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] *Institute of Econometrics and Statistics, University of Cologne, Germany*

March 9, 2015

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 statistical properties of the model and we discuss a test for the symmetry of the error term (no inefficiency). We provide a simulation study to show that our model delivers estimators of efficiency with smaller bias 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 technical efficiencies well below the frontier, while the classical one estimates no inefficiency in those years.

Keywords: Stochastic frontier model, production efficiency, skewness, testing symmetry
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 the 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.). If we want a model allowing for inefficiency, we need to specify a model allowing to observe production plans (x_i, y_i) below this optimal frontier. The interesting feature of SFM (as opposed to deterministic frontier models) is that the model permits 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 Almanidis and Sickles (2011) and Almanidis et al. (2014) consider a doubly 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 $\widehat{\mu}_{3,n} = n^{-1} \sum_{i=1}^{n} \widehat{\varepsilon}_{i,\mathrm{OLS}}^3$ 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, under the chosen model, u_i has positive skewness. To be clear, we follow in our approach the idea that the "wrong skewness" is a small sample problem but that can pose serious problems for practitioners when estimating the traditional SFA models.

For these reasons, solutions have been proposed for choosing distributions for U_i that allow for negative skewness. Examples are the aforementioned negative binomial model of Carree (2002) and the double truncated normal of Almanidis and Sickles (2011), Almanidis et al. (2014). While these approaches have their merits, there are potential drawbacks. In particular, 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 basic models and the extensions we suggest and discusses their properties. In Section 3 least-squares and maximum likelihood estimators are proposed and their properties are described. In particular, we are able to derive a simple likelihood ratio test for testing the symmetry of the error term. Section 4 reports results of a simulation study to illustrate the usefulness of the proposed model in situations where the true model is the classical one. In Section 5 we apply the model to analyze the efficiency of sub-sectors of the U.S. textile industry for 1958-2005. Finally, in Section 6 we conclude.

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. Technical efficiency is the given by $\exp(-U)$. 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 technical efficiency in the basic model is defined as $\text{E}[\exp(-U)]$, and technical efficiency for a given firm can be predicted using the conditional expectation given W

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

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

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

$$h(u) = \frac{2}{\sigma_u} \phi\left(\frac{u}{\sigma_u}\right), \quad u \geq 0. \tag{6}$$

where $\sigma_u > 0$, $\phi(\cdot)$ is the standard normal pdf, and the expectation of U is $\mu = \sigma_u\sqrt{2/\pi} > 0$. The density of W is then given by

$$g(w) = \frac{2}{\sigma}\phi\left(\frac{-w}{\sigma}\right)\Phi\left(\frac{-w}{\sigma}\frac{\sigma_u}{\sigma_v}\right) \tag{7}$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$ 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 $\mu_* = -w\sigma_u^2/\sigma^2$, $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$ 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 and similar expressions as above can be found in Kumbhakar and Lovell (2000, p.82). Note that both densities (half normal and exponential) belong to the class of one-parameter scale densities, which facilitates many computations. These densities can be written as

$$h(u) = \frac{1}{\eta}\tilde{h}\left(\frac{u}{\eta}\right), \quad u \geq 0, \tag{11}$$

where the only parameter $\eta > 0$ and where $\tilde{h}(\cdot)$ is a density on $\mathbb{R}^+$ ($\int_0^\infty \tilde{h}(t)dt = 1$). Then all the moments of U can be written as $\mathrm{E}(U^j) = \eta^j k_j$, for $j = 1, 2, \ldots$ where

$$k_j = \int_0^\infty t^j \tilde{h}(t)dt \tag{12}$$

are constants depending on the chosen basic density $\tilde{h}(\cdot)$.

The literature on SFM (see e.g. Kumbhakar and Lovell 2000 and the references therein) has also suggested more flexible two-parameter densities by adding a location parameter. Examples are the gamma or the truncated normal cases that could be applied to our extension ideas below. However, these models come at the cost of numerical problems due to the potential difficulties to identify all the parameters of the models in small samples (see the discussion in Ritter and Simar 1997). Therefore we will focus our presentation below on the one-parameter scale family and give all the analytical details for the two basic benchmark models, i.e., the half-normal and the exponential.

2.1 The extended stochastic frontier model

In the following extension of the classical stochastic frontier model, an important parameter will be called γ, whose absolute value is related to the scale parameter η of the previous section, and whose sign determines the sign of the skewness of U. For the basic models above, this skewness is positive and $\eta = \gamma > 0$ and the third central moment of U can be written as $\mathrm{E}\left[(U - \mathrm{E}(U))^3\right] = a_3^+\gamma^3 > 0$, where $a_3^+ = k_3 - 3k_1k_2 + 2k_1^3$ can easily be computed. A first modification of the basic model would be to maintain the production model (4) with composed error term

$$W = V - U \tag{13}$$

where U is a positive random variable with expectation $\mu \geq 0$. However, the distribution of U depends on the sign of γ. If $\gamma > 0$, then U has a classical density $h^+(u)$ with support $[0, \infty)$ such as the half-normal or exponential (one parameter scale family with $\eta = \gamma$).

If $\gamma < 0$ we build a density $h^-(u)$ for U as follows: first we take the same density as h^+ and we mirror it onto the negative axis by reflecting it at zero. Then we truncate this density at $-B$ and finally we shift the resulting density to the right by B such that now the support of U is $[0, B]$. Since we want to stay in the one-parameter scale family we will fix the truncation parameter such that the resulting random variable U has the same mean $\mathrm{E}(U) = \mu$ as in the case $\gamma > 0$. To be specific, we define

$$h^-(u) = \left[\int_0^B h^+(t)dt\right]^{-1} h^+(B - u) I(u \in [0, B]), \tag{14}$$

where B will be determined below and $I(\cdot)$ is the indicator function. It can be seen that if h^+ belongs to the one-parameter scale family with scaling parameter $\eta = |\gamma|$, i.e., $h^+(u) = (1/|\gamma|)\tilde{h}^+(u/|\gamma|)$, then h^- also belongs to the one-parameter scale family. We indeed have

$$\begin{aligned} h^-(u) &= \frac{1}{|\gamma|}\left[\int_0^{B/|\gamma|} \tilde{h}^+(t)dt\right]^{-1} \tilde{h}^+\left(\frac{B-u}{|\gamma|}\right) I(u/|\gamma| \in [0, B/|\gamma|]) \\ &= \frac{1}{|\gamma|}\tilde{h}^-(u/|\gamma|), \end{aligned} \tag{15}$$

where it can be shown that the basic density $\tilde{h}^-$ is such that $\int_0^\infty \tilde{h}^-(t)dt = 1$. The density $\tilde{h}^-$ only depends on $B/|\gamma|$ that turns out to be a constant given by equating the means of the two densities h^+ and h^-. In particular, we have the equation

$$\begin{aligned} k_1^+ &= \int_0^\infty t\tilde{h}^+(t)dt = \int_0^{B/|\gamma|} t\tilde{h}^-(t)dt = k_1^- \\ &= \left[\int_0^{B/|\gamma|} \tilde{h}^+(t)dt\right]^{-1} \int_0^{B/|\gamma|} t\tilde{h}^+(B/|\gamma| - t)dt. \end{aligned} \tag{16}$$

There is only one unknown, $B/|\gamma|$, which only depends on the choice of the basic $\tilde{h}^+$. We call the solution $a_0 = B/|\gamma|$, so that $B = a_0|\gamma| > 0$.

The intuition behind our construction of h^- is as follows: since the skewness of h^+ is positive, we change its sign by reflecting it at zero. Then we shift the density to the right by B and truncate it at zero to get a density on the positive real line. By doing so we obtain a density with negative skewness and that converges to a Dirac measure at zero when γ approaches zero. Finally, the constant B is fixed to ensure that the mean of U depends only on the absolute value of γ, which has the additional advantage of not having to estimate this parameter that may be difficult to identify numerically.

Obviously, model (13) reduces to the classical SFA model if $\gamma > 0$, but we gain flexibility by allowing γ to become negative. In any case, however, the mean of U is $\mu = k_1^+|\gamma| > 0$ and that of W is $-\mu$, so that the mean of W only depends on the size of γ, not its sign. The size of γ is related to measures of inefficiency, whereas the sign of γ gives flexibility to fit distributional properties of the data, such as positive or negative skewness. Full efficiency is attained when $\gamma = 0$, in which case U degenerates to a one-point distribution at zero and $W \sim N(0, \sigma_v^2)$ is symmetric.

