Performance estimation when the distribution of inefficiency is unknown

Mike G. Tsionas*

Abstract

We show how to compute inefficiency or performance scores when the distribution of the one-sided error component in Stochastic Frontier Models (SFMs) is unknown; and we do the same with Data Envelopment Analysis (DEA). Our procedure, which is based on the Fast Fourier Transform (FFT), utilizes the empirical characteristic function of the residuals in SFMs or efficiency scores in DEA. The new techniques perform well in Monte Carlo experiments and deliver reasonable results in an empirical application to large U.S. banks. In both cases, deconvolution of DEA scores with the FFT brings the results much closer to the inefficiency estimates from SFM.

Key Words: Productivity and Competitiveness; Stochastic Frontier Models; Data Envelopment Analysis; Fast Fourier Transform; Empirical Characteristic Function.

Acknowledgments: The author is grateful to three anonymous reviewers for many useful comments on an earlier version.

*Lancaster University Management School, LA1 4YX, U.K., m.tsionas@lancaster.ac.uk

1 Introduction

Stochastic Frontier Models (SFMs; Kumbhakar and Lovell, 2000, also known as Stochastic Frontier Analysis, SFA) as well as Data Envelopment Analysis (DEA, Charnes et al., 1978; Dyson et al., 2001) are essential in the measurement of performance and efficiency. However, SFMs rely on distributional assumptions about the two-sided error term that represents noise, and the one-sided error component that represents technical inefficiency. Although the normality assumption for the two-sided error is reasonable, specific distributional assumptions for the one-sided error component are mostly hard to defend. Such assumptions usually involve the half-normal, the truncated-normal, the exponential, less commonly the gamma distribution, etc. See also Adams and Sickles (2007).

For DEA, it is widely recognized that inefficiency scores incorporate random noise of measurement error and they are likely to overestimate inefficiency; see Johnson and Kuosmanen (2012), Cordero et al. (2015), Oh and Shin (2015), Holland and Lee (2002), Wanke et al. (2015), and Tsionas (2003). Since such distributional assumptions for the one-sided error in SFMs are more or less arbitrary and subject to researcher's preference, it would be important to develop a procedure that does not depend on such assumptions about the distribution. In this paper, we use the convolution of the two-sided (normal) and one-sided components to recover efficiency or performance estimates without specific assumptions on the one-sided error component. Specifically, we construct the empirical characteristic function of the composed error term, we recover the distribution of the one-sided disturbance through the Fourier transform, and we compute performance estimates using numerical integration. Essential to the numerical efficiency of this approach is the use of the Fast Fourier Transform (FFT).

The same method is extended to the case of DEA under the assumption that DEA scores can be decomposed into noise and pure inefficiency. Therefore, we "adjust" DEA scores for the presence of noise using the FFT deconvolution of DEA scores into measurement error and (in)efficiency. Other methods that address the problem of measurement error, include the Bayesian approach of Tsionas (2019), and the smooth monotone concave approximation of Tsionas and Izzeldin (2018), which is based on Allon, Beenstock, Hackman, Passy, and Shapiro (2007), Keshvari and Kuosmanen (2013) and Kuosmanen and Kortelainen (2012) and Lee et al. (2013). Of course, the presence of noise in obtaining performance scores from DEA is an important issue and much of the literature is devoted to addressing the issues from measurement problems (see, for example, Andor et al., 2019). In SFM, the main problem is to avoid assuming specific distribution for the one-sided error component; and for DEA the main issue is to perform deconvolution of (in)efficiency scores from the possible presence of measurement error without explicitly

specifying the way in which measurement error affects inputs or outputs. This is important because different ways of introducing measurement error in DEA problems may yield different efficiency scores; and choosing between different DEA models is an under-explored area. So, to summarize, in DEA our problem is to deconvolve "signal" (efficiency or inefficiency scores) from noise without specifically assuming how measurement error interacts with inputs and outputs (which can give rise to a whole constellation of possible models that would have to be considered). In SFM, our problem is that of obtaining (in)efficiency estimates (under normality, say, of the two-sided error component) without distributional assumptions about the one-sided error term. The new techniques are investigated using Monte Carlo simulations as well as in the context of panel data from large U.S. banks.

2 Model

Suppose $x_{it} \in \mathbb{R}^K$ is a vector of explanatory variables (say log input prices and log outputs), y_{it} is the dependent variable (typically log cost), and the model is

$$y_{it} = \alpha_i + x'_{it}\beta + v_{it} + u_{it},\ 1 \leq i \leq I, 1 \leq t \leq T, \tag{1}$$

where $v_{it} \sim i.i.d\mathcal{N}(0, \sigma_v^2)$, α_i represents firm effects to capture heterogeneity, u_{it} is a non-negative error component representing technical inefficiency, I is the number of decision-making units or firms, and T is the number of time periods. Alternative, we can define a production frontier of the form

$$y_{it} = \alpha_i + x'_{it}\beta + v_{it} - u_{it},\ 1 \leq i \leq I, 1 \leq t \leq T, \tag{2}$$

where x_{it} is the vector of log inputs, y_{it} is log output and the interpretation of v_{it} and u_{it} is the same as in (1). Typical assumptions about u_{it} include the exponential, half-normal, truncated-normal, gamma, etc. distributions. We make no specific distributional assumption about u_{it} other than that these error components are i.i.d and independent of the regressors. If the composed error term is denoted $\varepsilon_{it} = v_{it} + u_{it}$, then the characteristic function of ε_{it} is

$$\varphi_\varepsilon(\tau) = \mathbb{E}\left(e^{\iota\tau\varepsilon}\right) = \int_{-\infty}^{\infty} e^{\iota\tau\varepsilon} f_\varepsilon(\varepsilon)\,\mathrm{d}\varepsilon,$$

where $\mathbb{E}(\cdot)$ denotes expectation (with respect to ε in this instance whose density is denoted $f_\varepsilon(\varepsilon)$) and $\iota = \sqrt{-1}$. By standard properties of the characteristic function we have:

$$\varphi_\varepsilon(\tau) = \varphi_v(\tau) \cdot \varphi_u(\tau) = e^{-(\sigma_v^2/2)\tau^2} \varphi_u(\tau) \, \forall \tau \in \mathbb{R}. \tag{3}$$

In the case of production frontiers where the composed error is $\varepsilon = v - u$ the formula is modified to

$$\varphi_\varepsilon(\tau) = \varphi_v(\tau) \cdot \varphi_u(-\tau) = e^{-(\sigma_v^2/2)\tau^2} \varphi_u(-\tau) \, \forall \tau \in \mathbb{R}. \tag{4}$$

To recover the density from the characteristic function we have

$$f_u(u) = \tfrac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\iota u \tau} \varphi_u(\tau) \, \mathrm{d}\tau. \tag{5}$$

Given residuals[1] $\hat{\varepsilon}_{it}$ the empirical characteristic function is

$$\varphi_{\hat{\varepsilon}_{it}}(\tau) = (IT)^{-1} \sum_{i=1}^{I} \sum_{t=1}^{T} e^{\iota \tau \hat{\varepsilon}_{it}} \, \forall \tau \in \mathbb{R}. \tag{6}$$

Therefore, the characteristic function of u_{it} is given by

$$\varphi_{u_{it}}(\tau) = e^{-\frac{1}{2}\sigma_v^2 \tau^2} \varphi_{\hat{\varepsilon}_{it}}(\tau) \, \forall \tau \in \mathbb{R}. \tag{7}$$

The characteristic functions corresponding to several one-sided error distributions are presented in (A.6) of Appendix A. Taking natural logs we have

$$\ln \varphi_{\hat{\varepsilon}_{it}}(\tau) = -\tfrac{\sigma_v^2 \tau^2}{2} + \ln \varphi_{u_{it}}(\tau) \, \forall \tau \in \mathbb{R}. \tag{8}$$

Given $\varphi_{\hat{\varepsilon}_{it}}(\tau)$, σ_v^2 can be estimated as the coefficient of $-\tau_j^2/2$ in the following regression-like equation, see (A.5) in Appendix A:

$$\ln \varphi_{\hat{\varepsilon}_{it}}(\tau_j) = -\tfrac{\sigma_v^2 \tau_j^2}{2} + \ln \varphi_{u_{it}}(\tau_j) + \xi_j, \; j = 1, \ldots, J, \tag{9}$$

