Bounded Stochastic Frontiers with an Application to the US Banking Industry: 1984-2009 [1]

Pavlos Almanidis
Department of Economics
Rice University
pa1@rice.edu

Junhui Qian
School of Economics
Shanghai Jiao Tong University
jhqian@sjtu.edu.cn

Robin C. Sickles
Department of Economics
Rice University
rsickles@rice.edu

Previous version: October 1, 2008

This version: January 30, 2010

[1] Earlier versions of this paper were presented at the Tenth European Workshop on Efficiency and Productivity, Lille, France, June, 2007, and the Annual Texas Econometrics Camp XIII, Kerrville, Texas, February, 2008. The authors would like to thank C. A. Knox Lovell, Peter Schmidt, and Christopher Parmeter for their helpful comments and insights. The usual caveat applies.

Stochastic Frontiers with Bounded Inefficiency

Abstract

This paper introduces a new model of stochastic production frontier that incorporates an unobservable bound for inefficiency, which is naturally instituted by market competition. We consider doubly truncated normal, truncated half-normal, and truncated exponential distributions to model the inefficiency component of the error term. We derive the form of density function for the error term of each specification, expressions of the conditional mean of inefficiency levels, and provide proofs of local identifiability of these models under differing assumptions about the deep parameters of the distributions. We examine skewness properties of our new estimators and provide an explanation for the finding of "incorrect" skewness in many applied studies using the traditional stochastic frontier. We extend the model to the panel data setting and specify a time-varying inefficiency bound as well as time-varying efficiencies. A Monte Carlo study is conducted to study the finite sample performance of the maximum likelihood estimators in cross-sectional settings. Lastly we apply the model to a study of US banks from 1984 to 2009 using a recently developed panel of over 4000 banks and also compare out findings to those based on a set of competing specifications of the stochastic frontier model. We find substantial increases in efficiency after the regulatory reforms of the 1980's but also substantial backsliding during the 2005-2009 period presaging the financial meltdown experienced in the US and elsewhere in the last few years of the decade. This is the first study of which we are aware to examine such determinants of the recent financial malaise that has swept the international banking industry and international economies.

JEL classification codes: C13, C21, C23, D24, G21.

Key words and phrases: Stochastic frontier, bounded inefficiency, time-varying technical efficiency, doubly truncated normal, truncated half normal, truncated exponential, banking efficiency.

1 Introduction

The parametric approach to estimate stochastic production frontiers was introduced by Aigner, Lovell, and Schmidt (1977), Meeusen and van den Broeck (1977), and Battese and Corra (1977). These approaches specified a parametric production function and a two-component error term. One component, reflecting the influence of many unaccountable factors on production as well as measurement error, is considered "noise"and is usually assumed to be normal. The other component describes inefficiency and is assumed to have a one-sided distribution, of which the conventional candidates include the half normal (Aigner, et al., 1977), truncated normal (Stevenson, 1980), exponential (Meeusen and van den Broeck, 1977) and gamma (Greene 1980a,b, Stevenson, 1980).

In this paper we propose a new class of parametric stochastic frontier models with a more flexible specification of the inefficiency term. Instead of allowing unbounded support for the distribution of productive (cost) inefficiency term in the right (left) tail, we introduce an unobservable upper bound to inefficiencies or a lower bound to the efficiencies, which we call the *inefficiency bound*. The introduction of the inefficiency bound makes the parametric stochastic frontier model more appealing for empirical studies in at least two aspects. Firstly, it is plausible to allow only bounded support in many applications of stochastic frontier models. We may think of two scenarios. One is where we study a competitive industry or market from which the extremely inefficient firms are eliminated by competition. The other is where we study the efficiency profile of chain stores of large retail chains or franchises. Bounded inefficiency makes sense in this scenario since the extremely inefficient stores may be forced to closure by some fixed rule. In both scenarios, the individual production units constitute a truncated sample. The consequence is that even if we correctly specify a family of distributions for the inefficiency term, the stochastic frontier model may still be misspecified.

Secondly, the analysis of our model points to an explanation for the finding of "wrong "skewness in many applied studies using the traditional stochastic frontier and the potential for our bounded inefficiency model to explain these "incorrect"skewness findings. Researchers have often found positive instead of negative skewness in many samples examined in applied work, which may point to the stochastic frontier being incorrectly specified. However, we conjecture that the distribution of the inefficiency term may itself be negatively skewed, which may happen if there is an additional truncation on the right tail of the distribution. In particular, we propose a model where the distribution of the inefficiency term is doubly truncated normal, that is, a normal distribution truncated at a point on the right tail as well as at zero. As normal distributions are symmetric, the doubly truncated normal distribution may exhibit negative skewness if the truncation on the right is closer to the mode than that on the left.

In addition to the doubly truncated normal distribution, we also consider the truncated half normal distribution, which is a special case of the former, and the truncated exponential distribution. Although these two distributions are always positively skewed, the fact that there is a truncation on the right tail makes the skewness very hard to identify empirically. That is to say, when the true distribution of the one-sided inefficiency error is bounded (truncated), the extent to which skewness is present may be substantially reduced, often to the extent that negative sample skewness for the composite error is not statistically significant. Thus the finding of positive skewness may speak to the weak identifiability of skewness properties in a bounded frontier model.

We show that our models are identifiable, either locally or globally. We propose to use method of moments to obtain initial guess of parameters and to use maximum likelihood estimation to obtain more precise estimates. The analytic forms of the moments and the density functions of the composite error are provided. Simulation results show that the model parameters, including the inefficiency bound, can be consistently estimated by the maximum likelihood estimation. The inefficiency bound can naturally be used for gauging the tolerance for or ruthlessness against the inefficient firms. It is also worth mentioning that, using this bound as the "inefficient frontier," we may define "inverted" efficiency scores in the same spirit of "Inverted DEA" described in Entani, Maeda, and Tanaka (2002).

We also extend the model to the panel data setting and allow for time-varying inefficiency bound. By allowing the inefficiency bound to be time-varying, we in effect contribute another time-varying technical efficiency model. Our model differs from the existing literature in that, while previous time-varying efficiency models, notably Cornwell, Schmidt, and Sickles (1990), Kumbhakar (1990), Battese and Coelli (1992), and Lee and Schmidt (1993), are time-varying in the mean or intercept of individual effects, our model is time-varying in the lower support of the distribution of individual effects.

The outline of this paper is as follows. In Section 2 we present the new models and derive analytic formula for density functions and the calculation of inefficiencies. Section 3 deals with the "wrong" skewness issue inherent in traditional stochastic frontier model. Section 4 discusses the identification of the new models and the methods of estimation. Section 5 presents Monte Carlo results on the finite sample performance of the bounded inefficiency model vis-a-vis classical stochastic frontier estimators. The extension of the new models to panel data settings and specification of time-varying bound is presented in section 6. In Section 7 we give an illustrative study of the efficiency of US banking industry in 1984-2009. Section 8 concludes.

2 The Model

We consider the following Cobb-Douglas log-linear production model,

$$y_i = \alpha_0 + \sum_{k=1}^{K} \alpha_k x_{i,k} + \varepsilon_i \tag{1}$$

where

$$\varepsilon_i = v_i - u_i. \tag{2}$$

For every production unit i, y_i is the log output, x_{ik} the k-th log input, v_i the noise component, and u_i the inefficiency component. We maintain the usual assumption that v_i is iid $N(0, \sigma_v^2)$, u_i is iid, and v_i and u_i are independent from each other and from regressors.

As described in the introduction, our model differs from the traditional stochastic frontier model in that u_i is of bounded support. Additional to the lower bound, which is zero and which is the frontier, we specify an upper bound to the distribution of u_i. In particular, we assume that u_i is distributed as doubly truncated normal, the density of which is given by

$$f(u) = \frac{\frac{1}{\sigma_u}\phi(\frac{u-\mu}{\sigma_u})}{\Phi(\frac{B-\mu}{\sigma_u}) - \Phi(\frac{-\mu}{\sigma_u})} \mathbf{1}_{[0,B]}(u),\ \ \sigma_u > 0, B > 0, \tag{3}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and pdf of the standard normal distribution, respectively, and $1_{[0,B]}$ is an indicator function. It is a distribution obtained by truncating $N(\mu, \sigma_u^2)$ at zero and $B > 0$. The parameter B is the upper bound of the distribution of u_i and we may call it the inefficiency bound. The inefficiency bound may be a useful index of competitiveness of a market or an industry.

