MAXIMUM LIKELIHOOD ESTIMATION OF STOCHASTIC FRONTIER MODELS BY THE FOURIER TRANSFORM

by

Efthymios G. Tsionas*

DISCUSSION PAPER No. 172

March 2006

Department of Economics
Athens University of Economics and Business
76 Patission Str., Athens 104 34, Greece
Tel. (++30) 210-8203911 - Fax: (++30) 210-8203301

* Department of Economics, Athens University of Economics and Business.

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.
Phone: (++30210) 820 3388, Fax: (++30210) 820 3301
E-mail: tsionas@aueb.gr

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. 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.

Key Words: Stochastic frontiers; Moment generating function; Characteristic function; Fast Fourier transform; Maximum likelihood; Generalized method of moments.

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, Lovell, and

[1] The standard references are Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). For reviews see Bauer (1990), Greene (1993, 1999) and Kumbhakar and Lovel (2000).
[2] See Ritter and Simar (1997) for some disturbing properties of the normal-*gamma* model.

Schmidt (1977). But, for better or worse, the normal/half-normal model has dominated the received empirical literature" (Greene, 1999, p. 104).

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 (2004) 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.

The remaining 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 the Appendix 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]

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

where x_i is a $k \times 1$ vector of explanatory variables, β is a $k \times 1$ vector of parameters, $v_i \sim iidN(0, \sigma_v^2)$ and u_i s 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_i s. The probability density function of the composed error, $\varepsilon_i = v_i + u_i$, is given by $f_\varepsilon(\varepsilon;\gamma) = \int_0^\infty (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{(\varepsilon-u)^2}{2\sigma^2}\right] f(u \mid \gamma) du$. 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

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

It is known that $E(u) = \alpha\theta$ and $Var(u) = \alpha\theta^2$. 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 $E(z^{\alpha-1})$ when $z \sim N(-e_i - \theta^{-1}\sigma^2,\ \sigma^2)$ 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 $\varphi_X(t) = E[\exp(Itx)] = \int_{-\infty}^{\infty} \exp(Itx) f_X(x) dx$, for all $t \in \mathbb{R}$. Here, $I = \sqrt{-1}$ denotes complex unity. In the context of the normal-*gamma* stochastic

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

frontier model, the CF of the two-sided error is $\varphi_v(t) = \exp\left(-\frac{1}{2}\sigma_v^2 t^2\right)$, and the CF of the one-sided error is $\varphi_u(t) = (1 - It\theta)^{-\alpha}$. Therefore, the CF of the composed error is

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

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 denotes cost efficiency it is natural to assume[4] that this is *beta* distributed with parameters p and q, and density $f(r \mid p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} r^{p-1}(1-r)^{q-1}$, $0 \le r \le 1$, $p,q > 0$. 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

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

Next we turn attention to implementing ML estimation. Suppose $f_\varepsilon(e)$ 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 (Lucacs and Laha, 1964, p. 20) we have

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

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 $x_k = (k - 1 - \frac{N}{2})h$, $k = 1,...,N$, for some step length $h > 0$. If we set $t = 2\pi\omega$ and change variables then

[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.

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

This integral can be approximated using the trapezoidal rule to obtain

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

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

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

which can be computed efficiently using the FFT applied to the sequence $(-1)^{n-1}\varphi_\varepsilon(2\pi\omega_n)$, $n = 1,...,N$. 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. 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 $N = 2^{12} = 4{,}096$ 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 $\bar{e} = \max_{i=1,..,n} |e_i|$.

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. One such procedure has been proposed by Feuerverger and McDunnough (1981a,b) and involves fitting the

theoretical to the empirical CF using nonlinear generalized least squares[5]. Suppose $\varphi_\varepsilon(t)$ denotes the theoretical CF -of the composed error- and its empirical counterpart is

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

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

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

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

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

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

and

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

and collect them[6] in the vector $Y = Y(\Theta)$ where Θ denotes all unknown parameters. Suppose $Y(\Theta) = [Y_h(\Theta), h=1,...,2J]$ and $\widetilde{Y} = [\widetilde{Y}_h, h=1,...,2J]$, and consider the artificial regression