[1] These residuals can be obtained using various estimation techniques for (1) like least-squares-dummy-variables (LSDV), the random-effects formulation, the first-differencing approach, methods that can deal with endogeneity of the regressors, etc.

where $\{\tau_j, j = 1, \ldots, J\}$ is a set of J points around zero, and ξ_j is an error term. Of course, σ_v^2 must be restricted to be non-negative. Since the values of the dependent variable and $\ln \varphi_{u_{it}}(\tau_j)$ can be complex, we consider stacking the real and imaginary parts in (9) (Feuerverger and McDunnough, 1981a,b; see also Koutrouvelis, 1980, 1981). As an example, for the half-normal distribution we have

$$\varphi_u(\tau) = 2e^{-\frac{1}{2}\sigma_u^2\tau^2}\Phi\left(\iota\sigma_u\tau\right), \tag{10}$$

where $\Phi(\cdot)$ is the standard normal distribution function, see equation (A.6) in Appendix A. However, as in practice, we cannot obtain the empirical characteristic function of the one-sided error terms, we need to find a close approximation described in what follows. To introduce our procedures, we notice that the density function may be obtained as

$$p(z) = \tfrac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\iota\tau z}\varphi(\tau)\,\mathrm{d}\tau, \tag{11}$$

where $p(z)$ is the density of a random variable $\mathbb{Z}$ taking values z, $\varphi(\tau)$ is the respective characteristic function. The FFT-based algorithm is applied to derive the standardized density by applying the Fourier transform to the cdf. The integral in (11) is evaluated for N equally spaced points with distance h, that is

$$\chi_k = (k - 1 - \tfrac{N}{2})h,\ k = 1, ..., N. \tag{12}$$

If we set $\tau = 2\pi\omega$ we have

$$p(\chi_k) = \int_{-\infty}^{\infty} \varphi(2\pi\omega)e^{-\iota 2\pi\omega\chi_k}\,\mathrm{d}\omega. \tag{13}$$

This integral can be approximated by using the rectangle rule for N points with spacing s, i.e.

$$p(\chi_k) \simeq s\sum_{n=1}^{N} \varphi(2\pi\omega_n)e^{-\iota 2\pi\omega_n\chi_k}, \tag{14}$$

where $\omega_n = (n - 1 - \frac{N}{2})s$. Setting $s = \frac{1}{Nh}$ we obtain the following approximation:

$$p(\chi_k) \simeq s(-1)^{k-1-(N/2)}\sum_{n=1}^{N}(-1)^{n-1}\varphi(2\pi\omega_n)e^{-\frac{\iota 2\pi(n-1)(k-1)}{N}}. \tag{15}$$

This summation can be computed efficiently by applying the Fast Fourier Transform (FFT) to the sequence

$$(-1)^{n-1}\varphi(2\pi\omega_n),\ n=1,...,N.$$

Each element (say the kth one) is normalized by $s(-1)^{k-1-(N/2)}$ to obtain the density function at each point of the grid.

Any characteristic function satisfies the conditions: (i) $\varphi(0)=1$, (ii) $|\varphi(\tau)|\leq 1\ \forall\tau\in\mathbb{R}$, and (iii) $\varphi(-\tau)=\overline{\varphi(\tau)}\ \forall\tau\in\mathbb{R}$, where the overbar denotes complex conjugation. Since $\varphi_u(\tau)$ is a univariate function, we can use standard kernel smoothing techniques to estimate (9) as a semi-parametric linear model, which can be re-normalized so that properties (i), (ii) and (iii) hold. As this strategy seems prohibitively difficult, a simpler alternative is to use a **mixture-of-normal-distributions model (MNM)** for the characteristic function (see (A.5) in Appendix A as well as the additivity property of any characteristic function in (A.3)):

$$\varphi_u(\tau)=\sum_{g=1}^{G}p_g e^{\iota\tau\mu_g-(\sigma_g^2/2)\tau^2}, \tag{16}$$

where $\{\mu_g,\sigma_g,p_g,\ g=1,\ldots,G\}$ are unknown coefficients, $\sum_{g=1}^{g}p_g=1$, $p_g\geq 0$, $g=1,\ldots,G$, and G is the number of terms in the mixture. From (9) and (16) we obtain:

$$\ln\varphi_{\hat{\varepsilon}_{it}}(\tau_j)=-\tfrac{\sigma_v^2\tau_j^2}{2}+\ln\left\{\sum_{g=1}^{G}p_g e^{\iota\tau_j\mu_g-(\sigma_g^2/2)\tau_j^2}\right\}+\xi_j,\ 1\leq j\leq J. \tag{17}$$

In principle, this model approximates well any distribution of the one-sided error component, as long as G is large enough. The model is nonlinear but can be estimated using the method of nonlinear least squares. The parameters to be estimated are σ_v^2, $\{p_g,\mu_g,\sigma_g\}_{g=1}^{G}$. The data consist of points $\{\tau_j,\ j=1,\ldots,J\}$.

The selection of points $\{\tau_j,\ j=1,\ldots,J\}$ is guided by the consideration that sufficient points are to zero (as this determines the tail behavior of the corresponding density. We place N_0 points in the interval $\mathscr{I}=[-\epsilon,\epsilon]$ and another N_B points in the interval $[-B,B]$ excluding the interval $\mathscr{I}$) after checking that the results are insensitive to larger values of B. Of course, $J=N_0+N_B$. To select G, we use the standard Bayes Information Criterion (BIC). In any regression model with sum of squared residuals SSR, n observations and k regressors, the BIC is $BIC=n\ln\left(\frac{SSR}{n}\right)+k\ln n$.

The exact configuration of parameters, ϵ,B,N_0,N_B will be described as we proceed, but within certain bounds

their exact specification is immaterial.

To obtain the density we use the Fast Fourier Transform (FFT) in (5); see Tsionas (2012). Specifically, let $N = 2^{N'}$, and define the sequence

$$u_k = \left(k - 1 - \tfrac{N}{2}\right) h,\ 1 \leq k \leq N, \tag{18}$$

for some $h > 0$ and integer N'. The density can be expressed as

$$f_u(u_k) = \int_{-\infty}^{\infty} \varphi_u(2\pi\omega) e^{-2\pi\iota\omega u_k}\, \mathrm{d}\omega,\ 1 \leq k \leq N. \tag{19}$$

As this is the inverse Fourier transform, it provides the density once we know the characteristic function. This is based on equations (6) and (7) of Mittnik et al. (1999). An approximation is obtained as follows, by rectangle integration.

$$f_u(u_k) \simeq s \sum_{n=1}^{N} \varphi(2\pi\omega_n) e^{-2\pi\iota\omega_n u_k},\ 1 \leq k \leq N, \tag{20}$$

where $\omega_n = \left(n - 1 - \frac{N}{2}\right) s$ and $s = \frac{1}{Nh}$. Therefore,

$$f_u(u_k) \simeq s(-1)^{(k-1-N/2)} \sum_{n=1}^{N} (-1)^{n-1} \varphi(2\pi\omega_n) e^{-2\pi\iota(n-1)(k-1)/N}, \tag{21}$$

see Mittnik et al. (1999). This expression is essentially an application of rectangle integration. Based on the results of Doganoglu and Mittnik (1998) and Mittnik et al. (1999), the error of approximation is close to 10^{-7}. Therefore, an efficient way to compute the density is to apply the FFT to the sequence $(-1)^{n-1}\varphi(2\pi\omega_n)$.[2] The transform needs to be normalized by $s(-1)^{k-1-N/2}$. In our computations, we use $N' = 18$ (yielding $N = 262,144$) and we set $h = \frac{2\varepsilon^*}{N}$ where $\varepsilon^* = \max_{1\leq i\leq I, 1\leq t\leq T} \hat{\varepsilon}_{it}$ (Duhamel and Vetterli, 1990). To compute individual inefficiency scores, we need the distribution of inefficiency conditional on the data which is given by

$$p_{u_{it}}(u_{it}|D) \propto e^{-(r_{it}-u_{it})^2/(2\hat{\sigma}_v^2)} f_u(u_{it}), u_{it} \geq 0,\ 1 \leq i \leq I, 1 \leq t \leq T, \tag{22}$$

where D denotes the data on y_{it} and x_{it},and $r_{it} = y_{it} - \hat{\alpha}_i - x_{it}'\hat{\beta}$ given estimates $\hat{\sigma}_v^2$, $\hat{\alpha}_i$ and $\hat{\beta}$. The first term, $e^{-(r_{it}-u_{it})^2/(2\hat{\sigma}_v^2)}$, corresponds to the normality of v_{it} given u_{it}.The second term, $f_u(u_{it})$ is simply the density of u_{it}.

