I N S T I T U T D E S T A T I S T I Q U E

B I O S T A T I S T I Q U E E T

S C I E N C E S A C T U A R I E L L E S

(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2013/46

The "wrong skewness" problem in stochastic frontier models: A new approach

HAFNER, C., MANNER, H. and L. SIMAR

The "wrong skewness" problem in stochastic frontier models: A new approach

Christian M. Hafner*[1], Hans Manner[2], and Léopold Simar[1]

[1]*Institute of statistics, biostatistics and actuarial sciences, Université Catholique de Louvain, Belgium*

[2]*Department of Social and Economic Statistics, University of Cologne, Germany*

October 8, 2013

Abstract

Stochastic frontier models are widely used to measure, e.g., technical efficiencies of firms. The classical stochastic frontier model often suffers from the empirical artefact that the residuals of the production function may have a positive skewness, whereas a negative one is expected under the model, which leads to estimated full efficiencies of all firms. We propose a new approach to the problem by generalizing the distribution used for the inefficiency variable. This generalized stochastic frontier model allows the sample data to have the wrong skewness while estimating well-defined and non-degenerate efficiency measures. We discuss the properties of the model and its maximum likelihood estimator, and provide a simulation study to show that our model delivers estimators of efficiency with smaller mean squared error than those of the classical model even if the population skewness has the correct sign. Finally, we apply the model to data of the U.S. textile industry for 1958-2005, and show that for a number of years our model suggests efficiencies well below the frontier, while the classical one estimates no inefficiency in those years.

Keywords: Stochastic frontier model, production efficiency, skewness

JEL Classification: C13, C18, D24.

*Corresponding author, Institute of statistics, biostatistics and actuarial sciences, and CORE, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, email: christian.hafner@uclouvain.be. Ch. Hafner and L. Simar acknowledge research support by IAP Research Network P7/06 of the Belgian State (Belgian Science Policy).

1 Introduction

One of the most popular econometric models to estimate production frontier and firm efficiency is the parametric stochastic frontier model (SFM). The basic model was introduced by Aigner et al. (1977) and Meeusen and van den Broek (1977). The model assumes some functional form for the frontier which represents the locus of maximal achievable output $Y \in \mathbb{R}$ (production) for a given set of inputs $X \in \mathbb{R}^p$ (production factors such as labor, energy, capital, etc.). Since not all firms are efficient, we observe production plans (x_i, y_i) that may lie below this optimal frontier. The interesting feature of SFM (as opposed to deterministic frontier models) is that the model allows also the presence of the usual random noise. Thus, the error term in a SFM is a convolution of two terms: a one-sided inefficiency term plus a classical symmetric statistical noise, usually modeled by a normal distribution.

Several one-sided distributions have been proposed in the literature for the inefficiencies. The pioneering work of Aigner et al. (1977) suggests the use of an exponential or of a half-normal distribution. Other choices, e.g., two-parameter distributions such as the gamma (Greene 1990) or the truncated-normal (Stevenson 1980), have been proposed, see Kumbhakar and Lovell (2000), or Greene (2007) for detailed surveys. All of these one-sided distributions have a positive skewness, so Li (1996) considers the case of a uniform distribution and Carree (2002) a negative binomial allowing negative skewness. In the same spirit, Qian and Sickles (2009) and Almanidis and Sickles (2011) consider a double truncated normal distribution for the inefficiencies. The latter three approaches assume that the inefficiency term is bounded above and below.

Typically the basic model can be written as

$$Y_i = \alpha_0 + \alpha' X_i + W_i, \ i = 1, \ldots, n, \tag{1}$$

where $W_i = V_i - U_i$ with $V_i \sim N(0, \sigma_v^2)$ and U_i has one-sided parametric distribution on $\mathbb{R}_+$. We assume independence between V_i and U_i which are both i.i.d.[1]

It is well known (see, e.g., Greene 1990) that the third moment of W_i is given by

$$\mathrm{E}\left[(W_i - \mathrm{E}W_i)^3\right] = -\mathrm{E}\left[(U_i - \mathrm{E}(U_i))^3\right], \tag{2}$$

so that a positive skewness for U_i implies a negative skewness for W_i. A simple estimator of the parameters of the model is given by the Modified OLS (MOLS) approach (Olson

[1]In this paper we consider the production frontier case, but this can be easily translated to a cost frontier model where the error w_i would have the form $v_i + u_i$, $u_i \geq 0$ accounting for cost inefficiencies.

et al. 1980, Greene 1990) where a simple OLS procedure leads to consistent estimators of the slope parameters α of the following shifted model

$$Y_i = \alpha_0^* + \alpha' X_i + \varepsilon_i, \; i = 1, \ldots, n, \tag{3}$$

where $\alpha_0^* = \alpha_0 - \mathrm{E}(U_i)$ so that $\varepsilon_i = W_i + \mathrm{E}(U_i) = 0$.

Then the moments of OLS residuals are used to estimate the parameters of the distributions of U_i and V_i. If these distributions involve two unknown parameters (as in the normal/half-normal or normal/exponential cases), only the second and third empirical moments of $\widehat{\varepsilon}_{i,\mathrm{OLS}}$ are needed (one additional moment of higher order is needed if the distributions of u_i and v_i involve three unknown parameters, see Kumbhakar and Lovell 2000 for details).

From (2), it is clear that $\widehat{\mu}_{3,n} = n^{-1} \sum_{i=1}^{n} \widehat{\varepsilon}^3_{i,\mathrm{OLS}}$ is a consistent estimator of the negative of the third moment of U_i, which gives the sign of the skewness of U_i. It is well known and illustrated by numerous Monte-Carlo experiments (see Olson et al. 1980, or Simar and Wilson 2010) that very often, in finite samples, the sign of $\widehat{\mu}_{3,n}$ is positive, even though the opposite is expected. In this literature, researchers say that they observe the "wrong" skewness when the sign of the empirical skewness is positive. The consequence of a "wrong" skewness, as shown, e.g., by Waldman (1982), is that the MOLS and the MLE estimates of the slope are identical to the OLS slope, and there are no inefficiencies: the mean and the variance of U_i are estimated at zero and all the firms are supposed to be efficient, i.e. lying on the estimated frontier.

Long debates have appeared about this issue, see Carree (2002) and Almanidis and Sickles (2011) and the references therein for details. To summarize, the question whether the skewness is "wrong" or not is perhaps a misleading debate. The OLS residuals are what they are, and the wrong sign of the skewness is indeed unexpected when u_i has a positive skewness.

For these reasons, solutions have been proposed for choosing distributions for U_i allowing also negative skewness, for example, the negative binomial of Carree (2002) and the double truncated normal of Qian and Sickles (2009). While these approaches have their merits, there are potential drawbacks. In particular, first they provide distributions with bounded support, and second they do not nest the classical models (such as the normal/half-normal or normal/exponential). The traditional SFM may be correct but we are observing the unexpected skewness just by analyzing an unlucky sample. This is annoying as it plagues the estimation of the inefficiencies. The contribution of this paper is to extend the classical SFM to a model allowing the opposite skewness, but still nesting the traditional SFM. This is important because, when observing the "wrong" skewness, most researchers are tempted to believe that the model is wrong, and we know that even a correct SFM allowing inefficient firms may produce the wrong sign for the skewness. This happens more often with small sample sizes or when the ratio $\text{Var}(V)/\text{Var}(U)$ increases (see Simar and Wilson 2010 for a careful Monte-Carlo investigation).

The remainder of the paper is organized as follows. The following section presents the model, discusses its properties and its estimation. Section 3 reports results of a simulation study to illustrate the usefulness of the proposed model even in situations where the true model is the classical one. Finally, in Section 4 we apply the model to analyze the efficiency of sub-sectors of the U.S. textile industry for 1958-2005.

2 The model

Let us first recall the classical stochastic frontier model, starting from the basic model (1),

$$Y = \alpha_0 + \alpha' X + W, \tag{4}$$

where $W = V - U$, and U and V are independent random variables, the former representing inefficiency, and the latter statistical noise, which we assume is given by $V \sim N(0, \sigma_v^2)$. The positive random variable U is linked to the notion of inefficiency. The typical assumptions on the distribution of U imply that U has a positive skewness and W has negative skewness, which often leads to incompatibility with data when the sample skewness of residuals w is positive. Mean inefficiency in the basic model is usually measured by $\text{E}[\exp(-U)]$, and inefficiency for a given firm by

$$\tau_c = \text{E}[\exp(-U)|W], \tag{5}$$

such that $\tau_c \in [0, 1]$ by construction.

