Technical efficiency of rice production in Thailand: copula-based stochastic frontier model

Chonrada Nunti1,3, Petchaluck Boonyakunakorn2 and Songsak Sriboonchitta2

1Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
2Puey Ungphakorn Center of Excellence in Econometrics, Chiang Mai University, Chiang Mai, Thailand
3Email: nuntichonrada@gmail.com

Abstract. The study aims to investigate Thai rice production efficiency with the copula-based stochastic frontier model to draft general statistical data for Thai agriculture development in the future. The empirical results show that land and fertilizer have an indisputably positive and significant effect on rice production. The SFM with Gaussian copula has the lowest value regarding to AIC. The best average TE score belongs to the northern region followed by the central region, the north-eastern region, and the southern region with their average scores of 0.86, 0.819, 0.76 and 0.74, respectively. The interesting point is that, even with highest efficiency score, there is the largest gap between the highest and the lowest scores within central region provinces, followed by north-eastern, central and northern regions with their values of 0.2, 0.18, 0.15 and 0.08, respectively.

1. Introduction
Without doubt, Thailand’s main production sector is based on agricultural production concerning most population of the country. Amongst the agriculture sub-sectors, crop production accounts for 60% of the total agricultural production in which rice is the most fundamental commodities. Only until recent years, expansion of dry season cropping, especially on rice, has been more significantly observed. However, Singhapreecha [1] mentioned that there had been an increased growth on agriculture, yield has been observed as a constant but some crops have shown a downward trend.

In the light of more studies, according to the website of the government public relations department as of February 26, 2016 [2], announcement was made that rice production strategies were freshly included in Thailand’s agricultural reform scheduled to roll out in 2017. This reform is expected to stimulate Thai economy, specifically in agriculture sector, for the next period.

Subsequent to rice production efficiency evaluation, the study found that rice yield differs significantly across regions in Thailand. Generally, rice production is divided into four regions according to Thailand’s geography, namely north-eastern, northern, central and southern regions. In regard to aggregate production area, the highest rice production belongs to the northeast area with approximately 12 million tons, followed by the northern 7 million tons, central 4 million tons and southern regions 0.3 million tons. Nevertheless, the central region had the highest rice yield of 603 kilograms per rai followed by the northern, southern and north-eastern regions with 548, 401 and 350, respectively [3].

A few studies, such as [4] [5] and [6] studied Thai rice. For example, the study by Bravo-Ureta and Rieger [7] pointed out that measuring efficiency is crucial since it is the first step in a process leading
to substantial resource savings. Butso and Isvilanonda [8] investigated the returns on a scale of Thai rice productivity and found that they have decreased. The study by Changkid [9] claimed that the low production of Thai rice was the consequence of the low productivity of labor.

Figure 1 shows a comparison between top five exporters from 1980 to 2018, namely Burma, India, Pakistan, Thailand, and Vietnam. It indicates that Thailand had enjoyed its greatest achievement as the biggest rice exporter. Until 2012, when India replaced it as the world's top rice exporter and remained in this position throughout the rest of the time.

Analysis of rice production has significantly become an essential matter in this study. The stochastic frontier model (SFM) is applied for technical efficiency (TE) measurement, as initially proposed by Aigner et al. [10] and Meeusen et al. [11]. However, one of the main strong assumptions in SFM is the independence of the error components. Some researchers have questioned this issue. Pal and Sengupta [12] argued that the random factor such as weather might influence the managerial decisions. Another argument is that the misspecification of the model such as ignoring the significant variable can cause the dependence between error components. In 2008, Smith [13] proposed an introduction of the copula to investigate a relationship between them. Wiboonpongse et al [14] found the dependence between the error and suggested that the dependence should take into consideration. This is because the estimated TE is overestimated by using the standard SFM. Thus, this study will investigate TE by employing copula-based SFM.

