



Indirect inference estimation of stochastic production frontier models with skew-normal noise

Hung-pin Lai^{1,2}  · Subal C. Kumbhakar^{3,4}

Received: 21 June 2022 / Accepted: 12 March 2023 / Published online: 7 April 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

In this paper we consider a stochastic frontier model in which both the noise and inefficiency components are asymmetric, viz., the noise term is skew normal and the inefficiency term is half-normal. This formulation avoids the criticism that skewness of the composite error term (sum of the noise and inefficiency) cannot be an indicator of inefficiency because skewness can also arise from the noise term. Our estimator of inefficiency does not depend on skewness of the one-sided error alone; it controls for skewness in the noise term as well. We further generalize the model by introducing determinants of skewness of the noise term as well as determinants of inefficiency. Additionally, we test hypotheses that the noise term is either symmetric (normal) or has a constant skewness parameter. Instead of using the standard ML method, we use the indirect inference (II) approach to estimate the parameters of the proposed model. Formulae for predicting (in)efficiency are also provided. Finally, we provide both simulation and empirical results using the II estimation approach to showcase workings of our model.

Keywords Indirect inference estimation · Skew-normal error · Stochastic frontier model

JEL Classification C13 · C15 · D24

✉ Hung-pin Lai
ecdhpl@ccu.edu.tw
Subal C. Kumbhakar
kkar@binghamton.edu

¹ Department of Economics, National Chung Cheng University, Chiayi, Taiwan

² Research Center of Humanities and Social Sciences, Academia Sinica, Taipei, Taiwan

³ Department of Economics, State University of New York at Binghamton, Binghamton, NY 13902, USA

⁴ Inland Norway University of Applied Sciences, Lillehammer, Norway

1 Introduction

In stochastic frontier (SF) models, the noise term is almost universally assumed to be symmetric and normally distributed. The inefficiency term is assumed to have several competing distributions – the most common being half-normal, exponential and truncated normal. This makes the composite error term (noise minus inefficiency) in a production function model have a negatively skewed distribution. Because of this, the negative skewness is often used to test the presence of inefficiency (Ceolli 1995; Schmidt and Lin 1984) in a production function. However, skewness can be significant and of the wrong sign because of other things. For example, empirically skewness of the residual can be wrong (positive) in a production function even if the true model (or the data generating process, DGP) has the right skewness (negative).¹ Knowing that skewness can be wrong in the empirical model even though there is nothing wrong in the model or the DGP, we consider a model in which the skewness can be either positive or negative and allow the possibility of a skewed noise term. That is, we allow the noise term in the SF model to be positively or negatively skewed (asymmetric). This can be justified both theoretically and empirically (Badunenko and Henderson 2021). Thus, the presence of a skewed composite error term may or may not represent inefficiency. In the extreme case, inefficiency can be zero but the composite error term, which contains only the noise, can still be skewed. Our goal is to separate the skewness of the noise term so that it does not contaminate the estimate of inefficiency. Alternatively, we guard against labeling skewness of the composed error as an indicator of inefficiency because skewness can also arise from the noise term.

The generalization of symmetric noise to asymmetric noise in econometric models has been considered in many empirical applications. In financial studies, it is commonly observed that return of a market portfolio exhibits kurtosis and possesses skewness as well. Thus, a skew distribution is quite often used in financial applications. For instance, Adock (2010) uses a multivariate skew-Student-t distribution to study asset pricing and portfolio selection. Kraus and Litzenberger (1976) develop an asset pricing model that incorporates the effect of skewness on valuation. There are also some applications in macroeconomics studies. For instance, Busch et al. (2020) use the panel data on individuals and households from the United States, Germany and Sweden to study the income risk over the business cycle. They found that skewness of income growth is much more negative in recessions. In the studies on production and productivity, Qi et al. (2015) consider the asymmetric effects of the whether shocks on production; and Badunenko and Henderson (2021) consider a SF model with skew-normal noise and exponentially distributed inefficiency.

Perhaps the maximum likelihood (ML) method is the most commonly used approach to estimate models with skew noise. For instance, see Adock (2010) and Badunenko and Henderson (2021). Different from the existing literatures, in this paper we suggest using the indirect inference (II) approach, which was first introduced in Gourieroux et al. (1993) and then applied by Lai (2022) to estimate the SF model with skew-normal (SN) noise and half-normal inefficiency. The II approach can be easily

¹ Models are developed to address the so-called “wrong” skewness problem (Simar and Wilson 2010; Hafner et al. 2018), among others).

generalized to different combinations of the distributions of the random noise and inefficiency since there is no need to derive the probability density function. Similar to the simulated likelihood estimation, the II estimation is also a simulation-based approach. However, its implementation requires only the least squares estimation. Specifically, we first estimate the model by the least squares approach, obtain the coefficients and evaluate the empirical characteristic functions of the residuals. We then generate the random sample given the model parameter and conduct the same estimation procedure, including the least squares estimation and evaluating the empirical characteristic functions of the residuals using the generated sample. In the final step, we match the least squares coefficients and the empirical characteristic functions of the residuals from the real and simulated data. The parameter that minimizes the distance between the estimated coefficients and the empirical characteristic functions based on the real data and simulated data is the solution of the II estimation.

We also extend the model by using variables that are determinants of the shape parameter of the SN distribution. Similarly, inefficiency is also explained by some variables viewed as determinants of inefficiency. We provide simulation results to illustrate finite sample properties of the II estimation method in the case of SN noise and a half-normal (HN) inefficiency term. For this, we consider several settings of the data generating process. For example, we allow the skew-normal noise to have a positive shape parameter, negative shape parameter or heterogeneous shape parameter, and we examine whether the II estimation can correctly identify the direction of the shape of SN noise. In order to compare the finite sample performance of the II and ML estimators, we consider a SF model with a normal and HN distributions pair for the noise and inefficiency. Finally, an empirical application is provided to illustrate the workings of our model.

The rest of the paper is structured as follows. The model and estimation method are introduced in Sect. 2. Simulation results are discussed in Sect. 3. The empirical model and results are discussed in Sect. 4. Section 5 concludes the paper with a summary and future extensions.

2 The model formulation and estimation

We consider a cross-sectional production function with SN noise and HN technical inefficiency, viz.,

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

where y_i is log output, x_i consists of log inputs (for a Cobb-Douglas production function),² $v_i \sim SN(\sigma_v, \lambda)$, u is HN and is written as $u \sim N^+(0, \sigma_u^2)$. If v_i is normally distributed, the model in (1) is the standard SF model with N-HN distributions for v_i and u_i . Following the SF literature, we assume v_i and u_i to be independent. The input variables are assumed to be independent of v_i and u_i , i.e., we assume that inputs are exogenous.³

² For a translog production it contains the linear, square and cross products of log inputs.

³ The model can be extended to relax this assumption.

The probability density function (pdf) of the SN noise v_i is

$$f_v(v_i, \sigma_v, \lambda) = \frac{2}{\sigma_v} \phi\left(\frac{v_i}{\sigma_v}\right) \Phi\left(\frac{\lambda}{\sigma_v} v_i\right),$$

where σ_v is the scale parameter and λ is the shape parameter; $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of a standard normal variable. Two special cases of the shape parameter are the following: (i) if $\lambda = 0$, then $v_i \sim N(0, \sigma_v^2)$, and (ii) if $\lambda \rightarrow \infty$, $v_i \sim N^+(0, \sigma_v^2)$.

For simplicity, we use θ to denote the vector of the parameters of the model in (1). Thus $\theta = (\beta^T, \sigma_v, \lambda, \sigma_u)^T$ contains both the parameters in the frontier function and the parameters in the distributions of v and u . The model in (1) is usually estimated by the ML method. Here, we suggest using the II estimation method, where a linear regression model is used as an approximation of the model in (1).⁴ The estimation is based on the ordinary least squares (OLS) approach. This approximated model is referred to as the instrument model in our following discussion. It is worth mentioning that although we are making distributional assumptions of the error components, there is no need to evaluate the likelihood function for the II estimation.

