



The wrong skew problem in stochastic frontier models when inefficiency depends on environmental variables

Cheol-Keun Cho¹ · Peter Schmidt²

Received: 2 February 2018 / Accepted: 5 October 2018 / Published online: 16 October 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

A well-known result due to Waldman (J Econom 18:275–279, 1982) states that, in the standard normal/half-normal SFA model, estimated technical inefficiency will be zero if the OLS residuals are positively skewed. It is not clear how much this result generalizes. In this paper, we consider the normal/half-normal model in which the distribution of the half-normal error depends on explanatory variables. We consider estimation by nonlinear least squares and maximum likelihood. In both cases, we find a stationary point (zero derivatives) at parameter values that indicate zero inefficiency, a result similar to Waldman's. However, both for nonlinear least squares and for MLE, we show that in general the stationary point is neither a local minimum nor a local maximum.

Keywords Environmental variables · Stochastic frontier · Wrong skew

1 Introduction

Consider the stochastic frontier model

$$y_i = \alpha + x_i' \beta + v_i - u_i = \alpha + x_i' \beta + \varepsilon_i, \quad (1)$$

where x_i is “fixed” (independent of v_i and u_i); v_i is distributed as $N(0, \sigma_v^2)$; u_i is distributed as $N^+(0, \sigma_u^2)$, i.e., “half normal”; and v_i and u_i are independent. We will sometimes use the standard notation that $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\lambda = \sigma_u / \sigma_v$. This is the

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00181-018-1573-x>) contains supplementary material, which is available to authorized users.

✉ Peter Schmidt
schmidtp@msu.edu

¹ Korea Energy Economics Institute, Ulsan, South Korea

² Michigan State University, East Lansing, USA

model of Aigner et al. (1977) and Meeusen and van den Broeck (1977), and we will call it the standard SFA model.

The distribution of ε_i is “skew normal,” and the model can be estimated by MLE using the skew normal density. An alternative method of estimation is corrected OLS (COLS), which was suggested by Aigner, Lovell and Schmidt and further analyzed in Olson et al. (1980) and Waldman (1982), and which will be described below.

Under the assumptions made above, the third moment of ε_i is negative. However, its sample equivalent, the third moment of the OLS residuals, can be positive. This is the so-called *wrong skew* case. In the wrong skew case, two problems arise. The first, noted in the original Aigner, Lovell and Schmidt article, is that the COLS estimate does not exist. The second and more subtle problem, or set of problems, was noted by Waldman. The likelihood always (regardless of wrong or right skew) has a stationary point at the parameter values that reflect no inefficiency ($\lambda = 0$ and other parameters = the OLS estimates). At that stationary point, the Hessian is always singular. In the wrong skew case, this point is a local maximum of the likelihood function, and empirically in this case, this point is also the global maximum. Thus, with a positive probability the data will indicate no inefficiency. This complicates the process of inference, see Simar and Wilson (2009).

The question this paper addresses is whether similar problems occur in models in which the distribution of technical inefficiency depends on observable “environmental” variables. Specifically, we will consider a form of the RSCFG model of Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill et al. (1995). Here, we obtain results that are similar to but different from those for the standard SFA model. There is no COLS estimator, but the model can be estimated by nonlinear least squares, and the least squares criterion function always has a stationary point at the parameter values that reflect no inefficiency. The model can also be estimated by MLE, and the likelihood always has a stationary point and the Hessian is singular at the parameter values that reflect no inefficiency. In general, these stationary points are neither a local minimum nor a local maximum of the relevant criterion function (sum of squares or likelihood). None of these statements have any connection to the skew of the residuals.

This paper does not aim to give advice about how to proceed if in a particular data set the stationary point is the global minimum of the NLLS criterion or the maximum of the log-likelihood. Nor do we attempt to construct models in which a skew of the residuals of either sign is not wrong. Some papers that do these things will be discussed in the final section of the paper. We simply ask the questions of whether, for a given criterion function, there is always a stationary point and whether we can say that it is or is not a local maximum or minimum. These are rather specific questions, but as we will see it is not trivial to answer them.

2 More detail on the wrong skew problem in the standard SFA model

We will now be a little more precise than in the previous section about the nature of the wrong skew problem in the standard SFA model. We do this so that it is clear what results we might hope or expect to generalize the RSCFG model.

The model is as given in Eq. (1). It is well known that $E(u) \equiv \mu = \sqrt{\frac{2}{\pi}}\sigma_u$ so that $E(\varepsilon) = -\mu$. Also $\text{var}(u) = \frac{\pi-2}{\pi}\sigma_u^2$, $E(u^2) = \sigma_u^2$ and $\mu'_3 \equiv E(\varepsilon + \mu)^3 = -E(u - \mu)^3 = \sqrt{\frac{2}{\pi}}\left(\frac{\pi-4}{\pi}\right)\sigma_u^3$. Note that $\mu'_3 \leq 0$. OLS implicitly estimates $(\alpha - \mu)$ and β , and the OLS residuals e_i correspondingly are “estimates” of $v_i - (u_i - \mu)$. So

$$\hat{\sigma}^2 \equiv \frac{1}{N} \sum_i e_i^2 \rightarrow_p \sigma^2 = \sigma_v^2 + \sigma_u^2 \text{ and } \hat{\mu}'_3 \equiv \frac{1}{N} \sum_i e_i^3 \rightarrow_p \mu'_3 = \sqrt{\frac{2}{\pi}}\left(\frac{\pi-4}{\pi}\right)\sigma_u^3. \tag{2}$$

Therefore, we obtain consistent estimates of σ_u^2 and σ_v^2 as

$$\hat{\sigma}_u^2 = \left[\sqrt{\frac{\pi}{2}}\left(\frac{\pi}{\pi-4}\right)\hat{\mu}'_3 \right]^{-2/3}, \quad \hat{\sigma}_v^2 = \hat{\sigma}^2 - \hat{\sigma}_u^2. \tag{3}$$

If $\hat{\alpha}, \hat{\alpha}$ are the least squares estimates, the COLS estimates of α and β are $\tilde{\alpha} = \hat{\alpha} + \sqrt{\frac{2}{\pi}}\hat{\sigma}_u$ and $\tilde{\beta} = \hat{\beta}$. However, our interest in this paper is just in the estimates of σ_u^2 and σ_v^2 (as given in Eq. (3)) themselves. In the wrong skew case that $\hat{\mu}'_3 > 0$, in Eq. (3) we have $\hat{\sigma}_u < 0$ and $\hat{\sigma}_u^2$ is not really well defined. So the COLS method fails.