2. Methodology

2.1. Stochastic frontier model (SFM)

SFM is the statistical tool used to measure economic efficiency. The original model is proposed for measuring the productive efficiency of a firm when it combines resources in such a way as to produce a given output at the lowest possible average total cost. In contrast, productive inefficiency is defined as the stage when the firm does not provide at its lowest unit cost. In general, there are two ways to enhance productivity. One is upgrading technology. The second way is improvement procedures. For example, worker education improvement, workers apply optimal technology efficiency. Should firms perform well, they will operate more nearly to the existing frontier. Therefore, the increase of productivity may be attributed to efficiency improvement. The first SFM empirical departure point is formulated by Aigner, et al [10] and Meeusen et al. [11]. The stochastic frontier model (SFM) can be expressed as follows:

$$\log Y = x' \beta - U + V,$$

where $Y = y \in \mathbb{R}_+$ indicating the amount of output, $x$ is a vector $k \times 1$ of inputs, $\beta$ is a vector of unknown parameters. The error component $V \sim V(\mu, \sigma)$ is a zero-mean, symmetric error has. $G(\cdot) = Pr(\tilde{V} \leq \cdot)$ which is assumed to be continuous and independent of the variable $x$. However, the
error component depends upon unknown parameters that are collected into a vector $\delta$. The other error is $U + \epsilon$. The random variable with cdf $F(u) = \Pr(U \leq u)$. Its main characteristic of this error is non-negative error. In SFM, $U$ indicates the unobserved inefficiency. If the production has full efficiency, it can be represented as the frontier of maximal output for a given set of inputs $x$. On the other hand, if there are inefficiencies in production, $U$ will reduce output representing as the departure from the frontier.

Another aspect of SFM that some researchers have paid attention to is the assumption of the two error terms. They assume that the one-sided and two-sided error terms are independent because it is convenient to compute the joint distribution of these two errors. However, it is unrealistic in some case studies.

2.2. Copula models

A copula is a function which joins or couples a multivariate distribution function to its one-dimensional marginal distribution functions and can be displayed by the marginal probability distribution of a set of random variables along with different dependence among the random variables. Sklar [15] stated that a unique copula function can capture the dependence structure among the random variables and can also be used as a function of this copula function and marginal distribution functions of these random variables. Thus, any cumulative distribution function (CDF) $F(x_1, x_2)$ of two random variables $(X_1, X_2)$ can be stated as

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)), \quad (2)$$

where $F_1(x_1)$ and $F_2(x_2)$ are the marginal CDFs of variables $X_1$ and $X_2$, and $C$ is a bivariate copula function.

2.3. Copula-based SFM panel data

The strong assumption of the classical SFM is that the error components between $u$ and $v$ are independent. Then, Smith [8] relaxed this assumption by using the copula model to allow the dependence between the errors. The traditional SFM can be extended by using the copula. The joint density $f(u, \epsilon)$ can be derived from as follows:

$$f(u, \epsilon) = f(u, u+\epsilon) = f_u(u) f_\epsilon(u+\epsilon) c_{\theta}(F_u(u), F_\epsilon(u+\epsilon)). \quad (3)$$

Then, we marginalize out $U$ to obtain $f_\theta(\epsilon)$ as the following

$$f_\theta(\epsilon) = E_{\epsilon}(f_u(U+\epsilon) c_{\theta}(F_u(U), F_\epsilon(U+\epsilon))). \quad (4)$$

where $E_{\epsilon}()$ is the expectation regarding to the inefficiency $U$ and $\theta$ denotes the vector of marginals’ parameters and the copula function.