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

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 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_j s. To describe the estimator, the asymptotic form of the log-likelihood may be taken as

$$\ln L(\Theta) = const. + \tfrac{1}{2}\log|\,\boldsymbol{\Sigma}^{-1}(\Theta)\,| - \tfrac{N}{2}[Y(\Theta)-\widetilde{Y}]'\boldsymbol{\Sigma}^{-1}(\Theta)[Y(\Theta)-\widetilde{Y}] \tag{6}$$

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

or as just the last term, where $N^{-1}\boldsymbol{\Sigma}(\Theta)$ is the covariance matrix of $\widetilde{Y}$, and the elements of $\boldsymbol{\Sigma}(\Theta)$ are given in equation (2.2) of Feuerverger and McDunnough (1981a). 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 $[Y(\Theta)-\widetilde{Y}]'\boldsymbol{\Sigma}^{-1}(\Theta)[Y(\Theta)-\widetilde{Y}]$ or the quadratic form $[Y(\Theta)-\widetilde{Y}]'\widetilde{\boldsymbol{\Sigma}}^{-1}[Y(\Theta)-\widetilde{Y}]$ where $\widetilde{\boldsymbol{\Sigma}}$ is any consistent estimator of $\boldsymbol{\Sigma}^{-1}(\Theta)$. 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 $\boldsymbol{\Sigma}$ and use it to compute the 2SANF estimator. It has been found that if parameters and $\boldsymbol{\Sigma}$ are updated simultaneously in 2SANF then the $\boldsymbol{\Sigma}$ matrix can become non-positive-definite numerically[7] and convergence of the numerical optimization algorithm becomes impossible. Since the estimate of $\boldsymbol{\Sigma}$ obtained from the 1SANF estimator is consistent, the asymptotic distribution of the 2SANF estimator is the same no matter whether parameters and $\boldsymbol{\Sigma}$ 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 (5) the e_i s are not observed. Consider $y_i = \beta_0 + x_i'\alpha + \varepsilon_i$, where $\varepsilon_i = v_i + u_i$. Then, we can use $e_i = y_i - x_i'\tilde{\alpha}$, where $\tilde{\alpha}$ 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

---

[6] 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.

[7] The same computational experience is suggested by several other authors, see Feuerverger and McDunnough (1981) and Madan and Seneta (1982). Madan and Seneta (1982) suggest using a

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).

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 U.S. 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 (2004).

The five output variables are: 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: 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.

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[8]. Greene (1999) rightly argues that this is not a serious restriction since one can use LS to estimate cost function parameters other

generalized inverse of the $\mathbf{\Sigma}$ matrix. In the applications considered here this has been followed but it has not been found necessary since a simple inverse sufficed.

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 $\widetilde{\mathbf{\Sigma}}$ is computed and is used to implement 2SANF which is minimized with a conjugate gradient technique. Standard errors are computed from $N^{-1}\left(\mathbf{G}'\widetilde{\mathbf{\Sigma}}^{-1}\mathbf{G}\right)^{-1}$ where $\mathbf{G}=\dfrac{\partial Y(\Theta)}{\partial\Theta'}$ 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. $[Y(\Theta)-\widetilde{Y}]'\widetilde{\mathbf{\Sigma}}^{-1}[Y(\Theta)-\widetilde{Y}]$ 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 $2^{12}=4{,}096$ grid points for the characteristic function and step length $h=2\max|e_i|/2^{12}$ 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[9] 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. In **Table 2** reported are the results of estimation by ML-FFT 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 minutes on a PC with 800 MHz speed.

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

[8] 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.
[9] The column headed IO means "input-oriented". IO frontier models are discussed in section 5.

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

In **Table 4** 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 5**. 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[10].

Before proceeding it should be mentioned that a potentially important inefficiency distribution, the truncated normal, has been let 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, 2001), so it is useful to consider alternative estimators based on the CF. A detailed examination is contained in the Appendix 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.

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

[10] 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.

often than not are founded on the assumption of input-oriented (IO) efficiency. Kumbhakar and Tsionas (2004) propose a translog production frontier[11] 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

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

where Y_i is a scalar output, $X_i = (X_{1i},...,X_{Ji})$ is the vector of inputs actually used, $\Theta_i = X_{ji}^e / X_{ji} \le 1$ for $j = 1,...,J$, 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 $j = 1,...,J$ 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

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

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[12], 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 $g(\theta_i, x_i, T_i) = [\tfrac{1}{2}\theta_i^2 \Psi - \theta_i \Xi_i]$, $\Psi = 1_J'\Gamma 1_J$, and $\Xi_i = 1_J'(\beta + \Gamma x_i + \varphi T_i)$, $i = 1,...,n$.