Below we see now how to derive the density h^- and the value a_0 in the particular cases of our two benchmark models. In particular, we will see that the likelihood function for W is continuous for all $\gamma \in \mathbb{R}$ including $\gamma = 0$ ensuring the consistency of the MLE for all γ. We are also able to derive a LR test for testing the hypothesis $\gamma = 0$ in the full stochastic frontier model.

2.2 Examples

In the following we give the details of the computations for our two benchmarked basic densities that are often used in practice.

2.2.1 The extended normal-halfnormal distribution

Suppose U has a half-normal density given by

$$h^+(u) = \frac{2}{\gamma}\phi\left(\frac{u}{\gamma}\right), \quad u > 0, \tag{17}$$

where $\gamma > 0$. This is the classical normal-half-normal SFA model with moments of U given by $\mu = \mathrm{E}[U] = \sqrt{\frac{2}{\pi}}\gamma = \frac{\phi(0)}{\Phi(0)}\gamma$, $\mathrm{Var}(U) = a_2^+\gamma^2 = [(\pi-2)/\pi]\gamma^2$ and $\mathrm{E}[(U-\mu)^3)] = a_3^+\gamma^3 = \sqrt{2/\pi}\big((4-\pi)/\pi\big)\gamma^3 \approx 0.2180\gamma^3 > 0$.

We extend this to the possibility $\gamma < 0$. This is achieved when U has a half-normal distribution mirrored at zero, shifted to the right by $B = a_0|\gamma|$ to have the same mean as the corresponding half-normal distribution with positive γ, and truncated at zero to ensure that inefficiency can only be positive. This is in fact a special case of the doubly-truncated normal distribution of Almanidis and Sickles (2011) and Almanidis et al. (2014) with location parameter B, truncation at zero and B, and scale parameter given by $|\gamma|$. In this case the moments of U can be found analytically. Thus, U is a $N(B, \gamma^2)$ truncated at zero and at $B = a_0|\gamma|$. We have

$$\mu = \mathrm{E}[U] = |\gamma|\left[a_0 + \frac{\phi(a_0)-\phi(0)}{\Phi(a_0)-\Phi(0)}\right], \tag{18}$$

where $a_0 = 1.3892032925$ is the non-trivial solution of $\left[a_0 + \frac{\phi(a_0)-\phi(0)}{\Phi(a_0)-\Phi(0)}\right] = \frac{\phi(0)}{\Phi(0)}$ such that the expectation of U as a function of $|\gamma|$ is the same for the two densities. We also have

$$\mathrm{Var}(U) = a_2^-\gamma^2 \tag{19}$$

$$\mathrm{E}[(U-\mu)^3)] = a_3^-\gamma^3 < 0, \tag{20}$$

where the constants $a_2^- > 0$ and $a_3^- > 0$ are obtained from the formulae derived in Almanidis and Sickles (2011, p.211).[2] The density $h^-(u)$ is thus given by

$$h^-(u) = [\Phi(a_0)-\Phi(0)]^{-1}\frac{1}{|\gamma|}\phi\left(a_0 - \frac{u}{|\gamma|}\right), \quad 0 < u < a_0|\gamma|, \tag{21}$$

[2]We obtain

$$a_2^- = 1 - \left[\frac{\phi(a_0)-\phi(0)}{\Phi(a_0)-\Phi(0)}\right]^2 + \frac{-a_0\phi(a_0)}{\Phi(a_0)-\Phi(0)} \approx 0.14471441$$

$$a_3^- = -2\left[\frac{\phi(a_0)-\phi(0)}{\Phi(a_0)-\Phi(0)}\right]^3 + \left[\frac{\phi(a_0)-\phi(0)}{\Phi(a_0)-\Phi(0)}\right]\left(1 - 3\frac{a_0\phi(a_0)}{\Phi(a_0)-\Phi(0)}\right) - \frac{a_0^2\phi(a_0)}{\Phi(a_0)-\Phi(0)} \approx 0.016741474.$$

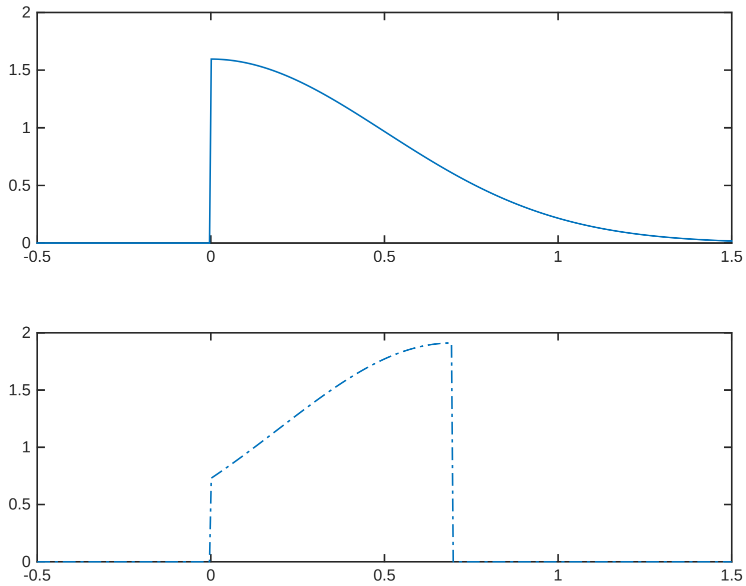

Figure 1: *The upper figure plots the density* $h^+(u)$ *of* U *in (17) (half-normal case) with* $\gamma = 0.5$. *The lower figure plots the density* $h^-(u)$ *of* U *with* $\gamma - 0.5$, *given in (21).*

where $\gamma < 0$. Note that in h^-, $|\gamma|$ can be replaced everywhere by $-\gamma$ to facilitate some derivations below. Figure 1 shows the density of U for the case of positive and negative skewness ($|\gamma| = 0.5$). Both distributions have the same mean $\mu = |\gamma|\sqrt{2/\pi} = 0.3989$ and are bounded below by zero. However, when $\gamma < 0$ the density is also bounded from above at $B = a_0|\gamma| = 0.6946$.

The density of W for the classical case $\gamma > 0$ is given by

$$g^+(w) = \frac{2}{\sigma}\phi\left(\frac{w}{\sigma}\right)\Phi\left(\frac{-w}{\sigma}\frac{\gamma}{\sigma_v}\right) \tag{22}$$

where $\sigma^2 = \gamma^2 + \sigma_v^2$. For the case $\gamma < 0$, the density of W can be shown to be[3]

$$g^-(w) = \frac{1}{\sigma(\Phi(a_0) - \Phi(0))}\phi\left(\frac{w - a_0\gamma}{\sigma}\right)\left[\Phi\left(A_w + \frac{a_0\sigma}{\sigma_v}\right) - \Phi(A_w)\right], \tag{23}$$

where

$$A_w = \frac{w - a_0\gamma}{\sigma}\frac{\gamma}{\sigma_v}. \tag{24}$$

[3]The density of W is the same as the one of ε in the first row of Table 1 of Almanidis et al. (2014) replacing μ by B, σ_u by $|\gamma|$ and setting $B = -a_0\gamma > 0$.

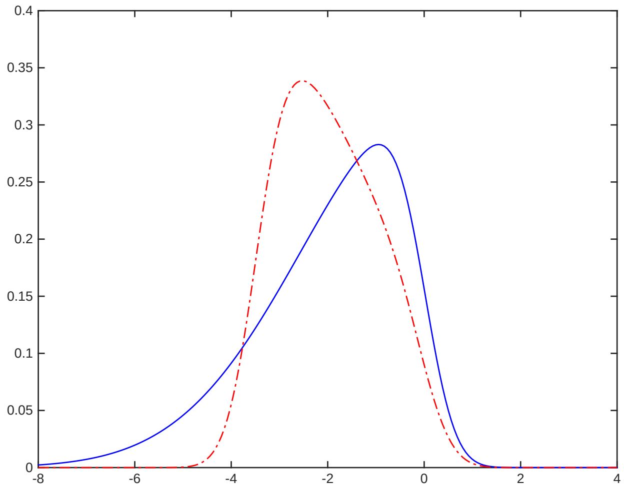

Figure 2: *Normal-Halfnormal case: Density* $g^+(w)$ *for* $\gamma = 2.5$ *(solid line) and* $g^-(w)$ *for* $\gamma = -2.5$ *(dash-dotted line), with* $\sigma_v = 0.5$. *Both densities have the same mean* $-\mu = -\sqrt{2/\pi}|\gamma| = -1.9947$.

In Figure 2 the density of W is plotted. For $\gamma > 0$ the distribution has negative skewness, whereas for $\gamma < 0$ its skewness is positive. Again, both distributions have the same mean.

