Thai Journal of Mathematics : (2017) 79–90
Special Issue on Entropy in Econometrics

http://thaijmath.in.cmu.ac.th
ISSN 1686-0209

Coffee Stochastic Frontier Model with Maximum Entropy

Chatchai Khiewngamdee[†,1], **Songsak Sriboonchitta**[‡,§], **Somsak Chanaim**[‡,§], **and Chongkolnee Rungruang**[¶]

[†]Department of Agricultural Economy and Development, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail : getliecon@gmail.com (C. Khiewngamdee)

[‡]Faculty of Economics, Chiang Mai University
Chiang Mai 50200, Thailand

[§]Center of Excellence in Econometrics, Chiang Mai University,
Chiang Mai 50200, Thailand
e-mail : songsakcon@gmail.com (S. Sriboonchitta)
somsak_ch@cmu.ac.th (S. Chanaim)

[¶]Faculty of Commerce and Management, Prince of Songkla University,
Trang Campus, Trang 92000, Thailand
e-mail : chongkolnee.r@psu.ac.th (C. Rungruang)

Abstract : In the classical maximum likelihood estimation of stochastic frontier model, a strong assumption on two error components, namely symmetric noise (V_j) and the non-negative inefficiency (U_j), are required. This could lead to non-reliable and erroneous interpretations when we misspecify the probability distribution of the error components. To overcome this problem, we apply the generalized maximum entropy (GME) approach to estimate the stochastic frontier model which allows us to avoid the need for making an ad hoc assumption about the distribution of the noise and inefficiency components. In this study, we investigate the technical efficiency of coffee production using generalized maximum entropy. The results show that the technical efficiency scores obtained from GME estimator are much smaller than ones from the maximum likelihood method, even though the estimated parameters are quite indifferent. In addition, we also find that the

[1]Corresponding author.

wider support value of the inefficiency component, the lower score of the estimated technical efficiency.

Keywords : technical efficiency; coffee output; inefficiency; generalized maximum entropy.
2010 Mathematics Subject Classification : 62P20; 91B84.

1 Introduction

Since Aigner et al. [1] and Meeusen and van den Broeck [2] introduced the stochastic frontier procedure to estimate the producers' efficiency and productivity using parametric econometric techniques, the stochastic frontier approach has been applied in several areas of study and has grown dramatically. Moreover, the basic idea of stochastic frontier has been extended in many directions, for instance, while Aigner et al. [1] and Meeusen and van den Broeck [2] assume the distribution of the efficiency error to be a half-Normal and an exponential, respectively, later studies include the truncated Normal and the gamma distributions.

In the past few decades, several papers have developed frameworks for applying in stochastic frontier model in order to relax the model's assumptions and improve the accuracy of estimated parameters of the stochastic frontier model. For example, Griffin and Steel [3] proposed a semiparametric Bayesian framework for stochastic frontiers analysis and efficiency measurement in order to deal with practically relevant sample sizes. Kumbhakar and Tsionas [4] employed a simulated maximum likelihood approach and a stochastic frontier model to estimate the input-oriented technical inefficiency and compare with the results from output-oriented model. Furthermore, Kumbhakar et al. [5] proposed a new approach based on local maximum likelihood technique to handle nonparametric stochastic frontier models. The methods presented to be useful for investigating the production efficiency.

In addition, Tsionas [6] investigated several kinds of stochastic frontier models, namely, input-oriented stochastic frontier models, output-oriented stochastic frontier models, two-tiered stochastic frontier models, latent class models with gamma distributed one-sided error term, and models with the distribution of the two-sided error component as stable Paretian and the one-sided error as gamma. Besides, Chen et al. [7] applied a Bayesian stochastic frontier model to analyze the technical efficiency of Chinese fossil-fuel electricity generation companies and Amsler et al. [8] studied a stochastic frontier model that contains environmental variables such as statistical noise, potentially endogenous regressors, and technical inefficiency.

