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

This version: January 31, 2011

[1] Earlier versions of this paper were presented at the Tenth European Workshop on Efficiency and Productivity, Lille, France, June, 2007, the Annual Texas Econometrics Camp XV, Montgomery, Texas, February, 2010, the Efficiency and Productivity Workshop, University of Auckland, New Zealand, February, 2010, the North American Productivity Workshop, Houston, Texas, June 2010, the International Econometrics Workshop, Guanghua Campus (SWUFE), Chengdu, August 12, 2010, and the 10th World Congress of the Econometric Society, Shanghai, August 21, 2010. We would like to thank participants of those conferences and workshops for their helpful comments and insights. We thank Carlos Martins-Filho for his helpful suggestions and criticism. The usual caveat applies. We would especially like to thank Robert Adams at the Board of Governors of the Federal Reserve System for his guidance and Rob Kuvinka for his excellent research assistance that was essential in our development of the banking data set that we analyze in our empirical illustration.

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

Abstract

This paper introduces a new model of the 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 our 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 potential 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). This stochastic frontier production function has become an iconic modeling pardigm in econometric research, rate making decisions in regulated industries across the world, in evaluating oucomes of market reforms in transition economies, and in establishing performance benchmarks for local, state, and federal governmental activities.

In this paper we propose a new class of parametric stochastic frontier models with a more flexible specification of the inefficiency term, which we view as improvement on the basic iconic stochastic frontier production model. 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. First, it is plausible to allow only bounded support in many applications of stochastic frontier models wherein the extremely inefficient firms in a competitive industry of market are eliminated by competition. Bounded inefficiency makes sense in this setting since the extremely inefficient stores will be forced to close and thus individual production units constitute a truncated sample.[2] This is consistent with the arguments of Alchian (1950) and Stigler (1958) wherein firms are at any point in time not in a static long run equilibirum, but rather are tending to that situation as they are buffeted by demand and cost shocks. As a consequence, even if we correctly specify a family of distributions for the inefficiency term, the stochastic frontier model may still be misspecified. This particular setting is one in which the inefficiency bound is informative as an indicator of competitive pressures and/or the extent of supervisory oversight by direct managment or by corporate boards. In settings in which firms can successfully differentiate their product, which is the typical market structure and not the exception, or where there are market concentrations that may reflect collusive behavior or conditions for a natural monopoly

[2] In addition, the frequent use of balanced panels in empirical studies would in effect eliminate those failing firms from the sample and thus would provide more merit to the bounded inefficiency model.

and regulatory oversight, incentives to fully exploit market power or to instead make satisficing decision are both possible outcomes. Much more likely is that it is not one or the other but some middle ground between the two extremes that would be found empirically.[3]

A second justification for our introduction of the inefficiency bound into the classical stochastic production frontier model is that our model points to an explanation for the finding of "wrong "skewness in many applied studies using the traditional stochastic frontier, and thus to the potential of 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. One such specification in which this is a natural consequence is when 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. 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 in any finite sample 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.

In addition to proposing new parametric forms for the classical stochastic production frontier model, we also show that our models are identifiable, and in which cases the identification is local or global. Initial consistent estimates are based on method of moments estimates, based on explicit analytic expressions which we derive, and which either can be used in a two-step method of scoring or as starting values in solving the normal equations for the relevant sample likelihood, based on

[3] "The quiet life hypothesis" (QLH) by Hicks (1935) argues that, due to management's subjective cost of reaching the optimal profits, firms use their market power to allow inefficient allocation of resources. Increasing competitive pressure is likely to force management to work harder to reach optimal profits. Another hypothesis that relates market power and efficiency is "the efficient structure hypothesis" (ESH) by Demsetz (1973). ESH argues that firms with superior efficiencies or technologies have lower costs and therefore higher profits. These firms are assumed to gain larger market shares which lead to higher concentration. Recently Kutlu and Sickles (2010) have constructed a model in which the dynamic game is played out and have tested for the alternative outcomes, finding support for the QLH in certain airlines city-pair markets and the ESH in others.

the parametric density functions whose expressions we also provide. As the regulatory conditions for maximum likelihood estimation method are satisfied, we employ it in order to obtain consistent and asymptotically efficient estimates of the model parameters, including this of the inefficiency bound. We conduct Monte Carlo experiments to study the finite sample behaviour of our estimators. We also extend the model to the panel data setting and allow for a time-varying inefficiency bound. By allowing the inefficiency bound to be time-varying, we contribute another time-varying technical efficiency model to the efficiency literature. Our model differs from those most commonly used in the literature, e.g., Cornwell, Schmidt, and Sickles (1990), Kumbhakar (1990), Battese and Coelli (1992), and Lee and Schmidt (1993) in that, while previous time-varying efficiency models 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. This explicitly allows for the level of competitive pressures on firms and other factors that may force the firms to exit the industry to change over time as demand and cost conditions change.

The outline of this paper is as follows. In Section 2 we present the new models and derive analytic formula for density functions and expressions that allow us to evaluate inefficiencies. Section 3 deals with the "wrong" skewness issue inherent in the 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 the 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 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 (nonnegative) 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. Clearly we can consider other more flexible functional forms for production (or cost) that are linear or linear in logarithms, such as the generalized Leontief or the transcendental logarithmic, or ones that are nonlinear. The only necessary asumption is that the error process ε_i is additively separable from the functional forms we employ in the stochastic production (cost) frontier.

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 the case of the cost frontier $\varepsilon_i = v_i + u_i$). In particular, we assume that u_i is distributed as doubly truncated normal, the density of which is given by

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

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.[4] In the banking industry, which we examine in section 7, the inefficiency bound may also represent factors that influence the financial health of the industry. It may be natural to extend this specification and treat the bound as a function of individual specific covariates z_i, such as $\exp(\delta' z_i)$, which would allow identification of bank-specific measures of financial health.

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 rather flexible. It nests the 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 truncation 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, " while in the truncated half normal model and the truncated exponential model (in which the distribution of u_i is truncated exponential), most firms are implicitly assumed to be relatively efficient.[5]

[4]The inefficiency bound has a natural role in gauging the tolerance for or ruthlessness against 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).

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

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,

$$f(u) = \frac{1}{\sigma_u(1 - e^{-B/\sigma_u})} e^{-\frac{u}{\sigma_u}} \mathbf{1}_{[0,B]}(u). \tag{3}$$

The truncated exponential distribution can be further generalized to the truncated gamma distribution, which shares the nice property with the doubly truncated normal distribution 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 $\gamma-$parametrization, which specifies

$$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}, \quad \gamma = \sigma_u^2/\sigma^2. \tag{4}$$

By definition $\gamma \in [0, 1]$, a compact support, which is desirable for the numerical procedure of maximum likelihood estimation. Another parametrization initially employed by Aigner et al. (1977) is the $\lambda-$parametrization

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

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).

