Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 14(1) 2021: 56-67
http://ejournal.undip.ac.id/index.php/media_statistika

UTILIZATION OF STUDENT'S T DISTRIBUTION TO HANDLE OUTLIERS IN TECHNICAL EFFICIENCY MEASUREMENT

Rizky Zulkarnain[1], Anik Djuraidah[2], I Made Sumertajaya[2], Indahwati[2]
[1] BPS-Statistics Indonesia
[2] Department of Statistics, IPB University

e-mail: zulqarnaen@bps.go.id

DOI: 10.14710/medstat.14.1.56-67

Article Info:
Received: 16 October 2020
Accepted: 7 June 2021
Available Online: 30 June 2021

Keywords:
Cobb-Douglas; *maximum simulated likelihood*; *robust*; *productivity*; *SFA*; *Translog*

Abstract: Stochastic frontier analysis (SFA) is the favorite method for measuring technical efficiency. SFA decomposes the error term into noise and inefficiency components. The noise component is generally assumed to have a normal distribution, while the inefficiency component is assumed to have half normal distribution. However, in the presence of outliers, the normality assumption of noise is not sufficient and can produce implausible technical efficiency scores. This paper aims to explore the use of Student's t distribution for handling outliers in technical efficiency measurement. The model was applied in paddy rice production in East Java. Output variable was the quantity of production, while the input variables were land, seed, fertilizer, labor and capital. To link the output and inputs, Cobb-Douglas or Translog production functions was chosen using likelihood ratio test, where the parameters were estimated using maximum simulated likelihood. Furthermore, the technical efficiency scores were calculated using Jondrow method. The results showed that Student's t distribution for noise can reduce the outliers in technical efficiency scores. Student's t distribution revised the extremely high technical efficiency scores downward and the extremely low technical efficiency scores upward. The performance of model was improved after the outliers were handled, indicated by smaller AIC value.

1. INTRODUCTION

Several methods have been proposed to measure technical efficiency of production units. The methods can be categorized into parametric and nonparametric approaches. Parametric approaches employ specific production function such as Cobb-Douglas or transcendental logarithmic (Translog) production functions, while nonparametric approaches do not use specific production function (Fusco, 2017). Free Disposal Hull (FDH) and Data Envelopment Analysis (DEA) are the examples of nonparametric approaches. Septianto and Widiharih (2010) employed DEA to analyze efficiency of rural bank in Semarang city. While the Deterministic Frontier Analysis (DFA) and Stochastic Frontier Analysis (SFA) are the examples of parametric approaches.

Stochastic frontier model is the frequently preferred method for measuring technical efficiency, especially in agricultural studies (see Heriqbaldi et al., 2015; Mariyono, 2018;

Siagian and Soetjipto, 2020; Sholikah and Kadarmanto, 2020). The primary characteristics of stochastic frontier model is that its error is partitioned into random shock (noise) and inefficiency components (Wheat, 2017). The noise component is generally assumed to have a normal distribution, while the inefficiency component could be exponential, half normal, truncated normal or gamma distributions (Wheat et al., 2019).

The normality assumption of noise is not a problem if the outlier does not exist. If the outliers are present, Wheat et al. (2019) showed that the standard stochastic frontier model can produce exaggrerated and implausible spreads of technical efficiency scores. Production frontier in stochastic frontier model consists of deterministic and stochastic parts. The deterministic part is determined by specified production function, while the stochastic part is determined by noise. The deviance of each production units to the production frontier is recorded as technical inefficiency. Therefore, the technical efficiency scores depend on the value of noise as well as the specification of production function. If the noise has outlier problems, the resulted technical efficiency scores could be affected.

This paper aims to utilize Student's t distribution for handling outlier problems in standard stochastic frontier model. Student's t distribution is one of the heavy-tailed distribution, so that the chance to cover extreme values is greater than the normal distribution. Wheat et al. (2019) had proposed the use of Student's t distribution to solve the outlier issues in England highways maintenance costs. However, this approach has not been widely explored in Indonesia, especially in agricultural studies. Zulkarnain and Indahwati (2021) showed that there were outlier issues in rice farming activities and employed Cauchy distribution to handle it. This study uses the same data with Zulkarnain and Indahwati (2021), but with different approaches. This study employs Student's t distribution for noise component to handle outlier issues. Moreover, this study uses likelihood ratio test to choose the best functional form: Cobb-Douglas or Translog production functions. The inefficiency component is assumed to have half normal distribution, so that the model is called Student's t – Half Normal model. The performance of Student's t – Half Normal model is compared to standard stochastic frontier model (Normal – Half Normal model) using Akaike Information Criterion (AIC).

