Efficiency Estimation using the Simulated Maximum Likelihood Approach: the Case of Polish Cooperative Banks

Tamar Khitarishvili
Department of Applied Economics, University of Minnesota

presented at the AAEA meetings in Long Beach, CA, July 2002

Abstract

The paper lays out a framework for studying firm efficiency using the Method of Simulated Maximum Likelihood. I develop the model using the assumption of gamma distributed inefficiency terms and apply it to the case of Polish cooperative banks. I derive preliminary results for the estimates of average and individual inefficiencies. Finally, I discuss the practical difficulties connected with this estimation method and suggest avenues for future research.

1 Introduction

Firm efficiency and productivity studies play a prominent role in current empirical literature. Results of these studies are used in policy recommendations and as such have important consequences. These results depend on the approach that was used and on how restrictive the assumptions of that approach are. Efficiency studies that use the Stochastic Frontier Approach use restrictive distributional assumptions. Until now, the attempts to use more flexible distributional assumptions have been hindered by technical difficulties. In this paper, I use the Simulated Maximum Likelihood approach first proposed for multinomial discrete choice models (Geweke, Keane and Runkle, 1994) and later applied to the stochastic frontier estimation by Greene (2000). This approach allows us to circumvent some of the difficulties related to integration, which has been the main reason for using simpler distributions.

I apply this approach to the dataset of 300 Polish cooperative banks. Cooperative banks play a critical part in financing agricultural production in Poland by providing about 50% of the loans to agricultural production, agri-food processing, agricultural product trade and other services to the rural community[1].

[1] materials of ACDI/VOCA conference "Transformation Process of Cooperative Banking Sector" held in Warsaw, February 1998 (p.23).

Given that they maintain a competitive position in the rural financial markets, their role in supporting agriculture will increase as the demand for credit increases. This is likely to happen due to the rural sector's high borrowing capacity. In 1996 about 82% of the potential borrowers did not borrow from the financial institutions at all, and about 60% of the small farmers had never used the banking services. About 97-99% of the total financing for the private farmers was equity financing. Recent developments in the Polish cooperative banking, as well as some preliminary tests, suggest that there are inefficiencies present in the Polish cooperative banking sector.

Recent developments in the structure of the Polish cooperative banking suggest the presence of inefficiencies. For example, the number of cooperative banks in the sector has been gradually declining. From 1510 cooperative banks in 1995 it dropped to 781 banks in 1999. While initially the major reason for such a decline was due to insolvency, starting with 1997 there has been a strong tendency towards mergers. In addition, there have been significant changes in the loan structure of the Polish cooperative banks. For example, the share of the net non-financial sector loans tototal assets had increased from 32.5% in 1993 to 49.9% in 1999. This may indicate that the cooperative banks are finding it more profitable (or less costly) to invest into non-financial ventures. These developments suggest that the occurence of a major restructuring of the Polish cooperative banking. In addition, preliminary normality tests also suggest the presence of inefficiencies.

The objectives of the paper are to lay out a framework for studying firm efficiency using the Method of Simulated Maximum Likelihood and to apply this approach to the case of Polish cooperative banks. I develop the model using the assumption of gamma distributed inefficiency terms and derive preliminary results for the estimates of average and individual inefficiencies. Finally, I discuss the practical difficulties connected with this estimation method and suggest avenues for future research.

2 Stochastic Frontier Approach

The variability in the performance of firms can be explained by factors that we believe determine their performance and by factors which we are not aware of or which we do not observe. For example, if we use profits as a measure of performance, the factors that do determine the performance of the firms are obtained from the definition of the indirect profit function, which is a function of input and output prices that the firms face. The variability due to other factors are attributed to inefficiencies. We may use different estimation methods to evaluate the causes of inefficiencies. Two of the common reasons are scale and scope economies. Scale economies occur when the fixed costs of running the operations determine the optimal size of operations.Scope economies occur when a firm could reduce per unit costs or increase per unit revenues by changing the input or output composition. In addition, exogenous factors such as the firm location and unobservable factors, such as managerial skills, may the contributors to inefficiencies.

Based on such an understanding of what constitutes an inefficiency, it becomes more clear that looking at the structure of the error terms (the unexplained factors) is a way of studying the inefficiency of a firm. The approach takes precisely this view of inefficiencies.

The stochastic frontier approach (SFA) separates the effect of independent variables X on the firms' measure of performance from the effect of unobserved factors, "hidden" in the error term. In the SFA, the error term is assumed to be composed of two components: a random error term v distributed $N\left(0, \sigma_v^2\right)$ and a one sided inefficiency term u.

More formally, the stochastic frontier that we analyze is

$$y = f(X, \beta) + \underbrace{v - u}_{\varepsilon}$$

where y is an $n \times 1$ vector of the dependent variable, X is an $n \times K$ matrix of independent variables, β is a $K \times 1$ vector of parameters.

