PANEL STOCHASTIC FRONTIER ANALYSIS WITH POSITIVE SKEWNESS

Rachida El Mehdi, Christian M. Hafner

REPRINT | 2024 / 31

UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA
Voie du Roman Pays 20 - L1.04.01
B-1348 Louvain-la-Neuve
Email : lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

Panel stochastic frontier analysis with positive skewness

Rachida El Mehdi[1*] and Christian M. Hafner[2†]

[1*]National School of Applied Sciences, SmartICT Laboratory, Mohammed First University, Alqods, Oujda, 60000, Oriental, Morocco.
[2]Institut de Statistique, Biostatistique et Sciences Actuarielles and LIDAM, Université Catholique de Louvain, Voie du Roman Pays, Louvain-la-Neuve, 1348, Wallonie, Belgium.

*Corresponding author(s). E-mail(s): r.elmehdi@ump.ac.ma;
Contributing authors: christian.hafner@uclouvain.be;
†This author contributed equally to this work.

Abstract

This paper focuses on solving the problem of technical efficiency estimation for panel data when residuals are right-skewed. Indeed, there is an ambiguity in stochastic frontier analysis when the residuals of the ordinary least squares estimates are right-skewed, which might indicate that either there is no inefficiency, or that the model is misspecified. To overcome and avoid this problem, we propose a panel model in which the inefficiency term has an extended half-normal distribution. Hence, our work is an extension of existing work for the cross-section case to panel data with time varying inefficiencies. We first propose estimators of the inefficiency under the extended half-normal distribution assuming independence between the noise and the inefficiency term. A simulation study illustrates the good performance of our procedure. An application to drinking water for forty-two Moroccan municipalities in the period 2017 to 2019 favors our extended model. Results reveal that the performance of this public sector is generally medium and therefore the waste was significant.

Keywords: Efficiency, Extended-half-normal, Panel data, Positive skewness, Stochastic frontier analysis.

Corresponding author ORCID:: https://orcid.org/0000-0002-9996-3960

1 Introduction

The basic idea of frontier analysis is the comparison between decision making units, DMU (firms, for example), in order to know how the inputs are used to have outputs through a technical efficiency (TE) score. The literature proposes so-called Farrell efficiency scores, which are between 0 and 1, and Shephard efficiency scores, which are just the inverse of Farrell scores and thus are greater than 1, see for example Aigner, Lovell, and Schmidt (1977) and Battese and Coelli (1988). In the classical stochastic frontier analysis (SFA), the error term, denoted here ϵ, is divided into two independent terms. Thus, it is given by $\epsilon = v - u$ where v is the statistical noise term and u is the inefficiency term.

In frontier analysis, several papers such as Koopmans (1951), Debreu (1951) and Farrell (1957) have developed measurements of technical efficiency. In the stochastic frontier approach, the first contributions in this context, Aigner et al. (1977) and Meeusen and Van Den Broeck (1977), have provided stochastic production models with composed error terms suggesting a half-normal and an exponential inefficiency term. Later, some works proposed a set of distributions for the inefficiency term and all deserve particular interest. We cite here, among others, Stevenson (1980), who develop a specification for a stochastic frontier model by suggesting a model based on the truncated normal distribution created by shifting the half normal one; Greene (1990) allows the one-sided part of the error term to have a more flexible distribution which is the two-parameter Gamma distribution; Deprins and Simar (1989) and Migon and Medici (2001) consider the lognormal and Weibull distributions for the inefficiencies in an SFA Bayesian approach; Hajargasht (2015) introduces a stochastic frontier model with a one-parameter distribution called the Rayleigh distribution and extends also the Rayleigh model to a model with environmental variables and to a model with time-varying features in panel stochastic frontier models.

Since the statistical noise is assumed to be symmetric about zero and the inefficiency is positive for the SFA production model, any skewness of the error term distribution is attributed to the inefficiency term as pointed out in Hollingsworth and Peacock (2008). In other words, except for the sign, the two skewnesses of the error and the inefficiency terms are the same. The classical SFA procedure leads generally to a negative skewness of the residuals resulting from the Ordinary Least Squares (ols) estimation, which consequently biases the efficiency estimate. When it leads to a positive skewness, that might indicate that there is no inefficiency, so all Decision Making Units (DMUs) are efficient, or that the model is misspecified.

Indeed, we were confronted with this kind of positive skewness problem in El Mehdi and Hafner (2014a) which handled, in the cross-section case, the nonparametric estimation of technical efficiencies of Moroccan municipalities regarding the relationship between local finances of the communal councils of the Eastern region and financial autonomy where the analysis revealed, contrary to reality, that all communal councils were efficient. The problem was solved by increasing the size of the dataset and comparing municipal councils at the national level instead of just the eastern region. We were fortunate to be able to expand the size of the dataset but unfortunately this is not always possible. The same problem happened in a bootstrap procedure for simulated data in El Mehdi and Hafner (2014b) which presented, in the cross-section case, a parametric estimation of technical efficiencies where the two components of the error term, statistical noise and inefficiency, are dependent and where this dependence is expressed by a Clayton copula function.

Some papers have proposed procedures to establish the efficiency estimate in the case of independence between the two error terms with positive skewness. In this context, Daniel, Hafner, Manner, and Simar (2011) present a way by changing the model specification to estimate the SFA model. It shows that one can have inefficiency with positive skewness and that, with its extended model, convergence problems of maximum likelihood estimators in cases of positive sample skewness reported in Aigner et al. (1977) are avoided, which is not possible in the classical model. As for Almanidis, Qian, and Sickles (2014), it introduces a stochastic frontier model with bounded inefficiencies which is a doubly truncated normal distribution. In the same context, Hafner, Manner, and Simar (2018) proposes, for the cross-section case, an estimation of the model by generalizing the inefficiency distribution.

For Hafner et al. (2018), an extended distribution for the inefficiency term was considered as a distribution mirrored at zero and then shifted to the right by some constant to have the same mean as the original distribution and truncated at zero to ensure that inefficiency can only be positive. It is characterized by a negative skewness as opposed to the positive skewness of the original distribution. Hence, the error term is given by $\epsilon = v - u$ where v is the statistical noise term and u is the inefficiency term with an extended distribution. Computations for the two basic extended half-normal and extended exponential are provided for the cross-sectional data.