The first problem is to compute the density $f_u(u_{it})$ at points that are different from the u_ks $(1 \leq k \leq N)$ used

[2]We use the `fortran` subroutine `ZFFT1I` in package `FFTPACK5`, by Paul Swarztrauber and Dick Valent.

in the FFT. The values of the density can be computed easily using interpolation (Tsionas, 2012). The second problem is to compute the normalizing constants using standard quadrature techniques:

$$C_{it} = \int_0^\infty e^{-(r_{it}-u)^2/(2\hat{\sigma}_v^2)} f_u(u)\,\mathrm{d}u, \tag{23}$$

In turn, inefficiency scores can be computed as

$$\hat{u}_{it} = \mathbb{E}(u_{it}|D) = C_{it}^{-1} \int_0^\infty u \cdot e^{-(r_{it}-u)^2/(2\hat{\sigma}_v^2)} f_u(u)\,\mathrm{d}u, \tag{24}$$

again, using quadrature techniques. The expression in (24) is simply the expected value of u from its posterior conditional distribution.

3 Deconvolution in DEA

Suppose $\{U_{it}, 1 \leq i \leq I, 1 \leq t \leq T\}$ is a set of inefficiency scores from (one from the several versions of) DEA. Clearly, a certain amount of noise (denoted v_{it}) is present, with the implication that the true inefficiency scores are $\{u_{it}, 1 \leq i \leq I, 1 \leq t \leq T\}$. Therefore, we can write the DEA scores as

$$U_{it} = v_{it} + u_{it},\ 1 \leq i \leq I, 1 \leq t \leq T. \tag{25}$$

We assume that $v_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_v^2)$. The distribution of actual inefficiency scores, under this assumption, has density

$$f_u(u) = \int_{-\infty}^\infty f_U(v+u)(2\pi\sigma_v^2)^{-1/2} e^{-v^2/(2\sigma_v^2)}\,\mathrm{d}v. \tag{26}$$

If σ were known, this integral would be computed using draws from a normal distribution, and the empirical density $\hat{f}_U(U)$ would be computed by standard kernel density techniques. To estimate σ we consider the log characteristic function

$$\log \hat{\varphi}_U(\tau) = -\tfrac{\sigma_v^2}{2}\tau^2 + \log \varphi_u(\tau) + \xi(\tau), \tag{27}$$

for some error term $\xi(\tau)$, where $\hat{\varphi}_U(\tau)$ is the empirical characteristic function of $\{U_{it}\}$:

$$\varphi_U(\tau) = (IT)^{-1} \sum_{i=1}^{I} \sum_{t=1}^{T} e^{\iota \tau U_{it}} \; \forall \tau \in \mathbb{R}. \tag{28}$$

We can approximate $\log \varphi_u(\tau)$ using the same methods as in (9) and (16). As this, in turn, provides an estimate of σ, we can recover the density of u_{it} using (5) together with the FFT transform as implemented in (18)-(21), in a flexible way.

4 Monte Carlo evidence

To assess our new techniques we consider a production function of the form

$$y_i = \beta_0 \prod_{k=1}^{K} x_i^{\beta_k} e^{v_i - u_i}, \; 1 \leq i \leq I, \tag{29}$$

where the x_is are generated from a lognormal distribution with location parameter 1 and standard deviation 2. We set $\beta_0 = 1$, $\beta_1 = \cdots \beta_K = \frac{1}{K}$, $v_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_v^2)$, $u_i \overset{iid}{\sim} \mathcal{N}_+(0, \sigma_u^2)$, defining the variance parameter $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and the signal-to-noise ratio $\lambda = \frac{\sigma_u}{\sigma_v}$.[3] We examine various values of I,λ and σ with $K = 10$ which give five input prices and five outputs. Since we have a production frontier we must take into account (4). In Table 1, Monte Carlo results for correlation coefficients between actual and predicted inefficiency are reported from SFM and output-oriented, variable-returns-to-scale DEA. We use 5,000 Monte Carlo replications. SFMs are estimated using the method of maximum likelihood for the MNM model, when the actual generating mechanism is a half-normal distribution, allowing for 500 maximum iterations. In case of non-convergence we draw another data set.[4]

[3]The original parameters may be recovered as $\sigma_v^2 = \frac{\sigma^2}{1+\lambda^2}$ and $\sigma_u^2 = \frac{\lambda^2 \sigma^2}{1+\lambda^2}$.

[4]The half-normal is an obvious and popular case to consider. Monte Carlo results for other distributions are reported below. Moreover, when the distribution of the one-sided error component is known, application of the FFT to recover the density function is closely approximated by the MNM with an error 10^{-7}.

Table 1: Monte Carlo results for FFT-SFM and FFT-DEA

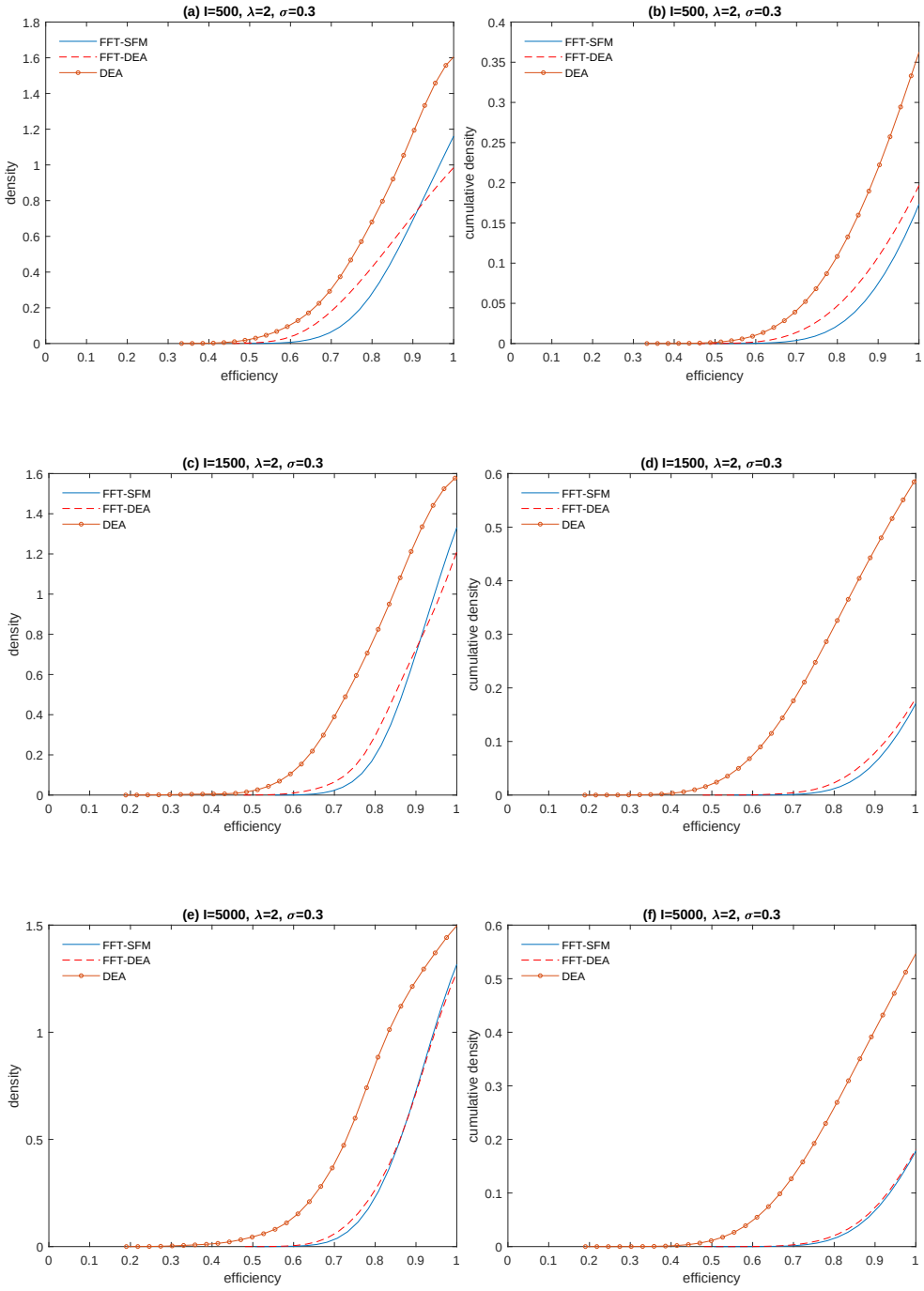
I=100	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 3$	$\lambda = 5$
$\sigma = 0.3$	0.516 *0.510*	0.589 *0.580*	0.912 *0.895*	0.980 *0.970*
$\sigma = 0.5$	0.518 *0.505*	0.588 *0.578*	0.913 *0.897*	0.982 *0.973*
$\sigma = 1$	0.519 *0.507*	0.589 *0.579*	0.912 *0.905*	0.983 *0.982*
I=500				
$\sigma = 0.3$	0.521 *0.519*	0.605 *0.604*	0.935 *0.934*	0.950 *0.949*
$\sigma = 0.5$	0.520 *0.518*	0.607 *0.607*	0.937 *0.934*	0.962 *0.963*
$\sigma = 1$	0.519 *0.517*	0.615 *0.613*	0.939 *0.935*	0.965 *0.965*
I=1,000				
$\sigma = 0.3$	0.525 *0.529*	0.625 *0.624*	0.955 *0.954*	0.960 *0.969*
$\sigma = 0.5$	0.530 *0.531*	0.626 *0.625*	0.944 *0.944*	0.982 *0.983*
$\sigma = 1$	0.533 *0.532*	0.635 *0.633*	0.959 *0.955*	0.975 *0.975*