Although several techniques have been applied in stochastic frontier model, there is still no conclusion for the dominant model since stochastic frontiers based on either maximum likelihood estimation or Bayesian method require ad-hoc assumptions on the distribution of the inefficiency component. Generalized maximum entropy approach, on the contrary, eases the restriction by avoiding the need for making assumptions on the inefficiency component. Generalized maximum entropy method was first introduced by Golan [9], since then it has proven to be robust under ill-posed and ill-conditioned problems and particularly suited in this context given the very large number of parameters to be estimated. Additionally, generalized maximum entropy provides a potential alternative frontier estimation approach that combines the strengths of both stochastic frontier analysis and data envelopment analysis as studied by Campbell et al. [10] who developed an efficiency estimation approach by utilizing generalized maximum entropy and compared with stochastic frontier analysis and data envelopment analysis approaches. Another example is from Tonini and Pede [11]. They developed a stochastic frontier model accounting for spatial dependency using generalized maximum entropy estimation approach. This method helps reduce multicollinearity issues among the exogenous variables since the generalized maximum entropy estimator allows the inclusion of prior information on parameters by defining support bounds and adding specific consistency constraints. This paper aims to investigate the technical efficiency of coffee production using generalized maximum entropy approach. We also compare the results with those from the maximum likelihood estimator.

The structure of this paper is organized as follows. Section 2 describes our methodology. The empirical results are presented in Section 3, and the final section contains conclusion.

2 Methodology

2.1 A Stochastic Frontier Model

The general form of the stochastic frontier model (SFM) can be written as

$$Y_j = f(X'_{jk}\beta) \cdot TE_j,\ j = 1, \cdots, N \tag{2.1}$$

or

$$\ln Y_j = X'_{jk}\beta + \varepsilon_j, \tag{2.2}$$

$$\varepsilon_j = V_j - U_j, \tag{2.3}$$

where Y_j represents the output variable and X'_{jk} denotes $N \times K$ matrix of K different input quantities. The term β is $(N \times K)$ matrix of estimated parameters of the input variables. The function $f(\cdot)$ is the functional form of the stochastic frontier model which is imposed to be the Cobb-Douglas production function. The term TE is technical efficiency and ε_j denotes the composite error term which consists of the noise, V_j, and the inefficiency, U_j. In general, the distribution assumptions in Eq.(2.3) are normal for V_j and half-normal for the random term U_j. In this study, since the entropy estimation is proposed to estimate the model, thus we ignore any distribution assumptions for V_j and U_j. Nevertheless, a restriction on U_j is given as $U_j \in [0, a], a < \infty$. These two error components are distributed identically and independently from each other and the regressor.

In the context of the stochastic frontier model, the technical efficiency (TE) is defined as the ratio of the observed output to the corresponding frontier output, conditional on the levels of inputs. Therefore, the technical efficiency or TE is given by

$$TE_j = \frac{\exp\{X_{jk} + V_j - U_j\}}{\exp\{X_{jk} + V_j\}} = \exp\{-U_j\} \tag{2.4}$$

2.2 Estimating Technique: Generalized Maximum Entropy (GME) Approach

We propose the use of maximum entropy estimator to estimate our unknown parameters in Eq.(2.2). The advantage and the properties of this estimator are provided at length in Golan [9] and Mittelhammer, Judge, and Miller [12], and more recently in Tonini and Pede [11].

In summary, the main advantages of the Generalized maximum entropy estimator are as follow: First, it efficiently takes into account all the information contained in each data point. Second, it is less affected by outlier since we can give the probability weight between signal and noise in the objective function. Third, it is a robust estimator as it does not require any assumptions regarding the error terms. Finally, the Generalized maximum entropy estimator does not require strong behavioral assumptions on the underlying data generating process.

