

Article

Technical Efficiency of Rice Production in the Upper North of Thailand: Clustering Copula-Based Stochastic Frontier Analysis

Yaovarate Chaovanapoonphol ¹, Jittima Singvejsakul ^{1,*} and Songsak Sriboonchitta ^{2,3}

¹ Department of Agricultural Economy and Development, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand

² Centre of Excellence in Econometrics, Chiang Mai University, Chiang Mai 50200, Thailand

³ Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand

* Correspondence: jittimasvsk@gmail.com

Abstract: This study examines the efficiency of rice production in Thailand, especially major rice, which is the main crop of farmers in all regions of Thailand and is still a pressing issue. Analyzing technical efficiency by using the appropriate analytical tools inevitably brings about determining the correct production efficiency measures. In this study, we applied the k-means algorithm and copula-based stochastic frontier model to cluster farmer groups in order to find the different factors that impact the group, and to relax the assumption of the two components of random error, which is that they are independent to each other; the correlation of the two components of random error is also represented by the estimation of copula. The findings from the k-means clustering algorithms applied in this study indicate that the production frontiers can be divided into two frontiers, with the number of farmers under the frontiers of such production differing from the number of farmers collected in each area. The production frontiers were obtained with 591 farmers under the first production frontier and 65 farmers under the second. In addition, the results reflected a correlation between the two error components U and V . This suggests inefficiencies and zero-mean, and that the symmetric error is not independent of each other. The findings from the application of the copula-based stochastic frontier production function models indicate that land, cost of chemicals, and labor inputs have significant positive effects on the mean output of major rice in both groups of farmers. Therefore, the results of this study indicate that the financial services in rural areas should be continuously promoted by governmental policy, particularly via agricultural loans, to rural people since the utilization of inputs affects the quantity of rice produced. Timely loans should be encouraged.

Keywords: clustering; copula; stochastic production frontier; rice production in Thailand



Citation: Chaovanapoonphol, Y.; Singvejsakul, J.; Sriboonchitta, S. Technical Efficiency of Rice Production in the Upper North of Thailand: Clustering Copula-Based Stochastic Frontier Analysis. *Agriculture* **2022**, *12*, 1585. <https://doi.org/10.3390/agriculture12101585>

Academic Editor: Johannes Sauer

Received: 7 June 2022

Accepted: 1 August 2022

Published: 1 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rice is a crucial economic crop for farmers in the northern regions of Thailand. The purpose of growing rice for farmers is both commercial and consumable, for both major and second rice. Major rice is a very important crop in the region, both in terms of planted area and yield, although the rice planted area in northern Thailand is likely to decrease from 15.154 million rai in 2011 (23.21% of the country's major rice planted area) to 14.136 million rai in 2020 (23.10% of the country's entire annual major rice planted area). However, the production of paddy increased from 7.121 million tons (27.52% of the country's annual major rice output) to 7.519 million tons (31.25% of the country's annual major rice output) in the same period. It can be said that the increase in rice production is the result of an increase in rice yield, namely the paddy yield in the north increased from 470 kgs per rai in 2011 (the country's paddy yield of 396 kgs per rai) to 532 kgs per rai in 2020 (the country's paddy yield of 393 kgs per rai) (Office of Agricultural Economics, 2022). However, compared with the yield of other Asian countries, the rice paddy yield in both the north and the whole country remains lower than in other Asian countries, with Japan and China producing the

highest rice paddy yield of 1065 and 1016 kgs per rai, respectively. Additionally, the rice paddy yield in Myanmar and Indonesia were 942.5 and 1280 kgs per rai, respectively.

Rice planted areas in northern Thailand are spread across 17 provinces, with the major rice planted areas of the upper north region including Chiang Rai, Phetchabun, Chiang Mai, Uttaradit, and Phayao, while the lower northern region includes Nakhon Sawan, Phichit, Kamphaeng Phet, Phitsanulok, Sukhothai, and Uthai Thani. Farmers prefer to grow Jasmine rice 105, one variety or only one other variety of rice, or grow both varieties of rice together. In addition to the Jasmine rice 105, there are native varieties including white rice AK 6, AK 15, Suphanburi 60, and Suphanburi 90 which are sensitive and non-sensitive to light range (Agricultural Information Center, Agricultural Economic Office). Moreover, farmers who grow only one variety of rice usually do so primarily for commercial purposes. Meanwhile, those who grow rice from two varieties often do so for consumption and commercial purposes. They often choose varieties that correspond to the terrain and climatic conditions. Therefore, the efficiency of rice production in Thailand is still a main issue, especially with regards to major rice, which is the main crop of farmers in all regions of Thailand. Increased efficiency will eventually lead to increased farmers' production and income. Analyzing technical efficiency by using analytical tools appropriate to the context inevitably brings about determining the correct production efficiency measures.

In the past, technical efficiency in agricultural production was analyzed using parametric frontiers with stochastic production frontier models, especially rice production. However, this analytical method is still somewhat popular for technical efficiency analysis. The stochastic frontier model's error structure is composed of a two-sided symmetric error and a one-sided component. The one-sided component indicates efficiency, but the two-sided error represents random effects beyond the producer's influence, such as measurement errors and other statistical noise common in empirical connections. Furthermore, production frontier analysis was used to calculate the performance of each production unit, and to investigate the factors that influence its efficiency on the assumption that a two-sided symmetric error and a one-sided component (inefficiency) are unrelated. Despite the fact that the original stochastic frontier models are widely utilized to assess technical efficiency and agricultural output, they have significant restrictions. First, a two-stage estimating strategy is frequently used to fit the model, implying the estimation methods may be inefficient. Furthermore, both the two error term components in the stochastic frontier equation are considered to be independent.