Using the usual nomenclature of stochastic frontier models, we may call the model described above the normal-doubly truncated normal model, or simply, the doubly truncated normal model. The doubly truncated normal model is very flexible. It nests truncated normal ($B = \infty$), half normal ($\mu = 0$ and $B = \infty$), and truncated half normal models ($\mu = 0$). One desirable feature of our model is that the doubly truncated normal distribution may be positively or negatively skewed, depending on the truncated parameter B. This feature provides us with an alternative explanation for the "wrong skewness" problem prevalent in empirical stochastic frontier studies. This will be made more clear later in this section. Another desirable feature of our model is that, like the truncated normal model, it can describe the scenario that only a few firms in the sector are efficient, a phenomenon that is described in the business press as "few stars, most dogs". [2] While in the truncated half normal model

[2] We thank C. A. K. Lovell for providing us this link between our econometric methdology and the business press.

and the truncated exponential model (in which the distribution of u_i is truncated exponential), most firms are implicitly assumed to be relatively efficient.

In Table 1 we provide detailed properties of our model. In particular, we present the density functions for the error term ε_i, which is necessary for maximum likelihood estimation, and the analytic form for $E[u_i|\varepsilon_i]$, which is the best predictor of the inefficiency term u_i under our assumptions, and the conditional distribution of u_i given ε_i, which is useful for making inferences on u_i. The results for the truncated half normal model, a special case of the doubly truncated normal model ($\mu = 0$), are also presented. Finally, we also provide results for the truncated exponential model, in which the inefficiency term u_i is distributed according to the following density function,

$$f(u) = \frac{1}{\sigma_u(1 - e^{-B/\sigma_u})} e^{-\frac{u}{\sigma_u}} \mathbf{1}_{[0,B]}(u). \tag{4}$$

The truncated exponential distribution shares with the doubly normal distribution the nice property that it may be positively or negatively skewed.

For the doubly truncated normal model and the truncated half normal model, the analytic forms of our results use the so-called $\lambda-$parametrization, which specifies

$$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}, \quad \lambda = \sigma_u/\sigma_v. \tag{5}$$

In practice we usually use another parameterization, called the $\gamma-$parametrization,

$$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}, \quad \gamma = \sigma_u^2/\sigma^2. \tag{6}$$

By definition $\gamma \in [0, 1]$, a compact support, which is desirable for the numerical procedure of maximum likelihood estimation.

We may check that when $B \to \infty$, the density function for ε_i in the doubly truncated normal model reduces to that of the truncated normal model introduced by Stevenson (1980). Furthermore, if $\mu = 0$, it reduces to the likelihood function for the half normal model introduced by Aigner, Lovell, and Schmidt (1977). Similarly, the truncated exponential model reduces to the exponential model introduced by Meeusen and van den Broeck (1977). page

3 The Skewness Issue

A common and important methodological problem encountered when dealing with empirical implementation of the stochastic frontier model is that the residuals may be skewed in the wrong direction. In particular, the OLS residuals may show positive skewness even though the composed error term $v-u$ should display negative skewness, in keeping with $u's$ positive skewness. This problem has important consequences for the interpretation of the skewness of the error term as a measure of technological

Table 1: Key Results

$f(\varepsilon)$ is the density of $\varepsilon = v - u$, $\mathbb{E}(u|\varepsilon)$ is the conditional mean of u given ε, and $f(u|\varepsilon)$ is the conditional density of u given ε. $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively. And $\mathbf{1}_{[0,B]}(\cdot)$ is an indicator function.

Model	$f(\varepsilon)$	$\mathbb{E}(u\|\varepsilon)$	$f(u\|\varepsilon)$
Doubly truncated normal	$\left[\Phi(\frac{B-\mu}{\sigma_u}) - \Phi(\frac{-\mu}{\sigma_u})\right]^{-1} \cdot \left[\frac{1}{\sigma}\phi(\frac{\varepsilon+\mu}{\sigma})\right] \cdot \left[\Phi(\frac{(B+\varepsilon)\lambda+(B-\mu)\lambda^{-1}}{\sigma}) - \Phi(\frac{\varepsilon\lambda-\mu\lambda^{-1}}{\sigma})\right]$	$\mu_* + \sigma_* \frac{\phi(-\frac{\mu_*}{\sigma_*})-\phi(\frac{B-\mu_*}{\sigma_*})}{\Phi(\frac{B-\mu_*}{\sigma_*})-\Phi(-\frac{\mu_*}{\sigma_*})}$	$\frac{\frac{1}{\sigma_*}\phi(\frac{u-\mu_*}{\sigma_*})}{\Phi(\frac{B-\mu_*}{\sigma_*})-\Phi(-\frac{\mu_*}{\sigma_*})}\mathbf{1}_{[0,B]}(u)$
	$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$, $\lambda = \sigma_u/\sigma_v$	$\mu_* = \frac{\mu\sigma_v^2 - \varepsilon\sigma_u^2}{\sigma^2}$, $\sigma_* = \frac{\sigma_u\sigma_v}{\sigma}$	
Truncated half normal	$[\Phi(\frac{B}{\sigma_u}) - 1/2]^{-1} \cdot \frac{1}{\sigma}\phi(\frac{\varepsilon}{\sigma}) \cdot [\Phi(\frac{(B+\varepsilon)\lambda+B\lambda^{-1}}{\sigma}) - \Phi(\frac{\varepsilon\lambda}{\sigma})]$	$\mu_* + \sigma_* \frac{\phi(-\frac{\mu_*}{\sigma_*})-\phi(\frac{B-\mu_*}{\sigma_*})}{\Phi(\frac{B-\mu_*}{\sigma_*})-\Phi(-\frac{\mu_*}{\sigma_*})}$	$\frac{\frac{1}{\sigma_*}\phi(\frac{u-\mu_*}{\sigma_*})}{\Phi(\frac{B-\mu_*}{\sigma_*})-\Phi(-\frac{\mu_*}{\sigma_*})}\mathbf{1}_{[0,B]}(u)$
		$\mu_* = -\frac{\varepsilon\sigma_u^2}{\sigma^2}$, $\sigma_* = \frac{\sigma_u\sigma_v}{\sigma}$	
Truncated exponential	$\frac{e^{\frac{\varepsilon}{\sigma_u}+\frac{\sigma_v^2}{2\sigma_u^2}}[\Phi(\frac{B+\varepsilon}{\sigma_v}+\frac{\sigma_v}{\sigma_u})-\Phi(\frac{\varepsilon}{\sigma_v}+\frac{\sigma_v}{\sigma_u})]}{\sigma_u(1-e^{-\sigma_u/B})}$	$\mu_* + \sigma_v \frac{\phi(-\frac{\mu_*}{\sigma_v})-\phi(\frac{B-\mu_*}{\sigma_v})}{\Phi(\frac{B-\mu_*}{\sigma_v})-\Phi(-\frac{\mu_*}{\sigma_v})}$	$\frac{\frac{1}{\sigma_v}\phi(\frac{u-\mu_*}{\sigma_v})}{\Phi(\frac{B-\mu_*}{\sigma_v})-\Phi(-\frac{\mu_*}{\sigma_v})}\mathbf{1}_{[0,B]}(u)$
		$\mu_* = -\varepsilon - \frac{\sigma_v^2}{\sigma_u}$	

inefficiency. It may imply that there had been an unfortunate sampling from an inefficiency distribution that has a correct population skewness. It may also be that positive skewness of the composed error indicates that there are no inefficiencies and that all firms are "super efficient", a term first used by Green and Mayes (1991). The later would suggest setting the variance of inefficiency term at zero, which would have problematic impacts on estimation and on inference. Carree (2002) considers one-sided distributions of inefficiencies (u_i) that can have negative or positive skewness. However, Carree (2002) uses the binomial distribution, which is a discrete distribution and which implicitly assumes that only a very small fraction of the firms attain a level of productivity close to the frontier, especially when u_i is negatively skewed.

