Journal of Econometrics 170 (2012) 234–248

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Maximum likelihood estimation of stochastic frontier models by the Fourier transform

Efthymios G. Tsionas *

Department of Economics, Athens University of Economics and Business, 76 Patission Street, 104 34 Athens, Greece

ARTICLE INFO

Article history:
Available online 30 May 2012

JEL classification:
C13
C16
C46

Keywords:
Stochastic frontiers
Moment generating function
Characteristic function
Fast Fourier transform
Maximum likelihood
Generalized method of moments

ABSTRACT

The paper is concerned with several kinds of stochastic frontier models whose likelihood function is not available in closed form. *First*, with output-oriented stochastic frontier models whose one-sided errors have a distribution other than the standard ones (exponential or half-normal). The *gamma* and *beta* distributions are leading examples. *Second*, with input-oriented stochastic frontier models which are common in theoretical discussions but not in econometric applications. *Third*, with two-tiered stochastic frontier models when the one-sided error components follow *gamma* distributions. *Fourth*, with latent class models with *gamma* distributed one-sided error terms. *Fifth*, with models whose two-sided error component is distributed as stable Paretian and the one-sided error is *gamma*. The principal aim is to propose approximations to the density of the composed error based on the inversion of the characteristic function (which turns out to be manageable) using the Fourier transform. Procedures that are based on the asymptotic normal form of the log-likelihood function and have arbitrary degrees of asymptotic efficiency are also proposed, implemented and evaluated in connection with output-oriented stochastic frontiers. The new methods are illustrated using data for US commercial banks, electric utilities, and a sample from the National Youth Longitudinal Survey.

1. Introduction

Although practical experience with stochastic frontier models whose one-sided error components follow exponential or half-normal distributions[1] has been quite good, complicated and more realistic distributions usually yield problems. The *gamma* distribution has been proposed as a reasonable description of technical inefficiency by Greene (1990). The likelihood function is not available in closed form because the convolution of the two error components of the stochastic frontier model involves an integral which cannot be evaluated analytically. Beckers and Hammond (1987) showed that the likelihood can be computed using special functions. Greene (2003) has proposed the method of maximum simulated likelihood (MSL) in this context.[2] The troublesome term in the likelihood function is the expectation of a truncated normal random variable when a certain exponent is real. Setting up the simulated likelihood involves obtaining a (potentially large) number of independent draws from this distribution and taking their empirical average to approximate the expectation unless α is integer in which case closed form expressions are available.

One objective of the present paper is to show how the fast Fourier transform (FFT) of the characteristic function (CF) can be used to compute the density accurately and efficiently and investigate the possibility of using certain simplifications without sacrificing asymptotic efficiency. The FFT is intuitively appealing in this instance since in spite of the fact that the density is not available in closed form, the characteristic function has a rather simple expression.

Moreover, one can think of alternative reasonable models for technical inefficiency whose convolution with noise does not yield to simple, analytical expressions for the likelihood function. One such example is the *beta* distribution for technical efficiency. Despite its importance in modeling bounded random variables like technical efficiency, surprisingly this distribution has not been used before in econometric efficiency and productivity studies. An objective of the paper is to show how the likelihood can be computed accurately and efficiently using the FFT as in the case of *gamma* stochastic frontiers. The scope for such extensions is of course that "…there is considerable scope for alternatives to the original model of Aigner et al. (1977). But, for better or worse, the normal/half-normal model has dominated the received empirical literature" (Greene, 1999, p. 104).

* Tel.: +30 210 820 3338; fax: +30 210 820 3301.
E-mail address: tsionas@aueb.gr.

[1] The standard references are Aigner et al. (1977) and Meeusen and van den Broeck (1977). For reviews see Bauer (1990), Greene (1993, 1999) and Kumbhakar and Lovell (2000).

[2] See Ritter and Simar (1997) for some disturbing properties of the normal-*gamma* model.

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.04.001

Given the CF two alternatives are proposed. *First*, the density is obtained from the CF by using the FFT as outlined above. *Second*, Fourier methods of inference are applied directly which exploit the asymptotic normal form of the log-likelihood function and have, at least in theory, asymptotic efficiency close to the Cramer–Rao bound. These methods of inference bypass completely the problem of obtaining the density of composed error term, and are asymptotically equivalent to estimation by the method of ML.

Another class of stochastic frontier models that are empirically relevant but estimation by direct use of ML is impossible is the class of models with input-oriented (IO) technical inefficiency which is rather common in theoretical discussions but is never implemented in practice with the exception of Kumbhakar and Tsionas (2006) who proposed the use of MSL. In this paper it is shown that ML can be implemented using the FFT and simulation can be avoided. Yet another class of models includes two-tiered stochastic frontier models introduced by Polachek and Yoon (1987, 1996) which are useful in modeling informational asymmetries in the labor and other markets. Estimation of such models has been restricted to the class of exponential distributions for the one-sided error components. In this paper it is shown that the characteristic function can be fruitfully employed to obtain ML parameter estimates in the class of *gamma* distributed error components. Moreover, we propose and estimate latent class models with *gamma* distributed inefficiencies and models with two-sided error components in the stable Paretian class and, again, *gamma* distributed inefficiencies.

The remaining part of the paper is organized as follows. ML estimation using the FFT is introduced in Section 2. In Section 3 we consider estimators that are simpler than ML-FFT but retain asymptotic efficiency. An empirical application to US commercial banks is described in Section 4. In Section 5 we describe FFT methods for ML estimation in the context of IO stochastic frontier models and in Section 6 we provide an empirical application for US electric utilities. In Section 7 we take up ML estimation in the two-tiered stochastic frontier model with gamma distributed one-sided error components. In Section 8, we consider the latent class models with *gamma* distributed inefficiencies and models with two-sided error component in the stable Paretian class. In Appendix B, special attention is devoted to output-oriented frontiers with truncated-normal inefficiency and various estimating techniques are implemented and compared using both real and artificial data.

2. ML estimation of output-oriented stochastic frontier models using the FFT

Consider the stochastic cost frontier model[3]

$$y_i = x_i'\beta + v_i + u_i, \quad i = 1, \ldots, n, \tag{1}$$

where x_i is a $k \times 1$ vector of explanatory variables, β is a $k \times 1$ vector of parameters, $v_i \overset{IID}{\sim} N(0, \sigma_v^2)$ and u_is are distributed *iid* according to a distribution with density $f(u_i; \gamma)$ with support $\mathbb{R}_+$ depending on the unknown parameters γ. The two error components are mutually independent and independent of x_is. The probability density function of the composed error, $\varepsilon_i = v_i + u_i$, is given by $f_\varepsilon(\varepsilon; \gamma) = \int_0^\infty (2\pi\sigma_v^2)^{-1/2} \exp\left[-\frac{(\varepsilon-u)^2}{2\sigma_v^2}\right] f_u(u; \gamma)du$. This density is needed in estimation of the parameters by ML but in certain instances it is not available in closed form. One such instance is the model proposed by Greene (1990), viz. when the one-sided error follows a *gamma* distribution with density

$$f_u(u; \alpha, \theta) = \theta^{-\alpha}\Gamma(\alpha)^{-1}u^{\alpha-1}\exp(-u/\theta), \quad \alpha, \theta > 0, u \geq 0.$$

It is known that $\mathbb{E}(u) = \alpha\theta$ and $\mathbb{V}\mathrm{ar}(u) = \alpha\theta^2$. Beckers and Hammond (1987) showed that the likelihood can be computed using special functions in the general case when α is not an integer. Greene (2003) has proposed the method of maximum simulated likelihood (MSL) in this context. The troublesome term in the likelihood function is $\mathbb{E}(z^{\alpha-1})$ when $z \sim N(\varepsilon_i - \theta^{-1}\sigma^2, \sigma^2)$ truncated below at zero ($z \geq 0$). Setting up the simulated likelihood involves obtaining a (potentially large) number of independent draws from this distribution and taking their empirical average to approximate the expectation. Although the approach is conceptually simple the problem is that, unfortunately, this has to be repeated for every observation in the sample. It is evident that if simulation can be avoided in the computation of the likelihood function, then it is never a method of choice.

To proceed, for a random variable X with distribution whose density is $f_X(x)$, its CF is defined as $\varphi_X(t) = \mathbb{E}[\exp(Itx)] = \int_{-\infty}^{\infty}\exp(Itx)f_X(x)dx$, for all $t \in \mathbb{R}$. Here, $I = \sqrt{-1}$ denotes complex unity. In the context of the normal-*gamma* stochastic frontier model, the CF of the two-sided error is $\varphi_v(t) = \exp\left(-\frac{1}{2}\sigma_v^2t^2\right)$, and the CF of the one-sided error is $\varphi_u(t) = (1 - It\theta)^{-\alpha}$. Therefore, the CF of the composed error is

$$\varphi_\varepsilon(t) = \exp\left(-\frac{1}{2}\sigma_v^2t^2\right)(1 - It\theta)^{-\alpha}, \quad t \in \mathbb{R}, \tag{2}$$

which is of particularly simple form and involves no special functions. Of course, it is a complex-valued function.

Another example of interest includes the *beta* distribution. If $r = \exp(-u)$ denotes cost efficiency it is natural to assume[4] that this is *beta* distributed with parameters p and q, and density $f_r(r; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}r^{p-1}(1 - r)^{q-1}$, $0 \leq r \leq 1$, $p, q > 0$. Technical inefficiency is $u = -\ln(r)$. The density of u can be obtained easily but its convolution with v does not yield a closed form expression. The CF of r is given by $\varphi_r(t) = {}_1F_1(p; p + q; It)$ where ${}_1F_1(a; b; z)$ is the confluent hypergeometric function of the first kind. The CF of u has a simpler form and it is easy to show that $\varphi_u(t) = \frac{B(p-It,q)}{B(p,q)}$, where $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is the *beta* function. The CF of the composed error in this case is

$$\varphi_\varepsilon(t) = \exp\left(-\frac{1}{2}\sigma_v^2t^2\right)\frac{B(p - It, q)}{B(p, q)}, \quad t \in \mathbb{R}. \tag{3}$$

Next we turn attention to implementing ML estimation. Suppose $f_\varepsilon(\varepsilon)$ is the density of the composed error and $\varphi_\varepsilon(t)$ is the corresponding CF shown in (2) or (3). By Levy's inversion theorem (Lukacs and Laha, 1964, p. 20) we have

$$f_\varepsilon(\varepsilon) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-I\varepsilon t)\varphi_\varepsilon(t)dt.$$

Therefore, the density can be obtained from the CF using an integral. Since the integral is not available in closed form it is natural to seek out a way to approximate it efficiently and accurately. Let $N = 2^{N'}$, and define a sequence of the form $x_k = (k - 1 - \frac{N}{2})h$, $k = 1, \ldots, N$, for some step length $h > 0$. If we set $t = 2\pi\omega$ and change variables then

$$f_\varepsilon(x_k) = \int_{-\infty}^{\infty}\varphi_\varepsilon(2\pi\omega)\exp(-I2\pi\omega x_k)d\omega.$$

This integral can be approximated using the trapezoidal rule to obtain

$$f_\varepsilon(x_k) \approx s\sum_{n=1}^{N}\varphi_\varepsilon(2\pi\omega_n)\exp(-I2\pi\omega_n x_k),$$

[3] A production frontier would have $v - u$ in place of $v + u$ as the composed error term.

[4] It is well known that the *beta* distribution is a quite flexible model for describing a random variable whose support is a bounded interval.

where $\omega_n = (n - 1 - \frac{N}{2})s$ and $s = (hN)^{-1}$. Therefore, we get

$$f_\varepsilon(x_k) \approx s(-1)^{k-1-(N/2)} \sum_{n=1}^{N} (-1)^{n-1} \varphi_\varepsilon(2\pi\omega_n) \times \exp\left(\frac{-I2\pi(n-1)(k-1)}{N}\right), \quad (4)$$

which can be computed efficiently using the FFT applied to the sequence $(-1)^{n-1}\varphi_\varepsilon(2\pi\omega_n)$, $n = 1, \ldots, N$. The (discrete) Fourier transform is a mapping of a (possibly complex) vector $(z_0, \ldots, z_{N-1})$ to a vector $(Z_0, \ldots, Z_{N-1})$, with $Z_i = \sum_{j=0}^{N-1} z_j \exp\left(-Ii\frac{2\pi j}{N}\right), i = 0, 1, \ldots, N-1$. An increase in efficiency can be achieved if $N = 2^{N'}$, that is when the dimension N is a power of 2. For more details, see Duhamel and Vetterli (1990).[5]

Below we provide generic MATLAB code to compute the FFT. The code needs a subprogram, *PHI*(t) that provides the value of the characteristic function.

```
function [x, p] = pdf(n, h)
n = 2^n;
x = (0:n-1)'*h-n*h/2;
s = 1/(h*n);
t = 2*pi*s*((0:n-1)'-n/2);
sgn = ones(n,1);
sgn(2:2:n) = -1*ones(n/2,1);
phi = sgn.*PHI(t);
p = s.*abs(fft(phi));
```

To obtain the density at the grid points x_k we have to normalize the transform by $s(-1)^{k-1-(N/2)}$. This procedure will provide the density at the grid points but *not* at points we actually need when, for example, FFT becomes part of a numerical optimization scheme that carries out ML. We can apply quadratic interpolation to approximate the density at the desired point. For a wide selection of parameter values I have found that this procedure preserves 8–9 significant digits so it can be used safely. In MATLAB, one case use `f = interp1(x, p, data)` to get the density `f` at the observed data `data` given the output `x` and `p` from the FFT.