In addition, researchers frequently evaluated the rice production statistics of farmers in numerous places, which are typically classified by subdivisions, such as provinces, considering that the same territories have the same production techniques and are subject to the same production frontier. Areas in different subdivisions show different production frontiers. However, using a zone or territory to divide the production frontier may not be a very appropriate approach, especially in countries with small areas and having similar climates within the country, as in Thailand (such as the Chiang Rai and Chiang Mai provinces, which are located in the upper north). Therefore, the separation of production frontier by provincial boundaries inevitably leads to the determination of different farmer productivity measures that may not be correct. In other words, even if farmers are in different territories, there is no need for different production frontiers, such as in Thailand, where the administrative zone is designated as a province, for convenience, but the rice cultivation of farmers living in different provinces does not reflect the different production practices. Moreover, provinces in Thailand often have similar topography and climate characteristics. Thus, the production frontier's differentiation by territory criteria may entail improper productivity enhancement policies.

At the present, although stochastic frontier analysis (SFA) is still a technical analysis technique used continuously, it has some drawbacks as mentioned above. Thus, sophisticated analytical techniques, appropriate for the context and issues, ought to be used to ensure accurate analysis results. All these techniques can explain the situation and bring about appropriate policies. Technical efficiency estimation techniques for rice production

have been developed and used in empirical studies including: the stochastic metafrontier approach [1–3]; the Bayesian stochastic frontier model [4,5]; the stochastic frontier analyses with a simultaneous (one-step) [6]; the copula-based stochastic frontier model [7]; the zero-inefficiency stochastic frontier [8]; the endogeneity-corrected stochastic frontier model [9]; the stochastic frontier production function with a sub-model of inefficiency effects [10]; and the trivariate Gaussian copula stochastic frontier model [11]. In addition, most of the recent studies investigate the technical efficiency in the agricultural sector by the traditional stochastic frontier method. Some examples include: the study of technical efficiency of maize in district Lakki Marwat of Khyber Pakhtunkhwa, Pakistan [12]; farm-level technical efficiency and its determinants of rice production in Indo-Gangetic plains: a stochastic frontier model approach [13]; and measuring the economic efficiency performance in Latin American and Caribbean countries: an empirical evidence from stochastic production frontier and data envelopment analysis [14]. This paper tries to fill the gap in the literature by using the K-means algorithm and copula-based stochastic frontier model, in which we relax the assumption that the two components of random error are independent to each other; furthermore, the correlation of the two components of random error is also represented by the estimation of copula. Regarding the study “modeling dependence between error components of the stochastic frontier model using copula: application to intercrop coffee production in northern Thailand [15], authors found that the technical efficiencies of the usual assumption of independence between the random error and inefficiency term result in a severe overestimation of technical efficiencies in the study.

Therefore, this article aims to estimate the technical efficiency of rice in the upper north of Thailand with clustering copula stochastic production frontier estimation techniques that overcome the drawbacks of historical estimation techniques, especially with regards to the issue of the two error term components in the stochastic frontier equation. This assumption can be relaxed by fitting the combined distribution of the two components of random error with a copula. Moreover, the K-means clustering is employed to cluster the production frontier of rice production to achieve accurate estimations that lead to improvements in the productivity of rice growers.

2. Materials

This study includes data collected in the provinces of Chiang Mai and Chiang Rai in Thailand's upper north in 2004. The Chiang Mai province encompasses 12.6 million rai, roughly 80% of which is surrounded by mountainous and forested areas and is unsuitable for arable agriculture. The Chiang Rai province has a total area of 7.3 million rai and has more arable land than the Chiang Mai province, since it has a forested area of only about 33% of the total area of the province (Department of Local Administration, 2003). Moreover, the climate of this province is similar to that of the Chiang Mai province since they are neighboring provinces.

Personal interviews were used to collect data from 656 farmers; 331 and 325 farmers from the provinces of Chiang Mai and Chiang Rai constituted the sample, respectively. Table 1 shows basic summary data for the important variables included in the models. These clearly show that Chiang Mai and Chiang Rai farmers grow rice differently. Chiang Mai farmers produced more than Chiang Rai farmers, with mean yields of 5383.27 and 8034.49 kg per rai, respectively. However, one farmer in Chiang Rai province had a yield that was only 224 kg per rai due to a flooding problem. For the Chiang Mai province, the lowest yield was 2000 kg per rai, since this farmer's farm is located in a non-irrigated area and there was less rain than normal in that area. In addition, the highest yields from the provinces of Chiang Mai and Chiang Rai were also different, namely, 49,000 and 23,000 kg per rai, respectively.

Table 1. Summary statistics of key variables for major rice farmers in Chiang Mai and Chiang Rai provinces.

| Variable | Sample Mean | | Sample Standard Deviation | | Minimum | | Maximum | |
|-----------------|-------------|-----------|---------------------------|-----------|---------|-----------|---------|--------|
| | CM | CR | CM | CR | CM | CR | CM | CR |
| TPP (kgs) | 5383.27 | 8034.49 | 4038.21 | 4606.06 | 224 | 2000 | 49,000 | 23,000 |
| AMR (rai) | 9.19 | 13.63 | 6.85 | 7.98 | 1 | 3 | 70 | 45 |
| TSD (kgs) | 74.86 | 104.75 | 62.51 | 58.25 | 5 | 20 | 560 | 285 |
| TCF (kgs) | 285.30 | 511.92 | 345.01 | 502.72 | 0 | 0 | 6200 | 3000 |
| TCP (baht) | 599.38 | 1060.85 | 728.77 | 953.94 | 0 | 0 | 8800 | 4110 |
| TLA (man-hours) | 392.57 | 490.95 | 555.64 | 520.38 | 8 | 60 | 6560 | 3360 |
| SPEN (baht) | 9152.71 | 15,965.38 | 11,215.88 | 15,168.70 | 0 | 0 | 100,000 | 61,250 |
| TVP (baht) | 3335.65 | 30,288.06 | 4137.65 | 14,310.13 | 0 | 14,314.28 | 18,000 | 76,430 |