We know that when $\gamma = 0$, the density of U is degenerate (a mass point) at zero and the density of W coincides with the one of V (typically a normal with zero mean and variance σ_v^2). So we will consider a density of W defined as follows:

$$g(w;\gamma) = g^+(w)I(\gamma > 0) + g^-(w)I(\gamma < 0) + \frac{1}{\sigma_v}\phi\Big(\frac{w}{\sigma_v}\Big)I(\gamma = 0). \tag{25}$$

Figure 3 displays the shape of the density of $g(w)$ when γ is varying from -2 to +2. The density seems to be continuous at $\gamma = 0$. However, on the floor of the picture, the contour plots indicates a non-smooth behavior near $\gamma = 0$. We will investigate this issue in detail in Section 3.

Technical efficiency for the case $\gamma < 0$[4] can be predicted given W using numerical integration:

$$TE = \mathrm{E}[\exp(-U)|W] = \int_0^{a_0|\gamma|} \exp(-u) f^-(u|w)du, \tag{26}$$

[4]The expressions for the case $\gamma > 0$ were given in (8)-(10) with $\sigma_u = \gamma$.

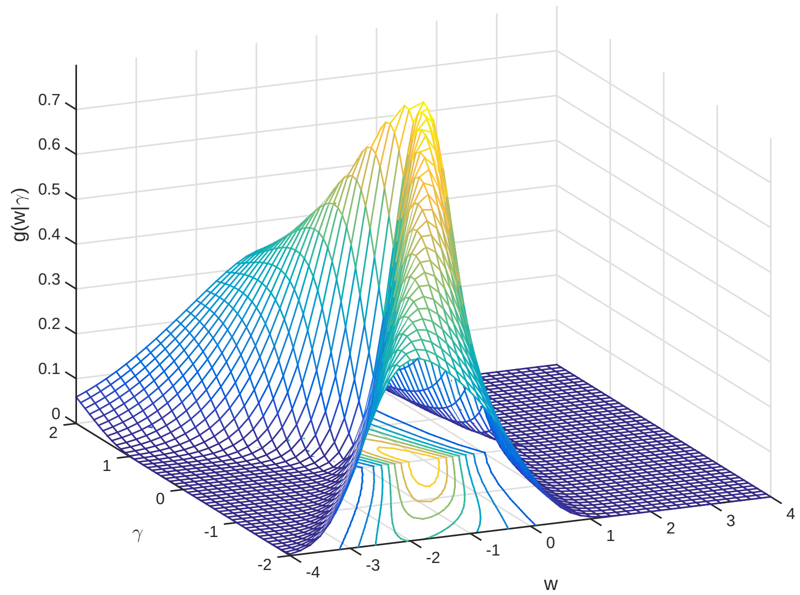

Figure 3: *Normal-Halfnormal case: Density* $g(w)$ *as a function of* $\gamma \in [-2, +2]$. *Here again* $\sigma_v = 0.5$.

where from the first row of Table 1 in Almanidis et al. (2014) we can recover

$$f^-(u|w) = \frac{1}{\sigma_*}\phi\left(\frac{u-\mu_*}{\sigma_*}\right)\left[\Phi\left(\frac{a_0|\gamma|-\mu^*}{\sigma_*}\right) - \Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right]^{-1} I(u \in [0, a_0|\gamma|]) \qquad (27)$$

with $\sigma_* = |\gamma|\sigma_v/\sigma$ and $\mu_* = -(a_0\gamma\sigma_v^2 + w\gamma^2)/\sigma^2$.

2.2.2 The extended exponential distribution

In the normal-exponential SFA model, the random variable U has density given by

$$h^+(u) = \frac{1}{\gamma}\exp\left(-\frac{u}{\gamma}\right), \quad u > 0 \qquad (28)$$

where $\mu = \mathrm{E}(U) = \gamma > 0$. We have also $\mathrm{Var}(U) = a_2^+\gamma^2 = \gamma^2$ and $\mathrm{E}\left[(U - \mathrm{E}(U))^3\right] = a_3^+\gamma^3 = 2\gamma^3 > 0$. The density of W is, see, e.g., Kumbhakar and Lovell (2000, p.80),

$$\begin{aligned} g^+(w) &= \frac{1}{\gamma}\exp\left(\frac{w}{\gamma} + \frac{\sigma_v^2}{2\gamma^2}\right)\Phi\left(-\frac{w}{\sigma_v} - \frac{\sigma_v}{\gamma}\right) \\ &= \frac{1}{\gamma}\frac{\Phi(-C_w)}{\phi(-C_w)}\phi\Big(\frac{w}{\sigma_v}\Big), \end{aligned} \qquad (29)$$

where $C_w = w/\sigma_v + \sigma_v/\gamma$ and where the latter expression is introduced to facilitate the analysis below.

As explained above we extend this to allow a density with negative skewness ($\gamma < 0$), in which case U has an exponential distribution mirrored at zero, shifted to the right by $B = a_0|\gamma|$, where $a_0 > 0$ is selected such that the density has the same positive mean as the corresponding original exponential distribution, $\mu = |\gamma| > 0$. We have

$$h^-(u) = \frac{e^{-a_0}}{\gamma(e^{-a_0} - 1)}\exp\left(-\frac{u}{\gamma}\right), \quad 0 < u < a_0|\gamma|. \qquad (30)$$

The expectation of the latter distribution is simply

$$\mathrm{E}[U] = k_1^-|\gamma|, \;\text{ where } k_1^- = \frac{e^{a_0}(a_0 - 1) + 1}{e^{a_0} - 1} > 0,$$

and a_0 has to be chosen such that $k_1^- = 1$. The non-trivial solution of this equation is $a_0 = 1.59362426$. The moments of the density h^- are easy to compute, and we have

$$\mathrm{Var}(U) = a_2^-\gamma^2 \qquad (31)$$

$$\mathrm{E}[(U-\mu)^3)] = a_3^-\gamma^3 < 0, \qquad (32)$$

where the constants[5] are given by

$$
\begin{aligned}
a_2^- &= k_2^- - 1 = (-a_0^2 + 2a_0 - 2 + 2e^{-a_0})e^{a_0}/(1 - e^{-a_0}) = 0.18724852 > 0 \\
a_3^- &= -(k_3^- - 3k_2^- + 2k_1^-) = (a_0^3 - 3a_0^2 + 6a_0 - 6 + 6e^{-a_0})e^{a_0}/(1 - e^{-a_0}) = 0.044214556 > 0.
\end{aligned}
\tag{33}
$$

For the density of W when $\gamma < 0$, we have

$$
g^-(w) = \frac{e^{-a_0}}{\gamma(e^{-a_0} - 1)} \frac{\Phi(-C_w) - \Phi(-C_w + a_0\gamma/\sigma_v)}{\phi(-C_w)} \phi\Big(\frac{w}{\sigma_v}\Big), \tag{34}
$$

where as above $C_w = w/\sigma_v + \sigma_v/\gamma$, but here $\gamma < 0$. Note that this corresponds to a reflection around its center of the truncated exponential distribution in Table 1 of Almanidis et al. (2014). Hence, ε is replaced by $-w - B = -w + a_0\gamma$ and σ_u by $-\gamma$.

As above, the density for W will be obtained by combining these results and defining

$$
g(w;\gamma) = g^+(w)I(\gamma > 0) + g^-(w)I(\gamma < 0) + \frac{1}{\sigma_v}\phi\Big(\frac{w}{\sigma_v}\Big)I(\gamma = 0). \tag{35}
$$

These densities are displayed in Figure 4 for two values of γ and in Figure 5 as a function of γ in the range $[-2, +2]$. These pictures are qualitatively very similar to the half normal case of the preceding section.

Finally, a predictor of the technical efficiencies for a given value of W when $\gamma < 0$ can be calculated numerically as in (26) with the conditional density given by[6]

$$
f^-(u|w) = \frac{1}{\sigma_v}\phi\left(\frac{u}{\sigma_v} + C_w\right)\left[\Phi\left(C_w - \frac{a_0\gamma}{\sigma_v}\right) - \Phi\left(C_w\right)\right]^{-1} I(u \in [0, a_0|\gamma|]). \tag{36}
$$

3 Statistical Properties

3.1 Least-squares estimators

Since it is easy to derive the moments of W we can define the MOLS estimators of the parameters $\theta = (\alpha_0, \alpha^T, \sigma_v, \gamma)^T$ in both our extended models. These are obtained from the OLS residuals having variance $\widehat{\mu}_{2,n}$ and third moment $\widehat{\mu}_{3,n}$ whatever being the sign of the latter.