The elements of the composite error term ε are the random error $v \tilde{} N[0, \sigma_v^2]$, and the inefficiency term $u \geq 0$. The inefficiency term u is nonnegative because the profit function is "the maximum profits the firm can make as a function of the vector of prices of the net outputs" (Varian p.40). Thus, on a consistent basis (i.e., ignoring random error and shock), the firm can only do at least worse than that. Also, it is assumed that $E[\varepsilon|X] = 0$, i.e. that the composite error term is uncorrelated with the matrix of regressors. This assumption is necessary for obtaining consistent coefficient estimates.

2.1 Distribution of the Inefficiency Term

The distribution of the inefficiency component has been an issue of concern in the literature. Ideally, it is desirable to assume a relatively flexible distribution, which can take different shapes depending on the sample. The attempts to use flexible distributions have been hampered by technical difficulties. As a result, the two most common distributional assumptions made in the literature are a truncated normal distribution and an exponential distribution, which are analytically straightforward to solve. However, both of them imply that observations in the sample are all concentrated near zero inefficiency, which may not always be a satisfactory restriction.

In this paper, I use the two-parameter gamma distribution assumption. The use of simulation methods developed in multinomial discrete choice literature and extended to the efficiency literature by Greene (2000) allows us to evaluate a continuous likelihood function of a composite error, which has been difficult or impossible using other methods.

The probability density function of the gamma distribution can be defined as $u\tilde{\ } G[\Theta, P]$:

$$f(u) = \frac{\Theta^P}{\Gamma(P)} u^{P-1} e^{-\Theta u},$$

where θ is the scale parameter and P is the skewness parameter.

2.2 Average and Individual Inefficiencies

In this study we are interested in obtaining an overall picture of average firm inefficiency, as well as their individual inefficiencies.

The average firm inefficiency can be measured by estimating the unconditional mean of $u, E[u]$. The mean and variance of a gamma distributed variable u are given by:

$$\begin{aligned} E[u] &= \frac{P}{\Theta} \\ var[u] &= \frac{P}{\Theta^2}. \end{aligned}$$

Note that the exponential distribution is a special case of the gamma distribution with $P = 1$, which allows us to evaluate whether the assumption of the exponential distribution is restrictive.

To obtain individual inefficiencies, $E[u_i|\varepsilon_i]$,we use the approach followed in Jondrow et al. (1992). The idea is to derive an explicit expression for the expectation of a random variable, conditional on the value of the residual for each individual observation.

The expectation of u_i conditional on the value of ε_i can be expressed as:

$$E[u_i|\varepsilon_i] = \int_0^\infty u_i f(u_i|\varepsilon_i) du_i.$$

The conditional density of u_i, $f(u_i|\varepsilon_i)$, in turn, can be expressed as (Lindgren, p.101):

$$f(u_i|\varepsilon_i) = \frac{f(u_i, \varepsilon_i)}{f(\varepsilon_i)}.$$

We use the expressions (2) and (4) derived in the appendix to substitute for $f(\varepsilon_i)$ and $f(u_i, \varepsilon_i)$, respectively, to obtain:

$$
\begin{aligned}
E[u_i|\varepsilon_i] &= \int_0^{\infty} u_i f(u_i|\varepsilon_i) du_i \\
&= \frac{\int_0^{\infty} e^{-\frac{(\varepsilon_i+u_i)^2}{2\sigma^2_{(\varepsilon+u)}}} u_i^P e^{-\Theta u_i} du_i}{\int_0^{+\infty} e^{-\frac{(\varepsilon_i+u_i)^2}{2\sigma_v^2}} u_i^{P-1} e^{-\Theta u_i} du_i} \\
&= \frac{h(P, \varepsilon_i)}{h(P-1, \varepsilon_i)},
\end{aligned}
$$

where

$$
h(P, \varepsilon_i) = \frac{\int_0^{\infty} e^{-\frac{(\varepsilon_i+u_i)^2}{2\sigma^2_{(\varepsilon+u)}}} u_i^P e^{-\Theta u_i} du_i}{\int_0^{\infty} e^{-\frac{(\varepsilon_i+u_i)^2}{2\sigma^2_{(\varepsilon+u)}}} e^{-\Theta u_i} du_i}.
$$

3 Maximum Likelihood Estimation

The method used to obtain the estimates of P, Θ, β,and σ_v^2 and, in turn, the estimates of average and individual inefficiencies, is the Maximum Likelihood method. From here on, we drop the i subscripts on u for notational simplicity. Note that i applies only to ε and u. All other variables and parameters are the same for all observations.