2. LITERATURE REVIEW

2.1. Production Function

Process of production describes the transformation of inputs (capital, labor, energy, material) into output. Production function is the description of this production processes in a mathematical form. In other words, production function represents a quantitative relationship between a bundle of inputs and the maximum possible output (Rasmussen, 2013; Ray and Kumbhakar, 2015). There are two kinds of theoretical production functions that are frequently used in literatures, namely Cobb-Douglas (CD) production function and transcendental logarithmic (Translog) production function. Specification of these production functions are as follows:

Cobb-Douglas (CD) production function:

$$\ln Y_i = \beta_0 + \sum_{j=1}^{P} \beta_j \ln X_{ij} \tag{1}$$

Transcendental Logarithmic (Translog) production function:

$$\ln Y_i = \beta_0 + \sum_{j=1}^{P} \beta_j \ln X_{ij} + \sum_{j=1}^{P} \delta_j \left(\ln X_{ij}\right)^2 + \sum_{j=1}^{P} \sum_{k>j}^{P} \tau_{jk} \left(\ln X_{ij}\right)(\ln X_{ik}) \tag{2}$$

where Y_i is the output of ith production unit, X_{ij} is the jth input of ith production unit, β_j is the linear component parameter, δ_j is the quadratic component parameter, τ_{jk} is the interaction component parameter and P is the number of production inputs.

Translog production function is the extension of Cobb-Douglas production function by incorporating quadratic and interaction components of inputs. Thus, Translog production function is better than Cobb-Douglas production function if the quadratic and interaction components are statistically significant. Conversely, Cobb-Douglas production function is better than Translog production function if the quadratic and interaction components are equal zero. Likelihood ratio test can be used to determine the best production function (Liu et al., 2019):

$$LR = 2(l_{TR} - l_{CD}) \sim \chi_k^2 \tag{3}$$

where LR is the log-likelihood ratio statistics that has a chi-square distribution, l_{TR} is the maximum log-likelihood value of the Translog production function, l_{CD} is the maximum log-likelihood value of the Cobb-Douglas production function, k is the number of quadratic and interaction components of inputs. If $LR > \chi^2_{\alpha;k}$ then Translog production function is better than Cobb-Douglas production function. If $LR < \chi^2_{\alpha;k}$ then Cobb-Douglas production function is better than Translog production function.

2.2. The Concept of Technical Efficiency

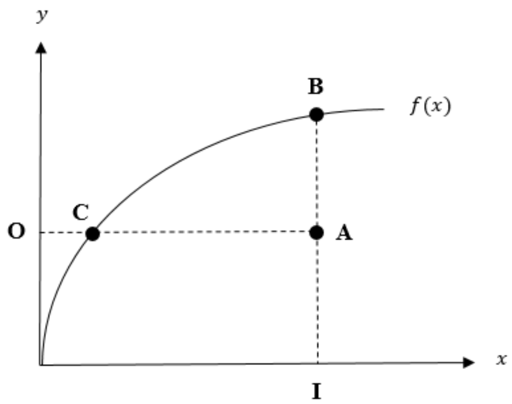

Figure 1. Technical Efficiency Illustration

Technical efficiency can be measured based on input-oriented or output-oriented basis. Input-oriented measures how much inputs can be proportionally cut down without altering the level of output, while output-oriented measures how much output can be increased for the given inputs (O'Donnell, 2018). Figure 1 illustrates the concept of technical efficiency. Suppose A, B and C are three production units with single output (y) and single input (x), where $f(x)$ is the production frontier. B and C are fully efficient as they are exactly in production frontier, while A is inefficient. According to input-oriented measure, the technical inefficiency of A is CA/OA (or technical efficiency of A is OC/OA). According to output-oriented measure, the technical inefficiency of A is AB/IB (or technical efficiency of A is IA/IB).

2.3. Stochastic Frontier Analysis (SFA)

Stochastic frontier model can be expressed as follow (Fusco, 2017):

$$\ln(y_i) = \ln[f(x_i;\beta)] + v_i - u_i \quad ; i = 1,2,\dots,n \tag{4}$$

where y_i is the output of ith production unit, $\boldsymbol{x}_i$ is the inputs vector of ith production unit, $\boldsymbol{\beta}$ is the vector of parameters, $\ln[f(x_i;\boldsymbol{\beta})]$ is the specified production function, v_i is the noise component, u_i is the inefficiency component, v_i and u_i are independent.

