Simulated Likelihood Estimation of the Normal-Gamma Stochastic Frontier Function

William H. Greene*
Stern School of Business
New York University

September 30, 2000

Abstract: The normal-gamma stochastic frontier model was proposed in Greene (1990) and Beckers and Hammond (1987) as an extension of the normal-exponential proposed in the original derivations of the stochastic frontier by Aigner, Lovell, and Schmidt (1977). The normal-gamma model has the virtue of providing a richer and more flexible parameterization of the inefficiency distribution in the stochastic frontier model than either of the canonical forms, normal-half normal and normal-exponential. However, several attempts to operationalize the normal-gamma model have met with very limited success, as the log likelihood is possesed of a significant degree of complexity. This note will propose an alternative approach to estimation of this model based on the method of simulated maximum likelihood estimation as opposed to the received attempts which have approached the problem by direct maximization.

I would like to thank the participants at the North American Productivity Workshop, held at Union College in Schenectady, New York, June 15-17, 2000, for helpful comments.

*Department of Economics, Stern School of Business, New York University, 44 West 4th St., New York, NY, 10012; http://www.stern.nyu.edu/~wgreene

1. Introduction

The stochastic frontier model was proposed (nearly simultaneously by researchers on three continents) in 1977 by Aigner, Lovell and Schmidt (1977), Meeusen and van den Broeck (1977) and Battese and Corra (1977). The original form of the model,

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u$$

where $\boldsymbol{\beta}'\mathbf{x} + v$ constitute a conventional regression model and u is a one side disturbance that is distributed either as half normal or exponential, has stood the test of nearly 25 years as the workhorse of the literature on frontier estimation. Notwithstanding the original model's longevity and distinguished service record, researchers have proposed many variants of the model in attempts to generalize the distribution of the inefficiency distribution, $f(u)$.

The normal-gamma stochastic frontier model was proposed in Greene (1990), Beckers and Hammond (1987) and Stevenson (1990) as an extension of the normal-exponential proposed in the original derivations of the stochastic frontier by Aigner, Lovell, and Schmidt (1977). The normal-gamma model provides a richer and more flexible parameterization of the inefficiency distribution in the stochastic frontier model than either of the canonical forms, normal-half normal and normal-exponential. However, several attempts to operationalize the normal-gamma model have met with very limited success, as the log likelihood is possesed of a significant degree of complexity. Greene (1990) attempted a direct, but crude maximization procedure which, as documented by Ritter and Simar (1997) was not sufficiently accurate to produce satisfactory estimates. (Ritter and Simar concluded from their work that even an accurate estimator of this model would suffer from significant identification problems.) Stevenson (1980) made note of the difficulties of estimation early on and proposed limiting attention to the Erlang from (noted below), which is a significant restriction of the model. Beckers and Hammond (1987) derived a form of the log likelihood which showed some potential, but, in the end, remained exceedingly complicated and was never operationalized. We note in passing at this point the work of Tsionas (2000), who made considerable progress on the model in a Bayesian framework, but provided estimates of the posterior distribution of u, rather than direct estimation of the parameters. This note will confine attention to classical, parametric analysis of the model.

We will propose an alternative approach to estimation of this model based on the method of simulated maximum likelihood estimation as opposed to the received attempts which have approached the problem by direct maximization. The previous work on this model has approached the estimation problem by direct maximization of what turns out to be a very complicated log likelihood function. As shown below, the log likelihood function for the normal-gamma model is actually the expectation of a random variable which can be simulated. This suggests the method of simulated maximum likelihood as a method of estimation.

In Section 2 below, we will briefly review the modeling framework for the stochastic frontier model leading to the normal-gamma model. Section 3 will analyze the problem of maximizing the log likelihood function. The simulated log likelihood function is derived, then two practical aspects of the estimation process are presented.

First, we note a useful method of producing the necessary random draws for the simulation - the model requires simulation of draws from a truncated distribution, which is, itself, a significant complication. Second, we introduce a new tool which has proved very useful in the simulated likelihood maximization of models for discrete choice, the Halton draw. Halton draws provide a method of dramatically speeding up the process of maximization by simulation. A brief application of the technique is presented in Section 4. Some conclusions are drawn in Section 5.

2. The Stochastic Frontier Model

The motivation for and mechanical details of estimation of the stochastic frontier model have appeared at many places in the literature, so it is not necessary to reiterate them here. For further details, the reader is referred, for example, to one of the many surveys on the subject, such as Kumbhakar and Lovell (2000). In this section, we will recall the aspects of the model that appear in the results below.

For convenience, we suppress the observation subscripts that denote an observed sample of N observations, $i = 1,...N$. The stochastic frontier model is

$$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$$

$$\varepsilon = v - u$$

$$v \sim N[0,\sigma_v^2]$$

$$u \geq 0, \text{ with continuous density, } f(u \mid \boldsymbol{\theta}), \text{ where } \boldsymbol{\theta} \text{ is a vector of parameters.}$$

The objective of estimation is $\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_v^2$ then u. As a practical matter, the firm specific inefficiency, u is the ultimate objective of the estimation. As has been widely documented, however, this is not feasible. Jondrow, Materov, Lovell, and Schmidt (1982) suggested the feasible alternative

$$E[u|\varepsilon] = g(\varepsilon, \boldsymbol{\theta})$$

which can be estimated using the estimated parameters and the observed data. [Properties of this estimator have been explored, e.g., by Horrace and Schmidt (1996).]

