Skewness issue in Stochastic Frontiers Models: Fact or Fiction?

Pavlos Almanidis
Departmrnt of Economics
Rice University

Robin C. Sickles
Department of Economics
Rice University

Abstract

Skewness plays an important role in the stochastic frontier model. Ever since the model was introduced by Aigner, Lovell and Schmidt (1977), Meeusen and van den Broeck (1977), and Battese and Cora (1977), researchers have often found the residuals estimated from these models displayed skewness in the wrong direction. In such cases applied researchers were faced with two often overlapping alternatives, respecify the model and/or obtain a new sample, neither of which are particularly appealing due to inferential problems introduced by such data-mining approaches. Recently, Simar and Wilson (2009) developed a bootstrap procedure to address the skewness problem in finite samples. Their findings points to the latter alternative as potentially the more appropriate–increase the sample size. That is, the skewness problem is a finite sample one and it often arises in finite samples from a data generating process based on the correct skewness. Thus the researcher should first attempt to increase the sample size instead of changing the model specification if she finds the "wrong" skewness in her empirical analyses. Our paper considers an alternative explanation to the "wrong" skewness problem and implicitly a new solution. We utilize the Qian and Sickles (2008) model in which an upper bound to inefficiencies or a lower bound to efficiencies is specified. Especially we consider the case there inefficiencies are assumed to be doubly-truncated normal, which allows the least square residuals to display skewness in both direction and nests the standard half-normal and truncated-normal inefficiencies models. We show and prove that finding incorrect skewness does not necessarily indicate that the model is misspecified and the only misspecification should arise from the fact that we might consider wrong distribution for inefficiency process. We also conduct exhaustive set of Monte Carlo experiments that confirm our general findings and show that "wrong" skewness is also a large sample issue and there is nothing wrong about this if one considers the bounded inefficiency approach. In this way the "wrong" skewness, while problematic in standard models, becomes a property of the samples in which the distribution of inefficiencies is bounded.

JEL Classification: C13, C21, C23, D24, G21.
Key words and phrases: Stochastic frontier, bounded inefficiency, skewness, identification

1 Introduction

Stochastic frontier model was first introduced by Aigner, Lovell and Schmidt, Meeusen and van den Broeck, Battese and Corra almost simultaneously in 1977. It assumes that parametric functional form exists between dependent and independent variables, as opposed to the alternative approach of data envelopment analysis (DEA) proposed by Charnes et al. (1978) and the free disposable hull (FDH) of Deprins, Simar,and Tulkens (1984). However, its great virtue essentially lies on the idea of defining an error term composed of two parts, one-sided error term that captures the effects of inefficiencies relative to the stochastic frontier and two-sided error term that captures random shocks, measurement error and other statistical noise, and allows random variation of frontiers across firms. This formulation proved to be more realistic than the deterministic frontier model proposed by Aigner and Chu (1968), since it acknowledges the fact that the entire deviation from frontier cannot be attributed solely to technical inefficiency which is under firm's control. Since that time a myriad of papers have emerged in the literature discussing either methodological or practical issues, as well as a series of applications of these models to the wide range of data sets. A recent detailed discussion of any innovations and empirical applications in this area is provided by Kumbhkar and Lovell (2000) and Greene (2007)[1].

As mentioned above, the literature on stochastic frontier models is very large and developments have been made in various directions regarding model specification and estimation techniques. There are two main methods of estimation that researchers adopt in general. One is based on traditional stochastic frontier models as they first have formulated and uses maximum likelihood techniques (ML). The other is the Bayesian estimation method introduced by van de Broeck et al. (1994) and Koop(1994), and Koop et al. (1995, 1997), which utilizes Gibbs sampling algorithm with data augmentation and Markov chain Monte Carlo (MCMC) techniques to estimate the model parameters and individual or mean inefficiencies. Kim and Schmidt (2000) provide review and empirical comparison of these two methods in panel data models. Regarding the model specification researchers attempted to relax the most restrictive assumptions of classical stochastic frontier model. In this paper we mainly interested in the specification of the distributions of the two-sided random noise and inefficiencies, and especially in the later one. Aigner et al. (1977) in their pioneering work proposed normal distribution for the random noise and half-normal for the inefficiency process. These random errors were assumed to be independent and identically distributed across observations and statistically independent of each other. At the same time, Meeusen and van de Broeck (1977) and Aigner et al. (1977) broaden the list of distribution of inefficiencies by including the exponential distribution as well. Other more flexible densities were introduced later, characteristic of which were the gamma distribution proposed by Greene (1980a, 1980b) and Stevenson

[1] see The Measurment of Productive Efficiency and Productivity Growth, Chapter 2

(1980) and truncated normal distribution introduced by Stevenson (1980). In light of Bayesian estimation techniques, other complex distributions such as the lognormal and Weibull distributions were also proposed[2]. In subsequent years researchers both, using classical and Bayesian methods, dealt with relaxing the assumption of specific distribution for inefficiencies, introducing time varying and firm-specific effects, and determining correlation structure to random errors. Sickles (2005) analyze in great details the implication of these issues in the panel data models.

A common problem that arises in fitting stochastic frontier models is that the least squares residuals estimated from these models may display skewness in the "wrong" direction. While the theory predicts they will be negatively (positively) skewed in production (cost) frontiers they are positively (negatively) skewed. As researchers consider this statistic to be an important indicator of the specification of the stochastic frontier model, whenever they find the least squares residuals skewed in the "wrong" direction in finite samples they tend to believe that the model is misspecified or the current data is inconsistent with this model. Two course of actions are oftentimes taken: recpecify the model and/or obtain a new sample which will result in the desired sign of skewness. In most of situations applied researchers proceed assuming the absence of inefficiencies and perform least squares estimation[3]. This weak point of stochastic frontier models is emphasized in a series of papers, some of which try to justify that this phenomenon might arise in finite samples even for models that are correctly specified. Of course this problem can be avoided in panel data models by either utilizing the fixed effects model of Schmidt and Sickles (1984) or time-varying models proposed by Cornwell et al. (1990, 1996) and Kumbhakar (1990, 1991) wherein no assumptions are made on the distribution of inefficiency term or the correlation between inefficiency term and the regressors.

This paper intends to illustrate how the bounded inefficiency formulation, proposed by Qian and Sickles (2008), overcomes the issue of the "wrong" skewness in stochastic frontier model. We first show and test that the imposition of an upper bound to inefficiency enables the distribution of one-sided inefficiency process to display positive and negative sign of skewness. This is particularly true for the flexible distributions such as truncated normal and gamma. We consider the former distribution in current paper for ease of illustration, although the analysis can be extended to include the gamma and the Weibull distributions as well. Imposing the bound on truncated normal density function apart from the zero yields skewness that can have both, positive or negative sign depending on the position of the bound in the support of inefficiency distribution and thus justifying the occurrence of the so called "wrong" skewness. We show and prove formally that normal-truncated normal model is capable to handle and estimate the model with the "wrong" skewness. While based

[2]see Deprins and Simar (1989b) and Migon and Medici (2001).

[3]This is in particular due to the results that Waldman et al. (1980) and Waldman (1982) obtain for stochastic frontier models when half-normal distribution for inefficiencies is specified.

on the Waldman proof many researchers consider the "wrong" skewness issue as a finite sample, we show that it is very reasonable to obtain it in large samples also. Truncating the right tail of inefficiency distribution, we also perform series of Monte Carlo experimentations and show that in the cases where we have positively skewed[4] distribution we could still get very reasonable MLE estimates of lamda as well as other parameters of the model. We also find that even in cases where OLS residuals display negative skewness, MLE values of lamda can be statistically insignificant (close to zero). In sum, we conjecture that there is no strong connection between skewness sign and MLE estimates of lamda. So, according to our findings, the only misspecification in stochastic frontier models when skewness is found to be in opposite direction can be in the distribution of inefficiency. To correct this, we simply suggest to use the model with the bounded inefficiency.

The present paper is divided into five sections. In section 2 the general problem of "wrong" skewness in stochastic frontier models and its implications is discussed, as well as solutions proposed in literature to solve it. Section 3 provides the main framework of stochastic frontier model with bounded inefficiency. In section 4 we check the validity of the bounded inefficiency model under the "wrong" skewness using simple models and generalize Waldman's proof to formally support the stochastic frontier models under these circumstances. Monte Carlo simulation results are also discussed. Section 6 contains our main conclusions.

2 Skewness issue in Stochastic Frontier Analysis

2.1 "Wrong" skewness and its importance in frontier models

As an important diagnostic test of stochastic frontier models, skewness statistic of least squares residuals has received considerable attention of theoretical and applied researchers all of these years. The error specification of models is $\varepsilon_i = \upsilon_i - u_i$[5] for production frontiers, where υ_i represents statistical noise and is assumed to be $i.i.d$ as $N(0, \sigma_\upsilon^2)$ and $u_i \geq 0$ represents technical inefficiency which is also $i.i.d$ random element that follows one-sided distribution. Under the usual assumptions υ_i and u_i are statistically independent of each other and from regressors. Given these assumptions, the distribution of the composed error is asymmetric and non-normal implying that least squares estimators, if applied, will be inefficient and will not provide us with any measurement of technical inefficiency. However, OLS provides consistent estimation of all parameters except the intercept, since $E(\varepsilon_i) = -E(u_i) \leq 0$. Moreover,

[4] We perform Monte Carlo simulations for production frontiers. The result can be easily extended to cost frontiers as well.