For the OLS estimation, we write the stochastic frontier model given in (1) as

$$y_i = a + x_i\beta + (v_i - u_i^*), \tag{2}$$

where $a = \alpha - \mathbb{E}u_i$, $u_i^* = u_i - \mathbb{E}u_i$, and the composite error $(v_i - u_i^*)$ has zero mean. Let \widehat{a}_N and $\widehat{\beta}_N$ denote the OLS estimators of the intercept a and the slope vector β . It can be shown that $\widehat{\beta}_N$ is a consistent estimator of β , but \widehat{a}_N is not a consistent estimator of a because $\mathbb{E}u_i \neq 0$.

Define $\xi_i = y_i - x_i\beta = a + v_i - u_i^* = \alpha + v_i - u_i$. Although the OLS estimator of a from (2) is not a consistent estimator of a , it is clear that the OLS estimators \widehat{a}_N and $\widehat{\beta}_N$ and the characteristic functions of $\widehat{\xi}_i$ are functions of the true parameter θ . In other words, the information contained in the moments of the residual $\widehat{\xi}_i$ can be used for identifying the parameters of the distributions of v_i and u_i .

Now consider the characteristic function of ξ , $\varphi_\xi(\cdot)$, which is defined as

$$\varphi_\xi(t) = \mathbb{E}\left(e^{it\xi}\right) = \mathbb{E}\left(e^{it(\alpha+v-u)}\right) = e^{it\alpha} \cdot \varphi_v(t) \cdot \varphi_u(t), \text{ where } t \in \mathbb{R}. \tag{3}$$

Using the OLS estimator $\widehat{\beta}_N$ and the observed sample, we obtain $\widehat{\xi}_i$ as

$$\widehat{\xi}_i = y_i - x_i\widehat{\beta}_N.$$

Given the set of points $t = t_1, \dots, t_q$, we can also compute the empirical characteristic function at t via

$$\widehat{\varphi}_{N,\xi}(t) = \frac{1}{N} \sum_{i=1}^N e^{it\widehat{\xi}_i}. \tag{4}$$

⁴ Note that the model in (1) is log-linear but one can specify a nonlinear function that can be approximated by a linear function.

We use $\widehat{\gamma}_N$ to denote a vector containing the OLS estimators of a and β and the empirical characteristic function evaluated at $t = t_1, \dots, t_q$, so

$$\widehat{\gamma}_N = \begin{pmatrix} \widehat{a}_N \\ \widehat{\beta}_N \\ \widehat{\varphi}_{N,\xi}(t_1) \\ \vdots \\ \widehat{\varphi}_{N,\xi}(t_q) \end{pmatrix}. \tag{5}$$

Alternatively, the estimator $\widehat{\gamma}_N$ is the solution of the moment conditions of the approximated model:

$$\mathbb{E}(e_i) = \mathbb{E}(\xi_i - a) = 0 \tag{6}$$

$$\mathbb{E}(e_i \cdot x_i) = \mathbb{E}[(\xi_i - a) \cdot x_i] = 0 \tag{7}$$

$$\mathbb{E}\left(e^{it\xi_i}\right) = \varphi_\xi(t), \text{ for } t = t_1, \dots, t_q, \tag{8}$$

where $e_i = (\xi_i - a)$. Equations (6) and (7) come from the first order conditions of the OLS estimation, and Eq. (8) is the sample counterpart of the characteristic function $\mathbb{E}(e^{it\xi_i})$. Under the distributional assumptions, the true value of γ is

$$\gamma = \begin{pmatrix} \alpha - \sigma_u \sqrt{\frac{2}{\pi}} \\ \beta \\ \mathbb{E}\left(e^{it_1 \xi_i}\right) \\ \vdots \\ \mathbb{E}\left(e^{it_q \xi_i}\right) \end{pmatrix}. \tag{9}$$

For the II estimation, it is required $q/2$ to be greater than the number of the parameters contained in the distributions of v and u so that we have enough identification conditions. Since the empirical characteristic function contains the real and imaginary parts, each point t gives two moment conditions. For the models in our simulation and empirical sections, we choose ten points, i.e., $q = 10$, with equal space in the interval $[-0.1, 0.1]$. See also Lai (2022) for more discussion about the choice of the points and the relevant literature.

The vector γ is a function of the parameter θ , and we denote it as

$$\gamma = \mathbb{B}(\theta), \tag{10}$$

where $\mathbb{B}(\theta)$ is called the binding function. It provides a link between the parameters of the true model and the approximated model. Since the characteristic function is unique and $t \in \mathbb{R}$, we can find infinite moment conditions for solving the true parameter θ . We discuss the procedure below:

Step 1: Given the approximate or instrumental model, we estimate γ by OLS using the observed data and denote it as

$$\widehat{\gamma}_N = (\widehat{a}_N, \widehat{\beta}_N^\top, \widehat{\varphi}_{N,\xi}(t_1), \dots, \widehat{\varphi}_{N,\xi}(t_q))^\top, \tag{11}$$

where $(\widehat{a}_N, \widehat{\beta}_N^\top)^\top$ is the OLS estimate from (2) using the observed data and $\widehat{\varphi}_{N,\xi}(t)$ is the empirical characteristic function of (4) evaluated at $t = t_1, t_2, \dots, t_q$.

Step 2: For a given θ , we can simulate values of the endogenous variable $\{y_i^m(\theta)\}_{i=1}^N$ using the true model given in (1). In other words, given the model specification and a value of the parameter θ , one can draw v and u from their distributions and then generate the simulated $\{y_i^m(\theta)\}_{i=1}^N$ conditional on the exogenous variables (x_1, \dots, x_N) . The estimator of γ in the m th replication is denoted as

$$\widehat{\gamma}_N^m(\theta) = (\widehat{a}_N^m(\theta), \widehat{\beta}_N^{m\top}(\theta), \widehat{\varphi}_{N,\xi}^m(t_1), \dots, \widehat{\varphi}_{N,\xi}^m(t_q))^\top, \tag{12}$$

where $(\widehat{a}_N^m(\theta), \widehat{\beta}_N^{m\top}(\theta))^\top$ is the solution of the problem

$$Q_N^m(b) = \min_b \frac{1}{N} \sum_{i=1}^N (y_i^m - a - x_i \beta)^2, \tag{13}$$

which is estimated based on the m th simulated data. Using the predicted OLS residual $\widehat{\xi}^m$, we can obtain the empirical characteristic function $\widehat{\varphi}_{\xi}^m(t)$ defined in (4) and (12) at the points $t = t_1, t_2, \dots, t_q$. With the simulated data $\{(y_i^m, x_i)\}_{i=1}^N$, one can then replicate M times such simulations and estimate the parameter γ .

Step 3: In this step, we then match the estimate of γ from the approximate model based on the real data and that from the simulated data. The II estimator of θ is defined as the one that has the minimum distance between the estimates based on the real data and simulated data, i.e.,

$$\widehat{\theta}_N^M(\Omega) = \arg \min_{\theta \in \Theta} \left(\widehat{\gamma}_N - \frac{1}{M} \sum_{m=1}^M \widehat{\gamma}_N^m(\theta) \right)^\top \Omega \left(\widehat{\gamma}_N - \frac{1}{M} \sum_{m=1}^M \widehat{\gamma}_N^m(\theta) \right), \tag{14}$$

where Ω is a symmetric non-negative matrix.

For simplicity, we choose the weighting matrix Ω as a diagonal matrix, which assigns 1 for each element of the OLS estimators $(\widehat{a}_N$ and $\widehat{\beta}_N)$ and $1/q$ for each element of the empirical characteristic functions $\widehat{\varphi}_{N,\xi}(t_1), \dots, \widehat{\varphi}_{N,\xi}(t_q)$.

It is worth mentioning that the characteristic function contained in the binding function is evaluated by the empirical characteristic function, so it does not matter what the form of the characteristic function is under the true distributions of v and u . In other words, we estimate the parameters contained in the distribution by matching the empirical characteristic functions from the observed data and the simulated data. In order to identify the parameter θ , it is required that the binding function $\mathbb{B}(\cdot)$ be a

function of the true parameter θ and the dimension of the auxiliary parameter γ must be larger than or equal to the dimension of θ .