For the normal gamma case, the likelihood function is:

$$
\begin{aligned}
L(\sigma, \Theta, P) &= \Pi_{i=1}^n f(\varepsilon_i) \\
&= \Pi_{i=1}^n \frac{1}{\sigma_v \sqrt{2\pi}} \frac{\Theta^P}{\Gamma(P)} \int_0^{\infty} e^{-\frac{(\varepsilon_i+u)^2}{2\sigma_v^2}} u^{P-1} e^{-\Theta u} du,
\end{aligned}
$$

since u can only be greater than or equal to zero.

Then the loglikelihood function is:

$$
\begin{aligned}
\ln L(\sigma, \Theta, P) &= l(\sigma, \Theta, P) = n \ln \left[\frac{1}{\sigma_v \sqrt{2\pi}} \frac{\Theta^P}{\Gamma(P)} \right] + \\
&\quad + \sum_{i=1}^{n} \ln \int_0^{\infty} e^{-\frac{(\varepsilon_i+u)^2}{2\sigma_v^2}} u^{P-1} e^{-\Theta u} du.
\end{aligned}
$$

After rearrangement (see Appendix), we obtain:

$$\ln(NG) = n\ln\Theta^P - n\ln\Gamma(P) + \sum_{i=1}^{n}\left(\Theta\varepsilon_i + \frac{\sigma_v^2\Theta^2}{2}\right) + \sum_{i=1}^{n}\ln h(P-1,\varepsilon_i) + \sum_{i=1}^{n}\ln\Phi\left(\frac{\mu_i}{\sigma_v}\right),$$

where $\mu_i = -(\varepsilon_i + \theta\sigma_v)$.

The term $h(P-1,\varepsilon_i)$ has been the cause of problems. Some papers have attempted to obtain a closed form solution (Simar, 2000), others - to find numerical approximations (Greene, 1990). The method used in this paper was first proposed for multinomial discrete choice models (Geweke. Keane and Runkle, 1994) and applied to the stochastic frontier estimation by Greene (2000).

3.1 Simulated Maximum Likelihood Estimation

The proposed SML method is based on noting that $h(P-1,\varepsilon_i)$ is the expectation of a truncated at zero normal variable u_i with mean $\mu_i = -\varepsilon_i - \sigma^2\Theta$ and variance σ_v^2, $E[u^{P-1}|u > 0, \varepsilon_i]$ (see Appendix for derivations). This point is critical because we have been considering a gamma-normal case, and we have obtained a variable, which is a truncated at zero normal variable.

We can estimate $h(P-1,\varepsilon_i)$ by taking the mean of a sample of draws from this distribution. For given values of ε_i and μ_i, by the Linberg-Levy variant of the Central Limit Theorem, $h(P-1,\varepsilon_i)$ would be consistently estimated by

$$h(P-1,\varepsilon_i) = \frac{1}{Q}\sum_{q=1}^{Q} u_{iq}^{P-1},$$

where u_{iq} is a random draw from the truncated normal distribution with mean μ_u and variance σ_u^2. In this case, $\mu_u = \mu_i$ and $\sigma_u^2 = \sigma_v^2$.

Applying this information to the gamma-normal loglikelihood function, we get

$$\ln(NG) = n\ln\Theta^P - n\ln\Gamma(P) + \sum_{i=1}^{n}\left(\Theta\varepsilon_i + \frac{\sigma_v^2\Theta^2}{2}\right) + \sum_{i=1}^{n}\ln\left(\frac{1}{Q}\sum_{q=1}^{Q} u_{iq}^{P-1}\right) + \sum_{i=1}^{n}\ln\Pr(u > 0|\varepsilon_i).$$

3.1.1 The draws

The next step is to clarify how to actually perform the draws. (for references here, see Greene textbook, p.169 and Geweke, Runkle and Keane 1994, p. 614, Stern, JEL 1997, p.2021) First of all, we use the fact that if we would like to obtain a standard normal random variable z such that $a < z < b$, we need to form

$$\begin{aligned}\xi &= \Phi^{-1}\left[\left[\Phi(b)-\Phi(a)\right]U+\Phi(a)\right] \\ &= \Phi^{-1}\left[\Phi(b)U-\Phi(a)U+\Phi(a)\right] \\ &= \Phi^{-1}\left[\Phi(b)U+\Phi(a)\left(1-U\right)\right],\end{aligned}$$

where the last expression becomes identical to Greene's (2000) when $b=\infty$ and $a=\frac{0-\mu_u}{\sigma_u}$, which holds in this case since we are concerned with a truncated at zero normal variable and Φ is a standard normal cdf.