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 ordinary least squares (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

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	$\frac{e^{\frac{\varepsilon}{\sigma_u}+\frac{\sigma_v^2}{2\sigma_u^2}}[\Phi(\frac{B+\varepsilon}{\sigma_v}+\frac{\sigma_v}{\sigma_u})-\Phi(\frac{\varepsilon}{\sigma_v}+\frac{\sigma_v}{\sigma_u})]}{\sigma_u(1-e^{-B/\sigma_u})}$	$\mu_* + \sigma_v \frac{\phi(-\frac{\mu_*}{\sigma_v})-\phi(\frac{B-\mu_*}{\sigma_v})}{\Phi(\frac{B-\mu_*}{\sigma_v})-\Phi(-\frac{\mu_*}{\sigma_v})}$	$\frac{\frac{1}{\sigma_v}\phi(\frac{u-\mu_*}{\sigma_v})}{\Phi(\frac{B-\mu_*}{\sigma_v})-\Phi(-\frac{\mu_*}{\sigma_v})}\mathbf{1}_{[0,B]}(u)$
		$\mu_* = -\varepsilon - \frac{\sigma_v^2}{\sigma_u}$	

term as a measure of technological inefficiency. It may imply that a nonrepresentative random sample had been drawn from an inefficiency distribution possessing the correct population skewness (see Carree, 2002; Greene, 2007; Almanidis and Sickles, 2009; Simar and Wilson, 2010). This is considered a finite sample "artifact" and the usual suggestion in the literature and by programs implementing stochastic frontier models is to treat all firms in the sample as fully efficient and proceed with straightforward ols based on the results of Olson et al. (1980) and Waldman (1982). As this would suggest setting the variance of the inefficiency term to zero, it would have problematic impacts on estimation and on inference. Simar and Wilson (2010) suggest a bagging method to overcome the inferential problems when a half-normal distribution for inefficiencies is specified. However, a finding of positive skewness in a sample may also indicate that inefficiencies are in fact drawn from a distribution which has positive skewness. 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 wherein continuous inefficiencies fall into discrete "inefficiency categories" 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 in the spirit of Carree (2002), 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

$$S_u = \frac{2\eta_0^3 - \eta_0(3\eta_1 + 1) + \eta_2}{\left(1 - \eta_0^2 + \eta_1\right)^{3/2}}. \tag{6}$$

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 this ambiguity theoretically could explain the prevalence of the “wrong” skewness problem in applied stochastic frontier research. Here, the term wrong is set in quotes to point out that the conventional wisdom that positive skewness is inconsistent with the standard stochastic frontier production model errors skewness is not necessarily the correct wisdom. When the underlying data generating process for u_i is based on the 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 also may be a large sample problem.

In finite samples, we may use simulations to show that our model is capable of generating residuals with "wrong" skewness with higher frequency than do traditional stochastic frontier models (Simar and Wilson, 2010). We generate samples of the residuals $\varepsilon = v - u$ with u being doubly truncated normal. We then 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 with 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 (2010) 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. In addition, although the truncated half normal and the truncated exponential models have correct (negative) skewness in the limit, the existence of the inefficiency bound reduces the identifiability of negative skewness in finite sample,

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

often to the extent that "wrong skewness" appears. 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 in which firms do not fall below some minimal level of efficiency in order to remain in the market or industry. Hence the traditional unbounded support for the inefficiency term would be misspecified and should be substituted with the the model of bounded inefficiency.

4 Estimation

4.1 Identification

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. For models without an intercept term the identification conditions for the first part are well known and are satisfied in most of the cases. The structural parameters can be consistently obtained by applying straightforward OLS. However, for models containing an intercept term there is a need to bias correction it using the distributional parameters since $E[\varepsilon] = -E[u] \neq 0$ (see Afriat, 1972 and Richmond, 1974). Therefore, the identification of the second part, which is based on method-of-moments requires a closer examination. Table 3 lists 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.

To examine the identification of the second part we note that under the assumption of independence of the noise and inefficiency term the following equality holds

$$E\left[(\varepsilon - E(\varepsilon))^4\right] - 3\left(E\left[(\varepsilon - E(\varepsilon))^2\right]\right)^2 = \psi_4 - 3\psi_2^2 = E\left[(u - E(u))^4\right] - 3\left(E\left[(u - E(u))^2\right]\right)^2$$

This is a measure of excess kurtosis and for the truncated half-normal model is derived as

$$\psi_4 - 3\psi_2^2 = \sigma_u^4(-\xi^3\tilde{\eta}_0 + 3\xi\tilde{\eta}_0 - 4\xi^2\tilde{\eta}_0^2 - 4\tilde{\eta}_0^2 - 3\xi^2\tilde{\eta}_0^2 - 12\xi\tilde{\eta}_0^3) \qquad (7)$$

where $\tilde{\eta}_0 = \frac{(2\pi)^{-1/2} - \xi\phi(\xi)}{\Phi(\xi) - \frac{1}{2}}$. Notice that for normal distribution $\tilde{\eta}_0 = 0$ and thus the excess kurtosis is also zero.

After multiplying (7) by $\psi_3^{-4/3}$ we eliminate σ_u and the resulting function, which we denote by g has only one argument ξ

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	$\sigma_u^4 \frac{-9e^{4\kappa}+36e^{3\kappa}-54e^{2\kappa}+36e^{\kappa}-9+6\kappa^2e^{\kappa}(e^{2\kappa}-2e^{\kappa}+1)+\kappa^4e^{\kappa}(e^{2\kappa}+e^{\kappa}+1)}{-e^{4\kappa}+4e^{3\kappa}-6e^{2\kappa}+4e^{\kappa}-1} +$
	$6\sigma_v^2\sigma_u^2 \frac{e^{2\kappa} - \left(\kappa^2+2\right)e^{\kappa}+1}{e^{2\kappa}-2e^{\kappa}+1} + 3\sigma_v^4, \quad \kappa = B/\sigma_u.$

$$g(\xi) = \frac{-\xi^3\tilde{\eta}_0 + 3\xi\tilde{\eta}_0 - 4\xi^2\tilde{\eta}_0^2 - 4\tilde{\eta}_0^2 - 3\xi^2\tilde{\eta}_0^2 - 12\xi\tilde{\eta}_0^3}{\left(2\tilde{\eta}_0^3 - 3\xi\tilde{\eta}_0^2 - \tilde{\eta}_0 + \xi^2\tilde{\eta}_0\right)^{-4/3}} \tag{8}$$

The weak law of large numbers implies that

$$\text{plim}\frac{1}{n}\sum\nolimits_i \hat{\varepsilon}_i^k = m_k = \psi_k$$

where m_k denotes the k^{th} central sample moments of the least squares residuals $\hat{\varepsilon}$. The first order moment is zero by definition and thus is not useful for identification purposes. By employing the Slutsky theorem we can specify the following function G

$$\begin{aligned} g(\xi) &= \frac{m_4 - 3m_2^2}{m^{4/3}} \\ &\Longrightarrow \\ G(\xi) &= g(\xi) - \frac{m_4 - 3m_2^2}{m^{4/3}} \end{aligned}$$