Source: Author’s survey. Note: CM is the abbreviation for Chiang Mai province and CR is the abbreviation for Chiang Rai province. TPP is the quantity of rice harvested for the sample farmer (kilograms); AMR is the total area planted to major rice (rai); TSD is the total amount of seed sown (kilograms); TCF is the amount of chemical fertilizer applied (kilograms); TCP is the total cost of chemicals (pesticides and herbicides) applied (baht); TLA is the total labor used in the cultivation of major rice (man-hours); SPEN is the total amount of any loans used in major rice production (baht); TPV is the present value of farming tools (baht).

3. Methods

In this paper, the K-Means clustering algorithm and copula-based stochastic frontier model are applied to investigate the technical efficiency of rice production in the upper north of Thailand. The K-Means algorithm is employed to cluster the farmer groups in order to find the different factors that impact the group; the copula-based stochastic frontier model is used to relax the assumption of the two components of random error—that they are independent to each other—and the correlation of the two components of random error is also represented by the estimation of copula. Additionally, this is the first paper which has applied these methods in order to investigate the technical efficiency of rice production, and the first use of a copula-based stochastic frontier method—which is a parametric model that consists of the two main components of random error terms—in an agricultural setting that is appropriate to the investigation factors, impacting the efficiency of output. The details of each estimation method are as follows.

3.1. The k-Means Clustering Algorithm of the Unsupervised Machine Learning

The K-Means method is specified to be a function of the available data type and the purpose of selective analysis [16]. The Euclidean distance is employed to be the criterion that calculates the hyperspace and two-dimensional space [17]. The development of the K-Means algorithms can be represented in five steps by following:

Step 1: clustering input data and minimizing the variance of the data to accept the number of group data and cluster the data into k groups.

Step 2: take the first random sampling or the first k instance of k elements by initializing the first k groups.

Step 3: the arithmetical mean of each cluster formed from the set of data is estimated to be designed as a gradient descent procedure.

Step 4: the initial cluster is a specified class by allocating the K-Means algorithms in each point of the dataset. The nearest cluster assigned each point through measurement of Euclidean distance.

Step 5: each point of recorded data is repeatedly assigned by the algorithm

3.2. Stochastic Frontier Model

The traditional production model considers maximum output as a function $h(x, \beta)$ of a vector x of inputs, which β is a vector of parameters, it can be written as

$$Y = h(x, \beta) \cdot TE, \tag{1}$$

where $TE < 1$, called technical efficiency, is the ratio of actual output $h(x, \beta)$. to maximum feasible output, it can be shown as

$$\ln Y = x' \beta - U, \tag{2}$$

where β is a vector of coefficients and $U = -\ln(TE)$ which is a non-negative error term. However, a theoretical issue with this method is that any measurement mistake is integrated in the disturbance estimation, which is extremely sensitive to outliers. To resolve this issue, symmetric random noise V is added to the right-hand side of (2) (Aigner et al.), resulting in the next equation,

$$\ln Y = x' \beta + \varepsilon, \tag{3}$$

$$\varepsilon = V - U, \tag{4}$$

where the two error components U and V are assumed to be independent. In the model. Therefore, the technical efficiency TE can then be represented as

$$TE = \frac{\exp(x'_i \beta + V - U)}{\exp(x'_i \beta + \beta)} = \exp(-U) \tag{5}$$

where the inefficiency error term U is assumed to be the half-normal distribution.

3.3. Copula

A copula joins a set of one-dimensional marginal distributions to generate a multi-variate joint distribution. Sklar’s theorem states that any cumulative distribution function (cdf) $F(x_1, x_2)$ of a two-dimensional random vector (X_1, X_2) can be expressed as

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)), \tag{6}$$

where $F_1(\cdot)$ and $F_2(\cdot)$ are the marginal of X_1 and X_2 and C is a bivariate function, called a copula.

According to the Frechet-Hoeffding theorem, any copula is contained within the following boundaries. C :

$$\max(u_1 + u_2 - 1, 0) \leq \min(u_1, u_2), \tag{7}$$

for any, $(u_1, u_2) \in [0, 1]^2$. The lower and upper frechet-Hoeffding bounds correspond to two extreme forms of dependence in which the two variables are, respectively, counter monotonic and comonotonic,

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,

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} \tag{8}$$

by the following formula,

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} \tag{9}$$

$$= \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} \frac{\partial F_1(x_1)}{\partial x_1} \frac{\partial F_2(x_2)}{\partial x_2} \tag{10}$$

$$= c(u_1, u_2) f_1(x_1) f_2(x_2) \tag{11}$$

where $f_1(x_1)$ and $f_2(x_2)$ are the marginal densities.

Pearson’s correlation coefficient is the most commonly used measure of random variable dependency. It only measures linear dependence, hence it is not particularly useful for asymmetric distributions. Rank correlation coefficients, such as Kendall’s tau

and Spearman’s rho, are better for measuring nonlinear dependence. Therefore, it can be written in terms of the copula as

$$\tau(X_1, X_2) = 4 \iint_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1 = 4E[C(u_1, u_2)] - 1, \tag{12}$$

$$\rho(X_1, X_2) = 12 \iint_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 3 = 12E[C(u_1, u_2)] - 3, \tag{13}$$

where (U_1, U_2) is a two-dimensional random vector with $C(u_1, u_2)$ as a cumulative distribution function (CDF).