To implement the FFT we need to choose N' and h. In practice I set $N' = 12$ (which implies that the density is approximated at $N = 2^{12} = 4096$ points). The selection of h can be based on a reasonable value for the maximum absolute residual. This value cannot be known in advance since in the course of ML iterations some residual can become larger in absolute value than the bound. In that case interpolation will be inaccurate and ML can break down. It has been found that a choice that works well is $h = 2\bar{e}/N$, where $\bar{e} = \max_{i=1,\ldots,n} |e_i|$.

The FFT method does not need large samples to work. Its purpose is to approximate intractable densities for *any* sample size. What is needed is a reasonably large value of N' to approximate the density well. Given this approximation, ML based on the FFT will provide the same results with ML based on the exact density or other approximations of the density, like for example maximum simulated likelihood (MSL). Moreover, the FFT method relies on specification of N' and h. It is not quite the case that h is like the bandwidth parameter in nonparametric kernel estimation. Given a large value for N' (say 12 or 16) if h is selected so that the resulting x_ks cover the support of the density, the FFT approximation to the density is quite accurate. This fact is known from FFT approximations to the density of stable Paretian laws. In my computational experience with models reported here, selecting $h = 2\bar{e}/N$ always produced accurate results.

It should be mentioned that once the density has been approximated using the FFT, computation of firm-specific technical inefficiency is possible. The well known Jondrow et al. (1982) measure can be computed from $\mathbb{E}(u|\varepsilon) = \frac{(2\pi\sigma_v^2)^{-1/2} \int_0^\infty u \exp\left[-\frac{(\varepsilon-u)^2}{2\sigma_v^2}\right] f(u|\gamma)du}{f_\varepsilon(\varepsilon)}$. The denominator is available from the FFT and the integral in the numerator can be computed easily using quadrature or Simpson's rule.

It is important to notice that heteroskedasticity can be treated using the techniques we have already proposed. If the one-sided component follows a half-normal or exponential distribution, the density of the composed error is available in closed form and implementation of ML is straightforward. If not, then we have to use the Fourier transform or other techniques. Consider the case of a normal-gamma model with heteroskedasticity: $v_i \sim IN\left(0, \sigma_{v,i}^2\right)$ with $\sigma_{vi}^2 = \exp\left(z_i'\gamma\right)$ and independently $u_i \overset{IID}{\sim} Gamma\,(\alpha, \theta)$. The characteristic function of the composed error $\varepsilon_i = v_i + u_i$ is $\varphi_{\varepsilon,i}(t) = \exp\left(-\frac{1}{2}\sigma_{v,i}^2 t^2\right)(1 - It\theta)^{-\alpha}$. *For each observation*, we can use the Fourier transform to obtain the density as $f_{\varepsilon_i}(\varepsilon) = (2\pi)^{-1} \int_{-\infty}^{\infty} \exp(-I\varepsilon t)\varphi_{\varepsilon,i}(t)dt$ over a grid of points for ε, and then compute $f_{\varepsilon_i}(\varepsilon_i)$ by interpolation, with $\varepsilon_i = y_i - x_i'\beta$. Alternatively, we can use MSL or Gaussian quadrature to obtain the density as $f_{\varepsilon_i}(\varepsilon_i) = \left(2\pi\sigma_{v,i}^2\right)^{-1/2} \theta^{-\alpha}\Gamma(\alpha)^{-1} \int_0^\infty u^{\alpha-1} \exp\left[-\frac{(\varepsilon_i-u)^2}{2\sigma_{v,i}^2} - \theta^{-1}u\right] du$, where $\sigma_{vi}^2 = \exp\left(z_i'\gamma\right)$.

To validate the FFT, in Appendix A we undertake a detailed comparison with conventional ML estimation for an exponential stochastic frontier model. For the exponential model, the likelihood function is available in closed form so a direct comparison is possible. The results are very encouraging in the sense that the FFT provides quite accurate approximations to the density and the finite sample distributions of ML estimators based on FFT or the exact likelihood are quite close.

3. Asymptotically normal form procedures

Next we turn attention to methods that are asymptotically as efficient as ML but are computationally less intensive. Given the CF we can implement estimation procedures *that attain the Cramer–Rao lower bound* without having to carry out the FFT to obtain the density and compute the ML estimate. In that sense such procedures will be called *efficient*. One such procedure has been proposed by Feuerverger and McDunnough (1981a,b,c) and Feuerverger and Mureika (1977). It involves fitting the theoretical to the empirical CF using nonlinear generalized least squares.[6] This procedure is also known as *k–L* and involves maximum likelihood based on the asymptotic distribution at k points of the empirical characteristic function. Suppose $\varphi_\varepsilon(t)$ denotes the theoretical CF of the composed error and its empirical counterpart is

$$\tilde{\varphi}_\varepsilon(t) = N^{-1} \sum_{i=1}^{N} \exp(Ite_i) = N^{-1} \sum_{i=1}^{N} \cos(te_i) + I \cdot N^{-1} \sum_{i=1}^{N} \sin(te_i), \quad (5)$$

[5] See also Mittnik et al. (1999).

[6] For an application of the empirical characteristic function to finite mixtures of normals, see Tran (1998).

where the e_is are observed. We remark again that $I = \sqrt{-1}$ is the complex unity. Given a grid of points $\{t_j, j = 1, \ldots, J\}$ consider the real and imaginary parts of the empirical CF,

$$\mathrm{Re}\tilde{\varphi}_\varepsilon(t_j) = N^{-1}\sum_{i=1}^{N}\cos(t_j e_i), \quad j = 1, \ldots, J,$$

$$\mathrm{Im}\tilde{\varphi}_\varepsilon(t_j) = N^{-1}\sum_{i=1}^{N}\sin(t_j e_i), \quad j = 1, \ldots, J,$$

and collect them in the vector $\tilde{Y} = [\mathrm{Re}\tilde{\varphi}_\varepsilon(t_1), \ldots, \mathrm{Re}\tilde{\varphi}_\varepsilon(t_J), \mathrm{Im}\tilde{\varphi}_\varepsilon(t_1), \ldots, \mathrm{Im}\tilde{\varphi}_\varepsilon(t_J)]'$. Similarly we can consider the real and imaginary parts of the theoretical CF, viz.

$$\mathrm{Re}\varphi_\varepsilon(t) = \mathbb{E}[\cos(te)] = \int_{-\infty}^{\infty}\cos(te)f_\varepsilon(e|\Theta)de,$$

and

$$\mathrm{Im}\varphi_\varepsilon(t) = \mathbb{E}[\sin(te)] = \int_{-\infty}^{\infty}\sin(te)f_\varepsilon(e|\Theta)de,$$

and collect them[7] in the vector $Y = Y(\Theta)$ where Θ denotes all unknown parameters. Suppose $Y(\Theta) = [Y_h(\Theta), h = 1, \ldots, 2J]$ and $\tilde{Y} = [\tilde{Y}_h, h = 1, \ldots, 2J]$, and consider the artificial regression

$$\tilde{Y}_h = Y_h(\Theta) + \xi_h, \quad h = 1, \ldots, 2J.$$

This regression can be estimated using nonlinear least squares to obtain consistent estimators for Θ, as long as $J \geq \frac{1}{2}\dim(\Theta)$. We call this the first-stage asymptotic normal form estimator (*1SANF*). The estimator was called "*harmonic regression*" by Feuerverger and McDunnough (1981a). The second-stage asymptotic normal form estimator (*2SANF*) attains asymptotic efficiency that can be made arbitrarily close to the Cramer–Rao lower bound by selection of the t_js. To describe the estimator known as *k–L* procedure, the asymptotic form of the log-likelihood may be taken as

$$\ln L(\Theta) = const. + \frac{1}{2}\log|\Sigma^{-1}(\Theta)| - \frac{N}{2}[Y(\Theta) - \tilde{Y}]'\Sigma^{-1}(\Theta)[Y(\Theta) - \tilde{Y}] \quad (6)$$

or as just the last term, where $N^{-1}\Sigma(\Theta)$ is the covariance matrix of $\tilde{Y}$, and the elements of $\Sigma(\Theta)$ are given in Eq. (2.2) of Feuerverger and McDunnough (1981a).[8] These elements depend only on the real and imaginary parts of the theoretical characteristic function. The asymptotic normal form can be treated as the log-likelihood function of any stochastic frontier model even when the density function of the composed error cannot be obtained in closed form. Maximization of this function is equivalent to minimizing $[Y(\Theta) - \tilde{Y}]'\Sigma^{-1}(\Theta)[Y(\Theta) - \tilde{Y}]$ or the quadratic form $[Y(\Theta) - \tilde{Y}]'\tilde{\Sigma}^{-1}[Y(\Theta) - \tilde{Y}]$ where $\tilde{\Sigma}$ is any consistent estimator of $\Sigma^{-1}(\Theta)$.

Computational experience suggests the following practical scheme. *First*, implement 1SANF which provides estimates that are consistent but not asymptotically efficient. *Second*, based on the final estimates of 1SANF compute an estimate of Σ and use it to compute the 2SANF estimator. It has been found that if parameters and Σ are updated simultaneously in 2SANF then the Σ matrix can become non-positive-definite numerically[9] and convergence of the numerical optimization algorithm becomes impossible. Since the estimate of Σ obtained from the 1SANF estimator is consistent, the asymptotic distribution of the 2SANF estimator is the same no matter whether parameters and Σ are updated simultaneously or not.

Finally, it should be mentioned that the efficiency result based on characteristic function procedures is asymptotic and therefore in finite samples some differences with ML are expected to arise. Indeed Madan and Seneta (1987) through a simulation study cast some doubt on the good performance of characteristic function procedures in small samples. We will see that the same message emerges from the present paper as well.

In connection with (5) and (6) the e_is are not observed. Consider $y_i = \beta_0 + x_i'\gamma + \varepsilon_i$, where $\varepsilon_i = v_i + u_i$. Then, we can use $e_i = y_i - x_i'\tilde{\gamma}$, where $\tilde{\gamma}$ is the LS estimator of the slope coefficients. Using (5) and (6) we can estimate the intercept β_0 and the shape and scale parameters of the distribution of ε_i, viz. σ_v^2, α and θ for the normal-*gamma*, and σ_v^2, p and q for the normal-*beta* model. An alternative is the efficient GMM technique of Carrasco and Florens (2002) who use a continuum of moment conditions. For an excellent review of this and other techniques in the context of the empirical CF, see Yu (2004). For an examination of GMM with a continuum of moments in the context of truncated normal stochastic frontiers, see Appendix B.

4. Empirical application for output-oriented stochastic frontier model

The new techniques are illustrated using a cost function for US commercial banks. The data for this study is taken from the commercial bank and bank holding company database managed by the Federal Reserve Bank of Chicago. It is based on the Report of Condition and Income (Call Report) for all US commercial banks that report to the Federal Reserve banks and the FDIC. In this paper I used the data for the year 2000 and selected a random sample of 500 commercial banks. The data set has been previously employed by Kumbhakar and Tsionas (2005a,b).[10]

The five output variables are as follows: Installment loans to individuals for personal/household expenses (y_1), real estate loans (y_2), business loans (y_3), federal funds sold and securities purchased under agreements to resell (y_4), other assets (assets that cannot be properly included in any other asset items in the balance sheet) (y_5). The five input variables are as follows: labor, capital, purchased funds, interest-bearing deposits in total transaction accounts, and interest-bearing deposits in total non-transaction accounts. For each input the price is obtained by dividing total expenses on it by the corresponding input quantity. Thus, for example, the price of labor (w_1) is obtained from expenses on salaries and benefits divided by the number of full time employees. The same approach is used to obtain w_2 through w_5. Total cost is then defined as the sum of the expenses on these five inputs. To impose the linear homogeneity restrictions, total cost and all the prices are normalized with respect to w_5. A Cobb–Douglas cost function has been used.

[7] Typically, these are not available in analytic form because the integrals cannot be evaluated in closed form. However, they can be computed numerically very easily.

[8] It should be noted that since $N^{-1}\Sigma$ is the covariance matrix of $\tilde{Y}$, and the elements of Σ are readily available, the problem of heteroskedasticity in "harmonic regression" or 1SANF can be dealt with in a straightforward manner. In the *k–L* or 2SANF procedure, this is automatically done.

[9] The same computational experience is suggested by several other authors, see Feuerverger and McDunnough (1981a,b,c) and Madan and Seneta (1987). Madan and Seneta (1987) suggest using a generalized inverse of the Σ matrix. In the applications considered here this has been followed but it has not been found necessary since a simple inverse sufficed.

[10] Clearly, use is made of the well known DR Y-9SP bank holding company data on large commercial banks. This is a smaller population, than the FDIC Statistics in Depository Institutions database. Moreover, it is important to mention that no violations of the monotonicity conditions (evaluated at each data point) have been noted in our estimations.