Similarly, we can derive the function G for the normal-truncated exponential model with function g expressed by

$$g(\xi)=\frac{36e^{2\xi}-24e^{\xi}-24e^{3\xi}+6e^{4\xi}-\xi^4e^{\xi}-4\xi^4e^{2\xi}-\xi^4e^{3\xi}+6}{(6e^{2\xi}-4e^{\xi}-4e^{3\xi}+e^{4\xi}+1)(-\frac{2e^{3\xi}-\left(\xi^3+6\right)e^{2\xi}+\left(6-\xi^3\right)e^{\xi}-2}{e^{3\xi}-3e^{2\xi}+3e^{\xi}-1})^{4/3}} \tag{9}$$

Both the truncated half normal model and the truncated exponential model are globally identified. To see this, we can examine the monotonicity of the function G with respect to the parameter ξ which will allow us to express this parameter (implicitly) as a function of sample moments and data. This condition provides the necessary and sufficient condition for global identification ala Rothenberg (1971). For the truncated half normal model, G is monotonically decreasing and for the truncated exponential model, G is monotonically increasing. Hence, in both cases, G is invertible and ξ can be identified. The identification of other parameters then follows from the third order moment of least squares residuals. Note, however, that for large values of ξ (e.g., $\xi > 5$ for the normal-truncated half-normal model and $\xi > 20$ for the normal-truncated exponential model), the curve $g(\xi)$ is nearly flat and gives poor identification. ξ can be large for two reasons: either σ_u goes to zero or the bound parameter is large. In the first case the distribution of the inefficiency process approaches the Dirac-delta distribution which makes it very hard for the distributional parameters to be identified. This limiting case is discussed in Wang and Schmidt (2008). In the second case the distribution of the inefficiency term becomes unbounded as in the standard stochastic frontier models for which it is straightforward to show that the model is globally identified (see Aigner et al., 1977 and Olson et al., 1980).

It is not clear, however, that the doubly truncated normal model is 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

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

If H is of full rank, 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, and I_4, the definitions of which are given as follows,

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

(ii) $I_2 \equiv \{(\mu, B) | \mu > 0, B \in (0, 2\mu)\}$

(iii) $I_3 \equiv \{(\mu, B) | B = 2\mu > 0\}$

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

The line $I_3 \equiv \{(\mu, B) | B = 2\mu > 0\}$ corresponds to the case where $B = 2\mu$ and $\psi_3 = 0$, hence the functions g_1 and g_2 are not continuous and the implicit function theorem is not applicable. 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. This may indicate that the restricted ($B = 2\mu$) model may be nested in the unrestricted model and the model is locally identifiable on $I_2 \bigcup I_3 \bigcup I_4$.

We may treat the doubly truncated normal model 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. For example, on the line of $\{(\mu, B) | \mu = 0, B > 0\} \subset I_1$, the doubly truncated normal model reduces to the truncated half normal model. As another useful example, the line I_2 corresponds to a sub-model that has positive skewness even asymptotically.

4.2 Method of Moment Estimation

The method-of-moments (Olson et al., 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 $\varepsilon_i^* = \varepsilon_i + (\mathbb{E}u_i)$ has zero mean and constant variance σ_ε^2. Hence OLS yields consistent estimates for ε_i^* and α_k, $k = 1, ..., K$. Equating the sample moments of estimated residuals ($\hat{\varepsilon}_i^*$) to the population moments, one can solve for the parameters associated with the distribution of (ε_i^*).

4.3 Maximum Likelihood Estimation

For more efficient estimation, we may use maximum likelihood estimation (MLE). Note that with the presence of a noise term v_i, the range of residual is unbounded and does not depend on the parameter. No other standard regularity conditions might be questioned. In the remainder of this section we provide the log-likelihood functions for the bounded inefficiency model for the three parametric distributions we have considered. 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. These are provided in the appendix.

In addition to the $\gamma-$parametrization discussed earlier, we re-parametrize the bound parameter with another parameter $\tilde{B} = \exp(-B)$. Unlike the bound, $\tilde{B}$ takes values in compact unit interval which facilitates the numerical procedure of maximum likelihood estimation as well as establishing the asymptotic normality of this parameter. When $\tilde{B}$ lies in the interior of parameter space, the MLE estimator is asymptotically normal (see Rao, 1973 and Davidson and MacKinnon, 1993 among others).

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

$$\begin{aligned}\ln L &= -n\ln\left[\Phi(\frac{-\ln\tilde{B}-\mu}{\sigma_u(\sigma,\gamma)}) - \Phi(\frac{-\mu}{\sigma_u(\sigma,\gamma)})\right] - n\ln\sigma - \frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n}\frac{(\varepsilon_i+\mu)^2}{2\sigma^2} \\ &+ \sum_{i=1}^{n}\ln\{\Phi(\frac{(-\ln\tilde{B}+\varepsilon_i)\sqrt{\gamma/(1-\gamma)} - (\ln\tilde{B}+\mu)\sqrt{(1-\gamma)/\gamma}}{\sigma} \\ &-\Phi(\frac{\varepsilon_i\sqrt{\gamma/(1-\gamma)} - \mu\sqrt{(1-\gamma)/\gamma}}{\sigma})\}, \qquad (10)\end{aligned}$$

where $\varepsilon_i = y_i - x_i\alpha$, $x_i = (1, x_{ik})$, and $\alpha = (\alpha_0, \alpha_k)'$.

$$\sigma_u(\sigma,\gamma) = \sigma\sqrt{\gamma}. \qquad (11)$$

This can be expressed in terms of the $\lambda-$parametrization as in Aigner et al., (1977) by substituting γ in (10) with

$$\gamma(\lambda) = \frac{\lambda^2}{1+\lambda^2}. \qquad (12)$$

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

$$\begin{aligned}\ln L &= -n\ln(\Phi(\frac{-\ln\tilde{B}}{\sigma_u(\sigma,\gamma)}) - \frac{1}{2}) - n\ln\sigma - \frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n}\frac{\varepsilon_i^2}{2\sigma^2} \qquad (13) \\ &+ \sum_{i=1}^{n}\ln\{\Phi(\frac{(-\ln\tilde{B}+\varepsilon_i)\sqrt{\gamma/(1-\gamma)} - \ln\tilde{B}\sqrt{(1-\gamma)/\gamma}}{\sigma}) - \Phi(\frac{\varepsilon_i\sqrt{\gamma/(1-\gamma)}}{\sigma})\},\end{aligned}$$

Again, substituting γ into (13) with $\gamma(\lambda)$ in (12), we get the log L with $\lambda-$parametrization.