Lai (2022) discussed some practical issues in implementing the II estimation and suggested selecting the points close together and equally spaced around zero. The main idea of the II approach is to use simulations performed under the ‘true’ model to correct for the asymptotic bias of the approximate model $\widehat{\gamma}_N$. In the above estimation procedure, steps 1 and 2 require the OLS estimation and step 3 requires nonlinear least squares estimation. Since the characteristic function is the Fourier transform of the probability density function, it completely defines the probability distribution of this random variable. The characteristic function provides an alternative route to estimate the distribution of the random variable. Therefore, $\widehat{\gamma}_N$ is the binding function used in the estimation of the SF model. After obtaining the II estimator $\widehat{\theta}_N^M(\Omega)$ from (14), we can compute the standard error using the approach discussed in Lai (2022). Alternatively, one can also use the bootstrap method to obtain the bootstrap standard error, which is easier to implement from the practical point of view.

Prediction of (in)efficiency: Once we obtain the estimate of θ , the prediction of inefficiency can be done using the simulated approach. Let $g(u)$ be an integrable function of u . For the SF analysis, we are particularly interested in $g(u) = u$ and $g(u) = \exp(-u)$. When $g(u) = u$, we have $\mathbb{E}[g(u)|\varepsilon] = \mathbb{E}[u|\varepsilon]$. If $g(u) = e^{-u}$, then $\mathbb{E}[g(u)|\varepsilon] = \mathbb{E}[e^{-u}|\varepsilon]$. By the definition of conditional expectation, we have

$$\mathbb{E}[g(u)|\varepsilon] = \int_0^\infty g(u) \cdot f_{u|\varepsilon}(u|\varepsilon) du \tag{15a}$$

$$= \int_0^\infty g(u) \cdot \frac{f_{u,\varepsilon}(u, \varepsilon)}{f_\varepsilon(\varepsilon)} du \tag{15b}$$

$$= \int_0^\infty g(u) \cdot \frac{f_{\varepsilon|u}(\varepsilon|u) f_u(u)}{\int_0^\infty f_{\varepsilon|u}(\varepsilon|u) \cdot f_u(u) du} du. \tag{15c}$$

It follows from (15c) that $\mathbb{E}[g(u)|\varepsilon]$ can be estimated using the simulated estimator

$$\mathbb{E}^s[g(u)|\varepsilon] = \frac{1}{R} \sum_{r=1}^R g(u^r) \left(\frac{f_v(\varepsilon + u^r)}{\sum_{r=1}^R f_v(\varepsilon + u^r)} \right). \tag{16}$$

Let $W_i^r = \frac{f_v(\varepsilon+u^r)}{\sum_{r=1}^R f_v(\varepsilon+u^r)}$ denote the weight based on the simulated probability. Then the simulated estimator $\mathbb{E}^s[g(u)|\varepsilon]$ takes the form of a weighted average of $g(u^r)$.

Under the distribution $v \sim SN(\sigma_v, \alpha)$, the pdf of v is

$$f_v(v; \sigma_v, \alpha) = \frac{2}{\sigma_v} \phi\left(\frac{v}{\sigma_v}\right) \Phi\left(\frac{\alpha}{\sigma_v} v\right). \tag{17}$$

Plugging (17) into (16), we obtain the simulated estimator of the inefficiencies and TEs. The simulated estimator for predicting TE and inefficiency has been applied in various stochastic frontier models, for example, the panel stochastic frontier models

in Lai and Kumbhakar (2018) and Lai and Kumbhakar (2021) and also the spatial autoregressive panel stochastic frontier model in Lai and Tran (2022).

Nonconstant σ_u : A simple way to model determinants of inefficiency is to make σ_u nonconstant – for example, $\sigma_{u,i} = \exp(\delta'w_i)$, where w is a vector of exogenous variables that can be interpreted as determinants of inefficiency. When u_i is HN, $\mathbb{E}u_i = \sqrt{2/\pi}\sigma_{u,i} = \sqrt{2/\pi} \exp(\delta'w_i)$. Consequently, the w variables can be viewed as determinants of inefficiency. With nonconstant σ_u the SF model given in (1) can be written as

$$y_i = [\alpha - h(w_i; \delta)] + x_i'\beta + v_i - u_i^*, \quad i = 1, \dots, N, \tag{18}$$

where $u_i^* = u_i - h(w_i; \delta)$ and $h(w_i; \delta) = \sqrt{2/\pi} \exp(\delta'w_i)$, which is nonlinear. Define $h_0 = \mathbb{E}[h(w_i; \delta)]$ and $a = \alpha - h_0$, then $[\alpha - h(w_i; \delta)]$ can be decomposed as

$$\alpha - h(w_i; \delta) = (\alpha - h_0) + [h_0 - h(w_i; \delta)],$$

where $\mathbb{E}[h_0 - h(w_i; \delta)] = 0$ by construction. Note that h_0 is the expectation of the function $h(w_i; \delta)$ taken over w , thus it is a constant. Let $a = \alpha - h_0$ be the intercept of the instrument model, then the above model in (18) can be represented as

$$y_i = a + x_i'\beta + [h_0 - h(w_i; \delta)] + v_i - u_i^* \tag{19}$$

$$= x_i'\beta + \xi_i, \tag{20}$$

where $\xi_i = y_i - x_i\beta = a + [h_0 - h(w_i; \delta)] + v_i - u_i^*$. Note that (20) has a similar form as Eq. (2). Thus, the corresponding binding function is

$$\gamma = \begin{pmatrix} \alpha - \sqrt{2/\pi} \mathbb{E}[\exp(\delta'w_i)] \\ \beta \\ \mathbb{E}(e^{it_1\xi_i}) \\ \vdots \\ \mathbb{E}(e^{it_q\xi_i}) \end{pmatrix}. \tag{21}$$

It follows that the II estimation of (18) can be done in the same way. One can apply OLS to (20) and obtain the estimators $\widehat{a}_N, \widehat{\beta}_N$ and the predicted $\widehat{\xi}_i$. The empirical characteristic function of ξ at points t_1, t_2, \dots, t_q is computed using (4). It is clear that each element of γ is a function of the true parameter θ . The estimation procedure is similar to the previous case of constant σ_u . By matching $\widehat{\gamma}_N$ from the observed data and $\widehat{\gamma}_N^m(\theta)$ from the simulated data, we obtain an estimate of θ , which minimizes the distance between the two estimates.

3 Simulation

In this section, we conduct some Monte Carlo experiments to investigate the finite sample performance of the II estimator. Furthermore, we are also interested in knowing whether the proposed estimator can correctly identify the direction of the skewness

from the data. In this regard, we consider opposite signs of the shape parameter that determines the skewness in the data generating process (DGP) in our experiments.

We consider the following specification of our DGP:

$$y_i = \beta_0 + \beta_1 x_i + v_i - u_i, \tag{22}$$

where

$$v \sim \text{SN}(\sigma_v, \lambda) \text{ and } u \sim N^+(0, \sigma_{u,i}^2). \tag{23}$$

In order to allow exogenous determinants in the inefficiency u , we specify $\sigma_{u,i}$ as $\sigma_{u,i} = \exp(\delta_0 + \delta_1 w_i)$. The true parameters are set as

$$\beta_0 = 0.5, \beta_1 = 1, \sigma_v = 0.15,$$

and

$$\delta_0 = -1 \text{ and } \delta_1 = -0.25.$$

The exogenous variables are randomly drawn as $x \stackrel{i.i.d.}{\sim} U(0, 3)$ and $w \stackrel{i.i.d.}{\sim} N(0, 1)$. For the skew-normal random variable v , we consider three different specifications. In the first two specifications, v has the constant shape parameter λ , where $\lambda = 3$ in Case 1 and $\lambda = -3$ in Case 2. Empirically, we do not know the direction of the skewness of v , which crucially depends on the sign of λ . Thus, we consider these two opposite cases in our simulation to explore how our approach performs and whether it can correctly identify the skewness of v when the shape parameter has opposite signs. In Case 3, we allow the shape parameter λ to be heteroscedastic and parameterize it as

$$\lambda_i = \lambda_0 + \lambda_1 z_i, \tag{24}$$

where

$$\lambda_0 = -0.1 \text{ and } \lambda_1 = -3.$$

We assume that the exogenous determinant variable z in the shape parameter follows a normal distribution $z_i \stackrel{i.i.d.}{\sim} N(0, 1)$. Under this setting, the scale parameter λ_i can be either positive or non-positive, and the sample means of λ_i in the generated random sample is about -3 .