Note 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

[11] 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.

[12] 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.

constant multiple of θ_i, (viz., $[\frac{1}{2}\theta_i^2\Psi - \theta\Xi_i] = -r\theta_i$), 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

$$y_i = z_i'\alpha + \tfrac{1}{2}\theta_i^2\Psi - \theta_i\Xi_i + v_i,\ i = 1,...,n. \tag{8}$$

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

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

where ϕ denotes the entire parameter vector. The integral above is not available in closed form so Kumbhakar and Tsionas (2004) proposed to approximate it using Monte Carlo methods resulting in the method of MSL[13]. 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 $\varphi_\varepsilon(t) = \exp(-\frac{1}{2}\sigma_v^2 t^2)\int_0^\infty \exp[It(\frac{1}{2}\Psi\theta^2 - \theta\Xi)]\, f_\theta(\theta)d\theta$ by theorem 1.5.9 in Lukacs and Laha (1964, p.22) which states that if the random variable X has

[13] 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/(Ξ-Ψθ(g)) and θ(g) can be obtained as the unique positive root of the polynomial (Ψ/2)θ²-Ξθ-g=0, provided g≤0 and Ψ<0. However, the problem is that the convolution of v and g is not available in closed form.

distribution function $F(x)$, and $Y = S(X)$ where S is finite, single-valued and Borel- measurable then $\varphi_Y(t) = E[\exp(ItS(x)] = \int \exp(ItS(x)]dF(x)$.

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

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

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

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 $E(g_i) = I^{-1}\left.\dfrac{d\varphi_g(t)}{dt}\right|_{t=0} = \frac{1}{2}\Psi\sigma_\theta^2 - (2/\pi)^{1/2}\sigma_\theta\Xi_i$

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[14]. 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

$$\varphi_\varepsilon(t) = (2\pi)^{1/2}\sigma_\theta\sigma_*\exp(`-\tfrac{1}{2}\sigma_v^2t^2 - \tfrac{1}{2}It\Psi m_*^2)\Phi(m_*/\sigma_*), \qquad (11)$$

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

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

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

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

[14] Instead, Kumbhakar and Tsionas (2004) resorted to numerical integration to compute the first moment of g.

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

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

where

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

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 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[15] 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

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

parameter vector. The FFT is applied to obtain the density for each value of Ξ in the grid and over the fixed grid for e s. 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[16].

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_i s. 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[17], difficulty to obtain convergence *etc*.

[16] 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.

[17] For this application timing was 0.88 for the OO model and 1.38 minutes for the IO model.

The results are reported in **Table 6** 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 characteristic function 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 given 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 modeling informational asymmetries in other contexts as well, an expansion of the distributional alternatives is necessary. Here, we consider *gamma* distributed error terms[18], that is

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

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

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

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

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

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 2,178 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 hours 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 7. 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.

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 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.

References

Aigner, D., C.A.K. Lovell, and P. Schmidt, 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., and C. Hammond, 1987, A tractable likelihood function for the normal-gamma stochastic frontier model, *Economics Letters* 24, 33-38.

Bellman, R., 1970, *Methods of nonlinear analysis*, volume I, Academic Press, New York.

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

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

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

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

Greene, W.H., 1993, The econometric approach to efficiency analysis, in: H.O. Fried, C.A.K. Lovell and S.S. Schmidt, eds., *The measurement of productive efficiency: Techniques and applications* (Oxford University Press, Oxford) 68-119.

Greene, W.H., 1999, Frontier production functions, in M. Hashem Pesaran and Peter Schmidt, *Handbook of Applied Econometrics*, volume II: Microeconomics, Blackwell, Malden, Massachusetts, pp. 81-166.

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

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

Kopp, R., and J. Mullahy, 1989, Moment-based estimation and testing of stochastic frontier models, Discussion Paper 89-10, Resources for the Future, Washington, DC.

Kumbhakar, S.C. and C.A.K Lovell, 2000, *Stochastic frontier analysis* (Cambridge University Press, New York).