Our present work is an extension of Hafner et al. (2018) to a panel data framework with time varying inefficiency when the inefficiency term has an extended half-normal distribution. The advantage of the extended model of Hafner et al. (2018) is that it nests the classical one which is not the case for the previous works and it allows for negative skewness which solves the positive skewness problem. The contribution of our paper is to propose a panel SFA version of the normal-extended half normal model. Our approach therefore consists in developing the theoretical procedure and its computation when the inefficiency has an extended-half-normal distribution as well as a numerical application for simulated and real panel data characterized by an OLS positive skewness.

In the following we present in Section 2 the error term density and the technical efficiency expressions in the extended half-normal (HN) distribution case where the time effect is a time varying inefficiency of Battese and Coelli (1992) in the panel SFA model when data are right skewed. A simulation and an empirical study are the subject of Section 3 and Section 4, respectively. Section 5 will conclude this paper.

2 The panel SFA model with extended half-normal distribution

A parametric approach to frontier analysis uses in particular the stochastic frontier analysis method. In the basic panel SFA, the model of the i^{th} DMU at the t^{th} time period, assuming that inefficiencies vary systematically with time, is specified as

$$y_{it} = f\left(\underline{x}_{it}, \beta\right) + \epsilon_{it}, i = 1, \ldots, n \ ; \ t = 1, \ldots, T \tag{1}$$

with input variables $\underline{X}_{it} \in \mathbb{R}_{+}^{p}$, output variable $Y_{it} \in \mathbb{R}_{+}$, and where $y_{it} = \log\left(Y_{it}\right)$, $\underline{x}_{it} = \log\left(\underline{X}_{it}\right)$, $\beta \in \mathbb{R}^{l+T}$ is a vector of unknown parameters to be estimated, ϵ_{it} is a stochastic error term, f is the frontier function, n is the number of DMUs under study and T is the number of time periods. If intercepts are constant over time, then $\beta \in \mathbb{R}^{l+1}$. The number of input variables is p, and l is the number of parameters to be estimated excluding potentially time-varying intercepts. The stochastic term ϵ_{it} contains information about both the noise and the inefficiency. It can be decomposed into two independent terms, a statistical noise v_{it} and a technical inefficiency u_{it} such that $\epsilon_{it} = v_{it} - u_{it}$ for the production frontier. We assume that v_{it} is a Gaussian error term, $v_{it} \sim N\left(0, \sigma_v^2\right)$, and u_{it} is an inefficiency term with non-negative support, $u_{it} \geq 0$.

In the presence of a right skewness, Hafner et al. (2018) proposes two extended distributions, among them the extended half-normal. Let us denote the half-normal density function as $f_{\gamma}^{+}\left(u_i\right) = \frac{2}{\gamma}\phi\left(\frac{u_i}{\gamma}\right)$ for $u_i > 0$, where $\gamma > 0$ and define the extended half-normal density function as $f_{\gamma}^{-}\left(u_i\right) = \left[\Phi\left(a_0\right) - \Phi\left(0\right)\right]^{-1}\frac{1}{|\gamma|}\phi\left(a_0 - \frac{u_i}{|\gamma|}\right)$ for $0 < u_i < B$, where $\gamma < 0$, a_0 is a constant that can be approximated by $a_0 = 1.3892032925$, Φ and ϕ are the standard normal probability density and the cumulative density functions, respectively, and $B = a_0\left|\gamma\right|$. Thus, our principal contribution consists in the proposal of a panel data version for the SFA model with the extended half-normal distribution of Hafner et al. (2018) under the independence hypothesis between the error terms and considering the inefficiency term $u_{it} = \eta\left(t\right)u_i$ proposed in Lee and Schmidt (1993), where $\eta\left(t\right)$ is a function of time. The special case $\eta\left(t\right) = \exp\{-\eta_1\left(t - T\right)\}$ of Battese and Coelli (1992), where η_1 is a parameter to be estimated, is adopted in our application study. We first provide theoretical expressions for the error term density and the technical efficiency using this extended half-normal distribution. We then apply the model to Moroccan drinking water data characterized by a positive skewness.

2.1 Error term density of the normal extended half-normal model

For panel data, assuming $v_{it} \sim N\left(0, \sigma_v^2\right)$ for the noise term and an extended half-normal distribution u_{it} for the inefficieny term, the density of the error term $\epsilon_i = \left(\epsilon_{i1}, \ldots, \epsilon_{it}, \ldots, \epsilon_{iT}\right)$ and technical efficiency TE_{it} for each DMU$_i$, $i = 1, 2, \ldots, n$ at time $t = 1, 2, \ldots, T$ are presented in the following Theorem 1 and Theorem 2, respectively. Proofs of the two theorems are given in the Appendices A and B.

Theorem 1. *Assume that the noise term $v_{it} \sim N\left(0, \sigma_v^2\right)$, that the inefficiency term $u_{it} = \eta\left(t\right)u_i$, where $\eta\left(t\right)$ is a function of time and where u_i follows an extended half-normal distribution with density $f_{\gamma}^{-}\left(u_i\right) = \left[\Phi\left(a_0\right) - \Phi\left(0\right)\right]^{-1}\frac{1}{|\gamma|}\phi\left(a_0 - \frac{u_i}{|\gamma|}\right)$ for*

$0 < u_i < B$, $\gamma < 0$, $a_0 = 1.3892032925$, $B = a_0 |\gamma|$, ϕ and Φ denote respectively the standard probability and the standard cumulative normal functions, and the error term $\epsilon_{it} = v_{it} - u_{it}$. Assume further that the errors ϵ_{it} are i.i.d., that v_{it} are i.i.d., that u_{it} are i.i.d. and that the two components v_{it} and u_{it} are independent for all $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$. Then, the error term density is given by

$$g^- (\epsilon_i) = \frac{\sigma_* \exp\left\{ -\frac{1}{2} a_{*i} \right\}}{(2\pi)^{T/2} \sigma_v^T [\Phi(a_0) - \Phi(0)] |\gamma|} \left[\Phi\left(\frac{B - \mu_{*i}}{\sigma_*} \right) - \Phi\left(-\frac{\mu_{*i}}{\sigma_*} \right) \right] \tag{2}$$

where

$$\sigma_*^2 = \frac{\sigma_v^2 \gamma^2}{\gamma^2 \sum_t \eta^2(t) + \sigma_v^2} \ ,$$
$$\mu_{*i} = \frac{-\gamma^2 \sum_t \epsilon_{it} \eta(t) + \sigma_v^2 B}{\gamma^2 \sum_t \eta^2(t) + \sigma_v^2} \ ,$$
$$a_{*i} = \frac{\gamma^2 \sum_t \epsilon_{it}^2 + \sigma_v^2 B^2}{\sigma_v^2 \gamma^2} - \frac{\mu_{*i}^2}{\sigma_*^2}.$$