[5]These constants can also be recovered from the formulae in Table 2, 3rd row, in Almanidis et al. (2014) with the appropriate changes.

[6]When $\gamma > 0$ the formulae are given e.g. in Kumbhakar and Lovell (2000). For $\gamma < 0$ the density can be recovered from Almanidis et al. (2014, Table 1) when realizing that u has to replaced by $B - u$, with $B = -a_0\gamma$.

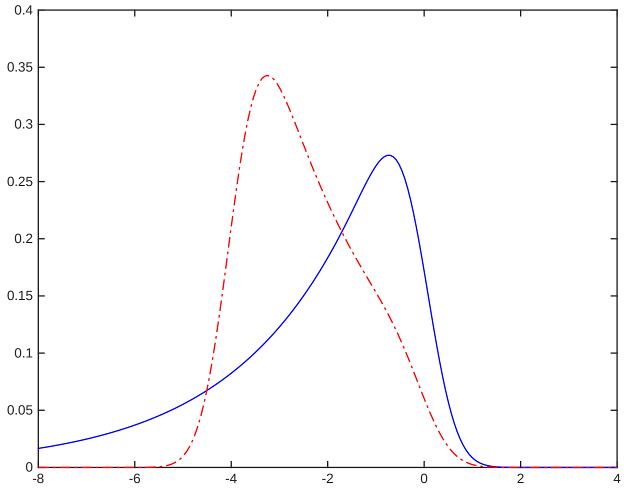

Figure 4: *Normal-Exponential case: Density $g(w)$ for $\gamma = 2.5$ (solid line) and $\gamma = -2.5$ (dash-dotted line), with $\sigma_v = 0.5$. Both densities have the same mean $-\mu = -|\gamma|$.*

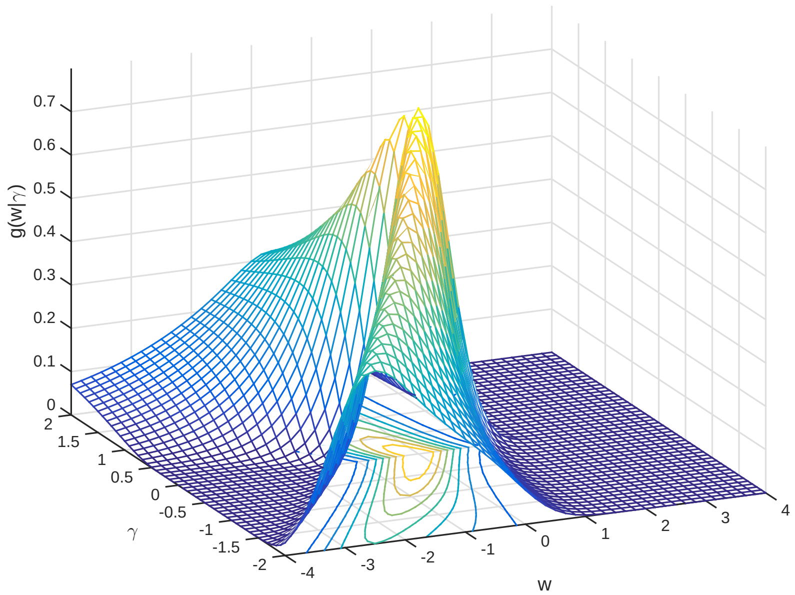

Figure 5: *Normal-Exponential case: Density $g(w)$ as a function of $\gamma \in [-2, +2]$. Here again $\sigma_v = 0.5$.*

Table 1: Constants for the central moments of U according to the sign of γ.

Model	a_2^+	a_2^-	a_3^+	a_3^-
norm-halfnormal	$(\pi-2)/\pi$	0.14471441	$\sqrt{2/\pi}\big((4-\pi)/\pi\big)$	0.016741474
norm-exponential	1	0.18724852	2	0.044214556

Indeed we have $\mathrm{Var}(U) = a_2\gamma^2$ and $\mathrm{E}\Big[(U - E(U))^3\Big] = a_3\gamma^3$ where a_2 and a_3 have a different value according to the sign of γ. For the two particular models detailed in the preceding section the values are summarized in Table 1. This allows to define the MOLS estimators as follows:

$$\text{if } \widehat{\mu}_{3,n} < 0, \quad \widehat{\gamma} = \left[-\frac{1}{a_3^+}\widehat{\mu}_{3,n}\right]^{1/3} > 0, \text{ and } \widehat{\sigma}_v^2 = \widehat{\mu}_{2,n} - a_2^+\widehat{\gamma}^2, \tag{37}$$

$$\text{if } \widehat{\mu}_{3,n} > 0, \quad \widehat{\gamma} = -\left[\frac{1}{a_3^-}\widehat{\mu}_{3,n}\right]^{1/3} < 0, \text{ and } \widehat{\sigma}_v^2 = \widehat{\mu}_{2,n} - a_2^-\widehat{\gamma}^2. \tag{38}$$

The other parameters are estimated as usual; $\widehat{\alpha}$ is directly given by the OLS, and $\widehat{\alpha}_0 = \widehat{\alpha}_{0,\mathrm{OLS}} + k_1\ |\widehat{\gamma}|$, where $k_1 = \sqrt{2/\pi}$ for the half normal model and $k_1 = 1$ for the exponential case. As already noticed in the related literature (see Almanidis and Sickles 2011, Almanidis et al. 2014), this provides a consistent estimator of the parameter vector θ.

3.2 Maximum likelihood estimators

In the extended models, the log-likelihood function for a sample (X_i, Y_i), $i = 1, \ldots, n$ is given by

$$\log L(\theta) = \sum_{i=1}^{n} \left[\log g^+(w_i;\theta)I(\gamma > 0) + \log g^-(w_i;\theta)I(\gamma < 0) + \log g^0(w_i;\theta)I(\gamma = 0)\right], \tag{39}$$

where $g^0(w_i;\theta) = (1/\sigma_v)\phi\big(w_i/\sigma_v\big)$ is the density in the case of full efficiency (only noise).

The two sub-models defining our model are "regular" models (meaning, in particular, differentiability of the likelihood) as long as either $\gamma > 0$ or $\gamma < 0$; see, e.g., Aigner et al. (1977), Almanidis and Sickles (2011), Almanidis et al. (2014). Thus, our model is regular for any $\gamma \neq 0$. The same is obviously true in the limiting case of $\gamma = 0$, in which case we have a standard regression model. The asymptotic properties of $\widehat{\theta} = \arg\max_\theta \log L(\theta)$, the MLE of θ, will depend on the value of $\gamma \in \mathbb{R}$. Obviously, when $\gamma \neq 0$, we have the usual asymptotic properties of the MLE coming from one of the two regular sub-models. However, in the degenerate case $\gamma = 0$, we know (see Lee 1993) that the situation is more complex and depends on the sign of $\widehat{\gamma}$. We now analyze the behavior of the likelihood as a function of γ near $\gamma = 0$. We show that the likelihood function in (39) is continuous at $\gamma = 0$, but that it is not continuously differentiable at this point.

Theorem 1 *For the extended normal-half-normal and for the normal-exponential models we have for all values of the other parameters $\xi = (\alpha_0, \alpha^T, \sigma_v)^T$ and for all w,*

$$\lim_{\gamma \to 0^+} g(w;\theta) = \lim_{\gamma \to 0^-} g(w;\theta) = \frac{1}{\sigma_v}\phi\left(\frac{w}{\sigma_v}\right). \tag{40}$$

The left and right derivatives of $\log g$ at $\gamma = 0$ are given by

$$\lim_{\gamma \to 0^+} \frac{\partial \log g(w;\theta)}{\partial \gamma} = -k_1 \frac{w}{\sigma_v^2} \tag{41}$$

$$\lim_{\gamma \to 0^-} \frac{\partial \log g(w;\theta)}{\partial \gamma} = k_1 \frac{w}{\sigma_v^2}, \tag{42}$$

where k_1, defined in(12), is the first moment of the basic underlying density of the one-parameter scale family, with $k_1 = \sqrt{2/\pi}$ for the half-normal and $k_1 = 1$ for the exponential cases.

The proof is given in Appendix A. This result has immediate consequences for the likelihood function. We have the following corollary.