Thus, we have

$$\xi=\Phi^{-1}\left[U+\Phi(\frac{-\mu_u}{\sigma_u})\left(1-U\right)\right].$$

Furthermore, since u is not a standard normal, but a truncated normal with mean $\mu_i=-\varepsilon_i-\sigma_v^2\Theta$ and variance σ_v^2 we have that (see Lindgren, p.179):

$$u_i=\mu_i+\sigma_v\xi_i,$$

where ξ_i is the standard normal variable we are drawing.

Taking all the above issues into consideration, $h(P-1,\varepsilon_i)$ can be written as:

$$\begin{aligned}h(P-1,\varepsilon_i) &= \frac{1}{Q}\sum_{q=1}^{Q}u_{iq}^{P-1} \\ &= \frac{1}{Q}\sum_{q=1}^{Q}\left(\mu_i+\sigma_v\xi_{iq}\right)^{P-1} \\ &= \frac{1}{Q}\sum_{q=1}^{Q}\left(\mu_i+\sigma_v\Phi^{-1}\left[U+\Phi(\frac{-\mu_i}{\sigma_v})\left(1-U\right)\right]\right)^{P-1}.\end{aligned}$$

3.1.2 How many draws to obtain

From several hundred to several thousand (Bhat, 1999, "Quasi-Random MSL Estimation of the Mixed Multinomial Logit Model," WP, Department of Civil Engineering, U of Texas, Austin).

3.1.3 Which method to use in obtaining them

We use the Gauss random number generator. Greene (2000) uses the Halton sequence of draws. This is an interesting question.

Once the method is determined, we produce Q draws for each observation and thus obtain Q values of ξ_{iq} and in turn $\frac{1}{Q}\sum_{q=1}^{Q} u_{iq}^{P-1}$ for each observation.

4 Dataset and the Functional Form

The study uses the 1997 National Bank of Poland Call Report data. The data were sorted according to the size (total assets). Out of 1295 cooperative banks that existed in 1997, three hundred banks were chosen as a sample.

The important assumption of the paper is that the banks are profit maximizers. Thus the function that we are estimating is the profit function. Moreover, it is assumed that the banks are price setters on the output side. Therefore they maximize their profit function with respect to output prices (rather than output quantities) and input quantities. The functional form used is the Fuss normalized quadratic function, which can be expressed as:

$$\begin{aligned}\frac{\pi(p,y,z)}{p_n} &= \sum_{i=1}^{n}\alpha_i\frac{p_i}{p_n}+\sum_{r=1}^{k}\beta_r y_r+\gamma z+\frac{1}{2}\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\delta_{ij}\frac{p_ip_j}{p_n^2}+ \\ &+\frac{1}{2}\sum_{r=1}^{k}\sum_{s=1}^{k}\phi_{rs}y_ry_s+\mu z^2+\frac{1}{2}\sum_{i=1}^{n-1}\sum_{r=1}^{k}\eta_{ir}\frac{p_i}{p_n}z+\sum_{r=1}^{k}\lambda_r y_r z.\end{aligned}$$

In the above equation, p_i is the price of input i, for $i=1,...,n$, y_r is quantity of output r, for $r=1,...,k$, and z is the fixed netput. We impose linear homogeneity in prices by normalizing the input prices by p_n. In our model, we use the price of deposits for normalization. Symmetry is imposed on the model by assuming that $\gamma_{ij}=\gamma_{ji}$, $\phi_{ij}=\phi_{ji},\eta_{ij}=\eta_{ji}$. Note also that $n=2$ and $k=3$.

I employ the intermediation approach for measuring inputs and outputs (Sealey and Lindley, 1977). The normalized quadratic profit function uses three outputs, two input prices, and one fixed netput. The outputs are financial loans, agricultural loans, nonagricultural nonfinancial loans. The input prices are the price of deposits, defined as expenses on deposits divided by the total amount of deposits, and the wages, defined as labor expenses divided by the number of employees. There is one fixed netput, capital, defined as physical assets. The dependent variable is net profits minus provisions for loan losses (NRP) . All variables are normalized. The price of deposits is used as a numeraire.

5 Results

The normality test is performed to identify indications of inefficiencies. The test suggests that the error terms are not normally distributed (the F -statistic is 203.39701).

We use the results of the OLS and the corrected OLS (Greene, 1990, Khitarishvili, 2000) as starting values for the ML estimation. The results of the OLS regression analysis suggest that capital is negatively related to average net profits. In other words, banks that occupy larger buildings are the ones with lower profits per unit of total assets. One of the reasons could be that the banks' operations are too small given the level of fixed assets that they own. If that is so, there must be some indications of the scale economies among the Polish cooperative banks. This indeed appears to be the case from the recent changes in the number of cooperative banks, as well as formal results of the previous study (Khitarishvili, 2000).