Let us now consider the error density depending on the parameter $\gamma \in \mathbb{R}$. The sign of γ reflects the skewness of u. If $\gamma > 0$, the inefficiency term u has the classical half-normal density; if $\gamma < 0$, the density of u is an extended half-normal; if $\gamma = 0$ there is no inefficiency and, consequently, in this case ϵ is symmetric given that it has the normal distribution of v. The error term density which depends on γ is defined in Hafner et al. (2018) as

$$g(\epsilon_i, \gamma) = g^+(\epsilon_i) I(\gamma > 0) + g^-(\epsilon_i) I(\gamma < 0) + g^0(\epsilon_i) I(\gamma = 0), \tag{3}$$

where I is the indicator function, $g^-(\epsilon_i)$ is the density defined in (2) and $g^+(\epsilon_i)$ for panel data is defined in Kumbhakar and Knox Lovell (2000), page 111 as

$$g^+(\epsilon_i) = \frac{2\sigma_{**} \exp\left\{ -\frac{1}{2} a_{**i} \right\}}{(2\pi)^{T/2} \sigma_v^T \sigma_u} \left(1 - \Phi\left(-\frac{\mu_{**i}}{\sigma_{**}} \right) \right) \tag{4}$$

with

$$\sigma_{**}^2 = \frac{\sigma_v^2 \sigma_u^2}{\sigma_u^2 \sum_t \eta^2(t) + \sigma_v^2} = \sigma_*^2 \ ,$$
$$\mu_{**i} = \frac{\sigma_v^2 \sum_t \eta(t) \epsilon_{it}}{\sigma_u^2 \sum_t \eta^2(t) + \sigma_v^2} \ ,$$
$$a_{**i} = \frac{1}{\sigma_v^2} \left[\sum_t \epsilon_{it}^2 - \frac{\sigma_u^2 \left(\sum_t \eta(t) \epsilon_{it} \right)^2}{\sigma_u^2 \sum_t \eta^2(t) + \sigma_v^2} \right].$$

and where for panel data

$$g^0(\epsilon_i) = \frac{1}{\sigma_v^T (2\pi)^{T/2}} \exp\left\{ -\frac{1}{2\sigma_v^2} \sum_t \epsilon_{it}^2 \right\}. \tag{5}$$

Therefore, assuming independence across DMUs, the log likelihood function, given in equation (39) of Hafner et al. (2018), can be written for $\vartheta = (\gamma, \sigma_v, \beta, \underline{\eta}_k)$, where $\underline{\eta}_k$ is a vector of k parameters in $\eta(t)$, as

$$\begin{aligned} l(\vartheta) &= \log L(\vartheta) \\ &= \sum_{i=1}^{n} \Big[\log\left(g^{+}(\epsilon_i)\right) I(\gamma > 0) + \log\left(g^{-}(\epsilon_i)\right) I(\gamma < 0) \\ &\quad + \log\left(g^{0}(\epsilon_i)\right) I(\gamma = 0) \Big]. \end{aligned} \tag{6}$$

The Maximum Likelihood (ML) estimator of ϑ is then defined as

$$\widehat{\vartheta}_{ML} = \arg\max_{\vartheta} l(\vartheta)$$

which is obtained by numerical maximization, as no analytical solution is available.

2.2 Technical efficiency expression

As usual, we define technical efficiency of the DMU_i at time t as $E(\exp\{-u_{it}\} \mid \epsilon_i)$. Depending on the observed data $\underline{x}_{it}$, y_{it} and on ϑ, an expression for technical efficiency TE_{it} for the extended half-normal model is provided in the following Theorem 2.

Theorem 2. *Under the same assumptions as in Theorem 1 and assuming $\gamma < 0$, technical efficiency is given by*

$$TE_{it} = E\Big(\exp\{-u_{it}\} \mid \epsilon_i\Big) = B_{it}/B_i = B_{it}/g^{-}(\epsilon_i) \tag{7}$$

where

$$B_{it} = \frac{\sigma_* \exp\left\{-\frac{1}{2}a_{*it}\right\}}{(2\pi)^{T/2}\sigma_v^T\left[\Phi(a_0) - \Phi(0)\right]|\gamma|}\left[\Phi\left(\frac{B-\mu_{*it}}{\sigma_*}\right) - \Phi\left(-\frac{\mu_{*it}}{\sigma_*}\right)\right],$$

$$a_{*it} = \frac{\gamma^2\sum_t \epsilon_{it}^2 + \sigma_v^2 B^2}{\sigma_v^2\gamma^2} - \frac{\mu_{*it}^2}{\sigma_*^2}.$$

$$\mu_{*it} = \frac{-\gamma^2\sum_t \epsilon_{it}\eta(t) + \sigma_v^2 B - \sigma_v^2\gamma^2\eta(t)}{\gamma^2\sum_t \eta^2(t) + \sigma_v^2}.$$

We note here that if $\gamma > 0$, the expression for technical efficiencies is given by equation (3.3.33), page 112 of Kumbhakar and Knox Lovell (2000). Technical efficiencies in this case are provided by the frontier package in R. In contrast, if γ is zero, the technical efficiencies are equal to one.

3 Simulation study

We provide a Monte Carlo simulation exercise to study the performance of our model for panel data presenting a positive skewness. Note that the positive skewness problem occurs mainly when the size of the database is small, so when both the number of DMUs and the number of time periods are small. Table 1 illustrates this phenomenon. It indicates that 38% of the simulated datasets have a positive skewness when $n = 20$ and $T = 2$, but that no positive skewness is found when $n = 100$ and $T = 10$.