[5] $\varepsilon_i = \upsilon_i + u_i$ for the case of cost frontiers.

$$E[(\varepsilon_i - E[\varepsilon_i])^3] = E[(\upsilon_i - u_i + E[u_i])^3] = -E[(u_i - E[u_i])^3] \quad (1)$$

which implies that the negative of the third moment of OLS residuals is a consistent estimator of the skewness of one-sided error.

The common distributions that appear in the literature are positively skewed reflecting the fact that the big proportion of the firms attain levels of production not very far from frontier. This theoretically means that, whenever we substract the positively skewed inefficiency component from symmetric error, the composite error should display negative skewness (positive skewness in case of cost frontiers). Based on these results, researchers find stochastic frontier models inappropriate to model the inefficiencies as they obtain residuals skewed in the "wrong" direction. The typical conclusion is that, either the model is messpecified or the data is not compatible with this model. However, there can be a third interpretation as well. This can be based on the fact that inefficiencies might nave been drawn from distribution which displays negative skewness.

Formally, the problem of skewness is discussed in Olson et al., (1980) in their derivation of corrected ordinary least squares (COLS) estimates as an alternative to maximum likelihood estimates. They refer to this problem as a "Type I" failure. COLS proceed the estimation of the slope parameters by OLS, which are unbiased and consistent. The OLS estimate of constant term is biased and can be corrected by simply adding $\sqrt{2/\pi}\ \hat{\sigma}_u$ to it. This term is the estimated bias of the OLS estimator of the constant term and can be easily calculated. The estimates of σ_υ^2 and σ_u^2 are derived by method of moments techniques using the second and third moments of OLS residuals. These are consistent, although not asymptotically efficient, and are given by

$$\hat{\sigma}_u^2 = [\sqrt{2/\pi}(\frac{\pi}{\pi - 4})\hat{\mu}_3]^{2/3} \quad (2)$$

and

$$\hat{\sigma}_\upsilon^2 = \hat{\mu}_2 - (\frac{\pi - 2}{\pi})\hat{\sigma}_u^2 \quad (3)$$

where $\hat{\mu}_2$ and $\hat{\mu}_3$ are the estimated second and third moments of the OLS residuals, respectively.

It is obvious from 2 that this method posses a serious flow whenever $\hat{\mu}_3$ is positive, since the estimated variance of inefficiencies becomes negative! In this case Waldman (1982) shows that MLE estimates of σ_u^2 is zero and that the model parameters can be efficiently estimated by OLS. Bellow we outline the main steps and results of Waldman's proof. Starting from log-likelihood function of normal-half-normal model which is written as

$$logL = nlog(\sqrt{2/\pi}) - nlog(\sigma) + \sum_{i=1}^{n} log[1 - \Phi(\frac{\varepsilon_i \lambda}{\sigma})] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\varepsilon_i^2 \tag{4}$$

where $\varepsilon_i = y_i - x_i\beta$, $\lambda = \frac{\sigma_u}{\sigma_v}$, $\sigma^2 = \sigma_v^2 + \sigma_u^2$, and $\Phi(\bullet)$ denotes the cdf of the standard normal distribution, Waldman notes there are two stationary points that characterize the log-likelihood function. Under the parameter vector $\theta' = (\beta', \sigma^2, \lambda)$, the first stationary point is the one for which the first-order derivatives of above function are zero and the second is the OLS solution for θ wherein the parameter λ is set to zero. The superiority of these two stationary points is then compared in cases of the wrong skewness. One way to do this is to examine the second-order derivative matrix of log-likelihood function at these two points. The Hessian matrix evaluated at OLS solution, $\theta^* = (b', s^2, 0)$, is

$$H(\theta^*) = \begin{bmatrix} -s^{-2}\sum_{i=1}^{n} x_i x_i' & \sqrt{2/\pi}s^{-1}\sum_{i=1}^{n} x_i & 0 \\ \sqrt{2/\pi}s^{-1}\sum_{i=1}^{n} x_i & -2n/\pi & 0 \\ 0 & 0 & -n/2s^4 \end{bmatrix} \tag{5}$$

where $b = (\sum_{i=1}^{n} x_i x_i^{'})^{-1}\sum_{i=1}^{n} x_i y_{i,}$ $s^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$ and e_i is the least squares residual.

This matrix is singular with $k+1$ negative characteristic roots and one zero root. The log-likelihood function is examined in the direction determined by the characteristic vector associated with the zero root which is given by the vector $z = (s\sqrt{2/\pi}, 1, 0)$. The term of interest is then the sign of

$$\begin{aligned} \Delta logL &= logL(\theta^* + \delta z) - logL(\theta^*) \qquad (6) \\ &= -\delta^2\frac{n}{\pi} + \sum_{i=1}^{n} log[2 - 2\Phi(e_i\delta s^{-1} - \delta^2\sqrt{2/\pi})] \end{aligned}$$

where $\delta > 0$ is an arbitrary small number. Using Taylor's series expansion it can be easily shown that

$$\Delta logL = (\delta^3/6s^3)\sqrt{2/\pi}[(\pi-4)/\pi]\sum_{i=1}^{n}\varepsilon_i^3 + O(\delta^4) \tag{7}$$

which clearly shows that if $\sum_{i=1}^{n}\varepsilon_i^3 > 0$ then maximum of the log-likelihood function is located at OLS solution and which is superior to MLE. Again this result suggests two strategies for practitioners: apply OLS whenever the least squares residuals display positive sign or increase the sample size, since

$$plim(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^3) = \sigma_u^3\sqrt{2/\pi}[(\pi-4)/\pi] < 0 \tag{8}$$

which implies that asymptotically the problem of the "wrong" skewness goes away.

This is true in particular if the inefficiencies are drawn from the half-normal distribution. What if they are not? What if they are drawn from the distribution which displays negative skewness as well? We will attempt to give answers to these questions.

The problem of the "wrong" skewness is also made apparent and emphasized by the two widely-used computer packages used to estimate stochastic frontiers. The first package LIMDEP 7.0, which is developed by Greene (1995), calculates and checks the skewness of the OLS residuals just before maximum likelihood estimation begins. In case the sign of the skewness statistic is positive, significantly or not, the message appears that warns user about the misspecification of the model and suggests to use OLS instead of MLE. The second software FRONTIER 4.1, produced by Coelli (1996), also obtains the OLS estimates first as a starting values for the grid search of starting value of the γ[6] parameter. If the skewness is positive, the final maximum likelihood value of this parameter is very close to zero, indicating no inefficiencies. More detailed description and comparison of these packages can be found in Sena (1999).

Based on these results, several parametric and non-parametric test statistics have been developed to check the skewness of least squares residuals in stochastic frontier models. Schmidt and Lin (1984) proposed the test statistic

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} \tag{9}$$

where m_2 and m_3 represent the second and the third moments of the empirical distribution of the least squares residuals. The distribution of $\sqrt{b_1}$ is not straightforward and the application of this test requires special tables provided by D'Agostino and Pearson (1973). Coeli (1995) proposed an alternative test statistic for testing whether the third moment of residuals is greater than or equal to zero

$$\sqrt{b_1^*} = \frac{m_3}{(6m_2^3/N)^{1/2}} \tag{10}$$

where N denotes the number of observations in sample. Under the null hypothesis of zero skewness, the third moment of OLS residuals is asymptotically distributed as a normal random variable with zero mean and variance $6m_2^3/N$. This implies that $\sqrt{b_1^*}$ is asymptotically distributed as a standard normal variable and one can consult the corresponding statistical tables for making an inference. These two tests, although easily computed and implemented, have unknown finite sample properties. Coeli (1995) conducts Monte Carlo experimentations and shows that $\sqrt{b_1^*}$ has correct size and good power in small samples. This is the reason why this test statistic is commonly accepted and used in applications.