Another important relationship is the joint capital-nonagricultural loans term, which has a positive sign. It indicates that the banks with larger fixed assets and the larger nonfinancial nonagricultural loans are also the ones with higher average net profits. This result may indicate some preference by larger banks to provide more nonagricultural nonfinancial loans. The regression results indicate that the banks with greater proportion of nonfinancial nonagricultural loans are also more profitable. This may suggest that the cooperative banks could increase their profitability by increasing the amount of loans given for nonfinancial nonagricultural purposes. This conjecture could be supported by comparing the returns on the three types of loans , which suggests that nonfinancial nonagricultural loans are characterized by the highest returns. Their "price" (interest income divided by the total volume) is 1.81798 zloty per unit of loan, as compared to 0.7547 for agricultural loans and 0.09849 for financial loans. It also is the case however that the standard deviation of nonfinancial nonagricultural loans is the greatest as well (10.25 as compared to 0.93 and 0.05 for agricultural and financial loans, respectively).

The OLS starting values and the COLS estimates of P and θ are as follows[2]:

[2] ** statistically significantly different from zero at 5%
***statistically significantly different from zero at 10%

variable	estimate
$const$	352.67
CAP	-2.509362^{**}
W	-1.438720
Y_f	-0.007200
Y_{ag}	0.343796
Y_{nag}	0.284014^{***}
W^2	-0.003273
WY_f	0.002486^{***}
WY_{ag}	0.001861
WY_{nag}	0.002164^{***}
Y_f^2	-0.000014
$Y_f Y_{ag}$	-0.000023
$Y_f Y_{nag}$	0.000009
Y_{ag}^2	0.000001
$Y_{ag} Y_{nag}$	-0.000048
Y_{nag}^2	-0.000029
CAP^2	0.009384
$CAPW$	0.000149
$CAPY_f$	0.000098
$CAPY_{ag}$	0.000366^{**}
$CAPY_{nag}$	-0.000420
P	0.0191
θ	2.98
σ_v^2	0.0171

As the estimates of P, θ and σ_v^2 are point estimates (method of moments estimates), we cannot conclude anything about how good estimates they are.

The ML estimated parameter values that correspond to the COLS corrected values are:

However, it also is the case that the ML estimates are sensitive to the starting values. This implies that while the gamma loglikelihood function is a globally concave function, the simulation component of it is introducing some nonconcavities (the function is still continuous though). Some authors do suggest however that "unless the sample size reaches several thousand observations the shape parameter of the gamma density is hard to estimate" (Ritter and Simar, 1997, p.167).

The second important result in the paper is that given the initial values suggested by Greene (2000), which are the COLS values (obtained in Khitarishvili (2000)), the convergent value of P is far from 1. I have not tested this result though. However, this would provide important implications for the future research because it would suggest that indeed the value of P may be significantly different from 1, which, in turn, suggests that most banks are concentrated near being efficient.

Finally, the distribution of the individual inefficiencies, which includes the random error component and the gamma distribution given a particular parameter choice are compared. What do the results suggest? The comparison once again suggests that more research is needed.

6 Conclusions and Further Research

The important conclusion is that the simulated maximum likelihood estimation presents a new avenue to approach problems that until now have been too complex to estimate. Being a new approach in the efficiency literature, this also implies that its properties are not yet fully understood and that it has not been applied in many applications yet. But the results of this study indicate that may be a value to using this approach. More specifically, the results show that the value of the P parameter is not unambigously greater than 1. More research is needed to determine the causes of some of the problems outlined above.

In addition to understanding better the properties of the simulated estimator, a natural extension of this research is to understand the relationship between the individual inefficiencies and their possible causes. Some of the work to identify these causes, more specifically the evaluatione of scale and scope economies has been done in Khitarishvili, 2000, using the OLS setup.

Another extension for future research is the identification of how measurement errors affect the values of estimated inefficiencies, and importantly how they affect the properties of the SML estimators.

7 Appendix

7.1 Distribution of the composite error term, ε

In this section[3], we evaluate the distribution of a composite error term $\varepsilon = v - u$.

Theorem 1 *If X and Y are independent, continuous variables, the probability density function of their sum is*

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) dy. \tag{1}$$

[3] see Lindgren, p.123 and 59

This result depends on the assumed independence between X and Y, and can be obtained by differentiating the cumulative distribution function, which is:

$$\begin{aligned} F_{X+Y}(z) &= \Pr(X+Y \leq z) = \int_{-\infty}^{+\infty}\int_{-\infty}^{z-y} f(x,y)dxdy \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{z-y} f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{z-y} f_X(x)dx\right] f_Y(y)dy \\ &= \int_{-\infty}^{+\infty} F_X(z-y)f_Y(y)dy. \end{aligned}$$

Once we differentiate this expression we obtain (1).