Table 1 Percentage of positive skewed panel datasets according n and T

Size n	$T = 2$	$T = 3$	$T = 5$	$T = 7$	$T = 10$
20	38	16	1	0	0
50	20	7	0	0	0
75	12	4	0	0	0
100	10	1	0	0	0

For the simulated data, we generate the exogenous variables as $X_{1it} \sim N(1.5, 0.3)$, $X_{2it} \sim N(1.8, 0.3)$, and fix the parameters as $\beta_0 = 0.9$, $\beta_1 = 0.6$, $\beta_2 = 0.5$, $\sigma_v = 0.25$. The time-varying efficiency function is specified as in Battese and Coelli (1992) as $\eta(t) = \exp\{-\eta_1(t - T)\}$, where $\eta_1 = 0.3$. The number of DMUs is set to $n = 50$, and the number of time periods to $T = 3$. Hence, $u \sim N^+\left(0, \sigma_u^2\right)$ with $E(u) = \mu = 0.2$ and $\sigma_u = \sqrt{\pi/2}\mu = 0.25$ for the ordinary half-normal distribution and with $u_{it} = \eta(t)\, u_i$, for all $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$. The time function $\eta(t) = \exp\{-\eta_1(t - T)\}$ is chosen for its simplicity and its widespread use in frontier analysis, but other functions could of course be used as well.

The model considered in this simulation study is defined as

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \epsilon_{it}, i = 1, \cdots, n\ ;\ t = 1, \cdots, T \tag{8}$$

where $y_{it} = \log(Y_{it})$, $x_{1it} = \log(X_{1it})$, $x_{2it} = \log(X_{2it})$ and the intercept β_0 is constant over time. In the classical model $\epsilon_{it} = v_{it} - u_{it}$ and for this simulation study we consider that $u_{it} = \eta(t)\, u_i$, such that $\epsilon_{it} = v_{it} - \eta(t)\, u_i$.

We first note that the OLS skewness of our simulated dataset in this case is 0.473, so it is right skewed. Results of the estimation of the classical normal-half-normal model with the frontier package of R for the time-varying efficiencies of Battese and Coelli (1992) are given in Table 2. However, we should be careful with these results given that data present a positive skewness and the classical SFA expects a negative one. It should be mentioned that the frontier package of R returns the estimate of the composite error variance σ_ϵ^2, or σ^2 for simplicity, and not those of the noise σ_v^2 and the inefficiency σ_u^2. In the stochastic frontier analysis, the expressions $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\omega = \frac{\lambda^2}{1+\lambda^2}$ and $\lambda = \frac{\sigma_u}{\sigma_v}$ allow us to deduce the estimated values of σ_v and σ_u reported in Table 2.

For the normal-extended half-normal model (the noise term is normal and the inefficiency term is an extended half-normal), OLS estimates are used as initial values to optimize the log-likelihood function. A grid of values for the time parameter η_1 was proposed and the model with the highest log-likelihood value was chosen. Model estimation is given in Table 3. It shows that all parameters are significant at a 5% significance level given that all t-statistics are greater than 1.96. Furthermore, all true values are in their associated confidence intervals except for σ_v which is almost equal to the upper bound.

One of the main goals of frontier analysis is to estimate technical efficiencies TE. Table 4 provides, for both the classical and the extended models, Farrell technical efficiency scores defined in Farrell (1957). A comparison between them indicates that for 98% DMUs efficiencies are overestimated in the classical model. Comparing the TE averages for each period, it is clear that they are not the same for the two models. The extended model has a better goodness-of-fit given that it has a higher likelihood value than the classical model.

We also compute the mean squared error (MSE) in order to evaluate and compare the performance of the classical and normal-extended half-normal models. Hence, a Monte Carlo design is established using Algorithm#3 of Simar and Wilson (2010) adapted to our panel data framework (see El Mehdi and Hafner (2021), step 2 of the algorithm in the independence case) and considering the expression $TE_{it} = E\Big(\exp\{-u_{it}\} \mid \epsilon_i\Big)$. Thus, for each replication a new response y is created by keeping the exogenous variables x_1 and x_2 from our original simulated database (associated to Tables 2-4) and by generating independently a noise variable v_{it} from a normal distribution and an inefficiency variable u_i from a half-normal distribution. For each new y available, the model is re-estimated, giving new estimates of $TE_{it}, t = 1, ..., T$ for each $DMU_i, i = 1, ..., n$. The experiment is repeated $M = 10000$ times, creating thus $T \times M$ technical efficiencies for each DMU_i (See Simar and Wilson (2010) for more details). In order to provide an MSE for each i, the TE_{it} average according to the T periods are calculated for the original model, denoted TE_i and also for each m of the Monte Carlo trials, denoted TE_i^m, and finally the $MSE(i)$ is evaluated as $MSE(i) = \sum_{m=1}^{M} \left(TE_i^m - TE_i\right)^2 / M$. The general MSE is obtained by the average of the $MSE(i)$ of the n $DMUs$ as $MSE = \sum_{i=1}^{n} MSE(i)/n$.

The MSE results of the OLS skewed samples (a proportion of 0.0123) reported in Table 4 indicate that for 43 DMUs, or 86%, the MSE of the classical normal-half normal model is higher than that of the extended one. Moreover, and regarding the general MSE, its value for the classical model (0.0212) shows that it is wider than that of the extended one (0.0110). Besides, the Student test for equality of two means clearly rejects the equality of the two general $MSEs$ at a significance level of 5% and a p-value=0.0001479, confirming the significant difference between the two mean square errors. Thus, in the half-normal case, when the residuals present positive skewness, our results strongly support the model with extended-half-normal distribution for the inefficiency term u compared with the classical half-normal model.