[6] $\gamma = \frac{\sigma_u}{\sigma_u + \sigma_v}$ which is another reparametrization used in stochastic frontier models

2.2 Solutions to the "wrong" skewness

Nonetheless, solutions considered in introduction for the "wrong" skewness essentially constitute no solutions with regard of the stochastic frontier model. Setting the variance of inefficiency process to be equal to zero simply because the OLS residuals happened to have an opposite sign does not seem to be a very convincing argument. This would imply that all the firms in the industry are fully efficient, fact that is not true if one examines the relative productivities of firms or countries over time. On the other hand, data-mining techniques will introduce inferential problems and possibly bias in parameters and their standard errors according to Leamer (1978). Most importantly, the availability of the data in economics is very limited and thus this alternative seems to be not realistic in most of the times. An alternative that argues that the inefficiencies are drawn from the distribution with negative skew should be interpreted with caution. The major problem with this assumption is that it implies that there is only a very small fraction of the firms that attain a level of productivity or cost close to the frontier. This fact is also falsified by the data. Carree (2002) considers a distribution for inefficiencies that allows for both, negative and positive skewness[7]. He proposes a binomial distribution $b(n, p)$ which for range of values of parameter p is negatively skewed. He derives method-of-moments estimators in the same way as Olson et al. (1980) and Greene (1990) do and gives explanation of how theoretically and empirically the "wrong" skewness issue may arise in stochastic frontier model. The shortcoming of this approach, however, is that MM-estimators may not be defined for some empirical values of the higher sample moments of the least squares residuals. Empirically, the use of the binomial distribution can be justified by the model in which cycle of innovations and imitations occurs. The negatively skewed distribution arises from the fact that few firms in the industry innovate and the rest try to imitate them. While the period of imitation extends a considerable large number of firms may experience very large inefficiencies. Based on the real-world examples, this model can be hardly applicable. In addition, this approach is well understood and applied for cross-sectional data, but it is not clear how it will work in panel data stochastic frontier models. As we show later in this paper, the bounded inefficiency formulation does not faces such problems and has a nice interpretation for an industry or countries in both, cross-sectional and panel data models.

On the other hand, Greene (2007) and more recently Simar and Wilson (2009) note that in the finite samples, even the correctly specified stochastic frontier model is capable to produce least squares residuals with opposite skewness sign with relatively high frequency. This fact is even better justified under the case of bounded

[7] Other autors also considered distributions with negative skew (see Johnson et al. 1992, 1994)

inefficiencies. We provide a simple Monte Carlo study in section 4 which confirms these results. In application of the stochastic frontier model to the airlines panel data, Greene (2007) obtains positively skewed residuals. He corrects this issue by reversing the sign of the third moment of the OLS residuals to compute the first-round method-of-moments estimators of λ and σ. Under this strategy, reasonable and statistically significant estimates of λ and σ are obtained, suggesting the evidence of technical inefficiency. Simar and Wilson (2009) develop a bootstrap procedure that produces inefficiency estimates even if the least squares residuals have "incorrect" skewness sign. Nevertheless, these findings also indicate that skewness statistic shouldn't be related so closely to the inefficiency concept and especially to the parameters λ and σ, since even in the cases of the "correct" skewness one may obtain MLE estimates of lamda close to zero.

3 Stochastic frontier model with bounded inefficiency

3.1 Model

In this section we briefly introduce the stochastic frontier model with bounded inefficiency proposed by Qian and Sickles (2008)[8]. The formulation of the model is similar to the traditional stochastic frontier model with only difference that an upper bound to inefficiencies or a lower bound to efficiencies is specified. In this way a second truncation point, other than zero, is imposed to the distribution of the inefficiency process. The model in Cobb-Douglas log-linear form can be written as

$$y_i = x_i\beta + \varepsilon_i \tag{11}$$

where

$$\varepsilon_i = \upsilon_i - u_i \tag{12}$$

y_i denotes the log output, x_i is the matrix of log of k inputs, υ_i is the random statistical noise, and u_i defines the inefficency component. Under the usual assumptions, $\upsilon_i \sim^{iid} N(0, \sigma_\upsilon^2)$, $u_i \sim^{iid} f_u(x)$ is non-negative random variable and $f(\cdot)$ is defined on positive domain. Three distributions are considered for the inefficiencies: *doubly truncated normal*, *truncated half − normal*, and *truncated exponential* distribution. Also, υ_i and u_i are assumed to be statistically independent from each other and from regressors.

[8]For detailed description as well as Monte Carlo experiments of this model see the corresponding paper.

The initial purpose of the bounded inefficiency model was to introduce a stochastic frontier model in which the bound can be used for gauging the tolerance for or ruthlessness against the inefficient firms and thus to serve as an index of competitiveness of an industry. At the same time it was introducing another time-varying technical efficiency model in the literature. In this paper we note another usefulness of this model which is reflected in the flexibility of one-sided distribution of inefficiencies. This flexibility especially enables the truncated-normal distribution with strictly positive mean to produce positive, negative, and zero skewness. This leads us to take a closer look at the *doubly truncated normal* inefficiencies whose density is given by

$$f_u(x) = \frac{\frac{1}{\sigma_u}\phi(\frac{x-\mu}{\sigma_u})}{\Phi(\frac{B-\mu}{\sigma_u}) - \Phi(\frac{-\mu}{\sigma_u})} I_{[0,B]}(x), \qquad \sigma_u > 0, B > 0 \tag{13}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and pdf of the standard normal distribution respectively, and $I(\cdot)$ denotes the indicator function.

It can be easily verified that this distribution generalizes the truncated-normal ($B = \infty$), truncated-half normal ($\mu = 0$), and the half-normal ($B = \infty, \mu = 0$) distributions. The same generalization and flexibility can be shown for other distributions used to model the inefficiencies in stochastic frontier models, such as the *gamma* distribution *e.t.c.*

3.2 Estimation

Under the appropriate λ parametrization used in Aigner at al. (1977), the log-likelihood function for the *doubly truncated normal* model is given by

$$\begin{aligned}\log(L) = & -n\ln[\Phi(\frac{B-\mu}{\sigma_u(\sigma,\lambda)}) - \Phi(\frac{-\mu}{\sigma_u(\sigma,\lambda)})] \\ & -n\ln\sigma - \frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n}\frac{(\varepsilon_i+\mu)^2}{2\sigma^2} \\ & +\sum_{i=1}^{n}\ln\{\Phi(\frac{(B+\varepsilon_i)\lambda + (B-\mu)\lambda^{-1}}{\sigma}) \\ & -\Phi(\frac{\varepsilon_i\lambda - \mu\lambda^{-1}}{\sigma})\}\end{aligned} \tag{14}$$

where $\sigma_u(\sigma,\lambda) = \frac{\sigma}{\sqrt{1+\frac{1}{\lambda^2}}}$, $\lambda = \frac{\sigma_u}{\sigma_v}$, and $\varepsilon_i = y_i - x_i\beta$.

This log-likelihood function can be maximized to obtain the MLE estimates of the model parameters along with the parameter that determines the bound of the one-sided distribution. While the support of the distribution of u depends on the bound, the support of the composite error is unbounded. Hence, the regularity conditions for MLE are not violated and we can get consistent and asymptotically efficient estimators. However, the global identifiability ala Rothenberg (1971) of this model fails, which is also true for the normal-truncated normal model, and we can identify some parameters only locally. We provide more discussion on identification in the appendix of this paper.

Moreover, the truncation parameter B makes more sence in highly deregulated and competitive markets and its estiamte can provide a useful index of competitiveness of a market or an industry. In addition

$$E[u_i|\varepsilon_i = \hat{\varepsilon}_i] \tag{15}$$

where $\hat{\varepsilon}_i = y_i - X_i\hat{\beta}$, can be used in the same spirit as in Jondrw et al. (1982) to derive individual and mean technical inefficiencies.

4 Skewness statistic under the bounded inefficiencies

4.1 Derivation of skewness and COLS estimates with doubly-truncated-normal inefficiencies

The location parameter of the $doubly-truncated-normal$ distribution, as a function of the imposed bound, mean and the variance of the normal distribution is given by

$$\psi_1(B, \mu, \sigma_u^2) = E(u) = \mu + \sigma_u\eta \tag{16}$$

with

$$\eta \equiv \frac{\phi(\xi_1) - \phi(\xi_2)}{\Phi(\xi_2) - \Phi(\xi_1)} \tag{17}$$

where $\xi_1 = \frac{-\mu}{\sigma_u}$, $\xi_2 = \frac{B-\mu}{\sigma_u}$, and μ is the mean of the normal distribution, while $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and pdf of the standard normal distribution, respectively. η

represents the inverse Mill's ratio and it is equal to $\sqrt{2/\pi}$ for the normal-half-normal model. It should be noted that ξ_1 and ξ_2 are lower and upper truncation points of the standard normal density, respectively.