The received literature has relied on two specific formulations of the inefficiency distribution,

half normal: $u = |U|$, $U \sim N[0,\sigma_u^2]$,

and

exponential $f(u) = \theta\exp(-\theta u)$, $\theta > 0$.

As noted, the parameters of the distributions are of secondary importance in the estimation process. What is actually of greatest interest is the inefficiency component of the underlying model and estimation of values that will enable the comparison of individuals in the sample. The Jondrow, et al. formulas for the two models suggested are

$$E[u|\varepsilon] = \sigma\lambda/(1+\lambda^2)\,[\phi(\varepsilon\lambda/\sigma) / \{1-\Phi(\varepsilon\lambda/\sigma)\} - \varepsilon\lambda/\sigma].$$

where

$$\varepsilon = y - \boldsymbol{\beta}'\mathbf{x},$$

$$\lambda = \sigma_u / \sigma_v,$$

$$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2},$$

$\phi(t)$ and $\Phi(t)$ = standard normal density and distribution functions, respectively,

for the normal-half normal model and

$$E[u|\varepsilon] = z + \sigma_v\phi(q/\sigma_v) / \Phi(q/\sigma_v)$$

$$q = \varepsilon - \theta\sigma_v^2.$$

for the normal-exponential model.

The appeal of these two distributions is their known and quite straightforward closed form. See, e.g., the original paper by Aigner et al. (1977) or the survey by Kumbhakar and Lovell (2000). Many extensions have been proposed which layer deeper parameterizations on the mean and variance components of the half normal random variable, such as replacing the zero mean in the half normal distribution with a regression,

$$E[U|z] = \boldsymbol{\delta}'\mathbf{z}.$$

or adding heteroscedasticity in the variances, as in

$$\sigma_v^2 = \exp(\boldsymbol{\alpha}'\mathbf{h}).$$

[Recent extensive surveys which discuss these are Coelli et al (1997) or Kumbhakar and Lovell (2000).] However, even with these extensions, the normal-half normal has remained the workhorse of the literature.

3. The Normal-Gamma Stochastic Frontier Model

The normal-gamma frontier model provides an extension to the normal-exponential model;

$$f(u) = \theta^P/\Gamma(P) \exp(-\theta u)\, u^{P-1}.$$

This distribution provides a more flexible parameterization of the distribution. Figure 1 below illustrates a case in which the exponential and gamma variates both have mean 1, and the shape paremeter of the gamma density is $P = 1.5$. In the exponential model, $\theta = 1$, while in the gamma model, $\theta = 1.5$. The value of P larger than 1 allows the mass of the distribution to move away from zero - values of P less than one produce densities that

resemble the exponential distribution. As can be seen, the prior assumption of a value of P (e.g., 1) amounts to a substantive assumption about the distribution of inefficiencies in the population.

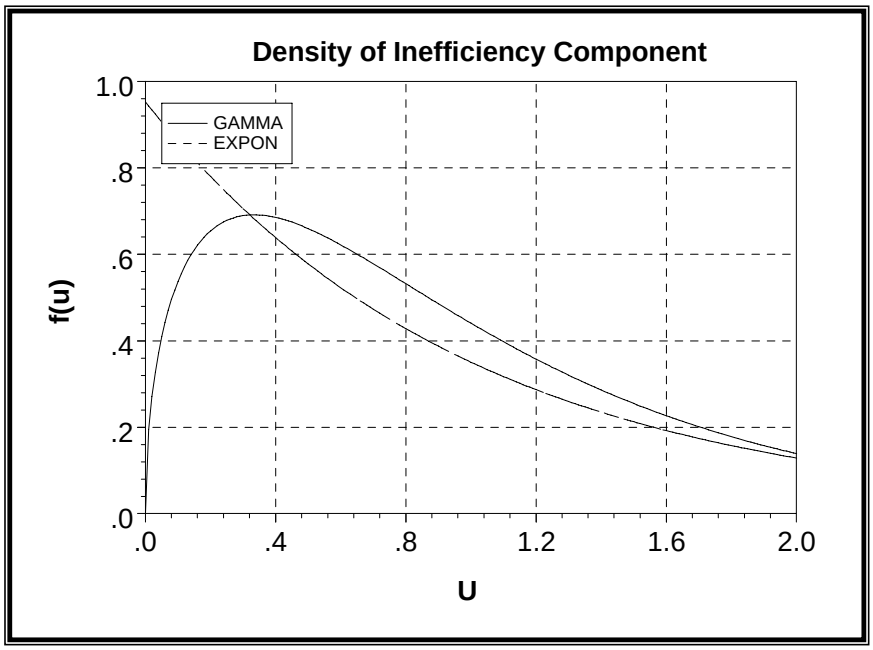


Figure 1. Illustrative Densities for Gamma and Exponential Models

3.1. Log Likelihood for the Normal - Gamma Stochastic Frontier Model

The log likelihood function for the normal-gamma model is derived in Greene (1990) and in a different form in Beckers and Hammond (1987). We will proceed directly to the end result here. For the normal-exponential (*NE*) model,