Table 2 ML estimators of the normal-HN model for simulated data

	Estimate	Std. Error	z-value	$Pr(>\|z\|)$	
β_0	0.832	0.078	10.666	$<$ 2.2e-16	***
β_1	0.684	0.094	7.310	2.68e-13	***
β_2	0.653	0.101	6.447	1.14e-10	***
σ^2	0.092	0.022	4.271	1.95e-05	***
ω	0.562	0.118	4.755	1.99e-06	***
η_1	0.238	0.079	2.997	0.002724	**
λ	1.133				
σ_u	0.228				
σ_v	0.201				
log likelihood value: -0.4423 (df=6)					

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 3 ML estimates of the normal-extended HN model and their confidence intervals for simulated data

	Estimate	Std. Error	z-value	Lower bound	Upper bound
γ	-0.364	0.065	-5.601	-0.494	-0.233
σ_v	0.200	0.014	14.198	0.172	0.228
β_0	0.914	0.088	10.339	0.736	1.092
β_1	0.733	0.092	7.964	0.548	0.919
β_2	0.659	0.104	6.314	0.448	0.869
η_1	0.191	0.062	3.064	0.065	0.317
log likelihood value: 0.0695656 (df=6)					

4 Application to Moroccan drinking water production

We now apply the proposed panel data model to a real data set collected from Moroccan statistical yearbooks concerning the production and sales of drinking water for 42 municipalities during a period of three years from 2017 to 2019. It is a sector managed by the government in order to provide for the citizen's needs. The response variable Y is the water production, while the exogenous variables are sales, denoted X_1, including sales to the subscribers and to the autonomous authorities, and the number of subscribers, denoted X_2. We noticed missing data in almost all sectors for other explanatory variables and therefore decided to keep this simple model with two variables. Hence, the model is a simple case of equation (8) defined for $n = 42$ DMUs and

Table 4 Technical efficiency of the normal-HN and normal-extended HN models for simulated data and MSE of the ols skewed samples

	Normal-HN model				Normal-Extended HN model			
DMU	Period 1	Period 2	Period 3	MSE	Period 1	Period 2	Period 3	MSE
1	0.529	0.605	0.672	0.0538	0.517	0.580	0.638	0.0204
2	0.756	0.801	0.839	0.0080	0.641	0.692	0.737	0.0112
3	0.597	0.665	0.724	0.0340	0.536	0.597	0.653	0.0179
4	0.702	0.755	0.801	0.0136	0.596	0.652	0.702	0.0138
5	0.771	0.814	0.849	0.0085	0.658	0.707	0.751	0.0098
6	0.895	0.915	0.932	0.0107	0.820	0.848	0.872	0.0060
7	0.678	0.735	0.784	0.0155	0.582	0.639	0.690	0.0141
8	0.889	0.911	0.929	0.0217	0.806	0.836	0.862	0.0066
9	0.768	0.811	0.847	0.0091	0.650	0.700	0.744	0.0113
10	0.615	0.680	0.737	0.0313	0.545	0.605	0.660	0.0176
11	0.846	0.876	0.900	0.0120	0.740	0.779	0.813	0.0080
12	0.892	0.913	0.930	0.0170	0.810	0.840	0.865	0.0056
13	0.680	0.737	0.786	0.0183	0.587	0.643	0.694	0.0139
14	0.883	0.906	0.925	0.0201	0.801	0.831	0.858	0.0057
15	0.558	0.630	0.694	0.0462	0.525	0.587	0.644	0.0194
16	0.883	0.906	0.925	0.0155	0.803	0.833	0.859	0.0057
...	...	...	...	...	...	...	...	...
40	0.643	0.705	0.758	0.0221	0.560	0.619	0.673	0.0156
41	0.870	0.896	0.916	0.0144	0.779	0.813	0.842	0.0065
42	0.928	0.942	0.954	0.0220	0.881	0.900	0.916	0.0039
43	0.909	0.927	0.942	0.0217	0.841	0.866	0.887	0.0055
44	0.922	0.938	0.950	0.0170	0.872	0.893	0.910	0.0034
45	0.768	0.811	0.847	0.0124	0.653	0.703	0.747	0.0102
46	0.760	0.804	0.842	0.0101	0.648	0.698	0.743	0.0113
47	0.684	0.740	0.788	0.0154	0.588	0.645	0.695	0.0132
48	0.914	0.931	0.945	0.0252	0.855	0.878	0.898	0.0037
49	0.813	0.848	0.878	0.0100	0.707	0.750	0.788	0.0079
50	0.846	0.875	0.900	0.0103	0.738	0.777	0.811	0.0085
Mean	0.758	0.801	0.837	0.0212	0.675	0.721	0.762	0.0110

$T = 3$, where $y_{it} = \log(Y_{it})$, $x_{1it} = \log(X_{1it})$, $x_{2it} = \log(X_{2it})$. The time function is defined as $\eta(t) = \exp\{-\eta_1(t-T)\}$.

Testing the skewness can be implemented by estimating the model using OLS as pointed out by Kumbhakar, Parameter, and Zelenyuk (2018). The OLS estimation for our model reveals that residuals are right skewed, meaning that either there is no inefficiency or the model is misspecified. However, given that the variable number of subscribers (X_2) is insignificant with a p-value equal to 0.53, we adopted in this study the model including just the exogenous variable sales (X_1) which is highly significant and where the skewness coefficient 2.798 is important and positive.

The positive skewness property for our panel data is not desirable in a classical SFA model because we would expect a negative skewness if the model was correctly specified. To solve this problem, our model will be estimated using the procedure developed in this paper. Estimation using the classical normal-half normal estimator is done using the frontier package of Coelli and Henningsen (2013), which is an improved R version of the Coelli (1996) software. It reveals that the intercept and the sales parameter are highly significant at a significance level of 5%, and that the time parameter is not significant as shown in Table 5, but with a warning stating that data are right skewed.

Table 5 ML estimates of the normal-HN model for water data

	Estimate	Std. Error	z-value	$Pr(>\|z\|)$	
β_0	6.625	0.934	7.090	1.345e-12	***
β_1	0.368	0.067	5.466	4.600e-08	***
σ^2	0.462	0.457	1.012	0.312	
ω	0.882	0.385	2.290	0.022	*
η_1	0.102	0.309	0.329	0.743	
λ	2.734				
σ_u	0.639				
σ_v	0.234				
log likelihood value: -74.725 (df=5)					

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

In order to overcome this positive skewness problem, we estimate the same model described in (8) but without the x_2 variable using our proposed procedure. In this normal-extended half-normal estimation, OLS estimates with a grid of values for the time parameter are used as initial values to optimize the log-likelihood function. Results are given in Table 6 which shows that the t-statistics for the parameters γ, σ_v, β_0 and β_1 are largely greater than 1.96 with 5% as a significance level. However, the time effect parameter is insignificant at the same level.