The classical normal- halfnormal SFM assumes that U has a halfnormal density given by

$$h_\mu(u) = \frac{2}{\mu\sqrt{\pi/2}} \phi\left(\frac{u}{\mu\sqrt{\pi/2}}\right), \quad u \geq 0. \tag{6}$$

where $\mu > 0$ is the expectation of U, and $\sigma_u = \mu\sqrt{\pi/2} > 0$ its standard deviation. The density of W is given by

$$g(w) = \frac{2}{\sigma}\phi\left(\frac{-w}{\sigma}\right)\Phi\left(\frac{-w}{\sigma}\,\frac{\mu\sqrt{\pi/2}}{\sigma_v}\right) \tag{7}$$

where $\sigma^2 = \mu^2\pi/2 + \sigma_v^2$ and $\Phi(\cdot)$ is the cdf of a standard normal random variable, see, e.g., Kumbhakar and Lovell (2000, p.75). They also give expressions for the conditional expectation of U and the inefficiency:

$$U|W = w \sim N^+(\mu_*, \sigma_*^2)) \tag{8}$$

$$\mathrm{E}(U|W = w) = \sigma_* \left[A + \frac{\phi(A)}{\Phi(A)}\right]) \tag{9}$$

$$\mathrm{E}(\exp(-U)|W = w) = [1 - \Phi(\sigma_* - A)]\,[\Phi(A)]^{-1} \exp(-\mu_* + \sigma_*^2/2)) \tag{10}$$

where $\mu_* = -w\mu^2\pi/(2\sigma^2)$, $\sigma_*^2 = \mu^2\sigma_v^2\pi/(2\sigma^2)$ and $A = \mu_*/\sigma_*$. These expressions can then be used for maximum likelihood estimation of the parameters and of the inefficiency measures.

Other distributions such as the exponential have been used for U, but the normal-halfnormal can be considered as the benchmark model. In the following, we are first proposing a general framework that includes the classical stochastic frontier model as a special case, and then give examples of extended normal-halfnormal and exponential distributions.

2.1 The extended stochastic frontier model

A first modification of the basic model would be to maintain the production model (4), but to replace the term $W = V - U$ by the following composed term:

$$W = V - U + 2\mathrm{E}[U]I(\mu < 0) \tag{11}$$

where $\mu \in \mathbb{R}$ is the expectation of U, $I(\cdot)$ is the indicator function, and the distribution of U depends on the sign of μ: If $\mu > 0$, then U has density with positive support, while if $\mu < 0$, then it has density with negative support.

Obviously, model (11) reduces to the classical SFA model if $\mu > 0$. However, if $\mu < 0$, the model is $W = V - U + 2\mathrm{E}[U]$, and the correction term $2\mathrm{E}[U]$ is shifting the distribution of $-U$ such that the mean of $-U + 2\mathrm{E}[U]$ with $\mu < 0$ is the same as the mean of $-U$ in the classical case with $\mu > 0$, namely $-|\mu|$ in both cases. Therefore, the mean of W only depends on the size of μ, not its sign. The size of μ is related to measures of inefficiency, whereas the sign of μ gives flexibility to fit distributional properties of the data, such as positive or negative skewness. Full efficiency is attained when $\mu = 0$, in which case U degenerates to a one-point distribution at zero and $W \sim N(0, \sigma_v^2)$ is symmetric. Efficiency is defined as a natural extension of (5),

$$\tau_I = \mathrm{E}[\exp(-U + 2\mathrm{E}[U]I(\mu < 0))|W], \tag{12}$$

which collapses to τ_c if $\mu > 0$. Note that, if $\mu < 0$, then U has negative support and there would be a small probability that τ_i is larger than one. Firms would be considered as super-efficient in this case, which may happen temporarily in situations of technological innovation as outlined by Carree (2002). This case is discussed in detail in Section 2.3.

Model (11) can be viewed as a threshold model with threshold at zero. Rather than a threshold, one might prefer a smooth transition between the two regimes $\mu > 0$ and $\mu < 0$ which would give additional flexibility. This could be achieved by the following extension:

$$W = V - U + 2\mathrm{E}[U](1 - G_\gamma(\mu)) \tag{13}$$

where $\mu \in \mathbb{R}$ is a parameter linked to the distribution of U, to be specified below, and $G_\gamma(\mu)$ is the logistic function defined by $G_\gamma(\mu) = (1 + \exp(-\gamma\mu))^{-1}$. In the limiting case $\gamma \to \infty$, for which $G_\gamma(\mu)$ converges to the indicator function $I(\mu > 0)$, so that the above model (11) is recovered as a special case.

We now define efficiency as a natural extension of (5) to be given by

$$\tau_\gamma = \mathrm{E}[\exp(-U + 2\mathrm{E}[U](1 - G_\gamma(\mu)))|W], \tag{14}$$

which collapses to τ_I if $\gamma \to \infty$, and to τ_c if $\mu > 0$ and $\gamma \to \infty$. Hence, this measure includes the classical efficiency measure as a special case. For $\gamma < \infty$, there is again a small probability of obtaining super-efficiency, i.e., $\tau_\gamma > 1$, both for positive and negative μ, which will be discussed in Section 2.3.

For the inefficiency variable U we assume that it has density given by

$$f_\mu(u) = h_\mu(u)\left\{G_\gamma(\mu)I(u \geq 0) + (1 - G_\gamma(\mu))I(u < 0)\right\} \tag{15}$$

where $I(\cdot)$ is the indicator function, and $h_\mu(u)$ is a function defined on $\mathbb{R}$ with the following properties:

$$\begin{aligned}
h_\mu(u) &\geq 0 && (16)\\
h_\mu(u) &= h_\mu(-u) && (17)\\
h_\mu(u) &= h_{-\mu}(u) && (18)\\
\int_0^\infty h_\mu(u)du &= 1 && (19)\\
\int_0^\infty u h_\mu(u)du &= \mu(2G_\gamma(\mu) - 1). && (20)
\end{aligned}$$

Equation (17) requires that h is an even function of u. The remaining conditions ensure that h is a density on $(0, \infty)$ with parameter μ and expectation $\mu(2G_\gamma(\mu) - 1)$.

Note that $f_\mu(u)$ is a well defined density function since $f_\mu(u) \geq 0$ and

$$\begin{aligned}
\int_{-\infty}^{\infty} f_\mu(u)du &= (1-G_\gamma(\mu))\int_{-\infty}^{0} h_\mu(u)du + G_\gamma(\mu)\int_{0}^{\infty} h_\mu(u)du \\
&= (1-G_\gamma(\mu))\underbrace{\int_{0}^{\infty} h_\mu(u)du}_{=1} + G_\gamma(\mu)\underbrace{\int_{0}^{\infty} h_\mu(u)du}_{=1} \\
&= 1 - G_\gamma(\mu) + G_\gamma(\mu) = 1
\end{aligned}$$

Moreover, note that by construction, $\mathrm{E}[U] = \mu(2G_\gamma(\mu)-1)^2$, since

$$\begin{aligned}
\int_{-\infty}^{\infty} uf_\mu(u)du &= (1-G_\gamma(\mu))\int_{-\infty}^{0} uh_\mu(u)du + G_\gamma(\mu)\int_{0}^{\infty} uh_\mu(u)du \\
&= -(1-G_\gamma(\mu))\underbrace{\int_{0}^{\infty} uh_\mu(u)du}_{=\mu(2G_\gamma(\mu)-1)} + G_\gamma(\mu)\underbrace{\int_{0}^{\infty} uh_\mu(u)du}_{=\mu(2G_\gamma(\mu)-1)} \\
&= \mu(2G_\gamma(\mu)-1)^2
\end{aligned}$$

Hence, the third term in (13) is given by

$$a(\mu) := 2\mathrm{E}[U](1-G_\gamma(\mu)) = 2\mu(2G_\gamma(\mu)-1)^2(1-G_\gamma(\mu)), \tag{21}$$

and

$$\mathrm{E}[W] = -\mathrm{E}[U] + a(\mu) = -\mu(2G_\gamma(\mu)-1)^3. \tag{22}$$

Due to the symmetry assumption (17), if h has positive skewness on $(0,\infty)$, then it will have negative skewness on $(-\infty, 0)$ with $\mu < 0$. Suppose that h has positive skewness on $(0,\infty)$, which is the typical case. Then W would have negative skewness if $\mu > 0$ and positive skewness if $\mu < 0$. Hence, inefficiency and skewness are no longer directly linked, and one can have, even asymptotically, inefficiency and positive skewness, which is not possible in the classical model. Another advantage of the extended model compared with the classical model is that convergence problems of maximum likelihood estimators in cases of positive sample skewness are avoided.

The following lemma gives the form of the density of W, used, e.g., for maximum likelihood estimation.

Lemma 1 *The density of W has the following form:*

$$g(w) = G_\gamma(\mu)g^+(-w+a(\mu)) + (1-G_\gamma(\mu))g^+(w-a(\mu)) \tag{23}$$

where

$$g^+(z) = \frac{1}{\sigma_v}\int_0^\infty h_\mu(u)\phi(\frac{z-u}{\sigma_v})du. \tag{24}$$

and $\phi(\cdot)$ is the standard normal density function.