3.4. Copula-Based Stochastic Frontier Model

In the conventional SFM, the error components V and U are assumed to be independent. Smith proposed to relax this assumption, by which the dependence between V and U are modelled using a copula. The classical model is resurrected as a special case for the product copula. Following Smith, the joint density $f(u, \varepsilon)$ can be obtained from $f(u, v)$ as

$$f(u, \varepsilon) = f(u, u + \varepsilon) = f_U(u) f_V(u + \varepsilon) c_\theta(F_U(u), F_V(u + \varepsilon)), \tag{14}$$

Marginalizing out U

$$f_\theta(\varepsilon) = \int_0^{+\infty} f(u, \varepsilon) du, \tag{15}$$

or, equivalently,

$$f_\theta(\varepsilon) = E_\theta[(f_V(U + \varepsilon) c_\theta(F_U(U), F_V(U + \varepsilon)))] \tag{16}$$

where $E_u[\cdot]$ denotes the expectation concerning the technical inefficiency U and θ represents the vector containing all marginal and copula function parameters. The log-likelihood function is produced by assuming the data consists of independent cross-sectional observations of individual or firms.

$$L(\beta, \sigma_u, \sigma_v, \theta) = \sum_{i=1}^n \log f_\theta(\varepsilon_i) = \sum_{i=1}^n \log f_\theta(y_i - x_i' \beta), \tag{17}$$

where y_i is the realization of the output from individual or firm i , x_i is an explanatory variable vector for individual i , and σ_u and σ_v are scale parameters of the marginal distribution of U and V , respectively.

Assuming U and V to have, respectively, half-normal and normal distributions, the density function of ε can be expressed as

$$f(\varepsilon) = \int_0^\infty f_U(u) f_V(u + \varepsilon) c_\theta(F_U(u), F_V(u + \varepsilon)) du, \tag{18}$$

$$= \int_0^\infty \frac{2 \exp(-\frac{u^2}{2\sigma_u^2})}{\sqrt{2\pi}\sigma_u} f_V(u + \varepsilon) c_\theta(F_U(u), F_V(u + \varepsilon)) du, \tag{19}$$

$$= \int_0^\infty \frac{2 \exp(-\frac{u^2}{2\sigma_u^2})}{\sqrt{2\pi}\sigma_u} f_V(\sigma_u u_0 + \varepsilon) c_\theta(F_U(\sigma_u u_0), F_V(\sigma_u u_0 + \varepsilon)) d\sigma_u u_0, \tag{20}$$

$$= \int_0^\infty \frac{2 \exp(-\frac{u^2}{2})}{\sqrt{2\pi}} f_V(\sigma_u u_0 + \varepsilon) c_\theta(F_U(\sigma_u u_0), F_V(\sigma_u u_0 + \varepsilon)) du_0, \tag{21}$$

It can then be approximated by

$$\tilde{f}(\varepsilon) = \frac{1}{N} \sum_{r=1}^N f_v(\sigma_u u_{0,r} + \varepsilon) c_\theta(F_U(\sigma_u u_{0,r}), F_V(\sigma_u u_{0,r} + \varepsilon)), \tag{22}$$

where $u_{0,r}, r = 1, \dots, N$, is a sequence of N random draws from the standard half-normal distribution. The simulated log-likelihood is

$$L_S(\beta, \sigma_u, \sigma_v, \theta) = \sum_{i=1}^n \log\left[\frac{1}{N} f_v(\sigma_u u_{0,r} + \varepsilon_i) c_\theta(F_U(\sigma_u u_{0,r}), F_V(\sigma_u u_{0,r} + \varepsilon_i))\right], \tag{23}$$

where the parameters (σ_W, σ_V) can be transformed to (λ, σ) with $\lambda = \sigma_W/\sigma_V$ and $\sigma = \sqrt{\sigma_W^2 + \sigma_V^2}$. The larger λ , the greater the inefficiency component in the model and the global inefficiency is measured by $\lambda = \sigma_W^2 / (\sigma_W^2 + \sigma_V^2)$. The values of λ and γ reveal whether inefficiency plays an important role in the composite error term.

In stochastic frontier analysis, the technical efficiency terms are of primary interest. The equation of technical efficiency conditional expectations given

$$TE_0 = E[\exp(-w)|\varepsilon] \tag{24}$$

$$= \frac{1}{f_\theta(\varepsilon)} \int_0^{+\infty} \exp(-u) f(u, \varepsilon) du, \tag{25}$$

$$= \frac{E_U[\exp(-U) f_V(U + \varepsilon) c_\theta(F_U(U), F_V(U + \varepsilon))]}{E_U[f_V(U + \varepsilon) c_\theta(F_U(U), F_V(U + \varepsilon))]}, \tag{26}$$

4. Results

Estimates for the K-Means clustering algorithm and copula-based stochastic frontier model are presented in 2 sections. Section 4.1 presents the results of the rice production data using the K-Means clustering algorithm. Section 4.2 is the parameters for the copula-based stochastic frontier model.