Finally, we consider Case 4, where $\lambda_1 = 0$ and $\lambda_0 = 0$, i.e., $\lambda_i = 0$ for all i . The rest of the setting of the DGP in Case 4 is the same as in the previous case. If $\lambda_i = 0$, then v_i follows a normal distribution $N(0, \sigma_v^2)$. The normal-half normal specification has been used widely in the empirical application of the stochastic frontier model. Thus, we are also interested in comparing the performances of the II estimator and the ML estimator.

In the II estimation, the total number of replications in all simulations is 1000. For each estimation, we consider the number of replications M to be $M = \{10, 100\}$ and investigate the efficiency gain as we increase the number of replications.

Table 1 Biases and RMSEs of Cases 1 and 2

Parameter \ N	Bias			RMSE		
	250	500	1000	250	500	1000
Case 1A: Positive shape parameter ($\alpha = 3$ and $M = 10$)						
β_1	-0.0046	-0.0022	-0.0011	0.0180	0.0131	0.0091
β_0	0.0073	0.0042	0.0027	0.0309	0.0214	0.0152
σ_v	0.0017	0.0019	0.0020	0.0056	0.0053	0.0046
λ	0.0011	0.0041	0.0014	0.0464	0.0555	0.0526
δ_1	-0.0134	-0.0178	-0.0183	0.0163	0.0189	0.0154
δ_0	0.0124	0.0109	0.0132	0.0176	0.0147	0.0147
Case 1B: Positive shape parameter ($\alpha = 3$ and $M = 100$)						
β_1	-0.0001	-0.0008	-0.0003	0.0180	0.0131	0.0091
β_0	0.0031	0.0049	0.0043	0.0304	0.0214	0.0155
σ_v	0.0016	0.0012	0.0014	0.0050	0.0046	0.0043
λ	-0.0001	0.0024	0.0015	0.0437	0.0515	0.0450
δ_1	-0.0097	-0.0137	-0.0148	0.0150	0.0181	0.0138
δ_0	0.0100	0.0099	0.0118	0.0170	0.0137	0.0157
Case 2A: Negative shape parameter ($\alpha = -3$ and $M = 10$)						
β_1	-0.0023	0.0000	0.0004	0.0184	0.0131	0.0093
β_0	0.0057	0.0030	0.0024	0.0313	0.0222	0.0158
σ_v	0.0006	0.0004	0.0004	0.0018	0.0024	0.0020
λ	0.0046	-0.0144	0.0027	0.1077	0.1862	0.1304
δ_1	-0.0125	-0.0167	-0.0196	0.0102	0.0136	0.0140
δ_0	0.0123	0.0113	0.0129	0.0105	0.0096	0.0104
Case 2B: Negative shape parameter ($\alpha = -3$ and $M = 100$)						
β_1	0.0006	0.0002	0.0004	0.0183	0.0128	0.0096
β_0	0.0049	0.0041	0.0043	0.0300	0.0217	0.0163
σ_v	0.0006	0.0004	0.0004	0.0019	0.0015	0.0017
λ	0.0019	-0.0018	-0.0056	0.0726	0.1047	0.0909
δ_1	-0.0123	-0.0110	-0.0148	0.0165	0.0113	0.0106
δ_0	0.0090	0.0087	0.0095	0.0113	0.0094	0.0085

The total number of replications is 1000

Table 1 summarizes the simulation results of the four cases, labeled as Case 1A, Case 1B, Case 2A and Case 2B. In Case 1, we have a positive value of λ , so that v has a positive skewness. We have a negative λ in Case 2, thus v has a negative skewness. Moreover, we labeled the cases with the letter ‘A’ when the number of replications $M = 10$ and ‘B’ when $M = 100$. Values in columns 2 to 4 give the biases when the sample sizes are 250, 500 and 1000. Columns 5 to 7 summarize the root mean squared error (RMSE). As shown in the table, all biases and RMSEs are in small magnitudes. Most of the biases slightly decrease as we increase the sample size from 250 to 1000. So do the RMSEs. On the other hand, by comparing Cases 1A (2A) and 1B (2B) we find that the RMSEs slightly decrease for most of the parameters as we increase M

from 10 to 100. For the negative skewness cases (2A and 2B), we find that the biases of λ are slightly larger than those of the cases with a positive skewness (1A and 1B). It could be due to the direction of the skewness. For example, consider three random noises v_1 , v_2 and v_3 , where v_1 has a positive skewness, v_2 has zero skewness and v_3 has negative skewness. It is clear that the skewness of the composite errors satisfies the relation $SK(v_1 - u) > SK(v_2 - u) > SK(v_3 - u)$ and $0 > SK(v_2 - u) > SK(v_3 - u)$. We conjecture that the bias of λ is slightly larger when v and u have the same direction of skewness. However, the bias decreases quickly as we increase the sample size.

In Case 3, we let the shape parameter be heteroscedastic and have it depend on an exogenous variable z . The simulation results are summarized in the upper panel of Table 2. As shown in the table, the pattern observed from the biases and RMSEs is quite similar to what we found in Table 1. In Case 4, we compare the biases and RMSEs of the ML estimator and the indirect estimators when $M = 10$ and $M = 100$. The results are summarized in the lower panel of Table 2. As shown in the table, the bias of the II estimator when $M = 10$ is slightly larger than that of the ML estimator. However, as we increase the number of replications to 100, the bias of the II estimator significantly decreases and performs quite similarly to the ML estimator. Furthermore, we find the RMSEs of the ML and II estimators when $M = 10$ and $M = 100$ are quite similar. There is no significant evidence showing that the ML estimator is better in terms of RMSE. The simulation results in Tables 1 and 2 show the consistency of our proposed estimator. Based on these results, we conclude that our estimator performs quite well in the small sample.

Below, we use an artificial random sample of $N = 500$ to compare the predicted inefficiencies and TEs, which are computed using Eq. (16) under 1000 draws. The sample is generated according to the DGP specified in Eq. (22). We specify $\lambda_0 = -3.5$ and $\lambda_1 = -1$, so that the average ratio of $\sqrt{\frac{\text{Var}(v)}{\text{Var}(u)}}$ is around $\frac{0.2275}{0.2586} = 0.88$. The remaining parameters are specified as before. The true distribution of the composite error is heteroscedastic SN and half-normal distribution. This random sample is estimated under three different specifications of the composite error. The three models are labelled as follows:

M1: heteroscedastic SN noise and half-normal inefficiency

M2: normal noise and half-normal inefficiency

M3: homoscedastic SN noise and half-normal inefficiency.

Thus, M1 is the true model. We compare the true values of the inefficiencies and TE (u and e^{-u}) and their expectations ($\mathbb{E}(u|\varepsilon)$ and $E(e^{-u}|\varepsilon)$) with their predictions under models M1-M3. The graphs of the true TEs, their expected values, and their predicted values are given in Fig. 1. As shown in the figure, the predicted TEs from M2 seem to be overestimated at the lower tail and underestimated at the right tail of the distribution of TE. This is because M2 imposes a symmetric distribution on the random noise v , and thus the prediction of TE ignores the asymmetric distribution of v . On the other hand, the predicted TEs from M1 and M3 have more similarities because both of them assume SN distributions on v . M1 has a heteroscedastic shape parameter and M3 assumes a homoscedastic shape parameter. The conditional expectation functions are evaluated using the true parameters. In order to compare the predicted inefficiencies and TEs