The central population moments up through order four as a functions of B, μ, and σ_u^2 are given by

$$\psi_2(B,\mu,\sigma_u^2) = \sigma_u^2(1-\eta^2+\frac{\xi_1\phi\,(\xi_1)-\xi_2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)}) \tag{18}$$

$$\begin{aligned}\psi_3(B,\mu,\sigma_u^2) &= \sigma_u^3(2\eta^3-[3(\frac{\xi_1\phi(\xi_1)-\xi_2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)})+1]\eta \\ &\quad +\frac{\xi_1^2\phi(\xi_1)-\xi_2^2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)})\end{aligned} \tag{19}$$

$$\begin{aligned}\psi_4(B,\mu,\sigma_u^2) &= \sigma_u^4(-3\eta^4+2[3(\frac{\xi_1\phi(\xi_1)-\xi_2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)}-1)]\eta^2 \\ &\quad -4\eta(\frac{\xi_1^2\phi(\xi_1)-\xi_2^2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)}) \\ &\quad +3(\frac{\xi_1\phi(\xi_1)-\xi_2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)})+\frac{\xi_1^3\phi(\xi_1)-\xi_2^3\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)}+3)\end{aligned} \tag{20}$$

Two special cases immediately arise from the doublyt runcated normal distribution, one is the normal distribution and the other is the half-normal distribution. If we let $\xi_1=-\infty$ and $\xi_2=\infty$ then η becomes zero and if we additionally use the L'Hospital's Rule, under which $\lim_{\xi_1\to-\infty}\xi_1\phi(\xi_1)=\lim_{\xi_2\to\infty}\xi_2\phi(\xi_2)=0$ and also $\lim_{\xi_1\to-\infty}\xi_1^2\phi(\xi_1)=\lim_{\xi_2\to\infty}\xi_2^2\phi(\xi_2)=0$, we obtain exactly the cumulants of the normal distribution. On the other hand, if only the lower truncation exists ($\xi_1=0$) and $\mu=0$ we obtain results for the half-normal distribution. It is also straightforward to obtain results for truncated normal distribution.

Using these expressions the skewness of the doubly-truncated-normal distribution is calculated then as

$$\begin{aligned}\gamma_1(B,\mu,\sigma_u^2) &= \frac{\psi_3}{\psi_{2^{3/2}}}=\frac{\psi_3}{\psi_2\sqrt{\psi_2}} \\ &= \frac{(2\eta^3-[3(\frac{\xi_1\phi(\xi_1)-\xi_2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)})+1]\eta++\frac{\xi_1^2\phi(\xi_1)-\xi_2^2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)})}{1-\eta^2+\frac{\xi_1\phi(\xi_1)-\xi_2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)}}\end{aligned} \tag{21}$$

This statistic along with the kurtosis are parameters which describe the shape of the distribution independent of location and scale. Most of the non-symmetric distributions have either positive or negative γ_1. In this particular case the sign of skewness is ambiguous. It is either positive whenever $B > 2\mu$ or negative otherwise. Note that this is true for μ strictly positive since bound is also by assumption a strictly positive value[9].

So now what are the implications for skewness having both signs in the stochastic frontier analysis framework? Theoretically, it justifies the fact that the least squares residuals can be skewed in both directions while the variance of inefficiency term is nonzero. Technically, in finite samples anything can happen and the skewness statistic is more variable and prone to outliers and the "wrong" skewness can be obtained with non-zero probability.

One could employ the COLS estimation method to obtain first round method-of-moments estimates of λ and σ in order to start the iterations. However, the second and higher order moments of ε under a doubly-truncated-normal model are non-linear functions of parameters. Thus the global identifiability of the model fails (Rothenberg, 1971), which is also the case for the truncated-normal model. Greene (1993), and Ritter and Simar (1997) provide more discussion on identification issues in SFM[10]. On the other hand, the Fisher's information matrix is nonsingular at any point of the parameter space Θ for which $\lambda > 0$ and is not too large. This may establish local identifiability of the model ala Rothenberg. As for $\lambda \to 0$ and $\lambda \to \infty$, Wang and Schmidt (2008) show that the distribution of $\hat{u}$ collapses at the point $E[u]$ and converges to the distribution of u, respectively. We do not further discuss these limiting cases in the current paper.

The COLS estimators of variances of inefficiency term and the noise are given by[11]

$$\hat{\sigma}_u^2 = [\frac{-\hat{m}_3}{2\eta^3 - [3(\frac{\xi_1\phi(\xi_1)-\xi_2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)}) + 1]\eta + \frac{\xi_1^2\phi(\xi_1)-\xi_2^2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)}}]^{2/3} \tag{22}$$

and

$$\hat{\sigma}_v^2 = \hat{m}_2 - \hat{\sigma}_u^2(1 - \eta^2 + \frac{\xi_1\phi\left(\xi_1\right) - \xi_2\phi(\xi_2)}{\Phi(\xi_2) - \Phi(\xi_1)}) \tag{23}$$

[9] see appendix for graphical representation of doubly truncated inefficiencies with $\mu > 0$.

[10] The identification in cases of normal-truncated normal and normal-doubly truncated models is discussed in the appendix

[11] The second and third central moments of OLS residuals are $\hat{m}_2 = \sigma_v^2 + \sigma_u^2(1-\eta^2+\frac{\xi_1\phi(\xi_1)-\xi_2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)})$ and $\hat{m}_3 = \sigma_u^3(2\eta^3 - [3(\frac{\xi_1\phi(\xi_1)-\xi_2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)})+1]\eta + \frac{\xi_1^2\phi(\xi_1)-\xi_2^2\phi(\xi_2)}{\Phi(\xi_2)-\Phi(\xi_1)})$ respectively. Solving these two equations we get COLS estimators of σ_v^2 and σ_u^2.

where $\hat{m}_2$ and $\hat{m}_3$ are the second and the third sample moments of least squares residuals.

For the illustration of the skewness problem we can fix the values of parameters B and μ and calculate σ_v^2 and σ_u^2 from (22) and (23)[12]. Since the negative of the third moment of the OLS residuals is an unbiased and consistent estimator of the skewness of inefficiencies, one can see that the estimate of the σ_u^2 can have positive sign even in the case of positive skewness as opposed to the standard models. Most importantly, the "type I" failure goes away asymptotically since positive $\hat{m}_3$ would imply that ψ_3 is negative, which is whenever $B < 2\mu$ and so $\hat{\sigma}_u^2$ cannot be never negative. In cases where we have $B = 2\mu$ the ratio in (22) is unidentified. By applying L'Hospital rule and evaluating the limits we can see that the variance of inefficiency term is strictly positive number. The only case that it is zero is when the bound is zero.

We can test the extent to which the distribution of unobservable inefficiencies can display negative or positive skewness using the observable residuals according to the expression in (1). For this purpose we can utilize the adjusted for skewness test statistic proposed by Bera and Premaratne (2001) , since the excess kurtosis is not zero. By using the standard test for skewness we will have either over-rejection or under-rejection of the null hypothesis of non-negative skewness and this will depend primarily on the sign of the kurtosis (excess). In addition, since there are two points at which the doubly-truncated-normal distribution has zero skewness the standard tests are not appropriate once they cannot distinguish these, fact that would lead researchers falsely to not reject the null hypothesis of zero variance. The modified likelihood ratio statistic (see Lee, 1993), which is asymptotically distributed as $\frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$ does not faces such problems.

We also conduct Monte Carlo experiment in the same spirit as in Simar and Wilson (2009) wherein they note that in the finite samples even the correctly specified stochastic frontier model is capable of generating least squares residuals with the "wrong" skewness with relatively high frequency. They calculate the proportion of samples with positively skewed residuals which converges to zero as the sample size grows large. They conduct a Monte Carlo experiment and calculate the proportion of samples with positively skewed residuals. This proportion converges to zero as the sample size grows large. We conduct the same experiment under current error specification and display the results in table 1 in the appendix. We set the parameter μ to 1 and use inverse CDF method to sample from convolution of inefficiency and the noise distributions by varying the bound parameter[13]. By this way we examine all the three cases of the skewness sign and compute the proportion of 1000 samples with positive skewness. Our findings again clearly indicate that the skewness issue is also

[12] In normal-half-normal model these values are also fixed ($B = \infty$, $\mu = 0$).

[13] see appendix for the sampling from the distribution of the composite error ε using the inverse CDF method

a large sample issue. This means that if the true DGP is based on inefficiencies that are drawn from doubly-truncated-normal distribution and researcher fails to see that and she finds the skewness statistic having the wrong direction then she will reject her model. Even worse, if there is a potentiality of increasing the sample size and researcher keeps increasing it and finds continuously positive signs of skewness then at the end of the day she will erroneously assume that all firms in her sample are supper efficient. The flexibility of the bounded inefficiency approach avoids this problem. We also conduct experiments for other error specification such as truncated-half-normal ($\mu = 0$) and truncated exponential model. To save space we do not report their results but it is worth to mention that in these case for certain levels of the bound the skewness statistic becomes statistically insignificant and so the null hypothesis of no inefficiencies cannot be rejected , even if λ is not zero, applying the standard tests. We also note in our Monte Carlo experiments that if the DGP from which a sample of data is drawn has bounded inefficiency then it will mask the true skewness. It is often the case in such settings that point estimates of skewness may have the "wrong" sign. However, this is simply due to the weak identifiability of skewness in a stochastic frontier with bounded inefficiency and the "wrong" sign is not "significantly wrong" in a statistical sense.