Finally, the log-likelihood function for the truncated exponential model with $\gamma-$parametrization is given by

$$\begin{aligned}\ln L &= -\frac{n}{2}\ln\gamma - n\ln\sigma - n\ln(1-e^{\frac{\ln\tilde{B}\gamma^{-1/2}}{\sigma}}) + \frac{n}{2}\frac{1-\gamma}{\gamma} + \frac{\gamma^{-1/2}}{\sigma}\sum_{i=1}^{n}\varepsilon_i \qquad (14)\\ &+ \sum_{i=1}^{n}\ln[\Phi(\frac{(-\ln\tilde{B}+\varepsilon_i)(1-\gamma)^{-1/2}}{\sigma} + \sqrt{\frac{1-\gamma}{\gamma}}) - \Phi(\frac{\varepsilon_i(1-\gamma)^{-1/2}}{\sigma} + \sqrt{\frac{1-\gamma}{\gamma}})],\end{aligned}$$

where $\varepsilon_i = y - x_i\alpha$

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

$$\hat{\varepsilon}_i = y_i - \hat{\alpha}_0 - \sum x_{i,k}\hat{\alpha}_k, i = 1, \cdots, n. \qquad (15)$$

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

One reasonable question is whether or not one can test for the absence or the presence of the bound ($H_0 : \tilde{B} = 0$ vs. $H_1 : \tilde{B} > 0$), which one may wish to test since this would suggest that the proper specification would be the standard SF model which assumes no bound as a special case of our more general bounded SF model. The test procedure is slightly complicated but still feasible. The first complication arises from the fact that $\tilde{B}$ lies on the boundary of the parameter space under the null. Second, it is obvious from the log-likelihood functions provided above that the bound is not identified in this case and it can be shown that any finite order derivative of the log-likelihood function with respect to $\tilde{B}$ is zero. Thus the conventional Wald and Lagrange Multiplier (LM) statistics are not defined and the Likelihood Ratio (LR) statistic has a nonstandard asymptotic distribution that strictly would dominate the $\chi^2_{(1)}$ distribution. Lee (1993) derives the asymptotic distribution of such an estimate as a mixture of χ^2 distributions under the null that its value is zero, focusing in particular on the SF model under the assumption of half-normally distributed inefficiencies. Here λ is globally identified, which can also be seen using the method-of-moments estimator provided in Aigner et al. (1977). Lee (1993) provides useful one-to-one re-parametrizations which transform the singular information matrix into a nonsingular one. However, since the bound in our model case is not identified in this situation, there is no such re-parametrization and hence this procedure cannot be used. An alternative is to apply the bootstrap procedure proposed by Hansen (1996, 1999) to construct asymptotically equivalent $p - values$ to make an inference. To implement the test we treat the $\hat{\varepsilon}_i$ ($i = 1, ..., n$) as a sample from which the bootstrap samples $\hat{\varepsilon}_i^{(m)}$ ($i = 1, ..., n; m = 1, ..., M$) are drawn with replacement. Using the bootstrap sample we estimate the model under the null and the alternative of bounded inefficiency and construct the corresponding LR statistic. We repeat this procedure M times and calculate the percentage of times the bootstrap LR exceeds the actual one. This provides us with the bootstrap estimate of the asymptotic $p-value$ of LR under the null.

5 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:

$$y_{it} = \alpha_0 + \sum_{k=1}^{K} \alpha_k x_{it,k} + \varepsilon_{it} \tag{16}$$

where

$$\varepsilon_{it} = v_{it} - u_{it}. \tag{17}$$

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,

$$B_t = \sum_{i=0}^{K} b_i (t/T)^i, \quad t = 1, ..., T, \tag{18}$$

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

6 Simulations

To examine the finite sample performance of the three 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).[6] 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 7 and 8 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 quantitatively. The first important conclusion that can be drawn is that the MLE estimators for technology parameters, α_1 and α_2, are accurately estimated. As the noise level increases, the mean squared error (mse) of these estimates increases only marginally. The second important observation is that the estimates of the inefficiency bound have relatively smaller mse's when the noise level is mild. When noise level is high, as when $\sigma_v = 0.5$, $\hat{B}$ becomes inaccurate. In table 7 distribution parameters, $\hat{\sigma}$ and $\hat{\gamma}$ display a significantly upward bias and large mse as the signal-to-noise ratio decreases.[7] Wang and Schmidt (2008) show that the distribution of u degenerates to a point mass at $E[u]$ as λ tends to zero. This is also the case for our model. When the noise level is high relative to the variance of inefficiencies the distributional parameters are very hard identified. On the other hand, as $\lambda \to \infty$ the variance of the noise is not identified (Deterministic Frontier). Table 8 shows that the problem is alleviated somewhat when the sample size increases.

[6] 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.

[7] It can be shown that in this case the Hessian is 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 in this setting.

We now look at the doubly truncated normal model. Table 9 and 10 reports Monte Carlo results with a sample size of 200 and 1000, respectively. For 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 and γ, are upward biased, especially when λ and N are relatively small. Their mse is low for low levels of noise. In addition, the parameter μ is accurately estimated, especially then the sample size is large. The inefficiency bound is significantly distorted when the signal-to-noise ratio decreases. 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 11 and 12 provide results for the truncated exponential model with a sample size of 200 and 1000, respectively. As with the previous models, the technology parameter estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are accurately estimated. Parameters of the one-sided error term have relatively low mse's when the noise level is mild. It can be seen from these tables how sensitive the parameter estimates $\hat{\sigma}_u$ and B are to the level of stochastic noise. The estimated values of these parameters are highly contaminated by the noise when this dominates the inefficiency term. As expected, the finite sample problems with $\hat{\sigma}_u$ and B are lessened when we consider the larger sample size of 1000.

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, and experienced an adverse economic environment in the last few turbulent years of the last decade.[8] Our analysis covers a lengthy period between 1984 and 2009. What is generally observed during this period is that the number of commercial banks substantially decreased through either mergers or failures. The cuurent number of banks is less than half of the number in 1984. It is characteristic of the fact that the number of failed banks in 2009 was about 2.75 times more than that of the period 2001-2008 due to banking crisis that was fired up in summer of 2007. As of the present time, the number of mergers and new charters has decreased, while the proportion of problematic banks has dramatically increased. However, the biggest failures occurred during the financial crisis of early 1990's where almost 400 banks failed within three years. All these facts have triggered the interest of researchers to

[8] 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.

analyze the US 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 US 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[9]

$$Y_{it} = Y_{it}^{*\prime}\gamma + X_{it}^{\prime}\beta + v_{it} - u_{it}, \tag{19}$$

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).[10] Y_{it}^{*} includes the log of commercial and industrial loans/real estate loans (ciln) and installment loans/real estate loans (inln). In order to account for the riskiness and heterogeneity of the banks we include the log of the ratio of equity to total assets (*eqrt*) which usually measures the risk of insolvency of the banks in banking literature.[11] The lower the ratio the more riskier a bank is considered. We assume the 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 5. Once the individual effects u_{it} are estimated, technical efficiency for a particular firm at time t is calculated as $TE = \exp(u_{it} - \max_{1\leq j\leq N} u_{jt})$.

We use US commercial banking data from 1984 quarter I through 2009 quarter III. The data is a balanced panel of 4193 commercial banks and was recently compiled from the Consolidated Reports of Condition and Income (Call Report) and the FDIC Summary of Deposits. The data set includes 431879 observations for 103 quarterly periods. This is a fairly long panel and thus the assumption of time-invariant inefficiencies does not seem tenable. 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, along with the baseline fixed effect estimator (FIX) of Schmidt and Sickles (1984). Descriptive statistics for the bank-level variables are given in table 6, where all nominal values are converted to reflect 2000 year values.

---

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

[10] 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 savings deposits.