4.1. The Estimation of the k-Mean Clustering Algorithm

In this section, the K-Means algorithm is used to cluster the rice production data. Based on the data collection, the data was collected using the territorial boundaries in the provinces of Chiang Mai and Chiang Rai. In the past, we considered the test of the null hypothesis: that the production frontier of rice in both provinces is the same. If the null hypothesis is rejected, different production frontiers will be presented. As a result, the production frontier separated provincially could lead to a separate technical efficiency upgrade approach that does not have a practical effect. The K-Means algorithm, which is selected by the elbow method to find the best value of k , is applied in this study to cluster the rice production data based on the similarity of the rice production of farmers. Based on the results of the K-Means algorithm presented in Table 2 and Figure 1, the rice production data of 656 sample farmers can be clustered into two groups including 591 observations for the first group and 65 observations which were predictively provided for the second group. Additionally, the descriptive statistics for the two major farmer groups are shown in Table 3.

Table 2. Clustering rice production in the Upper North of Thailand by the K-Means algorithm.

| Group | Observation |
|--------------|-------------|
| First group | 591 |
| Second group | 65 |

Source: Authors' estimation.

Table 3. Summary statistics of key variables for major rice farmers in first group and second group.

| Variable | Sample Mean | | Sample Standard Deviation | | Minimum | | Maximum | |
|-----------------|-------------|-----------|---------------------------|-----------|---------|--------|---------|--------|
| | First | Second | First | Second | First | Second | First | Second |
| TPP (kgs) | 5686.67 | 5612.58 | 4580.16 | 3708.59 | 500 | 224 | 49,000 | 23,400 |
| AMR (rai) | 9.13 | 10.15 | 7.27 | 6.86 | 1 | 2 | 70 | 38 |
| TSD (kgs) | 76.39 | 79.20 | 64.78 | 60.49 | 5 | 10 | 560 | 450 |
| TCF (kgs) | 276.60 | 339.54 | 286.83 | 435.65 | 0 | 0 | 2100 | 6200 |
| TCP (baht) | 785.69 | 502.95 | 769.56 | 735.06 | 0 | 0 | 4340 | 8800 |
| TLA (man-hours) | 328.25 | 477.64 | 490.50 | 600.13 | 8 | 16 | 5600 | 6560 |
| SPEN (baht) | 9503.84 | 10,135.62 | 12,986.23 | 10,523.47 | 0 | 0 | 100,000 | 67,500 |
| TVP (baht) | 5573.63 | 6482.72 | 8940.82 | 11,011.71 | 0 | 0 | 53,000 | 76,430 |

Source: Author’s survey. Note: First is the abbreviation for first group and Second is the abbreviation for second group.

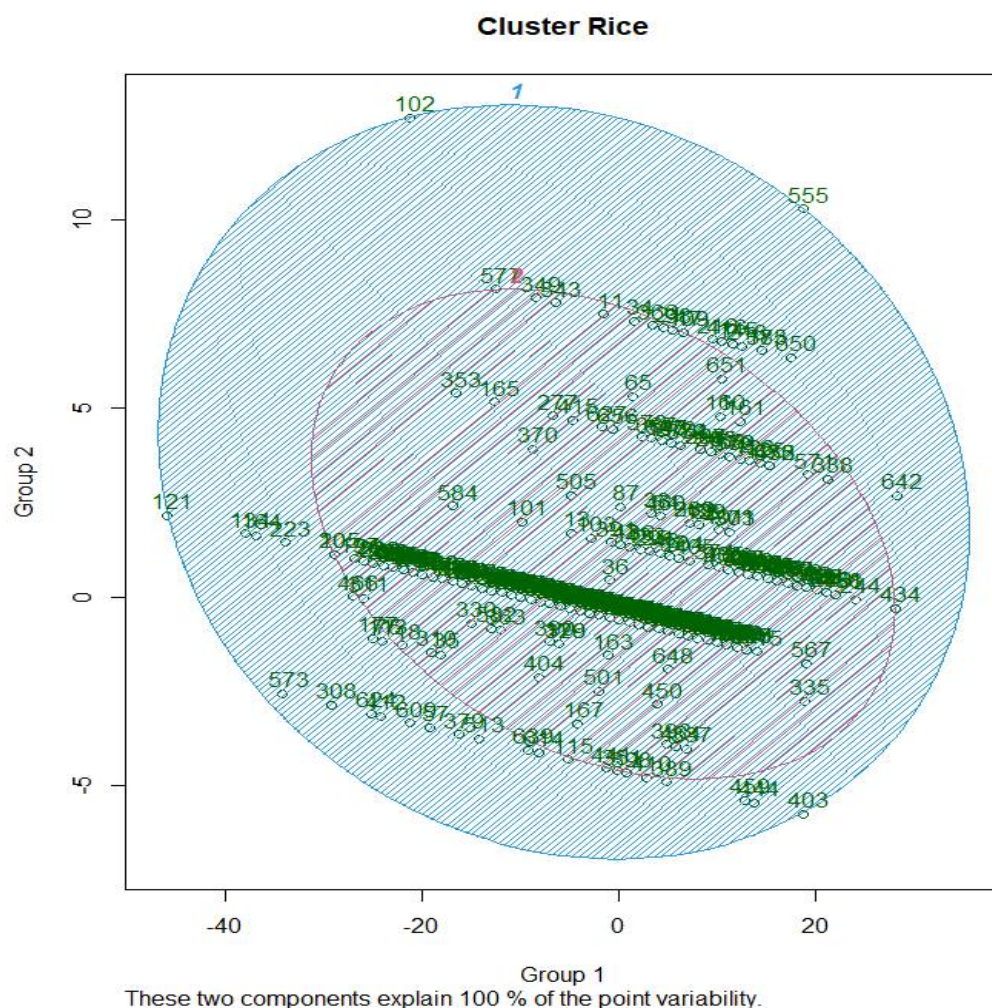


Figure 1. The K-Means clustering results. Source: Authors’ estimation.