4.2 Generalization of Waldman's proof

We can formally proof our findings by following the main arguments in the Waldman's proof. To compare and contrast the problem of the "wrong" skewness with the case of the normal-half-normal model, we will treat without loss of generality the values of parameters B and μ as fixed and consider the scores of parameter $\theta = (\beta', \sigma^2, \lambda)$ as a function of these parameters. Note that, the normal-half-normal model fixes these values at infinity and zero, respectively. In this case the second-order derivative matrix at OLS solution point, $\theta^* = (b', s^2, 0)$, is given by

$$H(\theta^*) = \begin{bmatrix} -s^{-2}\sum_{i=1}^{n} x_i x_i' & -\frac{1}{s^4}\sum_{i=1}^{n}(e_i-\mu)x_i & 0 \\ -\frac{1}{s^4}\sum_{i=1}^{n}(e_i-\mu)x_i & \frac{n}{2s^4} - \frac{1}{s^6}\sum_{i=1}^{n}(e_i-\mu)^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{24}$$

where $b = (\sum_{i=1}^{n} x_i x_i')^{-1} \sum_{i=1}^{n} x_i y_{i,}\ s^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$ and e_i is the least squares residual.

Obviously, $H(\theta^*)$ is singular with $k+1$ negative characteristic roots and one zero root. The eigenvector associated with this root is given by $z' = (\mathbf{0}, 0, 1)$. We search the sign of $\Delta logL = logL(\theta^* + \delta z) - logL(\theta^*)$ in the positive direction ($\delta > 0$), since λ is constrained to be non-negative. By expanding the $\Delta logL$ the first term in the series is zero since OLS is a stationary point. The second term also vanishes since

$|H(\theta^*)| = 0$. So, the only relative point that remains to be evaluated is the third derivative with respect of λ at OLS solution

$$\frac{1}{6}\delta^3 \frac{\partial^3 log(\theta^*)}{\partial\lambda^3} \tag{25}$$

Substituting for the third derivative and ignoring higher order terms, we get

$$\begin{aligned} \Delta logL \;\; &\cong \;\; \frac{\delta^3}{6s^3}\{-2[\varpi]^3 + [3(\frac{\omega_2\phi(\omega_2) - \omega_1\phi(\omega_1)}{\Phi(\omega_1) - \Phi(\omega_2)}) + 1]\varpi \\ &+ \frac{\omega_2^2\phi(\omega_2) - \omega_1^2\phi(\omega_1)}{\Phi(\omega_1) - \Phi(\omega_2)}\}\sum_{i=1}^{n} e_i^3 \end{aligned} \tag{26}$$

where $\omega_1 = \frac{B-\mu}{s}$, $\omega_2 = -\frac{\mu}{s}$, and $\varpi = \frac{\phi(\omega_1)-\omega\phi(\omega_2)}{\Phi(\omega_2)-\Phi(\omega_1)}$.

Now, it can be seen that the third order term of least squares residuals need not always have the opposite sign with $\Delta logL$. This will mainly depend on the relationship between the imposed bound and the mean of the normal distribution. For $B < 2\mu$, η is negative and the term in the curly brackets becomes positive. This implies that in finite samples whenever researcher finds positively skewed residuals it might be the case that the inefficiencies have been drawn from a distribution that has negative skew. For $B = 2\mu$, $\Delta logL = 0$. In this case MLE should be employed since it will be more efficient than OLS and will provide us with technical inefficiency estimates. Asymptotically the third order term of OLS residuals and the expression in curly brackets have the same sign since

$$\begin{aligned} plim(\frac{1}{n}\sum_{i=1}^{n} e_i^3) \;\; &= \;\; -\sigma_u^3(2\eta^3 - [3(\frac{\xi_1\phi(\xi_1) - \xi_2\phi(\xi_2)}{\Phi(\xi_2) - \Phi(\xi_1)}) + 1]\eta \\ &+ \frac{\xi_1^2\phi(\xi_1) - \xi_2^2\phi(\xi_2)}{\Phi(\xi_2) - \Phi(\xi_1)}) \end{aligned} \tag{27}$$

which implies that we can observe the "wrong" skewness even in large samples. So, based on this generalized proof we can argue that the problem of the "wrong" skewness is not a just finite sample issue. Positive or negative skewness of least squares residuals will always imply positive variance of inefficiency process in large samples. In finite samples, anything can happen. We can obtain negatively skewed residuals even if we sample from negatively skewed distribution of inefficiencies. It remains an empirical issue for the normal-doubly-truncated-normal model to not be able to identify positive and negative skewness cases.

5 Conclusions

Most of the distributions for inefficiencies considered in stochastic frontier models are positively skewed. Half-normal distribution is commonly used in literature and applications. Doubly truncated normal inefficiencies generalize the model in a way that allow the negative skewness, as well. This implies that finding incorrect skewness does not necessarily indicate that the model is misspecified. The only misspecification should arise from the fact that we might consider wrong distribution for inefficiency process which has the opposite skewness from the skewness of the least square residuals. We formally prove that normal-doubly truncated normal model can still be valid with the "wrong" sign of the skewness statistic and does not preclude its appearance in large samples. Therefore, we add an additional strategy for this case of conceived anomaly: to use bounded inefficiency approach.

6 Appendix

6.1 Truncated Normal and Doubly Truncated Normal Distributions

Below we depict four distributions for inefficiencies that are discussed in current paper. Namely, truncated normal and doubly truncated normal with zero, positive, and negative skewness. The skewness statistic of truncated distribution is

$$\begin{aligned}\gamma_1(B,\mu,\sigma_u^2) &= \frac{\sigma_u^3(2\eta^3 - [3(\frac{\xi_1\phi(\xi_1)}{1-\Phi(\xi_1)}) + 1]\eta + \frac{\xi_1^2\phi(\xi_1)}{1-\Phi(\xi_1)})}{1-\eta^2+\frac{\xi_1\phi(\xi_1)}{1-\Phi(\xi_1)}} \\ &= \frac{\sigma_u^3}{\psi_2}\eta[2\eta^2 - 3\xi_1 - 1 + \xi_1^2] \qquad (28)\end{aligned}$$

where $\eta \equiv \frac{\phi(\xi_1)}{1-\Phi(\xi_1)}$ and it has positive sign which means that γ_1 is positive, as well. Note that, as μ grows large η and so γ_1 tend to zero and the truncated normal distribution resembles the bell-shaped normal distribution.

As mentioned previously, we still might have the skewness OLS residual to be zero and the variance of inefficiency process strictly positive[14]. This is the case when $B = 2\mu$ and is depicted on the following graph.

The graphs of positively skewed ($B > 2\mu$) and negatively skewed ($B < 2\mu$) inefficiencies are provided below. Again it is clear that negative skewness does not imply the lack of inefficiencies

6.2 Inverse CDF method

In this part we describe the inverse CDF method which is used to sample from distribution of composed error term. Particularly, we know that the density function of doubly truncated normal inefficiency term is given by

$$f_u(u) = \frac{\frac{1}{\sigma_u}\phi(\frac{u-\mu}{\sigma_u})}{\Phi(\frac{B-\mu}{\sigma_u}) - \Phi(\frac{-\mu}{\sigma_u})}\mathbf{1}_{[0,B]}(x),\ \sigma_u > 0, B > 0, \qquad (29)$$

[14] Recall that in standard models the variance of inefficiency term should be zero and the conclusion is that there are no inefficient firms in the sample

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and pdf of the standard normal distribution respectively, and $\mathbf{1}[0,B]$ is an indicator function.

The probability density function for the composite terror term, ε, is derived as follows. Since u and v are independent, the joint density function of u and v is

$$\begin{aligned} f_{u,v}(u,v) &= \frac{[\frac{1}{\sigma_u}\phi(\frac{u-\mu}{\sigma_u})][\frac{1}{\sigma_v}\phi(\frac{v}{\sigma_v})]}{\Phi(\frac{B-\mu}{\sigma_u})-\Phi(\frac{-\mu}{\sigma_u})}\mathbf{1}_{[0,B]}(x) \\ &= \frac{\exp(-\frac{(u-\mu)^2}{2\sigma_u^2}-\frac{v^2}{2\sigma_v^2})\mathbf{1}_{[0,B]}(x)}{2\pi\sigma_u\sigma_v(\Phi(\frac{B-\mu}{\sigma_u})-\Phi(\frac{-\mu}{\sigma_u}))}. \end{aligned} \tag{30}$$

The joint density function of u and ε is then

$$f_{u,\varepsilon}(u,\varepsilon) = \frac{\exp(-\frac{(u-\mu)^2}{2\sigma_u^2}-\frac{(u+\varepsilon)^2}{2\sigma_v^2})\mathbf{1}_{[0,B]}(x)}{2\pi\sigma_u\sigma_v(\Phi(\frac{B-\mu}{\sigma_u})-\Phi(\frac{-\mu}{\sigma_u}))} \tag{31}$$

By integrating u out of (31), we get the marginal distribution of ε,

$$\begin{aligned} f_\varepsilon(\varepsilon) &= \left[\Phi(\frac{B-\mu}{\sigma_u})-\Phi(\frac{-\mu}{\sigma_u})\right]^{-1}\cdot\left[\frac{1}{\sigma}\phi(\frac{\varepsilon+\mu}{\sigma})\right]\cdot \\ &\quad \left[\Phi(\frac{(B+\varepsilon)\lambda+(B-\mu)\lambda^{-1}}{\sigma})-\Phi(\frac{\varepsilon\lambda-\mu\lambda^{-1}}{\sigma})\right], \end{aligned} \tag{32}$$

where

$$\begin{aligned} \sigma &= \sqrt{\sigma_u^2+\sigma_v^2} \\ \lambda &= \sigma_u/\sigma_v. \end{aligned} \tag{33}$$