Notes: Results for FFT-SFM are reported in regular font and FFT-DEA in italics. The actual generating mechanism is a half-normal distribution for the one-sided error and we use the MNM model in (16) with $G = 3$. This selection is based on the BIC criterion.

The important conclusion from the Monte Carlo simulations is that the performance of both SFM and DEA perform rises with the value of as λ. This is also true for sample size (I) and for values of $\lambda \geq 3$ where the results in terms of efficiency scores are accurate. For a direct comparison with the performance of standard DEA, we report results graphically in Figure 1; when $\lambda = 2$ and $\sigma = 0.3$ for I=500 (in panels (a) and (b)) and I=1500 (in panels (c) and (d)). These effects become noticeable when the sample size increases to I =5,000 (panels (e) and (f)) where the performance of standard DEA is inferior to that of DEA with smaller sample sizes.[5]

[5]To implement the deconvolution, we use $G = 3$ as, after experimentation, this was the most frequent value that performed well in simulations. Using $G = 4$ and $G = 5$, did not change our results in a substantive way.

Figure 1: Further Monte Carlo results corresponding to different values of λ and I.



The deterioration of the performance of DEA as the sample size increases is caused by additional noise as observations are added. Compared to earlier studies, Banker et al. (1993) find that DEA performs relatively well when the amount of measurement error is low but not when noise is relatively large. From the same study corrected

ordinary least squares (COLS) is shown to be more sensitive to data noise, so that the results are inferior to those of DEA (see, however, Ruggiero, 1999, on this point). For this reason we give no further consideration to COLS. In Table 2 we report mean absolute errors (MAE) between actual and estimated inefficiency.

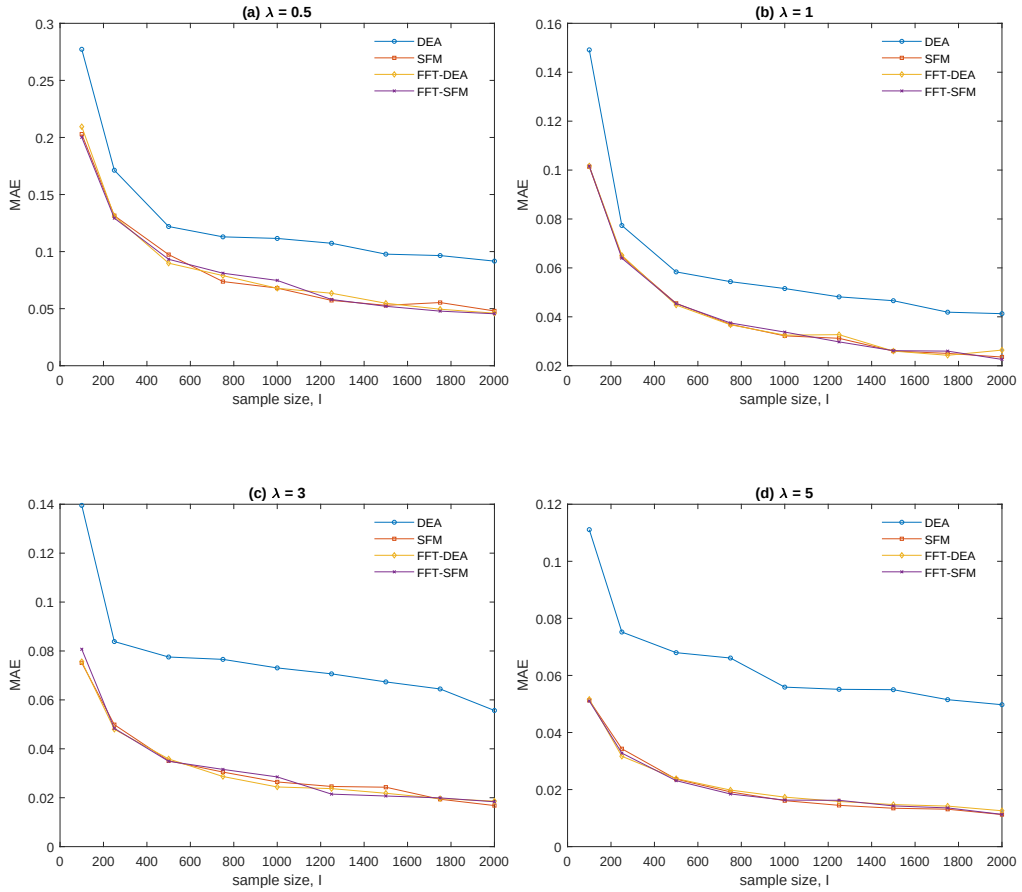
Table 2: Monte Carlo results (MAE) for FFT-SFM and FFT-DEA.

I=100	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 3$	$\lambda = 5$
$\sigma = 0.3$	0.093 *0.091*	0.077 *0.074*	0.051 *0.047*	0.030 *0.029*
$\sigma = 0.5$	0.090 *0.092*	0.075 *0.077*	0.050 *0.052*	0.029 *0.031*
$\sigma = 1$	0.088 *0.086*	0.074 *0.072*	0.048 *0.047*	0.027 *0.025*
I=500				
$\sigma = 0.3$	0.088 *0.090*	0.071 *0.075*	0.047 *0.050*	0.026 *0.029*
$\sigma = 0.5$	0.087 *0.089*	0.070 *0.074*	0.047 *0.049*	0.025 *0.028*
$\sigma = 1$	0.086 *0.087*	0.070 *0.072*	0.046 *0.047*	0.024 *0.025*
I=1,000				
$\sigma = 0.3$	0.075 *0.079*	0.062 *0.071*	0.041 *0.045*	0.020 *0.022*
$\sigma = 0.5$	0.074 *0.075*	0.060 *0.063*	0.040 *0.042*	0.020 *0.021*
$\sigma = 1$	0.073 *0.072*	0.059 *0.057*	0.040 *0.041*	0.020 *0.020*

Notes: Results for FFT-SFM are reported in regular font and FFT-DEA in italics. The actual generating mechanism is a half-normal distribution for the one-sided error and we use the MNM model in (16) with $G = 3$. This selection is based on the BIC criterion.