An important difference of ML based on FFT and 2SANF is that the former provides estimates of all parameters whereas the latter can be used to estimate only the intercept of the cost function and the shape or scale parameters in the distributions of one- and two-sided error terms.[11] Greene (1999) rightly argues that this is not a serious restriction since one can use LS to estimate cost function parameters other than the intercept and *one is mainly interested in the shape and scale parameters that are crucial in estimating technical inefficiency*. To implement 1SANF the grid of points for the computation of the characteristic function is a set of 100 equi-spaced points ranging from 10^{-4} to 40. Based on the converged parameter estimates the matrix $\tilde{\Sigma}$ is computed and is used to implement 2SANF which is minimized with a conjugate gradient technique. Standard errors are computed from $N^{-1}\left(G'\tilde{\Sigma}^{-1}G\right)^{-1}$ where $G = \frac{\partial Y(\Theta)}{\partial \Theta'}$ computed at the final estimates. The 2SANF estimates are obtained using Newton's method with starting values obtained from 1SANF until convergence is obtained. In practice the numerical difference between the two estimators has not been found appreciable. Finally, twice the number of grid points times the 2SANF objective function, viz. $[Y(\Theta) - \tilde{Y}]'\tilde{\Sigma}^{-1}[Y(\Theta) - \tilde{Y}]$ is Hansen's J-statistic which follows a chi-square with degrees of freedom $2\dim(Y(\Theta)) - \dim(\Theta)$. This can be used as a test for the number of over-identifying restrictions implied by 2SANF.

To implement ML based on FFT I have used $N = 2^{12} = 4096$ grid points for the characteristic function and step length $h = 2\max|e_i|/N$ where e_i is the cost function residual computed at the current parameter estimate. The step length is updated whenever a new parameter vector is given to the log-likelihood function. Starting values were obtained from OLS (for the slope parameters) and the 2SANF estimator (for all other parameters).

The empirical results are reported in Table 1. We see[12] that output cost elasticity is close to one suggesting nearly constant returns to scale, and standard errors from ML-FFT are in general smaller than standard errors from OLS. I have also tried when $N' = 12$ and $N' = 16$. These results are very nearly the same so – at least in this application – using $N' = 10$ is sufficient. Notice the great timing differences when $N' = 16$. The case $N' = 10$ took only 0.67 min on a PC with 800 MHz speed. Results using MSL (with 5000 replications) produced nearly the same results and are not reported to save space. Timing was 7.12 min, that is much larger compared to the FFT.

The estimate of α from ML-FFT is quite low, 0.379 with standard error 0.262 so it appears that this parameter is less than one, at least based on a 95% asymptotic confidence interval. Based on this estimate an exponential distribution for the one-sided error seems relatively implausible so it is worth the effort to estimate the normal-gamma model. In Table 2 reported are the estimation results from ML-FFT estimation of the normal-*beta* model for US commercial banks. I have used $N' = 12$ and a similar way to set up the t-grid as in the normal-*gamma* case. The parameter estimates for β_js are almost identical to those obtained for the normal-*gamma* model. The estimated shape parameters are $p = 3.888$ and $q = 0.374$ suggesting average technical efficiency (measured by $p/(p+q)$) close to 91.2%. The normal-*gamma* model implies average efficiency close to 89.5% so the two models give more or less the same implications as far as mean efficiency. The mean log-likelihood functions for the two models are very close (0.106383 for normal-*beta* and 0.106371 for normal-*gamma*) so since the two models have the same number of parameters it would be difficult, at least in this application, to discriminate among them using standard information criteria. However, the log-likelihood values show a – very slight – preference for the normal-*beta* model. *As a conclusion, application of the FFT in ML estimation is quite feasible and provides reliable results for quite complicated models like the normal-gamma and normal-beta whose likelihood function cannot be*

Table 1
Parameter estimates for normal-*gamma* cost function of US banks.

	OLS	2SANF	ML-FFT
Constant	1.333 (0.290)	1.087 (0.0097)	1.043 (0.249)
$\log w_1$	0.237 (0.044)	–	0.199 (0.038)
$\log w_2$	0.064 (0.018)	–	0.061 (0.015)
$\log w_3$	0.170 (0.029)	–	0.110 (0.025)
$\log w_4$	0.043 (0.024)	–	0.043 (0.024)
$\log y_1$	0.055 (0.015)	–	0.053 (0.014)
$\log y_2$	0.420 (0.015)	–	0.463 (0.014)
$\log y_3$	0.179 (0.021)	–	0.229 (0.020)
$\log y_4$	0.056 (0.008)	–	0.043 (0.007)
$\log y_5$	0.236 (0.020)	–	0.160 (0.019)
σ_v	–	0.174 (0.0097)	0.162 (0.014)
α	–	0.142 (0.057)	0.379 (0.262)
θ		0.490 (0.168)	0.276 (0.097)

Notes: Standard errors appear in parentheses. OLS denotes ordinary least squares, 2SANF denotes two-stage asymptotically normal form estimator and ML-FFT stands for maximum likelihood estimation based on the fast Fourier transform. The variance of the two-sided error term is σ^2, while α and θ denote the shape and scale parameter of the gamma distribution for the one-sided error term respectively. The density of gamma distribution is $f_u(u;\alpha,\theta) = \theta^{-\alpha}\Gamma(\alpha)^{-1}u^{\alpha-1}\exp(-u/\theta)$. The Hansen J-test is 95.50. The log-likelihood function from ML-FFT was 53.15. Execution time for ML-FFT was 2.29 min on a Pentium 800 MHz machine running Windows 98 under WinGauss 3.2.38. ML-FFT has been tried for $N' = 6$ but it was not possible to obtain convergence within 200 iterations. Computing times were 0.67, 2.29 and 212.3 min for $N' = 10$, 12 and 16 respectively. The results for $N' = 10$ and 16 are not reported since they are identical (to three decimal places) to results reported here for $N' = 12$. Computing time is CPU time in minutes required to obtaining full convergence. It excludes computing time required for obtaining second derivatives for approximation of asymptotic standard errors. MSL and ML based on quadrature have also been implemented. We have used 5000 replications for MSL and 40 base points for Gaussian quadrature. Timings were 231.12 and 1.97 respectively.

Table 2
Maximum likelihood FFT parameter estimates for cost function of US commercial banks using the normal-*beta* model.

	OLS	ML-FFT
Constant	1.333 (0.290)	1.043 (0.249)
$\log w_1$	0.237 (0.044)	0.199 (0.038)
$\log w_2$	0.064 (0.018)	0.061 (0.015)
$\log w_3$	0.170 (0.029)	0.110 (0.025)
$\log w_4$	0.043 (0.024)	0.043 (0.024)
$\log y_1$	0.055 (0.015)	0.053 (0.014)
$\log y_2$	0.420 (0.015)	0.463 (0.014)
$\log y_3$	0.179 (0.021)	0.229 (0.020)
$\log y_4$	0.056 (0.008)	0.043 (0.007)
$\log y_5$	0.236 (0.020)	0.160 (0.019)
σ	0.23	0.163 (0.014)
p	–	3.888 (1.173)
q	–	0.374 (0.261)

Notes: Standard errors appear in parentheses. OLS denotes ordinary least squares estimates which are provided for convenience. The variance of the two-sided error term is σ^2, and p and q denote the two shape parameters of the *beta* distribution for the one-sided error term. Execution time for ML-FFT was 1.87 min on a Pentium 800 MHz machine running Windows 98 under WinGauss 3.2.38. The mean log-likelihood is 0.106383. Execution time for MSL using 5000 replications was 12.12 min excluding the time to generate random numbers.

[11] It is possible to remove this deficiency of 2SANF by using the conditional (on the x_i) CF. The results were largely the same so I opted for presenting the simpler technique to focus on the substantive issues. The most striking feature of Table 1 is the difference in inefficiency estimates between 2SANF and ML-FFT. We have $E(u) = 0.069$ for 2SANF and $E(u) = 0.104$ for ML-FFT. The difference can be attributed to the placement and selection of grid points and results may differ more in other data sets. This plagues techniques that are based on characteristic functions and placement of grid points but not techniques based on the FFT provided N' is large and h is selected to cover the support.

[12] The column headed IO means "input-oriented". IO frontier models are discussed in Section 5.

Table 3
Various one-step efficient 2SANF estimates for cost function of US commercial banks.

	Case 1	*Case* 2	*Case* 3	*Case* 4	*Case* 5	*Case* 6
Constant	1.096 (0.097)	0.890 (167.76)	1.097 (0.289)	1.097 (0.170)	1.097 (0.163)	1.097 (0.154)
σ	0.186 (0.055)	0.152 (31.08)	0.178 (0.123)	0.178 (0.084)	0.178 (0.082)	0.178 (0.080)
α	0.083 (0.453)	2.157 (2271.9)	0.086 (1.47)	0.086 (0.816)	0.086 (0.769)	0.086 (0.717)
θ	0.632 (3.364)	0.113 (41.51)	0.786 (16.18)	0.782 (8.47)	0.782 (8.154)	0.782 (7.452)

Notes: Standard errors multiplied by $\sqrt{N}$ in parentheses. The specifications for the different cases correspond to differences in the grid points for the characteristic functions and they are as follows.
Case 1. 20 points in the interval from 0.0001 to 10.
Case 2. 20 points in the interval from 0.0001 to 3.
Case 3. 20 points in the interval from 0.0001 to 20.
Case 4. As in Case 3 with 50 points.
Case 5. As in Case 3 with 100 points.
Case 6. As in Case 3 with 200 points.

expressed in closed form. Execution times are trivial for sample sizes typically encountered in econometric applications.

In Table 3 reported are results from the one-step efficient version of the 2SANF estimator (viz. one iteration away from the 1SANF result) for different configurations of the grid points. The general rule is that arbitrarily high asymptotic efficiency requires a sufficiently fine and extended grid so the results in the last column headed Case 6 should be the most reliable. *Parameter estimates and their standard errors appear to be sensitive on placement of the grid points*. When the grid is not extended (Cases 1 and 2) parameter estimates differ compared to what they would be by extending the grid and standard errors can be huge (Case 2). If the grid is sufficiently extended (from 0.0001 to 20 is a reasonable choice) standard errors seem to converge provided we have more than 50 points but in certain parameters like θ standard errors can differ more than 10% if we compare 50 and 200 points (Cases 4 and 6). To conclude the comparison, it appears that the finite sample properties of 2SANF and ML can be very different and obtaining correct parameter estimates with the 2SANF requires some experimentation with placing the grid points.

Results using the fully converged 2SANF estimator for various choices for the grid points are reported in Table 4. The ML-FFT standard errors are greater than the corresponding standard errors from 2SANF for all parameters except θ and there are notable differences in parameter estimates for θ and P. This should be attributed to the small-sample behavior of 2SANF. The standard errors for θ are nearly 50% lower compared with the one-step-efficient 2SANF estimator and estimates of the shape parameter α almost double bringing them closer to the ML result while their standard errors are also much lower than the one-step-efficient 2SANF case. This clearly suggests the gain from iterating 2SANF to convergence and of course the gains that can be expected from application of ML estimation through the FFT. It is conjectured that updating the covariance matrix in the course of iterations of 2SANF would also produce some efficiency gains relative to fully converged 2SANF with the covariance held fixed at the estimate produced from 1SANF. Unfortunately, I have not been able to implement this estimator since the covariance matrix behaves very badly.[13]

Before proceeding it should be mentioned that a potentially important inefficiency distribution, the truncated normal, has been left without discussion so far. Although the likelihood function is available in closed form and therefore ML can be implemented without trouble, in practice many numerical problems are encountered (see Greene (1999) and Tsionas (2001a)), so it is useful to consider alternative estimators based on the CF. A detailed examination is contained in Appendix B of this paper. The conclusion is that although various moment estimators and estimators based on the ANF are easy to implement, in practice the configuration of grid points is a non-trivial issue and certain matrices involved in the computation are very badly conditioned. From that point of view the truncated-normal and the numerical problems associated with implementing estimation and inferences are very much an open issue.

Table 4
Various fully converged 2SANF estimates for cost function of US commercial banks.

	Case 1	*Case* 2	*Case* 3	*Case* 4
Constant	1.087 (0.217)	1.087 (0.276)	1.088 (0.202)	1.087 (0.217)
σ	0.174 (0.095)	0.174 (0.105)	0.174 (0.090)	0.174 (0.095)
α	0.144 (1.270)	0.137 (1.624)	0.136 (1.149)	0.143 (1.267)
θ	0.488 (3.717)	0.509 (5.412)	0.513 (3.860)	0.490 (3.760)

Notes: Standard errors multiplied by $\sqrt{N}$ in parentheses. The specifications for the different cases correspond to differences in the grid points for the characteristic functions and they are as follows.
Case 1. 50 points in the interval from 0.0001 to 20.
Case 2. 100 points in the interval from 0.0001 to 20.
Case 3. 200 points in the interval from 0.0001 to 20.
Case 4. 100 points in the interval from 0.0001 to 40. Using 200 points for this case did not produce results due to problems with numerical optimization owing to the bad conditioning of the covariance matrix.

5. Stochastic production frontiers with input oriented technical inefficiency

Production frontiers are invariably estimated under the assumption that technical inefficiency is output-oriented (OO) contrary to theoretical investigations which more often than not are founded on the assumption of input-oriented (IO) efficiency. Kumbhakar and Tsionas (2006) propose a translog production frontier[14] with IO technical efficiency and show that parameter estimates can be obtained by the method of MSL. In this section we show that generating function methods can be implemented, their application is quite easy, the results are robust, and there is no need to resort to simulation to compute the likelihood function in implementing estimation by the ML method.

The production technology with the IO measure of technical inefficiency can be expressed as

$$Y_i = f(\Theta_i X_i), \quad i = 1, \ldots, n,$$

[13] An idea is to iterate the 2SANF estimator itself when the covariance matrix is not updated jointly with the parameters but it is held fixed to the value taken when computed at the previous vector of parameters. Since the difference between 1SANF and 2SANF is not appreciable I did not give further consideration to implementing this scheme.