Expressed in our notation, we have

$$f_{v-u}(\varepsilon) = \int_{-\infty}^{+\infty} f_v(\varepsilon+u)f_{-u}(-u)d(-u).$$

Use once again the transformation formula on page 64 of Lindgren. Let $w = a + bu = -u$, where $a = 0$ and $b = -1$. Then

$$f_W(w) = f_U\left(\frac{w-0}{-1}\right)\frac{1}{-1} = -f_U(u).$$

and

$$\begin{aligned} f_{v-u}(\varepsilon) &= \int_{-\infty}^{+\infty} f_v(\varepsilon+u)f_{-u}(-u)d(-u) \\ &= \int_{-\infty}^{+\infty} f_v(\varepsilon+u)f_u(u)du. \end{aligned}$$

For the normal-gamma case, we have

$$\begin{aligned} f_{v-u}(\varepsilon) &= \int_{-\infty}^{+\infty} f_v(\varepsilon+u)f_u(u)du \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-\frac{(\varepsilon+u-\mu_{(\varepsilon+u)})^2}{2\sigma^2_{\varepsilon+u}}} \frac{\Theta^P}{\Gamma(P)} u^{P-1}e^{-\Theta u}du \\ &= \frac{\Theta^P}{\Gamma(P)\sqrt{2\pi\sigma_v^2}} \int_{-\infty}^{+\infty} e^{-\frac{(\varepsilon+u)^2}{2\sigma_v^2}} u^{P-1}e^{-\Theta u}du. \end{aligned} \tag{2}$$

We use several observations in the steps above. Denote $\sigma^2_{\varepsilon+u}$ as σ^2_v. Also note that since $\varepsilon = v - u$, it is the case that $E[\varepsilon] = E[v-u] = -E[u]$, assuming that v is a standard normal variable. Then $E[\varepsilon+u] = E[\varepsilon] + E[u] = 0$.

7.2 The joint distribution of u and ε, $f(u, \varepsilon)$

Following the logic of Greene, use the assumption that u and v are independent. Then, their joint distribution can be expressed as the product of their marginal distributions:

$$\begin{aligned} f(v,u) &= f(v)f(u) \\ &= \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-\frac{v^2}{2\sigma_v^2}} \frac{\Theta^P}{\Gamma(P)} u^{P-1} e^{-\Theta u}. \end{aligned} \tag{3}$$

To obtain $f(\varepsilon, u)$, we substitute $\varepsilon + u$ instead of v in (3):

$$f(\varepsilon,u) = \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-\frac{(\varepsilon+u)^2}{2\sigma^2_{(\varepsilon+u)}}} \frac{\Theta^P}{\Gamma(P)} u^{P-1} e^{-\Theta u}. \tag{4}$$

Then,

$$\begin{aligned} f(u|\varepsilon) &= \frac{f(\varepsilon,u)}{f(\varepsilon)} \\ &= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\varepsilon+u)^2}{2\sigma^2_{(\varepsilon+u)}}} \frac{\Theta^P}{\Gamma(P)} u^{P-1} e^{-\Theta u}}{\frac{\Theta^P}{\Gamma(P)\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{(\varepsilon+u)^2}{2\sigma_v^2}} u^{P-1} e^{-\Theta u} du} \\ &= \frac{e^{-\frac{(\varepsilon+u)^2}{2\sigma^2_{(\varepsilon+u)}}} u^{P-1} e^{-\Theta u}}{\int_{-\infty}^{+\infty} e^{-\frac{(\varepsilon+u)^2}{2\sigma_v^2}} u^{P-1} e^{-\Theta u} du}. \end{aligned}$$

Then, in turn,

$$\begin{aligned} E[u|\varepsilon] &= \int_0^{\infty} u f(u|\varepsilon) du \\ &= \int_0^{\infty} u \frac{e^{-\frac{(\varepsilon+u)^2}{2\sigma^2_{(\varepsilon+u)}}} u^{P-1} e^{-\Theta u}}{\int_{-\infty}^{+\infty} e^{-\frac{(\varepsilon+u)^2}{2\sigma_v^2}} u^{P-1} e^{-\Theta u} du} du \\ &= \frac{\int_0^{\infty} e^{-\frac{(\varepsilon+u)^2}{2\sigma^2_{(\varepsilon+u)}}} u^{P} e^{-\Theta u} du}{\int_0^{+\infty} e^{-\frac{(\varepsilon+u)^2}{2\sigma_v^2}} u^{P-1} e^{-\Theta u} du}. \end{aligned}$$

This can also be expressed as:

$$E[u|u > 0, \varepsilon] = \frac{h(P,\varepsilon)}{h(P-1,\varepsilon)},$$

since

$$\frac{\int_0^\infty e^{-\frac{(\varepsilon+u)^2}{2\sigma^2(\varepsilon+u)}} u^P e^{-\Theta u} du}{\int_0^{+\infty} e^{-\frac{(\varepsilon+u)^2}{2\sigma_v^2}} u^{P-1} e^{-\Theta u} du} = \frac{\int_0^\infty u^P \phi(u) du}{\int_0^\infty u^{P-1} \phi(u) du}$$