The function $g^+(\cdot)$ is typically known or can be calculated. In the limiting case $\gamma \to \infty$ and $\mu > 0$, we have $G_\gamma(\mu) = 1$, $a(\mu) = 0$, and the density of w would simplify to $g(w) = g^+(-w)$.

If $h_\mu(u)$ is twice continuously differentiable w.r.t. μ, then so is $g(w)$ and the likelihood function, and standard MLE theory applies. To test the null hypothesis $H_0 : \mu = 0$, the classical likelihood ratio statistic follows a chi-square distribution asymptotically. However, for the limiting case $G_\gamma(\mu) = I(\mu > 0)$, the likelihood function is not continuously differentiable at zero, and hence the LR test follows non-standard asymptotics as in Lee (1993).

2.2 Examples

In the following we give some examples of distributions that are often used in practice.

2.2.1 The extended normal-halfnormal distribution

Suppose U has density given by (15) with $h_\mu(u)$ given by

$$h_\mu(u) = \frac{2}{\tilde{\mu}\sqrt{\pi/2}}\phi\left(\frac{u}{\tilde{\mu}\sqrt{\pi/2}}\right) \tag{25}$$

where $\tilde{\mu} = \mu(2G_\gamma(\mu) - 1)$.

For the case $\gamma \to \infty$ and $\mu \geq 0$, U has a positive half-normal distribution with support $(0, \infty)$ with parameter $\sigma_u = \mu\sqrt{\pi/2} > 0$. This is the classical normal-half-normal SFA model. If $\gamma \to \infty$ and $\mu < 0$, then U has a negative half-normal distribution with support $(-\infty, 0)$ and parameter $\sigma_u = -\mu\sqrt{\pi/2} > 0$.

In general, we have $\tilde{\mu} = \sqrt{2/\pi}\sigma_u$, and hence $\mathrm{E}[U] = \tilde{\mu}(2G_\gamma(\mu) - 1) = (2G_\gamma(\mu) - 1)\sqrt{2/\pi}\sigma_u$.

The density of W can be computed using (23) with the expression for $g^+(z)$ given by

$$g^+(z) = \frac{2}{\sigma}\phi\left(\frac{z}{\sigma}\right)\Phi\left(\frac{z}{\sigma}\frac{\tilde{\mu}\sqrt{\pi/2}}{\sigma_v}\right) \tag{26}$$

where $\sigma^2 = \tilde{\mu}^2\pi/2 + \sigma_v^2$ and $\Phi(\cdot)$ is the cdf of a standard normal random variable. For the classical case $\mu \geq 0$ and $G_\gamma(\mu) = I(\mu > 0)$, this density has been given, e.g., by Kumbhakar and Lovell (2000, p. 75).

Note that this distribution is closely related to the skewed normal distribution of Azzalini (1985). A random variable Z is said to follow a skewed-normal distribution, $Z \sim SN(\lambda)$, if its density function is given by $\psi(z;\lambda) = 2\phi(z)\Phi(\lambda z)$, where $\lambda \in \mathbb{R}$, $\phi(\cdot)$ and $\Phi(\cdot)$ are pdf and cdf, respectively, of the standard normal distribution. Consider the limiting case $\gamma \to \infty$. If $\mu > 0$, then $g(w) = g^+(-w) = \frac{1}{\sigma}\psi(-\frac{w}{\sigma};\lambda) = \frac{1}{\sigma}\psi(\frac{w}{\sigma};-\lambda)$ with $\lambda = \mu\sqrt{\pi/2}/\sigma_v$. Thus, in this case, W follows a scaled skewed-normal distribution with negative skewness. If $\mu < 0$, then $g(w) = g^+(w-2\mu) = \frac{1}{\sigma}\psi(\frac{w-2\mu}{\sigma};\lambda)$ with $\lambda = -\mu\sqrt{\pi/2}/\sigma_v > 0$. In this case, W follows a scaled and shifted skewed normal distribution with positive skewness.

It can also be checked using the expression for the mean of a skewed-normal distribution given by Azzalini (1985) that $\mathrm{E}[W]$ only depends on the absolute value of μ, not on its sign. In particular, if $Z \sim SN(\lambda)$, then $\mathrm{E}[Z] = \sqrt{2/\pi}\frac{\lambda}{\sqrt{1+\lambda^2}}$, and straightforward calculations yield $\mathrm{E}[W] = -|\mu|$, which corresponds to the limiting case of (22) for $\gamma \to \infty$.

Computation of inefficiencies The conditional density of $U|W$ can be calculated as $p(u|w) = f(u,w)/g(w)$, with joint density $f(u,w)$ and marginal density $g(w)$ given by

$$\begin{aligned} f(u,w) &= \frac{1}{\tilde{\mu}\sigma_v\sqrt{\pi^3/2}} \exp\left\{-\frac{u^2}{\tilde{\mu}^2\pi} - \frac{1}{2}\left(\frac{w+u-a(\mu)}{\sigma_v}\right)^2\right\} H(u) \\ g(w) &= G_\gamma(\mu)g^+(-w+a(\mu)) + (1-G_\gamma(\mu))g^+(w-a(\mu)) \end{aligned}$$

where $g^+(\cdot)$ is given in (26) and $H(u) = G_\gamma(\mu)I(u \geq 0) + (1-G_\gamma(\mu))I(u<0)$. Using the density $p(u|w)$, we can calculate technical efficiency $\mathrm{E}[\exp(-U+a(\mu))|W]$ numerically.

Consider the asymptotic case with $\gamma \to \infty$, such that $G_\gamma(\mu) = I(\mu \geq 0)$. If $\mu > 0$, then $a(\mu) = 0$, $g(w) = g^+(-w)$, and we recover the classical formulae of Kumbhakar and Lovell (2000, p.74-78) given in (8)-(10).

2.2.2 The extended exponential distribution

The random variable U has density given by (15) with

$$h_\mu(u) = \frac{1}{\tilde{\mu}}\exp\left(-\frac{|u|}{\tilde{\mu}}\right) \tag{27}$$

where $\tilde{\mu} = \mu(2G_\gamma(\mu) - 1)$.

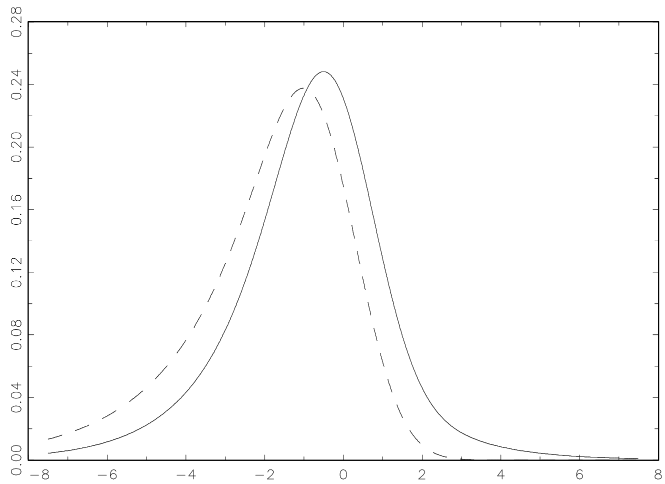

Figure 1: *Density $g_\mu(w)$ in (23) with $h_\mu(\cdot)$ given by (27), $\mu = 2$, and $\gamma = 1$ (solid line) and $\gamma = 10$ (dashed line).*

The function $g^+(z)$ in the expression for $g(w)$ in (23) is given by

$$g^+(z) = \frac{1}{\tilde{\mu}} \exp\left(-\frac{z}{\tilde{\mu}} + \frac{\sigma_v^2}{2\tilde{\mu}^2}\right) \Phi\left(\frac{z}{\sigma_v} - \frac{\sigma_v}{\tilde{\mu}}\right), \tag{28}$$

where $\Phi(\cdot)$ is the cdf of a standard normal random variable. For the classical case $\mu \geq 0$ and $\gamma \to \infty$, we obtain $G_\gamma(\mu) = 1$ and $a(\mu) = 0$, such that one obtains from the expression for $g(w)$ in (23) $g(w) = g^+(-w)$, and this density has been given, e.g., by Kumbhakar and Lovell (2000, p.80).

To illustrate the shape of this distribution as a function of μ and of γ, we fix $\mu = 2$ in Figure 1 and let γ be either 1 or 10, while in Figure 2 we fix $\gamma = 10$ and let μ be either 2 or -2. Note that in Figure 2, both densities share the same mean, $\mathrm{E}[W] = -\mu(2G_\gamma(\mu)-1)^3 \approx -2$.