With respect to the MLE, the situation is more complicated. Waldman (1982) showed that the point $\hat{\alpha}, \hat{\beta} = \text{OLS}, \hat{\sigma}_u^2 = 0, \hat{\sigma}_v^2 = \hat{\sigma}^2$ is always a stationary point of the likelihood. He also showed that the information matrix is singular at this point. Finally, Waldman showed that the stationary point given above is a local maximum of the likelihood when $\hat{\mu}'_3 > 0$. It is generally thought that it is also the global maximum.

The wrong skew problem occurs most frequently when the sample size is small and the population value of $\lambda = \sigma_u/\sigma_v$ is small. Many people might find the frequency with which it occurs to be surprising. For example, Simar and Wilson (2009, Table 1, p. 71) report simulations in which the probability of a wrong skew is 0.301 when $n = 100$ and $\lambda = 1$; it is 0.320 when $n = 500$ and $\lambda = 0.5$; and it is 0.386 when $n = 10,000$ and $\lambda = 0.1$.

3 The RSCFG model

We will now consider the case that the distribution of technical inefficiency (u_i) depends on some observable “environmental variables” z_i that may or may not affect the level of the frontier but that do affect the level of technical inefficiency. A possible example of such a variable in an agricultural setting would be ownership of the farm (private vs. state-owned).

The most commonly assumed case is that the distribution of u_i is truncated normal. In standard notation, u_i is distributed as $N^+(\mu_i, \sigma_i^2)$. When $\mu_i = 0$ and σ_i^2 is constant (does not depend on i), we have the standard stochastic frontier model of the previous section. When μ_i and σ_i^2 are constant, we have the truncated normal model of Stevenson (1980). However, here we are interested in models in which μ_i and/or σ_i^2 depend

on environmental variables z_i . For example, in the RSCFG model of Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill et al. (1995), $\mu_i = 0$ and σ_i^2 is a function of z_i and some parameters. In the KGMHLBC model of Kumbhakar et al. (1991), Huang and Liu (1994) and Battese and Coelli (1995), σ_i^2 is constant (does not depend on i) and μ_i is a function of z_i and parameters. In the model of Wang (2002), both μ_i and σ_i^2 depend on z_i and parameters. In the model of Alvarez et al. (2006), there is a “scaling function” $g(z_i, \theta)$ such that $\mu_i = \mu \cdot g(z_i, \theta)$ and $\sigma_i = \sigma \cdot g(z_i, \theta)$. A related model is the model of Amsler et al. (2015) in which the post-truncation mean and variance of u_i are parameterized.

In this paper, we will consider the specific case of the RSCFG model with $\sigma_i = \sigma_u \exp(z_i' \delta)$. We treat x_i and z_i as “fixed” (independent of v_i and u_i) so that the model is:

$$y_i = \alpha + x_i' \beta + v_i - u_i, \quad v_i \sim N(0, \sigma_v^2), \quad u_i \sim N^+(0, \sigma_u^2 \exp(2z_i' \delta)). \quad (4)$$

This is a straightforward extension of the standard stochastic frontier model because u_i is still half normal. We will use the notation $d_z = \text{dimension}(z_i)$ and $d_x = \text{dimension}(x_i)$.

4 Nonlinear least squares estimation of the RSCFG model

Given the RSCFG model with $\sigma_i = \sigma_u \exp(z_i' \delta)$, we have $E(y_i | x_i, z_i) = \alpha + x_i' \beta - \sqrt{\frac{2}{\pi}} \sigma_u \exp(z_i' \delta)$. This suggests a nonlinear least squares (NLLS) estimator that minimizes (with respect to α , β , σ_u and δ) the criterion function

$$\text{SSE} = \sum_i \left[y_i - \alpha - x_i' \beta + \sqrt{\frac{2}{\pi}} \sigma_u \exp(z_i' \delta) \right]^2. \quad (5)$$

We will denote the NLLS estimates as $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\sigma}_u$ and $\tilde{\delta}$.

We can note a few points about identification of the model based on this criterion function. First, obviously δ is not identified when $\sigma_u = 0$. Second, α and σ_u are not separately identified when $\delta = 0$. Third, this criterion function does not lead directly to an estimate of σ_v^2 , but we can derive an estimate based on the NLLS estimates. Note that if $u_i \sim N^+(0, \sigma_i^2)$, then $E(u_i^2) = \sigma_i^2$ and so $E(\varepsilon_i^2) = \sigma_v^2 + \sigma_i^2 = \sigma_v^2 + \sigma_u^2 \exp(2z_i' \delta)$. If $\tilde{\varepsilon}_i = y_i - \tilde{\alpha} - x_i' \tilde{\beta}$, this leads to the estimate

$$\tilde{\sigma}_v^2 = \frac{1}{n} \sum_i \tilde{\varepsilon}_i^2 - \tilde{\sigma}_u^2 \frac{1}{n} \sum_i \exp(2z_i' \tilde{\delta}). \quad (6)$$

This is similar in spirit to the estimator given in (3) for the standard SFA model.

Let $\psi = \begin{bmatrix} \alpha \\ \beta \\ \sigma_u \\ \delta \end{bmatrix}$, the parameters that we seek to estimate by NLLS. Let $\psi^* = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \\ 0 \\ 0 \end{bmatrix}$, where $\hat{\alpha}, \hat{\beta} = \text{OLS of } y \text{ on intercept and } x$. Obviously, ψ^* is a set of parameters

that indicate no inefficiency. Note that we carefully said “a set of parameters that indicate no inefficiency” rather than “the set of parameters that indicate no inefficiency” because when $\sigma_u = 0$ we have no inefficiency regardless of the value of δ .