The inverse CDF method requires that ε is a continuos random variable with cdf F and pdf f, respectively. Then, if $U \sim U(0,1)$ is uniform random variable on the unit interval then the sample draws can be obtained from $\varepsilon = F^{-1}(U)$. This of course requires that the distribution function F can be easily calculated. To obtain the cdf of ε we need to integrate its pdf which is not an easy task. However, since, u and v are assumed to be independent random variables we can independently draw them and use $\varepsilon = v - u$ to construct the values of the overall error term. Hence, what we need is to apply inverse CDF method to draw samples from doubly-truncated-normal distribution. The cdf of this distribution is given by

$$\begin{aligned} F(u) &= [\Phi(\frac{B-\mu}{\sigma_u})-\Phi(\frac{-\mu}{\sigma_u})]^{-1}\frac{1}{\sigma_u}\int_0^u \phi(\frac{u-\mu}{\sigma_u})dt \\ &= \frac{\Phi(\frac{u-\mu}{\sigma_u})-\Phi(\frac{-\mu}{\sigma_u})]}{\Phi(\frac{B-\mu}{\sigma_u})-\Phi(\frac{-\mu}{\sigma_u})} \end{aligned} \tag{34}$$

$$U = \frac{\Phi(\frac{u-\mu}{\sigma_u}) - \Phi(\frac{-\mu}{\sigma_u})]}{\Phi(\frac{B-\mu}{\sigma_u}) - \Phi(\frac{-\mu}{\sigma_u})} \tag{35}$$

By inverting (35) we can get the desired draws from

$$u = \mu + \sigma_u F^{-1}(\Phi(\frac{-\mu}{\sigma_u}) + U[\Phi(\frac{B-\mu}{\sigma_u}) - \Phi(\frac{-\mu}{\sigma_u})]) \tag{36}$$

Having these tools at hand we could investigate various DGP's with different values of the parameters λ and B.

6.3 Identification in normal-doubly-truncated-normal model

Stochastic Frontier Models are likelihood-based models and such as require careful examination of identification of parameters and inefficiencies scores. Very few studies consider these issues in SFM, among them Greene (1993, 2007) and Ritter and Simar (1997). Rothenberg (1971), defines two kind of parametric identifications in models based on likelihood inference: the global and the local identification.

Definition 1. *Two parameter points (structures) θ^1 and θ^2 are said to be observationally equivalent if $f(y,\theta^1) = f(y,\theta^2)$ for all y in R^n. A parameter point $\theta^0 \in \Theta$ is said to be identifiable if there is no $\theta \in \Theta$ which is observationally equivalent. For exponential family densities, this definition is equivalent to have the Fisher's information matrix nonsingular for every convex set of parameter points containing Θ. For nonexponential family densities the condition requires that every parameter can be expressed only as a function of sample moments of the corresponding probability distribution. That is, suppose there exist p known functions $g_1(Y), ..., g_p(Y)$ such that for all θ in Θ*

$$\theta_i = E[g_i(Y)] \qquad\qquad i = 1, 2, ..., p$$

Then every θ in Θ is identifiable.

Definition 2. *A parameter point (structure) θ^0 is said to be locally identified if there exists an open neighborhood of θ^0 containing no other θ in Θ which is observationally equivalent. Equivalently, let θ^0 be a regular point of Fisher's information matrix $I(\theta)$[15]. Then θ^0 is locally identifiable if and only if $I(\theta^0)$ is nonsingular. In other words, if $I(\theta)$ is nonsingular for $\theta^0 \in \Theta$, then there exists a neighborhood $N(\theta^0) \subset \Theta$ of θ^0 in which no θ is equivallent to θ^0.*

In deriving the COLS estimators we observed that the sample moments of least squares residuals are complicated functions of model parameters and due to symmetry of the standard normal distribution there is no way for us to uniquely express the later in terms of formers and the data. Hence, the normal-doubly trunctated normal model is not globally identified. The same holds for normal-truncated normal model. It is straightforward, however, to show the global identification of normal-half normal model. Thus, failing to obtain the results for global identification, we need to check the conditions for local identification. We do so for both models.

The log-likelihood function (14) for single observation is given by

[15] see Rothenberg 1971, pp 579 for the relevant definition of the regular point

$$
\begin{aligned}
\ln(L) \;=\; & -\ln[\Phi(\frac{B-\mu}{\sigma}(\lambda^{-2}+1)^{1/2}) - \Phi(\frac{-\mu}{\sigma}(\lambda^{-2}+1)^{1/2})] \\
& -\ln\sigma - \frac{1}{2}\ln(2\pi) - \frac{(\varepsilon_i+\mu)^2}{2\sigma^2} \\
& +\ln\{\Phi(\frac{(B+\varepsilon_i)\lambda + (B-\mu)\lambda^{-1}}{\sigma}) \\
& -\Phi(\frac{\varepsilon_i\lambda - \mu\lambda^{-1}}{\sigma})\}
\end{aligned} \tag{37}
$$

which for $B = \infty$ reduces to the log-likelihood function of normal-truncated-normal model

$$
\begin{aligned}
\ln(L) \;=\; & -\ln[\Phi(\frac{\mu}{\sigma}(\lambda^{-2}+1)^{1/2})] \\
& -\ln\sigma - \frac{1}{2}\ln(2\pi) - \frac{(\varepsilon_i+\mu)^2}{2\sigma^2} \\
& +\ln\{\Phi(\frac{\mu\lambda^{-1} - \varepsilon_i\lambda}{\sigma})\}
\end{aligned} \tag{38}
$$

The first order conditions for maximization of (37) with respect to its parameters are

$$
\frac{\partial \ln(L)}{\partial \beta} = \frac{(\varepsilon_i+\mu)x_i}{\sigma^2} + \frac{\lambda x_i}{\sigma}\frac{\phi(\Xi_4)-\phi(\Xi_3)}{\Phi(\Xi_3)-\Phi(\Xi_4)} \tag{39}
$$

$$
\frac{\partial \ln(L)}{\partial \sigma^2} = \frac{1}{2\sigma^2}\frac{[(\Xi_1\phi(\Xi_1)-\Xi_2\phi(\Xi_2)]}{\Phi(\Xi_1)-\Phi(\Xi_2)} - \frac{1}{2\sigma^2} + \frac{(\varepsilon_i+\mu)^2}{2\sigma^4} + \frac{1}{2\sigma^2}\frac{[\Xi_4\phi(\Xi_4)-\Xi_3\varphi(\Xi_3)]}{\Phi(\Xi_3)-\Phi(\Xi_4)} \tag{40}
$$

$$
\begin{aligned}
\frac{\partial \ln(L)}{\partial \lambda} \;=\; & \frac{(\lambda^{-2}+1)^{-1}}{\lambda^3}\frac{[(\Xi_1\phi(\Xi_1)-\Xi_2\phi(\Xi_2)]}{\Phi(\Xi_1)-\Phi(\Xi_2)} + \\
& \frac{1}{\sigma}\frac{[((B+\varepsilon_i)\lambda + (B-\mu)\lambda^{-2})\phi(\Xi_3) - (\varepsilon_i\lambda - \mu\lambda^{-2})\phi(\Xi_4)]}{\Phi(\Xi_3)-\Phi(\Xi_4)}
\end{aligned} \tag{41}
$$

$$
\frac{\partial \ln(L)}{\partial \mu} = \frac{(\lambda^{-2}+1)^{1/2}}{\sigma}\frac{\phi(\Xi_1)-\phi(\Xi_2)}{\Phi(\Xi_1)-\Phi(\Xi_2)} - \frac{(\varepsilon_i+\mu)}{\sigma^2} + \frac{1}{\lambda\sigma}\frac{\phi(\Xi_4)-\varphi(\Xi_3)}{\Phi(\Xi_3)-\Phi(\Xi_4)} \tag{42}
$$

$$\frac{\partial \ln(L)}{\partial B} = -\frac{(\lambda^{-2}+1)^{1/2}}{\sigma}\frac{\phi(\Xi_1)}{\Phi(\Xi_1)-\Phi(\Xi_2)} + (\frac{\lambda}{\sigma}+\frac{1}{\lambda\sigma})\frac{\phi(\Xi_3)}{\Phi(\Xi_3)-\Phi(\Xi_3)} \tag{43}$$

where $\Xi_1 = \frac{(B-\mu)}{\sigma}(\lambda^{-2}+1)^{1/2}$, $\Xi_2 = \frac{-\mu}{\sigma}(\lambda^{-2}+1)^{1/2}$, $\Xi_3 = \frac{(B+\varepsilon_i)\lambda+(B-\mu)\lambda^{-1}}{\sigma}$, and $\Xi_3 = \frac{\varepsilon_i\lambda-\mu\lambda^{-1}}{\sigma}$.

