



HAL
open science

Modeling dependence between error components of the stochastic frontier model using copula: Application to intercrop coffee production in Northern Thailand

Aree Wiboonpongse, Jianxu Liu, Songsak Sriboonchitta, Thierry Denoeux

► To cite this version:

Aree Wiboonpongse, Jianxu Liu, Songsak Sriboonchitta, Thierry 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*, 2015, 65, pp.34-44. 10.1016/j.ijar.2015.04.001 . hal-01294274

HAL Id: hal-01294274

<https://hal.science/hal-01294274>

Submitted on 29 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modeling dependence between error components of the stochastic frontier model using copula. Application to Intercrop Coffee Production in Northern Thailand

Aree Wiboonpongse^{1,2}, Jianxu Liu², Songsak Sriboonchitta², Thierry Denoeux³

Abstract

In the standard stochastic frontier model, the two-sided error term V and the one-sided technical inefficiency error term W are assumed to be independent. In this paper, we relax this assumption by modeling the dependence between V and W using copulas. Twelve copula families are considered, and their parameters are estimated using maximum simulated likelihood. The best model is then selected using the AIC and BIC criteria. This methodology was applied to coffee production data from Northern Thailand. For these data, the best model was the one based on the rotated Gumbel copula. The main finding of this study is that the dependence between V and W is significant and cannot be ignored. In particular, the standard stochastic frontier model with independence assumption grossly overestimated the technical efficiency of coffee production. In this study, about 38% of farmers were found to have low technical efficiency and most of them had technical efficiencies between 0.2 and 0.3. These results suggest that a considerable amount of productivity is lost due to inefficiency.

Keywords: Stochastic Frontier, Copula, Technical Efficiency, Thailand.

1. Introduction

Stochastic frontier models (SFMs) have been proved very useful to measure technical efficiency of production units. The stochastic frontier production model for a cross-section of observations was independently proposed by Aigner et al. (1977) and Meeusen and Van den Broeck (1977). This is essentially a linear regression model with two independent error components: a two-sided one that captures random variation of the production frontier across firms and a one-sided one that measures inefficiency relative to the frontier. In recent decades, most studies about production, cost or profit efficiency have used the conventional stochastic frontier model (see, e.g., Rahman et al (2012,2009,2008), Sriboonchitta (2012), Wiboonpongse et al. (2012a,

¹ Faculty of Economics, Prince of Songkla University, Songkla, Thailand

² Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand

³ UMR 7253 Heudiasyc, Université de Technologie de Compiègne, CNRS, France

Corresponding author: Songsak Sriboonchitta, Fax: +6653942202

Email address: Songsakecon@gmail.com

2012b, 2008a, 2008b, 2005), Wiboonpongse and Sriboonchitta (2004), Sriboonchitta and Wiboonpongse (2006, 2005, 2004), Aigner et al. (1977), Jondrow et al. (1982), Coelli et al. (1998, 2002) Kumbhakar and Lovel (2000)). In all these studies, it is assumed that the one-sided and two-sided error terms are independent, so that the joint distribution of these two random variables can easily be computed. Based on this assumption, the parameters of the SFM can be estimated using the Corrected Ordinary Least Squares (COLS) or Maximum Likelihood (ML) methods. However, the impact of the independence assumption on technical efficiency estimation has long remained an open issue. Fortunately, this strong assumption can be relaxed by using copula to fit the joint distribution of the two random error components more appropriately.

Smith (2008) first proposed a SFM allowing for dependence between the two error components using copula functions. Copula functions can be used to capture rank correlation and tail dependence between the two error components, thus making the stochastic frontier analysis much more flexible. However, the log likelihood function in the copula-based stochastic frontier model generally does not have a closed form, which makes its maximization numerically complex. In this paper, we proposed to use the maximum simulated likelihood method, which has numerical and computational advantages over numerical integration method used by Smith (2008). Furthermore, to explore the dependence structure of the error components in the SFM, we consider several copula families including the Student-t, Clayton, Gumbel and Joe families as well as their relevant rotated versions. The model with the best fit/complexity trade-off is selected using the AIC and BIC criteria. This approach was applied to cross-sectional data of coffee production in Thailand. A comparison between technical efficiencies computed with and without the independence assumption (considering the best copula model) reveals that the standard approach grossly overestimates efficiency, which has important implication for production analysis using the SFM.

The remainder of this paper is organized as follows. Section 2 introduces the necessary background on the SFM and copula. Section 3 presents the copula-based stochastic frontier approach. Empirical results with this model applied to coffee production data are reported in Section 4. Finally, Section 5 concludes the paper.

2. Background and theory

The SFM is a regression-like model with a disturbance function that is asymmetric and distinctly non-normal. This model will first be briefly in Section 2.1. Some background on copula will then be recalled in Section 2.2. These are the two building blocks of the copula-based model introduced in Section 3.

2.1 Stochastic frontier model

The SFM aims at representing the “ideal” relation between inputs x_j and output y of a production process by a production function, seen as a theoretical ideal. The observed deviations from the production function could arise from two sources:

1. Productive inefficiency, resulting in a non-negative error term;
2. Firm-specific effects, which can enter the model with either signs.

Assuming the relation between the logarithm of production $\log(y)$ and the input vector x to be linear, we obtain the following equation:

$$\begin{aligned}\log y_i &= x_i' \beta + \varepsilon_i \\ \varepsilon_i &= v_i - w_i\end{aligned}\tag{1}$$

for $i=1, 2, \dots, N$, where N is the number of firms, y_i is the production output for firm i , x_i is a $k \times 1$ vector of inputs for firm i (including a constant component $x_{i1}=1$), β is a $k \times 1$ vector of unknown parameters, v_i is a random error, and w_i is a non-negative unobservable random variable representing production inefficiency. The two error components w_i and v_i are assumed to be independent. The technical efficiency TE_i for firm i can then be defined as the ratio between actual to theoretically ideal production:

$$TE_i = \frac{\exp(x_i' \beta + v_i - w_i)}{\exp(x_i' \beta + v_i)} = \exp(-w_i).\tag{2}$$

It is comprised between 0 and 1. The inefficiency error term w is usually assumed to have a gamma, exponential or half-normal distribution (defined as the distribution of the absolute value of a normal variable) (Cullinane and Song, 2006). In contrast, the symmetric error term v is usually assumed to have a normal or logistic distribution.