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

Kumbhakar, S.C., and E.G. Tsionas, 2004, Estimation of stochastic frontier production functions with input-oriented technical inefficiency, State University of New York at Binghamton, working paper.

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

Lukacs, E. and R.G. Laha, 1964, *Applications of characteristic functions*, Charles Griffin and Company Limited, London.

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

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

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

Ritter, C., and L. Simar, 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., 2000, Full likelihood inference in normal-gamma stochastic frontier models, *Journal of Productivity Analysis* 13, 179-201.

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

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

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)	0.379 (0.262)

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 $f(u;\alpha,\theta) = \theta^{-\alpha}\Gamma(\alpha)^{-1}u^{\alpha-1}\exp(-u/\theta)$. The Hansen J-test is 95.50. The log-likelihood function from ML-FFT was 53.15. Execution time for ML-FFT was 2.93 minutes on a Pentium 800 MHz machine running Windows 98 under WinGauss 3.2.38.

Table 2. Sensitivity of ML-FFT for normal-gamma model of US banks

	N =10	N =12	N =16
Constant	1.042 (0.249)	1.043 (0.249)	1.043 (0.249)
$\log w_1$	0.199 (0.038)	0.199 (0.038)	0.199 (0.038)
$\log w_2$	0.061 (0.015)	0.061 (0.015)	0.061 (0.015)
$\log w_3$	0.110 (0.025)	0.110 (0.025)	0.110 (0.025)
$\log w_4$	0.043 (0.024)	0.043 (0.024)	0.043 (0.024)
$\log y_1$	0.053 (0.014)	0.053 (0.014)	0.053 (0.014)
$\log y_2$	0.463 (0.014)	0.463 (0.014)	0.463 (0.014)
$\log y_3$	0.229 (0.020)	0.229 (0.020)	0.229 (0.020)
$\log y_4$	0.043 (0.007)	0.043 (0.007)	0.043 (0.007)
$\log y_5$	0.160 (0.019)	0.160 (0.019)	0.160 (0.019)
σ	0.162 (0.014)	0.162 (0.014)	0.162 (0.014)
α	0.379 (0.262)	0.379 (0.262)	0.379 (0.262)
θ	0.276 (0.097)	0.276 (0.097)	0.276 (0.097)
Computing time (min)	0.67	2.29	212.3

Notes: ML-FFT has been tried for N=6 but it was not possible to obtain convergence within 200 iterations. 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. The results for N=12 are the same as those reported in Table 1.

Table 3. 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 minutes on a Pentium 800 MHz machine running Windows 98 under WinGauss 3.2.38. The mean log-likelihood is 0.106383.

Table 4. 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.

Table 5. 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.

Table 6. 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 minutes.

Table 7. 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
½ (education)^2	-0.0052	0.0012	-0.0052	0.0011
½ (experience)^2	-0.0094	0.0004	-0.0094	0.0004
½ (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	

APPENDIX. Performance of various procedures for the normal/truncated-normal stochastic frontier model.

A potentially important distribution in applications is the ***truncated normal*** for technical inefficiency, $u \sim N(\mu, \sigma_u^2)$ 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 (2001) 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 informative to aid practical identification of the model. 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

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

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

The ML estimates of a truncated normal frontier are reported in **Table A1**. 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. 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 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[19] an artifact. I have contacted 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 (2001). 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 to 20 equi-distant 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 A2** 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

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

empirical log MGF using the US commercial banking data. The estimator is not very sensitive to how the grid points are selected 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 estimates of ω range from -0.0062 when G =4 to 0.0165 when G =50.

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 $\sqrt{Var(u)/Var(v)}$ 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 1,000 Monte Carlo replications, with the endpoint being T =2 using 10 points from -T to T. The sampling distributions are reported in **Figure A1**. 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.

It should be mentioned that our objective was not (and indeed could not be) to improve on the ML estimator by proposing a GMM estimator with better small sample behavior in the context of the truncated normal model. If the model *is* identified in practice then ML will behave nicely provided the sample is "large enough". The purpose was only to propose a method of moments estimator that overcomes certain practical problems associated with the ML estimator. Under ideal conditions GMM like ML gives reasonable and useful results but of course far less than ideal conditions are encountered in reality so -as is well known- working with the truncated normal model is extremely difficult.