We use the MLE to estimate the copula-based SFM following the study by Wiboonpongse et al. [14]. In this case, we apply panel data. The Hausman test is applied to choose between fixed effects and random effects. The result suggests that the fixed effects are more appropriate for the data. To eliminate the individual effect, we demean the data by subtracting within individual means. Thus, the likelihood equation can be expressed as

$$L(\beta, \sigma_u, \sigma_v, \theta) = \prod_{i=1}^a f_\theta(\log y_{it} - x_{it}^\prime \beta), \quad (5)$$

where $y_{it}$ is the output of firm $i$, $t=1,...,T$. $x_{it}$ is the vector of regressors. $\sigma_u$ and $\sigma_v$ are scale parameters of marginal $U$ and $V$. Smith [13] also demonstrated that there are a few density functions of $\epsilon$ which have a close-form expression. Thus, the estimated copula-based SFM can use equation (5) instead of equation (4) and use Monte Carlo simulation. Since, the inefficiency error $U$ is unobserved, we maximize the log-likelihood (5) by using the maximum simulated likelihood method. In the case
that \( U \) is assumed to be half-normal and \( V \) to be normal distribution, then the pdf of \( \varepsilon \) can be written as:

\[
f(\varepsilon) = \int_0^\infty f_U(u) \times f_V(u + \varepsilon) \times c_\varepsilon(F_U(u), F_V(u + \varepsilon)) du
\]

\[
= \int_0^\infty 2 \exp \left(-\frac{u^2}{2\sigma_u^2}\right) \times f_V(u + \varepsilon) \times c_\varepsilon(F_U(u), F_V(u + \varepsilon)) du
\]

\[
= \int_0^\infty 2 \exp \left(-\frac{\sigma_u^2 u_0^2}{2\sigma_u^2}\right) \times f_V(\sigma_u u_0 + \varepsilon) \times c_\varepsilon(F_U(\sigma_u u_0), F_V(\sigma_u u_0 + \varepsilon)) d\sigma_u u_0
\]

(6)

\[
= \int_0^\infty 2 \exp \left(-\frac{u_0^2}{2}\right) \times f_V(\sigma_u u_0 + \varepsilon) \times c_\varepsilon(F_U(\sigma_u u_0), F_V(\sigma_u u_0 + \varepsilon)) du_0
\]

where \( u_{0,r} \), \( r = 1, \ldots, N \) denotes a sequence of random draws from the distribution of \( U \). Then, the simulated log-likelihood of the copula-based SFM is expressed as

\[
L(\beta, \sigma_u, \sigma_v, \theta) \approx \sum_{r=1}^N \log \left( \frac{1}{N} \sum_{r=1}^N f_V(\sigma_u u_{0,r} + \varepsilon) \times c_\varepsilon(F_U(\sigma_u u_{0,r}), F_V(\sigma_u u_{0,r} + \varepsilon)) \right)
\]

(7)

Following the study of Battese and Coelli [16], the parameters \((\sigma_u, \sigma_v)\) can be transformed into 
\((\lambda, \sigma)\) with \( \lambda = \sigma_u / \sigma_v \) and \( \sigma = \sqrt{\sigma_u^2 + \sigma_v^2} \). The larger \( \lambda \) is, the higher is the inefficiency component in the model (Greene [17]). We can also measure the inefficiency by \( \gamma = \sigma_u^2 / (\sigma_u^2 + \sigma_v^2) \). The values of \( \lambda \) and \( \gamma \) may reveal whether inefficiency is significant in the statistical error. In SFM, the technical inefficiency can estimate their conditions of expectation given \( \varepsilon \),

\[
TE_\varepsilon = E[\exp(-U) | \varepsilon] = \frac{1}{f_\varepsilon(\varepsilon)} \int_0^\infty \exp(-u) f(u, \varepsilon) du
\]

(8)

3. Data and empirical results

3.1. Data

Data are obtained from The Rice Department of Thailand database and divided into 76 provinces according to Thailand’s geography. They are based on a yearly basis ranging from 2009 to 2017. Four variables consist of rice production as the dependent variable, labor, fertilizer and land as the explanatory variables.

3.2. Empirical results

3.2.1. Panel Unit root test. In the process of checking panel unit root test, Levin, Lin & Chu t (LLC) test, Im, Pesaran and Shin W-stat (IPS) test, Augmented Dickey-Fuller (ADF) test and Phillips-Perron
The Second International Conference on Physics, Mathematics and Statistics  
IOP Conf. Series: Journal of Physics: Conf. Series 1324 (2019) 012107  
doi:10.1088/1742-6596/1324/1/012107

5

(PP) test are employed. The results provide that all the variables with logarithm transformation are stationary.