Corollary 1 *In the extended normal-half-normal and normal-exponential models, the likelihood function is continuous at all values of θ and $\widehat{\theta} \xrightarrow{p} \theta$ as $n \to \infty$.*

Proof. The continuity of the likelihood function derives from the regularity of the two sub-models for $\gamma > 0$ and for $\gamma < 0$ and from the continuity at $\gamma = 0$ obtained in Theorem 1. The consistency of the MLE estimator follows directly; see, e.g., Amemiya (1985, Section 4.2.2). ■

For both the normal-half-normal and normal-exponential models it is straightforward to derive the score function for each sub-model by applying the formulae available in the literature; see, e.g., Almanidis et al. 2014 and using the result obtained in Theorem 1 above. By doing so we will see that the information matrix at $\gamma = 0$ is singular and that for each sub-model we are in the situation described in detail by Lee (1993). A remarkable consequence of Theorem 1 is that the score with respect to the scale parameter γ has opposite signs when taking left and right limits, but that in absolute value these limits are identical. In the following, we derive the expressions of the score with respect to $\beta = (\alpha_0, \alpha^T)^T$ and to γ in the extended normal-half normal model, which are those that determine the singularity of the information matrix. To save space we do not report the score w.r.t. to σ_v as it does not play any particular role. The results for the normal-exponential model are similar and available upon request.

Consider first the case $\gamma > 0$ and let $\tilde{x}_i = (1, x_i^T)^T$. Then, we have

$$\begin{aligned}
\frac{\partial \log L(\theta)}{\partial \beta} &= \sum_{i=1}^{n} \frac{w_i}{\sigma^2}\tilde{x}_i + \frac{\gamma}{\sigma\sigma_v}\sum_{i=1}^{n} \eta_i(\theta)\tilde{x}_i \\
\frac{\partial \log L(\theta)}{\partial \gamma} &= \frac{\gamma}{\sigma^2}\sum_{i=1}^{n}\Big(\frac{w_i^2}{\sigma^2} - 1\Big) - \frac{\sigma_v}{\sigma^3}\sum_{i=1}^{n} \eta_i(\theta) w_i
\end{aligned}$$

where $\eta_i^+(\theta) = \dfrac{\phi\big(-(w_i\gamma)/(\sigma\sigma_v)\big)}{\Phi\big(-(w_i\gamma)/(\sigma\sigma_v)\big)}$. Clearly, when $\gamma \to 0^+$ we have $\sigma \to \sigma_v$ and $\eta_i^+(\theta) \to \phi(0)/\Phi(0) = k_1$ and, denoting $\xi = (\alpha_0, \alpha^T, \sigma_v)^T$, we obtain

$$\frac{\partial \log L(\xi, 0)}{\partial \beta} = \sum_{i=1}^{n} \frac{w_i}{\sigma_v^2}\tilde{x}_i \tag{43}$$

$$\frac{\partial \log L(\xi, 0)}{\partial \gamma} = -\frac{1}{\sigma_v^2} k_1 \sum_{i=1}^{n} w_i. \tag{44}$$

We see that $\dfrac{\partial \log L(\xi, 0)}{\partial \theta} S(\xi) = 0$ where $S(\xi) = (k_1\ell_1^T, 0, 1)^T$ with ℓ_1 being the first column of the identity matrix of order $p+1$. Thus, up to a different parametrization, we are in the same situation as the one described by Lee (1993): we obtain a singular information matrix in the sub-model where $\gamma > 0$ in the limit as $\gamma \to 0^+$.

A similar situation arises in the case $\gamma < 0$ for which

$$\begin{aligned}
\frac{\partial \log L(\theta)}{\partial \beta} &= \sum_{i=1}^{n} \frac{w_i - a_0\gamma}{\sigma^2}\tilde{x}_i + \frac{\gamma}{\sigma\sigma_v}\sum_{i=1}^{n} \eta_i^-(\theta)\tilde{x}_i \\
\frac{\partial \log L(\theta)}{\partial \gamma} &= \frac{n\gamma}{\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}\Big[\sigma^2 a_0(w_i + a_0\gamma) - \gamma(w_i + a_0\gamma)^2\Big] \\
&\quad - \frac{1}{\sigma^3}\sum_{i=1}^{n}(w_i + a_0\gamma)\sigma_v\eta_i^-(\theta) - \frac{a_0\gamma}{\sigma_v\sigma}\sum_{i=1}^{n}\frac{\phi(A_{w_i})}{\Phi\big(A_{w_i} + a_0\sigma/\sigma_v\big) - \Phi\big(A_{w_i}\big)},
\end{aligned}$$

where $\eta_i^-(\theta) = \dfrac{\phi\big(A_{w_i} + a_0\sigma/\sigma_v\big) - \phi\big(A_{w_i}\big)}{\Phi\big(A_{w_i} + a_0\sigma/\sigma_v\big) - \Phi\big(A_{w_i}\big)}$ with the notation A_w introduced in (24).

Again it is easy to check that as $\gamma \to 0^-$, $A_{w_i} \to 0$, $\sigma \to \sigma_v$, so that

$$\frac{\partial \log L(\xi,0)}{\partial \beta} = \sum_{i=1}^{n} \frac{w_i}{\sigma_v^2} \tilde{x}_i \quad (45)$$

$$\frac{\partial \log L(\xi,0)}{\partial \gamma} = \sum_{i=1}^{n} \frac{w_i}{\sigma^2} \left[a_0 + \frac{\phi(a_0) - \phi(0)}{\Phi(a_0) - \Phi(0)} \right] = \frac{1}{\sigma_v^2} k_1 \sum_{i=1}^{n} w_i. \quad (46)$$

The last equality is due to our way of choosing the truncation point of the double truncated normal. Again, the same singularity issue of the information matrix appears as before, where now the S-vector is defined as $S(\xi) = (k_1 \ell_1^T, 0, -1)^T$.

Note that the score w.r.t. β in (43) is the same as that in (45), i.e., it does not depend on the sign of γ. This could be used to derive the asymptotic distribution of the estimator of β under $\gamma = 0$. However, the joint asymptotic distribution of the estimator of the full parameter vector is nonstandard under $\gamma = 0$. Lee (1993) uses a re-parametrization to obtain a non-singular information matrix, which allows deriving the asymptotic distribution of the MLE under $\gamma = 0$. It is a mixture of two non-standard conditional distributions depending on the condition that the MLE of γ is either positive or zero. Lee considers only the traditional model with the half-normal U, so when $\gamma = 0$, the probability of $\widehat{\gamma} > 0$ is 1/2 and the same for the probability of $\widehat{\gamma} = 0$. In the extended model developed above this is avoided because $\mathrm{P}(\widehat{\gamma} = 0) = 0$. In our case, we only have the asymptotic conditional distribution obtained for the two sub-models when $\widehat{\gamma} > 0$ corresponding to the model g^+ and when $\widehat{\gamma} < 0$, corresponding to the model g^-, each having a probability 1/2. As we will see in the following, to derive a test for $\gamma = 0$ we do not need an explicit expression for the distribution under $\gamma = 0$, but our previous results allow determining the distribution of the likelihood ratio (LR) test by adapting Lee's results to our framework.

3.3 Testing symmetry of W

As a side result of the derivations in the last section, we can build a likelihood ratio test for testing the symmetry of the error term W. This corresponds to the case where there is no inefficiency and $P(U = 0) = 1$, so we only have noise in the model and $W \sim N(0, \sigma_v^2)$. In our approach this corresponds to the case $\gamma = 0$. Testing $H_0 : \gamma = 0$ against the alternative $H_1 : \gamma \neq 0$ can be achieved with a likelihood ratio statistic that turns out to follow a standard chi-square distribution with one degree of freedom. We indeed avoid the difficulties described in Lee (1993) because in our extended model we are not testing a boundary value of the parameter of interest.

Table 2: Size of the LR test for symmetry

α/n	50	100	200	500	1000	2000
0.1	0.2065	0.1324	0.113	0.1037	0.1015	0.0964
0.05	0.1522	0.0755	0.0607	0.0541	0.0529	0.0491
0.01	0.0975	0.0228	0.0131	0.0116	0.0107	0.0100

Note: Empirical size of the likelihood ratio test for symmetric errors. The data is generated from the model in (48) with $\sigma_v = 1$ and $\gamma = 0$. The number of Monte Carlo replications is equal to 100,000.