Finally, I have implemented the 2SANF method based on the FFT for the truncated normal stochastic frontier. As in the normal-*gamma* model, the 1SANF is implemented first and the covariance matrix is estimated and used subsequently in 2SANF. The method is found to be more satisfactory compared to GMM[20]. The results for various configurations of the grid points are reported in **Table A3**. The final estimates are not very sensitive as has been found to be the case with GMM so the 2SANF appears somewhat robust at least in this application. However, standard errors depend on the configuration of the grid points. With too few grid points standard errors are inflated owing to the bad conditioning of $\boldsymbol{\Sigma}$, *i.e.* the covariance matrix of the characteristic function. As the number of grid points increases the standard errors become more reasonable. It should be mentioned that the computational requirements of 2SANF are somewhat excessive in that an ill-conditioned $G \times G$ has to be inverted (once). So for T =5 it is necessary to invert a 500×500 matrix. Here, I have used a pseudo-inverse and numerical gradients to obtain the asymptotic standard errors for the 2SANF estimator. As a general conclusion it might be conjectured that if an extended grid is needed then the grid must be extremely fine -so fine that the dimension of the covariance matrix becomes too large to manage or to ensure that it behaves nicely in the numerical sense. From the results reported it seems that ω is between -0.773 and -0.682 but it is definitely statistically insignificant at conventional levels of significance. Based on these results the assumption of half-normality for the one-sided error term does not seem to be at odds with the data.

Following are some technical details regarding Table A3. Although parameter estimates are reasonably close when the grid configurations vary -and therefore the 2SANF estimator can be useful in practice- standard errors computed from inverting the ill-conditioned matrix

[20] It should be kept in mind that both 2SANF and MGF-based procedures can be thought of as GMM since 2SANF is GMM when the moment conditions are obtained from the characteristic function.

$\mathbf{G}'\boldsymbol{\Sigma}^{-1}\mathbf{G}$ are very sensitive. Useful results can be obtained by applying Tychonov regularization (see Bellman, 1970, page 86) to the matrix[21]. In practice the regularizing parameter is set to $\frac{1}{100}\min\left(1, \max_{i=1,...,k} a_{ii}\right)$ where $A=[a_{ij}, i, j=1,....,k]$ is the matrix in question. The results indicate that the standard errors are robust in two cases, viz. in the interval 0.1 to 10 provided we have more than 10 points and in the interval 0.0001 to 5 (or 10) with nearly 100 points. When the endpoints are 5 or 1.5 or when too few points are considered the standard errors we get are different so it seems yet again that the grid must be sufficiently fine and extended. With 500 points in the interval from 0.0001 to 5 we get some differences compared to results when we have 20 or 100 points especially in the estimation of standard error for σ_v.

I have tried other ways to compute standard errors. One approach is to compute the numerical Hessian of the asymptotic normal form (using a numerical gradient of the characteristic function to approximate the elements of $\mathbf{G}$. Another way is to approximate the Hessian of the log-likelihood corresponding to asymptotic normal form using the estimated Hessian from the last update of Newton-Raphson iterations that was used to minimize the quadratic form in the asymptotic normal form. Both approaches do not seem to be useful in practice. In the first approach, usually, very few Newton iterations are taken and the approximation to the Hessian is not accurate enough. In the second approach the results are sensitive to the step size in computing numerical derivatives and quite often, in the banking data, the diagonal element corresponding to parameter σ_u was found to be too close to zero for the Hessian of the quadratic form to be positive definite. It seems that without due care the computation of standard errors from 2SANF can be a problem but parameter estimation is more or less immune to how the grid is specified or how certain matrices behave.

These results beg the question what is a reasonable approach to select the extent of the grid and the number of grid points when it comes to estimating the standard errors -the reader is reminded here that estimation of parameters themselves is not so sensitive to the configuration of the grid. In order to provide an answer I turn to the following experiment. Given the characteristic function and parameter values of the model, the covariance matrix of estimates from 2SANF can be computed for various grid configurations. It is then natural to look for the configuration that maximizes the information, *i.e.* maximizes $\mathbf{G}'\boldsymbol{\Sigma}^{-1}\mathbf{G}$ in some sense. One sense in which this can be maximized is to maximize the sum of its eigenvalues. The reader is reminded that this is *not* a Monte Carlo experiment because we do not generate artificial data because we use the characteristic function, not the likelihood function: The information matrix can be computed easily because in turn depends on the grid points only; its dependence on the actual data is irrelevant. It is not my intention to report detailed results here covering all possible parameterizations and all possible grid configurations but only to provide a rough idea of what the optimal grid selection might be like in practice.