If Normal – Half Normal model is used, that is if v_i is normally distributed and u_i has half normal distribution, then the density function of v_i and u_i are as follows:

$$f_v(v) = \frac{1}{\sqrt{2\pi}\sigma_v} exp\left\{-\frac{v^2}{2\sigma_v^2}\right\} \;;-\infty < v < \infty, \sigma_v^2 > 0 \tag{5}$$

$$f_u(u) = \frac{2}{\sqrt{2\pi}\sigma_u} exp\left\{-\frac{u^2}{2\sigma_u^2}\right\} \;;u > 0, \sigma_u^2 > 0$$

Since v_i and u_i are independent, the joint density of v_i and u_i is as follow:

$$f_{u,v}(u,v) = f_v(v).f_u(u) = \frac{1}{\pi\sigma_u\sigma_v} exp\left\{-\frac{u^2}{2\sigma_u^2}-\frac{v^2}{2\sigma_v^2}\right\} \tag{6}$$

The composed error ε_i can be expressed as $\varepsilon_i = v_i - u_i$, thus the equation (6) can be written as:

$$f_{u,\varepsilon}(u,\varepsilon) = \frac{1}{\pi\sigma_u\sigma_v} exp\left\{-\frac{u^2}{2\sigma_u^2}-\frac{(\varepsilon+u)^2}{2\sigma_v^2}\right\} \tag{7}$$

Marginal density function for ε can be derived by integrating equation (7) with respect to u. Fusco (2017) showed that the marginal density of ε is as follow:

$$f_\varepsilon(\varepsilon) = \int_0^\infty f_{u,\varepsilon}(u,\varepsilon)\; du = \frac{2}{\sqrt{2\pi}\sigma}\left[1-\Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right] exp\left\{-\frac{\varepsilon^2}{2\sigma^2}\right\} \tag{8}$$

where $\sigma = \sqrt{\sigma_u^2+\sigma_v^2}$, $\lambda = \sigma_u/\sigma_v$ and $\Phi(.)$ is the cumulative distribution function of the standard normal distribution. The likelihood (L) and log-likelihood functions (l) are further constructed as follows:

$$L = \prod_{i=1}^{n} f_\varepsilon(\varepsilon_i) = \left(\frac{2}{\sqrt{2\pi}\sigma}\right)^n \prod_{i=1}^{n}\left[1-\Phi\left(\frac{\varepsilon_i\lambda}{\sigma}\right)\right] exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\varepsilon_i^2\right\} \tag{9}$$

$$l = \ln(L) = \sum_{i=1}^{n}\left\{\frac{1}{2}ln\left(\frac{2}{\pi}\right) - ln(\sigma) + ln\left[\Phi\left(-\frac{\lambda\varepsilon_i}{\sigma}\right)\right] - \frac{\varepsilon_i^2}{2\sigma^2}\right\} \tag{10}$$

where L is the likelihood function, l is the log-likelihood function, n is the number of data, $\varepsilon_i = \ln(y_i) - \ln[f(x_i;\boldsymbol{\beta})], \sigma = \sqrt{\sigma_u^2+\sigma_v^2}$, $\lambda = \sigma_u/\sigma_v$ and $\Phi(.)$ is the cumulative distribution function of the standard normal distribution.

The parameters $\boldsymbol{\beta}$, λ and σ are estimated by maximizing likelihood or log-likelihood functions in equation (9) or (10). Solution of equation (9) or (10) can be produced using iteration methods such as Newton-Raphson or Fisher scoring. After the parameter estimates are derived, the technical efficiency scores can be calculated. Jondrow et al. (1982) defined the technical efficiency scores for each production units (TE_i) as:

$$TE_i = exp[-E(u_i|\varepsilon_i)] \;; i = 1,2,\dots,n \tag{11}$$

According to equation (11), the conditional density function $u|\varepsilon$ is required to derive technical efficiency scores. If we denote $\sigma_* = \sigma_u\sigma_v/\sigma, \mu_* = -\varepsilon\sigma_u^2/\sigma^2, \lambda = \sigma_u/\sigma_v$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$ then the conditional density function $u|\varepsilon$ is derived as follows:

$$
\begin{aligned}
f(u|\varepsilon) = \frac{f_{u,\varepsilon}(u,\varepsilon)}{f_\varepsilon(\varepsilon)} &= \frac{\frac{1}{\pi\sigma_u\sigma_v} exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon+u)^2}{2\sigma_v^2}\right\}}{\frac{2}{\sqrt{2\pi}\sigma}\left[1-\Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right] exp\left\{-\frac{\varepsilon^2}{2\sigma^2}\right\}} \\
&= \frac{1}{\sqrt{2\pi}\left(\frac{\sigma_u\sigma_v}{\sigma}\right)} \frac{1}{\left[1-\Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right]} exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon+u)^2}{2\sigma_v^2} + \frac{\varepsilon^2}{2\sigma^2}\right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_*} \frac{1}{\left[1-\Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right]} exp\left\{-\left(\frac{(\sigma_u^2+\sigma_v^2)u^2}{2\sigma_u^2\sigma_v^2}\right) - \frac{\varepsilon u}{\sigma_v^2} - \left(\frac{(\sigma^2-\sigma_v^2)\varepsilon^2}{2\sigma_v^2\sigma^2}\right)\right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_*} \frac{1}{\left[1-\Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right]} exp\left\{-\left(\frac{\sigma^2u^2}{2\sigma_u^2\sigma_v^2}\right) - \frac{\varepsilon u}{\sigma_v^2} - \left(\frac{\sigma_u^2\varepsilon^2}{2\sigma_v^2\sigma^2}\right)\right\} \qquad (12) \\
&= \frac{1}{\left[1-\Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right]} \frac{1}{\sqrt{2\pi}\sigma_*} exp\left\{-\frac{1}{2\sigma_*^2}(u-\mu_*)^2\right\} \\
&= \frac{\phi\left(\frac{u-\mu_*}{\sigma_*}\right)}{\Phi(\infty)-\Phi\left(-\frac{\mu_*}{\sigma_*}\right)} \quad ; u > 0
\end{aligned}
$$