$$\log L_{NE} = N\{\log\theta + \tfrac{1}{2}\theta^2\} + \sum_{i=1}^{N} \{\theta\varepsilon_i + \log\Phi[-(\varepsilon_i/\sigma_v + \theta\sigma_v)]\}$$

where $\varepsilon_i = y_i - \boldsymbol{\beta}'\mathbf{x}_i$ $\varepsilon_{i.}$ The log likelihood for the normal-gamma (*NG*) model is that for the normal-exponential model plus the term which has complicated the analysis to date;

$$\log L_{NG} = \log L_{NE} + N[(P\text{-}1)\log\theta - \log\Gamma(P)] + \sum_{i=1}^{N} \log h(P\text{-}1,\varepsilon_i)$$

$$h(r,\varepsilon_i) = \frac{\int_0^\infty z^r \left(\frac{1}{\sigma_v}\right)\phi\left(\frac{z-\mu_i}{\sigma_v}\right)dz}{\int_0^\infty \left(\frac{1}{\sigma_v}\right)\phi\left(\frac{z-\mu_i}{\sigma_v}\right)dz}, \quad \mu_i = -\varepsilon_i - \theta\sigma_v^2 .$$

Note if $P = 1$, then $h(r,\varepsilon_i) = 1$ and $\log h(r,\varepsilon_i) = 0$ which returns the log likelihood for the exponential model

3.2. The Simulated Log Likelihood

In principle, the parameters of the model are estimated by maximizing the log likelihood function. The 'problem' that has hetetofore complicated the matter is computing $h(P\text{-}1,\varepsilon_i)$. Stevenson simplified the problem by suggesting the Erlang form, that is the cases with integer P. This does produce a simpler model, but drastically restricts the model. Beckers and Hammond reformulated the function in terms of a series of Pochammer symbols. (The Pochammer symbol is $a_x = \Gamma(a+x)/\Gamma(a)$. Accurate computation when x is small relative to a requires special routines.) While the Beckers and Hammond (1987) formulation did provide a complete (if not closed) functional form for the integral, it was not operationalized Greene's original application of the model used a very crude approximation to the integral based on Simpson's rule and the areas of trapezoids. Simar showed in two later papers that this approach was insufficiently accurate. Simar's results might have implied a full implementation, however he did not do the analysis from that viewpoint, so a method of proceeding for the practitioner remains to be developed. The purpose of this note is to suggest such a procedure.

We base an alternative approach on the result that, as can be seen by inspection, $h(r,\varepsilon_i)$ is the expectation of a random variable;

$$h(r,\varepsilon_i) = E[z^r \mid z \geq 0] \text{ where } z \sim N[\mu_i, \sigma_v^2] \text{ and } \mu_i = -\varepsilon_i - \theta\sigma_v^2.$$

In words, $h(r,\varepsilon_i)$ is the expected value of z^r where z has a truncated at zero normal distribution with underlying mean μ_i and variance σ_v^2. We propose to estimate $h(r,\varepsilon_i)$ by using the mean of a sample of draws from this distribution. For given values of (i.e., conditioned on) ε_i and μ_i (i.e., y_i, $\mathbf{x}_i$, $\boldsymbol{\beta}$, σ_v, θ, r), by the Lindberg-Levy variant of the central limit theorem [see Greene(2000)], $h(r,\varepsilon_i)$ would be consistently estimated by

$$\hat{h}i = \frac{1}{Q}\sum_{q=1}^{Q} z_{iq}^r$$

where z_{iq} is a random draw from the truncated normal distribution with mean parameter μ_i and variance parameter σ_v. (The truncated normal distribution has finite moments of all orders, so this is an application of the most narrow version of the central limit theorem.) We propose, then, to maximize the simulated log likelihood function

$$\log L_{NG,S} = \log L(\text{exponential}) + N[(P\text{-}1)\log\theta - \log\Gamma(P)] + \sum_{i=1}^{N} \log \hat{h}(P\text{-}1,\varepsilon_i)$$

(Properties of this method of maximum likelihood estimation are discussed elsewhere, such as in the November, 1994 symposium in the *Review of Economics and Statistics*. The techniques has been widely used in estimation of multinomial probit models (see the aforementioned symposium) and in estimation of discrete choice models with random parameters (see Train (Revelt and 1999) for example).

3.3. Constructing the Simulated Log Likelihood Function

We now consider two practical issues, generating the random draws from the truncated distribution and the prior problem of producing the primitive draws for the simulation.