The maximum entropy concept consists of inferring the probability distribution that maximizes information entropy given a set of various constraints. Let p_k be a proper probability mass function on a finite set of β. Shannon [13] developed his information criteria and proposed a classical entropy, that is:

$$H(p) = -\sum_{k=1}^{K} p_k \log p_k \tag{2.5}$$

where $\sum_{k=1}^{K} p_k = 1$. The entropy measures the uncertainty of a distribution and reaches a maximum when p_k is uniformly distributed (Wu [14]). With the stochastic frontier model, the objective entropy function can be written as

$$\begin{aligned} H(p,v,w) &= H(p) + H(v) + H(w) \\ &= -\sum_{k=1}^{K}\sum_{m=1}^{M} p_{km}\log p_{km} - \sum_{i=1}^{N}\sum_{q=1}^{Q} v_{iq}\log v_{iq} - \sum_{i=1}^{N}\sum_{u=1}^{U} w_{iu}\log w_{w_{iu}} \end{aligned} \tag{2.6}$$

where p_{km}, v_{iu}, and w_{iq} are the probability of estimated parameters β_k, noise V_i, and inefficiency U_i. The constraint of this objective function is given by

$$Y_j = \sum_{k=1}^{K}\sum_{m=1}^{M} p_{km} z_{km} X_{jk} + \sum_{m=1}^{M} v_{jm} g_{jm} - \sum_{m=1}^{M} w_{jm} h_{jm}, \ j = 1, \cdots, N \tag{2.7}$$

and

$$\sum_{m=1}^{M} p_{km} = 1, \sum_{m=1}^{M} v_{km} = 1, \ \sum_{m=1}^{M} w_{km} = 1. \tag{2.8}$$

The support values z_{km}, g_{tm} , and h_{tm}, are needed to estimate the unknown parameters, noise V_j, and inefficiency U_j in the stochastic frontier model. Therefore,

$$\beta_k = \sum_{m=1}^{M} p_{km} z_{km} \tag{2.9}$$

$$V_j = \sum_{m=1}^{M} v_{tm} g_{tm} \tag{2.10}$$

$$U_j = \sum_{m=1}^{M} w_{tm} h_{tm} \tag{2.11}$$

Suppose that stochastic frontier model has one regressor $k = 1$; thus, the optimization problems from Eq. (2.7) and (2.9) are solved by the Lagrangian method, which takes the form

$$\begin{aligned} L =& H(p, v, w) + \lambda_1' \left(Y_j - \sum_{m=1}^{M} p_{km} z_{km} X_j^k - \sum_{m=1}^{M} v_{tm} g_{tm} + \sum_{m=1}^{M} w_{tm} h_{tm} \right) \\ &+ \lambda_2' \left(1 - \sum_{m=1}^{M} p_m \right) + \lambda_3' \left(1 - \sum_{m=1}^{M} v_m \right) + \lambda_4' \left(1 - \sum_{m=1}^{M} w_m \right) \end{aligned} \tag{2.12}$$

where $\lambda_i', \ i = 1, 2, 3, 4$ are the vectors of Lagrangian multiplier. The first-order conditions are:

$$\frac{\partial L}{\partial p_m} = -\log(p_m) - \sum_{m=1}^{M} \lambda_{1m} z_m X_t - \lambda_{2t} = 0 \tag{2.13}$$

$$\frac{\partial L}{\partial v_{tm}} = -\log(v_{tm}) - \sum_{m=1}^{M} \lambda_{1m} g_{tm} - \lambda_{3t} = 0 \tag{2.14}$$

$$\frac{\partial L}{\partial w_{tm}} = -\log(w_{tm}) - \sum_{m=1}^{M} \lambda_{1m} h_{tm} - \lambda_{4t} = 0 \tag{2.15}$$

$$\frac{\partial L}{\partial \lambda_1} = Y_t - \sum_{m=1}^{M} p_{km} z_{km} X_t^k - \sum_{m=1}^{M} v_{tm} g_{tm} + \sum_{m=1}^{M} w_{tm} h_{tm} = 0 \tag{2.16}$$

$$\frac{\partial L}{\partial \lambda_2} = 1 - \sum_{m=1}^{M} p_m \tag{2.17}$$

$$\frac{\partial L}{\partial \lambda_3} = 1 - \sum_{m=1}^{M} v_m \tag{2.18}$$

$$\frac{\partial L}{\partial \lambda_4} = 1 - \sum_{m=1}^{M} w_m \tag{2.19}$$

Solving the optimization of this problem, we yield the optimal and unique solution as in the following:

$$p_m = \frac{\exp[-z_m \sum_t \lambda_1 X_t]}{\sum_{m=1}^{M} \exp[-z_m \sum_t \lambda_1 X_t]} \tag{2.20}$$

$$v_{tm} = \frac{\exp[-\lambda_1 g_{tm}]}{\sum_{m=1}^{M} \exp[-\lambda_1 g_{tm}]} \tag{2.21}$$

$$w_{tm} = \frac{\exp[-\lambda_1 h_{tm}]}{\sum_{m=1}^{M} \exp[-\lambda_1 h_{tm}]} \tag{2.22}$$