2.2 Copula

A copula connects a given number of one-dimensional marginal distributions to form a joint multivariate distribution. In the following, we will restrict the presentation to bivariate copula, which will be used later. Sklar's theorem (Sklar, 1959) states that any bivariate cumulative distribution function (cdf) $F(x_1, x_2)$ can be expressed as

$$F(x_1, x_2) = C_\theta(F_1(x_1), F_2(x_2)),\tag{3}$$

where $F_1(\cdot)$ and $F_2(\cdot)$ and the marginal cdfs of X_1 and X_2 , C_θ is a bivariate copula function and θ is a parameter vector of the copula, commonly referred to as the dependence parameter vector. Bivariate copulas satisfy the following properties

1. $C(u_1, 0) = C(0, u_2) = 0$ for all u_1 and u_2 in $[0, 1]$;
2. $C(u_1, 1) = u_1$ and $C(1, u_2) = u_2$ for all u_1 and u_2 in $[0, 1]$;
3. For all $0 \leq u_1 \leq u_2 \leq 1$ and $0 \leq v_1 \leq v_2 \leq 1$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

If the random vector (X_1, X_2) has a joint density $f(x_1, x_2)$, it can be expressed as a function of the copula density by (Lee et al., 2009):

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} \frac{\partial F(x_1)}{\partial x_1} \frac{\partial F(x_2)}{\partial x_2} = c(u_1, u_2) f_1(x_1) f_2(x_2),\tag{4}$$

where $f_1(x_1)$ and $f_2(x_2)$ are the marginal densities, the small letter c represents the density function of the copula.

The most common measure of dependence between random variables is Pearson's correlation coefficient, which measures only linear dependence and is not very informative for asymmetric distributions (Boyer et al., 1999). To measure nonlinear dependence, rank correlation coefficients such as Kendall's tau and Spearman's rho are more suitable. They can be expressed in terms of the copula as follows (see Nelsen, 2006):

$$\tau_{X_1, X_2} = 4 \iint_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1 = 4E[C(U_1, U_2)] - 1, \quad (5)$$

$$S_{\rho(X_1, X_2)} = 12 \iint_{[0,1]^2} u_1 u_2 dC(u_1, u_2) - 3 = 12 \iint_{[0,1]^2} C(u_1, u_2) du_1 du_2 - 3 = 12E(U_1 U_2) - 3. \quad (6)$$

Moreover, there is the possibility that the tail dependence may divide into upper tail and lower tail dependences, the definitions of which are written as

$$\lambda_L := \lim_{u \rightarrow 0^+} P[X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u)] = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u} \quad (7)$$

and

$$\lambda_U := \lim_{u \rightarrow 1^-} P[X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)] = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u}, \quad (8)$$

where F_1 and F_2 are the marginal cumulative distribution functions of X_1 and X_2 , respectively. Some copulas, such as the Gaussian and Frank copulas, possess the characteristic $\lambda_U = \lambda_L = 0$, while most copulas can capture upper tail and/or lower tail

dependence. For instance, Clayton copula can measure lower tail dependence, and Gumbel copula can measure upper tail dependence, and Joe copula families, which can represent upper tail dependence as well, while student-t copula reflects symmetric tail dependence (Brechmann and Schepsmeier, 2013). However, these copulas can only capture positive dependence except Gaussian, T and Frank copulas. Fortunately, these copulas may then be "rotated" and applied again. A thorough review of rotated copulas may be found in Christian Cech (2006), Brechmann and Schepsmeier, 2013 and Luo (2010). There are three rotated forms, rotated 90 degrees, rotated 180 degrees and rotated 270 degrees. When rotating them by 180 degrees, one obtains the corresponding survival copulas. Non-rotated Clayton copula can capture lower tail dependence, while survival Clayton copula can measure upper tail dependence. Also, rotation by 90 and 270 degrees allows for the modeling of negative dependence. The distribution functions of rotated copulas by 90, 180 and 270 degrees are given as follows:

$$C_{90}(u_1, u_2) = u_2 - C(1 - u_1, u_2), \quad (9)$$

$$C_{180}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2), \quad (10)$$

$$C_{270}(u_1, u_2) = u_1 - C(u_1, 1 - u_2). \quad (11)$$

Fréchet-Hoeffding bounds for copula functions are crucial to select an appropriate copula that should cover the sample space between the lower and the upper bounds and that as a parameter of copula approaches the lower (upper) bound of its permissible range. Fréchet-Hoeffding bounds for copula functions can be expressed as

$$W(u_1, u_2) = \max(u_1 + u_2 - 1, 0) \leq C(u_1, u_2) \leq \min(u_1, u_2) = M(u_1, u_2). \quad (12)$$

The copula approaches the Fréchet–Hoeffding lower (upper) bound. Therefore, we can make use of some copula families that cover a large dependence as much as possible, such as the (rotated) Clayton, (rotated) Gumbel and (rotated) Joe copula families, which can represent both negative and positive dependence. The copula families used in this paper are summarized below. We refer to the formulas and notations from Trivedi and Trivedi (2005) and Sriboonchitta et al. (2013).

(1) Independence copula

The independence copula is defined by $C_{IN} = u_1 u_2$. It connects the cdfs of two independent random variables.