From Table 2 and Figure 2, the FFT-DEA and SFM results in terms of MAEs are similar. Their decline with the square root of the sample size suggests that asymptotic normality of the estimators is possible. For a more realistic view of the relationship between DEA, FFT-DEA and FFT-SFM we use as our regressors the observations from the empirical application of the next section (2,397 observations, ten regressors, production SFM). The true coefficients are set to those estimated by least squares (LS). More specifically, we consider samples of size 100, 250, 500, 750, 1000, 1250, 1500, 1750, and 2000 for various values of λ fixing $\sigma = 0.3$. The results are reported in Figure 2.

Figure 2: Monte Carlo results using as regressors the actual data



The important conclusion from Figure 2, is the similar behavior of mean absolute errors for SFM, FFT-SFM and FFT-DEA which decrease approximately at a rate proportional to $\sqrt{I}$. The rate of convergence of DEA is, of course, quite different and its MAE is distinctly larger compared to the other techniques. Moreover, the FFT-based techniques were found to be robust not only to the number of points J but also to their placement around zero. This evidence is corroborates results reported previously in Tsionas (2012).

Another important issue is the sensitivity of results when the inputs and the two-sided error are correlated, i.e., there is an endogeneity problem (this comment was provided by an anonymous reviewer). The x_is are generated as previously, but they depend on the two-sided error term. Each regressor is then generated using $x_i = \tilde{x}_i + \epsilon_i$, where the $\tilde{x}_i$s are generated from a lognormal distribution with location parameter 1 and standard deviation 2, as before. The error term ϵ_i is normally distributed with zero mean and variance σ_ϵ^2, where σ_ϵ^2 depends on the correlation

coefficient ϱ based on $\tilde{x}_i$ and the two-sided error term. Our results are reported in Table 3, when $\lambda = 0.5$ and for different values of σ and ϱ. FFT-SFM and FFT-DEA are based on the MNM model defined in (16). The BIC criterion favored $G = 3$ normal components. Based on the results in Table 3, MAEs increase with ϱ and with σ but decrease with higher values of λ (Table 2). Even with $\varrho = 0.25$, the results are acceptable but as ϱ increases to 0.5 or 0.75 the MAEs are increasing and, therefore, the results are not to be trusted. This is reasonable as endogeneity that is ignored has material impact on coefficient estimates and functions of interest when ϱ exceeds 0.25 or so.

Table 3: Monte Carlo results (MAE) for FFT-SFM and FFT-DEA under endogeneity (correlation coefficients between actual and estimated.)

I=100	$\varrho = 0.1$	$\varrho = 0.25$	$\varrho = 0.50$	$\varrho = 0.75$
$\sigma = 0.3$	0.012 0.017	0.021 0.022	0.035 0.037	0.040 0.044
$\sigma = 0.5$	0.015 0.016	0.025 0.027	0.030 0.035	0.042 0.040
$\sigma = 1$	0.022 0.027	0.029 0.032	0.045 0.048	0.052 0.057
I=500				
$\sigma = 0.3$	0.017 0.019	0.022 0.025	0.045 0.049	0.051 0.055
$\sigma = 0.5$	0.021 0.025	0.025 0.029	0.049 0.052	0.062 0.065
$\sigma = 1$	0.025 0.29	0.033 0.037	0.054 0.057	0.079 0.081
I=1,000				
$\sigma = 0.3$	0.019 0.099	0.097 0.0991	0.101 0.105	0.067 0.075
$\sigma = 0.5$	0.021 0.025	0.105 0.109	0.112 0.118	0.075 0.088
$\sigma = 1$	0.025 0.029	0.105 0.115	0.122 0.129	0.092 0.094

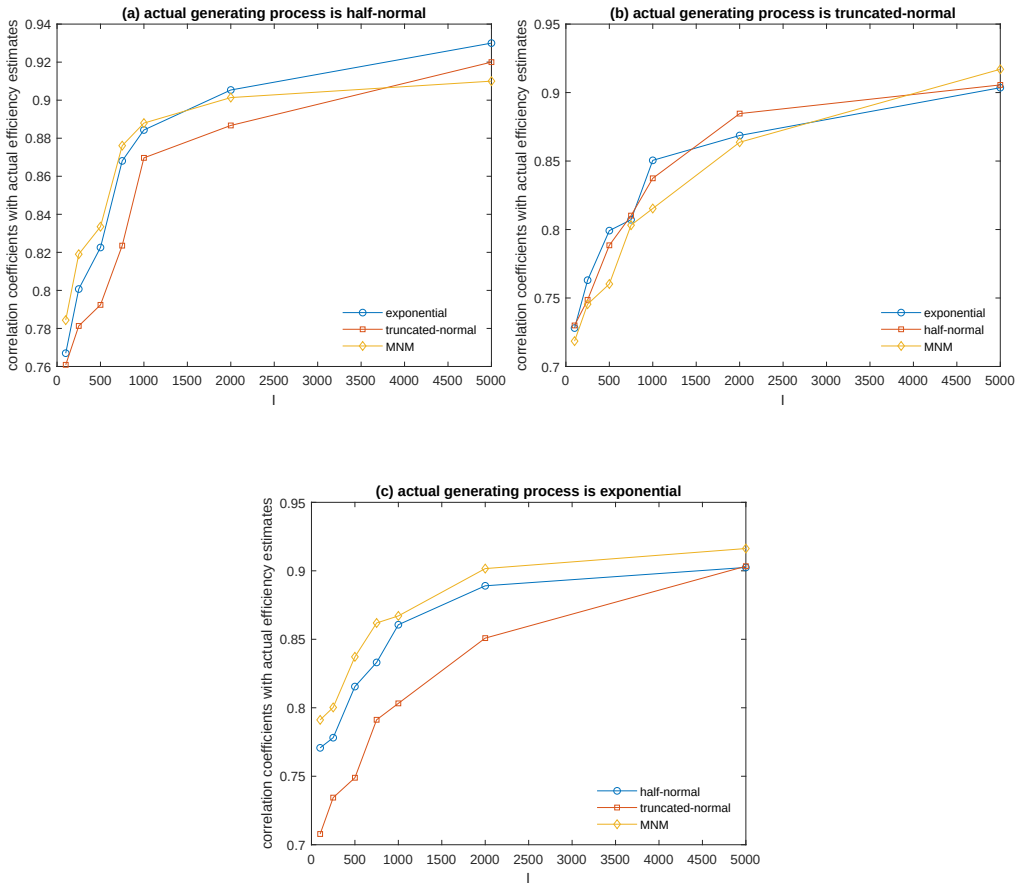
Notes: Results for FFT-SFM are reported in regular font and FFT-DEA in italics. MNM model is implemented with a half-normal distribution for the one-sided error term and $G = 3$ in (16).

Another issue (pointed out by an anonymous reviewer) is the performance of efficiency estimates when the wrong distribution is adopted for technical inefficiency. Again, we use the MNM model in (16) with $G = 3$. This selection is based on the BIC criterion. The Figures below, present results for different generating processes in comparison to all others. Our generating processes are the same as previously with the exception that technical inefficiencies are generated by different distributions. We choose $\lambda = 0.5$ and $\sigma = 0.3$. Moreover, we use sample sizes I=100, 250, 500, 750, 1000, 2000, and 5000. The results are reported in Figure 3.

5 Empirical application

We have an unbalanced panel with 2,397 bank-year observations for 285 large U.S. commercial banks operating in 2001-2010, whose total assets were more than one billion 2005 U.S. dollars over the first three years of observation. The data come from Call Reports available from the Federal Reserve Bank of Chicago as used in Malikov et al.

Figure 3: Correlation coefficients when the data generating process of technical inefficiency is misspecified



(2016). For detailed description of the data construction, see Section 5 of Malikov et al. (2016). The list of included variables is as follows: y_1 Consumer Loans, y_2 Real Estate Loans, y_3 Commercial & Industrial Loans, y_4 Securities, y_5 Off-Balance Sheet Activities Income, x_1 Labor, number of full-time employees, x_2 Physical Capital (Fixed Assets), x_3 Purchased Funds, x_4 Interest-Bearing Transaction Accounts, x_5 Non-Transaction Accounts.