Computation of inefficiencies To obtain measures of technical efficiency for this model, we calculate the conditional density $p(u|w) = f(u,w)/g(w)$, where the joint density

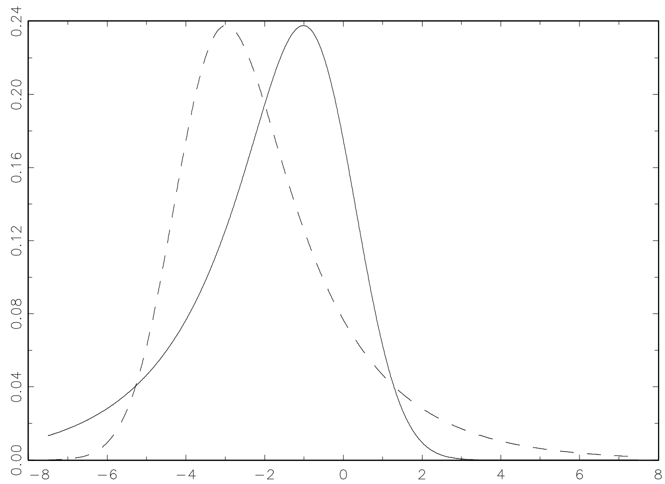

Figure 2: *Density $g_\mu(w)$ in (23) with $h_\mu(\cdot)$ given by (27), $\gamma = 10$, and $\mu = 2$ (solid line) and $\mu = -2$ (dashed line).*

$f(u, w)$ and marginal density $g(w)$ are given by

$$\begin{aligned} f(u,w) &= \frac{1}{\sqrt{2\pi}\sigma_v\tilde{\mu}} \exp\left(-\frac{(w+u-a(\mu))^2}{2\sigma_v^2} - \frac{|u|}{\tilde{\mu}}\right) H(u) \\ g(w) &= G_\gamma(\mu) g^+(-w+a(\mu)) + (1-G_\gamma(\mu)) g^+(w-a(\mu)) \end{aligned}$$

where $g^+(\cdot)$ is given by (28), and $H(u) = G_\gamma(\mu)I(u \geq 0) + (1 - G_\gamma(\mu))I(u < 0)$.

Again, we can use these expressions to calculate technical efficiency $\mathrm{E}[\exp(-U + a(\mu))|W]$ numerically.

2.3 Discussion

We see indeed that the model we propose can be seen as an extension of the traditional stochastic frontier model (we limited our presentation to the most popular normal/halfnormal or normal/exponential models). This appears clearly when $\gamma \to \infty$, which happens asymptotically if we choose, as in the next section, $\gamma = Cn^\nu$ for some constant C and some $\nu > 0$. So the standard SFM can be seen as a kind of "anchorage" model in which we believe, but we want to allow for the possibility that in a finite sample we have negative skewness of U_i.

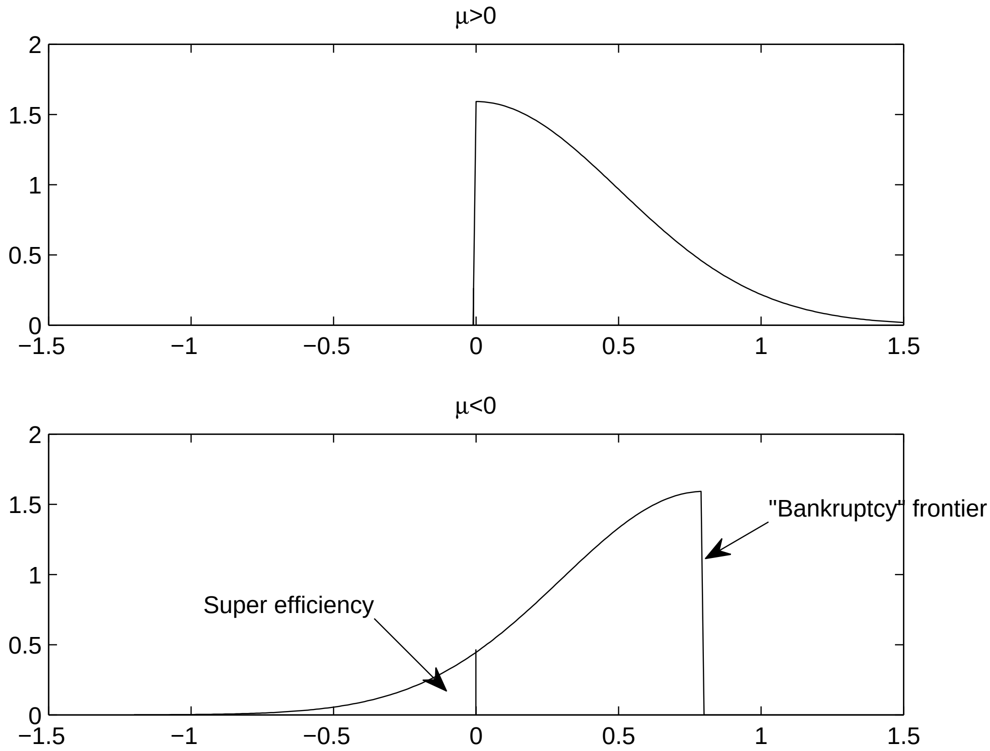

Figure 3: *The upper figure plots the density $f_\mu(u)$ of U in (15) with $h_\mu(u)$ given by (25) (half-normal case), $\gamma = \infty$ and $\mu = 0.2$. The lower figure plots the density of $U - 2E[U]$, where U has density $f_\mu(u)$ with $\mu = -0.2$. Both densities have a mean of 0.2.*

The easiest way to interpret the model is to analyze its components when $G_\gamma(\mu) = I(\mu > 0)$. Figure 3 illustrates this case by showing the two possibilities for the density of U. The upper panel corresponds to the classical stochastic frontier model, whereas the lower panel is the distribution underlying $U - 2\mathrm{E}[U]$ when it has negative skewness (note that the two distributions have equal means by construction).

In practice with our logistic weights we will have a continuous mixture of these two extreme cases that is illustrated in Figure 4. One clear advantage of doing this is that in the lower panel ($\mu < 0$) the size of the inefficiency is not bounded. However, the inefficiency can potentially be negative indicating the existence of "super efficient" firms. Below we discuss how this could be interpreted, which is mainly relevant for the case $\mu < 0$. For the case $\mu > 0$ this can be prevented by considering the limiting case of our

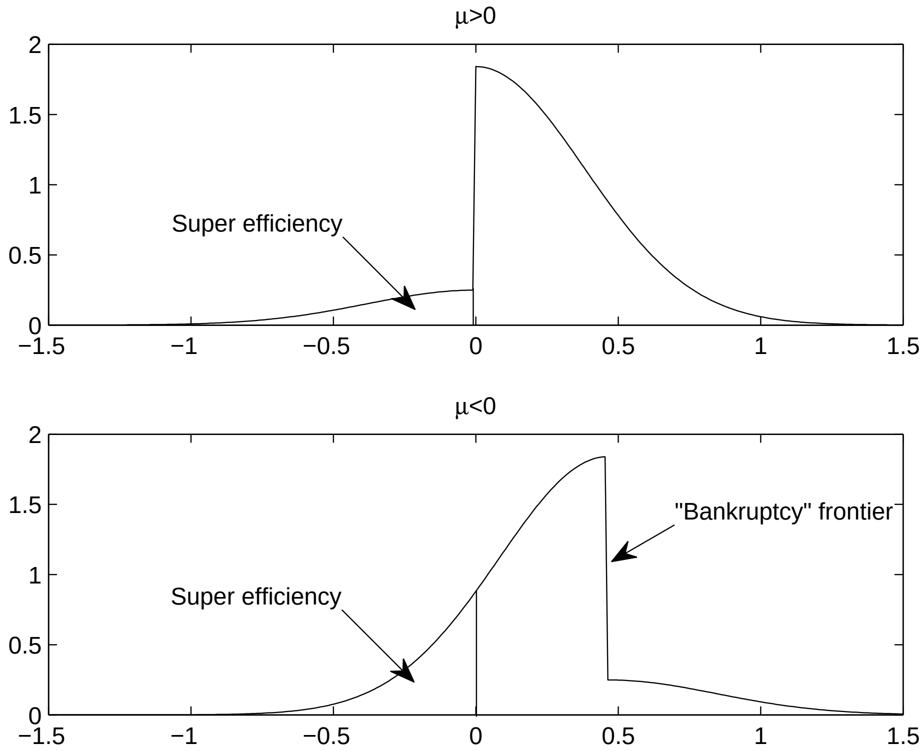

Figure 4: *The upper figure plots the density $f_\mu(u)$ of U in (15) with $h_\mu(u)$ given by (25) (half-normal case), $\gamma = 5$ and $\mu = 0.2$. The lower figure plots the density of $U - 2E[U]$, where U has density $f_\mu(u)$ with $\mu = -0.2$. Both densities have a mean of* $\mu(2G_\gamma(\mu) - 1)^2 \approx 0.043$.

model as $\gamma \to \infty$.