[11] We exclude from the sample banks with *eqrt* less that 0.02. Typically, these banks are close to failure and estimation of their efficiency scores require special treatments. (see Wheelock and Wilson, 2000 and Almanidis, 2010 for more discussion).

7.2 Results

Table 4 compares the parameter estimates of the bounded inefficiency (BIE) model with that of FIX, CSSW, and BC.[12] The structural parameters are statistically significant at the 1% 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. The negative value of the coefficient of the *eqrt* implies that riskier firms tend to produce more loans, and especially real estate loans that are considered of high risk. The positive sign of the estimate of the time trend shows technological progress on average. There is a slight difference between the distributional parameters of BIE and BC model which are also statstically significant at any conventional significance level. We also tested (not reported here) other distributional specifications for BIE discussed above. The distributional parameters obtained from normal-truncated half-normal model did not differ very much from that reported in the table, but those obtained from normal-truncated exponential model did. However, this is not a specific to bounded inefficiency models. Similar differences have been documented in unbounded SF models as well.

We test for the extent by which the distribution of inefficiencies displays positive skewness by testing the asymetry of the distribution of observable least squares residulas. We do so by utilizing the adjusted for skewness test statistic proposed by Bera and Premaratne (2001), which is suitable for testing for a symmetry of distributions with non-zero excess kurtosis. The test statistic is given by

$$S = \frac{n \cdot \psi_3^2}{\psi_2^3[9 + \psi_6 \psi_2^{-3} - 6\psi_4 \psi_2^{-2}]} \tag{20}$$

where ψ_i is the i^{th} population central moment of the least squares residuals and n is the number of observations in the sample. This statistic is asymptotically normally distributed and its value in our case is calculated to be 990.26, leading to rejection of the null hypothesis of symmetry at any conventional significance level. The asymmetry of the least squares residuals is also verified by quantile-quantile plot representation in figure 1.

We also estimate the time-varying inefficiency bound, B, using two approaches. First we estimate the bound for the 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. We choose to fit a fifth degree polynomial the coefficients of which are estimated by MLE along with the rest parameters of the model.[13] Both approaches

[12] 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.

[13] The choice of degrees of the time polynomial was based on a simple likelihood-ratio (LR) test and degrees of the polynomial ranging from 1 to 10. The maximum likelihood estimates of coefficients for this polynomial are given by

are illustrated in figure 3 with their respective 95% confidence intervals. It can be seen that the inefficiency bound has had a decreasing trend up to year 2005, when the financial crisis (informally) began, and then it is increasing for the remaining periods through 2009 quarter III. One interpretation of this trend can be that the deregulations in 1980's and 1990's increased competitive pressures and forced many inefficient banks to exit the industry, reducing the upper limit of inefficiency that banks could sustain and still remain in their particular niche market in the larger banking industry. The new upward trend can be attributed to the adverse economic environment and an increase in the proportion of banks that are characterized as "too big to fail."

Of course, for time-varying efficiency models such as CSSW, BC, and BIE, average efficiencies change over time.[14] These are illustrated in figure 2 along with their 95% confidence bounds. The BIE averaged efficiencies (panel 4) are significantly higher than those obtained from the fixed effect time-invariant model. However, the differences are small compared to BC and CSSW models. These small differences are not unexpected, however, since the existence of the inefficiency bound implies that the mean conditional distribution of inefficiencies is also bounded from above, resulting in higher average efficiencies. Failing to take the bound into account could possibly yield underestimated mean and individual efficiency scores (see table 1). We smooth the BIE averaged efficiencies by fitting ninth degree polynomial of time in order to capture their trend and also to be able to compare them with other two time-varying averaged efficiency estimates. These are represented by a curve labeled BIEsmooth. It can be seen that the efficiency trend for the BIE model is in close agreement with the CSSW model and better reflects the deregulatory reforms and consolidation of the US commercial banking industry. It is increasing initially and then falls soon after the saving and loans (S&L) crisis of early 90's began. It has the decreasing pace and reaches its minium in 1993 a year before Congress passed the Reigle-Neal Act which allowed commercial banks to merge with and acquire banks across the state lines. This spurred a new era of interstate banking and branching, which along with the Gramm-Leach-Billey Act that granted broad-based securities and insurance power to commercial banks, substantially decreased the number of banks operated in the US from 10,453 in 1994 to 8,315 by the end of the millennium. After 1994 the banking industry witnessed a rapid increase in averaged efficiencies of its institutions due in part to the disappearance of inefficient banks previously sheltered from competitive pressure and due to the expansion of large banks that both financially and geographically diversified their products. The increasing trend continues until the new recessionary period of 2001 and then steadily falls thereafter

---

$b_0 = -3.9477e{-}007$, $b_1 = 0.0039509^{**}$, $b_2 = -15.816^{***}$, $b_3 = 31656^{**}$, $b_4 = -3.168e{+}007^{*}$, $b_5 = 1.2682e + 010$

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

Table 4: 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	0.2407(0.0015)	0.2971(0.0014)	0.2284(0.0013)	0.2838(0.0012)
inln	0.2206(0.0013)	0.1715(0.0012)	0.2043(0.0013)	0.2609(0.0013)
dd	−0.0940(0.0024)	−0.0935(0.0020)	−0.1197(0.0024)	−0.0996(0.0020)
dep	−0.3999(0.0048)	−0.4037(0.0051)	−0.4368(0.0048)	−0.4053(0.0034)
lab	−0.3104(0.0046)	−0.2219(0.0042)	−0.1610(0.0044)	−0.1892(0.0020)
cap	−0.0460(0.0016)	−0.0464(0.0014)	−0.0510(0.0015)	−0.0965(0.0015)
purf	−0.1507(0.0034)	−0.1658(0.0029)	−0.1627(0.0034)	−0.1665(0.0031)
time	0.0057(0.0001)	–	0.0020(0.0001)	0.0021(0.0001)
eqrt	−0.1369(0.0045)	−0.1189(0.0041)	−0.0975(0.0044)	−0.1088(0.0039)
γ	0	0	0.7980(0.0115)	0, 7690(0.0058)
σ	0.2210(0.0034)	0.2070(0.0020)	0.2733(0.0045)	0.2712(0.0022)
μ	–	–	0.3240(0.0139)	0.3518(0.0630)
B	–	–	–	1.5186
ATE	0.5853	0.6470	0.6410	0.6998

until the rapid decline illustrating the effects of the 2007-2009 crisis. The CSSW model is able to show the weakness of the banking industry as early as 2005. This weakness is illustrated by the estimated inefficiency bound from the BIE model.

On the other hand, the BC model shows a slight, statistically non-significant, upward efficiency trend for all these periods ($\eta = 0.0066$). We also can look at the efficiency ranking of firms from these four estimators. Table 5 tabulates the Spearman rank correlations among different models, which shows that BIE efficiency ranking is in agreement with other estimators, especially with CSSW.

In sum, figures 2 and 3 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

Table 5: 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

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 the problems the financial services industry faces today.

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 settings. 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., and Schmidt, P., 1977, Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6, 21-37.