Computing the simulated log likelihood function will require Q draws from the truncated normal distribution for each observation. In principle, these can be drawn by the simple rejection method. Draws are taken from an untruncated distribution and rejected until a draw from the desired region is obtained. The problems with this approach (notwithstanding its raw inelegance) are that it can require huge numbers of draws if the desired region is near the tail of the underlying distribution, and, most importantly for maximm simulated estimation likelihood, has the result that different draws will be used for different computations of the log likelihood function. As such, the simulated log likelihood will no longer continuous be a continuous function of the parameters. The iterations will never converge. An alternative method used, e.g., in Gewekem Keane, and Runkle (1994) is a superior approach: This method requires only a single draw; the procedure is carried out as follows: Let

T = the truncation point.

μ = the mean of untruncated distribution

σ = the standard deviation of untruncated distribution

$P_L = \Phi[(T - \mu) / \sigma]$

F = a draw from the standard continuous uniform distribution $U[0,1]$

and

$z = \mu + \sigma\Phi^{-1}[P_L + F\times(1 - P_L)]$

Then,

z = the draw from the desired truncated distribution.

For implementing Geweke's method to compute h(r,ε), we have the following:

$T = 0,$

$\varepsilon_i = y_i - \boldsymbol{\beta}'\mathbf{x}_i,$

$\mu_i = -\varepsilon_i - \theta\sigma_v^2,$

$\sigma = \sigma_v,$

$P_L = \Phi[(\varepsilon_i + \theta\sigma_v^2)/\sigma_v] = \Phi(-\mu_i/\sigma).$

Collecting all terms, we have the simulated log likelihood function:

$$\text{Log } L = N\{\log\theta + \tfrac{1}{2}\sigma_v^2\theta^2\} + \sum_{i=1}^{N}\{\theta\varepsilon_i + \log\Phi(\mu_i/\sigma_v)\} + N[(P\text{-}1)\log\theta - \log\Gamma(P)]$$
$$+ \sum_{i=1}^{N}\log\left\{\frac{1}{Q}\sum_{q=1}^{Q}\left[\mu_i + \sigma_v\Phi^{-1}\left(F_{iq} + (1-F_{iq})\Phi\left(\frac{-\mu_i}{\sigma_v}\right)\right)\right]^{P-1}\right\}$$

The simulated log likelihood function is a smooth continuous function of the parameters, [β', σ_v, θ, P]. Derivatives are fairly simple as well. (They are presented in the Appendix.). Conventional maximization methods such as DFP or BFGS can be used for the optimization. We have used the BHHH (outer product of gradients) estimator to compute the asymptotic covariance matrix of the simulated maximum likelihood estimator.

For maximizing the simulated log likelihood we emphasize that the simulation is over F_{iq} draws from the standard uniform distribution. Use Q points for the simulation. In order to achieve continuity, it is necessary to use the same set of draws, [F_{i1},F_{i2},...,F_{iQ}], for every computation. Every sum over q = 1,...,Q uses the same set of random draws. (Each observation has its own unchanging vector of draws.) How this is done by researchers who employ this method varies from one application to another. Ruud and McFadden (1994) and Bhat (1999) recommend maintaining a fixed, indexed reservoir of random draws. For our simulations, we have, instead, controlled the draws by associating with each individual observation in the sample a unique seed for the random number generator, and restarting the generator at that value for each observation as needed. In connection to the point noted in the next section, we note that this method is slightly more time consuming than the fixed pool approach. But, it requires no additional computer memory while the fixed pool method will be profligate with memory when analyzing a large sample and using a large number of draws for the simulation.

3.4. Efficient Computation of the Simulated Log Likelihood Function

For implementation, there remains the practitional consideration of how best to obtain the underlying random draws, F_{iq}, from the U[0,1] distribution that drive the simulation. We consider two issues, how many draws to obtain and how to create them. On the first point, the literature varies. The glib advice, "the more the better," is not helpful when time becomes a consideration and, in fact, the marginal benefit of additional draws eventually becomes nil. Again, researchers differ, but received studies seem to suggest that several hundred to over 1,000 are needed. [See Bhat (1999), for example.]
The second consideration concerns how to obtain the draws. For most practitioners, the conventional approach amounts to relying on a random number buried within, say, Gauss, under the assumption that the draws it produces are truly random by the standards of received tests of randomness. A recently emerging literature [see, e.g., Bhat (1999) or Revelt and Train (1999), based on work in numerical analysis, suggests that this view of the process neglects some potentially large gains in computational efficiency. An alternative approach based on Halton draws (derived below) promises to improve the computations of simulated likelihoods such as ours.

The Halton sequence of draws is based on an 'intelligent' set of values for the simulation. The process is motivated by the idea that true randomness is not really the objective in producing the simulation. Coverage of the range of the variation with a sequence of draws that is uncorrelated with the variables in the model is the objective - the simulation is intended, after all, to estimate an integral, that is, an expectation. Numerical analysts have found that a small number of Halton draws is as effective as or more so than a large number of pseudorandom draws using a random number generator.