These give us five equations with five unknown parameters to be estimated. Again it can be seen that there is no closed-form and unique solution to these parameters.

We check the local identification of these two models by examining the Fisher's information matrix evaluated at a given parameter point. Ritter and Simar (1997) note that the distribution of the composite error in normal-gamma stochastic frontier model tends to the normal distribution as the shape parameter of the gamma distribution increases without bound and the scale parameter remains relatively low. In this case the model parameters and inefficiencies cannot be identified. This is also the case for normal-truncated-normal model, wherein for relatevily small values of parameter λ the distribution resembles the normal distribution as the mean of inefficiencies μ becomes relatively large. Therefore, this model fails to be locally identified in this particular case. On the other hand, the bounded inefficiency model is still capable to identify the model parameters even for large values of μ, since the existence of the bound will produce heavier right tails. We prove the local identification of the normal-doubly truncated normal model and in turn the local unidentifiabilty of the normal-truncated normal model for large μ.

The representation of these two cases is provided on below graphs, where the truncated normal density looks like the normal density for values of μ as low as 1, while keeping the variance of inrfficiencies to be half. On the other hand, doubly truncated distribution distinguishes itself from the normal distribution in the sense that its right tail will be shorter than its corresponding left tail. Henceforth, the model will be identified even for large values of the parameter μ. It is worth to mention that often the empirical distribution of inefficiencies looks like normal distribution. Therefore, without imposing the bound it is difficult to identify it from normally distributed noise term.

Claim: *There is a $\mu' \in R$ such that for any $\mu > \mu'$ the normal-truncated normal model is not locally identified. However, normal-doubly truncated normal model can still be identified.*

Proof. The Fisher's information matrix evaluated at $\mu > \mu'$ for the normal-trunctated normal is given by

$$I(\beta,\sigma^2,\lambda,\mu)=\begin{bmatrix} -\frac{1}{\sigma^2}\sum_{i=1}^n x_i x_i' & -\frac{1}{\sigma^4}\sum_{i=1}^n(\varepsilon_i+\mu)x_i & 0 & -\frac{1}{\sigma^2}\sum_{i=1}^n x_i \\ -\frac{1}{\sigma^4}\sum_{i=1}^n(\varepsilon_i+\mu)x_i & \frac{n}{2\sigma^4}-\frac{1}{\sigma^6}\sum_{i=1}^n(\varepsilon_i+\mu)^2 & 0 & -\frac{1}{\sigma^4}\sum_{i=1}^n(\varepsilon_i+\mu) \\ 0 & 0 & 0 & 0 \\ -\frac{1}{\sigma^2}\sum_{i=1}^n x_i & -\frac{1}{\sigma^4}\sum_{i=1}^n(\varepsilon_i+\mu) & 0 & -\frac{n}{\sigma^2} \end{bmatrix}$$

which is clearly singular. Hence, the model is not locally identified.

On the other hand, the information matrix evaluated at $\mu > \mu'$ for the normal-doubly truncated normal model is given by

$$I(\beta,\sigma^2,\lambda,\mu,B)=\begin{bmatrix} \frac{\partial^2\ln(L)}{\partial\beta\partial\beta'} & \frac{\partial^2\ln(L)}{\partial\beta\partial\sigma^2} & \frac{\partial^2\ln(L)}{\partial\beta\partial\lambda} & \frac{\partial^2\ln(L)}{\partial\beta\partial\mu} & \frac{\partial^2\ln(L)}{\partial\beta\partial B} \\ & \frac{\partial^2\ln(L)}{\partial(\sigma^2)^2} & \frac{\partial^2\ln(L)}{\partial\sigma^2\partial\lambda} & \frac{\partial^2\ln(L)}{\partial\sigma^2\partial\mu} & \frac{\partial^2\ln(L)}{\partial\sigma^2\partial B} \\ & & \frac{\partial^2\ln(L)}{\partial\lambda^2} & \frac{\partial^2\ln(L)}{\partial\lambda\partial\mu} & \frac{\partial^2\ln(L)}{\partial\lambda\partial B} \\ & & & \frac{\partial^2\ln(L)}{\partial\mu^2} & \frac{\partial^2\ln(L)}{\partial\mu\partial B} \\ & & & & \frac{\partial^2\ln(L)}{\partial B^2} \end{bmatrix}$$

where the elements of this matrix are given by the corresponding second-order derivatives of the log-likelihood function

$$\frac{\partial^2\ln(L)}{\partial\beta\partial\beta'}=-\frac{1}{\sigma^2}\sum_{i=1}^n x_i x_i'$$ [16]

$$\frac{\partial^2\ln(L)}{\partial\beta\partial\sigma^2}=-\frac{1}{\sigma^4}\sum_{i=1}^n(\varepsilon_i+\mu)x_i$$

$$\frac{\partial^2\ln(L)}{\partial\beta\partial\lambda}=0$$

$$\frac{\partial^2\ln(L)}{\partial\beta\partial\mu}=-\frac{1}{\sigma^2}\sum_{i=1}^n x_i$$

$$\frac{\partial^2\ln(L)}{\partial\beta\partial B}=0$$

$$\begin{aligned}\frac{\partial^2\ln(L)}{\partial(\sigma^2)^2} &= -n\{\frac{3(B-\mu)(\lambda^{-2}+1)^{1/2}[\sigma-\frac{1}{3}(B-\mu)(\lambda^{-2}+1)^{1/2}\Xi_1]}{4\sigma^6}\frac{\phi(\Xi_1)}{\Phi(\Xi_1)} \\ &\quad -\frac{(B-\mu)^2(\lambda^{-2}+1)}{4\sigma^6}\frac{\phi^2(\Xi_1)}{\Phi^2(\Xi_1)}\}+\frac{n}{2\sigma^4}-\frac{1}{\sigma^6}\sum_{i=1}^n(\varepsilon_i+\mu)^2\end{aligned}$$

$$\frac{\partial^2\ln(L)}{\partial\sigma^2\partial\lambda}=-n\{\frac{(B-\mu)(\lambda^{-2}+1)^{-1/2}[1-\frac{(B-\mu)}{\sigma}(\lambda^{-2}+1)^{1/2}\Xi_1]}{2(\lambda\sigma)^3}\frac{\phi(\Xi_1)}{\Phi(\Xi_1)}-\frac{(B-\mu)^2}{2\sigma^4\lambda^3}\frac{\phi^2(\Xi_1)}{\Phi^2(\Xi_1)}\}$$

[16] Note that large values of μ are followed by large values of bound so it can be assumed that $\Xi_3\phi(\Xi_3)\to 0$ as well

$$\frac{\partial^2 \ln(L)}{\partial\sigma^2\partial\mu} = -n\{\frac{(\lambda^{-2}+1)^{1/2}[1-(B-\mu)(\lambda^{-2}+1)^{1/2}\Xi_1]}{2\sigma^3}\frac{\phi(\Xi_1)}{\Phi(\Xi_1)} - \frac{(B-\mu)(\lambda^{-2}+1)}{2\sigma^4}\frac{\phi^2(\Xi_1)}{\Phi^2(\Xi_1)}\} - \frac{1}{\sigma^4}\sum_{i=1}^{n}(\varepsilon_i+\mu)$$

$$\frac{\partial^2 \ln(L)}{\partial\sigma^2\partial B} = n\{\frac{(\lambda^{-2}+1)^{1/2}[1-(B-\mu)(\lambda^{-2}+1)^{1/2}\Xi_1]}{2\sigma^3} - \frac{(B-\mu)(\lambda^{-2}+1)}{2\sigma^4}\frac{\phi^2(\Xi_1)}{\Phi^2(\Xi_1)}\}$$

$$\frac{\partial^2 \ln(L)}{\partial\lambda^2} = -n\{[\frac{6(2\lambda^{-6}+1)}{\lambda^{-7}} - \frac{(\lambda^{-6}+1)}{\lambda^6}(\frac{B-\mu}{\sigma})\Xi_1]\frac{\phi(\Xi_1)}{\Phi(\Xi_1)} + [\frac{(B-\mu)(\lambda^{-6}+1)}{\sigma\lambda^6}]^2\frac{\phi^2(\Xi_1)}{\Phi^2(\Xi_1)}\}$$

$$\frac{\partial^2 \ln(L)}{\partial\lambda\partial\mu} = -n\{\frac{(\lambda^{-2}+1)^{-1/2}[1-\frac{(B-\mu)}{\sigma}(\lambda^{-2}+1)^{1/2}\Xi_1]}{\sigma\lambda^3}\frac{\phi(\Xi_1)}{\Phi(\Xi_1)} + \frac{(B-\mu)}{\sigma^2\lambda^3}\frac{\phi^2(\Xi_1)}{\Phi^2(\Xi_1)}\}$$

$$\frac{\partial^2 \ln(L)}{\partial\lambda\partial B} = n\{\frac{(\lambda^{-2}+1)^{-1/2}}{\sigma\lambda^3}\frac{\phi(\Xi_1)}{\Phi(\Xi_1)}[1 - \frac{(B-\mu)}{\sigma}\frac{\phi(\Xi_1)}{\Phi(\Xi_1)}\}$$

$$\frac{\partial^2 \ln(L)}{\partial\mu^2} = n\{\frac{(\lambda^{-2}+1)\phi(\Xi_1)[\Xi_1\Phi(\Xi_1)+\phi(\Xi_1)]}{\sigma^2\Phi^2(\Xi_1)} - \frac{n}{\sigma^2}$$

$$\frac{\partial^2 \ln(L)}{\partial\mu\partial B} = -n\{\frac{(\lambda^{-2}+1)\phi(\Xi_1)[\Xi_1\Phi(\Xi_1)+\phi(\Xi_1)]}{\sigma^2\Phi^2(\Xi_1)}\}$$

$$\frac{\partial^2 \ln(L)}{\partial B^2} = n\{\frac{(\lambda^{-2}+1)\phi(\Xi_1)[\Xi_1\Phi(\Xi_1)-\phi(\Xi_1)]}{\sigma^2\Phi^2(\Xi_1)}\}$$