Result 1 The criterion SSE given in Eq. (5) has a stationary point at ψ^* .

Proof The derivatives of the NLLS criterion function with respect to the parameters in ψ are:

$$\nabla_{\alpha} \text{SSE} = -2 \sum_i \left[y_i - \alpha - x'_i \beta + \sqrt{\frac{2}{\pi}} \sigma_u \exp(z'_i \delta) \right], \tag{7A}$$

$$\nabla_{\beta} \text{SSE} = -2 \sum_i \left\{ \left[y_i - \alpha - x'_i \beta + \sqrt{\frac{2}{\pi}} \sigma_u \exp(z'_i \delta) \right] x_i \right\}, \tag{7B}$$

$$\nabla_{\sigma_u} \text{SSE} = 2 \sqrt{\frac{2}{\pi}} \sum_i \left\{ \left[y_i - \alpha - x'_i \beta + \sqrt{\frac{2}{\pi}} \sigma_u \exp(z'_i \delta) \right] \exp(z'_i \delta) \right\}, \tag{7C}$$

$$\nabla_{\delta} \text{SSE} = 2 \sqrt{\frac{2}{\pi}} \sigma_u \sum_i \left\{ \left[y_i - \alpha - x'_i \beta + \sqrt{\frac{2}{\pi}} \sigma_u \exp(z'_i \delta) \right] \exp(z'_i \delta) z_i \right\}. \tag{7D}$$

These derivatives are all zero at ψ^* . (The derivatives in (7A), (7B) and (7D) equal zero when $\alpha = \hat{\alpha}, \beta = \hat{\beta}$ and $\sigma_u = 0$ regardless of the value of δ , but $\delta = 0$ is required for the derivative in (7C) to equal zero.) So this point is a stationary point of the NLLS criterion function.

Next we ask whether there is any readily interpretable condition (analogous to “wrong skew” in the standard SFA model) such that this stationary point is a local minimum of the NLLS criterion function. The Hessian (second derivative) matrix H is messy and is given in Appendix 1. The Hessian evaluated at the stationary point ψ^* is simpler and is given by:

$$\frac{1}{2} H(\psi^*) = \begin{bmatrix} n & n\bar{x}' & -kn & 0 \\ n\bar{x} & \sum_i x_i x'_i & -kn\bar{x} & 0 \\ -kn & -kn\bar{x}' & k^2 n & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{8}$$

In this expression, $k = \sqrt{2/\pi}$. The four block rows (and columns) correspond to α, β, σ_u and δ , and they are of dimension 1, $d_x, 1$ and d_z , respectively. The matrix is singular, with rank equal to $d_x + 1$, so there are $d_z + 1$ eigenvalues equal to zero.

The Hessian $H(\psi^*)$ is positive semi-definite, which is a necessary condition for ψ^* to be a local minimum. However, because the Hessian is singular, this is not a sufficient condition. As in Waldman (1982), we need to examine the values of the criterion function in the directions that correspond to the eigenvectors associated with the zero eigenvalues. To elaborate on this point, consider the Taylor series expansion of SSE around the point ψ^* :

$$\text{SSE}(\psi) - \text{SSE}(\psi^*) = (\nabla_{\psi} \text{SSE}(\psi^*))'(\psi - \psi^*) + 1/2(\psi - \psi^*)' H(\psi^*)(\psi - \psi^*) + \text{higher-order terms.} \tag{9}$$

The first term on the right-hand side of (9) equals zero because $\nabla_{\psi} \text{SSE}(\psi^*) = 0$, so we consider the second term, $(\psi - \psi^*)' H(\psi^*)(\psi - \psi^*)$. A necessary condition for ψ^* to be a local minimum is that $H(\psi^*)$ is positive semi-definite, since, if there exists a vector g such that $g' H(\theta^*) g < 0$, then for small enough scalar $\tau > 0$, $\psi = \psi^* + \tau g$ will lead to a smaller SSE than ψ^* . However, for vectors that are linear combinations of the eigenvectors corresponding to the zero eigenvalues (i.e., vectors in the null space of $H(\psi^*)$) the second-order term above is zero, and we need to investigate the behavior of the criterion function in those directions, since the higher-order terms could be of either sign.

The eigenvectors that correspond to the zero eigenvalues, and which span the null space of $H(\psi^*)$, are as follows:

$$\begin{bmatrix} 1 \\ 0 \\ 1/k \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \iota_1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \iota_2 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \iota_{d_z} \end{bmatrix}. \tag{10}$$

Here $\frac{1}{k} = \sqrt{\frac{\pi}{2}}$; ι_j is a vector of zeroes except for a one in position j , and the dimensions of the four blocks in the vectors in (10) are 1, $d_x, 1$ and d_z . Therefore, a

vector in the null space of $H(\psi^*)$ will be of the form $w = \begin{bmatrix} 1 \\ 0 \\ 1/k \\ \delta \end{bmatrix}$ where in this expression δ is arbitrary (not necessarily the true parameter value).

Now we consider what happens to the least squares criterion function for small moves in this direction. That is, we consider a parameter value $\psi^o = \psi^* + \tau w$ for small $\tau > 0$. (We require $\tau > 0$ because $\sigma_u > 0$.) Then, we calculate the following:

$$\text{SSE}(\psi^*) = \text{usual least squares SSE} = \sum_i e_i^2, e_i = \text{OLS residuals}, \tag{11A}$$

$$\text{SSE}(\psi^o) = \sum_i [e_i + \tau (\exp(\tau z_i' \delta) - 1)]^2, \tag{11B}$$

$$\Delta \equiv \text{SSE}(\psi^o) - \text{SSE}(\psi^*) = \tau^2 \sum_i \left[\exp(\tau z'_i \delta) - 1 \right]^2 + 2\tau \sum_i e_i \exp(\tau z'_i \delta). \tag{11C}$$

For local movements (small τ), the term of order τ will dominate the term of order τ^2 , and so a necessary and sufficient condition for ψ^* to be a local minimum of SSE is:

$$2\tau \sum_i e_i \exp(\tau z'_i \delta) \geq 0 \text{ for all } \delta. \tag{12}$$

This is a strong requirement that intuitively should not be expected to hold. And it does not, as the following result shows.