The empirical results from the K-Means clustering indicate that the rice production data of farmers in the upper north region can be divided into two groups, with the number of members in each group unequal to the number of farmer samples in each province. This implies that territorial boundaries may not be an appropriate criterion for the production frontier differentiation test. This study differs from [18], which tested a hypothesis to isolate production frontier and the results showed a difference in production frontier by province, thereby guidelines for raising production efficiency by province was presented.

Moreover, this approach has been applied continuously to date. However, this study used the K-Means clustering to classify production function based on the similarity of production practices, a concept that has not yet appeared in Thailand's particular rice production efficiency studies. The results indicate that the similarity between farmers should be considered as the criterion for production frontier differentiation rather than using territory as the criterion. Therefore, farmers in each clustered group are similar in production practices and the use of inputs inevitably determines the way production efficiency can lead to farmers' practices.

4.2. The Estimation of the Copula-Based Stochastic Frontier Model

The estimates of parameters of the preferred stochastic frontier model of both clustered groups of farmers, clustered by the K-Means clustering algorithm, are presented in Table 4. The estimated ρ -parameter, associated with the relationship between the error terms of U and V , are estimated to be 0.737 and 0.990 for farmers in the first group and the second group, respectively. These results indicate that there is a correlation between the error term. Moreover, the results are consistent with [7], which is the only recent work using the copula-based SFA method in rice production in Thailand, that found that the parameter values were similarly significant, reflecting that the copula-based SFA is more appropriate than the standard SFA. The empirical results show that the estimates for both groups of farmers are significant at the one percent level and indicate there is an inefficiency in the production frontier in both clustered groups.

Table 4. Estimates for parameters of stochastic frontier models for major rice farmers in the upper north of Thailand.

| Variables | First Clustered Group | | | Second Clustered Group | | |
|------------|-----------------------|--------|------------|------------------------|--------|------------|
| | Coefficient | S.E. | p -Value | Coefficient | S.E. | p -Value |
| Constant | 7.107 *** | 0.094 | 0.0000 | 7.71 *** | 0.89 | 0.0000 |
| ln(AMR) | 0.853 *** | 0.034 | 0.0000 | 0.75 *** | 0.13 | 0.0000 |
| ln(TSD) | −0.013 | 0.027 | 0.6400 | 0.1100 | 0.12 | 0.3800 |
| ln(TCF) | −0.0068 | 0.0091 | 0.4540 | −0.0072 | 0.0293 | 0.8061 |
| ln(TCP) | 0.0131 *** | 0.0049 | 0.0073 | 0.042 ** | 0.018 | 0.0240 |
| ln(TLA) | 0.028 ** | 0.013 | 0.0340 | −0.0005 | 0.0521 | 0.9925 |
| ln(SPEN) | 0.0018 | 0.0027 | 0.5013 | −0.0085 | 0.0085 | 0.3166 |
| ln(TPV) | 0.0036 | 0.0033 | 0.2723 | 0.0089 | 0.0829 | 0.9141 |
| σ_v | 0.275 *** | 0.022 | 0.0000 | 1.506 *** | 0.213 | 0.0000 |
| σ_u | 0.7503 *** | 0.0047 | 0.0000 | 1.96 *** | 0.23 | 0.0000 |
| ρ | 0.737 *** | 0.025 | 0.0000 | 0.9900 *** | 0.0059 | 0.0000 |

Source: Authors' estimation. Note: significance at the 0.01, 0.05, and 0.10 levels is indicated by *** and **, respectively

The estimated coefficients of the stochastic frontier model generally have the expected signs. For the first clustered group, AMR, TCP, and TLA inputs have positive significant effects on major rice production. These results show that an increase of 1% in the total area planted for major rice, the total cost of chemicals (pesticides and herbicides) applied, and the total labor used in the cultivation of major rice will increase the quantity of rice harvested by 0.853, 0.0131 and 0.028%, respectively. For the second clustered group, there are only two inputs including AMR and TCP that show a positive significant impact on major rice production. This implies that when the total area planted for major rice and the total cost of chemicals are increased by 1%, it will lead to an increase in the quantity of rice harvested by 0.75 and 0.042%, respectively.

The empirical results from the production function indicate that land and cost of chemicals (pesticides and herbicides) are crucial factors for major rice production and the impacts of these two variables on the mean major rice outputs are very close in both clustered groups. The results of this study reflect that land and labor are commonly included in rice production studies and found an impact on rice output. This is consistent with rice growing patterns in Asian countries with small, planted areas. Thus, the use of

large machines is difficult and labor-intensive rice growing is widely used. The results of this study are consistent with several recent studies, which have indicated that land under rice had a strong significant impact on rice production, including [7,11]. Studies that included labor as an input variable in their production frontier found that it had significant effects, for example [11]. Meanwhile, the cost of chemicals variable was included in the production function of some previous studies. However, it was found that the cost of chemicals significantly affected rice production for both groups of farmers. This reflects that rice diseases and pests are important factors affecting the rice production of farmers in the upper north of Thailand.

4.3. The Estimation of the Technical Efficiency

Estimation of the technical efficiency was based on the Gaussian family copula of the crucial rice producers in the provinces of Chiang Mai and Chiang Rai. Table 5 shows the distribution of the predicted technical efficiencies of the sample rice farmers in both clustered groups. For the first group of farmers (591 farmers), the mean technical efficiency was estimated to be 0.579, with a maximum of 0.996 and a minimum of 0.113. This means that, given their input levels and existing technology, the large-rice farmers in the first group were generating around 57.9% of the potential (stochastic) frontier output levels. The technical efficiency of the second group (65 farmers) ranged from 0.333 to 0.956, with an estimated mean technical efficiency of 0.630. This indicates that the second group's main rice farmers generated around 63% of the projected frontier production levels.