Halton sequences are generated as follows: Let s be a prime number larger than 2. Expand the sequence of integers $g = 1,...$ in terms of the base s as

$$g = \sum_{i=0}^{I} b_i s^i$$

where by construction, $0 \leq b_i \leq s - 1$ and $s^I \leq g < s^{I+1}$. Then, the Halton sequence of values that corresponds to this series is

$$H(g) = \sum_{i=0}^{I} b_i s^{-i-1}$$

Halton values are contained in the unit interval. They are not random draws, but they are well spaced in the interval. A simple Box-Jenkins identification of the Halton sequence from base s shows large autocorrelation at lag ks. For example, Table 1 below shows the autocorrelations and partial autocorrelations for a Halton sequence for base 7.

Table 1. Autocorrelations and Partial Autocorrelations for the Halton 7 Sequence

Lag	Autocorrelation		Partial Autocorrelation	
1	.263*	***	.263*	***
2	-.229*	***	-.320*	****
3	-.476*	*****	-.374*	****
4	-.478*	*****	-.438*	*****
5	-.235*	***	-.486*	*****
6	.253*	***	-.305*	***
7	.983*	**********	.962*	**********
8	.249*	***	-.801*	*********
9	-.242*	***	-.107*	*
10	-.488*	*****	-.019	*
11	-.489*	*****	.001	*
12	-.246*	***	.013	*
13	.242*	***	.024	*
14	.972*	**********	.176*	**
15	.239*	***	-.333*	****
16	-.250*	***	-.104*	*
17	-.495*	*****	-.045*	*
18	-.495*	*****	-.035*	*
19	-.252*	***	-.034*	*
20	.236*	***	-.040*	*
21	.965*	**********	.063*	*
22	.234*	***	-.177*	**
23	-.253*	***	-.097*	*
24	-.497*	*****	-.062*	*
25	-.497*	*****	-.056*	*

Figures 2 and 3 below compare two sequences of 1000 pseudorandom values to the first 1000 values from Halton base 7 and Halton base 9

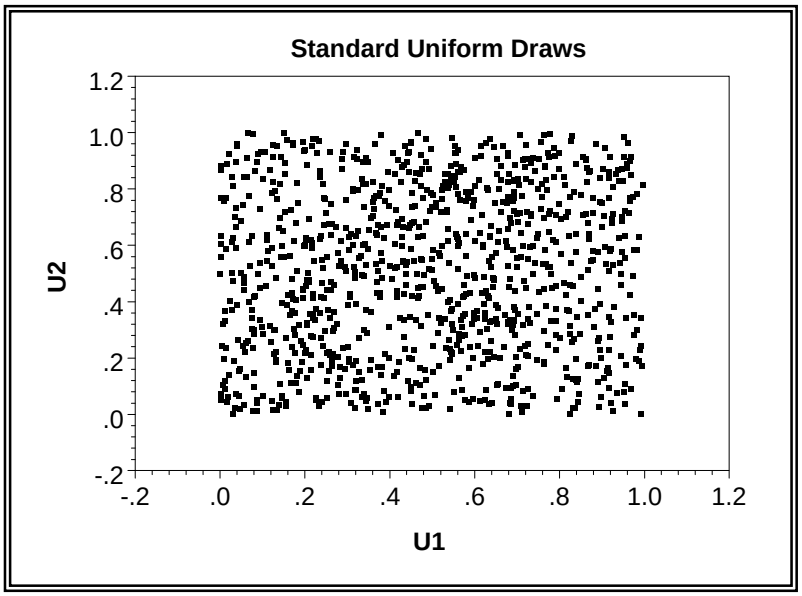


Figure 2. 1000 Pairs of Pseudorandom U[0,1] Draws

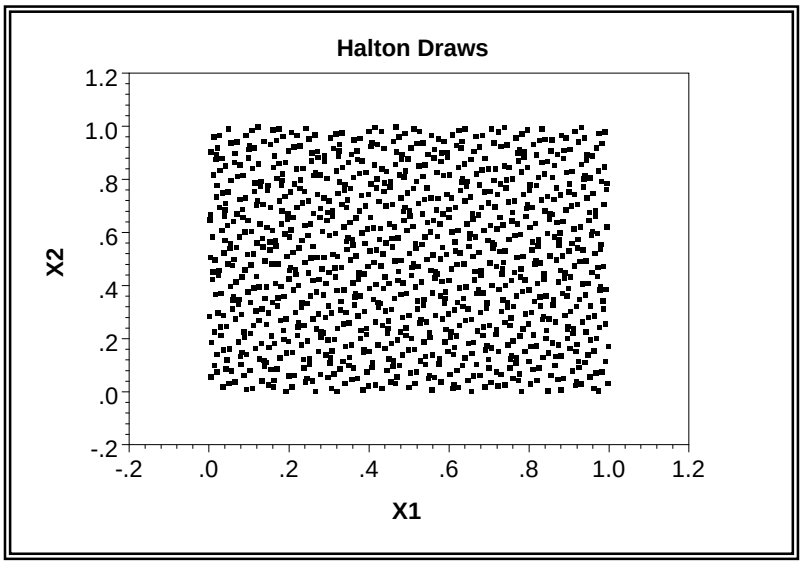


Figure 3. The First Halton 7 and Halton 9 Values

Note the clumping in Figure 2. This is what makes large numbers of draws necessary. The Halton sequences are much more efficient at covering the sample space. Halton sequences are being used with great success in estimating "mixed logit" models that require huge amounts of simulation. (See Train (1999), for example.) The computational efficiency compared to pseudo random draws appears to be at least 10 to 1. The same results are obtained with only 1/10 of the number of draws. Computation time in these models is roughly linear in the number of replications, so time savings are potentially very large.