There are two interpretations of the situation when the skewness of the residuals $\widehat{W}_i$ is positive. The first is that this is purely a problem in finite samples as demonstrated in Simar and Wilson (2010). Thus, the classical stochastic frontier model with U having positive skewness is considered as an appropriate model, but the sample available to the researcher still shows positive skewness of $\widehat{W}_i$. Applying the traditional Stochastic Frontier Model will estimate the mean of U_i to be equal to zero and technical efficiency cannot be estimated as it is equal to one for all firms. This clearly is unreasonable and researchers may infer that their model is wrong. In such situations our model offers a possibility to approximate the "true" Stochastic Frontier Model and to obtain reasonable estimates for the technical efficiency along with the possibility to rank firms according to their efficiency.

The second interpretation allows the skewness of U to be negative even in population. Some of the arguments developed in the literature to justify the possibility of a negative skewness of U are based on the assumption that at some point in time, certain firms could develop new skills (which Carree, 2002 defines as *innovation*) providing some super-efficient firms with respect to the current technology, pushing the distribution of U to the left, with negative skewness and potentially very small values for U. When this happens the (relatively few) innovative firms are highly efficient, even super-efficient with respect to the old frontier, whereas the firms that have not (yet) adopted the new technology are relatively inefficient. This period could be followed by a period of learning by the other firms in the industry (called *imitation* by Carree, 2002). The cross section picture could be done at one moment or another. During the transition phase the firms that have already adapted have a different production frontier than those who have not adapted. Estimating a model on both types of firms it is very likely to observe the wrong skewness in a sample. The negative binomial and the double truncated model are ways to deal with such situations. However, these approaches have some drawbacks. First, the support of U is bounded at the two boundaries (one being zero), which a priory restricts the permissible range of inefficiency and is an uncommon assumption in the literature. Second, these model do not nest the traditional stochastic frontier model, which may be disadvantageous in cases where the wrong skewness problem is not present.

Our model in (13) corrects for that and may be interpreted along similar lines, pushing the argument even more. This can be represented by the lower panels of Figures 3 and 4. Those super-efficient (innovative) firms can indeed be very efficient with no bounds for the values of U that can in principle even be negative. Thus there is also a positive probability of observing some super-efficient firms above the efficient frontier. Furthermore, in our model with $\mu < 0$ and $U_i < 0$, we have a second boundary which is the right endpoint of the distribution of $U - 2E[U]$ in the lower panel of Figure 3. We call this frontier the "bankruptcy frontier". In the presence of an innovation in the industry, firms that would stay below this frontier would disappear after some period from the market if they do not adapt in some form. The extended model with $\gamma < \infty$ in Figure 4 is more attractive in this situation, as it does allow some firms to be less efficient than the bankruptcy frontier. For example, this could happen when firms are about to go out of business surviving through bank loans or companies that are sacrificing efficiency in the present to implement innovations in order to be more efficient in the future.

2.4 Estimation theory

We now give results for the asymptotic distribution of the maximum likelihood estimator of μ and show that it collapses to the classical case if $\mu_0 > 0$ and $\gamma \to \infty$. Moreover, we can also give the results for $\mu_0 < 0$, which differ slightly from the classical case.

Suppose we have observations $w_1, \ldots, w_n$, and assume for simplicity that all parameters are known except μ. The log-likelihood function is given by

$$\log L(\mu) = \sum_{i=1}^{n} \log \left\{ Gg^{+}(-w_i + a(\mu)) + (1-G)g^{+}(w_i - a(\mu)) \right\}$$

Let the true parameter be μ_0. The MLE is defined as

$$\hat{\mu} = \arg\max \log L(\mu).$$

We have the following assumptions:

(A1) Let the parameter space Θ be an open subset of $\mathbb{R}$, and the true parameter $\mu_0 \neq 0$.

(A2) $\gamma = Cn^{\nu}$ for some positive constants C and ν, so that $\lim_{n\to\infty} G_{\gamma}(\mu) = I(\mu > 0)$.

(A3) The function $h_{\mu}(u)$ defined by (16)-(20) is twice continuously differentiable with respect to μ in an open neighborhood of μ_0.

Assumption (A1) avoids a singularity issue arising if the true μ_0 is equal to zero. This is a well known problem in classical SFM when using, for example, likelihood ratio tests of the hypothesis $\mu_0 = 0$, which has a non-standard distribution, see Lee (1993). Assumption (A2) ensures that asymptotically, we recover the classical SFM if $\mu_0 > 0$. Finally, (A3) is a technical condition to have a well-defined asymptotic variance. The following theorem gives the asymptotic normality of the maximum likelihood estimator of μ.

Theorem 1 *Under our assumptions,*

$$\sqrt{n}(\hat{\mu} - \mu_0) \to_d N(0, B(\mu_0)) \tag{29}$$

where

$$\begin{aligned} B(\mu_0) &= \lim E \frac{1}{n} \left(\left. \frac{\partial \log L(\mu)}{\partial \mu} \right|_{\mu_0} \right)^2 \\ &= E\left[\left(\frac{g^{+\prime}(-w_i)}{g^{+}(-w_i)} \right)^2 \right] I(\mu_0 > 0) + 4E\left[\left(\frac{g^{+\prime}(w_i - 2\mu_0)}{g^{+}(w_i - 2\mu_0)} \right)^2 \right] I(\mu_0 < 0) \end{aligned}$$

with $'$ denoting first derivative w.r.t. μ.

The asymptotic variance $B(\mu_0)$ is decomposed into two terms, the first one being well known from classical SFM, and the second one being relevant if the true parameter is negative. Note that $B(\mu_0)$ is continuously differentiable except at $\mu_0 = 0$. Therefore, classical likelihood-based tests such as likelihood ratio will face the same issues as in classical SFM, but similar remedies as in Lee (1993) can be established. Rather than imposing (A2), we could alternatively keep γ finite, which guarantees a smooth transition between the two components of the asymptotic variance, and classical tests would have standard distributions. However, we prefer to impose $\gamma \to \infty$ to recover the classical SFM asymptotically and to emphasize that the objective of our generalized model is mainly to deal with finite sample problems.

3 A simulation study

In this section we present the results of a Monte Carlo study to compare the behavior of our proposed model compared with the classical stochastic frontier model. We do this for the normal-exponential and the normal-half normal model. We are interested in how the models estimate the model parameters and average technical efficiency in small samples when it is likely to observe a sample that is characterized by the "wrong skewness" problem. Our data generating process allows for two production factors and is given by

$$Y_i = \alpha_0 + \alpha_1 \log X_{1i} + \alpha_2 \log X_{2i} + V_i - U_i, \tag{30}$$

where $V_i \sim N(0, \sigma_v^2)$, $U_i \sim \exp(\mu_u)$ or $U_i \sim N^+(0, \mu_u\sqrt{\pi/2})$, $\log X_{1i} \sim N(1.5, 0.3)$ and $\log X_{2i} \sim N(1.8, 0.3)$. The true parameters are $\alpha_0 = 0.9$, $\alpha_1 = 0.6$, $\alpha_2 = 0.5$. We let μ take on the values 0.2, 0.3, and 0.4 corresponding to varying degrees of average technical efficiency. The standard deviation of the two-sided error σ_v is chosen as $\sigma_v = 0.5$ for the normal-exponential model and $\sigma_v = 0.25$ for the normal-half normal model. Different values are chosen to ensure that a reasonable fraction of the samples has positive and negative skewness of the composite error. We consider the sample sizes $N = 50, 100, 200$ and let the smoothing parameter γ increase with the sample size taking on the values $\gamma = 30, 40, 50$.[2] We report bias and MSE for the parameter estimates, as well as for technical efficiency defined as $TE = \mathrm{E}(\exp(-U + a(\mu))) = \mathrm{E}(\mathrm{E}(\exp(-U + a(\mu))|W))$ (true values obtained by simulation) computed as described in Section 2.2 and estimated

[2] We also used different values for γ. Mostly, the results are robust w.r.t. this choice as long as γ is not too small. If the function G is too smooth, however, biased estimates can occur.

as $n^{-1}\sum_{i=1}^{n}\widehat{\mathrm{E}}(\exp(-U_i+a(\hat{\mu})|W_i)$. We also report the fraction of samples having positive skewness (row: "pos. skew."). Results for "One-sided" refer to the standard stochastic frontier model and for "Two-sided" to the extended version. The number of Monte Carlo replications was equal to 1000.