Under the null ($\gamma = 0$) we have a reduced set of parameters $\xi = (\alpha_0, \alpha^T, \sigma_v)^T$ of a regular standard regression model. Thus, under the null the MLE of ξ is given by $\widetilde{\xi}$ (which are the OLS estimators) and the LR test statistics can be written as

$$LR = 2(\log L(\widehat{\theta}) - \log L(\widetilde{\xi}, 0)). \tag{47}$$

Conditioning on the two events $\widehat{\gamma} > 0$ and $\widehat{\gamma} = 0$, Lee (1993) shows that under the null $\gamma = 0$, LR has an asymptotic distribution given by an equally weighted mixture of a chi-square distribution with zero degrees of freedom (i.e., a mass 1 at zero) and a chi-square with one degree of freedom (the number of restrictions of the null). The degenerate part is coming from the conditional distribution if $\widehat{\gamma} = 0$, which occurs with probability 1/2. As pointed out above, in our extended model this probability is zero and the conditioning event is now $\widehat{\gamma} < 0$ where the other sub-model is used, leading to the same χ_1^2 distribution with probability 1/2. Therefore, we can conclude that under the null, the LR statistics has an asymptotic χ^2 distribution with one degree of freedom. As usual, we reject the null if LR is larger than the critical value of this distribution depending on the level of the test.

Table 2 shows the size properties for the test in small samples based on 100,000 Monte Carlo replications. The data generating process is given by equation (48) below with $\sigma_v = 1$ and $\gamma = 0$. This corresponds to the setup in our Monte Carlo simulations in the next section. It can be seen that the test is oversized for small samples, but for $n = 200$ or larger the asymptotic distribution is an appropriate approximation for the distribution of the LR statistic.

4 A simulation study

In this section we present the results of a Monte Carlo study to compare the behavior of our proposed model with the classical stochastic frontier model. We do this for the normal-half normal and normal-exponential models. We are interested in how well the models estimate the location of the production frontier 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

$$Y_i = \alpha_0 + \alpha_1 \log X_{1i} + \alpha_2 \log X_{2i} + V_i - U_i, \; i = 1, \ldots, n, \tag{48}$$

where $V_i \sim N(0, \sigma_v^2)$, $U_i \sim \text{Exp}(\gamma)$ or $U_i \sim N^+(0, \gamma)$, $\log X_{1i} \sim N(1.5, 0.3)$ 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 γ take on the values 0.3, 0.4, and 0.5 corresponding to varying degrees of average technical efficiency. The standard deviation of the two-sided error σ_v is chosen as $\sigma_v = 0.25$ for the half-normal and as $\sigma_v = 0.5$ for the exponential case. Different values were chosen to ensure that the fraction of samples with positive skewness is similar in the two situations. We consider the sample sizes $n = 50, 100, 200$. We report the average bias and mean-square-error (MSE) for two parameters of interest. The first is the estimate of technical efficiency defined as $TE = \text{E}(\exp(-U))$ (true values obtained by simulation) and estimated as $n^{-1} \sum_{i=1}^{n} \widehat{\text{E}}(\exp(-U_i)|W_i)$. The second parameter of interest is the intercept α_0 as this represents the location of the frontier. We also report the fraction of samples having positive skewness (column: "pos. skew."). Results for "Standard" refer to the classical (one-sided) stochastic frontier model and for "Extended" to our extended (two-sided) version. The number of Monte Carlo replications is equal to 10,000.

The results are reported in Tables 3 to 6. For the chosen specifications the fraction of samples with positive skewness ranges from 31% to 0.3%. The advantage of our approach becomes apparent when looking at the bias for technical efficiency, which is always smaller for the extended model, except for the half-normal case with $n = 50$ and $\gamma = 0.3$. This can be explained by the fact that the classical model cannot handle positive skewness and estimates technical efficiency to be equal to one in these cases. For the settings for which the fraction of samples with positive skewness is close to zero, however, the two models basically give identical estimates. The MSE here gives a very similar picture, although for the exponential case there are several cases in which the MSE is actually smaller for the standard model.

The intercept α_0 is also estimated with a smaller bias for our model. This shows the

Table 3: Bias normal-half normal model

		bias TE		bias α_0	
$\gamma = 0.3$	pos. skew.	Standard	Extended	Standard	Extended
n=50	0.3103	0.044	-0.051	-0.045	0.084
n=100	0.2256	0.039	-0.023	-0.041	0.037
n=200	0.1405	0.029	-0.005	-0.032	0.011
$\gamma = 0.4$					
n=50	0.2083	0.047	-0.018	-0.048	0.042
n=100	0.1027	0.029	-0.001	-0.033	0.006
n=200	0.0349	0.016	0.007	-0.018	-0.007
$\gamma = 0.5$					
n=50	0.1358	0.044	0.006	-0.049	0.007
n=100	0.0466	0.022	0.009	-0.028	-0.011
n=200	0.0056	0.008	0.006	-0.012	-0.010
$\gamma = 0.3$		Only pos. skewed samples			
n=50		0.201	-0.102	-0.241	0.173
n=100		0.201	-0.071	-0.238	0.109
n=200		0.201	-0.047	-0.243	0.066
$\gamma = 0.4$					
n=50		0.254	-0.054	-0.314	0.119
n=100		0.254	-0.039	-0.319	0.067
n=200		0.254	0.012	-0.317	-0.011
$\gamma = 0.5$					
n=50		0.301	0.017	-0.395	0.017
n=100		0.301	0.034	-0.398	-0.035
n=200		0.301	0.030	-0.398	-0.043

Note: Table 3 presents Monte Carlo estimates of the bias for the estimates of average technical efficiency (TE) and the regression constant (α_0) for the model given in (48) with one-sided errors coming from a half-normal distribution. The column "Pos. skew" is the fraction of samples that are characterized by the wrong skewness problem. The columns labeled "Standard" refer to the classical stochastic frontier model, whereas "Extended" refers to the model introduced in Section 2.1. The lower panel only considers those samples characterized by positive skewness. The results are based on 10,000 Monte Carlo replications.

Table 4: MSE normal-half normal model

	MSE TE		MSE α_0	
$\gamma = 0.3$	Standard	Extended	Standard	Extended
n=50	0.0163	0.0116	0.0679	0.0755
n=100	0.0119	0.0069	0.0373	0.0345
n=200	0.0074	0.0039	0.0208	0.017
$\gamma = 0.4$				
n=50	0.0176	0.0092	0.0832	0.0846
n=100	0.0095	0.0049	0.0403	0.0359
n=200	0.0044	0.0027	0.0194	0.0175
$\gamma = 0.5$				
n=50	0.017	0.0093	0.1011	0.1009
n=100	0.0074	0.0044	0.0435	0.0406
n=200	0.0022	0.0017	0.0191	0.0183
$\gamma = 0.3$	Only pos. skewed samples			
n=50	0.0403	0.0251	0.0996	0.1244
n=100	0.0403	0.0185	0.075	0.0624
n=200	0.0403	0.0136	0.0673	0.039
$\gamma = 0.4$				
n=50	0.0642	0.0235	0.1545	0.1606
n=100	0.0642	0.0195	0.1239	0.0813
n=200	0.0642	0.0181	0.1053	0.0535
$\gamma = 0.5$				
n=50	0.0905	0.0323	0.2208	0.2194
n=100	0.0905	0.0279	0.1792	0.1168
n=200	0.0905	0.0183	0.1651	0.0543

Note: Table 4 presents Monte Carlo estimates of the mean-square-error for the estimates of average technical efficiency (TE) and the regression constant (α_0) for the model given in (48) with one-sided errors coming from a half-normal distribution. The columns labeled "Standard" refer to the classical stochastic frontier model, whereas "Extended" refers to the model introduced in Section 2.1. The lower panel only considers those samples characterized by positive skewness. The results are based on 10,000 Monte Carlo replications.

Table 5: Bias normal-exponential model

		bias TE		bias α_0	
$\gamma = 0.3$	pos. skew.	Standard	Extended	Standard	Extended
n=50	0.2903	0.063	-0.056	-0.080	0.111
n=100	0.2084	0.050	-0.028	-0.063	0.053
n=200	0.1087	0.033	-0.004	-0.039	0.012
$\gamma = 0.4$					
n=50	0.1820	0.060	-0.006	-0.071	0.032
n=100	0.0825	0.035	0.011	-0.046	-0.011
n=200	0.0202	0.016	0.011	-0.023	-0.016
$\gamma = 0.5$					
n=50	0.1124	0.049	0.014	-0.069	-0.016
n=100	0.0308	0.025	0.017	-0.034	-0.023
n=200	0.0031	0.009	0.009	-0.014	-0.013
$\gamma = 0.3$		Only pos. skewed samples			
n=50		0.231	-0.177	-0.298	0.359
n=100		0.231	-0.144	-0.303	0.254
n=200		0.231	-0.102	-0.285	0.187
$\gamma = 0.4$					
n=50		0.286	-0.074	-0.373	0.191
n=100		0.286	-0.016	-0.399	0.029
n=200		0.286	0.018	-0.403	-0.038
$\gamma = 0.5$					
n=50		0.333	0.018	-0.480	-0.003
n=100		0.333	0.065	-0.470	-0.109
n=200		0.333	0.072	-0.400	-0.040

Note: Table 5 presents Monte Carlo estimates of the bias for the estimates of average technical efficiency (TE) and the regression constant (α_0) for the model given in (48) with one-sided errors coming from an exponential. The column “Pos. skew” is the fraction of samples that are characterized by the wrong skewness problem. The columns labeled “Standard” refer to the classical stochastic frontier model, whereas “Extended” refers to the model introduced in Section 2.1. The lower panel only considers those samples characterized by positive skewness. The results are based on 10,000 Monte Carlo replications.