(2) Gaussian copula

This copula takes the form

$$C_{Ga}(u_1, u_2; \rho) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right) dx_1 dx_2, \quad (13)$$

where $-1 < \rho < 1$ is Pearson's correlation coefficient and Φ is the cdf the standard normal distribution function. This is the copula pertaining to a bivariate normal distribution with standard normal marginals and Pearson's linear correlation coefficient ρ . Parameter ρ is related to the Kendall's tau τ and Spearman's rho

S_ρ coefficients by the following equations:

$$\tau = (2/\pi) \sin^{-1}(\rho) \quad \text{and} \quad S_\rho = (6/\pi) \sin^{-1}(\rho/2).$$

(3) T copula

As Gaussian copulas, T copulas belong to the class of elliptical copula (i.e., they are the copulas of elliptically contoured distributions). However, T copulas can capture tail dependence, and it is symmetric extreme dependence. They are defined as

$$C(u_1, u_2 | \rho, \nu) = \int_{-\infty}^{T_v^{-1}(u_1)} dx \int_{-\infty}^{T_v^{-1}(u_2)} dy \frac{1}{2\pi\sqrt{1-\rho^2}} \left\{ 1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)} \right\}^{-(\nu+2)/2}, \quad (14)$$

where ρ is the Pearson's correlation coefficient and T_ν is the cdf of a Student distribution with ν degrees of freedom. When ν tends to infinity, the T copula converges to the Gaussian copula. The symmetric tail dependence can be calculated as

a function of parameters ρ and v as follows:

$$\lambda_U = \lambda_L = 2T_{v+1}(-\sqrt{v+1}\sqrt{(1-\rho)/(1+\rho)}). \quad (15)$$

(4) Clayton copula

The copula $C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$ is called Archimedean and the function φ its generator. There are different Archimedean copulas according to different generators, such as Clayton, Gumbel, Frank, Joe and AMH copulas etc. Clayton copula family can reflect the lower tail dependence for $\theta > 0$. It is characterized by the following formula:

$$C_{Cl}(u_1, u_2 | \theta) = (u_1^\theta + u_2^\theta - 1)^{-1/\theta}. \quad (16)$$

The above copula can only capture a strong lower tail and positive dependence, but it can be rotated and used in capture negative dependence or reflect strong upper tail dependence (see Cech (2006) and Luo (2010)). The corresponding Kendall's tau measure is simply given by $\tau_{CL} = \theta / (\theta + 2)$. However, the relation between the copula parameter and the Spearman's rho is very complicated, and the lower tail dependence can be also simply calculated by $\lambda_L = 2^{-1/\theta}$.

(5) Frank copula

Copulas in the Frank family are the only Archimedean copulas that attain both lower and upper bounds, thus allowing for positive and negative dependence. The corresponding copula function is given by:

$$C_{Fr}(u_1, u_2; \theta) = -\frac{1}{\theta} \ln \left(1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)}{\exp(-\theta) - 1} \right), \quad (17)$$

where $\theta \in (-\infty, +\infty) \setminus \{0\}$; positive (resp., negative) values of θ correspond to positive (resp., negative) dependence. The independence copula is recovered in the limit when $\theta \rightarrow 0$. The rank correlation coefficients are given by:

$$\tau = 1 - \frac{4}{\theta} + 4 \frac{D_1(\theta)}{\theta} \quad \text{and} \quad S_\rho = 1 - \frac{12}{\theta} [D_1(\theta) - D_2(\theta)],$$

where $D_k(\theta)$ is Debye function

$$D_k(\theta) = \frac{k}{\theta^k} \int_0^\theta \frac{t^k}{e^t - 1} dt, k = 1, 2.$$

(6) Gumbel copula

The bivariate Gumbel copula is given by

$$C_{Gum}(u_1, u_2 | \theta) = \exp\left(-\left((-\ln u_1)^{1/\theta} + (-\ln u_2)^{1/\theta}\right)^\theta\right), \quad (18)$$

where $\theta \in (1, +\infty)$. It is an asymmetric copula of the Archimedean family, which

allows for strong upper tail dependence. The rotated Gumbel copula can be applied to capture negative dependence as well. The Kendall's tau of Gumbel copula is given by $\tau_{Gumb} = 1 - \theta^{-1}$, but the Spearman's rho does not have close form. Tail dependence measures can be expressed as $\lambda_U = 2 - 2^{1/\theta}$.

(7) Joe copula

The Joe copula is defined as follows:

$$C_{Joe}(u_1, u_2 | \theta) = 1 - [(1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta (1 - u_2)^\theta]^{1/\theta}, \quad (19)$$

where $\theta \geq 1$. This copula can capture upper tail dependence, which is similar to the Gumbel copula. But it can capture a the stronger upper tail dependence than Gumbel copula (see Bhat and Eluru, 2009), and $\lambda_U = 2 - 2^{1/\theta}$. The Kendall's tau is related to parameter θ by

$$\tau_{Joe} = 1 + \frac{4}{\theta^2} \int_0^1 t \log(t) (1-t)^{2(1-\theta)/\theta} dt \quad (20)$$

The relationship between Spearman's rho and parameter θ does not have a closed form expression. The rotated Joe copula can describe negative dependence as well.

To summarize, different copulas have different characteristics, such as upper tail dependence, lower dependence, positive and negative dependences, etc. Therefore, the above copula families and relevant rotated copula can potentially capture the appropriate dependence between two random variables. Other popular copula families, such as the Farlie–Gumbel–Morgenstern (FGM) copula and Ali-Mikhail-Haq (AMH) copula have been discarded because they cannot achieve Fréchet bounds and can only accommodate relatively weak dependence between the margins. The ranges of dependence of Kendall's tau and Spearman's rho for the FGM copula are $[-2/9, 2/9]$ and $[-1/3, 1/3]$, respectively; the Kendall's tau of the AMH copula is bounded to the interval $[-0.1817, 0.3333]$, and the range of Spearman's rho is $[-0.2711, 0.4784]$ (see Kumar, 2010; Smith, 2008).