We next show why that is the case

$$\begin{aligned} \frac{\int_0^\infty u^P \phi(u) du}{\int_0^\infty u^{P-1} \phi(u) du} &= \frac{\int_0^\infty u^P \phi(u) du}{\int_0^\infty \phi(u) du} \frac{\int_0^\infty \phi(u) du}{\int_0^\infty u^{P-1} \phi(u) du} \\ &= \frac{h(P,\varepsilon)}{h(P-1,\varepsilon)}. \end{aligned}$$

7.3 What is $E[u^{P-1}|u > 0, \varepsilon_i]$?

In this paper, we simulate $h(P-1, \varepsilon_i)$. In doing that, it is critical to note that

$$\int_0^\infty \left(\sigma_v \sqrt{2\pi}\right)^{-1} e^{-\frac{(\varepsilon_i + u + \sigma_v^2 \Theta)^2}{2\sigma_v^2}} du = \Pr(u > 0|\varepsilon_i),$$

where u is a normal truncated at zero variable with mean $\mu_i = -\left(\varepsilon_i + \sigma_v^2 \Theta\right)$ and variance σ_v^2[4].

As a result, it is also the case that

$$\frac{\int_0^\infty \left(\sigma\sqrt{2\pi}\right)^{-1} e^{-\frac{(\varepsilon_i + u_i + \sigma^2\Theta)^2}{2\sigma^2}} u^{P-1} du}{\int_0^\infty \left(\sigma\sqrt{2\pi}\right)^{-1} e^{-\frac{(\varepsilon_i + u_i + \sigma^2\Theta)^2}{2\sigma^2}} du} = E[u^{P-1}|u > 0, \varepsilon_i].$$

[4] To clarify more, we know that for the case of a normal variable with mean $\mu_u = \varepsilon_i + \sigma_v^2 \Theta$ and variance σ_v^2, we have

$$\Pr(u > a|\varepsilon_i) = 1 - \Phi(\frac{a - \mu_u}{\sigma_v^2}).$$

Use the fact that in this case $a = 0$ and that by symmetry $\Phi(-A) = 1 - \Phi(A)$. Then

$$\Pr(u > 0|\varepsilon_i) = 1 - \Phi(\frac{-\mu_u}{\sigma_v^2}) = \Phi(\frac{\mu_u}{\sigma_v^2}).$$

This is so since

$$\begin{aligned}
E[u^{P-1}|u &> 0,\varepsilon_i] = \int_0^\infty u^{P-1} f(u|u>0)du \\
&= \int_0^\infty u^{P-1}\frac{f(u)}{\Pr(u>0)}du \\
&= \frac{\int_0^\infty u^{P-1} f(u)du}{\Pr(u>0)} \\
&= \frac{\int_0^\infty u^{P-1} f(u)du}{\int_0^\infty f(u)du} \\
&= \frac{\int_0^\infty u^{P-1} \phi(u)du}{\int_0^\infty \phi(u)du}.
\end{aligned}$$

Note also that

$$E[u^{P-1}|u>0,\varepsilon_i] = h(P-1,\varepsilon_i),$$

as defined above.

7.4 Putting loglikelihood function into an appropriate shape

The gamma-normal log likelihood function can be expressed as:

$$\begin{aligned}
l(\sigma_v,\Theta,P) &= n\ln\Theta^P - n\ln\Gamma(P) - n\ln\sigma_v\sqrt{2\pi} \\
&\quad + \sum_{i=1}^n \ln\int_0^\infty e^{-\frac{(\varepsilon_i+u)^2}{2\sigma_v^2}-\Theta u} u^{P-1}du \\
&= n\ln\Theta^P - n\ln\Gamma(P) - n\ln\sigma_v\sqrt{2\pi} \\
&\quad + \sum_{i=1}^n \ln\int_0^\infty e^{\Theta\varepsilon_i+\frac{\sigma_v^2\Theta^2}{2}-\frac{(\varepsilon_i+u+\sigma_v^2\Theta)^2}{2\sigma_v^2}} u^{P-1}du \\
&= n\ln\Theta^P - n\ln\Gamma(P) + n\ln\left(\sigma_v\sqrt{2\pi}\right)^{-1} \\
&\quad + \sum_{i=1}^n \ln e^{\Theta\varepsilon_i+\frac{\sigma_v^2\Theta^2}{2}}\int_0^\infty e^{-\frac{(\varepsilon_i+u+\sigma_v^2\Theta)^2}{2\sigma_v^2}} u^{P-1}du \\
&= n\ln\Theta^P - n\ln\Gamma(P) + \sum \ln e^{\Theta\varepsilon_i+\frac{\sigma_v^2\Theta^2}{2}} \\
&\quad + \sum_{i=1}^n \ln\int_0^\infty \left(\sigma_v\sqrt{2\pi}\right)^{-1} e^{-\frac{(\varepsilon_i+u+\sigma_v^2\Theta)^2}{2\sigma_v^2}} u^{P-1}du \\
&= n\ln\Theta^P - n\ln\Gamma(P) + \sum_{i=1}^n \Theta\varepsilon_i + n\frac{\sigma_v^2\Theta^2}{2} \\
&\quad + \sum_{i=1}^n \ln\int_0^\infty \left(\sigma_v\sqrt{2\pi}\right)^{-1} e^{-\frac{(\varepsilon_i+u+\sigma_v^2\Theta)^2}{2\sigma_v^2}} u^{P-1}du,
\end{aligned}$$