Our model addresses the "wrong skewness" problem similarly with Carree (2002)[3], but with a more appealing distributional specification on the efficiency term. For the doubly truncated normal model, let $\xi_1 = \frac{-\mu}{\sigma_u}$, $\xi_2 = \frac{B-\mu}{\sigma_u}$, and $\eta_k \equiv \frac{\xi_1^k \phi(\xi_1) - \xi_2^k \phi(\xi_2)}{\Phi(\xi_2) - \Phi(\xi_1)}$, $k = 0, 1, ..., 4$. Note that η_0 is the inverse Mill's ratio and it is equal to $\sqrt{2/\pi}$ in the half normal model, and that ξ_1 and ξ_2 are the lower and upper truncation points of the standard normal density, respectively. The skewness of the doubly truncated normal distribution is given by

$$S_u = \frac{2\eta_0^3 - \eta_0(3\eta_1 + 1) + \eta_2}{(1 - \eta_0^2 + \eta_1)^{3/2}}. \tag{7}$$

It can be checked that when $B > 2\mu$, S_u is positive. And when $B < 2\mu$, S_u is negative. Since $B > 0$ by definition, it is obvious that only when $\mu > 0$ is it possible for u_i to be negatively skewed. And the larger μ is, the larger range of values B may take such that u_i is negatively skewed. Consider the limiting case where a normal distribution with $\mu \to \infty$ is truncated at zero and $B > 0$. An infinitely large μ means that there is effectively no truncation on the left at all and that any finite truncation on the right gives rise to a negative skewness. Finally, for both the truncated half normal model ($\mu = 0$) and the truncated exponential model, the skewness of u_i is always positive.

Consequently, the doubly truncated normal model has a residual that has an ambiguous sign of the skewness, which depends on an unobservable relationship between the truncation parameter B and μ. We argue that the ambiguity theoretically explains the prevalence of the "wrong" skewness problem in applied stochastic frontier research. When the underlying data generating process for u_i is based on doubly truncated normal distribution, increasing sample size does not solve the "wrong skewness" problem. The skewness of the OLS residual ε may be positively skewed even when sample size goes to infinity. Hence the "wrong" skewness problem may also be a large sample problem.

[3] We thank C. A. Knox Lovell for his observation, which he made at the Tenth European Workshop on Efficiency and Productivity, Lille, France, June, 2007, that there was potential for our bounded frontier to address the skewness problem inherent in the use of the Aigner, Lovell and Schmidt stochastic frontier model

In finite samples, we may use simulations to show that our model is capable of generating residuals with "wrong" skewness with higher frequency than traditional stochastic frontier models do (Simar and Wilson, 2009). We generate samples of the residuals $\varepsilon = v - u$ with u being doubly truncated normal. And we calculate the proportion of samples with positively skewed residuals in 1000 repeated experiments. We set the parameter μ to 1 and examine the proportions of positive skewness when B is 1, 2, 5, and 10. We also experiment different values of λ and sample sizes from 50 to 10^5. The results are reported in Table 2.

The first column ($B = 1$) shows that the proportion of the samples with the positive ("wrong") skewness increases as the sample size gets larger. It appears to converge to one as the sample size increases, especially when the signal-noise-ratio λ is large. The second column corresponds to the case where $B = 2\mu$. In this case there is about a 50-50 chance that we generate a sample with positive skewness. In other words, the positive skewness appears to be statistically insignificant in most of the cases. The third column ($B = 5$) and the fourth column ($B = 10$) correspond to the case where the distribution of inefficiencies is positively skewed. In particular, the results in the fourth column are similar with those reported in Simar and Wilson (2009) for traditional stochastic frontier models.

Our simulation results confirm that the skewness issue is also a large sample issue, since for $B < 2\mu$ the proportion of the samples with positive skewness converges to one. This would mean that if the true data generating process is based on inefficiencies that are drawn from a doubly truncated normal distribution, and if a researcher fails to recognize this and finds a skewness statistic with the wrong sign, then she may erroneously reject her model. Moreover, if there is the potential for increasing sample size and the researcher keeps increasing it and finds continuously positive signs of skewness, then she may erroneously conclude that all firms in her sample are super efficient. The bounded inefficiency model, the doubly truncated normal model in particular, avoids this problem.

As a conclusion of this section, the doubly truncated normal model generalizes the stochastic frontier model in a way that allows for positive as well as negative skewness for the residual. This implies that finding a "wrong" skewness does not necessarily mean that the stochastic frontier model is inapplicable. It may only be that we are studying a market or an industry that is instituted an inefficiency bounded. Hence the traditional unbounded support for the inefficiency term is misspecified and the model of bounded inefficiency should be used instead.

Table 2: Proportion of Positive Skewness for Simulated Residuals in the Doubly Truncated Normal Model. The residuals are generated as in (2) with u being doubly truncated normal with $\mu = 1$. All figures in the table are proportions of experiments in 1000 repeated runs that show positive or "wrong" skewness of the residual.

	n	$B = 1$	$B = 2$	$B = 5$	$B = 10$
	50	0.519	0.505	0.480	0.509
	100	0.481	0.501	0.516	0.520
	200	0.495	0.473	0.514	0.493
$\lambda = 0.1$	500	0.487	0.503	0.539	0.507
	10^3	0.520	0.516	0.510	0.494
	10^4	0.504	0.483	0.512	0.498
	10^5	0.532	0.492	0.437	0.405
	50	0.517	0.485	0.503	0.510
	100	0.545	0.491	0.459	0.479
	200	0.551	0.490	0.486	0.466
$\lambda = 0.5$	500	0.520	0.488	0.431	0.459
	10^3	0.564	0.514	0.453	0.435
	10^4	0.684	0.491	0.397	0.318
	10^5	0.759	0.496	0.107	0.092
	50	0.565	0.536	0.367	0.383
	100	0.524	0.513	0.317	0.335
	200	0.529	0.512	0.224	0.245
$\lambda = 1$	500	0.567	0.514	0.155	0.122
	10^3	0.576	0.524	0.063	0.051
	10^4	0.709	0.501	0	0
	10^5	0.943	0.503	0	0

Table 3: Central Moments of ε.

Moment	Doubly-truncated-normal
ψ_1	$-\mu-\sigma_u\eta_0$
ψ_2	$\sigma_u^2\left(1-\eta_0^2+\eta_1\right)+\sigma_v^2$
ψ_3	$-\sigma_u^3\left(2\eta_0^3-3\eta_1\eta_0-\eta_0+\eta_2\right)$
ψ_4	$\sigma_u^4\left(3+3\eta_1+\eta_3-2\eta_0^2-4\eta_0\eta_2+6\eta_0^2\eta_1-3\eta_0^4\right)+6\sigma_u^2\sigma_v^2\left(1-\eta_0^2+\eta_1\right)+3\sigma_v^4$
ψ_5	$-10\sigma_v^2\sigma_u^3\left(2\eta_0^3-3\eta_1\eta_0-\eta_0+\eta_2\right)$ $-\sigma_u^5\left(\eta_4+4\eta_2-5\eta_0\eta_3+10\eta_0^2\eta_2-10\eta_0^3\eta_1+10\eta_0^3-15\eta_0\eta_1+4\eta_0^5-7\eta_0\right)$
	See the text for the definitions of η_k, $k=0,...,4$.
	Truncated-exp.
ψ_1	$-\sigma_u\left(1-\frac{\kappa}{e^\kappa-1}\right)$
ψ_2	$\sigma_v^2+\sigma_u^2\frac{e^{2\kappa}-\left(\kappa^2+2\right)e^\kappa+1}{e^{2\kappa}-2e^\kappa+1}$
ψ_3	$-\sigma_u^3\frac{2e^{3\kappa}-\left(\kappa^3+6\right)e^{2\kappa}+\left(6-\kappa^3\right)e^\kappa-2}{e^{3\kappa}-3e^{2\kappa}+3e^\kappa-1}$
ψ_4	$\sigma_u^4\frac{-9e^{4\kappa}+36e^{3\kappa}-54e^{2\kappa}+36e^\kappa-9+6\kappa^2e^\kappa(e^{2\kappa}-2e^\kappa+1)+\kappa^4e^\kappa(e^{2\kappa}+e^\kappa+1)}{-e^{4\kappa}+4e^{3\kappa}-6e^{2\kappa}+4e^\kappa-1}+$ $6\sigma_v^2\sigma_u^2\frac{e^{2\kappa}-\left(\kappa^2+2\right)e^\kappa+1}{e^{2\kappa}-2e^\kappa+1}+3\sigma_v^4,\quad \kappa=B/\sigma_u.$

4 Estimation

4.1 Identification

As will be made more clear later, the identification of our model may be done in two parts. The first part is concerned with the parameters describing the technology, and the second part identifies the distributional parameters using the information contained in the distribution of the residual. The identification conditions for the first part are well known and are satisfied in most of the cases. The second part deserves a close examination. In Table 3, we list the population (central) moments of (ε_i) for the doubly truncated normal model and the truncated exponential model. The moments of the truncated half normal model can be obtained by setting $\mu=0$ in the doubly truncated normal model. These results are essential for the discussion of identification and the method of moments estimation.