In this study, the supports for z_{km} are given as

$$[\widehat{\beta_k} - 3a_k,\ \widehat{\beta_k} - 1.5a_k,\ \widehat{\beta_k},\ \widehat{\beta_k} + 1.5a_k,\ \widehat{\beta_k} + 3a_k], \tag{2.23}$$

the supports for h_{tm} as

$$[-3a_v,\ -1.5a_v,\ 0,\ 1.5a_v,\ 3a_v], \tag{2.24}$$

and the supports for g_{tm} as

$$[0,\ .75a_u,\ 1.5a_u,\ 2.25a_u,\ 3a_u]. \tag{2.25}$$

We estimate $\widehat{\beta}_k, a_k, a_v$ and a_u by using package frontier in program R [15] for estimating the parameters in stochastic frontier model via maximum likelihood approach, where $\widehat{\beta}_k$ is the parameter estimation of β_k. a_k is the standard error of $\widehat{\beta}_k$, a_v is the σ_v of $v \sim N(0, \sigma_v^2)$ and a_u is the σ_u of $u \sim HN(0, \sigma_v^2)$. We can find σ_v and σ_u from the relationship that

$$\gamma = \frac{\sigma_u}{\sigma_v},$$
$$\sigma^2 = \sigma_v^2 + \sigma_u^2.$$

However, if the values of a_k, a_v, and a_u are too small, we will consider the values of a_k, a_v, a_u such that a_k is greater than standard error of $\widehat{\beta}_k$, $a_v > \sigma_v$ and $a_u > \sigma_u$.

By maximum entropy we can estimate standard error (σ_{β_k})of each $\beta_k,\ k = 0, 1, 2, \cdots, 5$ by

$$\sigma_{\beta_k} = \sqrt{\sum_{m=1}^{5} (z_{km} - \beta_k)^2 p_{km}}.$$

3 Empirical Results

In this section, we investigate the technical efficiency of coffee output by using stochastic frontier model. Then, we compare the results of technical efficiency estimated from the maximum likelihood method and generalized maximum entropy (GME) approach. The data used in this study were collected from 376 coffee farmers in Chiang Mai, Thailand, by Wiboonpongse et al. [16]. The descriptive statistics of this data set is presented in Table 1.

Table 1: Descriptive statistics

	$\ln(Y)$	$\ln(L)$	$\ln(F)$
mean	10.5535	1.9375	6.1697
median	10.5699	1.9081	6.2201
std	0.6999	0.9526	0.666
min	8.4124	−0.5108	4.0518
max	12.4868	4.8097	8.0392
skewness	0.1483	0.1482	−0.0277
kurtosis	2.8955	2.7825	2.8399
obs.		376	

Following Wiboonpongse et al. [16], the translog form for coffee production model can be written as

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(L_i) + \beta_2 \ln(F_i) + \beta_3 \frac{(\ln(L_i))^2}{2} + \beta_4 \frac{(\ln(F_i))^2}{2} + \beta_5(\ln(F_i) \times \ln(L_i)) + V_i - U_i$$

where Y_i represents coffee output of farmer i, which depends on fertilizer F_i and labor L_i, respectively. V_i and U_i are statistical noise and inefficiency terms, which are assumed to be normally and half-normally distributed, respectively.