The results are reported in Tables 1 to 4. First of all note that the results for the normal-exponential model and the normal-half normal model do not differ qualitatively. Looking at the bias for μ it is apparent that the two-sided model has a much larger bias, but that does not come as a surprise. The reason is that whenever the sample skewness of the residuals is positive, μ is estimated to be negative, so naturally the bias is inflated by this. The bias for σ_v is also a bit larger for our model. On the other hand, the intercept α_0 is estimated with a smaller bias for our model. For the remaining parameters the bias for the two-sided model is of similar size or smaller than for the one-sided model. The advantage of our approach becomes apparent when looking at the bias for technical efficiency, which is always smaller for the two-sided model due to the fact that the classical model cannot handle positive skewness and estimates technical efficiency to be equal to one in these cases. Note that as μ and the sample size increase, the fraction of samples with positive skewness becomes smaller and approaches zero. In these cases the two models basically give identical estimates. In terms of MSE, again both models perform equally well for most parameters, notably also for σ_v for which the bias was larger for our model. For technical efficiency, however, the MSE of the two-sided model is always smaller. Thus we can conclude that our approach yields reliable estimates for the model parameters without suffering from some of the drawbacks of the classical model when the residuals have positive skewness.

4 Application

We illustrate the advantages of our model for estimating technical efficiency using data from the NBER manufacturing productivity database (Bartelsman and Gray 1996). This database contains annual information on US manufacturing industries and contains data since 1958. Output is measured as total value added and as input factors we use total employment, cost of materials, energy cost and capital stock. In particular, we consider 54 sub-sectors from the textile industry over the years 1958-2005. We proceed by consecutively estimating the model on the cross-sectional data for each year. As a starting point the model is estimated by OLS to analyze the signs of the skewness of the residuals. It

Table 1: Bias normal-exponential model

	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.2$	-0.0475	-0.1403	-0.0478	-0.1217	-0.0422	-0.0968
$\sigma_v = 0.5$	-0.0257	-0.0615	-0.0105	-0.0316	-0.0048	-0.0174
$\alpha_0 = 0.9$	-0.0418	0.0470	-0.0498	0.0225	-0.0404	0.0148
$\alpha_1 = 0.6$	-0.0050	-0.0051	-0.0055	-0.0049	-0.0023	-0.0023
$\alpha_2 = 0.5$	0.0018	0.0018	0.0033	0.0032	0.0004	0.0006
TE=0.8333	0.0489	-0.0247	0.0467	-0.0141	0.0405	-0.0064
Pos. skew.		0.411		0.359		0.304
	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.3$	-0.0813	-0.1497	-0.0665	-0.1058	-0.0438	-0.0625
$\sigma_v = 0.5$	-0.0194	-0.0452	-0.0024	-0.0136	0.0022	-0.0023
$\alpha_0 = 0.9$	-0.0910	-0.0269	-0.0760	-0.0368	-0.0484	-0.0284
$\alpha_1 = 0.6$	0.0021	0.0025	0.0054	0.0052	0.0040	0.0042
$\alpha_2 = 0.5$	0.0042	0.0047	0.0021	0.0023	-0.0016	-0.0016
TE=0.7692	0.0684	0.0143	0.0547	0.0223	0.0356	0.0185
Pos. skew.		0.304		0.198		0.113
	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.4$	-0.0720	-0.1096	-0.0531	-0.0698	-0.0204	-0.0238
$\sigma_v = 0.5$	-0.0238	-0.0396	-0.0031	-0.0085	-0.0042	-0.0048
$\alpha_0 = 0.9$	-0.0794	-0.0389	-0.0613	-0.0422	-0.0210	-0.0181
$\alpha_1 = 0.6$	0.0069	0.0069	0.0060	0.0061	-0.0020	-0.0019
$\alpha_2 = 0.5$	0.0000	-0.0002	0.0005	0.0007	0.0022	0.0023
TE=0.7143	0.0569	0.0254	0.0412	0.0252	0.0158	0.0131
Pos. skew.		0.172		0.095		0.022

Note: This table presents Monte Carlo estimates of the bias for the parameter estimates for the model given in (30) with one-sided errors coming from an exponential distribution. The columns labeled "one-sided" refer to the classical stochastic frontier model, whereas "two-sided" refers to the model introduced in Section 2.1. The entry "Pos. skew" is the fraction of samples that are characterized by the wrong skewness problem. The smoothing parameter γ was chosen to be 30, 40 and 50 for sample sizes 50, 100 and 200, respectively. The results are based on 1000 Monte Carlo replications.

Table 2: MSE normal-exponential model

	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.2$	0.0297	0.0949	0.0220	0.0699	0.0176	0.0511
$\sigma_v = 0.5$	0.0092	0.0121	0.0039	0.0042	0.0023	0.0021
$\alpha_0 = 0.9$	0.1673	0.1639	0.0806	0.0727	0.0484	0.0377
$\alpha_1 = 0.6$	0.0225	0.0232	0.0097	0.0097	0.0048	0.0048
$\alpha_2 = 0.5$	0.0210	0.0220	0.0100	0.0102	0.0048	0.0048
TE	0.0174	0.0111	0.0131	0.0050	0.0108	0.0033
	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.3$	0.0464	0.108	0.0293	0.0629	0.0181	0.0337
$\sigma_v = 0.5$	0.0125	0.0127	0.0055	0.0049	0.0028	0.0024
$\alpha_0 = 0.9$	0.2111	0.2010	0.1032	0.0906	0.0529	0.0450
$\alpha_1 = 0.6$	0.0263	0.0272	0.0122	0.0122	0.0054	0.0054
$\alpha_2 = 0.5$	0.0275	0.0282	0.0117	0.0117	0.0058	0.0058
TE	0.0250	0.0138	0.0150	0.0064	0.0092	0.0042
	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.4$	0.0560	0.1006	0.0325	0.0510	0.0129	0.0161
$\sigma_v = 0.5$	0.0163	0.0154	0.007	0.0063	0.0034	0.0033
$\alpha_0 = 0.9$	0.2315	0.2168	0.1147	0.1055	0.0550	0.0534
$\alpha_1 = 0.6$	0.0291	0.0300	0.0133	0.0135	0.0066	0.0066
$\alpha_2 = 0.5$	0.0288	0.0291	0.0131	0.0132	0.0068	0.0068
TE	0.0252	0.0152	0.0133	0.0073	0.0047	0.0036

Note: This table presents Monte Carlo estimates of the mean squared error for the parameter estimates for the model given in (30) with one-sided errors coming from an exponential distribution. The columns labeled "one-sided" refer to the classical stochastic frontier model, whereas "two-sided" refers to the model introduced in Section 2.1. The smoothing parameter γ was chosen to be 30, 40 and 50 for sample sizes 50, 100 and 200, respectively. The results are based on 1000 Monte Carlo replications.

Table 3: Bias normal-half normal model

	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.2$	-0.0369	-0.1225	-0.0345	-0.0919	-0.0374	-0.0755
$\sigma_v = 0.25$	-0.0269	-0.0561	-0.0122	-0.0274	-0.0025	-0.0108
$\alpha_0 = 0.9$	-0.0387	0.0345	-0.0342	0.0204	-0.0373	-0.0007
$\alpha_1 = 0.6$	0.0001	-0.0001	-0.0008	-0.0008	0.0005	0.0005
$\alpha_2 = 0.5$	0.0004	0.0013	0.0010	0.0010	-0.0008	-0.0008
TE=0.8278	0.0399	-0.0226	0.0354	-0.0117	0.0351	0.0032
Pos. skew.		0.377		0.309		0.238
	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.3$	-0.0529	-0.1090	-0.0372	-0.0620	-0.0185	-0.0246
$\sigma_v = 0.25$	-0.0266	-0.0434	-0.0086	-0.0148	-0.0042	-0.0053
$\alpha_0 = 0.9$	-0.0535	-0.0040	-0.0334	-0.0106	-0.0146	-0.0091
$\alpha_1 = 0.6$	0.0018	0.0012	0.0000	-0.0003	-0.0010	-0.0010
$\alpha_2 = 0.5$	-0.0011	-0.0003	-0.0020	-0.0021	-0.0011	-0.0011
TE=0.7586	0.0486	0.0074	0.0326	0.0138	0.0161	0.0113
Pos. skew.		0.229		0.126		0.044
	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.4$	-0.0434	-0.0794	-0.0247	-0.0321	-0.0082	-0.0089
$\sigma_v = 0.25$	-0.0333	-0.0368	-0.0094	-0.0102	-0.0048	-0.0046
$\alpha_0 = 0.9$	-0.0394	-0.0151	-0.0141	-0.0089	-0.0145	-0.0148
$\alpha_1 = 0.6$	-0.0005	-0.0013	-0.0010	-0.0010	-0.0008	-0.0008
$\alpha_2 = 0.5$	-0.0009	-0.0011	-0.0039	-0.0041	0.0032	0.0032
TE=0.6988	0.0387	0.0195	0.0199	0.0158	0.0060	0.0062
Pos. skew.		0.121		0.033		0.002

Note: This table presents Monte Carlo estimates of the bias for the parameter estimates for the model given in (30) with one-sided errors coming from a half-normal distribution. The columns labeled "one-sided" refer to the classical stochastic frontier model, whereas "two-sided" refers to the model introduced in Section 2.1. The entry "Pos. skew" is the fraction of samples that are characterized by the wrong skewness problem. The smoothing parameter γ was chosen to be 30, 40 and 50 for sample sizes 50, 100 and 200, respectively. The results are based on 1000 Monte Carlo replications.