Since the results depend on the value of β_0 it does not seem worthwhile to report results from a full-blown experiment covering all possible values of the parameters. For this reason I set $\sigma_v = 0.15$ and $\sigma_u = 0.28$, the values from ML estimation for the US commercial banking data. Apparently these values are "empirically plausible". Then I examine the configuration of the grid that maximizes information for different combinations of β_0 and ω. I examine the case when T is 1, 5, 10, 15, 20, 50, 100 and 150 and k is 5, 10, 15, 20, 25, 50.

The results are reported in **Table A4**. If the true value of β_0 is close to 0.7 (as ML estimation of the half-normal model suggests) then if ω is close to unity we should choose $T=20$ and $k=10$. For other values of ω the choice of T is 150 or as large as possible subject to numerical considerations. I have tried larger values for T to see whether we can

[21] In the econometrics literature this is known as ridge regression.

obtain an interior maximum for (T,k) but unfortunately the maximum obtains at so large values of T (up to 600) that numerical problems begin to arise. What also seems to be clear is that the choice of k is not so difficult and values around 20 would be good in practice. It is the choice of T that requires some experimentation in practice at least over a set of parameter values which seem possible based on a priori consideration or results from ML estimation of related models.

Table A1. 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 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. GMM is implemented using fitting of the empirical to the theoretical MGF using nonlinear least squares with 20 equi-distant points from -2 to 2. The Hansen J-statistic (N times the objective function) was 0.00041.

Table A2. 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 gradients 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 taken Table 6.

Table A3. Sensitivity of 2SANF to configuration of grid points

	Parameter estimate, standard error and "precision"			
Grid endpoints and number of endpoints	Constant	σ_v	σ_u	ω
0.1 to 5, G=20	1.039 (25125.) (216.8) (0.239)	0.011 (4751.) (8.86) (1.127)	0.215 (241.4) (194.2) (0.058)	-0.612 (117233.) (46.7) (1.103)
0.1 to 5, G=50	1.072 (21.12) (222.8) (0.226)	0.011 (4.21) (9.32) (1.033)	0.215 (0.215) (190.0) (0.053)	-0.766 (98.18) (47.90) (1.043)
0.1 to 5, G=100	1.074 (0.077) (228.3) (0.06)	0.011 (0.013) (79.84) (0.013)	0.214 (0.005) (197.8) (0.005)	-0.775 (0.359) (49.05) (0.278)
0.1 to 10, G=10	1.059 (151.6) (219.4) (0.131)	0.012 (9068.) (15.11) (0.660)	0.201 (532.8) (265.3) (0.039)	-0.702 (34.50) (44.04) (0.649)
0.1 to 10, G=20	1.067 (148.1) (306.2) (0.089)	0.013 (9380.) (22.18) (0.450)	0.201 (584.3) (364.9) (0.028)	-0.743 (7.03) (61.61) (0.442)
0.1 to 10, G=50	1.06 (10269.) (320.07) (0.089)	0.012 (---) (22.60) (0.446)	0.201 (---) (383.7) (0.027)	-0.713 (50996.) (64.44) (0.438)
0.1 to 10, G=100	1.072 (3.919) (320.8) (0.088)	0.012 (0.981) (22.74) (0.406)	0.201 (0.060) (381.5) (0.025)	-0.771 (19.49) (64.51) (0.438)
0.0001 to 5, G=20	1.030 (27141.) (216.16) (0.239)	0.011 (6724.) (8.84) (1.129)	0.215 (341.7) (197.9) (0.006)	-0.569 (126205.) (46.54) (1.106)
0.0001 to 5, G=100	1.074 (0.159) (228.18) (0.088)	0.011 (0.059) (19.87) (0.057)	0.214 (0.006) (197.85) (0.006)	-0.775 (0.742) (48.96) (0.410)
0.0001 to 5, G=500	1.073 (0.165) (229.72) (0.082)	0.011 (0.009) (117.39) (0.009)	0.214 (0.005) (200.6) (0.005)	-0.773 (0.768) (49.32) (0.38)