We use an input distance function (IDF; Kumbhakar and Lovell, 2000, pp. 28–30) defined as

$$D(x,y) = \max\left\{\lambda : \left(\lambda^{-1}x, y\right) \text{ is technically feasible}\right\}. \tag{30}$$

The IDF is homogeneous of degree one in inputs, increasing in inputs, decreasing in outputs, concave in inputs, and $D(x,y) \leq 1$. These constraints are enforced at ten randomly selected observed points and the geometrical means of the data and then we check whether they hold at the majority of observations. Suppose $\mathbf{x}_{it} \in \mathbb{R}^K_+$ and $\mathbf{y}_{it} \in \mathbb{R}^M_+$ are, respectively, logs of inputs and outputs, a translog ODF is given as follows:

$$x_{it,1} = a_i + \boldsymbol{\beta}_1'\mathbf{z}_{it} + \tfrac{1}{2}\mathbf{z}_{it}'\boldsymbol{B}\mathbf{z}_{it} + \varepsilon_{it},\ \varepsilon_{it} \equiv v_{it} + u_{it}, \tag{31}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{B}$ are a vector and a symmetric matrix of parameters, respectively, and a_i denotes bank effects. Let us define

$$\mathbf{z}_{it} = [x_{it,2} - x_{it,1}, \ldots, x_{it,K} - x_{it,1}, y_{it,1}, y_{it,2}, \ldots, y_{it,M}, \mathscr{T}_{it}]', \tag{32}$$

where $\mathscr{T}_{it}$ denotes an approximation to temporal effects:

$$\mathscr{T}_{it} = \sum_{\tau=1}^{T} \alpha_{i\tau}\mathbb{I}(TR_{it} = \tau), \tag{33}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, and $TR_{it} = t$ (for all i and t). The coefficients in (33) are firm specific and we assume the following parameterization:

$$\alpha_{i\tau} = \tilde{\mathbf{z}}_{i\tau}'\boldsymbol{\varrho},\ \forall i, \tau \tag{34}$$

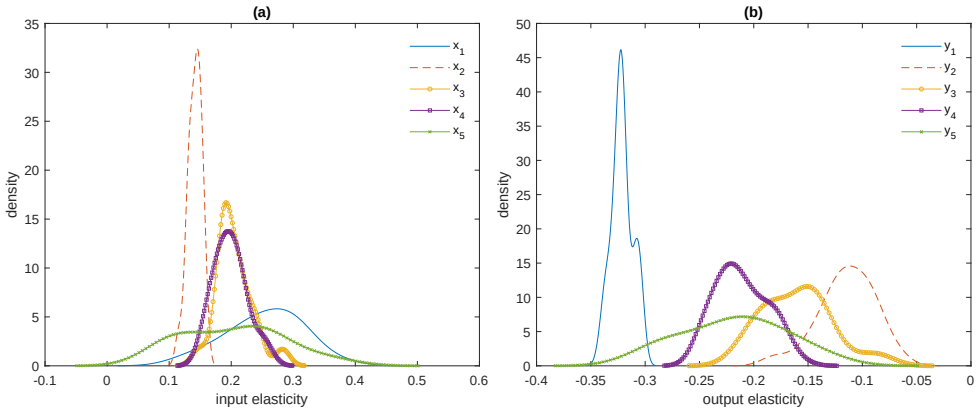
where $\tilde{\mathbf{z}}_{i\tau} = [x_{i\tau,2} - x_{i\tau,1}, \ldots, x_{i\tau,K} - x_{i\tau,1}, y_{i\tau,1}, y_{i\tau,2}, \ldots, y_{i\tau,M}]'$, and $\boldsymbol{\varrho}$ is a conformable vector. The specification in (33) resembles the Baltagi and Griffin (1988) general index of technological change. The difference is that the

coefficients are firm-specific due to the Baltagi in (34). This makes it possible to estimate firm-specific as well as time-varying technical change (TC_{it}) given as

$$TC_{it} = \alpha_{it} - \alpha_{i,t-1}. \tag{35}$$

The differencing of inputs in (32) is due to homogeneity of degree one of (31). The model in (31) is estimated using the maximum likelihood approach with the difference that the theoretical properties of the IDF are enforced at the means of the data and ten other randomly selected points. At the remaining points, the IDF performs extremely well in terms of satisfying theoretical properties. Indeed, the sample distributions of input and output elasticities, reported in panels (a) and (b) of Figure 4 satisfy the monotonicity properties. This means that an IDF should be increasing in other inputs and decreasing in outputs (Kumbhakar and Lovell, 2000, p. 30).

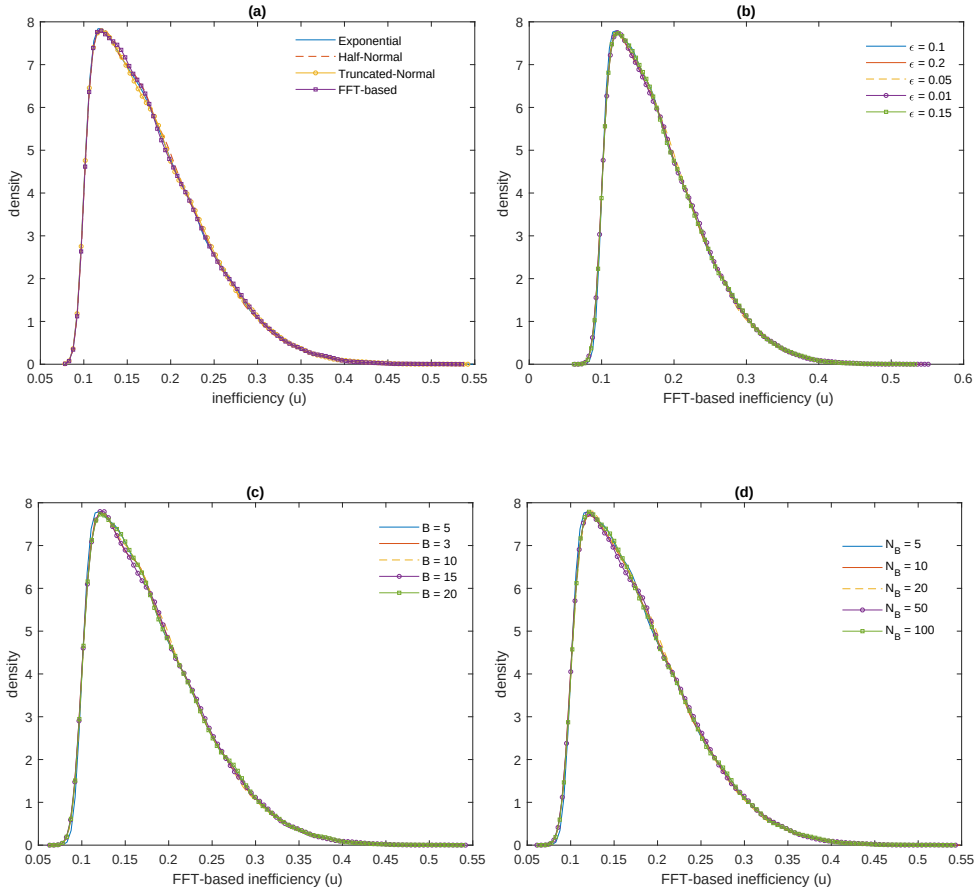
Figure 4: Sample distributions of input and output elasticities.