Result 2 If $\sum_i z'_i e_i \neq 0$, the point ψ^* is neither a local minimum nor a local maximum of SSE.

Proof We have $\Delta = 2\tau \sum_i e_i \exp(\tau z'_i \delta) + \tau^2 \sum_i \left[\exp(\tau z'_i \delta) - 1 \right]^2$ as given in Eq. (11C). (For small τ , the first term will be the one that matters, which the calculation we are about to do will verify.) We make use of the Taylor series expansion $\exp(x) = 1 + x + 1/2x^2 + \text{h.o.t.}$ (higher-order terms), which yields

$$\Delta = 2\tau \sum_i e_i \left[1 + \tau z'_i \delta + 1/2\tau^2 (z'_i \delta)^2 \right] + \tau^2 \sum_i \left[\tau z'_i \delta + 1/2\tau^2 (z'_i \delta)^2 \right]^2 + \text{h.o.t.} \tag{13}$$

Since $\sum_i e_i = 0$, when $\sum_i z'_i e_i \neq 0$ the dominant term is $2\tau^2 \sum_i (z'_i e_i) \delta$.

This term can be made of either sign by appropriate choice of δ . For example, suppose that the first nonzero element of $\sum_i z'_i e_i$ is in position “j” and that it is positive. Then, if we pick δ to be equal to zero except for a value of one in position “j,” this term will be positive, and if instead we put a value of minus one in position “j,” the term will be negative. (Reverse the signs when the first nonzero element of $\sum_i z'_i e_i$ is negative.) So Δ can be of either sign, and the condition in Eq. (12) cannot hold.

If $\sum_i z'_i e_i = 0$, the dominant term in the expression for Δ will be $\tau^3 \sum_i e_i (z'_i \delta)^2$, and we cannot find anything useful to say about the sign of that term. However, we note that $\sum_i z'_i e_i = 0$ is an event of probability zero unless z_i is made up of linear combinations of x_i . This is formally possible but very unlikely in empirical practice, so this case is not of particular practical importance.

5 MLE of the RSCFG model

We now consider the MLE of the RSCFG model. The model is as given in Eq. (4).

We define the following notation:

$$\begin{aligned} \sigma_i &= \sigma_u \exp(z_i' \delta), \lambda_i = \sigma_i / \sigma_v, \omega_i^2 = \sigma_i^2 + \sigma_v^2, c_i = -\varepsilon_i \lambda_i / \omega_i, \\ \varphi_i &= \varphi(c_i), \Phi_i = \Phi(c_i), \end{aligned} \tag{14}$$

where φ is the standard normal density and Φ is the standard normal cdf. Then, the log-likelihood is given by

$$\ln L = \text{constant} - \frac{1}{2} \sum_i \ln \omega_i^2 - \frac{1}{2} \sum_i \frac{\varepsilon_i^2}{\omega_i^2} + \sum_i \ln \Phi_i, \tag{15}$$

where $\varepsilon_i = y_i - \alpha - x_i' \beta$.

Let $\theta = \begin{bmatrix} \alpha \\ \beta \\ \sigma_u \\ \delta \\ \sigma_v^2 \end{bmatrix}$, the vector of the parameters we wish to estimate. This differs

from ψ of the previous section because it includes σ_v^2 . Let $\theta^* = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \\ 0 \\ 0 \\ \hat{\sigma}_v^2 \end{bmatrix}$, our potential

stationary point, where $\hat{\alpha}$ and $\hat{\beta}$ are the OLS estimates and $\hat{\sigma}_v^2 = \frac{1}{n} \sum_i e_i^2$ where the e_i^2 are the OLS residuals. Then, we have the following result.

Result 3 The log-likelihood given in Eq. (14) has a stationary point at θ^* .

Proof The derivatives of the log-likelihood with respect to the parameters in θ are:

$$\nabla_\alpha \ln L = \sum_i \left[\frac{\varepsilon_i}{\omega_i^2} + \frac{\varphi_i \lambda_i}{\Phi_i \omega_i} \right], \tag{16A}$$

$$\nabla_\beta \ln L = \sum_i \left[\frac{\varepsilon_i}{\omega_i^2} + \frac{\varphi_i \lambda_i}{\Phi_i \omega_i} \right] x_i, \tag{16B}$$

$$\begin{aligned} \nabla_{\sigma_u} \ln L &= \sum_i \left[-\frac{\sigma_u}{\omega_i^2} \exp(2z_i' \delta) + \frac{\sigma_u}{\omega_i^4} \exp(2z_i' \delta) \varepsilon_i^2 - \frac{\varphi_i}{\Phi_i} \frac{1}{\sigma_v} \frac{\varepsilon_i}{\omega_i} \exp(z_i' \delta) + \sigma_u \frac{\varphi_i \lambda_i \varepsilon_i}{\Phi_i \omega_i^3} \exp(2z_i' \delta) \right], \end{aligned} \tag{16C}$$

$$\nabla_{\delta} \ln L = \sum_i z_i \left[-\frac{\varphi_i}{\Phi_i} \frac{\varepsilon_i \lambda_i}{\omega_i} - \frac{\sigma_u^2}{\omega_i^2} \exp(2z_i' \delta) + \sigma_u^2 \frac{\varepsilon_i^2}{\omega_i^4} \exp(2z_i' \delta) + \sigma_u^2 \frac{\varphi_i}{\Phi_i} \frac{\lambda_i \varepsilon_i}{\omega_i^3} \exp(2z_i' \delta) \right], \tag{16D}$$

$$\nabla_{\sigma_v^2} \ln L = \frac{1}{2} \sum_i \left[\frac{\varphi_i}{\Phi_i} \frac{1}{\sigma_v^2} \frac{\lambda_i \varepsilon_i}{\omega_i} - \frac{1}{\omega_i^2} + \frac{\varepsilon_i^2}{\omega_i^4} + \frac{\varphi_i}{\Phi_i} \frac{\lambda_i \varepsilon_i}{\omega_i^3} \right]. \tag{16E}$$