Both the truncated half normal model and the truncated exponential model are globally identified. To see this, we check that for both models, $\psi_3^{-4/3}(\psi_4-3\psi_2^2)$ is a function of $\kappa=B/\sigma_u$ only, which we denote as $g(\kappa)$. The forms of g are complicated and hence omitted. For the truncated half normal model, g is monotonically decreasing; and for the truncated exponential model, g is monotonically increasing. In both cases, g is invertible and κ can be identified. The identification of other parameters then follows. Note, however, for large values of κ (e.g., $\kappa > 5$ for the truncated normal model and $\kappa > 20$), the curve $g(\kappa)$ is nearly flat and gives poor identification.

Unfortunately, the doubly truncated normal model is not globally identifiable. However, local identification can be verified. We may examine $\psi_3^{-4/3}(\psi_4 - 3\psi_2^2)$ and $\psi_3^{-5/3}(\psi_5 - 10\psi_2\psi_3)$, both of which are functions of ξ_1 and ξ_2 only and we denote them as $g_1(\xi_1, \xi_2)$ and $g_2(\xi_1, \xi_2)$, respectively. Let $\hat{g}_1$ and $\hat{g}_2$ be the sample versions of g_1 and g_2, respectively, we have the following system of identification equations,

$$
\begin{aligned}
G_1(\xi_1, \xi_2) &\equiv g_1(\xi_1, \xi_2) - \hat{g}_1 &= 0 \\
G_2(\xi_1, \xi_2) &\equiv g_2(\xi_1, \xi_2) - \hat{g}_2 &= 0.
\end{aligned}
$$

By the implicit function theorem, the identification of ξ_1 and ξ_2 depends on the matrix

$$
H = \begin{pmatrix} \frac{\partial g_1}{\partial \xi_1} & \frac{\partial g_1}{\partial \xi_2} \\ \frac{\partial g_2}{\partial \xi_1} & \frac{\partial g_1}{\partial \xi_2} \end{pmatrix}.
$$

If H is invertible (the determinant is nonzero), then ξ_1 and ξ_2 can be written as functions of $\hat{g}_1$ and $\hat{g}_2$; the identification of the model then follows. The analytic form of H is very complicated, but we may examine the invertibility of H by numerically evaluating g_1 and g_2 and inferring the sign of each element in H. It can be verified that the determinant of H is nonzero in neighborhoods within I_1, I_2, I_4, and I_6, the definitions of which are given as follows,

(i) $I_1 \equiv \{(\mu, B) | \mu \leq 0, B > 0\}$

(ii) $I_2 \equiv \{(\mu, B) | \mu > 0, B \in (0, \mu)\}$

(iii) $I_3 \equiv \{(\mu, B) | B = \mu > 0\}$

(iv) $I_4 \equiv \{(\mu, B) | \mu > 0, B \in (\mu, 2\mu)\}$

(v) $I_5 \equiv \{(\mu, B) | B = 2\mu > 0\}$

(vi) $I_6 \equiv \{(\mu, B) | \mu > 0, B > 2\mu\}$.

It can be verified that the model is not locally identifiable on the line that corresponds to I_3. This is also confirmed by simulation results (not shown in this paper). However, if we restrict $B = \mu$ a priori, we obtain a sub-model that is globally identifiable. By the shape of the doubly truncated normal distribution when $B = \mu$, we may call this special sub-model the "inverted" truncated half normal model. Unlike the truncated half normal model, the inverted truncated half normal model describes markets that have relatively few efficient firms close to the efficiency frontier.

The line $I_5 \equiv \{(\mu, B) | B = 2\mu > 0\}$ corresponds to the case where $B = 2\mu$ and $\psi_3 = 0$, hence the above strategy fails. Nonetheless, simulation results in the next section show that when the true values of B and μ satisfy $B = 2\mu$, both B and μ are consistently estimated without putting in the restriction $B = 2\mu$. This indicates that the restricted ($B = 2\mu$) model may be nested in the unrestricted model and the model is locally identifiable on $I_4 \bigcup I_5 \bigcup I_6$.

Strictly speaking, the doubly truncated normal model should be understood as a collection of different sub-models corresponding to the different domains of parameters. Treated separately, each of the sub-models is globally identified. In maximum likelihood estimation, the separate treatment is easily achieved by constrained optimization on each parameter subset. Finally, note that on the line $\{(\mu, B)|\mu = 0, B > 0\} \subset I_1$, the doubly truncated normal model reduces to the truncated half normal model.

4.2 Method of Moment Estimation

The method of moments (Olson, Schmidt, and Waldman, 1980) may be employed to estimate our model or to obtain initial values for maximum likelihood estimation. In the first step of this approach, OLS is used to obtain consistent estimates of the parameters describing the technology, apart from the intercept. In the second step, using the distributional assumptions on the residual, equations of moment conditions are solved to obtain estimates of the parameters describing the distribution of the residual.

More specifically, we may rewrite the production frontier model in (1) and (2) as

$$y_i = (\alpha_0 - \mathbb{E}u_i) + \sum_{k=1}^{K} \alpha_k x_{i,k} + \varepsilon_i^*,$$

where $\varepsilon_i^* = v_i - (\mathbb{E}u_i)$. The error term ε_i^* is of zero mean and constant variance. Hence the OLS yields consistent estimates for ε_i^* and α_k, $k = 1, ..., K$. Equating the sample moments of $(\hat{\varepsilon}_i^*)$ to the population moments, we solve for the parameters that are associated with the distribution of (ε_i^*). Note that since $E\varepsilon_i^* = 0$, the central moments of ε_i are equal to the moments of ε_i^*.

4.3 Maximum Likelihood Estimation

For more efficient estimation, we may use maximum likelihood estimation. Note that with the presence of a noise term v_i, the range of residual unbounded and does not depend on the parameter. In the remaining of this section we list the log-likelihood functions of all three models. Note that in practice we may also need the gradients of the log likelihood function. The gradients are complicated in form but straightforward to derive and we omit them here.

The log-likelihood function for the doubly truncated normal model with λ parameterization is given by

$$\begin{aligned}
\log L &= -n\log\left[\Phi(\frac{B-\mu}{\sigma_u(\sigma,\lambda)}) - \Phi(\frac{-\mu}{\sigma_u(\sigma,\lambda)})\right] \\
&\quad -n\log\sigma - \frac{n}{2}\log(2\pi) - \sum_{i=1}^{n}\frac{(\varepsilon_i+\mu)^2}{2\sigma^2} \\
&\quad +\sum_{i=1}^{n}\log\{\Phi(\frac{(B+\varepsilon_i)\lambda + (B-\mu)\lambda^{-1}}{\sigma}) \\
&\qquad\qquad -\Phi(\frac{\varepsilon_i\lambda - \mu\lambda^{-1}}{\sigma})\},
\end{aligned} \tag{8}$$

where $\varepsilon_i = y - \alpha_0 - \sum x_{i,k}\alpha_k$ and

$$\sigma_u(\sigma,\lambda) = \frac{\sigma}{\sqrt{1+1/\lambda^2}}. \tag{9}$$

It is easy to get $\log L$ in $\gamma-$parametrization. We can substitute λ in (8) with

$$\lambda(\gamma) = \sqrt{\frac{\gamma}{1-\gamma}}. \tag{10}$$

The log-likelihood function for the truncated half normal model is

$$\begin{aligned}
\log L &= -n\log(\Phi(\frac{B}{\sigma_u(\sigma,\lambda)}) - \frac{1}{2}) \\
&\quad -n\log\sigma - \frac{n}{2}\log(2\pi) - \sum_{i=1}^{n}\frac{\varepsilon_i^2}{2\sigma^2} \\
&\quad +\sum_{i=1}^{n}\log\{\Phi(\frac{(B+\varepsilon_i)\lambda + B\lambda^{-1}}{\sigma}) - \Phi(\frac{\varepsilon_i\lambda}{\sigma})\},
\end{aligned} \tag{11}$$

where $\varepsilon_i = y - \alpha_0 - \sum x_{i,k}\alpha_k$ and $\sigma_u(\sigma,\lambda)$ is defined in (9). Again, substitute λ in (11) with $\lambda(\gamma)$ in (10), we get $\log L$ of γparametrization.

Finally, the log-likelihood function for the truncated exponential model is

$$\begin{aligned}
\log L &= -n\log\sigma_u - n\log(1-e^{-\sigma_u/B}) \\
&\quad +\frac{n\sigma_v^2}{2\sigma_u^2} + \frac{1}{\sigma_u}\sum_{i=1}^{n}\varepsilon_i && (12) \\
&\quad +\sum_{i=1}^{n}\log[\Phi(\frac{B+\varepsilon_i}{\sigma_v} + \frac{\sigma_v}{\sigma_u}) - \Phi(\frac{\varepsilon_i}{\sigma_v} + \frac{\sigma_v}{\sigma_u})], && (13)
\end{aligned}$$

where $\varepsilon_i = y - \alpha_0 - \sum x_{i,k}\alpha_k$.