4. An Application

To illustrate the technique, we have applied the preceding to Christensen and Greene's (1976) electricity data. We used the 1970 sample that contains 158 firms and holding companies. The regression function is a simple Cobb-Douglas cost function with a quadradic term in log output;

$$\log(C/Pf) = \beta_1 + \beta_2 \log(Pl/Pf) + \beta_3 \log(Pk/Pf) + \beta_4 \log Y + \beta_5 \log^2 Y + v + u$$

where C is total cost of generation, Y is output, and Pf, Pk and Pl are the unit prices of fuel, capital and labor, respectively. (Translation of the original model to a cost function requires only a trivial change of sign of some terms in the log likelihood and its derivatives.) Table 2 presents the estimation results. The random draws approach is based on $Q = 500$. The Halton results are based on $Q = 50$.

Table 2. Estimated Stochastic Frontier Functions

	Exponential		**Gamma- U[0,1]**		**Gamma-Halton**	
Parameter	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.
β_1	-7.0345	0.207	-7.0338	0.2094	-7.0337	0.1308
β_2	0.1449	0.0421	0.1450	0.0421	0.1449	0.0419
β_3	0.1391	0.0390	0.1390	0.0389	0.1384	0.0387
β_4	0.4413	0.0302	0.4416	0.0304	0.4431	0.0310
β_5	0.0286	0.00208	0.0286	0.00210	0.0285	0.00213
θ	11.012	2.697	10.832	3.0040	10.164	3.701
σ_v	0.1030	0.0127	0.1033	0.0131	0.1038	0.0133
P	1.0000	0.0000	0.9620	0.3517	0.8422	0.5265
Log Likelihood	95.05542		93.06719		93.11514	
Estimated Standard Deviations of the Underlying Variables						
v	0.10296		0.10325		0.10383	
u	0.09083		0.09055		0.9028	

The estimated indefficiencies from the three sets of estimates are very similar, as the last two rows of Table 2 would suggest. Also, the estimate of P in the gamma models is not particularly large and, moreover, is less than one which if anything exaggerates the effect of packing the observations close to the origin as the exponential model does. Table 3 lists the descriptive statistics and correlations of the Jondrow et al. estimator of $E[u|\varepsilon]$ for the three models. The JLMS efficiency measure has the simple form

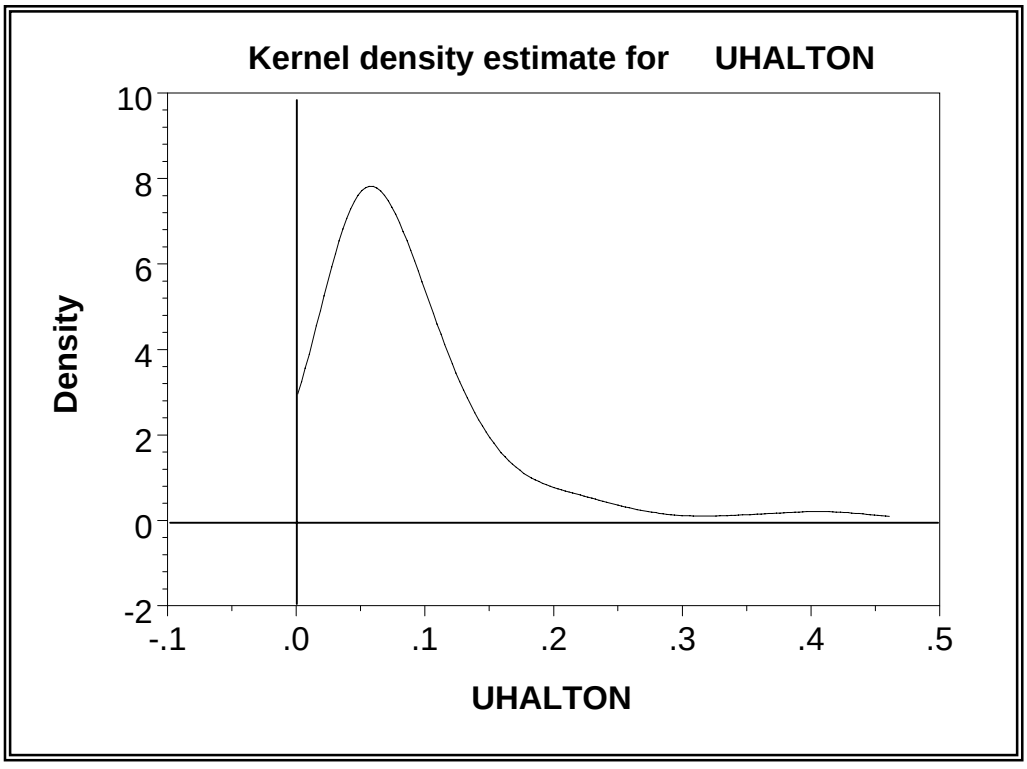
$$E[u|\varepsilon] = h(P,\varepsilon_i) / h(P-1,\varepsilon_i).$$

for the normal-gamma model.