3.2.2. The estimated results of SFM. We analyze the stochastic frontier model with copulas corresponding to 5 different copula families. We selected the best model based on the lowest Akaike information criterion (AIC). The results in Table 1 indicate that the SFM with Gaussian copula has the lowest value in term of AIC. With its values of -146.61 compared with the other copula families, the SFM with Gaussian is selected as the most appropriate model for this study. We also can see that the SFM with Gaussian provides the better-fitted model compared to the standard SFM based on AIC. Thus, the copula-based SFM improves the estimated results.

| Table 1. Information criterion for model selection. |
|-----------------------------------------------|
| Gaussian | Student T | Clayton | Gumbel | Joe | Standard SFM |
| AIC   | -146.61*  | -54.07  | -63.34 | -106.92 | -107.15 | -96.123 |
| LL    | 80.31     | 35.03   | 38.67  | 60.46  | 60.57  | 54.06 |

Note that: * denotes the lowest value of AIC

Table 2 shows the estimated parameters of SFM based on Gaussian copula. The independent variables (land, labor, and fertilizer) have a positive and significant effect on rice production, and the values of parameters $\beta$ equal to 1.614, 0.406, 0.555 and 0.029 respectively. This implies that an increase 1% in the land, labor and fertilizer will increase rice production by 0.406 %, 0.555 % and 0.029 %, respectively. The estimate of $\sigma_u$ is significant at the 1% level which indicates there is inefficiency in rice production. The estimated parameter of Gaussian copula is 0.334 that is significant at the 1% level, which confirms the noise-inefficiency dependence. Since we find the correlation between the error, it indicates that the use of copula with SFM is more appropriate than the standard SFM.

| Table 2. The estimated parameters of copula-based SFM |
|-----------------------------------------------|
| Input variables | SFM based on Gaussian copula | Bays factor |
| Estimates | S.E. | |
| Constant | 1.614**** | 0.128 | 0.000 |
| ln(land) | 0.406**** | 0.043 | 0.000 |
| ln(fertilizer) | 0.555**** | 0.027 | 0.000 |
| ln(labor) | 0.029 | 0.022 | 0.358 |
| $\sigma_v$ | 0.188* | 0.030 | 0.019 |
| $\sigma_u$ | 0.295**** | 0.010 | 0.000 |
| $\rho$ | 0.334* | 0.150 | 0.111 |

Note that: *,**,***,**** denote the weak, evidence, strong evidence, and very strong evidence, respectively

3.2.3. The technical efficiency (TE). Figure 2 shows the rice production technical efficiency based on the Gaussian copula in different regions in Thailand. The best average TE score belongs to the northern region with its highest value of 0.86. Mae Hong Son province is the best one in this region with its highest TE of 0.89. Meanwhile, the lowest in this region with TE score of 0.81 belongs to Nakhon Sawan province. The central region tends to be the second best with its average TE score of 0.82. Samut Prakan province has the highest in the central region with TE of 0.88, which is significantly different from Sa Kaeo province where it has the lowest in this province with TE scores of 0.68. Moreover, the north-eastern region is ranked as the third followed by the southern region with the average TE scores of 0.76 and 0.74, respectively. Loei province has the highest TE score of 0.85.
and Nakhon Ratchasima has the lowest one of 0.67 in the north-eastern region. Meanwhile, in the southern region, Ranong province has the highest score of 0.82 and Phangnga province has the lowest score of 0.67.

The empirical results also point out an interesting issue which is that the central region is the one that has the biggest gap between the highest and the lowest scores, followed by north-eastern, central and northern regions with their values of 0.2, 0.18, 0.15 and 0.08 respectively.