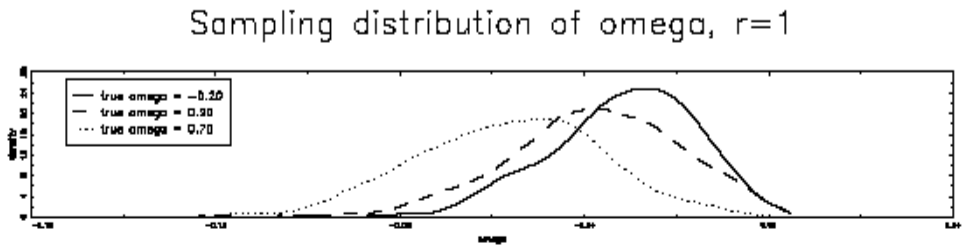
0.0001 to 10, G=20	1.15 (104.3) (308.5) (0.088)	0.014 (11293.) (24.11) (0.446)	0.201 (758.) (362.) (0.030)	-1.151 (51.75) (62.07) (0.438)
0.0001 to 10, G=100	1.074 (10.53) (320.8) (0.088)	0.012 (1.33) (22.70) (0.408)	0.201 (0.081) (22.70) (0.025)	-0.780 (52.41) (64.51) (0.438)
0.0001 to 1.5, G=10	0.97 (19710.) (57.05) (2.28)	0.011 (5568.) (0.85) (10.27)	0.226 (273.3) (45.45) (0.577)	-0.267 (86608.) (12.88) (10.04)
0.0001 to 1.5, G=50	0.97 (0.144) (68.86) (0.144)	0.012 (0.134) (9.57) (0.043)	0.225 (0.043) (53.80) (0.043)	-0.304 (0.656) (15.55) (0.621)
0.0001 to 1.5, G=100	1.321 (0.75) (79.42) (0.124)	0.016 (0.011) (92.83) (0.007)	0.226 (0.036) (70.43) (0.035)	-1.843 (3.341) (17.86) (0.528)

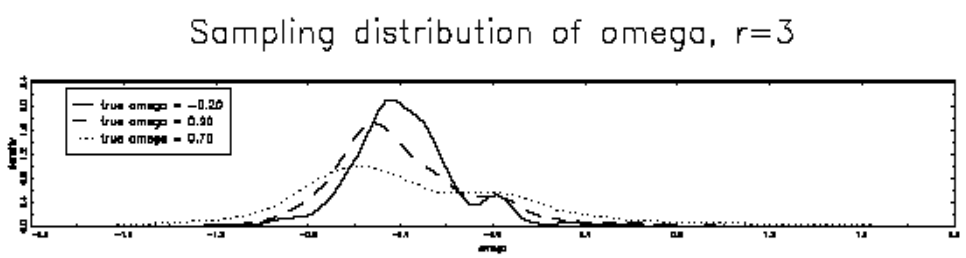
Notes: G denotes the number of grid points to implement the 2SANF estimator. The first number in parentheses below estimate is the standard error computed from the square root of the diagonal elements of $\frac{1}{N}(\mathbf{G}'\boldsymbol{\Sigma}^{-1}\mathbf{G})^{-1}$. The second number in parentheses is the square root of the corresponding diagonal element of $\mathbf{G}'\boldsymbol{\Sigma}^{-1}\mathbf{G} = N \cdot \mathbf{P}$, the precision (inverse covariance) matrix of 2SANF. The third number is the standard error computed as follows: First, matrix $\mathbf{P}$ is subjected to Tychonov regularization, and then it is inverted. The square roots of the diagonal elements of $(1/N)$ times the inverse are reported. Also, (---) indicates that the corresponding standard error could not be computed because the diagonal element of the covariance matrix was negative due to bad conditioning.