Table 3. Descriptive Statistics for JLMS Estimated Inefficiencies

	Mean	Std. Deviation	Minimum	Maximum
Exponential	0.090813	0.067581	0.022991	0.443508
Random *U*[0,1]	0.083121	0.067682	0.020464	0.438584
Halton	0.088334	0.068167	0.019775	0.430003
Correlations				
	Exponential	**RandomU[0,1]**	**Halton**	
Exponential	1.00000			
Random U[0,1]	.99801	1.00000		
Halton	.98431	.98283	1.00000	

Finally, in order to suggest what the overall results look like, Figure 4 below presents a kernel density estimator of the underlying density of the inefficiencies. While it is suggestive, unfortunately, the figure illustrates one of the shortcomings of the JLMS computation. As we can see from the results above, the estimated distribution of *u* for these data resemples the exponential, with mode at zero. But, the estimates computed using the residuals, have *v* mixed in them As the kernel density estimator suggests, this suggests a helpful, but obviously distorted picture. This may call the JLMS estimator into question, but that is beyond the scope of this paper.



5. Conclusions

The preceding has proposed a method of estimating the normal-gamma frontier model. The operational aspects of the proposed method are fairly straightforward, and implementation should be relatively simple. The estimator used here was built into *LIMDEP* (2000, forthcoming release) but could easily be programmed in *Gauss* or in a low level language if the analyst prefers. Experience with the estimator is limited, but the results do suggest a useful extension of the stochastic frontier model. If it is established that the estimator actually does work well in practice, then familiar extensions might be added to it. For example - and this could be added to the exponential model though we have not seen it - the location parameter can be parameterized to include heterogeneity, in the form $\theta_i = \exp(\boldsymbol{\delta}'\mathbf{z}_i)$ for example.

Whether Simar's observations about the (non)identifiability of the normal-gamma model prove general is an empirical question. His result was a matter of degree, not a definitive result. That is, he found that identification of the model would be 'difficult,' not impossible. As such, as might be expected, further research is called for.

References

Aigner, D., K. Lovell and P. Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, 1977, 6, pp. 21-37.

Battese, G., and G. Corra, "Estimation of a Production Frontier Model, with Application to the Pastoral ZOne off Eastern Australia," *Australian Journal of Agricultural Economics*, 1977, 21, pp. 169-179.

Beckers, D. and C. Hammond, "A Tractable Likelihood Function for the Normal-Gamma Stochastic Frontier Model," *Economics Letters*, 1987, 24, pp. 33-38.

Bhat, C., "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," Working paper, Department of Civil Engineering, University of Texas, Austin, 1999.

Christensen, L. and W. Greene, "Economies of Scale in U.S. Electric Power Generation," *Journal of Political Economy*, 1976, 84, pp. 655-676.

Coelli, T., D. Rao, and G. Battese, *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers, Boston, 1997.

Geweke, J., M. Keand, and D., Runkle, "Statistical Inference in the Multinomial, Multiperiod Probit Model," *Review of Economics and Statistics*, 1994, 76, pp. 609-632.

Greene, W., "A Gamma Distributed Stochastic Frontier Model," *Journal of Econometrics*, 1990, 46, pp. 141-164.

Greene, W., *Econometric Analysis*, 4th Ed. , Prentice Hall, Upper Saddle River, NJ, 2000.

Horrace, W. and P. Schmidt, "Confidence Statements for Efficiency Estimates form Stochastic Frontier Models," *Journal of Productivity Analysis*, 1996, 7, pp. 257-282.

Jondrow, J., I. Materov, K. Lovell and P. Schmidt, "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model," *Journal of Econometrics*, 1982, 19, pp. 233-238.

Kumbhakar, S., and K. Lovell, *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge, 2000

McFadden, D., and P. Ruud, "Estimation by Simulation," *Review of Economics and Statistics*, 1994, 76, pp. 591-608.

Meeusen, W. and J. van den Broeck, "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review*, 1977, 18, pp. 435-444.

Revelt, D., and K. Train, "Customer Specific Taste Parameters and Mixed Logit Models," Working Paper, Department of Economics, University of California, Berkeley, 1999.

Ritter, C. and L. Simar, "Pitfalls of Nornmal-Gamma Stochastoc Frontier Models," *Journal of Productivity Analysis*, 1997, 8, pp. 167-182.

Stevenson , R., "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics*, 1980, 13, pp. 57-66.

Tsionas, E., "Full Likelihood Inference in Normal-Gamma Stochastic Frontier Models," *Journal of Productivity Analysis*, 2000, 13, pp. 183-296.