After estimating the model, we can estimate the composed error term ε_i:

$$\hat{\varepsilon}_i = y_i - \hat{\alpha}_0 - \sum x_{i,k}\hat{\alpha}_k, i = 1, \cdots, n. \quad (14)$$

From this we can estimate the inefficiency term u_i using the formula for $E(u_i|\varepsilon_i)$ in Table 1.

5 Simulations

To examine the finite sample performance of the MLE estimators we run a series of Monte Carlo experiments for the standard cross-sectional stochastic frontier model. The data generating process is (1) and (2) with $\alpha_0 = 0$ and $K = 2$ (two regressors and no constant term)[4]. Throughout we set $\alpha_1 = 0.6$, $\alpha_2 = 0.5$. We set $\sigma_u = 0.3$ in all three submodels. To examine how the noise level (σ_v) affects the quality of estimation, we vary σ_v from 0.1, 0.2, to 0.5. In the other dimension, we change the inefficiency bound from 0.8, 1.0, to 1.2, to examine its impact on estimation. For both normal-truncated half normal and normal-doubly truncated normal models we use the γ-parameterization, and thus the parameters to be estimated are σ and γ as well as the production parameters. For the normal-truncated exponential model we report the estimates of parameters σ_u and σ_v themselves.

Tables 4 and 5 report results from the normal-truncated half normal model with a sample size of 200 and 1000, respectively. The results from these two tables differ only in quantitative manner. The first important conclusion that can be drawn is that the MLE estimators for technology parameters, α_1 and α_2, are accurate. As noise level increases, the MSE of these estimators only slightly increases. The second important observation is that the estimator for the inefficiency bound has small MSE when the noise level is mild. When noise level is high, as when $\sigma_v = 0.5$, $\hat{B}$ becomes inaccurate. In table 4 distribution parameters, $\hat{\sigma}$ and $\hat{\gamma}$ display a significantly upward bias and large MSE as the signal-to-noise ratio decreases[5]. Table 5 shows that the problem is alleviated somewhat when the sample size increases.

We now look at the doubly truncated normal model. Table 6 and 7 reports Monte Carlo results with a sample size of 200 and 1000, respectively. At both sample sizes, the technology parameter estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are quite accurate. In order to identify the distribution parameters we employ the restrictions that arise form the identification discussion of section 4. Now, the estimates of distribution parameters, σ

[4] The results does not change very much if we include the constant term. We ommit it to save space. The results with constant term are available upon the request.

[5] It can be shown that in this case the inverse Hessian becomes close to singular, which makes the estiamates of the model paramers less accurate. To our best knoweldge this pathology is shared by all likelihood based stochastic frontier models.

Table 4: Monte Carlo results for Truncated Half Normal model. The number of repetitions $M = 1000$. Sample size $N = 200$.

			$B = 0.8$		$B = 1.0$		$B = 1.2$	
		True	AVE	MSE	AVE	MSE	AVE	MSE
	$\hat{\sigma}$	0.3	0.3308	0.0022	0.3288	0.0013	0.3282	0.0011
	$\hat{\gamma}$	0.9	0.9079	0.0026	0.9101	0.0025	0.9107	0.0021
$\sigma_v = 0.1$	$\hat{B}$		0.7987	0.0148	0.9223	0.0309	0.9600	0.0923
	$\hat{\alpha}_1$	0.6	0.6004	0.0009	0.6008	0.0009	0.6003	0.0009
	$\hat{\alpha}_2$	0.5	0.5016	0.0008	0.5022	0.0007	0.5021	0.0007
	$\hat{\sigma}$	0.4	0.4967	0.1325	0.4464	0.0656	0.4466	0.0723
	$\hat{\gamma}$	0.7	0.7440	0.0451	0.7432	0.0354	0.7344	0.0412
$\sigma_v = 0.2$	$\hat{B}$		0.8429	0.0860	0.9585	0.0990	0.9790	0.1604
	$\hat{\alpha}_1$	0.6	0.6045	0.0023	0.6024	0.0021	0.6030	0.002
	$\hat{\alpha}_2$	0.5	0.5029	0.0021	0.5061	0.0020	0.5039	0.0021
	$\hat{\sigma}$	0.6	0.8399	0.3356	0.8538	0.3604	0.8662	0.3905
	$\hat{\gamma}$	0.3	0.4570	0.1525	0.4621	0.1636	0.4538	0.1616
$\sigma_v = 0.5$	$\hat{B}$		1.0780	0.6185	1.1966	0.6121	1.2083	0.5521
	$\hat{\alpha}_1$	0.6	0.6114	0.0100	0.6169	0.0108	0.6117	0.0112
	$\hat{\alpha}_2$	0.5	0.5202	0.0116	0.5176	0.0125	0.5210	0.0127

Table 5: Monte Carlo results for Truncated Half Normal model. The number of repetitions $M = 1000$. Sample size $N = 1000$.

			$B = 0.8$		$B = 1.0$		$B = 1.2$	
		True	AVE	MSE	AVE	MSE	AVE	MSE
	$\hat{\sigma}$	0.3	0.3191	0.0002	0.3188	0.0002	0.3191	0.0001
	$\hat{\gamma}$	0.9	0.9020	0.0005	0.9019	0.0004	0.9027	0.0004
$\sigma_v = 0.1$	$\hat{B}$		0.8045	0.0049	0.9889	0.0170	1.0918	0.0502
	$\hat{\alpha}_1$	0.6	0.6005	0.0002	0.5993	0.0002	0.6001	0.0002
	$\hat{\alpha}_2$	0.5	0.5001	0.0002	0.5013	0.0002	0.5004	0.0002
	$\hat{\sigma}$	0.4	0.3806	0.0042	0.3735	0.0010	0.3724	0.0009
	$\hat{\gamma}$	0.7	0.7111	0.0094	0.7156	0.0056	0.7125	0.0050
$\sigma_v = 0.2$	$\hat{B}$		0.8692	0.0589	1.0351	0.0808	1.1169	0.1044
	$\hat{\alpha}_1$	0.6	0.6010	0.0005	0.6027	0.0005	0.6020	0.000
	$\hat{\alpha}_2$	0.5	0.5023	0.0004	0.5021	0.0004	0.5021	0.0004
	$\hat{\sigma}$	0.6	0.6597	0.0407	0.6568	0.0355	0.6573	0.0373
	$\hat{\gamma}$	0.3	0.3580	0.0609	0.3568	0.0597	0.3565	0.0589
$\sigma_v = 0.5$	$\hat{B}$		0.9995	0.4273	1.1713	0.5255	1.2536	0.5442
	$\hat{\alpha}_1$	0.6	0.6020	0.0031	0.6028	0.0031	0.6042	0.0028
	$\hat{\alpha}_2$	0.5	0.5056	0.0028	0.5059	0.0029	0.5052	0.0026

Table 6: Monte Carlo results for Doubly Truncated Normal model. The number of repetitions $M = 1000$. Sample size $N = 200$.

			$B = 0.8$		$B = 1.0$		$B = 1.2$	
		True	AVE	MSE	AVE	MSE	AVE	MSE
	$\hat{\sigma}$	0.3	0.4141	0.0628	0.4227	0.0753	0.3899	0.0424
	$\hat{\gamma}$	0.9	0.9463	0.0072	0.9511	0.0085	0.9442	0.0091
$\sigma_v = 0.1$	$\hat{\mu}$	0.5	0.5894	0.0329	0.5367	0.0427	0.4758	0.0377
	$\hat{B}$		0.8452	0.0239	1.0207	0.0278	1.1797	0.0411
	$\hat{\alpha}_1$	0.6	0.6046	0.0020	0.6035	0.0024	0.5990	0.0028
	$\hat{\alpha}_2$	0.5	0.5084	0.0020	0.5032	0.0023	0.5006	0.0027
	$\hat{\sigma}$	0.4	0.4627	0.0658	0.5182	0.1026	0.4853	0.0759
	$\hat{\gamma}$	0.7	0.7937	0.0524	0.8300	0.0564	0.8173	0.0603
$\sigma_v = 0.2$	$\hat{\mu}$	0.5	0.6057	0.0627	0.5875	0.0923	0.5306	0.0958
	$\hat{B}$		0.9296	0.1035	1.0963	0.1205	1.2538	0.1421
	$\hat{\alpha}_1$	0.6	0.6122	0.0040	0.6123	0.0046	0.6093	0.005
	$\hat{\alpha}_2$	0.5	0.5189	0.0045	0.5079	0.0050	0.5076	0.0058
	$\hat{\sigma}$	0.6	0.7444	0.1064	0.7397	0.0973	0.7756	0.1238
	$\hat{\gamma}$	0.3	0.5174	0.1381	0.5338	0.1486	0.5542	0.1734
$\sigma_v = 0.5$	$\hat{\mu}$	0.5	0.4491	0.1187	0.5125	0.1635	0.5647	0.2265
	$\hat{B}$		1.1524	0.6325	1.3944	0.8184	1.5888	0.9906
	$\hat{\alpha}_1$	0.6	0.6155	0.0133	0.6179	0.0150	0.6205	0.0173
	$\hat{\alpha}_2$	0.5	0.5193	0.0156	0.5172	0.0157	0.5287	0.0189

