, I. S., 1981. On full identification of the stochastic production frontier model. Ekonomika i Matematicheskie Metody 17, 784–788.

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18 (2), 435–44.

Mester, L. J., 1993. Efficiency in the savings and loan industry. Journal of Banking & Finance 17 (2–3), 267–286.

Nakabayashi, K., Tone, K., 2006. Egoist's dilemma: a DEA game. Omega 34 (2), 135–148.

Nallari, R., Bayraktar, N., 2010. Micro efficiency and macro growth, the World Bank Policy Research Working Paper No. 5267.

Olson, J. A., Schmidt, P., Waldman, D. M., 1980. A Monte Carlo study of estimators of the stochastic frontier production function. Journal of Econometrics 13 (1), 67–82.

Park, B. U., Jeong, S.-O., Simar, L., 2010. Asymptotic distribution of conical-hull estimators of directional edges. Annals of Statistics 38 (3), 1320–1340.

Park, B. U., Sickles, R. C., Simar, L., 1998. Stochastic panel frontiers: A semiparametric approach. Journal of Econometrics 84 (2), 273–301.

Park, B. U., Sickles, R. C., Simar, L., 2003. Semiparametric-efficient estimation of AR(1) panel data models. Journal of Econometrics 117 (2), 279–309.

Park, B. U., Sickles, R. C., Simar, L., 2007. Semiparametric efficient estimation of dynamic panel data models. Journal of Econometrics 136 (1), 281–301.

Park, B. U., Simar, L., Weiner, C., 2000. The FDH estimator for productivity efficiency scores: Asymptotic properties. Econometric Theory 16 (6), 855–877.

Park, B. U., Simar, L., Zelenyuk, V., 2008. Local likelihood estimation of truncated regression and its partial derivatives: Theory and application. Journal of Econometrics 146 (1), 185–198.

Park, B. U., Simar, L., Zelenyuk, V., 2015. Categorical data in local maximum likelihood: Theory and applications to productivity analysis. Journal of Productivity Analysis 43 (2), 199–214.

Petersen, N. C., 1990. Data envelopment analysis on a relaxed set of assumptions. Management Science 36 (3), 305–314.

Pham, M. D., Zelenyuk, V., 2017. Convexity, disposability and returns to scale in production analysis, cEPA Working Paper No. WP04/2017.

Pham, M. D., Zelenyuk, V., 2018. Slack-based directional distance function in the presence of bad outputs: Theory and application to Vietnamese banking. Empirical Economics 54 (1), 153–187.

Pitt, M. M., Lee, L.-F., 1981. The measurement and sources of technical inefficiency in the Indonesian weaving industry. Journal of Development Economics 9 (1), 43–64.

Podinovski, V. V., 2015. Dea models with production trade-offs and weight restrictions. In: Zhu, J. (Ed.),). Data Envelopment Analysis: A Handbook of Models and Methods. New York, NY: Springer, pp. 105–144.

Podinovski, V. V., Bouzdine-Chameeva, T., 2013. Weight restrictions and free production in data envelopment analysis. Operations Research 61 (2), 426–437.

Podinovski, V. V., Kuosmanen, T., 2011. Modelling weak disposability in data envelopment analysis under relaxed convexity assumptions. European Journal of Operational Research 211 (3), 577–585.

Qian, J., Sickles, R. C., 2008. Stochastic frontiers with bounded inefficiency, unpublished manuscript, Shanghai Jiao Tong University, Shanghai, CN.

Rezvanian, R., Ariss, R. T., Mehdian, S. M., 2011. Cost efficiency, technological progress and productivity growth of Chinese banking pre-and post-WTO accession. Applied Financial Economics 21 (7), 437–454.

Schmidt, P., Sickles, R. C., 1984. Production frontiers and panel data. Journal of Business & Economic Statistics 2 (4), 367–374.