At θ^* , the following simplifications occur: $\sigma_u = 0, \lambda_i = 0, c_i = 0, \omega_i^2 = \sigma_v^2, \frac{\varphi_i}{\Phi_i} = \sqrt{\frac{2}{\pi}}$ and $\exp(z_i' \delta) = 1$. Then, we have $\nabla_{\alpha} \ln L = \frac{1}{\sigma_v^2} \sum_i \varepsilon_i, \nabla_{\beta} \ln L = \frac{1}{\sigma_v^2} \sum_i x_i \varepsilon_i, \nabla_{\sigma_u} \ln L = -\sqrt{\frac{2}{\pi}} \frac{1}{\sigma_v^2} \sum_i \varepsilon_i, \nabla_{\delta} \ln L = 0$ and $\nabla_{\sigma_v^2} \ln L = \frac{1}{2} \sum_i \left[-\frac{1}{\sigma_v^2} + \frac{1}{\sigma_v^4} \varepsilon_i^2 \right]$. All of these expressions equal zero when evaluated at the OLS values e_i and $\hat{\sigma}_v^2$.

Next we ask whether we can identify cases such that this may be a local maximum of the likelihood. We use a Taylor series expansion (similar to Eq. (9)) of $\ln L$ around the point θ^* :

$$\ln L(\theta) - \ln L(\theta^*) = (\nabla_{\theta} \ln L(\theta^*))'(\theta - \theta^*) + 1/2(\theta - \theta^*)' H(\theta^*)(\theta - \theta^*) + \text{higher-order terms.} \tag{17}$$

Here $H(\theta^*)$ is the Hessian (second derivative matrix) evaluated at θ^* . The first term on the right-hand side of (17) equals zero because $\nabla_{\theta} \ln L(\theta^*) = 0$, so we need to consider the second term, $(\theta - \theta^*)' H(\theta^*)(\theta - \theta^*)$. A necessary condition for θ^* to be a local maximum of $\ln L$ is that $H(\theta^*)$ is negative semi-definite, since, if there exists a vector v such that $v' H(\theta^*) v > 0$, then for small enough scalar $\tau > 0, \theta = \theta^* + \tau v$ will lead to a higher likelihood value than θ^* .

The Hessian is complicated and is given in a supplemental file available from the authors. However, at θ^* the simplifications listed above hold, and we obtain the following expression for the Hessian evaluated at θ^* .

$$H = -\frac{1}{\sigma_v^2} \begin{bmatrix} G & 0 \\ 0 & \frac{n}{2\sigma_v^2} \end{bmatrix}, \tag{18A}$$

$$G = \begin{bmatrix} A_{11} & A_{12} & -kA_{11} & 0 \\ A_{21} & A_{22} & -kA_{21} & 0 \\ -kA_{11} & -kA_{12} & k^2A_{11} & \xi \\ 0 & 0 & \xi' & 0 \end{bmatrix}. \tag{18B}$$

Here, $k = \sqrt{\frac{2}{\pi}}; \xi = k \sum_i z_i' e_i$ (a row vector); and

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} n & n\bar{x}' \\ n\bar{x} & \sum_i x_i x_i' \end{bmatrix} \tag{18C}$$

which is positive definite.

Suppose first that $\sum_i z'_i e_i \neq 0$ ($\xi \neq 0$), which will be the case with probability one unless z_i is made up of linear combinations of x_i . When $\xi \neq 0$, G and H are nonsingular if $d_z = 1$ (so that ξ is a scalar), and they are singular if $d_z \geq 2$. Regardless of the dimension of ξ , H will be negative semi-definite if and only if G is positive semi-definite. We now show that this is not the case.

Result 4 If $\xi \neq 0$, then G is neither positive semi-definite nor negative semi-definite.

Proof (i) Let $\eta_1 = [k, 0, 1, \xi]'$. Then, $\eta'_1 G \eta_1 = 2\xi\xi' > 0$ so G cannot be negative semi-definite. (ii) Now let $\eta_2 = [k, 0, 1, -\xi]'$. Then, $\eta'_2 G \eta_2 = -2\xi\xi' < 0$ so G cannot be positive semi-definite.

This shows that when $\sum_i z'_i e_i \neq 0$, θ^* cannot be a local maximum of the likelihood. (The necessary condition for θ^* to be a local maximum, namely that H should be negative semi-definite, fails when G is not positive semi-definite.) It also cannot be a local minimum.

Finally, we consider the case that $\xi = 0$ ($\sum_i z'_i e_i = 0$), which essentially means that z_i is made up of linear combinations of x_i . This case is very similar to the case considered in Sect. 4. The eigenvectors that correspond to the zero eigenvalues, and which span the null space of $H(\psi^*)$, are as follows:

$$\begin{bmatrix} 1 \\ 0 \\ 1/k \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ t_1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ t_2 \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ t_{d_z} \\ 0 \end{bmatrix}. \tag{19}$$

So a vector in the null space of $H(\psi^*)$ will be of the form $w = \begin{bmatrix} 1 \\ 0 \\ 1/k \\ \delta \\ 0 \end{bmatrix}$ for

arbitrary δ . Now we consider what happens to the log-likelihood function for small movements in this direction. That is, we consider a parameter value $\theta^\circ = \theta^* + \tau w$ for small $\tau > 0$. It is easy to calculate

$$\ln L(\theta^*) = \text{constant} - \frac{n}{2} - n \ln 2 - \frac{n}{2} \ln \hat{\sigma}_v^2. \tag{20}$$

However, $\ln L(\theta^* + \tau w)$ yields an explicit but impenetrable expression, which we give in Appendix 2. We do not see any way to say anything meaningful in this case. This is arguably not troublesome, because (as noted above) $\sum_i z'_i e_i = 0$ is a zero-probability event unless the elements of z_i are linear combinations of x_i .