4. Conclusions

This paper used the SFM based on copulas model to analyse the TE of rice production in Thailand. The best copula model is selected by the lowest AIC with Gaussian copula. From the estimation, the results show that land and fertilizer have a positive and significant effect on rice production.

Furthermore, the best average TE score belongs to the northern region followed by the central region, the north-eastern region and the southern region with their averages scores of 0.86, 0.819, 0.76 and 0.74, respectively. The empirical results also show an interesting point that the central region is the one that has the biggest gap between the highest and the lowest scores, followed by north-eastern, central and the northern regions with their values of 0.2, 0.18, 0.15 and 0.08, respectively.

On a global level, the index mundi [18] shows that the average rice yield of Thailand is ranked as the fifty-second among eighty-one countries, which is a relatively low level compared to other
countries. With vary rice production efficiency in each province, the variable is needed to investigate to improve overall Thai’s rice production. The TE result may indicate the province with maximum TE to be a case study for further investigation of what variable detail could be applied or improved for another province. In opposition, lowest TE score province could be emphasized of necessity of development. To the best of benefits, this is the first study of TE of Thai rice productivity evaluated by province.

References
[1] Singhapreecha C 2014 Economy and Agriculture in Thailand FFTC Agriculture Policy Platflom
[2] The Government Public Relations Department: New Rice Strategies to be Included in Thailand’s Agricultural Reform, http://thailand.prd.go.th/ewt_news.php?nid=2767&filename=index. Accessed 25 Aug 2018
[3] Rice department: Rice situation report, http://www.ricethailand.go.th/web/. Accessed 25 Aug 2018
[4] Sribbhoonchita S and Wiboonpongse A 2005 On the Estimation of Stochastic Production Frontiers with Self-Selectivity: Jasmine and Non-Jasmine Rice in Thailand CMU Journal 4 105-124
[5] Rahman S, Wiboonpongse A, Sribbhoonchita S and Chaovanapoonphol Y 2009 Production Efficiency of Jasmine Rice Producers in Northern and North-eastern Thailand Journal of Agricultural Economics 60(2) 419-435
[6] Songsirirot N and Singhapreecha C 2007 Technical Efficiency and Its determinants on Conventional and Certified Organic Jasmine Rice Farms in Yasothon Province: Thammasat Economic Journal 25(2)
[7] Bravo-Ureta B and Rieger L 1991 Dairy Farm Efficiency Measurement Using Stochastic Frontiers and Neoclassical Duality American Journal of Agricultural Economics 73 421-428
[8] Butso O and Isvilanonda S 2010 Two Decades of the Rice Economy of Thailand Applied Economics Journal 17(1) 70-92
[9] Changkid N 2014 Labour use Efficiency of Rice farming in Thailand with Emphasis on The Central Plain Journal of Humanities and Social Sciences SRU 73-82
[10] Aigner D J, Lovell C A K and Schmidt P 1977 Formulation and Estimate of Stochastic Frontier Models Journal of Econometrics 6 21-37
[11] Meeusen W, Broeck J V D 1977 Efficiency estimation from Cobb-Douglas production functions with composed error International Economic Review 18 435-444
[12] Pal M and Sengupta A 1999 A Model of FPF with Correlated Error Componenets: An Application to Indian Agriculture. Sankhya The Indian Journal of Statistics, Series B 337-350
[13] Smith M D 2008 Stochastic frontier models with dependent error components, Econometrics Journal 11 172-192
[14] Wiboonpongse A, Liu J, Sribbhoonchita S and Denoeux T. 2015 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 34-44
[15] Sklar A 1959 Fonctions de répartition à n dimensions et leurs marges Publ. Inst. Statist. Univ. Paris 8 229-231
[16] Battese G E, Coelli T J 1992 Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India The journal of productivity analysis 3 153-169
[17] Greene W 2004 Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization’s panel data on national health care systems Health Econ. 13 959-980
[18] Indexmundi: Milled Rice Yield by Country, https://www.indexmundi.com/agriculture/?commodity=milled-rice&graph=yield. Accessed 25 Aug 2018