Afriat, S. N., 1972, Efficiency estimation of production functions. International Economic Review, 13, 568-598.

Alchian, A.A., 1950, Uncertainty, evolution, and economic theory. Journal of Political Economy, 58, 211-21.

Almanidis, P., 2010, Banking Crises, Early Warning Models, and Efficiency. Working Paper.

Almanidis, P., and Sickles, R.C., 2009, Skewness problem in Stochastic Frontier Models: Fact or Fiction? Invited submission to Exploring Research Frontiers in Contemporary Statistics and Econometrics: A Festschrift in Honor of Leopold Simar, Ingrid Van Keilegom and Paul Wilson (eds.), Springer Publishing: New York, 2010.

Battese, G.E., and 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., and 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.

Bera, A.K., and Premaratne, G., 2001, Adjusting the tests for skewness and kurtosis for distributional misspecifications. UIUC-CBA Research Working Paper No. 01-0116

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

Davidson, R., and MacKinnon, J.G., 1993, Estimation and Inference in Econometrics, New York, Oxford University Press

Demsetz, H., 1973, Industry structure, market rivalry, and public policy, Journal of Law and Economics 16, 1-9.

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

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.

Greene, W. H., 2007, The Econometric Approach to Efficiency Analysis, in H. O. Fried, C. A. K. Lovell and S. S. Schmidt, eds., The Measurement of Productive Efficiency: Techniques and Applications. New York: Oxford University Press.

Hansen, B.E., 1996, Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413-430.

Hansen, B.E., 1999, Threshold effects in non-dynamic panels: Estimation, testing, and inference. Journal of Econometrics 93, 345-368.

Hicks, J. R. (1935), Annual survey of economic theory: the theory of monopoly, Econometrica 3, 1-20.

Jondrow, J., Lovell, C.A.K., Materov I.S., and 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., and 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.

Kutlu, L., and Sickles, R.C., 2010, Testing the Quiet Life and Efficient Structure Hypotheses: A Dynamic Model of Market Power. Rice University, Mimeo.

Lee, L., 1993, Asymptotic distribution of the Maximum Likelihood Estimator for a Stochastic Frontier Function Model with a Singular Information Matrix. Econometric Theory, 9, 413-430.

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

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

Rao, C.R., 1973, Linear Statistical Inference and its Applications, 2nd edition New York: Wiley.

Richmond, J., 1974, Estimating the Efficiency of Production. International Economic Review, 15, pp. 515–521.

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

Rothenberg, T.J., 1971, Identification in parametric models. Econometrica 39 (3), 577–591.

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

Simar, L., and Wilson, P.W., 2010, Estimation and Inference in Cross-Sectional, Stochastic Frontier Models. Econometric Reviews, 29, 62-98.

Stigler, G.S., 1958, The economies of scale. Journal of Law and Economics, 1, 54-71.

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

Waldman, D.M., 1982, A stationary point for the stochastic frontier likelihood. Journal of Econometrics 18, 275-279.

Wang, W.S., and Schmidt, P., 2008, On the distribution of estimated technical efficiency in stochastic frontier models. Journal of Econometrics 148, 36-45.

Wheelock, D and Wilson, P., 2000, Why do banks disappear? The determinants of US bank failures and acquisitions. Review of Economics and Statistics 82(1), 127–138.

9 Appendix: First-order derivatives of log-likelihood function

The first-order derivatives for normal-doubly-truncated normal model are calculated based on log-likelihood function (10) under $\gamma-$parametrization and $\tilde{B}-$parametrization and are given by

$$\frac{\partial \ln L}{\partial a}=\sum_{i=1}^{n}\frac{(\varepsilon_i+\mu)x_i}{\sigma^2}+\frac{\sqrt{\gamma/(1-\gamma)}}{\sigma}\sum_{i=1}^{n}x_i\frac{\phi(z_{4i})-\phi(z_{3i})}{\Phi(z_{3i})-\Phi(z_{4i})}$$

$$\frac{\partial \ln L}{\partial \sigma^2}=\frac{n}{2\sigma^2}\frac{[(z_1\phi(z_1)-z_2\phi(z_2))]}{\Phi(z_1)-\Phi(z_2)}-\frac{n}{2\sigma^2}+\sum_{i=1}^{n}\frac{(\varepsilon_i+\mu)^2}{2\sigma^4}+\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{[z_{4i}\phi(z_{4i})-z_{3i}\varphi(z_{3i})]}{\Phi(z_{3i})-\Phi(z_{4i})}$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \lambda} &= \frac{n}{\gamma}\frac{[(z_1\phi(z_1)-z_2\phi(z_2))]}{\Phi(z_1)-\Phi(z_2)}+\frac{1}{\sigma}\sum_{i=1}^{n}\frac{1}{\Phi(z_{3i})-\Phi(z_{4i})}\{((-\ln(\tilde{B})+\varepsilon_i)\frac{1}{(1-\gamma)^2}\sqrt{(1-\gamma)/\gamma} \\ &+(\ln(\tilde{B})+\mu)\frac{1}{\gamma^2}\sqrt{\gamma/(1-\gamma)})\phi(z_{3i})-(\varepsilon_i\frac{1}{(1-\gamma)^2}\sqrt{(1-\gamma)/\gamma}-\mu\lambda\frac{1}{\gamma^2}\sqrt{\gamma/(1-\gamma)})\phi(z_{4i})\}\end{aligned}$$

$$\frac{\partial \ln(L)}{\partial \mu}=\frac{n}{\sigma\sqrt{\gamma}}\frac{\phi(z_1)-\phi(z_2)}{\Phi(z_1)-\Phi(z_2)}-\sum_{i=1}^{n}\frac{(\varepsilon_i+\mu)}{\sigma^2}+\frac{\sqrt{(1-\gamma)/\gamma}}{\sigma}\sum_{i=1}^{n}\frac{\phi(z_{4i})-\varphi(z_{3i})}{\Phi(z_{3i})-\Phi(z_{4i})}$$

$$\frac{\partial \ln(L)}{\partial \tilde{B}}=\frac{n}{\tilde{B}\sigma\sqrt{\gamma}}\frac{\phi(z_1)}{\Phi(z_1)-\Phi(z_2)}-\frac{1}{\tilde{B}\sigma\sqrt{(1-\gamma)\gamma}}\sum_{i=1}^{n}\frac{\phi(z_{3i})}{\Phi(z_{3i})-\Phi(z_{3i})}$$

where $\varepsilon_i=y_i-x_i\alpha,\ z_1=-\frac{(\ln(\tilde{B})+\mu)}{\sigma\sqrt{\gamma}},\ z_2=\frac{-\mu}{\sigma\sqrt{\gamma}},\ z_{3i}=-\frac{(\ln(\tilde{B})-\varepsilon_i)\sqrt{\gamma/(1-\gamma)}+(\ln(\tilde{B})+\mu)\sqrt{(1-\gamma)/\gamma}}{\sigma}$, and $z_{4i}=\frac{\varepsilon_i\sqrt{\gamma/(1-\gamma)}-\mu\sqrt{(1-\gamma)/\gamma}}{\sigma}$. The first-order derivatives of log-likelihood function of normal-truncated half-normal model are obtained after substituting $\mu = 0$ in the above expressions.