Seiford, L. M., Zhu, J., 2002. Modeling undesirable factors in efficiency evaluation. European Journal of Operational Research 142 (1), 16–20.

Shephard, R. W., 1953. Cost and Production Functions. Princeton, NJ: Princeton University Press.

Shephard, R. W., 1970. Theory of Cost and Production Functions. Princeton studies in mathematical economics. Princeton, NJ: Princeton University Press.

Shephard, R. W., 1974. Indirect production functions. In: Mathematical Systems in Economics. Vol. 10. Hain, Meisenheim am Glan.

Sickles, R., Zelenyuk, V., 2018. Measurement of Productivity and Efficiency: Theory and Practice. New York, NY: Cambridge University Press, forthcoming.

Simar, L., 2007. How to improve the performances of DEA/FDH estimators in the presence of noise? Journal of Productivity Analysis 28 (3), 183–201.

Simar, L., Van Keilegom, I., Zelenyuk, V., 2017. Nonparametric least squares methods for stochastic frontier models. Journal of Productivity Analysis 47 (3), 189–204.

Simar, L., Wilson, P. W., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics 136 (1), 31–64.

Simar, L., Wilson, P. W., 2008. Statistical inference in nonparametric frontier models: Recent developments and perspectives. In: Harold Fried, C. A. K. L., Schmidt, S. (Eds.), The Measurement of Productive Efficiency, 2nd Edition. New York, NY: Oxford University Press, pp. 421–521.

Simar, L., Wilson, P. W., 2011. Two-stage DEA: Caveat emptor. Journal of Productivity Analysis 36 (2), 205–218.

Simar, L., Wilson, P. W., 2013. Estimation and inference in nonparametric frontier models: Recent developments and perspectives. Foundations and Trends in Econometrics 5 (3–4), 183–337.

Simar, L., Zelenyuk, V., 2006. On testing equality of distributions of technical efficiency scores. Econometric Reviews 25 (4), 497–522.

Simar, L., Zelenyuk, V., 2007. Statistical inference for aggregates of Farrell-type efficiencies. Journal of Applied Econometrics 22 (7), 1367–1394.

Simar, L., Zelenyuk, V., 2011. Stochastic FDH/DEA estimators for frontier analysis. Journal of Productivity Analysis 36 (1), 1–20.

Simar, L., Zelenyuk, V., 2017. Central limit theorems for aggregate efficiency. Operations Research, Forthcoming.

Stevenson, R. E., 1980. Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13 (1), 57–66.

Sueyoshi, T., Yuan, Y., Goto, M., 2017. A literature study for DEA applied to energy and environment. Energy Economics 62, 104–124.

Thompson, R. G., Langemeier, L. N., Lee, C.-T., Lee, E., Thrall, R. M., 1990. The role of multiplier bounds in efficiency analysis with application to Kansas farming. Journal of Econometrics 46 (1-2), 93–108.

Tsionas, E. G., Kumbhakar, S. C., 2014. Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random-effects model. Journal of Applied Econometrics 29 (1), 110–132.

Tsionas, E. G., Papadakis, E. N., 2010. A bayesian approach to statistical inference in stochastic dea. Omega 38 (5), 309–314.

Tyteca, D., 1996. On the measurement of the environmental performance of firms—a literature review and a productive efficiency perspective. Journal of Environmental Management 46 (3), 281–308.

von Neumann, J., 1945. A model of general equilibrium. Review of Economic Studies 13 (1), 1–9.

Wang, H.-J., Schmidt, P., 2002. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. Journal of Productivity Analysis 18 (2), 129–144.

Yao, S., Han, Z., Feng, G., 2008. Ownership reform, foreign competition and efficiency of Chinese commercial banks: A non-parametric approach. The World Economy 31 (10), 1310–1326.

Zelenyuk, V., Zheka, V., 2006. Corporate governance and firm's efficiency: The case of a transitional country, Ukraine. Journal of Productivity Analysis 25 (1), 143–157.