Table 2 presents the estimated results under the maximum likelihood method. The overall results show that all of the parameters are statistically significant. We find that an increase in the number of labors creates a positive impact on the coffee production and leads to higher output, whereas increasing use of fertilizer just creates a negative effect on the output. However, the parameter γ, which is equal to 1.24×10^{-5}, is close to boundary, indicating that there is no inefficiency, therefore the estimated technical efficiency (TE) score is equal to 1. This is very difficult to, or perhaps never, happen in reality. Thus, as an alternative method, we perform this experiment again using the generalized maximum entropy estimator. The results are discussed as follows.

Table 2: Estimated results from Maximum likelihood

	Estimate	**Std. Error**	**z value**	$Pr(>\|z\|)$
β_0	8.8033	0.9154	9.6170	2.2×10^{-16} $***$
β_1	0.74593	0.1839	4.0570	4.972×10^{-5} $***$
β_2	-0.68577	0.3399	-2.0176	0.0436$*$
β_3	0.0088	0.0030	2.9111	0.0036 $**$
β_4	0.3179	0.0064	4.9780	6.425×10^{-7} $***$
β_5	-0.1443	0.0036	-4.0036	6.238×10^{-5} $***$
σ^2	0.0032	0.0002	13.5876	$< 2.2 \times 10^{-16}$ $***$
γ	1.2478×10^{-5}	0.0006	0.0023	0.9982
Log likelihood	111.4526			
RMSE	0.1801			

Table 3: Estimated results from Maximum Entropy

	model 1		**model 2**		**model 3**	
	Est. coef.	**std**	**Est. coef.**	**std**	**Est. coef.**	**std**
β_0	11.5641	0.6434	9.3827	2.0695	9.2848	2.0856
β_1	1.3459	0.0040	0.8309	0.4187	0.7670	0.4239
β_2	-1.4689	0.5899	-0.6965	0.8485	-0.5791	0.8442
β_3	0.2070	0.1898	0.1031	0.2118	0.0939	0.2121
β_4	0.4350	0.1905	0.3187	0.2121	0.3007	0.2117
β_5	-0.2527	0.1936	-0.1594	0.2118	-0.1485	0.2121
Entropy	1,148.42		1,195.75		1,200.85	
RMSE	0.2184		0.1808		0.1803	
	model 4		**model 5**			
	Est. coef.	**std**	**Est. coef.**	**std**		
β_0	9.3570	2.0740	9.4458	2.0575		
β_1	0.7619	0.4241	0.7640	0.4240		
β_2	-0.5250	0.8386	-0.4801	0.8322		
β_3	0.0925	0.2121	0.0918	0.2121		
β_4	0.2941	0.2113	0.2897	0.2109		
β_5	-0.1483	0.2121	-0.1494	0.2121		
Entropy	1,202.48		1,203.19			
RMSE	0.1804		0.1808			

As maximum entropy allows for containing prior information on parameters, this study makes use of this approach and modifies the support values of the inefficiency component. Here we have five different support spaces for model 1 to model 5, and the estimated results of these models are presented in Table 3. From Eq. (2.23), we define $a_k = (a_0, a_1, \cdots, a_5) = (1, .2, .4, .2, .1, .1)$ for the model 1 to model 5 according to the standard error of β_k obtained from maximum likelihood estimator. In addition, $a_v = a_u = 0.18$ for model 1, $a_v = a_u = 0.36$ for model 2, $a_v = a_u = 0.54$ for model 3, $a_v = a_u = 0.72$ for model 4, and $a_v = a_u = 0.90$ for model 5. In other words, we let model 5 have the widest range of the support values of inefficiency and model 4 hold the second widest range of the supports, followed by model 3, 2, and 1, respectively. The estimated coefficients appear to be stable to different choices of support values except for model 1.

The modified supports of the inefficiency component appear to influence the estimated technical efficiency (TE) scores, which are displayed in Figure 1 and Figure 2. We find that the highest TE score is obtained from model 1, followed by the scores obtained from model 2, 3, and 4, respectively, while the lowest TE score is measured from model 5. This indicates that the wider support values of the inefficiency component, the lower score of the estimated technical efficiency.