[14] They have also derived expressions for the cost and profit functions corresponding to production functions with IO technical inefficiency and they have shown that they have the same functional form as the translog. So restricting attention to the production function case involves no real loss of generality.

where Y_i is a scalar output, $X_i = (X_{1i}, \ldots, X_{Ji})$ is the vector of inputs actually used, $\Theta_i = X_{ji}^e/X_{ji} \leq 1$ for $j = 1, \ldots, J$, is input-oriented efficiency (a scalar), and $X_i^e = \Theta_i X_i$ is the input vector in efficiency units. The IO technical inefficiency for firm i is defined as $1 - \Theta_i = (X_{ji} - X_{ji}^e)/X_{ji}$ for $j = 1, \ldots, J$ and is interpreted as the rate at which all the inputs could be reduced without reducing output. If we use a lower case letter to indicate the log of a variable, and assume that the production function has a translog form then it is easy to show that the production function can be written as

$$y_i = \beta_0 + x_i'\beta + \frac{1}{2}x_i'\Gamma x_i + \beta_T T_i + \frac{1}{2}\beta_{TT}T_i^2 + x_i'\varphi T_i + g(\theta_i, x_i, T_i) + v_i, \quad (7)$$

where y_i is the log of output, θ_i is IO technical inefficiency, 1_J denotes the $J \times 1$ vector of ones, x_i is the $J \times 1$ vector of inputs in log terms, T_i is the trend/shift variable,[15] β_0, β_T and β_{TT} are parameters, β, φ are $J \times 1$ parameter vectors, and Γ is a $J \times J$ symmetric matrix containing parameters. Here, λ and θ are non-negative. Finally, we denote $g(\theta_i, x_i, T_i) = [\frac{1}{2}\theta_i^2\Psi - \theta_i\Xi_i]$, $\Psi = 1_J'\Gamma 1_J$, and $\Xi_i = 1_J'(\beta + \Gamma x_i + \varphi T_i)$, $i = 1, \ldots, n$.

Notice that if the production function is homogeneous of degree r, then $\Gamma 1_J = 0$, $1_J'\beta = r$, and $1_J'\varphi = 0$. In such a case the $g(\theta_i, x_i)$ function becomes a constant multiple of θ_i, (viz., $[\frac{1}{2}\theta_i^2\Psi - \theta_i\Xi_i] = -r\,\theta_i$), and consequently, the IO model cannot be distinguished from the OO model. In the non-homogeneous case, the (negative of the) $g(\theta_i, x_i)$ function shows the percent by which output is lost due to technical inefficiency. For a well-behaved production function we should have $g(\theta_i, x_i) \leq 0$ for each i.

The expression under the first bracket in (5) is the standard translog production function and g_i is similar to $-u_i$ in an OO model. We write (7) more compactly as

$$y_i = z_i'\alpha + \frac{1}{2}\theta_i^2\Psi - \theta_i\Xi_i + v_i, \quad i = 1, \ldots, n. \quad (8)$$

Both Ψ and Ξ_i are functions of the parameters, and Ξ_i also depends on the regressors (x_i and T_i). Under the assumption that $v_i \sim IN(0, \sigma_v^2)$, and θ_i is distributed independently of v_i, according to a distribution with density $f_\theta(\theta_i; \omega)$ where ω is a parameter, the distribution of y_i has density

$$f(y_i; \phi) = (2\pi\sigma_v^2)^{-1/2} \times \int_0^\infty \exp\left[-\frac{\left(y_i - z_i'\alpha - \frac{1}{2}\theta_i^2\Psi + \theta_i\Xi_i\right)^2}{2\sigma_v^2}\right] \times f_\theta(\theta_i; \omega)\, d\theta_i, \quad i = 1, \ldots, n, \quad (9)$$

where ϕ denotes the entire parameter vector. The integral above is not available in closed form so Kumbhakar and Tsionas (2006) proposed to approximate it using Monte Carlo methods resulting in the method of MSL.[16] One promising alternative is ML based on the FFT. This is likely to be better than MSL based on what has already been said in connection with Greene's implementation of MSL for the normal-*gamma* stochastic frontier model.

Suppose $\varepsilon = v + g$ is the composed error for the IO model and g denotes the inefficiency term. Standard arguments lead to the fact that $\varphi_\varepsilon(t) = \varphi_v(t)\varphi_g(t)$ due to independence and $\varphi_\varepsilon(t) = \exp(-\frac{1}{2}\sigma_v^2t^2)\int_0^\infty \exp[It(\frac{1}{2}\Psi\theta^2 - \theta\Xi)]f_\theta(\theta)d\theta$ by Theorem 1.5.9 in Lukacs and Laha (1964, p. 22) which states that if the random variable X has distribution function $F(x)$, and $Y = S(X)$ where S is finite, single-valued and Borel-measurable then $\varphi_Y(t) = \mathbb{E}[\exp(ItS(x))] = \int \exp(ItS(x))dF(x)$.

The integral $\varphi_\varepsilon(t)$ can be evaluated analytically for the leading cases of interest when $f_\theta(\theta)$ corresponds to half-normal, truncated normal or exponential distributions. Indeed, for the *half-normal* case, $f_\theta(\theta; \sigma_\theta) = (\frac{\pi}{2}\sigma_\theta^2)^{-1/2}\exp\left(-\frac{\theta^2}{\sigma_\theta^2}\right)$, it is easy to obtain the CF

$$\varphi_\varepsilon(t) = 2\exp\left(-\frac{1}{2}t^2[\sigma_v^2 + \sigma_*^2\Xi^2]\right) \times (1 - It\sigma_\theta^2\Psi)^{-1/2}\Phi(-\sigma_* It\Xi), \quad (10)$$

where $\sigma_*^2(t) = \frac{\sigma_\theta^2}{1 - It\sigma_\theta^2\Psi}$.

The CF of g_i is seen to be $\varphi_g(t) = 2\exp(-\frac{1}{2}t^2\sigma_*^2\Xi^2)(1 - It\sigma_\theta^2\Psi)^{-1/2}\Phi(-\sigma I_* t\Xi)$ so the first moment of g_i can be computed as $E(g_i) = I^{-1}\frac{d\varphi_g(t)}{dt}|_{t=0} = \frac{1}{2}\Psi\sigma_\theta^2 - (2/\pi)^{1/2}\sigma_\theta\Xi_i$ taking account of the fact that σ_*^2 also depends on t. Since $-g_i$ is the IO technical inefficiency measure in *output* terms it follows that this can be computed easily in *closed form* once parameter estimates have been obtained.[17] Since $\Psi < 0$ and $\Xi_i > 0$ by the regularity conditions on technology, it follows that the technical inefficiency measure $-E(g_i)$ is positive.

For the *exponential* case, $f_\theta(\theta; \sigma_\theta) = \sigma_\theta\exp(-\sigma_\theta\theta)$, we have

$$\varphi_\varepsilon(t) = (2\pi)^{1/2}\sigma_\theta\sigma_*\exp\left(-\frac{1}{2}\sigma_v^2t^2 - \frac{1}{2}It\Psi m_*^2\right) \times \Phi(m_*/\sigma_*), \quad (11)$$

where $\sigma_*^2 = -\frac{1}{It\Psi}$, and $m_* = \frac{It\Xi + \sigma_\theta}{It\Psi}$, with the restriction $\Psi < 0$.

For the *truncated normal* distribution, $\theta \sim N(\mu, \sigma_\theta^2)$, $\theta \geq 0$, the corresponding expression is

$$\varphi_\varepsilon(t) = \Phi(\mu/\sigma_\theta)^{-1}(1 - \sigma_\theta^2It\Psi)^{-1/2} \times \exp\left[-\frac{1}{2}\sigma_v^2t^2 + \frac{1}{2}(m_*/\sigma_*)^2 - \frac{1}{2}(\mu/\sigma_\theta)^2\right] \times \Phi(m_*/\sigma_*), \quad (12)$$

where $\sigma_*^2 = \frac{\sigma_\theta^2}{1 - It\sigma_\theta^2\Psi}$, $m_* = \sigma_*^2(\mu/\sigma_\theta^2 - It\Xi)$, subject to the restriction $\Psi < 0$.

For the *gamma* distribution, $\theta \sim Gamma(p, q)$, we have

$$\varphi_\varepsilon(t) = q^{-p}\Gamma(p)^{-1}(2\pi\sigma_*^2)^{1/2} \times \exp\left[-\frac{1}{2}\sigma_v^2t^2 + \frac{1}{2}m_*\right]Q(m_*, \sigma_*), \quad (13)$$

where $\sigma_*^2 = (It\Psi)^{-1}$, $m_* = \sigma_*^2(q^{-1} + It\Xi)$, and $Q(m_*, \sigma_*) = \mathbb{E}[z^{p-1}|z \sim N(m_*, \sigma_*^2), z \geq 0]$.

The expectation involved in the Q function cannot be evaluated in closed form. In fact, estimation of the IO model is no more difficult to perform than estimation of the OO normal-*gamma* or normal-*beta* models. However, *one distinguishing feature of the IO model, unlike the OO case, is that the MGF and CF depend on the particular observation* through the Ξ_is, so in order to obtain the density the FFT must be applied n times where n is the sample size. *This task is hopeless when the sample size is large* and one might as

[15] Although this is a cross-sectional model we include a trend variable because we are using a panel data in which each firm is observed over a number of years. In this paper we are treating the panel as a cross-section.

[16] It is possible to obtain the distribution of g from the distribution of θ in closed form. Specifically, the Jacobian of transformation from θ to g is $1/(\Xi - \Psi\theta(g))$ and $\theta(g)$ can be obtained as the unique positive root of the polynomial $(\Psi/2)\theta^2 - \Xi\theta - g = 0$, provided $g \leq 0$ and $\Psi < 0$. However, the problem is that the convolution of v and g is not available in closed form.

[17] Instead, Kumbhakar and Tsionas (2006) resorted to numerical integration to compute the first moment of g.

well resort to quadrature (for each observation) to approximate the density. Indeed, the FFT technique is useful only when just a few transforms have to be performed for each parameter configuration since then the density can be obtained at the specific residuals using interpolation so that the operations do not depend heavily on the sample size-apart, of course, from the trivial consideration that larger samples imply that more points have to be interpolated. For example in OO models with gamma or beta inefficiency we saw that the FFT has to be performed only once since it depends only on the parameter values and then interpolation can be used to obtain the density of the composed error at a potentially very large number of residuals.

A computationally efficient solution to the problem is to approximate the log-density over a bivariate $N \times G$ grid of e and Ξ points (so the number of FFTs that must be applied for each parameter vector is G) and use bivariate linear interpolation to obtain the log-density[18] evaluated at the particular residual e_i and the associated value of Ξ_i. Since Ξ_i measures returns to scale the Ξ_is should not be expected to be too far apart and to cluster around a number reasonably close to unity. More specifically, when the density is to be evaluated, the residuals and the Ξ_is are computed. Based on their extreme values the bivariate grid is constructed (so that no residual – or Ξ_is – fall outside the grid). Apparently, the grid changes with each new parameter vector. The FFT is applied to obtain the density for each value of Ξ in the grid and over the fixed grid for es. Bivariate linear interpolation is then applied to the logarithms of the density approximations resulting from the FFT.

6. Empirical application of ML-FFT estimation in the IO frontier model

To apply the new technique I have used a data set from Lee (2002) for 56 privately owned, coal-dependent electric utilities over the period 1975–1990. The total number of observations is 896. Output is net steam generation in MWH, and the inputs are capital, labor, and sulfur coal (the aggregate of high- and low-sulfur coal). The reader is referred to Lee (2002) for details concerning the construction of these variables. A trend variable, its square and interactions with the inputs is also included in the translog production function. Inputs can be considered endogenous and output can be considered exogenous in this case, which is why the IO technical efficiency concept is more appropriate than the OO concept. One reason for choosing this data set is that the empirical third moment of LS residuals is quite small (−0.00072) so we do not expect the presence of large OO technical inefficiency in the sector and, therefore, fitting the OO or IO models by ML will be a challenge.[19]

To estimate the IO model with a half-normal distribution for technical inefficiency, 256 points were used to perform the FFT and 20 points were used for intermediate interpolation with respect to the Ξ_is. The results were not sensitive to changing these values and, as a matter of fact, good results have been obtained in this application even by using only 64 points for the FFT and 5 points for interpolation. Since convergence has been found difficult when the number of points for either FFT or interpolation increases, the second-order coefficients have been first fixed to their LS values and I iterated to full convergence of the first-order coefficients. Next, starting from these estimates the ML-FFT procedure is applied again to optimize all parameters. The ML-FFT procedure has not been found significantly more difficult to implement than usual ML estimation of the OO model in terms of timing,[20] difficulty to obtain convergence *etc*.

Table 5
Empirical results for the OO and IO frontier models of US utilities.