Table 6: MSE normal-exponential model

	MSE TE		MSE α_0	
$\gamma = 0.3$	Standard	Extended	Standard	Extended
n=50	0.0253	0.0317	0.1846	0.2998
n=100	0.0155	0.0167	0.0950	0.1329
n=200	0.0093	0.0080	0.0489	0.0577
$\gamma = 0.4$				
n=50	0.0279	0.0233	0.2118	0.2742
n=100	0.0124	0.0092	0.1044	0.1143
n=200	0.0049	0.0038	0.0487	0.0490
$\gamma = 0.5$				
n=50	0.0260	0.0203	0.2339	0.2635
n=100	0.0087	0.0066	0.1093	0.1124
n=200	0.0024	0.0022	0.0498	0.0497
$\gamma = 0.3$	Only pos. skewed samples			
n=50	0.0533	0.0752	0.2216	0.613
n=100	0.0533	0.0594	0.1521	0.3438
n=200	0.0533	0.0414	0.1123	0.1929
$\gamma = 0.4$				
n=50	0.0816	0.0568	0.3119	0.6456
n=100	0.0816	0.0440	0.2279	0.3459
n=200	0.0816	0.0317	0.1761	0.1875
$\gamma = 0.5$				
n=50	0.1111	0.0552	0.4023	0.6969
n=100	0.1111	0.0467	0.3386	0.4329
n=200	0.1111	0.0426	0.2911	0.2560

Note: Table 6 presents Monte Carlo estimates of the mean-square-error for the estimates of average technical efficiency (TE) and the regression constant (α_0) for the model given in (48) with one-sided errors coming from an exponential distribution. The columns labeled “Standard” refer to the classical stochastic frontier model, whereas “Extended” refers to the model introduced in Section 2.1. The lower panel only considers those samples characterized by positive skewness. The results are based on 10,000 Monte Carlo replications.

problem that a wrong skewness affects the estimation of the location of the frontier as well and that this problem can again be mitigated using our extension. In terms of the MSE the results of both models are quite similar.

In the lower panels of Tables 3 and 6 we report the results only for samples that were characterized by positive skewness. It can be seen that the extended model performs quite well in terms of bias, whereas the classical model gives biased estimates for average technical efficiency, which for these samples is always estimated to be 1, and for the location of the frontier. It should be noted that for small γ and small n, TE is underestimated, which is expected because in case of a wrong skewness only few observations are allowed to be close to the frontier by the construction of the model extension; see the lower panel of Figure 1. Therefore it is remarkable that for larger γ and n this bias is small when considering only the samples characterized by the "wrong skewness". In terms of the MSE the extended model clearly has a superior performance for estimating technical efficiency. The MSE for estimating a_0 is smaller when using the extended model in the normal-half normal case, but mostly larger in the normal-exponential case.

The bias for the remaining parameters are not reported. Nevertheless, we mention that the slope coefficients α_1 and α_2 are estimated without bias for both models. Finally, the bias for σ_v is roughly the same for the two models.

5 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 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.

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 48 years $\widehat{\mu}_{3,n}$ 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 6 and 7 show plots of the estimated average 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. However, it is striking that the estimates of technical efficiencies in these years are quite small. This is likely caused by the fact that under the wrong skewness less firms are allowed near the frontier, resulting in a potential underestimation of technical efficiency[7]. A possible explanation for our findings 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.

6 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

[7] We would like to thank an anonymous referee for pointing this out.

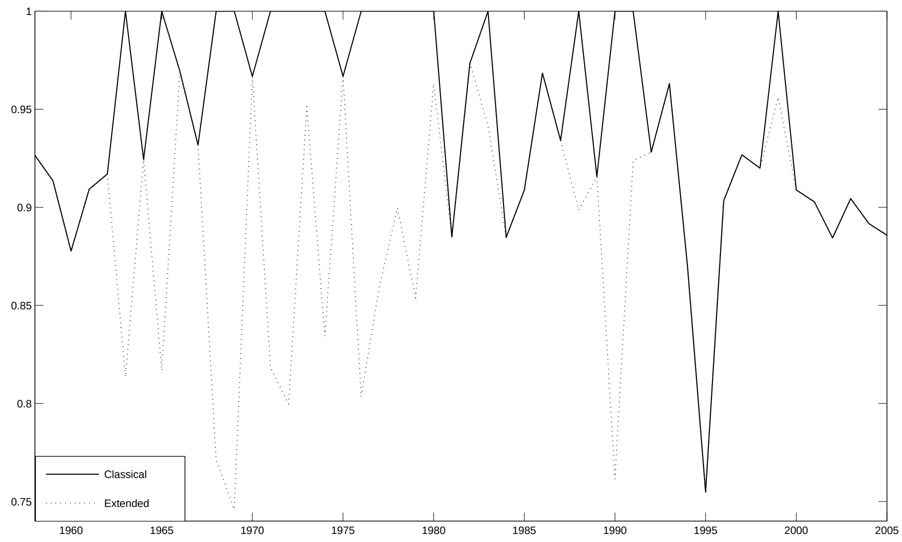

Figure 6: *Technical efficiency estimated by the normal-exponential stochastic frontier model (solid line) and our extended model (dashed line) for the years 1958-2005.*

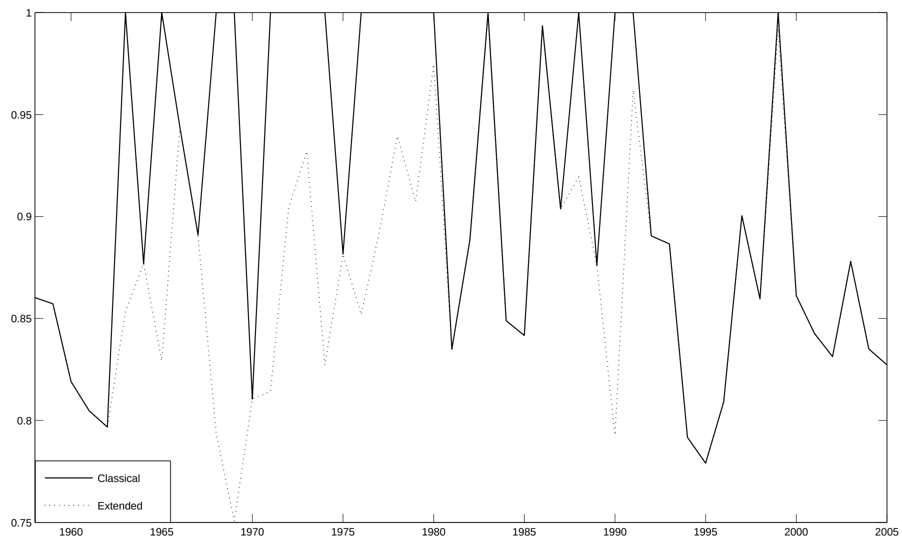

Figure 7: *Technical efficiency estimated by the normal-half normal stochastic frontier model (solid line) and our extended model (dashed line) for the years 1958-2005.*

and nesting, as a particular case, the traditional SFA model. The statistical properties of our model include continuity of the log-likelihood for any value of the scale parameter of the inefficiency distribution, in particular also at zero. As a consequence the MLE is consistent for any parameter value and standard likelihood ratio tests can be applied to test for the presence of inefficiencies. Future research may compare this test to alternative tests in the literature; see, e.g., Ahmad and Li (1997) or Kuosmanen and Fosgerau (2009). Our Monte-Carlo experiments show that the extended model performs favorably for estimating the technical efficiencies compared with the traditional SFA model, as it provides reasonable estimates of firm efficiencies in the presence of "wrong skewness". Therefore the model we propose enriches the toolbox of researchers for performing efficiency analyses with parametric SFA models.