and γ, are upward biased, especially when λ and N are relatively small. Their MSE is low for low levels of noise. In addition, parameter μ is accurately estimated, especially then the sample size is large. The inefficiency bound is significantly distorted when signal-to-noise ratio decreases. It is worth noting at this point that parameter μ is very hard to be identified in the standard normal-truncated normal model and then its empirical value is large enough, there is a problem of identification of the rest distributional parameters as well. This is avoided in the normal-truncated normal model. Finally, the case of $B = 0.8$ corresponds to the case of the "wrong skewness". Clearly there is no problem of estimation and identification of the model parameters for this particular case.

Tables 8 and 9 show the results for truncated exponential model with a sample size

Table 7: Monte Carlo results for Doubly Truncated Normal model. The number of repetitions $M = 1000$. Sample size $N = 1000$.

		True	$B = 0.8$		$B = 1.0$		$B = 1.2$	
			AVE	MSE	AVE	MSE	AVE	MSE
	$\hat{\sigma}$	0.3	0.3487	0.0108	0.3336	0.0058	0.3229	0.0010
	$\hat{\gamma}$	0.9	0.9155	0.0013	0.9142	0.0015	0.9125	0.0019
$\sigma_v = 0.1$	$\hat{\mu}$	0.5	0.5419	0.0088	0.5053	0.0035	0.4997	0.0044
	$\hat{B}$		0.8100	0.0056	1.0067	0.0094	1.2116	0.0157
	$\hat{\alpha}_1$	0.6	0.6025	0.0004	0.5995	0.0005	0.6014	0.0006
	$\hat{\alpha}_2$	0.5	0.5008	0.0004	0.5024	0.0005	0.5006	0.0006
	$\hat{\sigma}$	0.4	0.4305	0.0285	0.4128	0.0236	0.3874	0.0079
	$\hat{\gamma}$	0.7	0.7348	0.0208	0.7346	0.0181	0.7260	0.0139
$\sigma_v = 0.2$	$\hat{\mu}$	0.5	0.5735	0.0257	0.5298	0.0163	0.4977	0.0137
	$\hat{B}$		0.8268	0.0240	1.0441	0.0455	1.2619	0.1089
	$\hat{\alpha}_1$	0.6	0.6020	0.0008	0.6034	0.0010	0.6011	0.001
	$\hat{\alpha}_2$	0.5	0.5029	0.0008	0.5020	0.0010	0.5015	0.0012
	$\hat{\sigma}$	0.6	0.6780	0.0400	0.7115	0.0555	0.7100	0.0527
	$\hat{\gamma}$	0.3	0.4227	0.0721	0.4601	0.0872	0.4527	0.0884
$\sigma_v = 0.5$	$\hat{\mu}$	0.5	0.4771	0.0630	0.5361	0.0827	0.5253	0.0973
	$\hat{B}$		1.1325	0.7649	1.2912	0.6360	1.2649	0.6243
	$\hat{\alpha}_1$	0.6	0.6076	0.0032	0.6075	0.0032	0.6038	0.0033
	$\hat{\alpha}_2$	0.5	0.5076	0.0032	0.5086	0.0033	0.5069	0.0036

Table 8: Monte Carlo results for Truncated Exponential model. The number of repetitions $M = 1000$. Sample size $N = 200$.

		True	$B = 0.8$ AVE	$B = 0.8$ MSE	$B = 1.0$ AVE	$B = 1.0$ MSE	$B = 1.2$ AVE	$B = 1.2$ MSE
	$\hat{\sigma}_u$	0.3	0.3062	0.0017	0.3084	0.0011	0.3006	0.0006
	$\hat{\sigma}_v$	0.1	0.0986	0.0001	0.0992	0.0001	0.0985	0.0001
$\sigma_v = 0.1$	$\hat{B}$		0.7991	0.0018	0.9876	0.0027	1.1934	0.0053
	$\hat{\alpha}_1$	0.6	0.6005	0.0004	0.6038	0.0005	0.5968	0.0005
	$\hat{\alpha}_2$	0.5	0.4999	0.0003	0.4966	0.0003	0.5022	0.0004
	$\hat{\sigma}_u$	0.3	0.3334	0.0117	0.3147	0.0048	0.3133	0.0020
	$\hat{\sigma}_v$	0.2	0.1940	0.0004	0.1962	0.0003	0.1955	0.0003
$\sigma_v = 0.2$	$\hat{B}$		0.8191	0.0066	1.0150	0.0091	1.2008	0.0113
	$\hat{\alpha}_1$	0.6	0.6020	0.0014	0.6001	0.0008	0.6032	0.001
	$\hat{\alpha}_2$	0.5	0.5026	0.0009	0.5016	0.0007	0.5004	0.0008
	$\hat{\sigma}_u$	0.3	1.0081	7.0210	0.9838	4.8403	0.7934	1.5996
	$\hat{\sigma}_v$	0.5	0.5009	0.0210	0.4869	0.0037	0.4824	0.0033
$\sigma_v = 0.5$	$\hat{B}$		1.0335	0.3644	1.1115	0.3053	1.2878	0.3050
	$\hat{\alpha}_1$	0.6	0.5942	0.0108	0.6034	0.0058	0.6088	0.0060
	$\hat{\alpha}_2$	0.5	0.5166	0.0082	0.5025	0.0045	0.5266	0.0053

of 200 and 1000, respectively. As with the previous models, the technology parameter estimates, $\hat{\alpha}_1$ and $\hat{\alpha}_2$, are accurate. Parameters of one-sided error term are accurate when the noise level is mild. It can be seen from these tables how sensitive to the noise parameters $\hat{\sigma}_u$ and B are. The estimated values of these parameters are highly contaminated by the noise when this dominates the inefficiency term. As expected, the finite sample problem with $\hat{\sigma}_u$ and B is lessened when we have a larger sample size of 1000.

Table 9: Monte Carlo results for Truncated Exponential model. The number of repetitions $M = 1000$. Sample size $N = 1000$.

		True	$B = 0.8$		$B = 1.0$		$B = 1.2$	
			AVE	MSE	AVE	MSE	AVE	MSE
	$\hat{\sigma}_u$	0.3	0.3019	0.0008	0.3014	0.0004	0.3021	0.0003
	$\hat{\sigma}_v$	0.1	0.0992	0.0001	0.0990	0.0001	0.0988	0.0001
$\sigma_v = 0.1$	$\hat{B}$		0.7992	0.0009	0.9972	0.0015	1.1975	0.0024
	$\hat{\alpha}_1$	0.6	0.5995	0.0002	0.6001	0.0002	0.6001	0.0003
	$\hat{\alpha}_2$	0.5	0.5005	0.0002	0.5003	0.0002	0.5004	0.0002
	$\hat{\sigma}_u$	0.3	0.3186	0.0069	0.3112	0.0022	0.3074	0.0011
	$\hat{\sigma}_v$	0.2	0.1983	0.0002	0.1984	0.0002	0.1981	0.0001
$\sigma_v = 0.2$	$\hat{B}$		0.8091	0.0047	1.0038	0.0060	1.1988	0.0088
	$\hat{\alpha}_1$	0.6	0.6002	0.0005	0.6014	0.0005	0.6001	0.001
	$\hat{\alpha}_2$	0.5	0.5010	0.0004	0.5004	0.0004	0.5014	0.0004
	$\hat{\sigma}_u$	0.3	0.6274	1.2594	0.5654	0.7008	0.4593	0.3063
	$\hat{\sigma}_v$	0.5	0.5044	0.0099	0.4963	0.0034	0.5005	0.0144
$\sigma_v = 0.5$	$\hat{B}$		0.9446	0.4032	1.1498	0.3430	1.2166	0.3058
	$\hat{\alpha}_1$	0.6	0.5934	0.0055	0.6028	0.0043	0.6000	0.0041
	$\hat{\alpha}_2$	0.5	0.4996	0.0048	0.5032	0.0035	0.5047	0.0038

6 Panel Data

In the same spirit as Schmidt and Sickles (1984) and Cornwell, et al. (1990), we may specify a panel data model of bounded inefficiencies:

$$y_{it} = \alpha_0 + \sum_{k=1}^{K} \alpha_k x_{it,k} + \varepsilon_{it} \tag{15}$$

where

$$\varepsilon_{it} = v_{it} - u_{it}. \tag{16}$$

We assume that the inefficiency components (u_{it}) are positive, independent from the regressors, and are independently drawn from a time-varying distribution with upper bound B_t. We may set B_t to be time-invariant. However, it is certainly more plausible to assume otherwise, as the market or industry may well become more or less forgiving as time goes by, especially in settings in which market reforms are being introduced or firms are adjusting to a phased transition from regulation to deregulation.