Variable	*OO model*	*IO model*
Constant	2.53 (0.022)	2.53 (0.59)
k	0.87 (0.041)	0.87 (0.05)
l	−0.23 (0.044)	−0.23 (0.12)
z	0.42 (0.035)	0.42 (0.14)
$\frac{1}{2}k^2$	−0.34 (0.058)	−0.34 (0.08)
kl	0.08 (0.049)	0.08 (0.07)
kz	0.26 (0.041)	0.26 (0.035)
$\frac{1}{2}l^2$	−0.43 (0.093)	−0.43 (0.085)
lz	0.15 (0.056)	0.15 (0.052)
$\frac{1}{2}z^2$	−0.18 (0.044)	−0.18 (0.035)
t	−0.036 (0.006)	−0.040 (0.023)
$\frac{1}{2}t^2$	0.002 (0.0007)	0.002 (0.0007)
tk	−0.005 (0.004)	−0.005 (0.005)
tl	0.004 (0.004)	0.004 (0.005)
tz	0.00017 (0.003)	0.0002 (0.004)
σ_v	0.176 (0.0037)	0.177 (0.004)
σ_u	0.0036 (0.557)	–
σ_θ	–	0.0006 (0.564)

Notes: Standard errors in parentheses. k, l, and z denote respectively the log of capital, labor and coal input divided by their geometric means, and t is the trend variable. Convergence of ML for the OO model was not achieved. The slope coefficients reported and their standard errors are from LS. Estimates for intercept and standard deviation parameters are from ML estimation of the restricted OO model (restricted so that all slope coefficients are fixed at LS estimates). Execution time for the IO model was 1.38 min. MSL with 5000 replications took 12.12 min.

The results are reported in Table 5 and they are comparable with the OO case using a half-normal distribution for the one-sided disturbance in that very little evidence is found in favor of sizable inefficiencies as expected since the third moment of LS residuals is quite small. According to the results from the OO model, median inefficiency is only 0.013% while in the IO case the median is 0.017%. In the OO case the maximum inefficiency is hardly 0.5% and similar results are obtained in the IO case. Parameter estimates are similar but there are important differences in terms of standard errors, for the intercept and first-order terms for example.

7. Two-tiered stochastic frontier models

In this section it is shown how the CF of the composed error term can be used to facilitate ML estimation in the class of two-tiered stochastic frontier models. These models have been introduced by Polachek and Yoon (1987, 1996). The basic formulation is

$$y_i = x_i'\beta + v_i + u_i + w_i,$$

where $u_i \geq 0$ and $w_i \leq 0$ are independent error terms and $v_i \sim iidN(0, \sigma_v^2)$ independently of v_i and w_i. Only exponential distributions have been considered for the one-sided error terms because the likelihood function is not available in closed form under other stochastic specifications. Kumbhakar and Parmeter (2004) have expressed the likelihood function of the model under exponential one-sided errors in a slightly easier form to handle. In the context of the labor market, if the dependent variable is log wage, $-w_i$ represents employee deficiency in obtaining the maximum possible wage, and u_i represents employer deficiency to offer the minimum possible wage. Since the model is useful in

[18] An alternative is to use bivariate smoothing splines but linear interpolation has been found adequate here.

[19] This is because when the third moment is close to zero, the distribution of the composed error is approximately symmetric and σ_u in the OO model or σ_θ in the IO model will be close to zero. This will be a troublesome case to handle for numerical optimization algorithms.

[20] For this application timing was 0.88 for the OO model and 1.38 min for the IO model. MSL is possible to apply in this instance. Timing was 12.12 min using 1000 replications.

modeling informational asymmetries in other contexts as well, an expansion of the distributional alternatives is necessary. Here, we consider *gamma* distributed error terms,[21] that is

$$f(u_i) = [\theta_u \Gamma(\alpha_u)]^{-1} u_i^{\alpha_u - 1} \exp(-u_i/\theta_u),$$
$$\theta_u, \alpha_u > 0, u_i \geq 0,$$
$$f(w_i) = [\theta_w \Gamma(\alpha_w)]^{-1} w_i^{\alpha_w - 1} \exp(-w_i/\theta_w),$$
$$\theta_w, \alpha_w > 0, w_i \leq 0,$$

where α_u and α_w are shape parameters, and θ_u, θ_w are scale parameters. See also Tsionas (2000). As in the normal-*gamma* stochastic frontier model, when the shape parameters are equal to one we have the exponential two tier stochastic frontier model. If $\varepsilon_i = v_i + u_i + w_i$ is the composed error, the logarithm of the CF is given by

$$\log \phi_\varepsilon(t) = -\frac{1}{2}\sigma_v^2 t^2 - \alpha_u \ln(1 - It\theta_u) - \alpha_w \ln(1 + It\theta_w),$$

which is a complex-valued function but easy enough to compute and subject to the Fourier transform to obtain the corresponding density function.

For empirical application the data set constructed and provided by Koop and Tobias (2004) has been used. This is a data set of 2178 individuals from the National Youth Longitudinal Survey (NYLS) with a total of 17,919 observations. The sample is restricted by Koop and Tobias (2004) to white males who are at least 16 years of age, reported working at least 30 weeks a year and at least 800 h per year. The time varying characteristics include potential labor market experience, a time trend, education and the square of these variables. The time invariant characteristic used here is cognitive ability, measured by a standard test score. The dependent variable is the reported log hourly wage in the most recent employment converted to 1993 dollars.

We provide results for the *gamma* and exponential specifications for the one-sided error terms in Table 6. According to the exponential, employer deficiency averages 32.66% and employee deficiency averages 21.95% and both are highly statistically significant. From the *gamma* results, we see that employer deficiency cannot be described by an exponential distribution as α_u is statistically different from unity (0.655 with standard error 0.1381) although employee deficiency is reasonably close to the exponential model as α_w is not significantly different from unity. Average deficiencies from the *gamma* model are 25.3% in the employer side and 22% in the employee side so it turns out that they are more or less the same despite the prediction of the exponential model. As a result, the wage informational gap reduces from about 10% in the exponential model to just 3.3% in the *gamma* specification.

Table 6
ML results for the two-tiered frontiers, NYLS data.

	Exponential		*Gamma*	
	Estimate	Standard error	Estimate	Standard error
Constant	0.3364	0.1122	0.2690	0.1281
Education	0.1581	0.0166	0.1578	0.0154
Experience	0.1425	0.0039	0.1423	0.0039
Trend	−0.0627	0.0039	−0.0629	0.0041
1/2 (education)^2	−0.0052	0.0012	−0.0052	0.0011
1/2 (experience)^2	−0.0094	0.0004	−0.0094	0.0004
1/2 (trend)^2	0.0053	0.0005	0.0053	0.0005
Ability	0.0910	0.0043	0.0910	0.0042
σ_v	0.2600	0.0071	0.2829	0.0266
θ_u	0.3266	0.0058	0.3865	0.0324
θ_w	0.2195	0.0064	0.2105	0.0391
α_u	–	–	0.6550	0.1381
α_w	–	–	1.0461	0.5921
Average log lik.	−0.640701		−0.640512	

8. Non-normal frontier models with *gamma* one-sided error components

In this section we consider models which remove the normality assumption of the two-sided error term. We will consider latent class models which are useful in modeling heterogeneity in applied production analysis (Greene, 2002; Tsionas, 2001b) and models with a two-sided error distributed according to a stable Paretian distribution, useful when leptokurtosis is present. Both models remove the normality assumption for the two-sided disturbance and thus they are expected to be useful in applied econometric studies of production. Consider first the latent class model

$$y_i = x_i'\beta + v_i + u_i, \quad i = 1, \ldots, n,$$

where g denotes the specific group, $v_i \sim N(0, \sigma_g^2)$, $u_i \sim Gamma(\alpha_g, \theta_g)$, $g = 1, \ldots, G$, group g has probability λ_g, and, as usual, all error components are assumed *IID* and mutually independent. In this model, the technology is assumed common and heterogeneity is placed on the two error components.

Existing latent class models have been proposed and implemented so far for exponential and half-normal one-sided error terms but not for *gamma* distributed error terms due to the computational complexities associated with the fact that the likelihood function is not available in closed form. Clearly, for $\alpha_g = 1$ $(g = 1, \ldots, G)$ we nest the case of latent class models with exponential one-sided errors. It is not difficult to show that the CF of the overall error term $\varepsilon_i = y_i - x_i'\beta$ is given by

$$\phi(t) = \sum_{g=1}^{G} \lambda_g \exp\left(-\frac{1}{2}\sigma_{v,g}^2 t^2\right)(1 - I\theta_g t)^{-\alpha_g}.$$

Therefore, the CF is available in closed form and it is fairly easy to compute which means that ML estimation using the FFT is quite feasible for this model. A more general model results if we remove the assumption that technology is the same, resulting in

$$y_i = x_i'\beta_g + v_{i,g} + u_{i,g}, \quad i = 1, \ldots, n, g = 1, \ldots, G.$$

The composed error term for this model is $\varepsilon_{i,g} = v_{i,g} + u_{i,g}$, and the associated CF is

$$\phi_g(t) = \exp\left(-\frac{1}{2}\sigma_{v,g}^2 t^2\right)(1 - I\theta_g t)^{-\alpha_g}, \quad g = 1, \ldots, G.$$

Using the inverse Fourier transform we can obtain the density of $\varepsilon_{i,g}$, which we denote by $f(\varepsilon_{i,g}; \theta)$ for each group $g = 1, \ldots, G$. The conditional distribution of y_i given the covariates and the parameters can then be obtained simply as

$$f(y_i|x_i, \theta) = \sum_{g=1}^{G} \lambda_g f(y_i - x_i'\beta_g; \theta), \quad i = 1, \ldots, n.$$

The likelihood function is $L(\theta; y, X) = \prod_{i=1}^{n} f(y_i|x_i, \theta)$ and can be maximized with respect to θ to obtain ML parameter estimates. The major computational burden of this approach is that G inverse Fourier transforms have to computed, one for each component of the latent class model. Since the group characteristic functions depend on three parameters ($\sigma_{v,g}^2$, α_g, and θ_g) it seems possible to perform all Fourier inversions in advance, tabulate the results and then use interpolation with respect to these parameters to perform fast ML estimation as in the case of input-oriented stochastic frontier models.

Next, we consider a stable Paretian distribution as a model for the two-sided error term of the standard stochastic frontier model given by $y_i = x_i'\beta + v_i + u_i$. The stable distribution has been proposed by Mandelbrot (1963) and Fama (1965) to model excess

[21] Truncated normal distributions can be considered as well which include the half-normal as a special case. The details will be omitted here.

kurtosis in stock return behavior. It is usually expressed through its log CF, given by

$$\log \phi_v(t) = \begin{cases} -c|t|^{\lambda}\left(1 - Ib\mathrm{sgn}(t)\tan(\pi\lambda/2)\right), & \lambda \neq 1, \\ -c|t|\left(1 + Ib(2/\pi)\mathrm{sgn}(t)\log|t|\right), & \lambda = 1. \end{cases}$$

The parameters c, λ and b are respectively scale, index and skewness parameters. The range of the parameters is $0 < \lambda \leq 2$, $-1 \leq b \leq 1$, and $c > 0$. For $\lambda = 2$ and $b = 0$ it reduces to the normal distribution. For $\lambda = 1$ and $b = 0$ it reduces to the Cauchy distribution (Student-t with one degree of freedom). Models with $\lambda < 2$ have excess kurtosis relative to the normal distribution. There is a vast literature on stable distributions, see for example, Feuerverger and McDunnough (1981a,b), Koutrouvelis (1980), Press (1972), and Tsionas (1999) among many others.

In their application to stochastic frontiers, we need to assume symmetric two-sided errors in which case we set $b = 0$ to obtain the following simple expression for the CF,

$$\phi_v(t) = \exp(-c|t|^{\lambda}).$$

Combined with a *gamma* distribution for the one-sided error, $u_i \sim Gamma(\alpha, \theta)$, the characteristic function of the composed error, $\varepsilon_i = v_i + u_i$ is

$$\phi_\varepsilon(t) = \exp(-c|t|^{\lambda})(1 - It\theta)^{-\alpha}.$$

The CF is easy to compute and application of the inverse Fourier transform to obtain the density function is straightforward. The main advantage of this model is that it allows direct testing of the normality assumption by testing whether $\lambda = 2$.

To illustrate the new techniques we use the US banking data set analyzed in Section 4. Unlike the 500 banks used in Section 4, we now use the entire set available for 2000 consisting of 4985 banks. Since our attention is on the distributional assumption, we limit attention to a model with a homogeneous production technology but heterogeneous two- and one-sided error terms. The distributional possibilities for the one-sided error terms, include exponential and *gamma* distributed. In the first case the likelihood function is available in closed form but not in the second. The empirical results are reported in the first two columns of Table 7. The third column provides empirical results for the stable-*gamma* stochastic frontier.

According to the exponential model, the probability of classification (λ) is 0.6623 and highly statistically significant (standard error 0.0757), and the two groups are 8.65% and 24.25% inefficient on the average. This suggests the existence of two groups of banks which behave quite differently in terms of technical inefficiency. Removing the assumption of exponential inefficiencies, and introducing *gamma* one-sided errors results in a dramatic change of the results. The probability of classification (λ) is now 0.2221 and is not statistically significant (standard error 0.4397). Standard errors for the shape parameters (α_g) are also large not providing a clear rejection of the exponential model. Standard errors for the scale parameters (θ_g) are also larger in the *gamma* model reflecting the greater uncertainty introduced by assuming a model of inefficiencies which is hard to identify in practice. However, allowing for the possibility of *gamma* inefficiencies, lowers dramatically the probability of classification (λ) and casts some doubt on the existence of heterogeneity in the banking sector. It is not clear whether this is due to similar technological parameters in the banking practice or due solely to the fact that *gamma* distributions for technical inefficiency are notoriously difficult to identify in practice. Further work in this area would be an interesting extension for future research.