Additionally, the comparative performance of each model is evaluated through the root mean square error (RMSE), which is used to assess how effectively the model can explain the behavior of data set. As presented in Table 3, the lowest value of RMSE is in model 3, equals to 0.1803, meaning that this model is best explaining the empirically observed behavior over other models. Moreover, the RMSE value of model 3 is slightly different from one obtained from using the maximum likelihood estimator as well as the estimated coefficients (as shown in Table 2). The technical efficiency scores, however, present a huge difference.

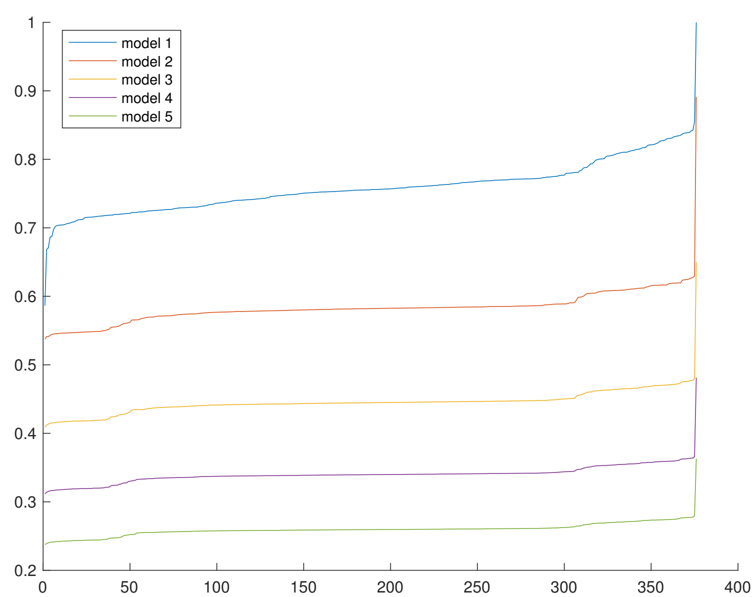

Figure 1: Technical efficiency calculation

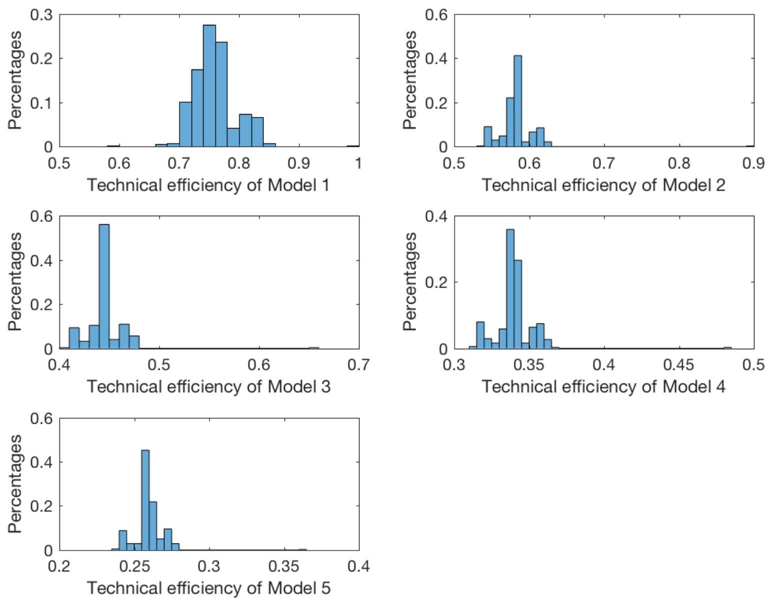

Figure 2: Technical efficiency from 5 models

4 Conclusions

The Stochastic Frontier Model (SFM) is a linear regression model in which the error consists of two components, namely the symmetric noise V_j and the non-negative inefficiency, U_j. In the classical maximum likelihood estimation for the SFM, the standard maximum likelihood requires ad-hoc assumptions on both the noise component and the inefficiency component. If the distributions are unknown and improperly specified, the results obviously are not reliable and lead to erroneous interpretations.