For certain bounds and by the property of the standard normal distribution, which assumes the existency of the bound whenever its deviation from the mean does not exceed three standard deviations[17], the determinant of the Fisher's information matrix is not zero, as long as the signal-to-noise ratio, λ, is relatively high. As $\lambda \to 0$, both models fail to be identified.

□

In sum, both normal-doubly truncated normal SFM and normal-truncated normal SFM fail to be globally identified, but are locally identified. However, if the true

[17] Otherwise we will reproduce the truncated normal distribution

density of inefficiencies has a large mean, the later model fails to be even locally identified, but the former model can still yield more precise and stable estimates of parameters and inefficiencies. Still it is not clear how can the statistical inference be validated in locally identified case. Shapiro (1986) and Dasgupta et al. (2007) discuss some cases there valid statistical inference can be obtained. Bayesian method could be utilized to identify the parameters through nonlinear constraints imposed by the skewness condition. We leave this possibility for the future work.

Table1: Proportion of Positive Skewness in Normal-Doubly Truncated Normal Model with $\mu = 1$

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

- The first column shows that the proportion of the samples with the positive ("wrong") skewness increases as the sample size grows larger. It converges to one as the variance of the one-sided inefficiency term becomes larger relative to the variance of two-sided error.
- In the second column we have the case where $B = 2\mu$. Under this case there is about a 50-50 chance that we generate a sample with positive skewness. In most of the cases, the positive skewness appears to be statistically insignificant.
- The third column presents the case where we have a positively skewed distribution of inefficiencies and as in Simar and Wilson (2009) the proportion decreases as the sample size and parameter λ increase.

References

- Aigner, D.J., Chu, S., 1968, On estimating the industry production function. American Economic Review 58, 826-839.

- Aigner, D.J., Lovell , C.A.K., Schmidt, P., 1977, Formulation and estimation of stochastic frontier models. Journal of Econometrics 6, 21-23.

- Battese, G.E., Cora, G.S., 1977. Estimation of a production frontier model: with application to the pastoral zone of eastern Australia. Australian Journal of Agricultural Economics 21, 169-179.

- Bera A.K., Premaratne, G., 2001, Adjusting the tests for skewness and kurtosis for distributional misspecifications. Working Paper.

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

- Charnes, A., Cooper, W.W., Rhodes, E.L., 1978, Measuring the efficiency of decision making units. European Journal of Operational Research 2, 429-444.

- Coeli, T., 1995, Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. Journal of Productivity Analysis 6, 247-268.

- Coeli, T., 1996, A guide to FRONTIER version 4.1: A computer program for stochastic frontier production and cost function estimation. CEPA working paper #96/07, Centre for Efficiency and Productivity Analysis, University of New England, Arimidale, NSW 2351, Australia.

- Cornwell, C., P. Schmidt and R. Sickles. Production Frontiers with Cross Sectional and Time Series Variation in Efficiency Levels. Journal of Econometrics, 1990, 46, pp. 185-200.

- Cornwell, C., P. Schmidt., 1996, Production Frontier and Efficiency Measurement, in L. Matyas and P. Sevestre (eds), Econometrics of Panel Data: Handbook of Theory and Application, Kluwer Academic Publishers.

- Dasguptaa, A., Steven, G., Selfb, and Das Gupta, S., 2007, Non-identifiable parametric probability models and reparametrization. Journal of Statistical Planning and Inference 137 (2007) 3380 – 3393

- D'Agostino, R.B., Pearson, E.S., 1973, Tests for departure from normality: Empirical results for the distribution of b_2 and $\sqrt{b_1}$. Biometrica 60, 613-622.

- Deprins, D., Simar, L., and Tulkens, H. ,1984, Measuring Labor Inefficiency in Post Offices. In the Performance of Public Enterprises: Concepts and Measurements. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243-267.

- Deprins, D., & Simar, L., 1989. "Estimation de frontieres diterministes avec facteurs exogunes d'inefficaciti," Annales d'Economie et de Statistique, ADRES, issue 14, pages 06, Avril-Jui.

- 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., 1993. "Frontier Production Functions," Working Papers 93-20, New York University, Leonard N. Stern School of Business, Department of Economics.

- Greene, W. H., 1995, LIMDEP Version 7.0 User's Manual, New York: Econometric Software, Inc.

- 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.

- 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.

- Johnson, N.L., S. Kotz and A.W. Kemp, 1992, Univariate Discrete Distributions, second edition, New York: John Wiley & Sons.

- Johnson, N.L., S. Kotz and N. Balakrishnan, 1994, Continuous Univariate Distributions, Vol. 1, second edition, New York: John Wiley & Sons.

- Kim, Y., Schmidt, P., 2000, A review and empirical comparison of bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data. Journal of Productivity Analysis 14, 91-118.

- Koop, G., 1994, "Recent Progress in Applied Bayesian Econometrics. Journal of Economic Surveys, 8, 1-34.

- Koop, G., 2001, Cross-sectorla paterns of efficiency and technological change in manufacturing. International Economic Review 42, 73-103.
- Koop, G., Osiewalski, J., and Steel, M. F. J., 1995, Posterior Analysis of Stochastic Frontier models Using Gibbs Sampling. Computational Statistics, 10, 353-373.
- Koop, G., Osiewalski, J., and Steel, M. F. J., 1997, Bayesian Efficiency Analysis Through Individual Effects: Hospital Cost Frontiers. Journal of Econometrics, 76, 77-105.
- Kumbhakar, S., 1990, Production Frontiers and Panel Data, and Time Varying Technical Inefficiency. Journal of Econometrics, 46, pp. 201-11.
- Kumbhakar, S., 1991, Estimation of technical inefficiency in panel data models with firm and time specific effects. Economic Letters, 36, 43-48.
- Kumbahckar, S.C., Lovell, C.A.K., 2000, Stochastic Frontier Analysis, Cambridge. Cambridge University Press.
- Leamer, E.E., 1978, Specification Searches: Ad hoc inference with nonexperimental data. New York, John Wiley and Sonc, Inc.
- Meeusen, W., van den Broeck , J., 1977, Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18, 435-444.
- Migon, H. S., Medici, E., 2001, A Bayesian Approach to Stochastic Production Frontier, Technical Report n. 105, DME-IM/UFRJ.
- 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.
- Qian, J., R. C. Sickles, 2008, Stochastic frontiers with bounded inefficiency. Working paper.
- Ritter, C., and Simar, L., 1997, Pitfalls of Normal-Gamma Stochastic Frontier Models. Journal of Productivity Analysis, 8, 167–182.
- Sena V., 1999., Stochastic frontier estimation: A review of the software options. Journal of Applied Econometrics 14, 579-586.
- Sickles, R.C., 2005, Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings. Journal of Econometrics 126, 305-334.

- Simar and Wilson, 2009, Estimation and Inference in Cross-Sectional,Stochastic Frontier Models, forthcoming in Econometric Reviews.
- Schmidt, P., Lin, T., 1984, Simple tests of alternative specifications in stochastic frontier models. Journal of Econometrics 24, 349-361.
- Schmidt, P., and R. Sickles. "Production Frontiers with Panel Data." Journal of Business and Economic Statistics, 1984, 2, pp. 367-74.
- Shapiro, A., 1986. Asymptotic theory of overparameterized structural models. Journal of Amererican Statistical Association 81, 142–149.
- Stevenson, R.E., 1980, Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13, 57-66.
- van den Broeck, J., Koop, G., Osiewalski, J., 1994, Stochastic frontier models: a Bayesian perspective. Journal of Econometrics 61, 273-303.
- Waldman, D., 1982, A stationary point for the stochastic frontier likelihood. Journal of Econometrics 18, 275-279.
- Waldman, D., 1984, Properties of technical efficiency estimators in the stochastic frontier model. Journal of Econometrics 25, 353-364.
- Wang, W. S., Schmidt, P., 2008, On the distribution of estimated technical efficiency in stochastic frontier models. Journal of Econometrics, 148, 36-45.