Table 4: MSE normal-half normal model

	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.2$	0.0234	0.0794	0.0174	0.0526	0.0129	0.0354
$\sigma_v = 0.25$	0.0072	0.0077	0.0028	0.0027	0.0014	0.0011
$\alpha_0 = 0.9$	0.0634	0.0530	0.0330	0.0245	0.0221	0.0144
$\alpha_1 = 0.6$	0.0068	0.0071	0.0029	0.0029	0.0015	0.0015
$\alpha_2 = 0.5$	0.0069	0.0071	0.0031	0.0031	0.0015	0.0015
TE	0.0152	0.0085	0.0118	0.0044	0.0089	0.0034
	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.3$	0.0301	0.0797	0.0182	0.0388	0.0084	0.0131
$\sigma_v = 0.25$	0.0103	0.0083	0.0041	0.0035	0.0018	0.0017
$\alpha_0 = 0.9$	0.0856	0.0715	0.0387	0.0327	0.0183	0.0162
$\alpha_1 = 0.6$	0.0095	0.0096	0.0038	0.0038	0.0020	0.0020
$\alpha_2 = 0.5$	0.0091	0.0092	0.0039	0.0039	0.0019	0.0019
TE	0.0175	0.0081	0.0105	0.0054	0.0046	0.0031
	N=50		N=100		N=200	
	One-sided	Two-sided	One-sided	Two-sided	One-sided	Two-sided
$\mu_u = 0.4$	0.0330	0.0688	0.0150	0.0219	0.0043	0.0046
$\sigma_v = 0.25$	0.0146	0.0110	0.0049	0.0045	0.0019	0.0019
$\alpha_0 = 0.9$	0.1011	0.0961	0.0416	0.0395	0.0177	0.0181
$\alpha_1 = 0.6$	0.0129	0.0126	0.0055	0.0055	0.0025	0.0025
$\alpha_2 = 0.5$	0.0115	0.0114	0.0050	0.0050	0.0023	0.0023
TE	0.0162	0.0101	0.0069	0.0053	0.0016	0.0016

Note: This table presents Monte Carlo estimates of the mean squared error for the parameter estimates for the model given in (30) with one-sided errors coming from a half-normal distribution. The columns labeled "one-sided" refer to the classical stochastic frontier model, whereas "two-sided" refers to the model introduced in Section 2.1. The smoothing parameter γ was chosen to be 30, 40 and 50 for sample sizes 50, 100 and 200, respectively. The results are based on 1000 Monte Carlo replications.

turns out that for a large number of years the OLS residuals have positive skewness. Thus, relying solely on the classical stochastic frontier model with exponential or half-normal inefficiencies one would not find any inefficiencies for the corresponding years. This seems highly unreasonable and our extended model is considered to be able to estimate technical efficiency in such cases.

Concerning the choice of the smoothing parameter γ we proceed as follows. We allow γ to take finite (possibly small) values by searching over the grid (3,5,8,10,20,50,100,500,1000) and selecting the value that gives the highest log-likelihood.[3] It turned out that generally very large values of γ are optimal. Especially in cases where the skewness has the expected sign the estimates from the classical and the extended model cannot be distinguished.

We consider the models with exponential and half-normal distribution for the inefficiency terms. For each year we estimate the classical stochastic frontier model and our extended model. For 18 out of the 47 years the skewness of the OLS residuals is positive, in which case the classical model estimates the absence of technical inefficiency for all industries. Detailed estimation results are not reported, but are available from the authors upon request. Based on the estimated models we compute the average technical efficiency for each year. Figures 5 and 6 show plots of the estimated technical efficiencies over time for the exponential and half-normal model, respectively. In both cases the results for the years characterized by "wrong skewness" are much more reasonable for the extended models. In fact, the estimates for technical efficiency appear relatively stable over time.

Along the lines of our arguments in Section 2.3 above, there are two possible explanations for our findings. The first is that the classical stochastic frontier model is indeed a reasonable approximation for the data generating process, but that by chance we observe the "wrong skewness" in a number of years. This is likely to happen for samples of such a small size. The second explanation could be that changes in the industry due to competition from abroad and changes in technology have led to an adaptation of firms. Periods with wrong skewness may represent times when some firms have already adapted, whereas others are still in the process of adapting to the new conditions.

[3] In principle one could estimate γ by MLE, but the likelihood function is very flat w.r.t. γ and it is therefore generally difficult to estimate.

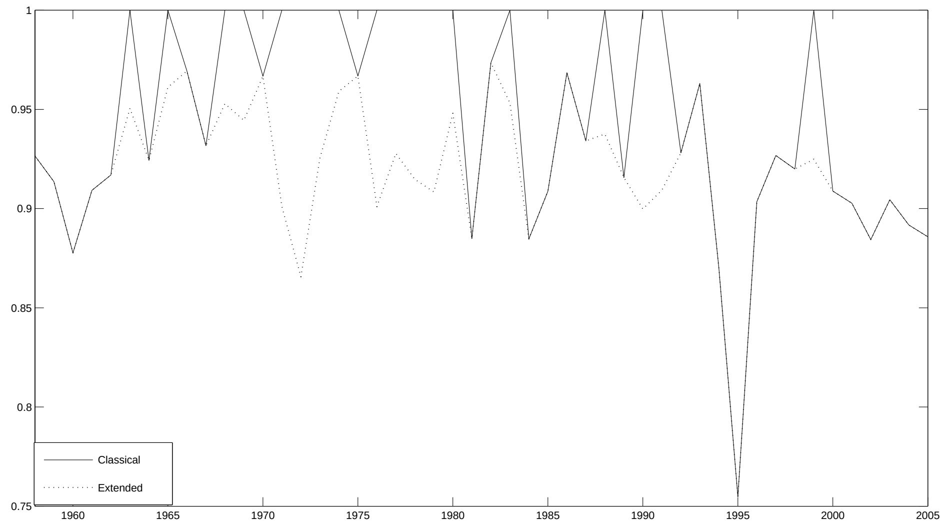

Figure 5: *Technical efficiency estimated by the normal-exponential stochastic frontier model (solid line) and our extended model (dashed line) for the years 1958-2005.*

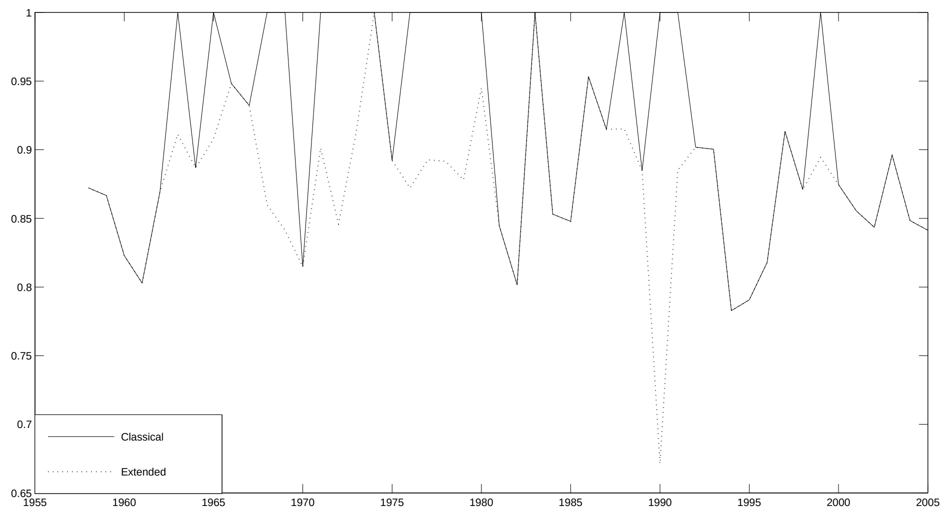

Figure 6: *Technical efficiency estimated by the normal-half normal stochastic frontier model (solid line) and our extended model (dashed line) for the years 1958-2005.*

5 Conclusions

In traditional SFA models, the "wrong skewness" problem is not a small issue because first, it plagues the estimation of the inefficiencies and second because researchers will often be tempted to change their model until they will observe the expected skewness. Classical inference assumes that the model specification is chosen independently of any estimates that are obtained. Specification-searching introduces problems of bias in both parameter estimates as well as variance-covariance estimates and we know from various simulation studies that the wrong skewness may appear even when the model is correctly specified.