Table 2 Biases and RMSEs of Cases 3 and 4

Parameter \ N	Bias			RMSE		
	250	500	1000	250	500	1000
Case 3A: ($M = 10$)						
β_1	-0.0024	0.0000	0.0004	0.0184	0.0131	0.0092
β_0	0.0061	0.0034	0.0025	0.0310	0.0223	0.0159
σ_v	0.0006	0.0005	0.0004	0.0021	0.0018	0.0021
λ_1	0.0012	0.0004	0.0024	0.1109	0.0513	0.0877
λ_0	0.0041	0.0026	-0.0006	0.1727	0.1123	0.1276
δ_1	-0.0134	-0.0176	-0.0202	0.0118	0.0140	0.0145
δ_0	0.0130	0.0121	0.0131	0.0114	0.0100	0.0110
Case 3B: ($M = 100$)						
β_1	-0.0000	-0.0006	-0.0001	0.0184	0.0131	0.0093
β_0	0.0044	0.0062	0.0055	0.0307	0.0218	0.0155
σ_v	0.0004	0.0004	0.0004	0.0016	0.0015	0.0018
λ_1	-0.0015	-0.0017	-0.0016	0.0695	0.0589	0.0644
λ_0	0.0024	-0.0011	0.0035	0.0746	0.0882	0.0785
δ_1	-0.0092	-0.0137	-0.0169	0.0115	0.0137	0.0122
δ_0	0.0091	0.0097	0.0104	0.0111	0.0096	0.0090
Case 4A: ($M = 10$)						
β_1	-0.0042	-0.0020	-0.0016	0.0196	0.0143	0.0100
β_0	0.0008	-0.0014	-0.0029	0.0334	0.0242	0.0169
σ_v	0.0138	0.0101	0.0123	0.0231	0.0205	0.0198
δ_1	-0.0058	-0.0073	-0.0080	0.0083	0.0097	0.0097
δ_0	0.0092	0.0087	0.0084	0.0094	0.0079	0.0061
Case 4B: ($M = 100$)						
β_1	-0.0001	-0.0006	-0.0006	0.0195	0.0143	0.0100
β_0	0.0024	0.0037	0.0031	0.0330	0.0241	0.0168
σ_v	0.0068	0.0087	0.0084	0.0256	0.0227	0.0295
δ_1	-0.0042	-0.0049	-0.0067	0.0073	0.0107	0.0123
δ_0	0.0069	0.0068	0.0069	0.0084	0.0078	0.0074
Case 4C: Maximum likelihood estimation						
β_1	0.0003	-0.0005	-0.0006	0.0183	0.0124	0.0088
β_0	-0.0028	-0.0004	0.0003	0.0480	0.0310	0.0204
σ_v	-0.0035	-0.0012	-0.0009	0.0271	0.0170	0.0119
δ_1	-0.0087	-0.0015	-0.0034	0.0717	0.0453	0.0307
δ_0	-0.0233	-0.0111	-0.0042	0.1490	0.0862	0.0558

The total number of replications is 1000

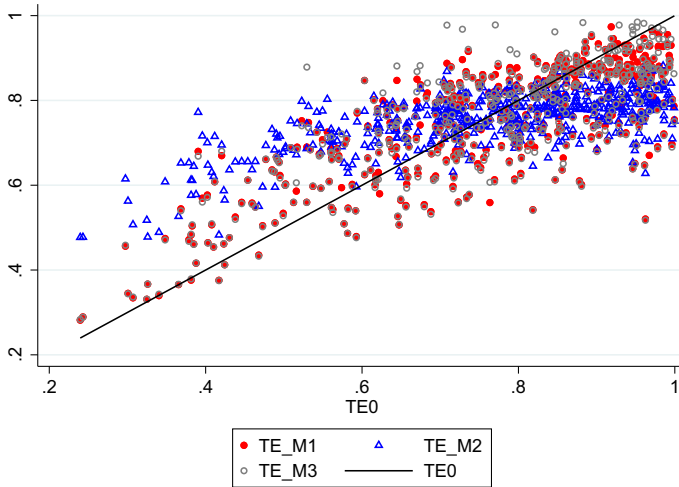


Fig. 1 The predicted TEs of Models M1-M3

under models M1-M3, we define the absolute values of the biases and prediction errors of inefficiencies and TEs as

$$\begin{aligned}
 \text{absolute bias of } \mathbb{E}(u|Mj) &= \left| \widehat{\mathbb{E}}(u|\varepsilon; Mj) - \mathbb{E}(u|\varepsilon) \right| \\
 \text{absolute bias of } \mathbb{E}(e^{-u}|Mj) &= \left| \widehat{\mathbb{E}}(e^{-u}|\varepsilon; Mj) - \mathbb{E}(e^{-u}|\varepsilon) \right| \\
 \text{absolute prediction error of } \mathbb{E}(u|Mj) &= \left| \widehat{\mathbb{E}}(u|\varepsilon; Mj) - u \right|
 \end{aligned}$$

and

$$\text{absolute prediction error of } \mathbb{E}(e^{-u}|Mj) = \left| \widehat{\mathbb{E}}(e^{-u}|\varepsilon; Mj) - e^{-u} \right|,$$

where $j = 1, 2, 3$. We summarize the descriptive statistics of the true and predicted values of u , TE, and the absolute biases and absolute prediction errors in Table 3. The means of the true values of u and $\mathbb{E}(u|\varepsilon)$ are about the same (0.30), but their standard deviations and ranges are different (ranges of u and $\mathbb{E}(u|\varepsilon)$ are 1.43 and 1.21, respectively). The means of the predicted values of $\mathbb{E}(u|\varepsilon)$ for the three models are almost the same as the true mean of the values of $\mathbb{E}(u|\varepsilon)$. Similarly, the means of the predicted values of $\mathbb{E}(e^{-u}|\varepsilon)$ from M1-M3 are identical and the same as the mean of the true values of $\mathbb{E}(u|\varepsilon)$ (all about 0.75). However, the standard deviation of predicted $\mathbb{E}(u|\varepsilon)$ from M2 (0.1055) is much smaller than those from M1 and M3 (which are around 0.20 ~ 0.21). It seems that the predicted $\mathbb{E}(u|\varepsilon)$ from M2 has the problem of under-dispersion. A similar finding also appears in the predictions of $\mathbb{E}(e^{-u}|\varepsilon)$. Furthermore, M2 has absolute biases that are quite different from the other two models. Similarly, the means of absolute prediction errors of $\mathbb{E}(u|\varepsilon)$ and $\mathbb{E}(e^{-u}|\varepsilon)$ from M1 and M3 are very similar and are different from M2. These results show that although there are some similarities, model M2 (which ignores the skewness in the

Table 3 The predicted inefficiency and TE

Variable	Mean	Std. dev.	Min	Max
True value of u and $\mathbb{E}(u \varepsilon)$:				
u	0.2984	0.2586	0.0006	1.4287
$\mathbb{E}(u \varepsilon)$	0.3030	0.2008	0.0234	1.2332
Predicted $\mathbb{E}(u \varepsilon; M_j)$ for $j = 1, 2, 3$				
M1	0.3144	0.2068	0.0270	1.2962
M2	0.3091	0.1055	0.1309	0.8117
M3	0.3100	0.2105	0.0154	1.2877
Absolute bias $ \mathbb{E}(u \varepsilon; M_j) - \mathbb{E}(u \varepsilon) $:				
M1	0.0134	0.0122	3.80e-06	0.0779
M2	0.0901	0.0702	0.0004	0.4374
M3	0.0237	0.0272	0.0001	0.2262
Absolute prediction error $ \mathbb{E}(u \varepsilon; M_j) - u $:				
M1	0.1209	0.0981	0.0002	0.6377
M2	0.1551	0.1198	0.0005	0.6847
M3	0.1223	0.1007	0.0004	0.6425
True values of e^{-u} and $\mathbb{E}(e^{-u} \varepsilon)$:				
e^{-u}	0.7644	0.1680	0.2396	0.9994
$\mathbb{E}(e^{-u} \varepsilon)$	0.7594	0.1284	0.2994	0.9771
Predicted $\mathbb{E}(e^{-u} \varepsilon; M_j)$ for $j = 1, 2, 3$				
M1	0.7519	0.1301	0.2818	0.9736
M2	0.7537	0.0662	0.4770	0.8816
M3	0.7553	0.1341	0.2835	0.9849
Absolute bias $ \mathbb{E}(e^{-u} \varepsilon; M_j) - \mathbb{E}(e^{-u} \varepsilon) $:				
M1	0.0090	0.0071	0.0000	0.0288
M2	0.0624	0.0422	0.0003	0.1923
M3	0.0169	0.0202	0.0001	0.1683
Absolute prediction error $ \mathbb{E}(e^{-u} \varepsilon; M_j) - e^{-u} $:				
M1	0.0865	0.0670	0.0003	0.4421
M2	0.1095	0.0738	0.0007	0.3815
M3	0.0875	0.0694	0.0006	0.4450

The results are based on a random sample of size 500

noise term) is different from models M1 and M3. We find similar results from models M1-M3 in the application with real data.