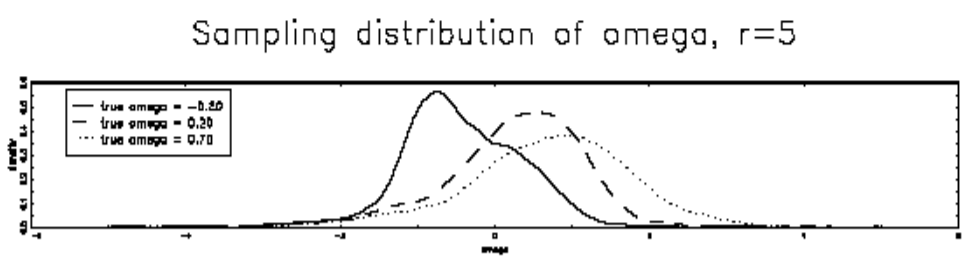
Table A4. Selection of grid points by maximum information

	β_0						
ω	-0.7	-0.2	0.0	0.2	0.7	1.0	1.5
-2.0	150,10	---	100,10	150,10	10,25	150,10	150,10
-1.0	100,5	150,10	150,50	50,50	150,5	50,5	150,10
0.0	50,20	20,50	150,15	150,5	150,5	100,5	50,20
1.0	20,50	50,5	150,10	15,25	20,10	150,5	150,15
2.0	50,5	---	50,20	10,20	150,10	150,5	150,10

Notes: The table reports the pairs of (T,k) that are optimal in maximizing sample information when $\sigma_v = 0.15$ and $\sigma_u = 0.28$, the values from ML estimation of the normal-half-normal model for the US commercial banking data. A (---) indicates that numerical problems prevented obtaining the optimal pair.


Sampling distribution of omega, r=1
true omega = -0.20
true omega = 0.20
true omega = 0.70


Sampling distribution of omega, r=3
true omega = -0.20
true omega = 0.20
true omega = 0.70


Sampling distribution of omega, r=5
true omega = -0.20
true omega = 0.20
true omega = 0.70

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF ECONOMICS

D I S C U S S I O N P A P E R S

(An electronic version of the paper may be downloaded from: www.de.aueb.gr)

162. **George D. Demopoulos**, European Chair Jean Monnet and AUEB, **E. Fratzeskos**, Bank of Cyprus and AUEB and **P. Kapopoulos**, Emporiki Bank and AUEB. FISCAL ADJUSTMENT AND UNEMPLOYMENT RATE: THE EUROPEAN EXPERIENCE. (Nov. 2004).
163. **Nikolaos Floros,** Cooperative Bank of Chania **and Andriana Vlachou,** Athens University of Economics and Business. ENERGY DEMAND AND ENERGY-RELATED CO_2 EMISSIONS IN GREEK MANUFACTURING: ASSESSING THE IMPACT OF A CARBON TAX. Forthcoming in: *Energy Economics*.
164. **George C. Bitros.** MODELS FOR RANKING EUROPEAN INSTITUTIONS OF HIGHER LEARNING WITH AN APPLICATION TO DATA FROM GREECE. (June 2005).
165. **Kostas Axarloglou**, Democritus University of Thrace, **and Mike Pournarakis**. CAPITAL INFLOWS IN THE BALKANS: FORTUNE OR MISFORTUNE? (June 2005).
166. **Emmanuel Drandakis.** THE ENVELOPE THEOREM AT WORK: A. COST MINIMIZATION IN THE SHORT AND LONG RUN. (June 2005).
167. **Kostas Axarloglou**, Democritus University of Thrace, **and Mike Pournarakis**. DO ALL FOREIGN DIRECT INVESTMENT INFLOWS BENEFIT THE LOCAL ECONOMY? (July 2005).
168. **Nicholas C. Baltas.** THE IMPACT OF ENLARGEMENT ON EU AGRICULTURE. Forthcoming in: C. Stefanou (Ed.), Book: *Adjusting to the EU Enlargement: Recurring Issues in a New Setting*. Cheltenham, Elgar Edward Publishing Co., 2006, ch.3, pp. 49-61.
169. **George C. Bitros, and Epaminondas E. Panas.** THE INFLATION-PRODUCTIVITY TRADE-OFF REVISITED. (Nov. 2005).
170. **George C. Bitros and Elias G. Flytzanis.** ON THE OPTIMAL LIFETIME OF ASSETS. (Dec. 2005).
171. **George C. Bitros.** SOME NOVEL IMPLICATIONS OF REPLACEMENT AND SCRAPPING. (Dec. 2005).
172. **Efthymios G. Tsionas**. MAXIMUM LIKELIHOOD ESTIMATION OF STOCHASTIC FRONTIER MODELS BY THE FOURIER TRANSFORM. (March 2006).