Results for the stable-*gamma* model are reported in the third column of Table 7. The characteristic exponent, λ, is quite close to 2 (the ML estimate is 1.988 with standard error 0.0174) suggesting that a normal-gamma model is reasonable and the two-sided error term does not have fat tails. Although the model has been estimated using all 4985 banks available for the year 2000, the results for the normal-*gamma* model with the 500 banks reported in Table 2 are more or less the same. Based on these results, the stable-*gamma* and normal-*gamma* do not provide a fundamentally different view of inefficiency in the banking sector, which is expected given the fact that the characteristic exponent is not statistically different from 2. Given the flexibility and ease of implementation, more applications of the stable model would seem interesting in applied econometric analyses of production. Routine direct testing of the normality assumption especially in large data sets would provide more needed evidence on the validity of normality, a fundamental and common assumption in stochastic frontier analyses. Another important avenue for future research would be the introduction of a stable distribution for the one-sided error term. It can be shown that random variables having stable distributions with $\lambda < 1$ and $b = 1$ are always non-negative, lending themselves naturally as candidate models for inefficiency in stochastic frontiers with possible fat tails in the inefficiency distribution. The combination of stable distributions for both error components of the model is also quite flexible and easy to estimate given numerical inverse Fourier transforms of the characteristic function of the composed error.

Table 7
Maximum likelihood FFT parameter estimates for cost function of US commercial banks using the latent class *gamma* and stable-*gamma* models.

	Exponential	*Gamma*	Stable-*gamma*
Constant	0.8111 (0.0813)	0.8294 (0.1703)	0.8359 (0.0789)
$\log w_1$	0.2531 (0.0118)	0.2550 (0.0119)	0.2534 (0.0118)
$\log w_2$	0.0134 (0.0042)	0.0133 (0.0042)	0.0125 (0.0042)
$\log w_3$	0.0686 (0.0077)	0.0691 (0.0077)	0.0688 (0.0077)
$\log w_4$	0.0347 (0.0070)	0.0348 (0.0070)	0.0348 (0.0070)
$\log y_1$	0.0680 (0.0041)	0.0684 (0.0041)	0.0678 (0.0041)
$\log y_2$	0.4630 (0.0043)	0.4624 (0.0043)	0.4639 (0.0043)
$\log y_3$	0.1847 (0.0055)	0.1838 (0.0055)	0.1850 (0.0055)
$\log y_4$	0.0413 (0.0022)	0.0413 (0.0022)	0.0413 (0.0022)
$\log y_5$	0.1901 (0.0057)	0.1911 (0.0057)	0.1888 (0.0056)
$\sigma_{v,(1)}$	0.1442 (0.0092)	0.1155 (0.0517)	0.1618 (0.0045)
$\sigma_{v,(2)}$	0.2090 (0.0168)	0.1780 (0.0305)	–
$\alpha_{(1)}$	1.00	0.6829 (1.1179)	0.3839 (0.0704)
$\alpha_{(1)}$	1.00	0.4527 (0.5376)	–
$\theta_{(1)}$	0.0865 (0.0354)	0.0756 (0.5129)	0.2935 (0.0277)
$\theta_{(2)}$	0.2424 (0.0216)	0.2891 (0.0659)	–
λ	0.6623 (0.0757)	0.2221 (0.4397)	1.988 (0.0174)
Average logL	0.0879	0.0882	0.0875

Notes: Standard errors appear in parentheses. Estimations used the entire set of 4985 banks available for 2000. The model estimated has common technological parameters, heterogeneous parameters of the two- and one-sided error components, and two latent classes are assumed. The two-sided error term is normal and the two one-sided components are *gamma* distributed. For the stable-gamma model, $\sigma_{v,(1)}$ is the scale parameter of the two-sided error, $\theta_{(1)}$ is the scale parameter of the *gamma* distribution, $\alpha_{(1)}$ is the shape parameter of the gamma distribution, and λ is the characteristic exponent. For the latent class models, λ is the probability of classification in the first group.

9. Conclusions

The purpose of the paper was to introduce the FFT as a way of obtaining ML estimates of the parameters of complicated output- or input-oriented stochastic frontier models and investigate the performance of certain GMM and other procedures that are based on the asymptotic normal form (ANF) of the likelihood function derived from the Fourier representation. It has been found that the FFT is a robust and computationally efficient way to implement ML estimation. ANF estimators are found to be critically dependent on certain parameters of the computational environment (like number and placement of grid points) and this dependence becomes rather restrictive when practical identification problems arise as for example in the case on truncated normal stochastic frontier models. Moreover, it is not entirely clear what configuration of grid

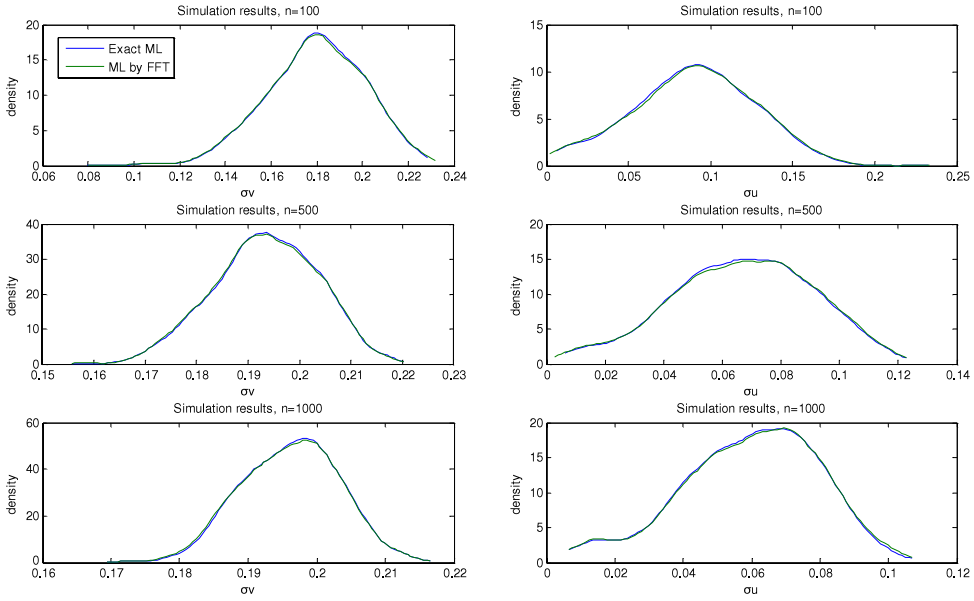

Fig. A.1. Sampling distributions of parameters for exponential specification ($\sigma_u = 0.05$).

points is optimal because certain matrices needed in implementing an investigation of the optimal scheme are very badly conditioned as has been found in computational experiments. Overall, it appears that in connection with ML estimation of either output- or input-oriented stochastic frontier models, the FFT is a viable alternative to currently used simulated likelihood methods. Moreover, the use of FFT enriches significantly the class of models that can be estimated easily in stochastic frontier analysis and thus, offers significantly more alternatives to researchers engaged in applied production analysis and efficiency/productivity analysis.

Acknowledgments

The author wishes to thank the Editor A. Ronald Gallant, an Associate Editor, and three anonymous referees for many useful comments on an earlier version.

Appendix A. Monte Carlo experiments for the ML-FFT procedure

To address the question of whether or not Fourier methods of inference are reasonable in stochastic frontier models, we consider the following Monte Carlo experiment. The true model is $y_i = -1 + x_i + v_i + u_i, i = 1, \ldots, n$, where the x_is are fixed in repeated samples and constructed from a standard normal distribution, $v_i \overset{IID}{\sim} N\left(0, \sigma_v^2\right)$ with $\sigma_v = 0.2$ and u_i follows an exponential distribution with density $f_u(u) = \sigma_u^{-1} \exp\left(-\sigma_u^{-1} u\right), u \geq 0$.

The characteristic function of the exponential is $\ln \varphi_u(t) = -\ln(1 - It\theta)$. In the Monte Carlo experiments we consider several configurations of σ_u and sample size, n. For each experiment 10,000 replications were used and the standard BFGS algorithm has been used to maximize the likelihood functions (either exact or approximated by the FFT) (see Table A.1).

The sampling distributions of ML estimators of σ_v and σ_u when $\sigma_u = 0.05$ are shown in Fig. A.1. The results are reasonable in the sense that the FFT approximation to the likelihood function is extremely accurate.

One important question is what happens with specifications that do not yield closed form expressions for the likelihood function, like the normal-gamma. To answer this question we use the FFT and Gaussian quadrature to approximate the density of a normal-gamma density with $\sigma_v = 0.2, \alpha = 0.35$ and $\sigma_u = 0.2/\alpha = 0.571$. It has been found (but the details are omitted) that for α greater than about 0.7, the FFT and quadrature results (with as few as 10 or 20 base points) are very similar, so the real question is how the FFT behaves for lower values of α, like for example 0.35. In Fig. A.2 we report the FFT approximations and "exact" results based on quadrature with 10, 100 and 500 base points.

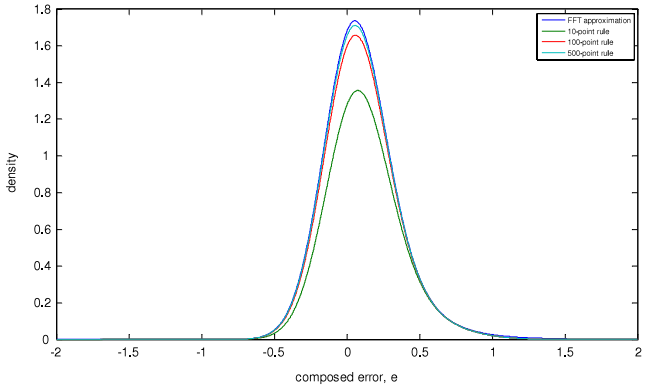

Fig. A.2. FFT and quadrature approximations to normal-*gamma* density ($\sigma_v = 0.2, \alpha = 0.35, \sigma_u = 0.571$).

It turns out that the FFT is far more accurate than quadrature and, in fact, a large number of base points is needed to yield an accurate approximation to the FFT—to which quadrature apparently converges! Of course, the FFT approximation is computed at a fraction of time compared to quadrature or, naturally, simulation estimation of the density. Therefore, in practice, it seems that the FFT is the only available fast and accurate way to approximate the density of a normal-gamma stochastic frontier model. Equally accurate results need a large number of Gaussian quadrature base points and an even greater number of replications for simulation-based estimation. The same is the case with the normal-beta model as well; see Fig. A.3.

Appendix B. Performance of procedures for the normal/truncated-normal frontier model

A potentially important distribution in applications is the *truncated normal* for technical inefficiency, $u \sim N(\mu, \sigma_u^2)$

Table A.1
Monte Carlo results for the exponential specification.

σ_u	$n = 100$			$n = 500$			$n = 1000$		
0.05	Exact	ML	Means	Exact	ML	Means	Exact	ML	Means
	0.1796	0.08971		0.1934	0.06661		0.1962	0.05947	
	FFT	ML	Means	FFT	ML	Means	FFT	ML	Means
	0.1795	0.08984		0.1933	0.06684		0.1961	0.05981	
	Exact	ML	S.d.	Exact	ML	S.d.	Exact	ML	S.d.
	0.02171	0.03706		0.01021	0.02435		0.007133	0.02006	
	FFT	ML	S.d.	FFT	ML	S.d.	FFT	ML	S.d.
	0.02200	0.03769		0.01028	0.02471		0.007229	0.02027	
	Exact	ML	MSE	Exact	ML	MSE	Exact	ML	MSE
	0.0008883	0.002949		0.0001472	0.0008683		6.515E−5	0.0004916	
	FFT	ML	MSE	FFT	ML	MSE	FFT	ML	MSE
	0.0009042	0.003007		0.0001501	0.0008935		6.746E−5	0.0005069	
0.10	Exact	ML	Means	Exact	ML	Means	Exact	ML	Means
	0.1878	0.1124		0.1987	0.09992		0.1998	0.09790	
	FFT	ML	Means	FFT	ML	Means	FFT	ML	Means
	0.1876	0.1127		0.1985	0.1002		0.1997	0.09820	
	Exact	ML	S.d.	Exact	ML	S.d.	Exact	ML	S.d.
	0.02475	0.04214		0.01183	0.02487		0.008422	0.01823	
	FFT	ML	S.d.	FFT	ML	S.d.	FFT	ML	S.d.
	0.02488	0.04250		0.01191	0.02508		0.008457	0.01825	
	Exact	ML	MSE	Exact	ML	MSE	Exact	ML	MSE
	0.0007613	0.001928		0.0001414	0.0006181		7.090E−5	0.0003365	
	FFT	ML	MSE	FFT	ML	MSE	FFT	ML	MSE
	0.0007713	0.001966		0.0001438	0.0006285		7.156E−5	0.0003360	
0.20	Exact	ML	Means	Exact	ML	Means	Exact	ML	Means
	0.1938	0.1968		0.1994	0.1995		0.1996	0.1994	
	FFT	ML	Means	FFT	ML	Means	FFT	ML	Means
	0.1936	0.1971		0.1993	0.1996		0.1995	0.1995	
	Exact	ML	S.d.	Exact	ML	S.d.	Exact	ML	S.d.
	0.03154	0.05218		0.01376	0.02175		0.009476	0.01572	
	FFT	ML	S.d.	FFT	ML	S.d.	FFT	ML	S.d.
	0.03156	0.05225		0.01376	0.02167		0.009475	0.01572	
	Exact	ML	MSE	Exact	ML	MSE	Exact	ML	MSE
	0.001032	0.002731		0.0001894	0.0004726		8.989E−5	0.0002472	
	FFT	ML	MSE	FFT	ML	MSE	FFT	ML	MSE
	0.001036	0.002736		0.0001896	0.0004691		8.993E−5	0.0002471	
0.30	Exact	ML	Means	Exact	ML	Means	Exact	ML	Means
	0.1938	0.2962		0.1991	0.2999		0.1992	0.3000	
	FFT	ML	Means	FFT	ML	Means	FFT	ML	Means
	0.1937	0.2964		0.1991	0.2999		0.1992	0.3000	
	Exact	ML	S.d.	Exact	ML	S.d.	Exact	ML	S.d.
	0.03694	0.05719		0.01527	0.02369		0.01076	0.01744	
	FFT	ML	S.d.	FFT	ML	S.d.	FFT	ML	S.d.
	0.03688	0.05704		0.01528	0.02369		0.01077	0.01746	
	Exact	ML	MSE	Exact	ML	MSE	Exact	ML	MSE
	0.001402	0.003281		0.0002338	0.0005606		0.0001162	0.0003040	
	FFT	ML	MSE	FFT	ML	MSE	FFT	ML	MSE
	0.001398	0.003264		0.0002340	0.0005606		0.0001166	0.0003047	
0.40	Exact	ML	Means	Exact	ML	Means	Exact	ML	Means
	0.1922	0.3971		0.1989	0.4001		0.1989	0.4002	
	FFT	ML	Means	FFT	ML	Means	FFT	ML	Means
	0.1922	0.3972		0.1989	0.4001		0.1989	0.4002	
	Exact	ML	S.d.	Exact	ML	S.d.	Exact	ML	S.d.
	0.03964	0.06279		0.01682	0.02719		0.01193	0.02003	
	FFT	ML	S.d.	FFT	ML	S.d.	FFT	ML	S.d.
	0.03965	0.06281		0.01682	0.02717		0.01193	0.02003	
	Exact	ML	MSE	Exact	ML	MSE	Exact	ML	MSE
	0.001631	0.003947		0.0002837	0.0007384		0.0001433	0.0004008	
	FFT	ML	MSE	FFT	ML	MSE	FFT	ML	MSE
	0.001632	0.003949		0.0002837	0.0007377		0.0001433	0.0004008	