6 Practical implications

It is certainly possible for the residuals (OLS or NLLS or MLE) to have a positive (wrong) skew. The main practical implication of our results is that this does not mean that there is a problem. There is always a stationary point (ψ^* , above) of the NLLS criterion that would indicate zero inefficiency, and similarly there is always a stationary point (θ^*) of the log-likelihood that would indicate zero inefficiency. However, in general these points are not a local minimum of the sum of squares or a local maximum of the log-likelihood. This does not appear to have anything to do with the skew of the residuals.

We will illustrate these issues with a small simulation. The DGP will be as simple as possible: $y_i = \alpha + v_i - u_i$, $i = 1, \dots, n$, where v_i is $N(0, \sigma_v^2)$ and (conditional on z_i) u_i is half normal with pre-truncation variance $\sigma_u^2 \cdot \exp(2z_i\delta)$. Here, z_i is $N(0,1)$, $\alpha = 0$, $\sigma_v^2 = 1$ and $n = 100$. The number of replications is 1000.

We use three different choices for δ and σ_u^2 . DGP1 has $\delta = 0$ and $\sigma_u^2 = 1$. This is the same DGP as for the standard SFA model with $\lambda = 1$. DGP2 has $\delta = 0.5$ and $\sigma_u^2 = 0.6065$ ($\sigma_u = 0.7788$), and DGP3 has $\delta = 1$ and $\sigma_u^2 = 0.1353$ ($\sigma_u = 0.3678$). Using the properties of the lognormal distribution, we have $E(u^2) = \sigma_u^2 e^{2\delta^2}$, and our values of δ and σ_u^2 are picked such that δ varies but $E(u^2) = 1$ for each of the DGPs.

For each replication, we estimate the model in three ways. (i) OLS, which estimates α and $\sigma_\varepsilon^2 \equiv \text{var}(\varepsilon_i) = \sigma_v^2 + \frac{\pi-2}{\pi}\sigma_u^2 e^{\delta^2}$; (ii) NLLS, which estimates α , δ and σ_u^2 ; and (iii) MLE, which estimates α , δ , σ_u^2 and σ_v^2 . We then calculate the mean, bias, variance and MSE of each of the estimates. We count the number of times that each of the various residuals has the wrong skew. Finally, we count how often the NLLS estimator equals ψ^* and how often the MLE equals θ^* . We need to be explicit about how this is defined. For example, the event “MLE is equal to θ^* ” is the complement of the event “MLE is not equal to θ^* ,” and this latter event occurs if and only if the likelihood evaluated at the MLE is larger than the likelihood evaluated at θ^* . That is, we compare likelihood values, rather than comparing the MLE parameter estimates to θ^* , and similarly for NLLS. Inequality is easier to verify or disprove numerically than equality.

The results strongly support the supposition that the residuals can have a wrong skew, but that this does not mean that there is a wrong skew *problem*. The proportions of replications in which the OLS residuals have a wrong skew (positive third moment) are 0.304, 0.104 and 0.064 for DGP1, DGP2 and DGP3, respectively. (As an aside, the value of 0.304 is very close to the value of 0.301 reported by Simar and Wilson (2009), Table 1, for the same parameter values.) The proportions of replications with a wrong skew for the NLLS residuals are 0.019, 0.001 and 0.001, while for the MLE residuals they are 0.072, 0.003 and 0.002. So a wrong skew can occur. However, it was never the case, for any of the 1000 replications for any of the three DGPs, that the NLLS estimator was equal to ψ^* or that the MLE was equal to θ^* . This is obviously a striking difference between the RSCFG model and the standard SFA model, where a wrong skew of the OLS residuals implies that the MLE is degenerate at a value that implies no inefficiency.

Although parameter estimation is not the focus of this paper, some evidence on the performance of the various estimators was generated in the course of the simulations.

Table 1 DGP1: $\delta = 0, \sigma_u^2 = 1$

OLS	$\hat{\alpha}$		$\hat{\sigma}_\varepsilon^2$	
True value	0		1.3633	
Mean	-0.7982		1.3465	
Bias	-0.7982		-0.01685	
Variance	0.01238		0.04031	
MSE	0.6495		0.04060	
NLLS	$\hat{\alpha}$	$\hat{\sigma}_u^2$	$\hat{\delta}$	
True value	0	1	0	
Mean	123.33	35023	-0.05740	
Bias	123.33	35022	-0.05740	
Variance	6887	6.120×10^8	1.3011	
MSE	22099	1.838×10^9	1.3044	
MLE	$\hat{\alpha}$	$\hat{\sigma}_u^2$	$\hat{\delta}$	$\hat{\sigma}_v^2$
True value	0	1	0	1
Mean	-0.1322	0.9748	-0.05270	0.9694
Bias	-0.1322	-0.02516	-0.05270	-0.03056
Variance	0.1969	0.8849	1.1847	0.1134
MSE	0.2144	0.8856	1.1875	0.1143

In Tables 1, 2 and 3 (for DGP1, DGP2 and DGP3, respectively), we give the mean, bias, variance and MSE of the OLS, NLLS and MLE estimators.

Consider first the results in Table 1. In DGP1, where $\delta = 0$ and therefore $\exp(z_i'\delta)$ is constant, the OLS estimate of α is the sample mean, and it correctly estimates the population mean which equals $\alpha - \sqrt{\frac{2}{\pi}}\sigma_u = -0.7979$. NLLS fails to give any coherent results because α and σ_u are not separately identified by the first moment (conditional on z) of the data. There is no apparent problem with MLE, for which the composed error distributional assumption yields identification. So the RSCFG model gives sensible MLE results even in this case where the RSCFG model is not needed.

Similar comments apply to the results for the other two DGPs (Tables 2 and 3). The only surprise (to us) is how poorly the NLLS estimator performs. With $\delta \neq 0$, we have identification from the first conditional moment but still have a great deal of difficulty of sorting out α from σ_u . Again, MLE is much better, and so the NLLS estimator is not recommended for practical use in this model.