4 Empirical application

In this section, we empirically demonstrate the estimation approach using real data (an unbalanced panel of 59 US electricity transmission companies for the period 2001–2009). The main objective is to showcase workings of our model instead of a thorough

empirical analysis of efficiency of the electricity transmission sector as in Llorca et al. (2016). Instead of searching for the best model in terms of their specifications of inefficiency, our objective is to illustrate the working of the II estimation with SN noise and half-normal inefficiency. Having said that, we want to point out that our approach can be extended to accommodate the more general case where the mean and variance of inefficiency can be functions of exogenous variables that can potentially explain inefficiency. Thus, our model clearly extends the existing SF models by allowing an asymmetric noise term where the asymmetry (the shape parameter of the SN variable) is explained by some exogenous variables. In our application, we use weather variables as determinants of inefficiency via the variance of the half-normal inefficiency term. The shape parameter in the SN distribution of the noise term is allowed to depend on peak load. In fact, one can use other variable(s) as determinants of the shape parameters also.

Since electricity distribution/transmission are services in which outputs are considered exogenous, a cost or input distance is commonly used. Here, we use a cost function approach in which the cost variable (C) is total cost (Totex). Totex is defined as the sum of Opex, which is the total of the operation and maintenance expenses, and Capex, which is the sum of annual depreciation on capital assets and the annual return on the balance of capital. Opex, Capex and Totex are measured in year 2000 US dollars. The output variables are Y_1 – electricity delivered (del), Y_2 – total capacity of substations (cap_subs), and Y_3 – network length (pole) measured in miles. The input variables are labor and capital. Price of labor (LPR) is defined as the average annual wage for the electric power transmission and distribution industry by state. As in the case of Totex, this variable is also measured in year 2000 US dollars. The producer price index for power transmission is used as a proxy for capital price (KPR). Since we use a cost function to represent the technology, the dependent variable y is log of Totex, and the x variables are log of input prices and outputs. The inefficiency term u appears with a positive sign in the cost function because inefficiency increases cost. There is no such requirement for the noise term. Input prices and Totex are normalized by KPR to impose the linear homogeneity (in input prices) requirement on the cost function. Details on these variables listed below can be found in Llorca et al. (2016).

The weather variables used to explain cost inefficiency are annual minimum temperature in Fahrenheit degrees (tmin), average of the daily mean wind speeds in knots (wind), average of the daily precipitation in inches (prcp), and Capex/Opex Ratio (cor). We also include positive and negative average demand growth (pogrowth and negrowth) as two separate determinant variables. The peak load demand is used to explain skewness of the noise term, v .

Finally, we control regional variations in cost by including the following regional dummies in the cost function: serc - Reliability Corporation, spp - Southwest Power Pool, wecc - Electricity Coordinating Council, npcc - Northeast Power Coordinating Council, rfc - ReliabilityFirst Corporation, mro - Midwest Reliability Organization, and ercot - Electric Reliability Council of Texas. We report summary statistics of the variables and their definitions in Table 4. Some of these variables are constructed using the original variables. For instance, we used the variable tmincor, which is defined as the product of tmin (Temperature) and cor (ratio of Capex and Opex). All the variables used in the translog cost function and the skewness and inefficiency functions as well

Table 4 Variable definitions and sample statistics

Variable	Mean	Std. dev.	Definition
<i>I. Variables used in the cost frontier function:</i>			
Cost, outputs, input prices and time trend			
Totexd	145,111,000	120,627,000	Total cost (US\$)
del	6,279,733	8,872,918	Electricity delivered (MWh)
cap_subs	27821.41	22719.78	Total capacity of substations
pole	4073.052	3263.2	Network length measured in miles
price	5.8419	0.1944	Labor price
ppi	179.2072	21.3485	Producer price index
t	4.8557	2.5176	Time
Regional dummies			
serc	0.4030	0.4911	SERC Reliability Corporation
spp	0.2189	0.4140	Southwest Power Pool
wecc	0.2612	0.4398	Electricity Coordinating Council
npcc	0.0448	0.2071	Northeast Power Coordinating Council
rfe	0.2488	0.4328	ReliabilityFirst Corporation
mro	0.1393	0.3467	Midwest Reliability Organization
ercot	0.0448	0.2071	Electric Reliability Council of Texas
<i>II. Exogenous determinants of skewness and inefficiency, v and u:</i>			
peak	6208.02	5539.243	peak load
tmin	-10.3453	16.5734	Annual minimum temperature in Fahrenheit degrees
wind	6.8257	1.0109	Average of the daily mean wind speeds in knots
prcp	0.0712	0.0339	Average of the daily precipitation in inches
cor	1.1805	0.7019	Capex/Opex Ratio, where Capex is the sum of annual depreciation on capital assets and the annual return on the balance of capital and Opex is the total of the operation and maintenance expenses
pogrowth	1.8048	7.0760	Positive average demand growth
negrowth	-1.7729	4.4181	Negative average demand growth

The sample size is 402

as the regional dummies, along with their means and standard deviations, are reported in Table 4.

The model we discussed so far is for cross-sectional data. However, our data is an unbalanced panel of 59 US electricity transmission companies for the period 2001–2009 – giving us a total of 402 observations. Since this is the first paper on the use of SN noise with determinants of the shape parameters while using the II approach to estimate the parameters of the stochastic frontier model, we have not used company effects and other possible panel extensions in the present model. Instead, we use a pooled cross-sectional model with a time trend as a covariate and determinant of

inefficiency.⁵ Using panel notations, the translog cost function, which we estimate, is written as

$$\ln C_{it} = \beta_0 + \sum_m \beta_m \ln Y_{mit} + \gamma_1 \ln W_{it} + \frac{1}{2} \sum_m \sum_n \beta_{mn} \ln Y_{mit} \ln Y_{nit} + \frac{1}{2} \gamma_2 \ln W_{it} \ln W_{it} + \beta_t t + \sum \delta_d D_d + v_{it} + u(w_{it}), t = 1, 2, \dots, T_i, i = 1, 2, \dots, 59, \tag{25}$$

where $Y_m, m = 1, 2, 3$, are outputs, D_d are regional dummies, and t is the time trend. Cost (C) and input price (W) are normalized by the price of capital. Since there are only two inputs, we have one normalized input price (labor), which is W . The shape parameter λ is a function of z_{it} variables. Here, z_{it} includes only peak. The vector w_{it} contains t , tmin, wind, prcp, tmin \times cor, wind \times cor, prcp \times cor, pogrowth and negrowth, which are used as determinants of inefficiency modeled in terms of σ_u^2 . That is, we model inefficiency via $\sigma_{u, it} = \exp(\delta' w_{it})$, where $u_{it} \sim N^+(0, \sigma_{u, it}^2)$. We include neutral technical change by including the time trend variable (t) in the cost function. We also include t in w_{it} to allow inefficiency to change over time in a systematic manner.