The scores for normal-truncated exponential model are derived from (14) as

$$\frac{\partial \ln L}{\partial a} = -\frac{\gamma^{-1/2}}{\sigma}\sum_{i=1}^{n} x_i + \frac{(1-\gamma)^{-1/2}}{\sigma}\sum_{i=1}^{n}\frac{\phi(\tilde{z}_{2i}) - \phi(\tilde{z}_{1i})}{\Phi(\tilde{z}_{1i}) - \Phi(\tilde{z}_{2i})}x_i$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \sigma} &= -\frac{n}{\sigma} - \frac{n\ln\tilde{B}\gamma^{-1/2}}{\sigma^2}\frac{e^{\frac{\ln\tilde{B}\gamma^{-1/2}}{\sigma}}}{1-e^{\frac{\ln\tilde{B}\gamma^{-1/2}}{\sigma}}} \\ &\quad + \frac{(1-\gamma)^{-1/2}}{\sigma^2}\sum_{i=1}^{n}\{\frac{\phi(\tilde{z}_{2i}) - \phi(\tilde{z}_{1i})}{\Phi(\tilde{z}_{1i}) - \Phi(\tilde{z}_{2i})}\varepsilon_i + \frac{\phi(\tilde{z}_{1i})}{\Phi(\tilde{z}_{1i}) - \Phi(\tilde{z}_{2i})}\ln\tilde{B}\}\end{aligned}$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \sigma} &= -\frac{n}{2\gamma} - \frac{n\ln\tilde{B}}{2\gamma^{3/2}}\frac{e^{\frac{\ln\tilde{B}\gamma^{-1/2}}{\sigma}}}{1-e^{\frac{\ln\tilde{B}\gamma^{-1/2}}{\sigma}}} - \frac{n}{2\gamma^2} - \frac{1}{2\gamma^{3/2}}\sum_{i=1}^{n}\varepsilon_i \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\{\frac{\phi(\tilde{z}_{2i}) - \phi(\tilde{z}_{1i})}{\Phi(\tilde{z}_{1i}) - \Phi(\tilde{z}_{2i})}\left(\frac{\varepsilon_i}{\sigma(1-\gamma)^{3/2}} - \frac{1}{\gamma^2}\sqrt{\frac{\gamma}{1-\gamma}}\right) - \frac{\ln\tilde{B}}{\sigma(1-\gamma)^{3/2}}\frac{\phi(\tilde{z}_{1i})}{\Phi(\tilde{z}_{1i}) - \Phi(\tilde{z}_{2i})}\end{aligned}$$

$$\frac{\partial \ln L}{\partial \tilde{B}} = \frac{n\gamma^{-1/2}}{\sigma\tilde{B}}\frac{e^{\frac{\ln\tilde{B}\gamma^{-1/2}}{\sigma}}}{1-e^{\frac{\ln\tilde{B}\gamma^{-1/2}}{\sigma}}} - \frac{(1-\gamma)^{-1/2}}{\tilde{B}\sigma}\sum_{i=1}^{n}\frac{\phi(\tilde{z}_{1i})}{\Phi(\tilde{z}_{1i}) - \Phi(\tilde{z}_{2i})}$$

where $\varepsilon_i = y_i - x_i\alpha$, $\tilde{z}_{1i} = \frac{(-\ln\tilde{B}+\varepsilon_i)(1-\gamma)^{-1/2}}{\sigma} + \sqrt{\frac{1-\gamma}{\gamma}}$ and $\tilde{z}_{1i} = \frac{\varepsilon_i(1-\gamma)^{-1/2}}{\sigma} + \sqrt{\frac{1-\gamma}{\gamma}}$.