3. Copula-based stochastic frontier model

The classical SFM makes the strong assumption that the error components v and w are independent. Smith (2008) proposed to relax this assumption and to model the dependence between v and w using copula. The classical model is recovered as a special case corresponding to the product copula. First, the density function of $f(w, v)$ can be transformed into $f(w, \varepsilon)$ whose formula follows from Smith (2008).

$$f(w, v) = f(u, u + \varepsilon) = f_U(u) \times f_V(w + \varepsilon) \times c_\theta(F_W(w), F_V(w + \varepsilon)) \quad (21)$$

Then, the probability density function of ε can be expressed as

$$f_\Theta(\varepsilon) = \int_{\mathbb{R}_+} f(w, \varepsilon) dw \quad (22)$$

$$\text{or } f_\Theta(\varepsilon) = E_w[f_V(W + \varepsilon) c_\theta(F_W(W), F_V(W + \varepsilon))], \quad (23)$$

where $E_w[\cdot]$ denotes the expected value pertaining to the distribution of technical inefficiency W , Θ represents all parameters that are from marginals and copula function. We use the maximum likelihood method to estimate the copula-based stochastic frontier model. Assume that we have cross-sectional observations of n individuals or firms. The likelihood function is then given by

$$L(\beta, \sigma_w, \sigma_v, \theta) = \prod_{i=1}^n f_{\Theta}(\varepsilon_i) = \prod_{i=1}^n f_{\Theta}(y_i - x_i' \beta) \quad (24)$$

where y_i is the output of individual i or firm i , x_i is explanatory variable vector, σ_w

and σ_v are from marginal distributions of W and V , respectively. Smith (2008)

showed that there are very few density function of ε for which the maximum likelihood estimate MLE has a closed form expression. Then, an alternate approach, maximum simulated likelihood method by Kao et al. (2001), is used to estimate the unknown parameters in copula-based stochastic frontier model (see Green (2010), Burns (2004)). The principle of maximum simulated likelihood is that computing the simulated log likelihood function will require R draws from the corresponding distribution of W for each observation. If W and V are assumed to be half-normal and normal distribution, then the density function of ε can be transformed by using simulation to approximate the integration as follows:

$$\begin{aligned} f(\varepsilon) &= \int_0^{\infty} f_w(w) \times f_v(w + \varepsilon) \times c_{\theta}(F_w(w), F_v(w + \varepsilon)) dw \\ &= \int_0^{\infty} \frac{2 \exp(-w^2 / 2\sigma_w^2)}{\sqrt{2\pi}\sigma_w} \times f_v(w + \varepsilon) \times c_{\theta}(F_w(w), F_v(w + \varepsilon)) dw \\ &= \int_0^{\infty} \frac{2 \exp(-(\sigma_w w_0)^2 / 2\sigma_w^2)}{\sqrt{2\pi}\sigma_w} \times f_v(\sigma_w w_0 + \varepsilon) \times c_{\theta}(F_w(\sigma_w w_0), F_v(\sigma_w w_0 + \varepsilon)) d\sigma_w w_0 \\ &= \int_0^{\infty} \frac{2 \exp(-w_0^2 / 2)}{\sqrt{2\pi}} \times f_v(\sigma_w w_0 + \varepsilon) \times c_{\theta}(F_w(\sigma_w w_0), F_v(\sigma_w w_0 + \varepsilon)) dw_0 \\ &\approx \frac{1}{R} \sum_{r=1}^R f_v(\sigma_w w_{0,ir} + \varepsilon_i) \times c_{\theta}(F_w(\sigma_w w_{0,ir}), F_v(\sigma_w w_{0,ir} + \varepsilon_i)) \end{aligned} \quad (25)$$

where $w_{0,ir}$ is a sequence of R random draws from the standard half normal distribution. The simulated log-likelihood of the copula-based stochastic frontier approach can be expressed as follows:

$$L_s(\beta, \sigma_w, \sigma_v, \theta) \approx \sum_{i=1}^N \log \left\{ \frac{1}{R} \sum_{r=1}^R f_v(\sigma_w w_{0,ir} + \varepsilon_i) \times c_{\theta}(F_w(\sigma_w w_{0,ir}), F_v(\sigma_w w_{0,ir} + \varepsilon_i)) \right\}, \quad (26)$$

where $\varepsilon_i = y_i - \beta' x_i$. In addition, following Battese and Coelli (1995), the variance

terms are parameterized by replacing the relevant parameters, λ , γ and σ^2 , can be

defined by $\lambda = \sigma_w / \sigma_v$, $\gamma = \sigma_w^2 / (\sigma_w^2 + \sigma_v^2)$ and $\sigma^2 = \sigma_w^2 + \sigma_v^2$. The larger is λ , the greater is the inefficiency component in the model (Greene, 2004). The values of λ and γ may reveal whether inefficiency plays an important role in the composite error term and postulate the choice of the stochastic frontier approach in the present study (Chen, 2007). Note that the standard errors of these parameters can be calculated by applying the invariance property.

The most important analysis of copula-based stochastic frontier model is the technical efficiency that is specified with

$$\begin{aligned}
TE_\theta &= E[\exp(-W) | \xi = \varepsilon] \\
&= \frac{1}{f_\theta(\varepsilon)} \int_{R_+} \exp(-w) f(w, \varepsilon) du \\
&= \frac{E_w[\exp(-W) f_V(W + \varepsilon) c_\theta(F_W(W), F_V(W + \varepsilon))]}{E_w[f_V(W + \varepsilon) c_\theta(F_W(W), F_V(W + \varepsilon))]} \\
&\approx \frac{\frac{1}{R} \sum_{r=1}^R \exp(-\sigma_w w_{0,ir}) \times f_V(\sigma_w w_{0,ir} + \varepsilon_i) \times c_\theta(F_W(\sigma_w w_{0,ir}), F_V(\sigma_w w_{0,ir} + \varepsilon_i))}{\frac{1}{R} \sum_{r=1}^R f_V(\sigma_w w_{0,ir} + \varepsilon_i) \times c_\theta(F_W(\sigma_w w_{0,ir}), F_V(\sigma_w w_{0,ir} + \varepsilon_i))}
\end{aligned} \tag{27}$$

Since the maximum simulated likelihood estimation method in this study is applied, the technical efficiency can be calculated by simulated process as well.

4. Empirical results

4.1 The data

The data used in this study were collected by interviewing farmers in Chiang Mai province of Thailand. The data used in this study were from Kunnika Isarangkun na ayuttaya. The upland areas of Chiang Mai province in particular proved fertile areas for high quality coffee. The area has lower humidity, a shorter monsoon season, and a lower annual temperature, creating a micro-climate among the Chiang Mai mountains. Rain water from the mountains seeps into the soil, making it particularly rich in mineral content. The questionnaire was constructed to ask for details about the irrigated rice production at the farms. In particular, there was interest in the area grown, the yields obtained, the use of inputs, such as fertilizer and labor. Information was also obtained on social characteristics of the sample farmers. Data on a total of 111 sample farmers were obtained in the survey. Figure 1 describes the relationships between output and labor, output and fertilizer, labor and fertilizer. They show that there exist significant linear correlations by logarithm forms.