Since the model with asymmetric noise is new, we test the specification with some other models. We estimate three models, viz., M1-M3 mentioned in Sect. 3, and test the appropriateness of each of them in terms of the data used. Because models M2 and M3 are special cases of M1, first we select between models M1 and M2 using the Wald test. That is, we test the joint hypothesis $H_0 : \lambda_0 = 0$ and $\lambda_1 = 0$, which is equivalent to testing $\lambda_i = 0$ for all i , because the skew-normal distribution with a zero shape parameter degenerates to the normal distribution. Using the bootstrap standard error from model M1, we compute the Wald statistic, which follows a $\chi^2(2)$ distribution and gives a value of 12,880. Based on this result, we reject model M2 at the 1% level of significance. Similarly, a test between models M2 and M3 is performed using the null hypothesis $H_0 : \lambda_0 = 0$. Since the shape parameter λ (from Table 5) is significant at the 5% level, we reject model M2. Finally, selection between models M1 and M3 is made by testing the hypothesis $H_0 : \lambda_1 = 0$. Since the estimated coefficient of peak is 0.0302 with the standard error 0.0177, the corresponding p-value is 0.087. Therefore, model M3 is rejected at the 10% level of significance, but not rejected at the 5% level of significance. Based on these tests, we conclude that model M1 is supported by the data. That is, we find evidence of asymmetric noise (v) even after controlling for the inefficiency term (u), which is always asymmetric. In a standard SF model (model M2), the composite error term ($v + u$) is skewed and the skewness of the composite error term is viewed as an indicator of inefficiency. This assertion is wrong if the noise term is skewed. Models M2 and M3 show how one can test for skewed v and still identify inefficiency.

We report the parameter estimates of M1-M3 in Table 5. Given that the parameters of the translog cost function have no direct interpretation, we compute scale elasticity for each output ($\partial \ln C / \partial \ln Y_m, m = 1, 2, 3$) and elasticity with respect to input price. These are reported in Table 6. The mean values of these elasticities with respect to

⁵ We plan to include panel features such as the fixed/random effects in our model while estimating it using the II approach in a separate paper.

Table 5 Results of II estimation of Models M1-M3

Parameter	M1: Hetero SN-HN		M2: N-HN		M3: Homo SN-HN	
	coeff.	s.e.	coeff.	s.e. ^a	coeff.	s.e.
del	-0.0243	0.0119	-0.3679	0.0450	0.0825	0.0347
cap_subs	0.2382	0.0139	0.0927	0.0171	0.0026	0.0170
pole	-5.1725	0.0348	-2.7109	0.3153	-5.4967	2.0187
price	-7.7518	0.0507	-7.5355	0.8740	-7.2145	2.6517
0.5×del ²	-0.0198	0.0107	-0.0084	0.0117	-0.1739	0.0158
0.5×cap_subs ²	0.3091	0.0133	0.2414	0.0308	0.3391	0.1266
0.5×pole ²	0.5652	0.0119	0.3184	0.0399	5.1759	0.1911
0.5×price ²	1.5304	0.0159	1.2070	0.1416	1.4155	0.5217
del×cap_subs	0.0219	0.0109	0.0046	0.0113	-0.0148	0.0176
del×pole	0.0675	0.0113	0.0355	0.0118	0.0731	0.0331
del×price	-0.0395	0.0112	0.0291	0.0112	-0.0371	0.0230
cap_subs×pole	-0.1711	0.0113	-0.1612	0.0216	-0.1734	0.0647
cap_subs×price	-0.2307	0.0124	-0.1232	0.0174	-0.2190	0.0186
pole×price	0.3990	0.0119	0.2261	0.0289	0.3508	0.1307
serc	0.1385	0.0129	-0.1216	0.0192	0.1411	0.0547
spp	-0.0860	0.0113	-0.1743	0.0227	-0.0432	0.0234
wecc	-0.1180	0.0122	0.1366	0.0205	-0.1786	0.0676
npsc	-0.2452	0.0141	0.3650	0.0449	-0.2474	0.0931
rfc	-0.2992	0.0113	-0.0424	0.0119	-0.2934	0.1081
mro	-0.4363	0.0122	-0.0687	0.0126	-0.5059	0.1864
ercot	0.0485	0.0110	0.0196	0.0125	0.0834	0.0359
t	0.0223	0.0116	0.0026	0.0110	0.0321	0.0201
constant	49.649	0.3142	43.1962	5.0180	49.5852	18.2098

Table 5 continued

Parameter	M1: Hetero SN-HN		M2: N-HN		M3: Homo SN-HN					
	coeff.	s.e.	coeff.	s.e. ^a	coeff.	s.e.				
SN noise: ^b										
σ_v	$v_{it} \sim SN(\sigma_v, \lambda_{it})$	0.9827	***	$v_{it} \sim N(0, \sigma_v^2)$	0.2986	***	$v_{it} \sim SN(\sigma_v, \lambda_{it})$	0.7275	***	0.0711
$\lambda_{it} = \lambda_0 + \lambda_1 peak_{it}$										
peak	0.0302	0.0177	*	NA	NA	NA	NA	NA	0.5076	
constant	-1.8586	0.0164	***	NA	NA	NA	-1.3871	***	0.5076	
Inefficiency: $u_{it} \sim N^+(0, \sigma_{u,it}^2)$, where $\sigma_{u,it} = \exp(\delta' w_{it})$										
t	0.0059	0.0532		0.0027	0.0387		-0.0086		0.0490	
tmin	-0.0032	0.0310		0.0157	0.0209		-0.0033		0.0288	
wind	0.1444	0.0273	***	0.3444	0.0576	***	0.1086	**	0.0543	
prcp	-3.1308	0.0230	***	-14.6285	1.6978	***	-3.2248	***	1.1813	
tmin×cor	-0.0075	0.0187		0.0018	0.0264		-0.0104		0.0304	
wind×cor	-0.1112	0.0282	***	-0.3359	0.0434	***	-0.1173	**	0.0570	
prcp×cor	2.4085	0.0194	***	12.5826	1.4632	***	2.1205	***	0.7812	
pogrowth	0.0019	0.0351		0.0095	0.0555		0.0005		0.0267	
negrowth	-0.0429	0.0340		-0.0029	0.0294		-0.0386		0.0479	
constant	-0.2171	0.0132	***	-0.2139	0.0290	***	0.2020	**	0.0796	
	mean	s.d.		mean	s.d.		mean		s.d.	
Predicted inefficiency and TE:										
$E(u_{it} \varepsilon_{it})$	0.1070	0.1201		0.4845	0.2899		0.5019		0.2068	
TE	0.9032	0.0098		0.6550	0.1673		0.6501		0.1029	

1. The number of bootstrap repetitions of all models is 1000. 2. *, ** and *** denote 10%, 5% and 1% significance levels

Table 6 Estimated elasticities of Models M1-M3

	Mean	S.D	Min	Max
M1: Hetero SN-HN				
del	0.2127	0.0513	0.0985	0.3417
cap_subs	0.9039	0.2309	0.0166	1.4270
pole	1.0307	0.3504	0.2014	2.3985
price	1.5325	0.3691	0.6472	2.6465
M2: N-HN				
del	0.0090	0.0242	-0.3821	0.0747
cap_subs	0.5346	0.1678	-0.2200	0.8953
pole	0.1093	0.1849	-0.3712	1.0120
price	0.5527	0.2591	-0.0339	1.4585
M3: Homo SN-HN				
del	0.0500	0.0435	-0.0634	0.1609
cap_subs	0.4606	0.2290	-0.5577	0.9844
pole	0.1012	0.3226	-0.6914	1.4083
price	1.0605	0.3389	0.3420	2.1687

Total number of observations is 402

$Y_1 - Y_3$ (del, cap_subs and pole) are 0.2127, 0.9039 and 1.0307, respectively, in model M1. For models M2 and M3, these elasticities are very different - 0.009, 0.5346, and 0.1093 in M2 and 0.0500, 0.4606, 0.1012 in M3. These elasticities for each output measure percentage change in cost for a one percent change in output, and they are supposed to be non-negative. It can be seen from Table 6 that the estimated elasticities from M2 and M3 are negative for some observations although their mean values are positive. For example, the elasticities of "del" from M2 has 169 negative values out of 402 values. In contrast all the estimated elasticities from M1 are positive. Similarly, the input price elasticity of labor from M1 is positive for all observations (as should be the case since an increase in input prices increases cost, ceteris paribus), whereas in M2 there are some negative values. These results also vindicate model M1 in addition to its support from the statistical test discussed above.

Estimates of technical change ($TC = \partial \ln C / \partial t$) - the coefficient of t - is found to be positive but insignificant in all three models. A positive (negative) value of TC indicates technical regress (progress) - an increase (decrease) in cost over time - holding everything else unchanged. TC in all three models is assumed to be neutral.