which is the same as equation (30) in Greene (1990).

7.5 The Lindberg-Levy Central Limit Theorem

Theorem 2 *The Lindberg-Levy Central Limit Theorem (Greene, EA, 2000, p.117)*

If x_1, $x_2, ..., x_n$ are a random sample from a multivariate distribution with finite mean vector μ and finite positive definite covariance matrix Q, then

$$\sqrt{n}(\overline{x}_n - \mu) \xrightarrow{d} N\left[0, Q\right],$$

where

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

7.6 A short summary on truncated distributions (based on Greene, p.896)

7.6.1 Density of a Truncated Random Variable

If a continuous random variable u has a probability density function $f(u)$ and a is a constant, then

$$f(u|u > a) = \frac{f(u)}{\Pr(u > a)}.$$

More specifically, if we consider a truncated normal distribution, we have that (see p. 180 Lindgren)

$$\begin{aligned} \Pr(u &> a) = 1 - \Pr(u \le a) \\ &= 1 - \Pr(z \le \frac{a - \mu_u}{\sigma_u}) \\ &= 1 - \Phi(\frac{a - \mu_u}{\sigma_u}) \\ &= 1 - \Phi(\alpha), \end{aligned}$$

where μ_u and σ_u are the mean and the standard deviation of u, respectively, and $\alpha = \frac{a-\mu_u}{\sigma_u}$ and $\Phi(\frac{a-\mu_u}{\sigma_u}) = \int_{-\infty}^{\frac{a-\mu_u}{\sigma_u}} \phi(z)dz$ is the standard normal cumulative distribution function. Then the density of the truncated normal distribution is

$$\begin{aligned} f(u|u &> a) = \frac{f(u)}{1 - \Phi(\alpha)} \\ &= \frac{\phi(\frac{u-\mu_u}{\sigma_u})}{\sigma\left(1 - \Phi(\alpha)\right)} \end{aligned}$$

where $\phi(.)$ is the standard normal probability density function

$$\phi(z) = \frac{1}{\sigma_z\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma_z^2}},$$

and

$$f(u) = \frac{\phi(\frac{u-\mu_u}{\sigma_u})}{\sigma}$$

because by Theorem 7, Lindgren, p.64, the pdf of a linear transformation of $Z : U = a + bZ$ is

$$f_U(u) = f_Z(\frac{u-a}{b})\frac{1}{|b|}.$$

7.6.2 Moments of the Truncated Normal Distribution

If $u\tilde{}N[\mu_u, \sigma_u^2]$ and a is a constant, then

$$\begin{aligned} E[u|truncation] &= \mu_u + \sigma_u\lambda(\alpha) \\ var[u|truncation] &= \sigma_u^2[1 - \delta(\alpha)] \end{aligned}$$

where

$$\begin{aligned} \lambda(\alpha) &= \begin{cases} \frac{\phi(\alpha)}{1-\Phi(\alpha)}, & \text{if truncation is } u > a \\ -\frac{\phi(\alpha)}{\Phi(\alpha)}, & \text{if truncation is } u < a \end{cases} \\ \delta(\alpha) &= \frac{\lambda(\alpha)}{\lambda(\alpha) - \alpha}. \end{aligned}$$

Usually, the truncation point is $a = 0$ to represent the assumption that the error term u is positive. However, in principle a can be any number.

8 References

1. Varian

2. Jondrow, Materov

3. Lindgren, B.

4. Greene 1990

5. Greene 2000

6. Simar 2000

7. Khitarishvili, T. (2000)

8. Materials of ACDI/VOCA conference "Transformation Process of Cooperative Banking Sector" held in Warsaw, February 1998

9. Geweke, Runkle and Keane (1994)

10 Stern, FEL 1997.