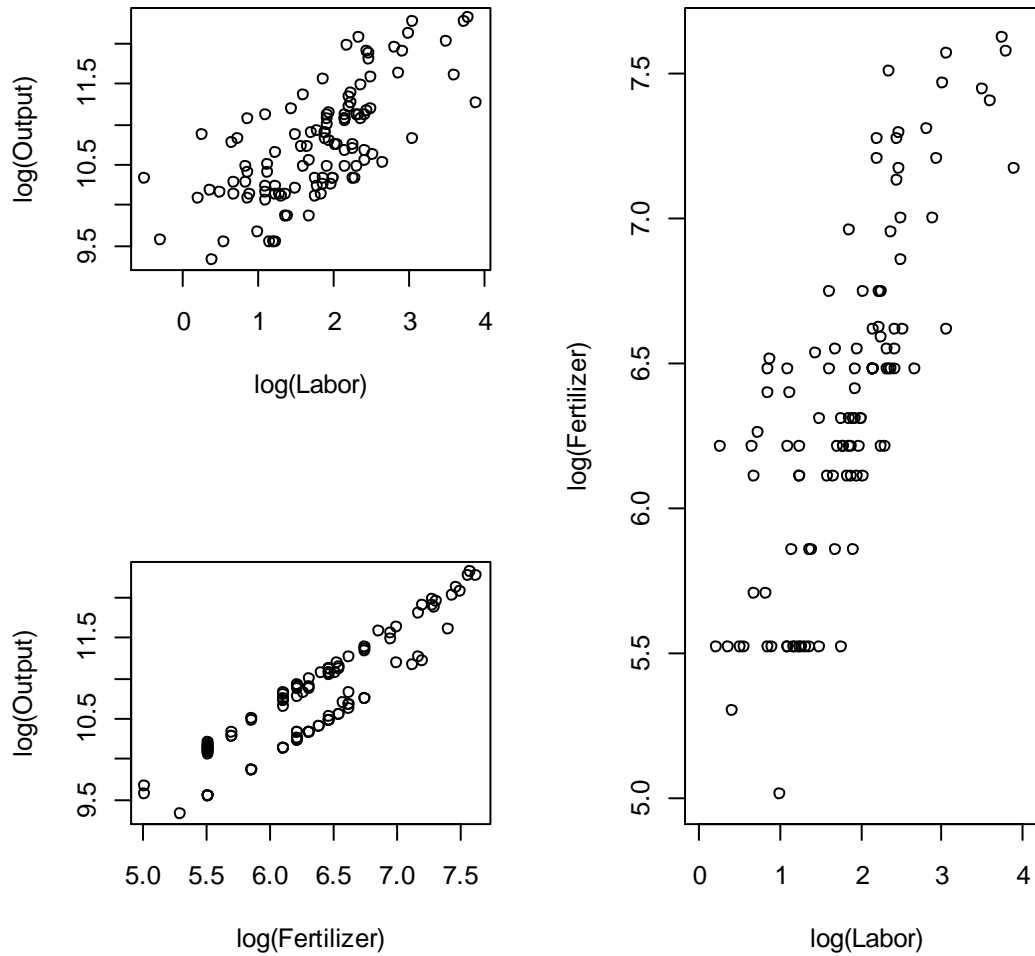


Fig. 1. The relationship between explanatory variables and explained variable.

4.2 Estimating parameters of copula-based stochastic frontier model

The copula-based stochastic frontier model is presented in the form of linear translog production model,

$$\log(Y_i) = \beta_0 + \beta_1 \log(labor) + \beta_2 \log(fertilizer) + \beta_3 (\log(labor))^2 / 2 + \beta_4 (\log(fertilizer))^2 / 2 + \beta_5 \log(labor) \log(fertilizer) + v_i - w_i \quad (28)$$

where Y_i represents intercrop coffee output of the i^{th} farmer. Consider the Log-term of Y is normal distribution, then, we assume that the v is normal distribution and the w is half-normal distribution.

We analyze the copula-based stochastic frontier model by setting R equals 111 (the length of sample), which corresponds to 10 different copula families including independent copula, Gaussian copula, T copula, Frank copula, Clayton copula, Gumbel copula, Joe copula, rotated Clayton copula (180°), rotated Gumbel copula (180°) and rotated Joe copula (180°). If there exist negative correlation between w

and v , the rotated copulas (90° and 270°) should be used in this model. The log-likelihood function in (26) was estimated by Nelder-Mead algorithm using starting values obtained from stochastic frontier model in R program.

Figure 2 describes the AIC and BIC are calculated from each copula-based stochastic frontier model. In terms of the values of AIC and BIC, the rotated Gumbel copula (180°) exhibits better explanatory ability than other dependence structures, while the independent copula performs the worst. Therefore, relaxing the assumption of independent between error components is effective, and is beneficial to improve the explanatory ability of the models. Moreover, the likelihood ratio (LR) test is used to compare independent copula with other copula models. The minimum LR statistics is from the independent copula and Gaussian copula, which equals to 3.294, and is significant at the 10% level. The maximum LR statistics exists between the independent copula and rotated Gumbel copula (180°), the value of which equals to 10.175 that is significant at the 1% level. Therefore, the results of LR statistics verify that the standard Normal-Half-Normal stochastic frontier (stochastic frontier based independent copula) model is inferior to the copula-based stochastic frontier models.