Appendix: Derivatives of the Simulated Log Likelihood

The *i*th term in the log likelihood for the normal-gamma model is

$$\log L_i = \log\theta + \tfrac{1}{2}\sigma_v^2\theta^2 + \theta\varepsilon_i + \log\Phi(\mu_i/\sigma_v) + (P\text{-}1)\log\theta - \log\Gamma(P)$$

$$+ \log\left\{\frac{1}{Q}\sum_{q=1}^{Q}\left[\mu_i + \sigma_v\Phi^{-1}\left(F_{iq} + (1-F_{iq})\Phi\left(\frac{-\mu_i}{\sigma_v}\right)\right)\right]^{P-1}\right\}.$$

It is convenient to gather the terms and collect a few to rewrite this as

$$\log L_i = P\log\theta - \log\Gamma(P) - \theta\sigma W_i - \theta^2\sigma^2/2 + \log\Phi(W_i)$$

$$+ \log\left\{\frac{1}{Q}\sum_{q=1}^{Q}\left[\sigma W_i + \sigma\Phi^{-1}\left(F_{iq} + (1-F_{iq})\Phi(-W_i)\right)\right]^{P-1}\right\}$$

where $W_i = \mu_i/\sigma$ and we have dropped the subscript on σ_v to simplify the notation. We have now written the function in terms of P, θ, σ and W_i which is a function of θ and σ as well as $\boldsymbol{\beta}$. It will also be convenient to define a symbol for the bracketed term in the summation, so we further compress the function to

$$\log L_i = P\log\theta - \log\Gamma(P) - \theta\sigma_v W_i - \theta^2\sigma_v^2/2 + \log\Phi(W_i) + \log\left\{\frac{1}{Q}\sum_{q=1}^{Q}[C_{iq}]^{P-1}\right\}$$

To avoid some notational clutter, let

$$H_i = \left\{\frac{1}{Q}\sum_{q=1}^{Q}[C_{iq}]^{P-1}\right\},\ \lambda_i = \frac{\phi(W_i)}{\Phi(W_i)},\ \text{and}\ D_{iq} = (P-1)[C_{iq}]^{P-2}$$

Recall that $W_i = \mu_i/\sigma = -\varepsilon_i/\sigma - \theta\sigma$. With these preliminaries, then,

$$\frac{\partial \log L_i}{\partial\theta} = \frac{P}{\theta} - \sigma W_i - \theta\sigma\frac{\partial W_i}{\partial\theta} - \theta\sigma^2 + \lambda_i\frac{\partial W_i}{\partial\theta} + \frac{1}{H_i}\frac{1}{Q}\sum_{q=1}^{Q} D_{iq}\frac{\partial C_{iq}}{\partial\theta},\quad \frac{\partial W_i}{\partial\theta} = -\sigma,$$

$$\frac{\partial \log L_i}{\partial\sigma} = \theta W_i - \theta\sigma\frac{\partial W_i}{\partial\sigma} - \theta^2\sigma + \lambda_i\frac{\partial W_i}{\partial\sigma} + \frac{1}{H_i}\frac{1}{Q}\sum_{q=1}^{Q} D_{iq}\frac{\partial C_{iq}}{\partial\sigma},\quad \frac{\partial W_i}{\partial\sigma} = \varepsilon_i/\sigma^2 - \theta,$$

$$\frac{\partial \log L_i}{\partial\varepsilon_i} = \theta\sigma\frac{\partial W_i}{\partial\varepsilon_i} + \lambda_i\frac{\partial W_i}{\partial\varepsilon_i} + \frac{1}{H_i}\frac{1}{Q}\sum_{q=1}^{Q} D_{iq}\frac{\partial C_{iq}}{\partial\varepsilon_i},\quad \frac{\partial W_i}{\partial\varepsilon_i} = -1/\sigma,\quad \frac{\partial\varepsilon_i}{\partial\boldsymbol{\beta}} = -\mathbf{x}_i$$

$$\frac{\partial \log L_i}{\partial P} = \log\theta - \Psi(P) + \frac{1}{H_i}\frac{1}{Q}\sum_{q=1}^{Q}\left(\log C_{iq}\right)[C_{iq}]^{P-1}$$

To complete the derivation, we require the derivatives of

$$C_{iq} = \sigma W_i + \sigma\Phi^{-1}\left(F_{iq} + (1 - F_{iq})\Phi(-W_i)\right)$$

Let $a_{iq} = F_{iq} + (1\text{-}F_{iq})\Phi(\text{-}W_i)$ so $C_{iq} = \sigma[W_i + \Phi^{-1}(a_{iq})]$. Then,

$$\frac{\partial C_{iq}}{\partial W_i} = \sigma + \sigma\frac{F_{iq}\phi(-W_i)}{\phi(a_{iq})} = \sigma[1 + E_{iq}].$$

Then, inserting derivatives of W_i where needed, we have

$$\frac{\partial C_{iq}}{\partial \theta} = -\sigma[1 + E_{iq}](-\sigma)$$

$$\frac{\partial C_{iq}}{\partial \sigma} = [W_i + \Phi^{-1}(a_{iq})] + \sigma[1 + E_{iq}](\varepsilon_i / \sigma^2 - \theta)$$

$$\frac{\partial C_{iq}}{\partial \varepsilon_i} = \sigma[1 + E_{iq}](-1/\sigma)$$