In addition, the Farrell technical efficiencies estimates of the normal-half normal and the normal-extended half normal models are presented in Table 7. For the extended model, they increase but slowly over time because the time effect is rather weak. The cross-section averages are greater than those of the classical model, suggesting that the latter are slightly underestimated. They were underestimated for thirty three municipalities according to the TE means over time for each decision making unit, representing thus a percentage of 78.6% among all municipalities under study. In addition, results show that on average, the municipalities under study could have been efficient and ensured subscribers demand with just 57% of their drinking water production. Furthermore, the log-likelihood value of the extended half normal model is higher than the classical one, which favors the extended model.

Table 6 ML estimates of the normal-extended HN model for water data

	Estimate	Std. Error	z-value
γ	-0.761	0.019	40.068
σ_v	0.047	0.001	79.000
β_0	3.669	0.001	4586.125
β_1	0.687	0.001	1144.167
η_1	0.011	0.009	1.299
log likelihood value: 102.145 (df=5)			

In comparison with our empirical study on drinking water supply established in El Mehdi and Hafner (2021) for a panel database covering the period 2001/2007, we noticed, here again, the poor performance of the drinking water sector in Morocco. Even if technical efficiencies are on average low, a small progress is made by the National Office of Drinking Water (ONEP according to its French abbreviation) after more than ten years. So, the waste was and remains considerable.

Given that waste is significant, measures have been applied since 2022 by the government or by municipal councils (depending on the municipality performance in this area) in order to reduce the waste, such as cutting off the drinking water supply for certain periods of the day, closing three days a week for car wash companies and public baths, reducing the use of public fountains, creating desalination stations for watering public gardens and grounds, rapid repair of canals. The consequences of these measures can only be evaluated in the future.

5 Conclusion

The main contribution of our research is the theoretical challenge of solving the right skewness problem in the panel stochastic frontier analysis. Indeed, a version of the model using the extended half-normal distribution for the inefficiency term adapted to panel data case is developed. Expressions of the error term density and of technical efficiency are given considering this extended model in order to compare DMUs. A simulation study was made to illustrate the good performance of our procedure in the presence of a positive skewness. It showed that true values are in their associated confidence intervals except for the standard deviation parameter of the noise term σ_v which is almost equal to the upper bound. Finally an application to Moroccan drinking water production and sales revealed a positive skewness problem. Our results showed that the performance of this public sector regarding drinking water production and sales is on average medium and that water waste is unfortunately substantial. It indicates also that the time effect is very weak and that the most appropriate inefficiency distribution is that of the extended half-normal in comparison with the classical half-normal. Open topics include using other distributions such as the extended exponential distribution of Hafner et al. (2018), and other time functions. This is left for future research.

Table 7 Technical efficiency of the normal-HN and normal-extended HN models for water data

		Normal-HN model			Normal-Extended HN model		
No	DMU	2017	2018	2019	2017	2018	2019
1	Agadir-Ida-Ou-Tanane	0.707	0.730	0.752	0.569	0.573	0.576
2	Al Haouz	0.178	0.210	0.244	0.340	0.344	0.349
3	Al Hoceima	0.674	0.700	0.724	0.776	0.778	0.780
4	Assa-Zag	0.109	0.135	0.164	0.342	0.346	0.350
5	Azilal	0.461	0.496	0.530	0.647	0.650	0.654
6	Beni Mellal	0.948	0.953	0.957	0.749	0.751	0.754
7	Berkane	0.885	0.895	0.904	0.967	0.968	0.968
8	Boujdour	0.160	0.191	0.224	0.351	0.356	0.360
9	Boulemane	0.245	0.280	0.317	0.424	0.428	0.432
10	Chefchaouen	0.283	0.319	0.356	0.446	0.450	0.454
11	Chichaoua	0.227	0.261	0.297	0.389	0.393	0.397
12	Chtouka-Ait Baha	0.274	0.310	0.347	0.420	0.424	0.428
13	El Hajeb	0.908	0.916	0.924	0.643	0.646	0.650
14	El Kelaa des Sraghna	0.453	0.488	0.523	0.535	0.539	0.543
15	Errachidia	0.628	0.656	0.683	0.662	0.665	0.668
...	...	...	...	...	...	...	...
30	Ouarzazate	0.424	0.460	0.495	0.563	0.567	0.570
31	Oued Ed-Dahab	0.433	0.469	0.504	0.612	0.615	0.619
32	Oujda-Angad	0.352	0.389	0.426	0.415	0.419	0.423
33	Safi	0.606	0.636	0.664	0.556	0.560	0.563
34	Sefrou	0.426	0.462	0.498	0.492	0.496	0.500
35	Tetouan	0.983	0.984	0.986	1.000	1.000	1.000
36	Tanger-Assilah	0.621	0.649	0.676	0.351	0.355	0.359
37	Taounate	0.589	0.619	0.648	0.801	0.803	0.805
38	Taroudannt	0.909	0.917	0.924	0.763	0.765	0.767
39	Tata	0.172	0.204	0.238	0.366	0.370	0.374
40	Taza	0.517	0.550	0.583	0.629	0.632	0.635
41	Tiznit	0.503	0.537	0.570	0.708	0.711	0.713
42	Zagora	0.280	0.316	0.353	0.471	0.475	0.479
-	Mean	0.488	0.516	0.544	0.571	0.574	0.578

Appendix A Proof of Theorem 1: Error term density when the inefficiency component is an extended half-normal

Proof. Knowing that the inefficiency term u_i is an extended half-normal, that $0 < u_i < B$ and $\epsilon_{it} = v_{it} - \eta(t) u_i$ and given that $f_{\epsilon,u}$ is the (ϵ_i, u_i) joint density, f_1 is the v_{it} density function and f_γ^- is the u_i density function, we have