We now turn to the frontier estimates of the preferred model (rotated Gumbel copula-based stochastic frontier model). The estimated results are shown in Table 1. It shows that all parameters are significant at the 5% level, and the values of parameters β equal to 13.31, 1.03, -1.79, 0.09, 0.47 and -0.18 respectively. The variable labor has a positive and significant effect on yield, which implies a 1% increase in labor number allocated to coffee will increase production by 1.03%, but the impact of quantity of fertilizer per ha on coffee production shows a negative sign. This may be explained by the fact that in comparing the use of fertilizer farmers applied the high standard variance (433.5 kg/rai) with a range of 150 kg per acre to 2050 kg per rai. Using wide range of urea would show the inefficient use of fertilizer. It could be related to poor agronomic management which leads to nitrogen losses through NH_3 volatilization and denitrification. Both the estimates of σ_w and σ_v are significantly different from zero at the 1% level. The estimated parameter of rotated Gumbel copula equals to 9.37 that is significantly different from zero at the 1% level as well, which confirms that serious dependence between error components exists, thereby justifying the use of copula-based stochastic frontier model. The parameter γ is 0.9135 that indicates the technical inefficient term has significant impact on intercrop coffee production. We transform the estimated parameter of rotated Gumbel copula to Kendall's tau and Spearman's rho that equal to 0.89 and 0.98, respectively. Plus, there exist strong lower tail dependence that equals to 0.92 between error components, which implies the higher technical efficiency it is, the smaller random disturbance it is.

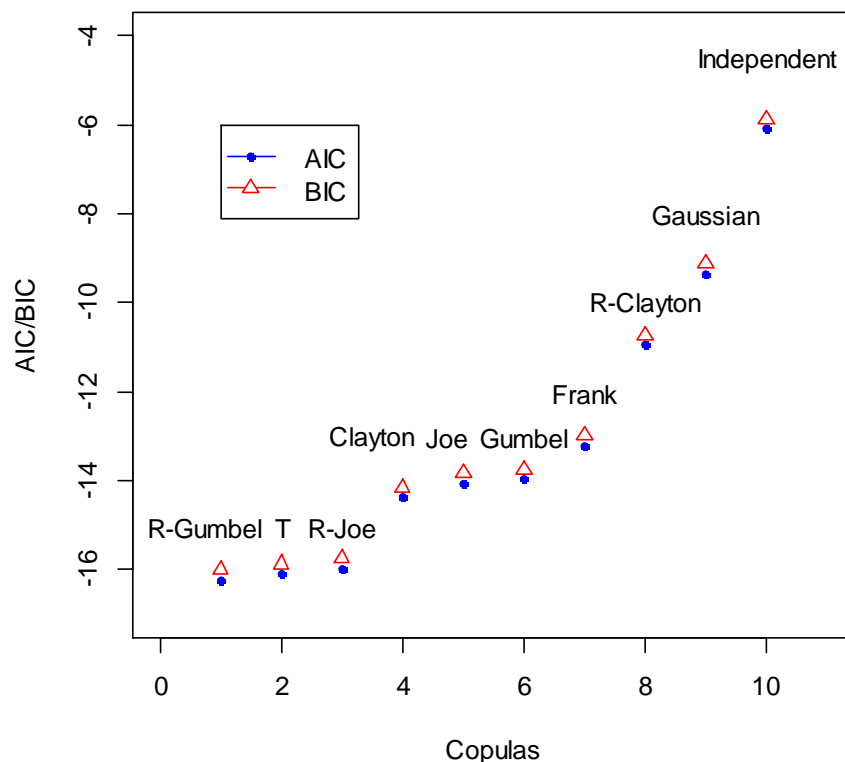


Fig. 2. AIC and BIC of each copula-based stochastic frontier model.

Talbe 1

The estimated parameters of rotated Gumbel copula

Parameters	β_0	β_1	β_2	β_3	β_4	β_5
Estimators	13.3080***	1.0300***	-1.7897***	0.0871***	0.4728***	-0.1763***
s.e	0.2000	0.0578	0.0425	0.0142	0.0059	0.0100
Parameters	σ_w	σ_v	θ	LogL	LR stats	K tau
Estimators	0.9046***	0.2782***	9.3726**	8.1962	10.1750**	0.8933***
s.e	0.0595	0.0252	3.2681			0.0212
Parameters	S rho	λ	σ^2	γ	λ_U	λ_L
Estimators	0.9835***	3.2515***	0.8958***	0.9135***	0	0.9232
s.e	0.0059	0.8151	0.0489	0.0396		

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

Source: computation.

4.3 Technical efficiencies

Figure 3 compares the preferred model (rotated Gumbel copula-based stochastic

frontier) with independent copula based model in term of technical efficiency. The line represents technical efficiencies rotated Gumbel copula based model, while the dotted line describes technical efficiencies of the independent copula-based model. Given the specification of the preferred model, all farmers have a range of 0.19 to 0.83 efficiency scores with an average efficiency 0.53, while the technical efficiency of all farmers are from 0.45 to 0.99 according to independent copula-based stochastic frontier model, and the average technical efficiency is 0.74. Obviously, the traditional independence assumption of random error term and technical inefficiency overestimate the technical efficiencies in this study. Thus it can be seen that employing copula can appropriately relax the independence assumption, thus avoiding the interdependence between W and V makes influence to technical efficiency. Figure 4 illustrates distribution of technical efficiency scores for all sample farmers based independent and rotated Gumbel copulas, respectively. More than half of the farmers have high technical efficiencies that are great than 0.7 and less than 0.99 in production of intercrop coffee based on traditional SFM. However, the technical efficiencies of the preferable model show that there are a half of farmers who have high technical efficiencies but they are less than 0.9. Under the present technology and the preferable model, there is a large proportion (38%) of the sample farmers who have lower levels (<0.30) of technical efficiency score. This suggests that considerable amount of productivity is lost due to inefficiency.