In this study, the generalized maximum entropy (GME) estimator is considered as an estimator for SFM in order to avoid the need for making unnecessary assumptions on both the noise and inefficiency components. Maximum entropy estimator allows us to estimate a model with a relatively large number of parameters and also include prior information on parameters by specifying the support bounds and adding some useful constraints. The advantage of this prior information is that it can help decrease multicollinearity among the exogenous variables and restrict the estimated parameters to be within the bounds. Hence, in this study, we employ an entropy based approach for the estimation of the SFM model on a different range of prior bound information.

We cast the SFM problem into the generalized maximum entropy framework.

The estimator allows us to express the SFM on a distribution-free basis, making it easy to work in any patterns of real data analysis. In this paper, we investigate the technical efficiency of coffee output by using stochastic frontier model. Furthermore, we compare the results of technical efficiency estimated from the maximum likelihood method and generalized maximum entropy (GME) approach. With the real data application, we found that the estimated coefficients and RMSE obtained from the primal GME estimator are quite similar to those obtained from the classical maximum likelihood estimator. Nevertheless, the technical efficiency scores from GME approach are much lower than ones obtained from maximum likelihood method. Finally, we also compare the performance of the estimator with different choices of support bound. The results reveal that the wider the support values of the inefficiency component, the lower the score of the estimated technical efficiency.

References

[1] D. Aigner, C. K. Lovell, P. Schmidt, Formulation and estimation of stochastic frontier production function models, Journal of Econometrics 6 (1) (1977) 21-37.

[2] W. Meeusen, J. Van Den Broeck, Efficiency estimation from Cobb-Douglas production functions with composed error, International economic review (1977) 435-444.

[3] J. E. Griffin, M. F. Steel, Semiparametric Bayesian inference for stochastic frontier models. Journal of econometrics 123 (1) (2004) 121-152.

[4] S. C. Kumbhakar, E. G. Tsionas, Estimation of stochastic frontier production functions with input-oriented technical efficiency, Journal of Econometrics 133 (1) (2006) 71-96.

[5] S. C. Kumbhakar, B. U. Park, L. Simar, E. G. Tsionas, Nonparametric stochastic frontiers: a local maximum likelihood approach, Journal of Econometrics 137 (1) (2007) 1-27.

[6] E. G. Tsionas, Maximum likelihood estimation of stochastic frontier models by the Fourier transform, Journal of Econometrics 170 (1) (2012) 234-248.

[7] Z. Chen, C. P. Barros, M. R. Borges, A Bayesian stochastic frontier analysis of Chinese fossil-fuel electricity generation companies, Energy Economics 48 (2015) 136-144.

[8] C. Amsler, A. Prokhorov, P. Schmidt, Endogenous environmental variables in stochastic frontier models, Journal of Econometrics (2017).

[9] A. Golan, Information and entropy econometrics-editor's view, Journal of Econometrics 107 (1) (2002) 1-15.

[10] R. Campbell, K. Rogers, J. Rezek, Efficient frontier estimation: a maximum entropy approach, Journal of Productivity Analysis 30 (3) (2008) 213-221.

[11] A. Tonini, V. Pede, A generalized maximum entropy stochastic frontier measuring productivity accounting for spatial dependency, Entropy 13 (11) (2011) 1916-1927.

[12] R. C. Mittelhammer, G. G. Judge, D. J. Miller, Econometric Foundations Pack with CD-ROM (Vol. 1), Cambridge University Press, 2000.

[13] C. E. Shannon, A mathematical theory of communication, ACM SIGMOBILE Mobile Computing and Communications Review 5 (1) 3-55 (2001).

[14] X. Wu, A weighted generalized maximum entropy estimator with a data-driven weight, Entropy 11 (4) (2009) 917-930.

[15] T. Coelli, A. Henningsen, frontier: Stochastic Frontier Analysis(2017), R package version 1.1-2. https://CRAN.R-Project.org/package=frontier.

[16] A. Wiboonpongse, J. Liu, S. Sriboonchitta, T. Denoeux, Modeling dependence between error components of the stochastic frontier model using copula: application to intercrop coffee production in Northern Thailand, International Journal of Approximate Reasoning 65 (2015) 34-44.

(Received 31 August 2017)
(Accepted 30 October 2017)