Notes: (1) represents Monte Carlo means of σ_v and σ_u using exact ML, (2) is the corresponding means using the FFT, (3) and (4) are Monte Carlo standard deviations from exact ML and FFT, (5) and (6) are RMSEs for the parameters using exact ML and FFT respectively.

truncated below at zero ($u \geq 0$) whose density is

$$f(u \mid \mu, \sigma_u) = (2\pi\sigma_u^2)^{-1/2} \exp\left[-\frac{(u-\mu)^2}{2\sigma_u^2}\right] \Phi(-\mu/\sigma_u)^{-1},$$

$u \geq 0$,

where Φ denotes the standard normal distribution function. See Stevenson (1980). For $\mu = 0$ it reduces to the half-normal distribution. The composition of v and u yields a closed form expression for the likelihood function but numerical experience with ML has generally been disappointing Greene (1999, p. 103). Tsionas (2001a,b) documents good performance of a Bayesian estimator using Markov Chain Monte Carlo methods but the presence of priors – especially a prior on μ or μ/σ_u – is likely to bother non-Bayesian researchers because these priors are

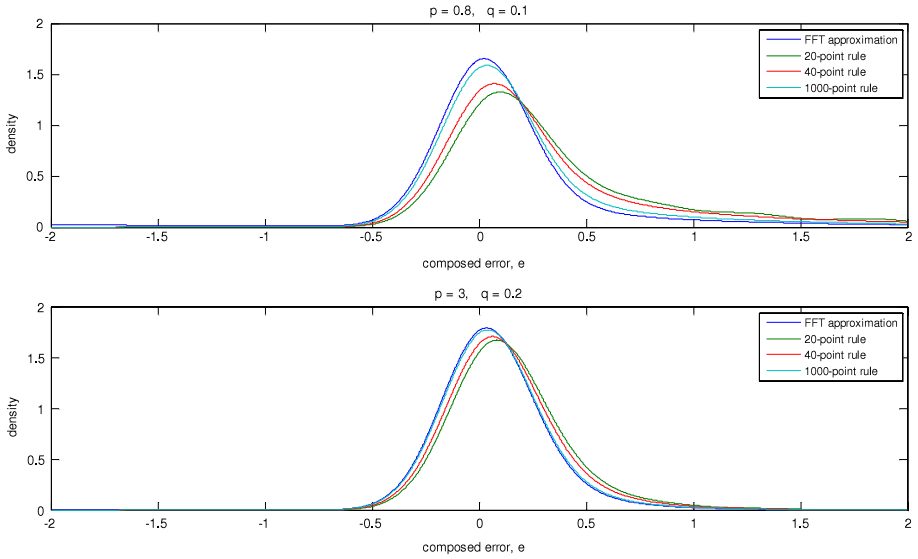

Fig. A.3. FFT and quadrature approximations to normal-*beta* density ($\sigma_v = 0.2$).

Table B.1
Parameter estimates for the normal-truncated normal model for US commercial banks.

	ML	GMM
Constant	0.957 (0.248)	0.869
σ_v	0.147 (0.014)	0.0957
σ_u	5.994 (289.3)	0.368
ω	−33.84 (1643.0)	−0.232

Notes: ML is implemented using a conjugate gradient technique, the starting parameter values are obtained from OLS for the regression coefficients, starting values of σ_v and σ_u are set to $s/2$ where s is the standard deviation of OLS residuals and the starting value for ω is set to 0.01. GMM is implemented using fitting of the empirical to the theoretical MGF using nonlinear least squares with 20 equidistant points from −2 to 2. The Hansen J-statistic (N times the objective function) was 0.00041.

informative to aid practical identification of the model.[22] It is natural to inquire whether simpler alternatives like GMM are available and whether they perform well in finite samples under reasonable parameterizations.

The log-CF of the truncated normal distribution is given by

$$\ln \varphi_u(t) = \omega\sigma_u It - \frac{1}{2}\sigma_u^2 t^2 + \ln \Phi(\omega - \sigma_u It) - \ln \Phi(\omega),$$

where $\omega = \mu/\sigma_u$.

The ML estimates of a truncated normal frontier are reported in Table B.1. Also reported are results from GMM whose implementation is described below. ML is implemented using a conjugate gradient techniques, the starting parameter values are obtained from OLS for the regression coefficients, starting values of σ_v and σ_u are set to $s/2$ where s is the standard deviation of OLS residuals and the starting value for ω is set to 0.01. The estimates are useless as quite often happens with ML estimation of this model. The estimate of ω is implausibly low and standard errors for several parameters are quite large suggesting practical non-identification problems with this model.

Next we turn attention to a GMM technique that involves fitting the empirical to the theoretical moment generating function (MGF) which can be obtained easily just like the CF. This is the k–L procedure proposed by Feuerverger and Mureika (1977). To resolve the choice of the grid points I have implemented fitting the empirical to the theoretical MGF (in levels, not in logs) using a weighting function of the form $\exp(-t^2)$, thus the criterion

Table B.2
Sensitivity of GMM for US commercial banks.

	$G = 4$	$G = 10$	$G = 20$	$G = 50$
Constant	0.880	0.869	0.872	0.870
σ_v	0.117	0.0957	0.097	0.097
σ_u	0.350	0.368	0.372	0.369
ω	−0.217	−0.232	−0.283	−0.247

Notes: The upper endpoint is $T = 2$ and G denotes the number of grid points. The grid is from $-T$ to T and a conjugate gradient algorithm has been used to implement nonlinear least squares fitting of the empirical to the theoretical log MGF. Initial values for the parameters were taken as in Table 5.

is an integral over the real line of the objective function times the weight. This integral can be computed easily using Gaussian quadrature. The resulting parameter estimates were invariant to the order of integration (I have tried 5, 20 and 100) and the estimate of ω was close to zero (0.0085) suggesting that a half-normal model might be appropriate. This, however, is unfortunately[23] an artifact. I have conducted a Monte Carlo experiment (details are omitted here) using reasonable parameterizations of the model and several different values of ω ranging from −3 to 3 as in Tsionas (2001a,b). This fitting function always produced estimates of ω quite close to zero so the placement of grid points must be responsible for this and Gaussian quadrature is not the right way to estimate the model. After other Monte Carlo experiments that I describe later I have found that placing the points in the interval from −2 to 2 using 10–20 equidistant points is a reasonable estimator. Application of this estimator to the US commercial banking data (with 10 grid points) produced the estimates that are reported in Table 6 under the heading "GMM". For ω the estimate appears more plausible relative to ML.

Before describing the Monte Carlo experiment with the GMM estimator, reported in Table B.2 are results from different configurations of the grid points with the nonlinear least squares fitting criterion that minimizes the sum of squared deviations between the theoretical and empirical log MGF using the US commercial banking data. The estimator is not very sensitive selection of the grid points when $T = 2$. For higher values of T I have had trouble obtaining convergence so it seems that the grid cannot be extended more. For $T = 1$ the results are different and

[22] It should be noted that in large samples the impact of priors will be small.

[23] The unfortunate fact here is that Gaussian weights provide an automatic way to implement GMM without having to configure the grid points.

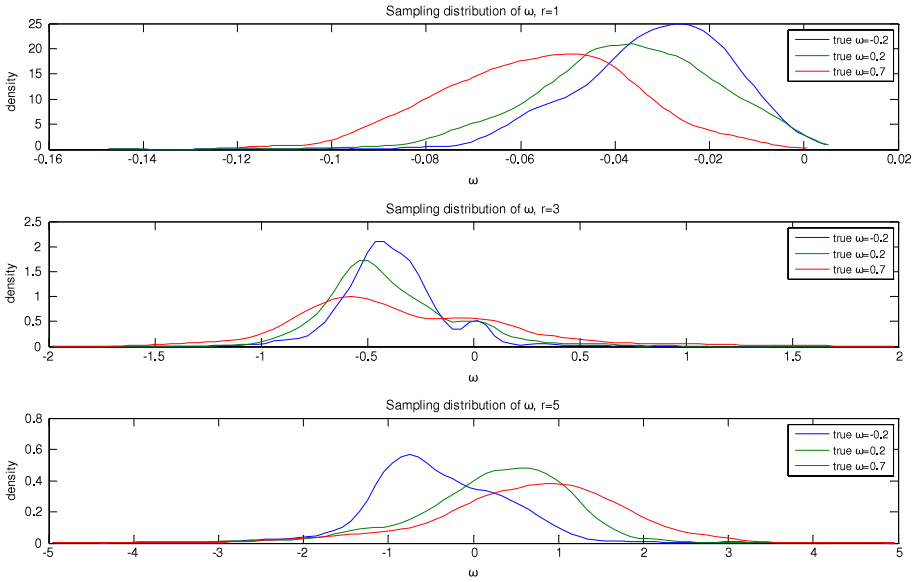

Fig. B.1. Sampling distributions of ω.

estimates of ω range from -0.0062 when $G = 4$ to 0.0165 when $G = 50$. I have also experimented with GMM procedures described in Carrasco and Florens (2002), Knight and Yu (2002), Yu (2004) and Paulson et al. (1975). The results are not reported to save space.

Next, I describe a Monte Carlo experiment to investigate the finite-sample performance of the GMM estimator. The focus here is not to present an exhaustive and detailed analysis but rather focus on a case of empirical relevance. Here we fit the theoretical to the empirical MGF using nonlinear LS. There are three regressors, the intercept (equal to 1) and two other variables whose coefficients are 0.5. The error variance is $\sigma_v^2 = 0.03$, ω can be -0.2, 0.2 or 0.7 and the sample size is $n = 500$. The ratio $r = \sigma_u/\sigma_v$ is 1, 3 or 5. The values of $\sqrt{\mathrm{Var}(u)/\mathrm{Var}(v)}$ are approximately 0.8, 2.0 and 3.0 respectively. Median technical inefficiency is from 10% to 17% in the first case, 30% to 52% in the second, and 50%–85% in the third case. I have used 1000 Monte Carlo replications, with the endpoint being $T = 2$ using 10 points from $-T$ to T. The sampling distributions are reported in Fig. B.1. For $r = 1$ the GMM estimator cannot distinguish effectively between the various values of ω. The situation improves as r increases but only at $r = 5$ the improvement becomes noticeable. *The sampling distributions are clearly far from normality so either huge samples or high values of r are needed to provide useful inferences in the truncated normal case*. Of course this is not a fault of GMM but it is due to the heavy requirements of the truncated normal model in terms of practical identification.

Appendix C. Monte Carlo simulation of the NG/ANF

Here, we describe Monte Carlo results for selection of grid points when implementing ANF procedures. Feuerverger and McDunnough (1981a,b,c) determine the number of grid points and the extent of the grid by minimizing the asymptotic variance of the estimators. Here, we follow a more direct approach which is suitable for finite samples. First we have to choose parameters σ_v, α, λ, and θ. Since average technical inefficiency is $\alpha\theta$ we can fix that (to 0.10 or 0.30) and examine a range of values for α, from 0.25 to 3.0, which contains the range of plausible values for this parameter. Given α and θ, we can determine σ_v by looking at the ratio of standard deviations for the two error components, that is $\lambda = \frac{\sigma_v}{\sqrt{\alpha\theta^2}}$. Given a sample size N, artificial data from a NG stochastic frontier model are simulated, the ANF procedure is applied using various grid configurations and the information matrix of parameters is computed. Finally, the grid is selected so that the determinant of the information matrix is maximal (see Table C.1).