In this application the value $B = 5$ was found adequate and using the BIC criterion we choose $G = 3$. Moreover, we set $\epsilon = 0.1$, $N_0 = 20$ and $N_B = 50$ as our benchmark case, and we examine sensitivity with respect to other values of these parameters. These values were used in order to reduce the approximation errors close to 10^{-7}. In panel (a) of Figure 5, we report the densities of estimated inefficiencies under four different assumptions based on a translog functional form for the IDF: The exponential, half-normal, truncated-normal and FFT-based inefficiencies. In panels (b), (c) and (d) we report sensitivity to other values of the benchmark parameters ϵ, B, and N_B. The results from panel (a) show that the distributions of inefficiency are in broad agreement except for the exponential

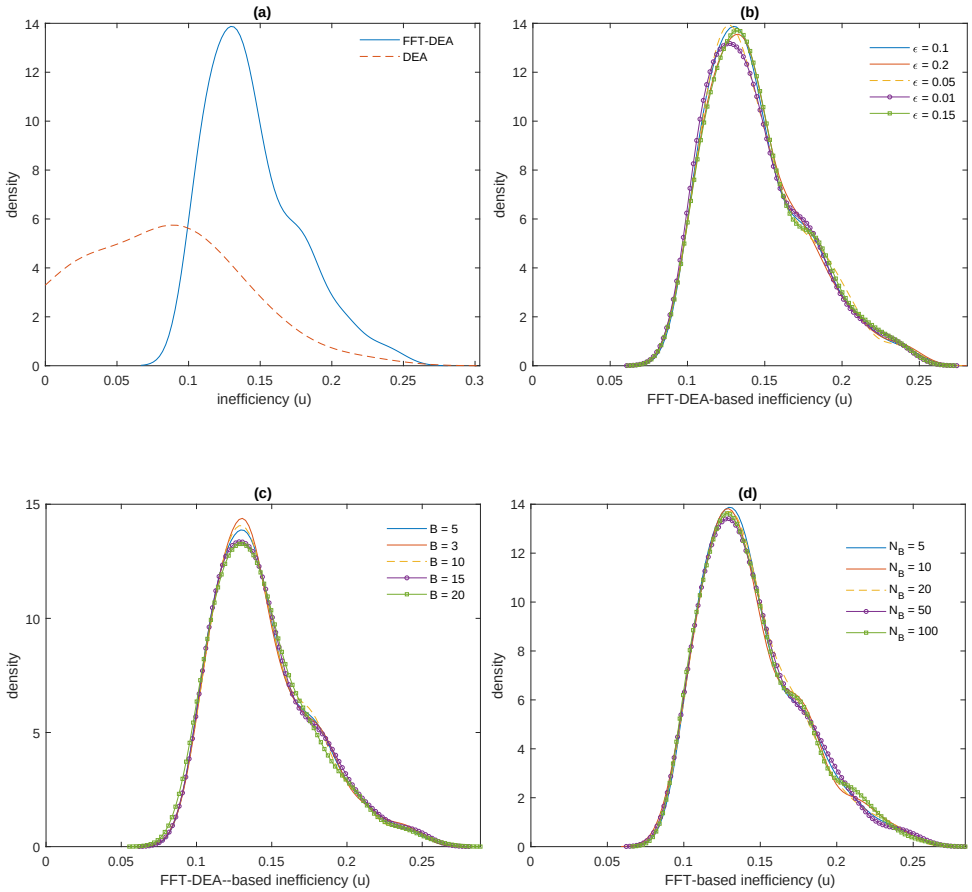
distribution. From panels (b), (c) and (d) it turns out that the FFT-based densities are more or less the same except when B is too small (viz. B=3).

Figure 5: Sampling distributions of inefficiency scores for SFM.



From (5) it is evident that the results are not sensitive to the exact specification of the parameters presented in panels (b), (c), and (d). The corresponding results for FFT-DEA and DEA are reported in Figure 6.

Figure 6: Sampling distributions of inefficiency scores for DEA and FFT-DEA

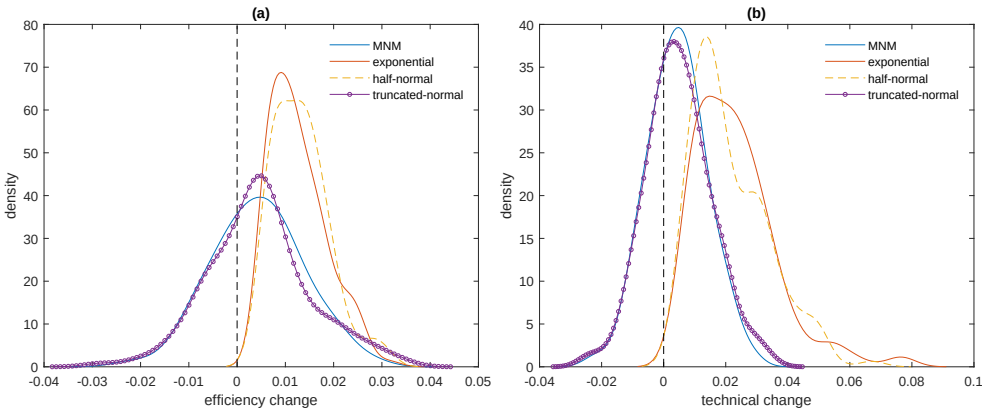


That FFT-DEA does not deliver inefficiency results drastically different to those from the SFM might be explained by the use of a flexible functional form. On the other hand, the results from standard output-oriented variable-returns-to-scale DEA are quite different predicting a substantial proportion of zero inefficiency scores as it is commonly the case with DEA. Deconvolution of DEA scores with the FFT brings the results much closer to the inefficiency estimates from SFM, as we see in Figure 6. Further aspects of the SFM model like technical change (TC), efficiency change (EC), and productivity growth (PG) are reported in Figure 7. TC is defined as the derivative of IDF with respect to time, EC is $EC_{it} = \frac{e^{-u_{it}} - e^{-u_{i,t-1}}}{\frac{1}{2}\left(e^{-u_{it}} + e^{-u_{i,t-1}}\right)}$ and $PG_{it} = TC_{it} + EC_{it}$.

TC averages –0.0091 (standard deviation, s.d. 0.0106), EC averages 0.0050 (s.d. 0.0106) and PG averages –0.0040 (s.d. 0.015) so, no performance improvements seem to take place on the average. Nevertheless, certain banks experienced TC as high as 8.5%, and EC as high as 4%. For other banks, TC and EC were as low as –4%.

All models yield different results for TC, and PC, and, therefore, for PG as well. We report sample distributions of Efficiency Change (EC) in panel (a) of Figure 7. In panel (b) we report sampling distributions of technical change (TC) using different distributional assumptions. The implication is that different specifications yield different results and, if we need to accept one model out of many, then we need to use the BIC or other techniques like, for example, Bayes factors. This is not attempted here as our main interest is in comparing different models rather than in selecting the "best" model.

Figure 7: Sample distributions of EC from SFM.



The MNM and truncated-normal for EC sample distributions are quite close, while the half-normal and exponential model are close but different compared to MNM and truncated-normal specifications. The intuition of this result is as follows. Given a flexible specification like the translog, deconvolution of the noise arising from the inefficiency in using the empirical characteristic function is reasonably accurate; and all remaining differences in functions of interest arise from different distributional assumptions relating to the one-sided error term. Clearly, this paves a way to combine DEA and SFMs in a manner not previously possible that have not been possible before; for example, using DEA from a certain sub-sample as an empirical Bayes prior of inefficiency distribution in SFMs (e.g., Tsionas, 2003).

6 Concluding remarks

The purpose of the paper is twofold. First, in the context of SFMs, we show how to compute inefficiency estimates when the distribution of the one-sided error component representing inefficiency is unknown. Second, in the context of DEA we show how to perform deconvolution of noise from inefficiency given the DEA efficiency scores. Our contributions are based on empirical characteristic functions and deconvolution, using the Fast Fourier Transform, and a generic new model of approximating any distribution of the one-sided error term via a mixture-of-normals model. The new techniques are shown to perform well in Monte Carlo experiments as well as in an application to large U.S. banks. The success of the new techniques suggests the usefulness of FFT, and empirical characteristic functions in future empirical work. In terms of limitations and future research, one can investigate different distributions of the two-sided error. For example, the Student-t would be an obvious candidate if there are outliers in the data. Second, adding environmental or contextual variables in inefficiency is an interesting avenue for further research that is not consider in this paper due to its complexity. Third, new challenges for the FFT are likely to arise if inefficiency is dynamic in both DEA and SFA. Finally, Bayesian inference procedures are likely to be important in the class of models considered in this paper. The major obstacle is that Bayesian analysis using characteristic functions instead of densities, is quite difficult. In this field, it would be useful to have results related to Bayesian updating of characteristic functions corresponding to the similar updating of densities.