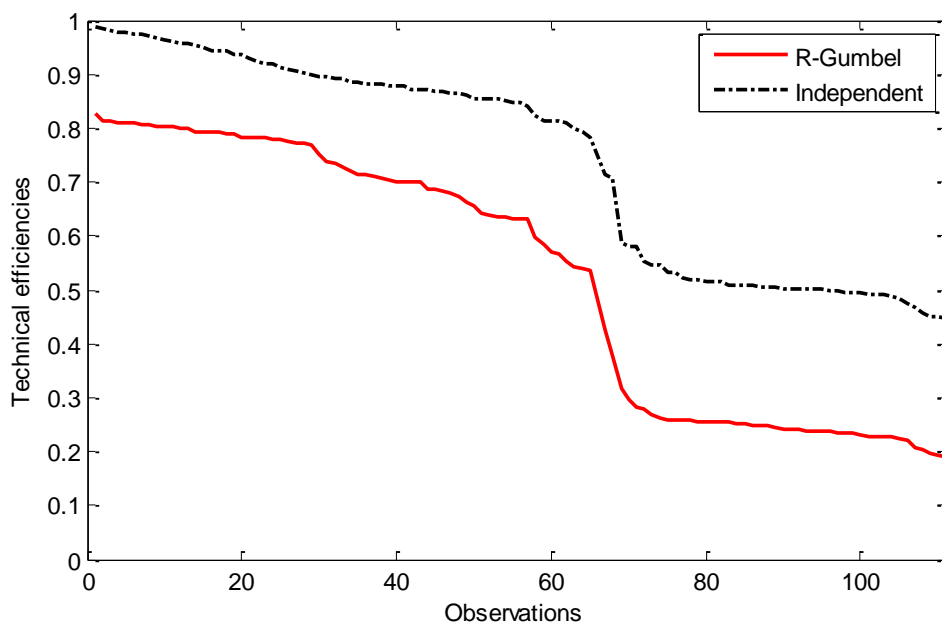


Fig. 3. Technical efficiencies based independent and Rotated Gumbel copulas.

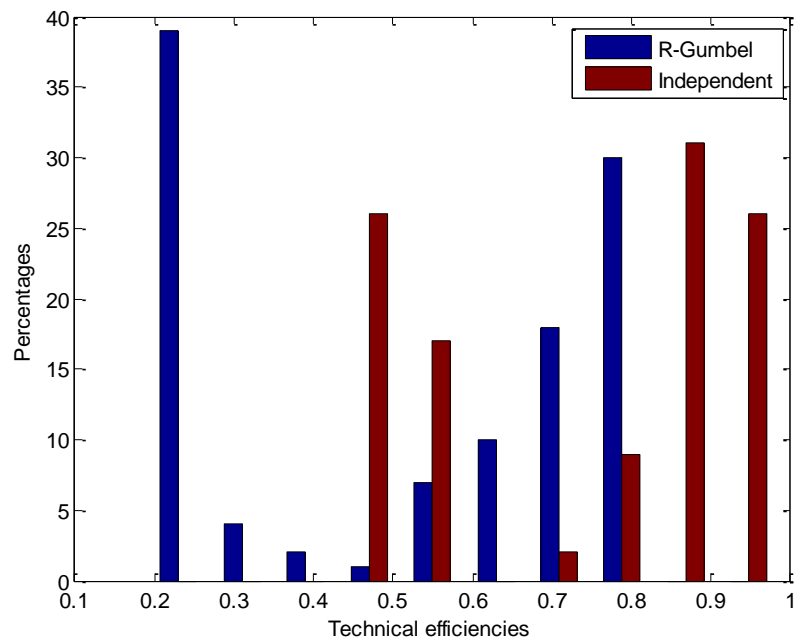


Fig. 4. Range of technical efficiencies and the percentage of sample farmers.

5. Conclusions

In this paper, we have summarized and extended the copula-based stochastic frontier model using maximum simulated likelihood estimation method that is used as substitute for numerical integral method thereby possibly generalizing the use of the copula-based SFA approach. The copula-based SFM allows the models of interest to capture the dependency between W and V . The information criteria, such as AIC and BIC, and LR test are applied to determine whether or not there exists any dependence between random error term and technical inefficiency. The most of copula families, basically, are used to combine with stochastic frontier model, such as Gaussian, T, Frank, Clayton, Gumbel, Joe, rotated Clayton (90° , 180° and 270°), rotated Gumbel (90° , 180° and 270°) and rotated Joe (90° , 180° and 270°) in this study. We used the copula-based stochastic frontier approach to analyze the intercrop coffee production associated with labor and fertilizer. The results showed that copula-based approaches had better performance than traditional stochastic frontier model which overestimate the technical efficiency in this study. This study found some significant results which showed that a 38% the coffee farmers had the production efficiencies are less than 0.3 from the maximum scale of 0.83 efficiency. This suggested that relatively considerable amount makers should seriously improve the efficiency of the intercrop coffee production.

Highlights

- A copula-based stochastic frontier method is used to examine technical

efficiency.

- The maximum simulated likelihood method is employed to estimate copula-based SFM.
- The copula-based approach allows us to capture dependency between U and V.
- We added eight candidates of the Archimedean copulas.
- The traditional stochastic frontier model overestimated the technical efficiency.

Acknowledgement

The support of the Thailand Research Fund (TRF) for the Faculty of Economics and the Faculty of Agriculture in Chiang Mai University is here gratefully acknowledged. We also gratefully thank Professor Murray D. Smith and Professor Hung T. Nguyen for their valuable comments and suggestions.

References

- [1] M.D. Smith, Stochastic frontier models with dependent error components, *Econometrics Journal* 11 (2008) 172–192.
- [2] W. Meeusen, J. van den Broeck, Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 18 (1977) 435–444.
- [3] G.E. Battese, T.J. Coelli, Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India, *The journal of productivity analysis* 3 (1992) 153–169.
- [4] S. Rahman, A. Wiboonpongse, S. Sriboonchitta, Y. Chaovanapoonphol, Production Efficiency of Jasmine Rice Producers in Northern and North-eastern Thailand, *Journal of Agricultural Economics* 60 (2) (2009) 419–435.
- [5] S. Sriboonchitta, H.T. Nguyen, A. Wiboonpongse, J. Liu, Modeling volatility and dependency of agricultural price and production indices of Thailand: Static versus time-varying copulas, *International Journal of Approximate Reasoning* 54 (2013) 793–808.
- [6] M. Sklar, Fonctions de repartition ‘an dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8 (1959) 229–231.
- [7] B.H. Boyer, M.S. Gibson, M. Loretan, Pitfalls in tests for changes in correlations. Federal Reserves Board, IFS Discussion Paper No 597R (1999).
- [8] C. Cech, Copula-based top-down approaches in financial risk aggregation, 2006. Available from: <<http://ssrn.com/abstract=953888>>

or<<http://dx.doi.org/10.2139/ssrn.953888>>.