When considered in each clustering group, the second group had a 49.23% increase in the number of farmers with technical efficiency score from 0.7 onwards, while the first group had 39.43%, indicating that the second group had more farmers near production frontier than the first group. Considering the characteristics of the average use of inputs of both groups of farmers, it is reflected that even the mean rice yields of both groups of farmers, presented in Table 5, are not that different, with mean yields of 5686.6650 and 5612.575 kgs per rai, respectively. However, it was found that the average use of inputs of both groups of farmers was different. The second group had higher average growing areas and also used all kinds of inputs, except the total cost of chemicals, more than the other group. This result indicates that the second group of farmers used significant inputs such as seed sown, chemical fertilizer, labor, and farming tools at a higher rate. In particular, the use of labor and farm tools suggests that the second group of farmers required these tools to help increase the amount of rice production. The results confirm that if farmers want to increase rice production efficiency, in addition to using inputs directly to the output, better efficiency is also associated with refinement in the rice production practice.

Table 5. Percentages of technical efficiencies based on the Gaussian family copula of major rice farmers within decile ranges.

| Interval | First Group (591) | Second Group (65) |
|------------|-------------------|-------------------|
| <0.50 | 143 (24.19%) | 9 (13.85%) |
| 0.50–0.60 | 81 (13.70%) | 13 (20.00%) |
| 0.60–0.70 | 133 (22.50%) | 11 (16.92%) |
| 0.70–0.80 | 215 (36.37%) | 29 (44.62%) |
| 0.80–0.90 | 18 (3.05%) | 2 (3.08%) |
| 0.90–1.00 | 1 (0.17%) | 1 (1.53%) |
| Mean TE | 0.579 | 0.630 |
| Maximum TE | 0.996 | 0.956 |
| Minimum TE | 0.133 | 0.333 |

Source: Computing, 2022.

5. Conclusions and Policy Recommendation

In the past, technical efficiency studies were based on the assumption that the two error components U and V were independent, which is the weakness of standard SFMs. This study then used copula to prove that there was an association between error terms, and the results obtained using Gaussian copulas showed there was a relationship between the error terms of U and V . In addition, past studies have used hypothetical testing methods on whether production data collected from different areas divided by subdivisions are on the same production frontier. However, different production frontiers do not reflect different production characteristics. In this study we applied the K-Means algorithm and a copula-based stochastic frontier model to cluster farmer groups in order to find the different factors that impact the groups, and to relax the assumption that the two components of random error are independent to each other; furthermore, the correlation of the two components of random error was also represented by the estimation of copula. The concept of segmentation by the K-Means clustering algorithm was then adopted and led to the division of farmers into two groups based on the similarity of the nature of farmers' production, which is a way to make the correct division of production borders and to set guidelines to proper improvements in rice productivity of farmers.

The findings from the K-Means clustering algorithms applied in this study indicate that the production frontiers can be divided into two frontiers, with the number of farmers under the frontiers of such production differing from the number of farmers collected in each area. In other words, data collected from farmers in the two areas included 331 farmers in Chiang Mai and 325 farmers in Chiang Rai, respectively. However, as a result of using the k -means clustering algorithm, two production frontiers were obtained with 591 farmers under the first production frontier and 65 farmers under the second production frontier. In addition, the results reflected a correlation between the two error components U and V . This suggests inefficiencies and zero-mean, and that the symmetric error is not independent of each other.

The Thai government has attempted to enhance rice production by expanding the area planted. Poor productivity, on the other hand, is a major concern in Thailand's main rice production. Farmers were encouraged by the government to use new technology, including high-yielding cultivars and improved agricultural tools, which, for example, steadily increased rice production. The use of copula-based stochastic frontier production function models reveals that land, chemical cost, and labor inputs have a significant positive effect on the mean output of major rice in both farmer groups. In Thailand, there is evidence that use of new technology in large scale rice production, such as the use of varieties, fertilizer, and pesticides, is already being used in financing to purchase rice cultivation inputs through rural financial institutions such as the Bank for Agriculture and Agricultural Cooperatives (BAAC). Despite the fact that rural financial institutions have little impact on rice output, loans from rural financial institutions may modify production practices or the time of input application, affecting major rice farmers' technical efficiency. As a consequence, providing financial services in rural areas should be continuously promoted by government policy, particularly via agricultural loans, to rural people. Furthermore, since the utilization of inputs affects the quantity of rice produced, timely loans should be encouraged.

However, the cost of rice production inputs particularly the cost of pesticides and herbicides, which have a significant impact on rice yields, will inevitably result in higher rice production costs, finally affecting farmers' profits. Managing to prevent disease and insects and reducing the use of purchased materials like chemicals, or the efficient use of fertilizer should be promoted to farmers by educating farmers on ways to reduce rice production costs. In addition, due to the different characteristics of the input use of the two clustering groups of farmers, the group of farmers with a higher average technical efficiency, in addition to using more average inputs, labor and machinery to take care of rice production. The results indicate that if farmers want to increase rice production efficiency,

in addition to using inputs directly to the output, better efficiency is also associated with refinements in the rice production practice.

6. Limitations and Future Recommendation

This study suffers some limitations. First, the empirical model was used in this study, and there is a possibility of the omitted variables problem, which may bias the estimation of time-invariant component of rice production efficiency and the sample size may limit the generalizability and estimation efficiency of this study. Finally, the model outlined following [10] assumed the dependency of the two-component error term of stochastic frontier model that the inefficiency effects are time invariant. This assumption may be questionable; thus, this study would like to recommend that future studies ought to estimate the model that assumes time-varying inefficiency for comparison purposes.