Appendix A. Mathematical Appendix

For any random variable X, the characteristic function is defined as

$$\varphi_X(\tau) = \mathbb{E}(e^{\iota\tau X}) = \int e^{\iota\tau x} f_X(x)\,\mathrm{d}x, \tag{A.1}$$

where $\iota = \sqrt{-1}$, $\tau \in \mathbb{R}$, and $f_X(x)$ is the probability density function of X evaluated at a point x. The characteristic function completely characterizes the distribution of X. If the characteristic function of a random variable X is integrable, then its distribution function is absolutely continuous, and therefore X has a probability density function,

which is given by inverting (A.1):

$$f_X(x) = \tfrac{1}{2\pi} \int_{-\infty}^{\infty} e^{\iota\tau x} \varphi_X(\tau)\, \mathrm{d}\tau,\ \forall x \in \mathbb{R}, \tag{A.2}$$

For a complete treatment, see Lukacs (1970). Some of the properties of the characteristic function are listed below.

$$\begin{gathered} \varphi_X(0) = 1, \\ |\varphi_X(\tau)| \leq 1,\ \forall \tau \in \mathbb{R}, \\ \text{for all random variables } X \text{ and } Y \text{we have :} \\ \varphi_{X+Y}(\tau) = \varphi_X(\tau)\varphi_Y(\tau),\ \forall \tau \in \mathbb{R}, \\ \varphi(-\tau) = \overline{\varphi_X(\tau)},\ \forall \tau \in \mathbb{R}, \\ \varphi_X(\tau) \text{ is always uniformly continuous,} \end{gathered} \tag{A.3}$$

where $\overline{\varphi_X(\tau)}$ denotes the complex conjugate. Any characteristic function admits the following expansion:

$$\log \varphi_X(\tau) = \sum_{n=0}^{\infty} \kappa_n \tfrac{(\iota\tau)^n}{n!}, \tag{A.4}$$

and the κ_ns are knows as cumulants. The characteristic function of a normal distribution, $\mathcal{N}(\mu, \sigma^2)$, is given by

$$\varphi_X(\tau) = e^{-\iota\mu\tau - (1/2)\sigma^2\tau^2},\ \forall \tau \in \mathbb{R}. \tag{A.5}$$

Finally, we provide some characteristics functions corresponding to well-known distributions below:

$$\begin{gathered} gamma(k, a) :\ \varphi(\tau) = (1 - \iota\tau a)^{-k}, \\ \text{exponential}(\lambda) :\ \varphi(\tau) = (1 - \iota\tau\lambda^{-1})^{-1} \\ \text{half-normal} :\ \varphi(\tau) = 2e^{-\frac{1}{2}\sigma^2\tau^2} \Phi\left(\iota\sigma\tau\right), \\ \text{truncated normal: See p. 7 of Abadir and Magdalinos (2002).} \end{gathered} \tag{A.6}$$

Here, $\Phi(\cdot)$ is the standard normal distribution function.

References

[1] Abadir, K., & Magdalinos, T., 2002. The Characteristic Function from a Family of Truncated Normal Distributions. Econometric Theory 18 (5), 1276–1287.

[2] Adams, R, M., & Sickles, R. C., 2007. Semiparametric Efficient Distribution Free Estimation of Panel Models. Communications in Statistics - Theory and Methods 36 (13), 2425–2442.

[3] Allon, G., M. Beenstock, S. Hackman, U. Passy, A. Shapiro, 2007. Nonparametric estimation of concave production technologies by entropic methods. Journal of Applied Econometrics, 22 (4), 795–816.

[4] Baltagi, B. H., and Griffin, J. M., 1988, A General Index of Technical Change, Journal of Political Economy 96 (1), 20–41.

[5] Banker, R.D., Gadh, V.M., Gorr, W.L., 1993. A Monte Carlo comparison of two production frontier estimation methods: Corrected ordinary least squares and data envelopment analysis. European Journal of Operational Research 67 (3) 332–343.

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

[7] Cordero, J. M., Santin, D., & Sicilia, G., 2015. Testing the accuracy of DEA estimates under endogeneity through a Monte Carlo simulation. European Journal of Operational Research 244 (2), 511–518.

[8] Doganoglu, T., & S. Mittnik, 1998. An approximation procedure for asymmetric stable Paretian densities. Computational Statistics, 13, 463–475.

[9] Duhamel, P., Vetterli, M., 1990. Fast Fourier transforms: a tutorial review. Signal Processing 19, 259–299.

[10] Dyson, R.G., R. Allen, A.S. Camanho, V.V. Podinovski, C.S. Sarrico, & E.A. Shale. 2001. Pitfalls and protocols in DEA. European Journal of Operational Research 132, 245–259.

[11] Feuerverger, A., & P. McDunnough, 1981a. On efficient inference in symmetric stable laws and processes, in M. Csorgo, ed., Statistics and Related Topics. New York: North-Holland, pp. 109–122.

[12] Feuerverger, A., and P. McDunnough, 1981b. On some Fourier methods for inference, Journal of the American Statistical Association 76, 379–387.

[13] Holland, D. S., & Lee, S. T., 2002. Impacts of random noise and specification on estimates of capacity derived from data envelopment analysis. European Journal of Operational Research 137 (1), 10–21.

[14] Johnson, A.L., & Kuosmanen, T., 2012. One-stage and two-stage DEA estimation of the effects of contextual variables. European Journal of Operational Research Volume 220 (2), 559–570.

[15] Keshvari, A., T. Kuosmanen, 2013. Stochastic non-convex envelopment of data: Applying isotonic regression to frontier estimation European Journal of Operational Research, 231 (2), 481–491.

[16] Koutrouvelis, I. A., 1980, Regression-type estimation of the parameters of stable laws, Journal of the American Statistical Association 75, 918–928.

[17] Koutrouvelis, I. A., 1981, An iterative procedure for estimation of the parameters of stable laws, Communications in Statistics - Simulation and Computation 10, 17–28.

[18] Kumbhakar, S. C., & Lovell, C.A.K., 2000. Stochastic Frontier Analysis. Cambridge, Cambridge University Press.

[19] Kuosmanen, T., & M. Kortelainen, 2012. Stochastic non-smooth envelopment of data: Semiparametric frontier estimation subject to shape constraints. Journal of Productivity Analysis, 38 (1) 11–28.

[20] Lee, C.Y., A.L. Johnson, E. Moreno-Centeno, & T. Kuosmanen, 2013. A more efficient algorithm for convex nonparametric least squares European Journal of Operational Research 227 (2), 391–400.

[21] Lukacs, E. (1970). Characteristic functions. London: Griffin.

[22] Malikov, E., S. C. Kumbhakar, & M. G. Tsionas, 2016. A cost system approach to the stochastic directional technology distance function with undesirable outputs: the case of U.S. banks in 2001-2010. Journal of Applied Econometrics, 31 (7), 1407–1429.

[23] Mittnik, S., Rachev, S. T., Doganoglu, & Chenyao, D. (1999). Maximum Likelihood of Stable Paretian Models. Mathematical and Computer Modelling 29, 276–293.

[24] Oh, S-C, & Shin, J., 2015. The impact of mismeasurement in performance benchmarking: A Monte Carlo comparison of SFA and DEA with different multi-period budgeting strategies. European Journal of Operational Research 240 (2), 518–527.

[25] Ruggiero, J., 1999. Efficiency estimation and error decomposition in the stochastic frontier model: A Monte Carlo analysis, European Journal of Operational Research 115, 555–563.

[26] Tsionas, M. G., 2003. Combining DEA and stochastic frontier models: An empirical Bayes approach. European Journal of Operational Research 147 (3), 499–510.

[27] Tsionas, M. G., 2012. Maximum likelihood estimation of stochastic frontier models by the Fourier transform. Journal of Econometrics 170, 234–248.

[28] Tsionas, M. G., 2019. A coherent approach to Bayesian Data Envelopment Analysis. European Journal of Operational Research 281 (2), 439–448.

[29] Tsionas, M. G., & Izzeldin, M., 2018. Smooth approximations to monotone concave functions in production analysis: An alternative to nonparametric concave least squares. European Journal of Operational Research 271 (3), 797–807.

[30] Wanke, P., Barros, C. P., & Emrouznejad, A., 2015. Assessing productive efficiency of banks using integrated Fuzzy-DEA and bootstrapping: A case of Mozambican banks. European Journal of Operational Research Volume 249 (1), 378–389.