Table C.1
Optimal number of grid points and upper endpoint for the NG/ANF procedure.

$\alpha\downarrow$	Ineff	N	$\lambda\to 0.25$	0.50	1.00	1.50	2.00
0.25	0.1	100	35, 5	10, 2.7	7, 1.9	15, 2.4	30, 3.7
		500	20, 2.1	50, 3.9	10, 5	15, 1.9	30, 1.1
		1000	35, 1.9	40, 3.2	45, 2.1	20, 2.1	35, 2.4
	0.2	100	35, 5	5, 2.9	20, 2.7	40, 1.3	30, 0.6
		500	5, 4.7	25, 2.9	20, 2.1	10, 1.6	20, 0.9
		1000	30, 4.7	5, 4.7	5, 1.7	40, 0.6	25, 1.4
	0.3	100	30, 1.7	20, 0.61	20, 0.61	40, 1.4	45, 0.9
		500	5, 4.7	35, 1.6	7, 0.9	10, 1.6	20, 0.9
		1000	5, 4.2	35, 3.2	30, 1.1	10, 0.6	35, 1.1
0.50	0.1	100	7, 2.42	15, 3.9	30, 4.2	7, 2.4	15, 1.9
		500	30, 0.36	10, 2.9	40, 4.5	45, 2.9	10, 1.4
		1000	7, 4.5	30, 4.7	10, 2.4	7, 4	30, 3.2
	0.2	100	45, 4.2	10, 0.35	7, 1.4	15, 1.6	50, 2.4
		500	50, 0.6	5, 2.4	45, 3.9	25, 1.6	45, 1.9
		1000	40, 0.4	7, 3.2	7, 2.4	25, 1.1	30, 3.1
	0.3	100	35, 1.4	50, 0.62	7, 1.4	50, 2.7	10, 0.62
		500	50, 1.9	5, 2.4	30, 1.6	15, 1.6	10, 1.4
		1000	50, 2.7	10, 0.9	15, 0.6	45, 0.4	25, 1.4
0.75	0.1	100	30, 3.7	45, 1.64	15, 4.5	30, 3.7	10, 5
		500	30, 1.4	40, 3.9	10, 5	50, 5	20, 4
		1000	5, 1.4	25, 4.7	45, 4.2	15, 4	15, 5
	0.2	100	10, 1.4	40, 2.2	5, 2.9	20, 2.4	30, 2.9
		500	50, 0.4	20, 0.9	35, 1.9	50, 2.4	30, 1.4
		1000	10, 4.5	45, 1.4	15, 1.9	10, 2.4	7, 2.4
	0.3	100	40, 3.5	20, 1.13	5, 1.39	25, 0.9	7, 1.13
		500	25, 2.9	25, 1.4	30, 1.1	5, 1.7	50, 2.4
		1000	35, 4	20, 1.1	15, 2.4	10, 2.4	35, 1.7
1.00	0.1	100	10, 1.13	50, 3.7	30, 5	10, 4.2	5, 4.7
		500	5, 2.7	30, 5	20, 4	35, 4.5	50, 4.7
		1000	20, 4.5	45, 5	5, 4.7	30, 3	40, 5
	0.2	100	45, 0.6	25, 2.9	50, 3.5	40, 4.2	10, 2.9
		500	20, 1.6	7, 2.9	45, 3.2	10, 3.7	30, 0.4
		1000	15, 1.9	7, 4.2	7, 2.4	35, 3.2	35, 2.2
	0.3	100	35, 0.61	10, 3.2	15, 3.7	25, 2.4	45, 1.13
		500	30, 0.9	45, 0.9	45, 2.4	45, 1.4	10, 1.7
		1000	45, 0.9	10, 0.9	15, 1.1	15, 1.9	40, 0.4
1.50	0.1	100	7, 3.5	20, 3.2	45, 2.4	45, 3.9	40, 3.7
		500	30, 4.7	50, 3.2	10, 1.6	50, 5	5, 4.2
		1000	10, 4.2	10, 2.7	15, 5	50, 4.7	40, 5

(*continued on next page*)

Table C.1 (*continued*)

$\alpha \downarrow$	Ineff	N	$\lambda \to 0.25$	0.50	1.00	1.50	2.00
	0.2	100	35, 1.1	45, 1.4	15, 5	40, 2.9	7, 3.2
		500	5, 1.65	35, 3.7	15, 4.2	20, 4.2	50, 4
		1000	25, 2.7	35, 1.9	50, 2.7	5, 4.2	30, 2.2
	0.3	100	35, 1.1	20, 1.1	7, 2.2	5, 2.2	7, 3.2
		500	50, 0.36	35, 3.7	10, 2.7	7, 2.4	20, 2.4
		1000	20, 2.2	7, 5	15, 3.7	35, 3.2	10, 2.7
2.00	0.1	100	50, 5	40, 3.5	40, 2.9	10, 5	15, 4.4
		500	10, 4.5	20, 5	25, 4.7	30, 4.7	45, 0.9
		1000	20, 2.2	35, 0.6	20, 2.4	20, 2.7	40, 2.2
	0.2	100	5, 2.7	15, 4.5	40, 3.5	7, 2.4	10, 4.5
		500	5, 1.1	35, 2.2	7, 2.7	5, 4.7	45, 5
		1000	45, 1.7	45, 5	5, 3.5	15, 4.7	10, 4.7
	0.3	100	40, 2.4	30, 4.7	45, 2.7	10, 3.5	15, 2.7
		500	45, 1.1	5, 1.4	45, 3.2	25, 2.7	7, 2.9
		1000	45, 1.7	30, 3.5	5, 3.2	7, 3	7, 2.2
2.50	0.1	100	5, 0.9	7, 2.9	5, 3.7	15, 2.2	10, 4.5
		500	15, 4.2	40, 3.9	40, 3.9	30, 4.2	5, 4.7
		1000	10, 2.2	40, 1.9	35, 4.5	7, 4.5	40, 5
	0.2	100	7, 1.6	15, 1.9	20, 3.2	25, 4.2	45, 1.9
		500	30, 2.4	35, 0.6	35, 5	30, 4.7	50, 1.9
		1000	30, 2.7	20, 4.7	30, 4.2	40, 4.2	30, 2.9
	0.3	100	40, 1.9	45, 3.7	20, 3.2	45, 3.5	25, 2.7
		500	20, 1.13	25, 4.7	35, 3.5	7, 4.7	40, 3.7
		1000	40, 4.5	25, 2.7	7, 2.7	20, 2.7	45, 3.2
3.00	0.1	100	10, 0.1	50, 2.9	7, 1.9	15, 4.7	45, 4.4
		500	45, 4.5	35, 3.2	5, 5	30, 2.1	50, 4.7
		1000	40, 2.9	10, 5	35, 3.5	10, 1.6	45, 5
	0.2	100	45, 1.4	10, 2.4	15, 2.9	35, 4.2	7, 1.9
		500	50, 2.1	25, 4	15, 5	30, 2.4	25, 4.5
		1000	30, 2.9	40, 4.7	20, 1.4	35, 3.5	30, 5
	0.3	100	45, 1.4	7, 3.5	30, 5	5, 4.7	30, 2.1
		500	50, 0.6	30, 4	50, 3	25, 4.7	40, 4.2
		1000	30, 2.9	20, 1.6	50, 1.6	50, 3.7	30, 1.1

Notes: "Ineff" denotes average technical inefficiency (which is equal to the product $\alpha\theta$), and N is the sample size. In each box we report pairs of the form (G, U) where G is the optimal number of grid points and U is the optimal upper endpoint. The lower endpoint is always set to 0.01.

Since the information matrix is not analytically available (due to the sampling expectation involved) it is computed using 100 alternative data sets and the expectation is approximated using the sample average across these artificial data sets.

References

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

Bauer, P.W., 1990. Recent developments in the econometric estimation of frontiers. Journal of Econometrics 46, 39–56.

Beckers, D., Hammond, C., 1987. A tractable likelihood function for the normal-gamma stochastic frontier model. Economics Letters 24, 33–38.

Carrasco, M., Florens, J., 2002. Efficient GMM estimation using the empirical characteristic function. Working Paper, Department of Economics. University of Rochester.

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

Fama, E., 1965. The behavior of stock prices. Journal of Business 47, 244–280.

Feuerverger, A., McDunnough, P., 1981a. On some Fourier methods for inference. Journal of the American Statistical Association 76 (374), 379–387.

Feuerverger, A., McDunnough, P., 1981b. On the efficiency of empirical characteristic function procedures. Journal of the Royal Statistical Society B43, 20–27.

Feuerverger, A., McDunnough, P., 1981c. On efficient inference in symmetric stable laws and processes. In: Csörgo, M., Dawson, D.A., Rao, J.N.K., Saleh, A.K.Md. (Eds.), Statistics and Related Topics. North-Holland, Amsterdam, pp. 109–122.

Feuerverger, A., Mureika, R.A., 1977. The empirical characteristic function and its applications. Annals of Statistics 5, 88–97.

Greene, W.H., 1993. The econometric approach to efficiency analysis. In: Fried, H.O., Lovell, C.A.K., Schmidt, S.S. (Eds.), The Measurement of Productive Efficiency: Techniques and Applications. Oxford University Press, Oxford, pp. 68–119.

Greene, W.H., 1999. Frontier production functions. In: Hashem Pesaran, M., Schmidt, Peter (Eds.), Handbook of Applied Econometrics, Volume II: Microeconomics. Blackwell, Malden, Massachusetts, pp. 81–166.

Greene, W.H., 1990. A gamma-distributed stochastic frontier model. Journal of Econometrics 46, 141–164.

Greene, W.H., 2003. Simulated likelihood estimation of the normal-gamma stochastic frontier function. Journal of Productivity Analysis 19, 179–190.

Greene, W.H., 2002. Alternative panel data estimators for stochastic frontier models. Working Paper. Department of Economics, Stern School of Business, New York University.

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, 233–238.

Knight, J.L., Yu, J., 2002. Empirical characteristic function in time series estimation. Econometric Theory 18, 691–721.

Koop, G., Tobias, J.L., 2004. Learning about heterogeneity in returns to schooling. Journal of Applied Econometrics 19, 827–849.

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

Kumbhakar, S.C., Lovell, C.A.K, 2000. Stochastic Frontier Analysis. Cambridge University Press, New York.

Kumbhakar, S.C, Parmeter, C.F., 2004. Estimation of employer and employee information asymmetries, manuscript.

Kumbhakar, S.C., Tsionas, E.G., 2006. Estimation of stochastic frontier production functions with input-oriented technical inefficiency. Journal of Econometrics 133, 71–96.

Kumbhakar, S.C., Tsionas, E.G., 2005a. Measuring technical and allocative inefficiency in the translog cost system: a Bayesian approach. Journal of Econometrics 126, 355–384.

Kumbhakar, S.C., Tsionas, E.G., 2005b. The joint measurement of technical and allocative inefficiency: an application of Bayesian inference in nonlinear random effects models. Journal of the American Statistical Association 100 (471), 736–747.

Lee, M., 2002. The effect of sulfur regulations on the US electric power industry: a generalized cost approach. Energy Economics 24, 491–508.

Lukacs, E., Laha, R.G., 1964. Applications of Characteristic Functions. Charles Griffin and Company Limited, London.

Madan, D.B., Seneta, E., 1987. Simulation of estimates using the empirical characteristic function. International Statistical Review 55, 153–161.

Mandelbrot, B., 1963. The variation of certain speculative prices. Journal of Business 36, 394–419.

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from Cobb–Douglas production functions with composed error. International Economic Review 8, 435–444.

Mittnik, S., Doganoglu, T., Chenyao, D., 1999. Computing the probability density function of the stable Paretian density. Mathematical and Computer Modelling 29, 235–240.

Paulson, A.S., Holcomb, E.W., Leitch, R.A., 1975. The estimation of the parameters of the stable laws. Biometrika 62, 163–170.

Polachek, S.W., Yoon, B.J., 1987. A two-tiered earnings frontier estimation of employer and employee information in the labor market. Review of Economics and Statistics 69, 296–302.

Polachek, S.W., Yoon, B.J., 1996. Panel estimates of a two-tiered earnings frontier. Journal of Applied Econometrics 11, 169–178.

Press, S.J., 1972. Estimation in univariate and multivariate stable distributions. Journal of the American Statistical Association 67, 842–846.

Ritter, C., Simar, L., 1997. Pitfalls of normal-gamma stochastic frontiers and panel data. Journal of Productivity Analysis 8, 167–182.

Stevenson, R.E., 1980. Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13, 58–66.

Tran, K.C., 1998. Estimating mixtures of normal distributions via empirical characteristic function. Econometric Reviews 17, 167–183.

Tsionas, E.G., 2001a. Posterior analysis of stochastic frontier models with truncated normal errors. Computational Statistics 16, 559–575.

Tsionas, E.G., 2001b. Finite mixture stochastic frontier models, manuscript, Department of Economics. Athens University of Economics and Business.

Tsionas, E.G., 1999. Monte Carlo inference in econometric models with symmetric stable disturbances. Journal of Econometrics 88, 365–401.

Tsionas, E.G., 2000. Full likelihood inference in normal-gamma stochastic frontier models. Journal of Productivity Analysis 13, 179–201.

Yu, J., 2004. Empirical characteristic function estimation and its applications. Econometric Reviews 23, 93–123.