We predict technical (in)efficiency using the formula in (16), which gives predicted values of both inefficiency (u_{it}) and TE ($\exp(-u_{it})$) for each company at every year it is observed. It is worth mentioning that one should use $v = \varepsilon - u^r$ in (16) since we have a cost frontier model. The inefficiencies and TEs are predicted by Eq. (16) based on 1000 draws. The mean values of inefficiency and TE in M1 are 0.1070 and 0.9032, meaning that the costs of the transmission companies, on average, are 10.7% higher because of inefficiency. Alternatively, their TE, on average, is 90.32%. However, the mean inefficiency and TE from M2 and M3 are very different. These models predict that TE, on average, is only 65%. Since models M2 and M3 are rejected and model

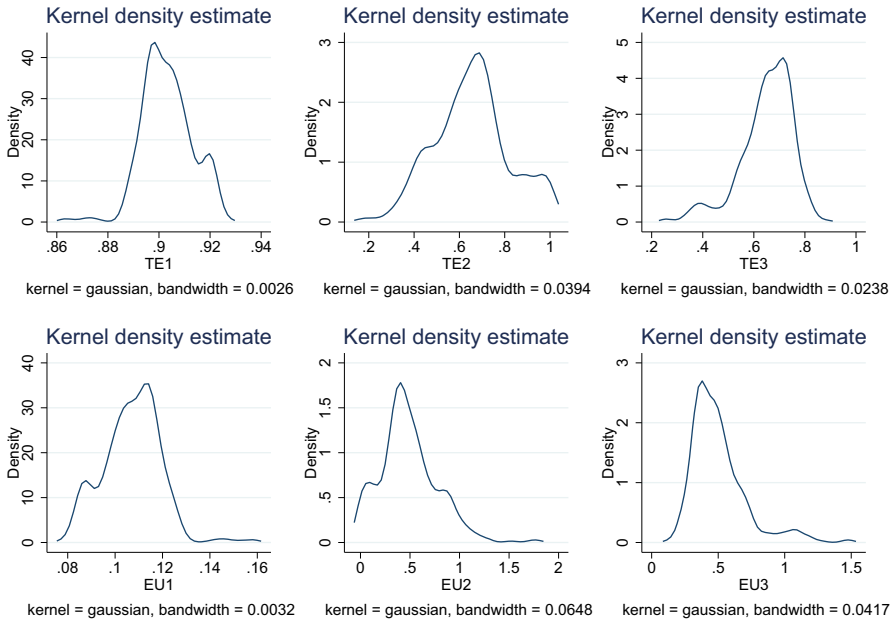


Fig. 2 The predicted inefficiencies and TEs of models M1-M3

M1 is supported by the data, we argue that the results from M1 are to be trusted more than those from the other two models for which mean inefficiencies, on average, are 48.45% and 50.19%. These numbers are too high to believe.

To get a closer look at the predicted inefficiency and TE, instead of their means, we report their density plots in Fig. 2. In model M1 most of the companies are more than 88% cost efficient (less than 10% inefficient). The spreads of cost inefficiency in M2 and M3 are much larger, resulting in estimated cost efficiency varying from less than 2% to almost 100% in model M2 and from 0.25% to 90% in model M3. Companies with very low cost efficiency in M2 and M3 drive the mean efficiency down, compared to M1. Model M1 shows that more than 50% of the companies are more than 90% efficient.

Since we include t as a determinant of inefficiency, we calculate technical efficiency change from $\Delta TE_{it} = -\partial(u_{it})/\partial t \approx -\partial(\mathbb{E}u_{it})/\partial t = \sqrt{2/\pi} \delta_t \times \exp(\delta' w_{it})$. Note that here we replace u_{it} by $\mathbb{E}u_{it}$ and then use the estimated values of the parameters to compute ΔTE_{it} . Recall that $\mathbb{E}u_{it} = \sqrt{2/\pi} \hat{\sigma}_{u, it} = \sqrt{2/\pi} \exp(\hat{\delta}'_t w_{it})$. Thus, the sign on δ_t indicates whether ΔTEC is increasing or decreasing. In models M1 and M2, $\hat{\delta}_t$ is positive, indicating a decreasing efficiency over time. In M3, $\hat{\delta}_t$ is negative, implying increasing cost efficiency. However, $\hat{\delta}_t$ is statistically insignificant in all three models.

5 Conclusion

This paper makes three contributions in the stochastic frontier (SF) models. First, the noise term is assumed to be asymmetric (skew normal) while the inefficiency term is assumed to be half-normal. This formulation avoids the criticism that skewness of the composite error term (sum of the noise and inefficiency) cannot be an indicator of inefficiency because skewness can arise from the noise term. Also, skewness in this model can be either positive or negative. Second, instead of using the standard ML method, we use the indirect inference (II) approach to estimate the parameters of the SF model, the parameters associated with the skewness function of the noise term, and determinants of inefficiency modeled via the variance of the half-normal inefficiency term. We tested this model against two special cases: (i) the noise term is symmetric and (ii) the skewness is a function of some covariates. Third, we generalize the SN and half-normal model even further in which the shape parameter of the SN model is made heteroscedastic (function of exogenous variables). Also, we added determinants of inefficiency by making the variance of the HN distribution heteroscedastic. We provide simulation results using the II estimation approach to show small sample performance of our model. Finally, we use a stochastic cost frontier model to showcase the workings of our model using an unbalanced panel of 59 US electricity transmission companies observed for the period 2001–2009. Test results show that our model M1 with heterogeneity in inefficiency and the skewness parameter is favored by the data against the two simpler models that are special cases of our model.

One caveat that we plan to address in the future is that our theoretical model is cross-sectional but the application uses a panel data. We used a pooled model without fixed/random effects to control for company-specific heterogeneity. Another future extension in the panel model would be to accommodate persistent inefficiency along with transient inefficiency with or without determinants.

References

- Adcock CJ (2010) Asset pricing and portfolio selection based on the multivariate extended skew-Student-t distribution. *Ann Oper Res* 176:221–234
- Badunenko O, Henderson DJ (2021) Production Analysis with Asymmetric Noise. MPRA_paper_110888.pdf
- Busch C, Domeij D, Guvenen F, Madera R (2020) Skewed idiosyncratic income risk over the business cycle: sources and insurance. *Am Econ J Macroecon* 14(2):207–42
- Coelli T (1995) Estimators and hypothesis tests for a stochastic frontier function: a monte carlo analysis. *J Prod Anal* 6:247–268
- Gourieroux C, Monfort A, Renault E (1993) Indirect inference. *J Appl Economet* 8(S1):S85–S118
- Hafner CM, Manner H, Simar L (2018) The “wrong skewness” problem in stochastic frontier models: a new approach. *Economet Rev* 37(4):380–400. <https://doi.org/10.1080/07474938.2016.1140284>
- Kraus A, Litzenberger RH (1976) Skewness preference and the valuation of risk assets. *J Financ* 31:1085–1100
- Lai H-P (2022) Indirect Inference of Stochastic Frontier Models, working paper
- Lai H-P, Kumbhakar SC (2018) Panel data stochastic frontier model with determinants of persistent and transient inefficiency. *Eur J Oper Res* 271(2):746–755
- Lai H-P, Kumbhakar SC (2021) Panel stochastic frontier model with endogenous inputs and correlated random components. *J Bus Econ Stat*, forthcoming

- Lai H-P, Tran K (2022) Persistent and transient inefficiency in a spatial autoregressive panel stochastic frontier model. *J Product Anal*, forthcoming
- Llorca M, Orea L, Pollitt MG (2016) Efficiency and environment factors in the US electricity transmission industry. *Energy Econ* 55:234–246
- Qi L, Bravo-Ureta BE, Cabrera VE (2015) From cold to hot: climatic effects and productivity in Wisconsin dairy farms. *J Dairy Sci* 98:8664–8677
- Schmidt P, Lin T-F (1984) Simple tests of alternative specifications in stochastic frontier models. *J Econom* 24:349–361
- Simar L, Wilson PW (2010) Inferences from cross-sectional stochastic frontier models. *Economet Rev* 29:62–98

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.