Previous approaches to handle this issue involve the choice of densities for the efficiencies that are bounded below (by zero) and above. These approaches have their own merits but also some drawbacks. They restrict a priori the admissible range for the efficiency, which is rather unusual in this literature and these models do not nest the traditional SFA models.

Our approach extends the SFA model, allowing to disentangle inefficiency and skewness and nesting, as a particular case, the traditional SFA model. The asymptotic theory indicates that we obtain the usual properties of MLE estimators. Our Monte-Carlo experiments show also that we have better estimates of the technical efficiencies than the traditional SFA model, even with data generated under the latter. So the model we propose enriches the toolbox of the researcher for investigating efficiency analysis with parametric SFA models.

Appendix

Proof of Lemma 1.

To show (23), note that if U has density given by (15), then $-U$ has density given by

$$f_\mu(u) = h_\mu(u)\left\{G_\gamma(\mu)I(u<0) + (1-G_\gamma(\mu))I(u \geq 0)\right\}$$

Denote $\phi_{\sigma_v}(z) = \frac{1}{\sigma_v}\phi(\frac{z}{\sigma_v})$. The density of $V-U$ is given by the convolution of the densities of $-U$ and V,

$$\begin{aligned}
(f * \phi_{\sigma_v})(z) &= \int_{-\infty}^{\infty} f(u)\phi_{\sigma_v}(z-u)du \\
&= \frac{1}{\sigma_v}\int_{-\infty}^{0} f(u)\phi_{\sigma_v}(z-u)du + \int_{0}^{\infty} f(u)\phi_{\sigma_v}(z-u)du \\
&= G_\gamma(\mu)\int_{-\infty}^{0} h_\mu(u)\phi_{\sigma_v}(z-u)du + (1-G_\gamma(\mu))\int_{0}^{\infty} h_\mu(u)\phi_{\sigma_v}(z-u)du \\
&= G_\gamma(\mu)\int_{0}^{\infty} h_\mu(u)\phi_{\sigma_v}(-z-u)du + (1-G_\gamma(\mu))\int_{0}^{\infty} h_\mu(u)\phi_{\sigma_v}(z-u)du \\
&= G_\gamma(\mu)g^+(-z) + (1-G_\gamma(\mu))g^+(z)
\end{aligned}$$

The density of W is then given by $g(w) = (f * \phi_{\sigma_v})(w - a(\mu))$, where $a(\mu)$ is the third term in (13), which only depends on μ. □

Proof of Theorem 1.

We have to check the following conditions:

(C1) $\frac{\partial^2 \log L(\mu)}{\partial \mu^2} < \infty$ and is continuous in an open, convex neighborhood of μ_0.

(C2) $\frac{1}{n}\left.\frac{\partial^2 \log L(\mu)}{\partial \mu^2}\right|_{\mu^*} \to_p A(\mu_0) = \lim E\frac{1}{n}\left.\frac{\partial^2 \log L(\mu)}{\partial \mu^2}\right|_{\mu_0}$, for any sequence μ^* such that $\mu^* \to_p \mu$.

(C3) $\frac{1}{\sqrt{n}}\left.\frac{\partial \log L(\mu)}{\partial \mu}\right|_{\mu_0} \to_d N(0, B(\mu_0))$

Let $G = G_\gamma(\mu)$. We have

$$\begin{aligned}\frac{\partial \log L(\mu)}{\partial \mu} &= \sum_{i=1}^n \frac{g_i'}{g_i} \\ &= G' \sum_{i=1}^n \frac{g^+(-w_i + a(\mu)) - g^+(w_i - a(\mu))}{g_i} \\ &+ G \sum_{i=1}^n \frac{\partial g^+(-w_i + a(\mu))/\partial \mu}{g_i} \\ &+ (1-G) \sum_{i=1}^n \frac{\partial g^{+\prime}(w_i - a(\mu))/\partial \mu}{g_i}\end{aligned}$$

where $g_i = Gg^+(-w_i + a(\mu)) + (1-G)g^+(w_i - a(\mu))$ and

$$g_i' = \partial g_i/\partial \mu = G'\{g^+(-z_i) - g^+(z_i)\} + a'(\mu)\{Gg^{+\prime}(-z_i) - (1-G)g^{+\prime}(z_i)\},$$

with $z_i = w_i - a(\mu)$. The first term is negligible as $G'(\mu_0)$ converges to zero exponentially. If $\mu_0 > 0$, then the third term is negligible as $1 - G(\mu_0)$ converges to zero exponentially. In the second term, G converges to one exponentially and $\lim \partial g^+(-w_i + a(\mu))/\partial \mu = g^{+\prime}(-w_i)$. Hence, it suffices to show a Lindeberg-Levy CLT for $\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{g^{+\prime}(-w_i)}{g_i}$. If $\mu_0 < 0$, then the second term converges to zero, $\lim \partial g^+(w_i - a(\mu))/\partial \mu = \lim g^{+\prime}(w_i - a(\mu))a'(\mu) = 2g^{+\prime}(w_i - 2\mu)$ and a CLT is established for the term $\frac{2}{\sqrt{n}} \sum_{i=1}^n \frac{g^{+\prime}(w_i - 2\mu)}{g_i}$, which shows (C3).

Furthermore,

$$\frac{\partial^2 \log L(\mu)}{\partial \mu^2} = \sum_{i=1}^n \frac{g_i''}{g_i} - \left(\frac{g_i'}{g_i}\right)^2 \tag{31}$$

where

$$\begin{aligned}g_i'' &= G''\{g^+(-z_i) - g^+(z_i)\} + 2G'a'(\mu)\{g^{+\prime}(z_i) - g^{+\prime}(-z_i)\} \\ &+ a'(\mu)^2\{Gg^{+\prime\prime}(-z_i) + (1-G)g^{+\prime\prime}(z_i)\} + a''(\mu)\{Gg^{+\prime}(-z_i) - (1-G)g^{+\prime}(z_i)\},\end{aligned}$$

and $z_i = w_i - a(\mu)$, which shows that, using assumption (A3), (31) is finite and continuous in an open, convex neighborhood of μ_0, and hence condition (C1) is satisfied.

Next, note that $\mathrm{E}[g_i''/g_i] = \int g_i'' dz = \partial^2/\partial\mu^2 \int g_i dz = 0$, so that $n^{-1} \sum_{i=1}^n g_i''/g_i - (g_i'/g_i)^2$ converges in probability, uniformly in an open neighborhood of μ_0, to $-E[(\partial \log g^+(-w_i)/\partial \mu)^2]$ if $\mu_0 > 0$, and to $-4E[(\partial \log g^+(w_i - 2\mu_0)/\partial \mu)^2]$ if $\mu_0 < 0$, since $\lim_{n\to\infty} a(\mu) = 2\mu I(\mu < 0)$. Thus, $A(\mu_0) = -B(\mu_0)$, and condition (C2) holds by Theorem 4.1.5 of Amemiya (1985). Finally, (C1), (C2) and (C3) imply (29) by Theorem 4.1.3. of Amemiya (1985). □

References

Aigner, D. J., C. A. K. Lovel, and P. J. Schmidt (1977). Formulation and estimatoin of stochastic frontier production function models. *Journal of Econometrics 6*, 21–37.

Almanidis, P. and R. C. Sickles (2011). The skewness issue in stochastic frontier models: Fact of fiction? In I. van Keilegom and P. W. Wilson (Eds.), *Exploring Research Frontiers in Contemporary Statistics and Econometrics*. Springer Verlag, Berlin Heidelberg.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics 12*, 171–178.

Bartelsman, E. J. and W. Gray (1996). The NBER manufacturing productivity database. Nber technical working paper.

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

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*. Oxford University Press, New York.

Kumbhakar, S. C. and C. A. K. Lovell (2000). *Stochastic Frontier Analysis*. Cambridge University Press.

Lee, L. F. (1993). Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singluar information matrix. *Econometric Theory 9*, 413–430.

Li, Q. (1996). Estimating a stochastic production frontier when the adjusted error is symmetric. *Economics Letters 52*, 221–228.

Meeusen, W. and J. van den Broek (1977). Efficiency estimation from Cobb-Douglas production functions with composed errors. *International Economic Review 18*, 435–444.

Olson, J. A., P. Schmidt, and D. M. Waldman (1980). A monte carlo study of estimators of the stochastic frontier production function. *Journal of Econometrics 13*, 67–82.

Qian, J. and R. C. Sickles (2009). Stochastic frontiers with bounded inefficiency. Mimeo, rice university.

Simar, L. and P. W. Wilson (2010). Inference from cross-sectional stochastic frontier models. *Econometric Review 29*, 62–98.

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.