7 Concluding remarks

In the standard normal/half-normal SFA model, the “wrong skew” problem refers to a set of problems that occur in the case that the OLS residuals are positively skewed. In

Table 2 DGP2: $\delta = 0.5$, $\sigma_u^2 = 0.6065$

OLS	$\hat{\alpha}$		$\hat{\sigma}_\varepsilon^2$	
True value	0		1.2830	
Mean	-0.7036		1.4765	
Bias	-0.7036		0.1935	
Variance	0.01423		0.07324	
MSE	0.5092		0.1107	
NLLS	$\hat{\alpha}$	$\hat{\sigma}_u^2$	$\hat{\delta}$	
True value	0	0.6065	0.5	
Mean	41.842	12053	0.7876	
Bias	41.842	12053	0.2876	
Variance	5865	5.120×10^8	1.5544	
MSE	7616	6.573×10^8	1.6371	
MLE	$\hat{\alpha}$	$\hat{\sigma}_u^2$	$\hat{\delta}$	$\hat{\sigma}_v^2$
True value	0	0.6065	0.5	1
Mean	0.01884	0.8176	0.6456	0.9054
Bias	0.01884	0.2111	0.1456	-0.09458
Variance	0.1124	0.5601	0.4390	0.07429
MSE	0.1127	0.6047	0.4602	0.08323

this case, Waldman (1982) showed that two things happen. First, the COLS estimator does not exist. Second, there is always a stationary point of the likelihood at a parameter value that corresponds to no inefficiency, and in the wrong skew case, this is a local maximum of the likelihood.

In this paper, we investigated the extent to which these results generalize to related but different models. Specifically, we considered the RSCFG model in which the (pre-truncation) variance of the half-normal error depends on explanatory variables. There is no equivalent of the COLS estimator for this model, but we considered NLLS as well as MLE. We found that there is always a stationary point of the criterion function (for both NLLS and MLE) at a parameter value that corresponds to no inefficiency, thus generalizing one of Waldman's results. For NLLS, we show that this parameter value is generally neither a local minimum nor a local maximum of the sum of squares. Similarly, for MLE, this parameter value is in general neither a local minimum nor a local maximum of the likelihood. Some limited simulations indicate that these stationary points are not a global minimum (for NLLS) or a global maximum (for MLE). So some of Waldman's results generalize to the RSCFG model and some do not.

If nothing else, this paper shows the inherent limitations of proceeding on a case-by-case basis. There are some other related results in models similar to the normal/half-normal stochastic frontier model. For example, there is no analogue to Waldman's results in the normal/exponential stochastic frontier model. However, Rho and Schmidt (2015) show that essentially the same results as Waldman's hold in

Table 3 DGP3: $\delta = 1, \sigma_u^2 = 0.1353$

OLS	$\hat{\alpha}$		$\hat{\sigma}_\varepsilon^2$	
True value	0		1.1336	
Mean	-0.4830		1.7616	
Bias	-0.4830		0.6280	
Variance	0.01753		2.1539	
MSE	0.2508		2.5484	
NLLS	$\hat{\alpha}$	$\hat{\sigma}_u^2$	$\hat{\delta}$	
True value	0	0.1353	1	
Mean	10.733	3026	1.2721	
Bias	10.733	3026	0.2721	
Variance	1803	1.535×10^8	1.7982	
MSE	1918	1.627×10^8	1.8723	
MLE	$\hat{\alpha}$	$\hat{\sigma}_u^2$	$\hat{\delta}$	$\hat{\sigma}_v^2$
True value	0	0.1353	1	1
Mean	0.07512	0.3524	1.0335	0.9089
Bias	0.07512	0.2171	0.03351	-0.09104
Variance	0.07546	0.2472	0.3736	0.04460
MSE	0.08111	0.2944	0.3748	0.05289

the zero inefficiency stochastic frontier model of Kumbhakar et al. (2013). Horrace and Wright (forthcoming) show that results similar to Waldman's hold for a "delta sequence" family of distributions of inefficiency, for which the inefficiency distribution converges to a Dirac delta function located at the origin as the variance of inefficiency goes to zero. This is a considerable generalization because this family includes many (but not all) commonly assumed distributions. They indicate that their results extend to models with determinants of inefficiency, like the RSCFG model, but they do not give details. It does not appear that our Results 1 or 2 (which have to do with NLLS, not MLE) or Result 4 (the stationary point of the log-likelihood is neither a local min nor a local max) follow from their results. Some more general results would be very useful to understand why results like Waldman's hold in one specific model and not in other closely related models.

An alternative direction of research is to find alternatives to the normal/half-normal distributional assumption that allow both positive and negative skewness, so that the skew cannot be "wrong." Papers that do this include Carree (2002), Griffin and Steel (2008), Almanidis and Sickles (2011), Almanidis et al. (2014), Bonanno et al. (2017) and Hafner et al. (2018). A somewhat similar model is the Laplace model of Horrace and Parmeter (2018), for which the population skewness is negative but a wrong skew of the OLS or LAD residuals does not cause difficulties in inference. None of these models include environmental variables, but they could be modified to do so.

Our paper does not address the important practical question of what to do if in an empirical setting one encounters a stationary point that is the global minimum (for NLLS) or maximum (for MLE). Simar and Wilson (2009) provide guidance for the normal/half-normal model without determinants of inefficiency. Extending their results to other models like the RSCFG model is a reasonable task but one that is beyond the scope of the present paper.