A Appendix: Proof of Theorem 1

A.1 The Normal-Halfnormal Case

The analytical derivations in this case are rather straightforward. When $\gamma > 0$ our model of W is

$$g^+(w) = \frac{2}{\sigma}\phi\left(\frac{w}{\sigma}\right)\Phi\left(\frac{-w}{\sigma}\frac{\gamma}{\sigma_v}\right)$$

where $\sigma^2 = \gamma^2 + \sigma_v^2$. For the case $\gamma < 0$, the density of W is

$$g^-(w) = \frac{1}{\sigma(\Phi(a_0) - \Phi(0))}\phi\left(\frac{w - a_0\gamma}{\sigma}\right)\left[\Phi\left(A_w + \frac{a_0\sigma}{\sigma_v}\right) - \Phi(A_w)\right],$$

where $A_w = \frac{w-a_0\gamma}{\sigma}\frac{\gamma}{\sigma_v}$. It is very easy to check that, $\lim_{\gamma\to 0^+} g^+(w) = \lim_{\gamma\to 0^-} g^(w) = (1/\sigma_v)\phi(w/\sigma_v)$. So defining the density $g(w)$ at $\gamma = 0$ as this common normal limiting distribution, the continuity of $g(w)$ at $\gamma = 0$ follows.

Concerning the score in γ of the two sub-models, the partial derivatives of the $\log(g)$ for the two sub-models can be computed directly. Careful derivations lead to

$$\begin{aligned}
\frac{\partial \log g^+(w)}{\partial\gamma} &= -\frac{\gamma}{\sigma^2} + \frac{\gamma w^2}{\sigma^4} - \frac{w\sigma_v}{\sigma^3}\frac{\phi\big(-(w\gamma)/(\sigma\sigma_v)\big)}{\Phi\big(-(w\gamma)/(\sigma\sigma_v)\big)};\\
\frac{\partial \log g^-(w))}{\partial\gamma} &= \frac{\gamma}{\sigma^2} + \frac{\left[\sigma^2 a_0(w + a_0\gamma) - \gamma(w + a_0\gamma)^2\right]}{\sigma^4}\\
&\quad -\frac{1}{\sigma^3}(w + a_0\gamma)\sigma_v\eta^-(\theta) - \frac{a_0\gamma}{\sigma_v\sigma}\frac{\phi(A_w)}{\Phi\big(A_w + a_0\sigma/\sigma_v\big) - \Phi\big(A_w\big)},
\end{aligned}$$

where $\eta^{-}(\theta) = \dfrac{\phi\big(A_w + a_0\sigma/\sigma_v\big) - \phi\big(A_w\big)}{\Phi\big(A_w + a_0\sigma/\sigma_v\big) - \Phi\big(A_w\big)}$. Then it is easy to see that

$$\lim_{\gamma\to 0^+} \frac{\partial \log g^+(w)}{\partial \gamma} = - \lim_{\gamma\to 0^-} \frac{\partial \log g^-(w)}{\partial \gamma} = -\frac{w}{\sigma_v^2}\frac{\phi(0)}{\Phi(0)},$$

where the last equality comes from the definition of a_0 below equation (18) for the two densities of U having the same mean. This completes the proof.

A.2 The Normal-Exponential Case

The analytical derivations for the Normal-Exponential case are more tedious due to the presence of several indeterminacy involving the use of the L'Hospital rule. We have for $\gamma > 0$ and $\gamma < 0$, respectively,

$$g^+(w) = \frac{1}{\gamma}\frac{\Phi(-C_w)}{\phi(-C_w)}\phi\Big(\frac{w}{\sigma_v}\Big),$$
$$g^-(w) = \frac{e^{-a_0}}{\gamma(e^{-a_0} - 1)}\frac{\Phi(-C_w) - \Phi(-C_w + a_0\gamma/\sigma_v)}{\phi(-C_w)}\phi\Big(\frac{w}{\sigma_v}\Big),$$

where $C_w = w/\sigma_v + \sigma_v/\gamma$. It can be seen that $\lim_{\gamma\to 0^+} g^+(w) = \phi(w/\sigma_v) \lim_{\gamma\to 0^+} \dfrac{\Phi(-C_w)/\gamma}{\phi(-C_w)}$. By the L'Hospital rule the latter limit converges to $1/\sigma_v$, so we have

$$\lim_{\gamma\to 0^+} g^+(w) = (1/\sigma_v)\phi(w/\sigma_v).$$

Similarly it can be found that $\lim_{\gamma\to 0^-} g^-(w) = c_0\phi(w/\sigma_v) \lim_{\gamma\to 0^-} \dfrac{\Phi(-C_w) - \Phi(-C_w + a_0\gamma/\sigma_v)}{\gamma\phi(-C_w)}$. Again by L'Hospital's rule and after some analytical derivations, the latter limit can be shown to converge to $(1 - e^{a_0})/\sigma_v = (e^{-a_0} - 1)/(e^{-a_0}\sigma_v)$. Thus we again have

$$\lim_{\gamma\to 0^-} g^-(w) = (1/\sigma_v)\phi(w/\sigma_v),$$

which proves the continuity of $g(w)$ defined in (35) w.r.t. γ, even at $\gamma = 0$ by defining $g(w;\gamma = 0) = g^0(w) = (1/\sigma_v)\phi(w/\sigma_v)$.

For the score function w.r.t. γ the easiest way to organize the calculations is to start with the densities. For $\gamma > 0$ we have the right derivative

$$\left.\frac{\partial g^+(w;\gamma)}{\partial \gamma}\right|_{0^+} = \lim_{\gamma\to 0^+} \frac{g^+(w;\gamma) - g^0(w)}{\gamma}.$$

Simple calculations lead to

$$\left.\frac{\partial g^+(w;\gamma)}{\partial \gamma}\right|_{0^+} = \phi(w/\sigma_v) \lim_{\gamma\to 0^+} \frac{(1/\gamma)\Phi(-C_w) - (1/\sigma_v)\phi(-C_W)}{\gamma\phi(-C_w)}.$$

L'Hospital's rule and the fact that the Mill's ratio $\Phi(-C_w)/\phi(-C_w) \to 0$ when $\gamma \to 0^+$, give for the latter limit (and after some careful calculations) the value $-w/\sigma_v^3$. So we have

$$\left.\frac{\partial g^+(w;\gamma)}{\partial \gamma}\right|_{0^+} = -\frac{w}{\sigma_v^3}\phi\Big(\frac{w}{\sigma_v}\Big).$$

When $\gamma < 0$ we have for the left derivative at zero by similar but more tedious derivations that we summarize below:

$$\left.\frac{\partial g^-(w;\gamma)}{\partial \gamma}\right|_{0^-} = \lim_{\gamma\to 0^-} \frac{g^(w;\gamma) - g^0(w)}{\gamma}.$$

This can easily be transform in

$$\left.\frac{\partial g^-(w;\gamma)}{\partial \gamma}\right|_{0^-} = \phi(w/\sigma_v) \lim_{\gamma\to 0^-} \frac{(c_0/\gamma)\left[\Phi(-C_w) - \Phi(-C_w + a_0\gamma/\sigma_v)\right] - (1/\sigma_v)\phi(-C_W)}{\gamma\phi(-C_w)},$$

where $c_0 = e^{-a_0}/(e^{-a_0} - 1)$. This latter limit can be handled with L'Hospital's rule leading to the value $-\frac{w}{\sigma^3}\left[\frac{a_0}{e^{-a_0} - 1} + 1\right]$. Now by the definition of a_0 in (30) we see that $a_0 + 2(e^{-a_0} - 1) = 0$, so $\frac{a_0}{e^{-a_0} - 1} = -2$. Finally, for $\gamma < 0$ we have the left derivative

$$\left.\frac{\partial g^-(w;\gamma)}{\partial \gamma}\right|_{0^-} = +\frac{w}{\sigma_v^3}\phi\Big(\frac{w}{\sigma_v}\Big).$$

For getting the scores derived in Theorem 1, we just divide these derivatives by $g^0(w) = (1/\sigma_v)\phi(w/\sigma_v)$ and this completes the proof.

References

Ahmad, I. A. and Q. Li (1997). Testing the symmetry of an unknown density function by the kernel method. *Journal of Nonparametric Statistics 7*, 279–293.

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., J. Qian, and R. S. Sickles (2014). Stochastic frontier models with bounded inefficiency. In R. C. Sickles and W. C. Horrace (Eds.), *Econometric Methods and Applications*. Springer Science & Business Media, New York.

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.

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.

Kuosmanen, T. and M. Fosgerau (2009). Neoclassical versus frontier production models? testing for the skewness of regression residuals. *Scandinavian Journal of Economics 111*, 351–367.

Lee, L. F. (1993). Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singular 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.

Ritter, C. and L. Simar (1997). Pitfalls of normal-gamma stochastic frontier models. *Journal of Productivity Analysis 8*, 167–182.

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.