$$
\begin{aligned}
g^-(\epsilon_i) &= \int_0^B f_{\epsilon,u}(\epsilon_i, u_i)\, du_i \\
&= \int_0^B f_{\epsilon,u}(\epsilon_{i1}, \ldots, \epsilon_{it}, \ldots, \epsilon_{iT}, u_i)\, du_i \\
&= \int_0^B \prod_t f_1(\epsilon_{it} + \eta(t).u_i)).f_\gamma^-(u_i)\, du_i \\
&= \int_0^B \frac{1}{(2\pi)^{T/2}\sigma_v^T} \exp\left\{-\frac{1}{2}\left[\frac{\sum_t(\epsilon_{it}+\eta(t).u_i)^2}{\sigma_v^2}\right]\right\} \\
&\qquad .\left[\Phi(a_0) - \Phi(0)\right]^{(-1)} \frac{1}{|\gamma|}\phi\left(a_0 - \frac{u_i}{|\gamma|}\right) du_i \\
&= \int_0^B \frac{1}{(2\pi)^{T/2}\sigma_v^T} \exp\left\{-\frac{1}{2}\left[\frac{\sum_t \epsilon_{it}^2 + \sum_t \eta^2(t)\, u_i^2 + 2\sum_t \epsilon_{it}\eta(t).u_i}{\sigma_v^2}\right]\right\} \\
&\qquad .\left[\Phi(a_0) - \Phi(0)\right]^{(-1)} \frac{1}{|\gamma|}\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(a_0 - \frac{u_i}{|\gamma|}\right)^2\right\} du_i \\
&= \frac{1}{(2\pi)^{T/2}\sigma_v^T\left[\Phi(a_0)-\Phi(0)\right]|\gamma|\sqrt{2\pi}} \\
&\qquad .\int_0^B \exp\left\{-\frac{1}{2}\left[\frac{\sum_t \epsilon_{it}^2 + \sum_t \eta^2(t)\, u_i^2 + 2\sum_t \epsilon_{it}\eta(t).u_i}{\sigma_v^2} + \frac{(a_0|\gamma| - u_i)^2}{\gamma^2}\right]\right\} du_i \\
&= \frac{1}{(2\pi)^{T/2}\sigma_v^T\left[\Phi(a_0)-\Phi(0)\right]|\gamma|\sqrt{2\pi}} \\
&\qquad .\int_0^B \exp\left\{-\frac{1}{2}\left[\frac{\sum_t \epsilon_{it}^2 + \sum_t \eta^2(t)\, u_i^2 + 2\sum_t \epsilon_{it}\eta(t).u_i}{\sigma_v^2} + \frac{B^2 + u_i^2 - 2Bu_i}{\gamma^2}\right]\right\} du_i \\
&= \frac{1}{(2\pi)^{T/2}\sigma_v^T\left[\Phi(a_0)-\Phi(0)\right]|\gamma|\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{\gamma^2\sum_t \epsilon_{it}^2 + \sigma_v^2 B^2}{\sigma_v^2\gamma^2}\right]\right\} \\
&\qquad .\int_0^B \exp\left\{-\frac{1}{2}\left[\frac{\gamma^2\sum_t \eta^2(t)\, u_i^2 + 2\gamma^2\sum_t \epsilon_{it}\eta(t).u_i + \sigma_v^2 u_i^2 - 2B\sigma_v^2 u_i}{\sigma_v^2\gamma^2}\right]\right\} du_i \\
&= \frac{\sigma_*}{(2\pi)^{T/2}\sigma_v^T\left[\Phi(a_0)-\Phi(0)\right]|\gamma|} \exp\left\{-\frac{1}{2}\left[\frac{\gamma^2\sum_t \epsilon_{it}^2 + \sigma_v^2 B^2}{\sigma_v^2\gamma^2}\right]\right\}
\end{aligned}
$$

$$\cdot \int_0^B \frac{1}{\sigma_* \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left[\frac{\gamma^2 \sum_t \eta^2(t) + \sigma_v^2}{\sigma_v^2 \gamma^2} u_i^2 - 2 \frac{-\gamma^2 \sum_t \epsilon_{it} \eta(t) + B\sigma_v^2}{\sigma_v^2 \gamma^2} u_i \right] \right\} du_i$$

$$= \frac{\sigma_*}{(2\pi)^{T/2} \sigma_v^T [\Phi(a_0) - \Phi(0)] |\gamma|} \exp\left\{ -\frac{1}{2} \left[\frac{\gamma^2 \sum_t \epsilon_{it}^2 + \sigma_v^2 B^2}{\sigma_v^2 \gamma^2} \right] \right\}$$

$$\cdot \int_0^B \frac{1}{\sigma_* \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left[\frac{1}{\sigma_*^2} u_i^2 - 2 \frac{\mu_{*i}}{\sigma_*^2} u_i + \frac{\mu_{*i}^2}{\sigma_*^2} - \frac{\mu_{*i}^2}{\sigma_*^2} \right] \right\} du_i$$

$$= \frac{\sigma_*}{(2\pi)^{T/2} \sigma_v^T [\Phi(a_0) - \Phi(0)] |\gamma|} \exp\left\{ -\frac{1}{2} \left[\frac{\gamma^2 \sum_t \epsilon_{it}^2 + \sigma_v^2 B^2}{\sigma_v^2 \gamma^2} - \frac{\mu_{*i}^2}{\sigma_*^2} \right] \right\}$$

$$\cdot \int_0^B \frac{1}{\sigma_* \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left[\frac{1}{\sigma_*^2} (u_i - \mu_{*i})^2 \right] \right\} du_i$$

$$= \frac{\sigma_* \exp\left\{ -\frac{1}{2} a_{*i} \right\}}{(2\pi)^{T/2} \sigma_v^T [\Phi(a_0) - \Phi(0)] |\gamma|} \left[\Phi\left(\frac{B - \mu_{*i}}{\sigma_*} \right) - \Phi\left(-\frac{\mu_{*i}}{\sigma_*} \right) \right]$$

□

Appendix B Proof of Theorem 2: Technical efficiency when the inefficiency component is an extended half-normal

Proof. If we denote f_2 the density of $(u_i \mid \epsilon_i)$, we have

$$\begin{aligned}
TE_{it} &= E\Big(\exp\{-u_{it}\} \mid \epsilon_i \Big) \\
&= E\Big(\exp\{-\eta(t) u_i\} \mid (\epsilon_{i1}, \ldots, \epsilon_{it}, \ldots, \epsilon_{iT}) \Big) \\
&= \int_0^B \exp\{-\eta(t) u_i\} f_2(u_i \mid (\epsilon_{i1}, \ldots, \epsilon_{it}, \ldots, \epsilon_{iT})) du_i \\
&= \int_0^B \exp\{-\eta(t) u_i\} \frac{f_{\epsilon,u}(\epsilon_{i1}, \ldots, \epsilon_{it}, \ldots, \epsilon_{iT}, u_i)}{g^-(\epsilon_i)} du_i \\
&= \frac{1}{g^-(\epsilon_i)} \int_0^B \exp\{-\eta(t) u_i\} f_{\epsilon,u}(\epsilon_{i1}, \ldots, \epsilon_{it}, \ldots, \epsilon_{iT}, u_i) du_i \\
&= \frac{\sigma_* \exp\left\{ -\frac{1}{2} a_{*it} \right\}}{(2\pi)^{T/2} \sigma_v^T [\Phi(a_0) - \Phi(0)] |\gamma|} \left[\Phi\left(\frac{B - \mu_{*it}}{\sigma_*} \right) - \Phi\left(-\frac{\mu_{*it}}{\sigma_*} \right) \right] \Big/ g^-(\epsilon_i)
\end{aligned}$$