Note that since u_{it} is time-varying, the above panel data model is in effect a time-varying technical efficiency model. Our model differs from the existing literature in that, while previous time-varying efficiency models, notably Cornwell, Schmidt, and Sickles (1990), Kumbhakar (1990), Battese and Coelli (1992), and Lee and Schmidt (1993), are time-varying in the mean or intercept of individual effects, our model is time-varying in the upper support of the distribution of inefficiency term u_i.

The assumption that u_{it} is independent over time simplifies estimation and analysis considerably. In particular, the covariance matrix of $\varepsilon_i \equiv (\varepsilon_{i1}, ..., \varepsilon_{iT})'$ is diagonal. This enables us to treat the panel model as a collection of cross-section models in the chronological order. We may certainly impose more structure on the sample path of the upper bound of u_{it}, B_t, without incurring heavy costs in terms of analytic difficulty. For example, we may impose smoothness conditions on B_t. This is empirically plausible, indeed, since changes in the market competitive conditions may come gradually. And it is also technically desirable, since imposing smoothness conditions gives us more degree of freedom in estimation, hence better estimators of model parameters. A natural way of doing this is to let B_t be a sum of weighted polynomials,

$$B_t = \sum_{i=0}^{K} b_i (t/T)^i, \quad t = 1, ..., T, \tag{17}$$

where (b_i) are constants. We may also use trigonometric series, splines, among others, in the modeling of B_t.

7 Efficiency Analysis of Banking Industry

7.1 Empirical Model and Data

We now apply the bounded inefficiency (BIE) model to an analysis of the US banking industry, which underwent a series of deregulatory reforms in the early 1980's and 1990's[6], and experienced an adverse economic environment in the last few turbulent years of the last decade. Our analysis covers an extensive period between 1984 and 2009. What is generally observed during this period is that the number of commercial banks has substantially decreased through either mergers or failures. It is characteristic the fact that the number of failed banks in 2009 was about 2.75 times more than that of the period 2001-2008. As of the present times, the number of mergers and new charters has decreased, while the proportion of problematic banks has dramatically increased. However, the biggest failures occurred in early 1990's where almost 4000 banks failed within three years (source FDIC). All these facts have triggered the interest of researchers to analyze the U.S. commercial banking industry more closely and especially the performance of its institutions and their market behavior. The primary aim of our model is to capture the efficiency trends of the U. S. banking sector during all of these years until the present time, as well as to identify the toughness of the market against very inefficient firms.

Here we extend our model to the panel setting and, following Adams, Berger, and Sickles (1999) and Kneip, Sickles, and Song(2005), we specify a multiple output and input Cobb-Douglas stochastic output distance frontier model as follows[7],

$$Y_{it} = Y_{it}^{*\prime}\gamma + X_{it}'\beta + v_{it} - u_{it}, \tag{18}$$

where Y_{it} is the log of real estate loans ; X_{it} is the negative of log of inputs, which include demand deposit (DD), time and savings deposit (DEP), labor (lab), capital (cap), and purchased funds (purf)[8]; and Y_{it}^* includes the log of commercial and industrial loans/real estate loans (ciln) and installment loans/real estate loans (inln). All nominal values are converted to reflect 2000 year values. We assume (v_{it}) are iid across i and t, and for each t, u_{it} has a upper bound B_t. Then we can treat this model as a generic panel data bounded inefficiency model as discussed in Section 6. Once the individual effects u_{it} are estimated, technical efficiency for a particular firm at time t is calculated as $TE = \exp(u_{it} - \max_{1\leq j\leq N} u_{jt})$.

[6] These deregulations gradually allowed banks in different states to merge with other banks across the state borders. Reigle-Neal Act that was passed by the Congress in 1994 also allowed the branching by banks across the state lines.

[7] For more discussion on stochastic distance frontiers see Lovell, Richardson, Travers, and Wood (1994).

[8] Purchased funds include federal funds purchased and securities sold under agreements to repurchase, time deposits in $100K denominations, mortgage debt, bank's liability on acceptances, and other liabilities that are not demand deposits and retail time and saving deposits.

We use U.S. commercial banking data from 1984 through 2009[9] . The data is quarterly balanced panel of 4250 commercial banks and was recently compiled from the Report of Condition and Income (Call Report) and the FDIC Summary of Deposits. The data set includes 437750 observations for 103 quarterly periods. This is a fairly long panel and thus the assumption of time-invariant inefficiencies should be abandoned herein. For this reason we compare the estimates from BIE model to the estimates from other time-varying models such as CSSW (Cornwell, Schmidt, and Sickles, 1990) and BC (Battese and Coelli, 1992) models. Fixed effect estimator (FIX) of Schmidt and Sickles (1984) is also considered for illustration purposes.

7.2 Results

Table 10 compares the parameter estimates of the bounded inefficiency (BIE)[10] model with that of FIX, CSSW, and BC. Technology parameters are statistically significant at 1% confidence level and have the expected sign for all four models. The technology parameters from BIE model are somewhat different from those obtained from other models. However, except for the coefficients of *ciln* and *cap*, the rest coefficients are in accordance with those of FIX and BC model. The striking difference, however, is between distributional parameters of BIE and BC model. Parameter μ in BC model has very small and insignificant value relative to that of BIE model. This difference is mainly due to the fact that there is not enough information in normal-truncated normal model to identify this parameter. In normal-doubly truncated normal model we examine the sign of OLS residuals (which is negative here) and by employing the method of moments we find that $\mu > 0$ and $B > 2\mu$. This corresponds to the set I_5 and hence we can identify this parameter. Any value from this set can be served as a starting value for MLE to consistently estimate μ, fact that is also shown in our Monte Carlo results.

We also estimate time-varying inefficiency bound, B, using two approaches. First we estimate the bound for panel data model without imposing any restriction on its sample path. In the second approach we specify the bound as a sum of weighted time polynomials and since the data set is large enough there are a lot of degrees of freedom to allow us to fit a fifth degree polynomial. The coefficients of the polynomial are estimated via maximum likelihood method along with the rest parameters. Both approaches are illustrated in figure 2. It can be seen that the inefficiency bound has decreasing trend up to year 2005 then its increasing throughout. One interpretation of this trend can be that the deregulations in 1980's and 1990's increased the competition and forced many inefficient banks to exit. This obviously reduced the upper limit of inefficiency that bank can sustain in order to not be ruled out from the market. The

[9] Data includes up to third quarter of 2009.

[10] We estimate the normal-doubly truncated normal model in order to be able to compare it with the BC model which specifies the inefficiencies to follow the truncated normal distribution.

new upward trend can be attributed to adverse economic environment and to the fact that the proportion of banks which are characterized as "too big too fail " has increased.

Of course, for time-varying efficiency models such as CSSW,BC, and BIE, the average efficiency changes over time[11]. This is illustrated in Figure 1. The BIE average efficiency is significantly higher than what the fixed effect model obtains. However, the difference is small for BC and CSSW models. This difference is not unexpected, however, since the existence of inefficiency bound implies that the mean conditional distribution of inefficiencies is also bounded from above, resulting to higher average efficiency. Failing to take the bound into account can possibly yield underestimated mean and individual efficiency scores (see table 1). $BIE2$ curve represents the fitted time polynomial of inefficiencies obtained from the bounded inefficiency model. It can be seen that the efficiency trend for BIE model is most of the times in line with that of CSSW model. BC model shows an upward efficiency trend for all these periods ($\eta = 0.067$). We also look at the efficiency ranking of firms. Table 11 tabulates the Spearman rank correlations among different models. It is clear that the BIE efficiency ranking is in agreement with previous estimations, especially with CSSW. This comes to the argument of Ritter and Simar (1997), that if researcher believes that the inefficiency bound is relatively large and if there is no "wrong " skewness issue then simple densities, such as half-normal, can be utilized to estimate parametric stochastic frontier model.