[9] J. Luo, Stepwise estimation of D-Vines with arbitrary specified copula pairs and EDA Tools, Diploma thesis, Technische Universitaet Muenchen, 2010.

[10] W. Greene, A stochastic frontier model with correction for sample selection, *Journal of Productivity Analysis* Springer 34(1) (2010) 15–24.

[11] A. Wiboonpongse, Quantitative Methods for Analyzing Factors Affecting Farmers' Adoption of Innovation, *Kasetsart Journal (Social Science)* 16 (1995) 53–65.

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

[13] S.C. Kumbhakar, C.A.K. Lovell, *Stochastic Frontier Analysis*. Cambridge UK: Cambridge University Press, 2000.

[14] T. Coelli, S. Rahman, C. Thirtle, Technical, allocative, cost and scale efficiencies in Bangladesh rice cultivation: A non-parametric approach, *Journal of Agricultural Economics* 53 (2002) 607–626.

[15] T. Coelli, D.S.P. Rao and G.E. Battese, *An Introduction to Efficiency and Productivity Analysis*. Boston: Kluwer. (1998).

[16] J. Jondrow, C.A.K. Lovell, I.S. Materov, P. Schmidt, On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics* 19 (1982) 233–38.

[17] R. Burns, *The Simulated Maximum Likelihood Estimation of Stochastic Frontier Models with Correlated Error Components*. Sydney: The University of Sydney. (2004).

[18] S. Sriboonchitta, A. Wiboonpongse, On the Estimation of Stochastic Production Frontiers with Self-Selectivity: Jasmine and Non-Jasmine Rice in Thailand, *CMU. Journal* 4 (2005) 105–124.

[19] A. Wiboonpongse, S. Sriboonchitta, The Effects of Production Inputs, Technical Efficiency and Other Factors on Jasmine and Non-jasmine Rice Yield in Production Year 1999/2000 in Thailand, *Asia-Pacific Productivity Conference at Brisbane, Australia* (2004) 14–16.

[20] S. Sriboonchitta, Evaluation of Cost Efficiency of Thai Public Universities, *International Journal of Intelligent Technologies & Applied Statistics* 5 (2012) 361–374.

- [21] S. Rahman, A. Wiboonpongse, S. Sriboonchitta, K. Kanmanee, Total factor productivity growth and convergence in Northern Thai agriculture, *African Journal of Agricultural Research* 17(7) (2012) 2689–2700.
- [22] A. Wiboonpongse, S. Sriboonchitta, S. Rahman, P. Calkins, Thanet Sriwichailumphun, Joint determination of the choice of planting season and technical efficiency of potato in northern Thailand: A comparison of Greene's versus Heckman's sample selection approach, *African Journal of Business Management* 6 (12) (2012a) 4504–4513.
- [23] S. Rahman, M.K. Hasan, Impact of environmental production conditions on productivity and efficiency: A case study of wheat farmers in Bangladesh, *Journal of environmental management* 88(4) (2008) 1495–1504.
- [24] A. Wiboonpongse, Y. Chaovanapoonphol, G.E. Battese, Inefficiency in Agricultural Markets: Asymmetric Rice Price Transmission in Thailand, AEE working paper (2012b).
- [25] A. Wiboonpongse, S. Sriboonchitta, "Impact of Technical Skill on Potato Production Efficiency in Thailand". presented and contributed paper for Asia Pacific Productivity Conference 2008, Academic Sinica, Taipei, Taiwan, July 17–19, 2008a.
- [26] A. Wiboonpongse, S. Sriboonchitta, "Agricultural Production Growth Assessment for Agro-economic Zones in Northern Thailand Using Statistical Data". presented and contributed paper for Asia Pacific Productivity Conference 2008, Academic Sinica, Taipei, Taiwan, July 17–19, 2008b.
- [27] A. Wiboonpongse, S. Sriboonchitta, E.G. Battese, The Effects of Production Input, Technical Inefficiency and Biological Risk on Jasmine and Non-Jasmine Rice Yields in Thailand, *The Empirical Economics Letters* 4 (3) (2005).
- [28] S. Sriboonchitta, A. Wiboonpongse, "Technical Efficiency of Rural Micro and Community Enterprises in the Upper North of Thailand". presented at Asia-Pacific Productivity Conference, Seoul National University, Seoul, Korea (2006) 17–19.
- [29] S. Sriboonchitta, A. Wiboonpongse, "The Effects of Production Inputs, Technical Efficiency and Other Factors on Jasmine and Non-Jasmine Rice Yields in Thailand". paper presented at Asia-Pacific Productivity conference, Brisbane, Australia (2004) 14–16.
- [30] K. Cullinane, D.W. Song. Estimating the Relative Efficiency of European Container Ports: A Stochastic Frontier Analysis. *Research in Transportation Economics*. Volume 16 (2006) 85–115.

- [31] T.H. Lee, X. Long, Copula-based multivariate GARCH model with uncorrelated dependent errors, *Journal of Econometrics* 150 (2009) 207–218.
- [32] E.C. Brechmann, U. Schepsmeier, Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine, *Journal of Statistical Software* 52 (3) (2013) 1–29.
- [33] C.R. Bhat, N. Eluru. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling, *Transportation Research Part B: Methodological* 43 (7) (2009) 749–765.
- [34] P. Kumar. Probability Distributions and Estimation of Ali-Mikhail-Haq Copula, *Applied Mathematical Sciences* 4 (14) (2010) 657–666.
- [35] Greene, W. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization’s panel data on national health care systems, *Health Econ.* 13 (2004) 959–980.
- [36] C.F. Chen. Applying the stochastic frontier approach to measure hotel managerial efficiency in Taiwan, *Tourism Management* 28 (2007) 696–702.
- [37] C. Kao, L. Lee, M.M. Pitt. Simulated Maximum Likelihood Estimation of The Linear Expenditure System with Binding Non-negativity Constraints, *Annals of Economics and Finance* 2 (2001) 203–223.