Author Contributions: Conceptualization, S.S.; data curation, Y.C.; methodology, S.S. and J.S.; writing—original draft, Y.C. and J.S.; writing—review & editing, Y.C. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: Centre of Excellence in Econometrics, Chiang Mai University, Chiang Mai, Thailand.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research is partly supported by the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Mabe, F.N.; Donkoh, S.A.; Al-Hassan, S. Accounting for Rice Productivity Heterogeneity in Ghana: The Two-Step Stochastic Meta Frontier Approach. *Int. Sch. Sci. Res. Innov.* **2018**, *12*, 223–232.
- Bravo-Ureta, B.E.; Higgins, D.; Arslan, A. Irrigation Infrastructure and Farm Productivity in the Philippines: A stochastic Meta-Frontier Analysis. *World Dev.* **2020**, *135*, 105073. [[CrossRef](#)]
- Obianefo, C.A.; Ng’ombe, J.N.; Mzyece, A.; Masasi, B.; Obiekwe, N.J.; Anumudu, O.O. Technical Efficiency and Technological Gaps of Rice Production in Anambra State. *Niger. Agric.* **2021**, *11*, 1240. [[CrossRef](#)]
- Valerien, O.P.; Areal, F.J.; Singbo, A.; McKinley, J.; Kei Kajisae, K. Spatial Dependency and Technical Efficiency: An Application of a Bayesian Stochastic Frontier Model to Irrigated and Rainfed Rice Farmers in Bohol. *Philipp. Agric. Econ.* **2018**, *49*, 3.
- Singvejsakul, J.; Intapan, C.; Chaiboonsri, C.; Permsiri, R. Bayesian Stochastic Frontier Analysis of Agricultural productivity efficiency in CLMV. *J. Phys. Conf. Ser.* **2021**, *1936*, 012006. [[CrossRef](#)]
- Pedroso, R.; Tran, D.H.; Viet, T.Q.; Van Le, A.; Dang, K.T.; Le, K.P. Technical efficiency of rice production in the delta of the Vu Gia Thu Bon river basin, Central Vietnam. *World Dev. Perspect.* **2018**, *9*, 18–26. [[CrossRef](#)]
- Nunti, C.; Boonyakunakorn, P.; Sriboonchitta, S. Technical Efficiency of Rice Production in Thailand: Copula-based Stochastic Frontier Model. *J. Phys. Conf. Ser.* **2019**, *1324*, 012107. [[CrossRef](#)]
- Jaehyun, K.; Donghwan, A. Measuring the Effect of Extreme Weather on Rice Production Efficiency using Zero-inefficiency Stochastic Frontier Model, Agricultural and Applied Economics Association. In Proceedings of the 2020 Annual Meeting, Kansas City, MO, USA, 26–28 July 2020.
- Ojo, T.O.; Baiyegunhi, L.J.S. Impact of Climate Change Adaptation Strategies on Rice Productivity in South-West, Nigeria: An Endogeneity Corrected Stochastic Frontier Model. *Sci. Total Environ.* **2020**, *745*, 141151. [[CrossRef](#)]
- Wu, W. Estimation of Technical Efficiency and Output Growth Decomposition for Small-Scale Rice Farmers in Eastern India. *J. Agribus. Dev. Emerg. Econ.* **2020**, *10*, 2.
- Liu, J.; Sriboonchitta, S.; Wiboonpongse, A.; Denceux, T. A Trivariate Gaussian Copula Stochastic Frontier Model with Sample Selection. *Int. J. Approx. Reason.* **2021**, *137*, 181–198. [[CrossRef](#)]
- Khan, D.; Ali, S.; Khan, A.; Waqas, M.; Khan, S.U. Technical Efficiency of Maize in District Lakki Marwat of Khyber Pakhtunkhwa, Pakistan. *Sarhad J. Agric.* **2020**, *36*, 402. [[CrossRef](#)]
- Bahadur, R.; Chandel, S.; Khan, A.; Li, X.; Xia, X. Farm-Level Technical Efficiency and Its Determinants of Rice Production in Indo-Gangetic Plains: A Stochastic Frontier Model Approach. *Sustainability* **2022**, *14*, 2267.
- Koengkan, M.; Fuinhas, J.A.; Kazemzadeh, E.; Osmani, F.; Alavijeh, N.K.; Teixeira, M. Measuring the economic efficiency performance in Latin American and Caribbean countries: An empirical evidence from stochastic production frontier and data envelopment analysis. *Int. Econ.* **2022**, *169*, 43–54. [[CrossRef](#)]

15. Wiboonpongse, A.; Liu, J.; Sriboonchitta, S.; Denoeux, T. Modeling dependence between error components of the stochastic frontier model using copula: Application to intercrop coffee production in Northern Thailand. *Int. J. Approx. Reason.* **2015**, *65*, 34–44. [[CrossRef](#)]
16. Oyelade, O.J.; Oladipupo, O.O.; Obagbuwa, I.C. Application of k-Means Clustering algorithm for prediction of students' academic performance. *Int. J. Comput. Sci. Inf. Secur.* **2010**, *7*, 292–295.
17. Li, Y.; Wu, H. A clustering method based on K-Means algorithm. *Phys. Procedia.* **2012**, *25*, 1104–1109. [[CrossRef](#)]
18. Chaovanapoonphol, Y. The Impact of Financial Services on the Performance of Rice Farmers in the Upper North of Thailand. Ph.D. Thesis, University of New England, Armidale, Australia, 2006.