□

Declarations

Funding No funding was received for conducting this study.
Conflict of interest/Competing interests Authors have no competing interests to declare that are relevant to the content of this article.
Ethics approval The authors declare that this research doesn't involve humans and/or animals.
Consent to participate Not applicable because no individual consent to participate is necessary for our study.
Consent for publication Not applicable because no individual consent to participate is necessary for our study.
Availability of data and materials There are two types of data supporting the results of our work. Our simulated data were generated as described in section 3 of this manuscript. As for our real data, we collected them ourselves from Moroccan statistical yearbooks. So, both of them are available, under request, from the corresponding author.
Code availability The code is available, under request, from the corresponding author.
Authors' contributions The two authors contributed equally to this work.
Financial interests/Non-financial interests The authors have no relevant financial or non-financial interests to disclose.

References

Aigner, D.J., Lovell, C.A.K., Schmidt, P.J. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*, 21–37, https://doi.org/10.1016/0304-4076(77)90052-5

Almanidis, P., Qian, J., Sickles, R.C. (2014). Stochastic frontier models with bounded inefficiency. R.C. Sickles & W.e. Horrace (Eds.), *Econometric Methods and Applications* (pp. 47–81). New York: Springer Science & Business Media.

Battese, G.E., & Coelli, T. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, *38*, 387–399, https://doi.org/10.1016/0304-4076(88)90053-X

Battese, G.E., & Coelli, T. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in india. *The Journal of Productivity Analysis*, *3*, 153–169, https://doi.org/10.1007/BF00158774

Coelli, T.J. (1996). *A guide to frontier version 4.1: A computer program for stochastic frontier production and cost function estimation.* (Working Paper No. 7/96, Centre for Efficiency and Productivity Analysis (CEPA), University of New England, Australia. Preprint at http://www.uq.edu.au/economics/cepa/frontier.php)

Coelli, T.J., & Henningsen, A. (2013). *frontier: Stochastic frontier analysis. r package version 1.1.* (Preprint at http://CRAN.R-Project.org/package=frontier)

Daniel, B.C., Hafner, C.M., Manner, H., Simar, L. (2011). *Asymmetries in business cycles and the role of oil production.* ((ISBA Discussion Paper; 2011/32), 23 p. Preprint at http://hdl.handle.net/2078.1/92536)

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, *19*, 273–292, https://doi.org/10.2307/1906814

Deprins, D., & Simar, L. (1989). Estimating technical inefficiencies with corrections for environmental conditions with an application to railway companies. *Annals of Public and Cooperative Economics*, *60*, 81–102, https://doi.org/10.1111/j.1467-8292.1989.tb02010.x

El Mehdi, R., & Hafner, C.M. (2014a). Local government efficiency: The case of moroccan municipalities. *African Development Review*, *26*, 88–101, https://doi.org/10.1111/1467-8268.12066

El Mehdi, R., & Hafner, C.M. (2014b). Inference in stochastic frontier analysis with dependent error terms. *Mathematics and Computers in Simulation (MATCOM)*, *102*, 104–116, https://doi.org/10.1016/j.matcom.2013.09.008

El Mehdi, R., & Hafner, C.M. (2021). Panel stochastic frontier analysis with dependent error terms. *International Econometric Review (IER)*, *13*, 24–40, https://doi.org/10.33818/ier.1033722

Farrell, M.J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, *120*, 253–290, https://doi.org/10.2307/2343100

Greene, W.H. (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics*, *46*, 141–163, https://doi.org/10.1016/0304-4076(90)90052-U

Hafner, C.M., Manner, H., Simar, L. (2018). The "wrong skewness"problem in stochastic frontier models: A new approach. *Econometric Reviews*, *37*, 380–400, https://doi.org/10.1080/07474938.2016.1140284

Hajargasht, G. (2015). Stochastic frontiers with a rayleigh distribution. *Journal of Productivity Analysis*, *44*, 199–208, https://doi.org/10.1007/s11123-014-0417-8

Hollingsworth, B., & Peacock, S. (2008). *Efficiency measurement in health and health care, first ed.* New York: Routledge.

Koopmans, T.C. (1951). An analysis of production as an efficient combination of activities. T.C. Koopmans (Ed.), *Activity Analysis of Production and Allocation* (pp. 33–97). New York: Wiley.

Kumbhakar, S.C., & Knox Lovell, C.A. (2000). *Stochastic Frontier Analysis, first ed.* United Kingdom: Cambridge University Press.

Kumbhakar, S.C., Parameter, C.F., Zelenyuk, V. (2018). *Stochastic frontier analysis: Foundations and advances.* (CEPA Working Paper Series, No. WP02/2018, Centre for Efficiency and Productivity Analysis. Preprint at https://ideas.repec.org/p/qld/uqcepa/123.html)

Lee, Y.H., & Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical inefficiency. H. Fried, C. Lovell, & S. Schmidt (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications* (pp. 237–255). New York: Oxford University Press.

Meeusen, W., & Van Den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *Econometric Reviews*, *18*, 435–444, https://doi.org/10.2307/2525757

Migon, H.S., & Medici, E. (2001). *A bayesian approach to stochastic production frontier.* (Technical Report n. 105, DME-IM/UFRJ)

Simar, L., & Wilson, P.W. (2010). Inferences from cross-sectional, stochastic frontier models. *Econometric Reviews*, *29*, 62–98, https://doi.org/10.1080/07474930903324523

Stevenson, R.E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, *13*, 57–66, https://doi.org/10.1016/0304-4076(80)90042-1