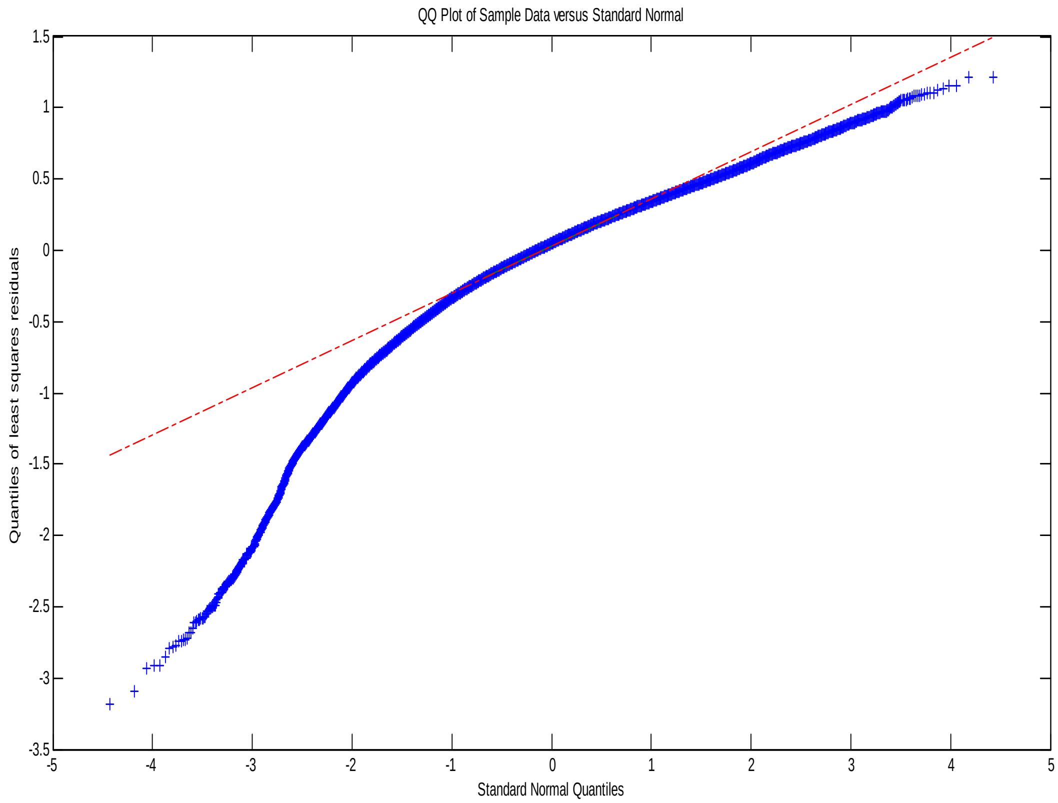

Figure 1: Quantile-Quantile plot

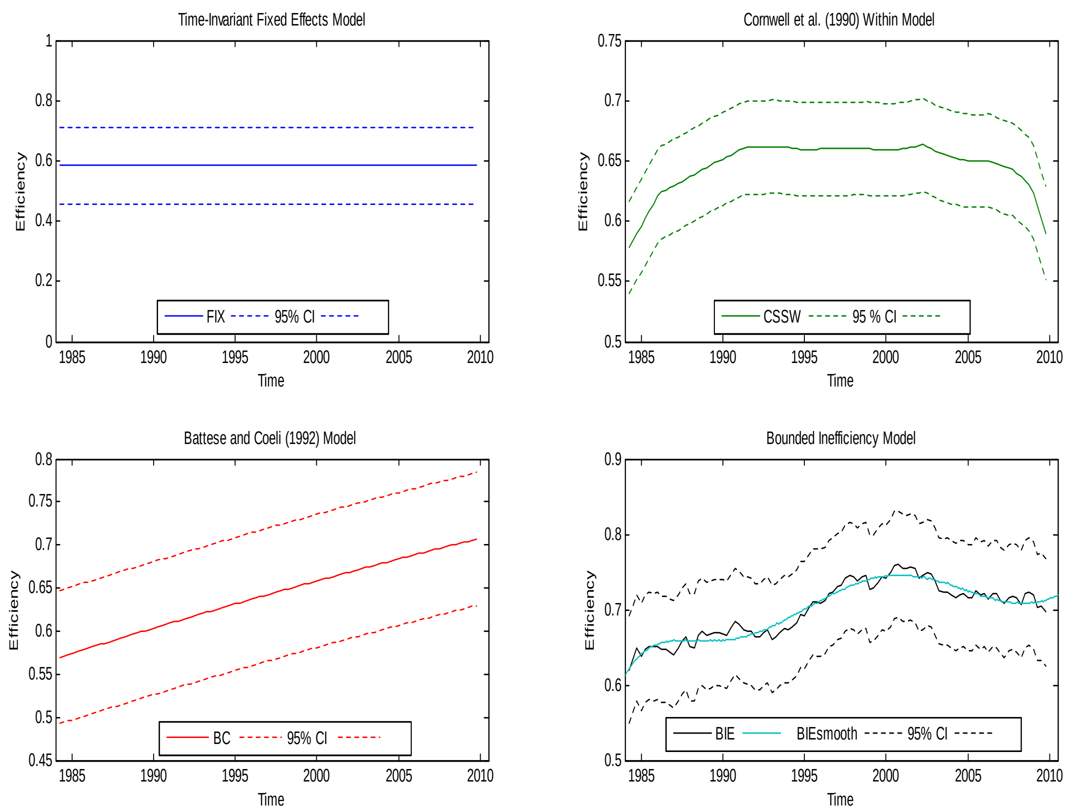

Figure 2: Averaged Efficiencies from each estimator

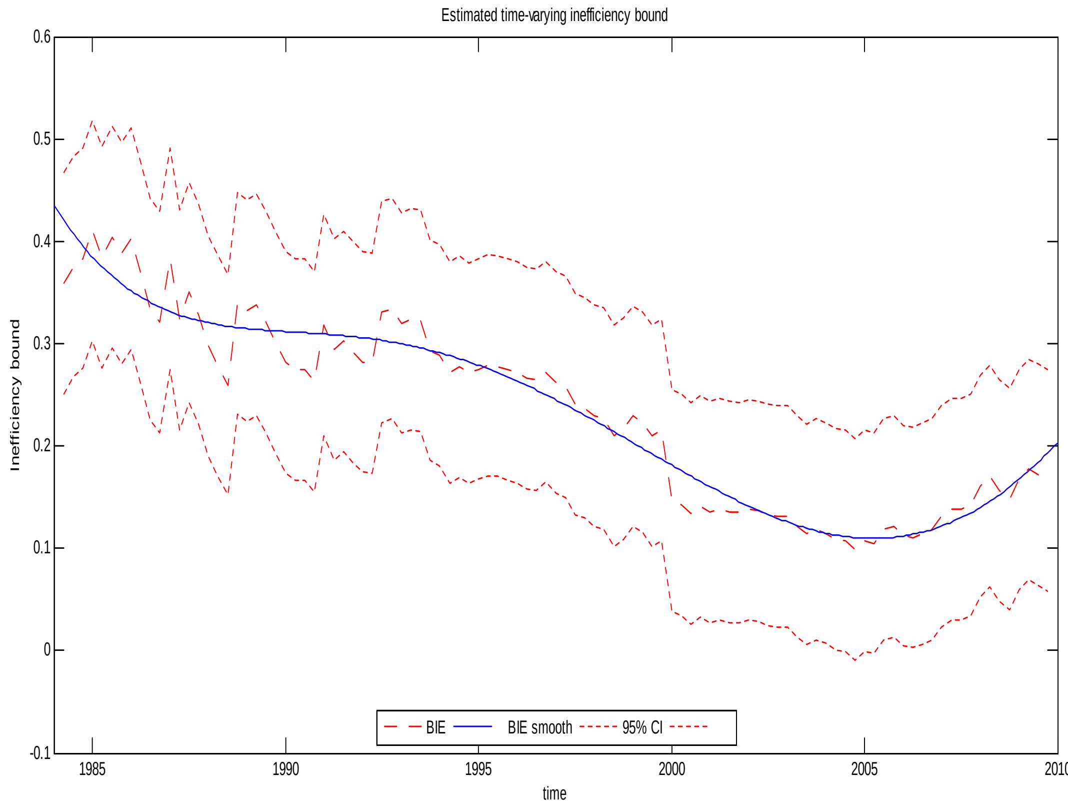

Figure 3: Estimated Inefficiency Bound

Table 6: Descpriptive statistics for bank-specific variables

Variable Name	Mean	Median	Standard Deviation	Min	Max
Real Estate loans	212968	17549	4341501	145	4.61E+08
Commercial and Industrial loans	103272	4908	2143974	46	1.82E+08
Installment loans	58869	4360	1417908	86	1.51E+08
Demand Deposits	54913	7282	912761	186	1.03E+08
Time and Savings Deposits	449003	46954	1.00E+07	1446	9.93E+08
Labor	186	29	2960	4	215670
Capital	8196	913	129778	9	1.16E+07
Purchased Funds	163785	13698	3322838	286	3.37E+08
Ratio of Equity to Total Assets	0.1007	0.0936	0.0312	0.0210	0.7459

Table 7: 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 8: Monte Carlo results for Truncated Half Normal model. The number of repetitions $M = 1000$. Sample size $N = 1000$.

		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
	$\hat{\sigma}$	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 9: 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

Table 10: 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
$\sigma_v = 0.1$	$\hat{\sigma}$	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
	$\hat{\mu}$	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
$\sigma_v = 0.2$	$\hat{\sigma}$	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
	$\hat{\mu}$	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
$\sigma_v = 0.5$	$\hat{\sigma}$	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
	$\hat{\mu}$	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 11: Monte Carlo results for Truncated Exponential model. The number of repetitions $M = 1000$. Sample size $N = 200$.

		True	$B = 0.8$		$B = 1.0$		$B = 1.2$	
			AVE	MSE	AVE	MSE	AVE	MSE
$\sigma_v = 0.1$	$\hat{\sigma}_u$	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
	$\hat{B}$		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
$\sigma_v = 0.2$	$\hat{\sigma}_u$	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
	$\hat{B}$		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
$\sigma_v = 0.5$	$\hat{\sigma}_u$	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
	$\hat{B}$		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

Table 12: 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