In sum, figures 1 and 2 display an interesting findings: on one hand, an upward trend is observed for the average efficiency of the industry, presumably benefiting from the deregulations in the 1980's and 1990's; on the other hand, the industry appears to be more "tolerant"of less efficient banks in the last decade. Possibly, these banks have a characteristic that we have not properly controlled for and we are currently examining this issue. Given the recent experiences in the credit markets due in part to the poor oversight lending authorities gave in their mortgage and other lending activities, our results also may be indicative of a backsliding in the toleration of inefficiency that could have contributed to problems the financial services industry faces today.

[11] We trimmed out the top and bottom 5% of inefficiencies to remove the effects of outliers.

Table 10: Comparisons of Various Estimators. Estimates and standard errors (in parentheses) for each model parameters from competing models (FIX, CSSW, BC, BIE).

	FIX	$CSSW$	BC	BIE
ciln	0.1915(0.0018)	0.1668(0.0016)	0.1809(0.0017)	0.2589(0.0018)
inln	0.2639(0.0018)	0.2822(0.0019)	0.2547(0.0017)	0.2905(0.0015)
DD	−0.0791(0.0032)	−0.0835(0.0029)	−0.0615(0.0031)	−0.0873(0.0026)
DEP	−0.5064(0.0058)	−0.4402(0.0062)	−0.5745(0.0057)	−0.5757(0.0031)
lab	−0.2605(0.0059)	−0.2486(0.0054)	−0.2053(0.0056)	−0.1904(0.0025)
cap	−0.0461(0.0020)	−0.0490(0.0019)	−0.0450(0.0019)	−0.0888(0.0019)
purf	−0.1108(0.0040)	−0.1616(0.0036)	−0.1345(0.0040)	−0.1279(0.0032)
time	0.0052(0.0001)	–	0.0023(0.0001)	–
γ	0	0	0.7947(0.0067)	0.7690(0.0058)
σ	0.243(0.0037)	0.214(0.0023)	0.2355(0.0015)	0.2733(0.0021)
μ	–	–	0.0127(0.0032)	0.4937(0.069)
B	–	–	–	1.5343
ATE	0.5614	0.6693	0.6574	0.6815

Table 11: Spearman Rank Correlations of Efficiencies

	FIX	CSSW	BC	BIE
FIX	1	·	·	·
CSSW	0.8556	1	·	·
BC	0.9662	0.8231	1	·
BIE	0.6919	0.7942	0.7168	1

8 Conclusions

In this paper we have introduced a series of parametric stochastic frontier models that have upper (lower) bounds on the inefficiency (efficiency). The model parameters can be estimated by maximum likelihood, including the inefficiency bound. The models are easily applicable for both cross-section and panel data setting. In the panel data setting, we set the inefficiency bound to be varying over time, hence contributing another time-varying efficiency model to the literature. We have examined the finite sample performance of the maximum likelihood estimator in the cross-sectional setting. We also have showed how the "wrong "skewness problem inherent in traditional stochastic frontier model can be avoided when the bound is taken into account. An empirical analysis of US banking industry using the new model revealed interesting trends in efficiency scores. We concluded with an empirical puzzle that we leave for the future analysis.

References

Adams, R.M., Berger, A.N., Sickles, R.C., 1999, Semiparametric approaches to stochastic panel frontiers with applications in the banking industry. Journal of Business and Economic Statistics 17, 349-358.

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

Battese, G.E., Coelli, T.J. 1992, Frontier production functions, technical efficiency and panel data, with application to paddy farmers in India. Journal of Productivity Analysis 3, 153-169.

Battese, G. E., Corra, G. 1977, Estimation of a production frontier model: with application to the pastoral zone of eastern Australia, Australian Journal of Agricultural Economics 21, 167-179.

Carree, M. A., 2002, Technological inefficiency and the skewness of the error component in stochastic frontier analysis, Economics Letters 77, 101-107.

Cornwell, C., Schmidt, P., Sickles, R.C., 1990, Production frontiers with cross-sectional and time series variation in efficiency levels. Journal of Econometrics 46, 185-200.

Entani, T., Maeda Y., Tanaka H., 2002, Dual models of interval DEA and its extension to interval data, European Journal of Operational Research 136 32-45.

Green, A., Mayes, D., 1991, Technical inefficiency in manufacturing industries, Economic Journal, 101, 523-538.

Greene, W.H., 1980a, Maximum likelihood estimation of econometric frontier functions. Journal of Econometrics 13, 27-56.

Greene, W.H., 1980b, On the estimation of a flexible frontier production model. Journal of Econometrics 13, 101-115.

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

Jondrow, J., Lovell, C.A.K., Materov I.S., Schmidt, P., 1982, On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19, 233-238

Kneip, A., Sickles, R.C., Song W., 2005, A new panel data treatment for heterogeneity in time trends. Mimeo, Rice University.

Kumbhakar, S.C., 1990, Production frontiers, panel data, and time-varying technical efficiency. Journal of Econometrics 46, 201-212.

Lee, Y.H., Schmidt, P., 1993, A production frontier model with flexible temporal variation in technical efficiency. In: Fried, H.O, Lovell, C.A.K., Schmidt, P. (Ed.), The measurement of productive efficiency: Techniques and Applications, Oxford University Press.

Lovell, C.A.K., Richardson, S., Travers, P., Wood, L., 1994, Resources and Functionings: A new view of inequality in Australia. Models and Measurements of Welfare and Inequality, W. Eichhorn, Berlin: Springer-Verlag.

Meeusen, W., van den Broeck, J., 1977, Efficiency estimation from Cobb-Douglas production function with composed error. International Economic Review 18, 435-444.

Olson, J.A., P. Schmidt, D.M. Waldman, 1980, A Monte Carlo study of estimators of the stochastic frontier production function. Journal of Econometrics 13, 67-82.

Ritter, C., and Simar, L., 1997, Pitfalls of Normal-Gamma Stochastic Frontier Models. Journal of Productivity Analysis, 8, 167–182.

Schmidt, P., Sickles, R.C., 1984, Production frontiers and panel data. Journal of Business and Economic Statistics 2, 367-374.

Simar and Wilson, 2009, Estimation and Inference in Cross-Sectional,Stochastic Frontier Models, forthcoming in Econometric Reviews.

Stevenson, R.E., 1980, Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13, 57-66.

Figure 1: Averaged efficiencies from each estimator

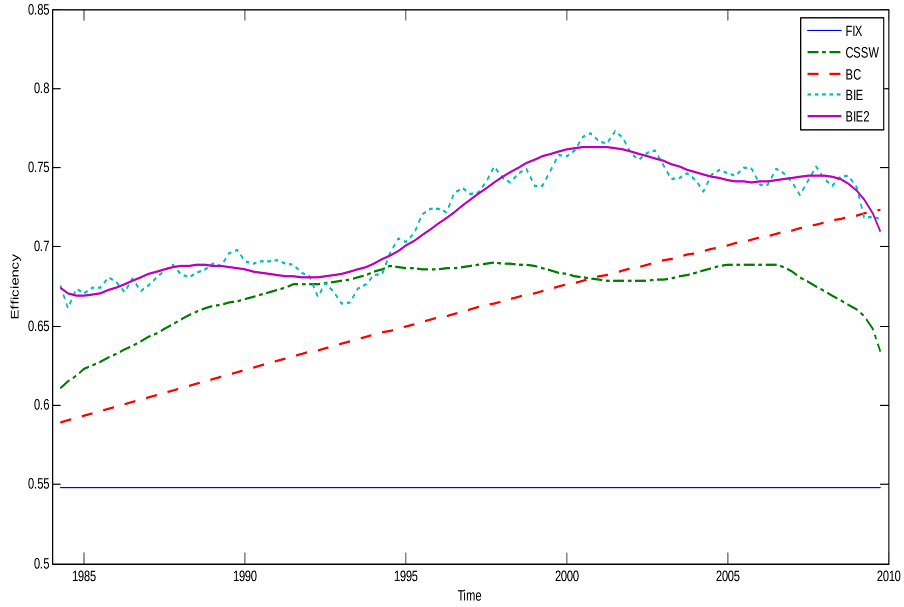

Figure 2: Estimated time-varying inefficiency bound