Compliance with ethical standards

Conflict of interest Both authors declare that they have no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix 1

Hessian for the NLLS problem

$$\nabla_{\alpha\alpha}\text{SSE} = 2n,$$

$$\nabla_{\alpha\beta}\text{SSE} = 2n\bar{x}',$$

$$\nabla_{\alpha\sigma_u}\text{SSE} = -2k \sum_i \exp(z_i' \delta),$$

$$\nabla_{\alpha\delta}\text{SSE} = -2k\sigma_u \sum_i \exp(z_i' \delta) z_i',$$

$$\nabla_{\beta\beta}\text{SSE} = \sum_i x_i x_i',$$

$$\nabla_{\beta\sigma_u}\text{SSE} = -2k \sum_i \exp(z_i' \delta) x_i,$$

$$\nabla_{\beta\delta}\text{SSE} = -2k\sigma_u \sum_i \exp(z_i' \delta) x_i z_i',$$

$$\nabla_{\sigma_u\sigma_u}\text{SSE} = 2k^2 \sum_i \exp(2z_i' \delta),$$

$$\nabla_{\sigma_u\delta}\text{SSE} = 2k \sum_i \left(y_i - \alpha - x_i' \beta + 2k\sigma_u \exp(z_i' \delta) \right) \exp(z_i' \delta) z_i',$$

$$\nabla_{\delta\delta} \text{SSE} = -2k\sigma_u \sum_i \left(y_i - \alpha - x_i' \beta + 2k\sigma_u \exp(z_i' \delta) \right) \exp(z_i' \delta) z_i z_i'$$

Appendix 2

Expression for $\ln L(\theta^o)$

We define $\theta^o = \theta^* + \tau w = \begin{bmatrix} \hat{\alpha} + \tau \\ \hat{\beta} \\ \tau/k \\ \tau\delta \\ \hat{\sigma}_v^2 \end{bmatrix}$. We also define the following notation:

$$\varepsilon_i^o = y_i - (\hat{\alpha} + \tau) - x_i' \hat{\beta} = \hat{\varepsilon}_i - \tau,$$

$$(\sigma_i^o)^2 = \left(\frac{\tau}{k}\right)^2 \exp(2\tau z_i' \delta),$$

$$(\omega_i^o)^2 = (\sigma_i^o)^2 + \hat{\sigma}_v^2,$$

$$\lambda_i^o = \sigma_i^o / \hat{\sigma}_v,$$

$$c_i^o = -\varepsilon_i^o \lambda_i^o / \omega_i^o.$$

Then, the log-likelihood is given by

$$\ln L(\theta^o) = \text{constant} - \sum_i \ln \omega_i^o - 1/2 \sum_i \left[(\varepsilon_i^o)^2 / (\omega_i^o)^2 \right] + \sum_i \ln \Phi(c_i^o).$$

This would need to be compared to $\ln L(\theta^*)$ as given in Eq. (20) of the text. No easy comparison is possible due to the various functions (logarithm, exponential, normal cdf) present in the likelihood expression above. One could attempt Taylor series expansions of these functions around the value $\tau = 0$, which is similar in spirit to what was done in Waldman (1982) and what we do in Sect. 4 for the least squares problem, but that did not get us any useful result for the least squares problem, which is simpler, so we will not follow that path here.

References

- Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. *J Econom* 6:21–37
- Almanidis P, Sickles RC (2011) The skewness issue in stochastic frontier models: fact or fiction? In: van Keilegom I, Wilson PW (eds) *Exploring research frontiers in contemporary statistics and econometrics*. Springer, Berlin

- Almanidis P, Qian J, Sickles RC (2014) Stochastic frontier models with bounded inefficiency. In: Sickles RC, Horrace WC (eds) *Econometric methods and applications*. Springer, New York
- Alvarez A, Amsler C, Orea L, Schmidt P (2006) Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *J Prod Anal* 25:201–212
- Amsler C, Tsay W-J, Schmidt P (2015) A post-truncation parameterization of truncated normal technical inefficiency. *J Prod Anal* 44:209–220
- Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empir Econ* 20:325–332
- Bonanno G, DeGiovanni D, Domma F (2017) The ‘wrong skewness’ problem: a re-specification of stochastic frontiers. *J Prod Anal* 47:49–64
- Carree MA (2002) Technological inefficiency and the skewness of the error component in stochastic frontier analysis. *Econ Lett* 77:101–107
- Caudill SB, Ford JM (1993) Biases in frontier estimation due to heteroskedasticity. *Econ Lett* 41:17–20
- Caudill SB, Ford JM, Gropper DM (1995) Frontier estimation and firm-specific inefficiency measures in the presence of heteroskedasticity. *J Bus Econ Stat* 13:105–111
- Griffin JE, Steel MFJ (2008) Flexible mixture modelling of stochastic frontiers. *J Prod Anal* 29:33–50
- Hafner CM, Manner H, Simar L (2018) The ‘wrong skewness’ problem in stochastic frontier models: a new approach. *Econom Rev* 37:380–400
- Horrace WC, Parmeter CF (2018) A laplace stochastic frontier model. *Econom Rev* 37:260–280
- Horrace WC, Wright IA (forthcoming) Stationary points for parametric stochastic frontier models. *J Bus Econ Stat* 36
- Huang CJ, Liu J-T (1994) Estimation of a non-neutral stochastic frontier production function. *J Prod Anal* 5:171–180
- Kumbhakar SC, Ghosh S, McGuckin JT (1991) A generalized production frontier approach for estimating determinants of inefficiency in U.S. Dairy Farms. *J Bus Econ Stat* 9:279–286
- Kumbhakar SC, Parmeter CF, Tsionas EG (2013) A zero inefficiency stochastic frontier model. *J Econom* 172:66–76
- Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. *Int Econ Rev* 18:435–444
- Olson JA, Schmidt P, Waldman DM (1980) A Monte Carlo study of estimators of stochastic frontier production functions. *J Econom* 13:67–82
- Reifschneider D, Stevenson R (1991) Systematic departures from the frontier: a framework for the analysis of firm inefficiency. *Int Econ Rev* 32:715–723
- Rho S, Schmidt P (2015) Are all firms inefficient? *J Prod Anal* 43:327–349
- Simar L, Wilson PW (2009) Inferences from cross-sectional stochastic frontier models. *Econom Rev* 29:62–98
- Stevenson RE (1980) Likelihood functions for generalized stochastic frontier estimation. *J Econom* 13:57–66
- Waldman DM (1982) A stationary point for the stochastic frontier likelihood. *J Econom* 18:275–279
- Wang H-J (2002) Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *J Prod Anal* 18:241–253