where $\phi(.)$ and $\Phi(.)$ denote the probability density function and cumulative distribution function of standard normal distribution, respectively.

Density function $f(u|\varepsilon)$ in equation (12) is the form of truncated normal distribution with parameters $\mu = \mu_*$, $\sigma = \sigma_*, \alpha = -\mu_*/\sigma_*$ and $\beta = \infty$. Expected value of a truncated normal random variable is $\mu + \frac{\phi(\alpha)-\phi(\beta)}{\Phi(\beta)-\Phi(\alpha)}\sigma$. Thus, the expected value of $u|\varepsilon$ is given by:

$$
\begin{aligned}
E(u|\varepsilon) = \mu_* + \frac{\phi(-\mu_*/\sigma_*) - \phi(\infty)}{\Phi(\infty) - \Phi(-\mu_*/\sigma_*)}\sigma_* &= \mu_* + \frac{\phi(-\mu_*/\sigma_*) - 0}{1 - \Phi(-\mu_*/\sigma_*)}\sigma_* \qquad (13) \\
&= \mu_* + \frac{\phi(-\mu_*/\sigma_*)}{\Phi(\mu_*/\sigma_*)}\sigma_*
\end{aligned}
$$

Finally, technical efficiency scores in equation (11) can be expressed as:

$$TE_i = exp[-E(u_i|\varepsilon_i)] = exp\left\{-\mu_{*i} - \sigma_*\left[\frac{\phi(-\mu_{*i}/\sigma_*)}{\Phi(\mu_{*i}/\sigma_*)}\right]\right\} \; ; i = 1,2,\dots,n \qquad (14)$$

where $\mu_{*i} = -\varepsilon_i\sigma_u^2/\sigma^2$, $\sigma_* = \sigma_u\sigma_v/\sigma$, $\phi(.)$ is the density function of standard normal distribution and $\Phi(.)$ is the cumulative distribution function of standard normal distribution.

2.4. Student's t – Half Normal Model

If noise component (v_i) is assumed to have a non-standardized Student's t distribution – which includes a scale parameter σ_v – and inefficiency component (u_i) has half normal distribution (Student's t – Half Normal model), then the density function of v_i and u_i are as follows (Wheat et al., 2019):

$$f_v(v) = \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v}\left[1 + \frac{1}{a}\left(\frac{v}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}} \; ; -\infty < v < \infty, \sigma_v > 0, a > 0$$

$$f_u(u) = \frac{2}{\sqrt{2\pi}\sigma_u} exp\left\{-\frac{u^2}{2\sigma_u^2}\right\} \; ; u > 0, \sigma_u^2 > 0 \tag{15}$$

where a is the shape parameter that specifies the degree of kurtosis of the Student's t distribution, σ_v is the scale parameter and $\Gamma(.)$ is the gamma function. Since v_i and u_i are independent, the joint density of v_i and u_i is given by:

$$f_{u,v}(u, v) = \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v}\left[1 + \frac{1}{a}\left(\frac{v}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}} \frac{2}{\sqrt{2\pi}\sigma_u} exp\left\{-\frac{u^2}{2\sigma_u^2}\right\} \tag{16}$$

The composed error ε_i can be expressed as $\varepsilon_i = v_i - u_i$, thus the equation (16) can be written as:

$$f_{u,\varepsilon}(u, \varepsilon) = \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v}\left[1 + \frac{1}{a}\left(\frac{\varepsilon + u}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}} \frac{2}{\sqrt{2\pi}\sigma_u} exp\left\{-\frac{u^2}{2\sigma_u^2}\right\} \tag{17}$$

Marginal density function for ε can be derived by integrating equation (17) with respect to u:

$$f_\varepsilon(\varepsilon) = \int_0^\infty \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v}\left[1 + \frac{1}{a}\left(\frac{\varepsilon + u}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}} \frac{2}{\sqrt{2\pi}\sigma_u} exp\left\{-\frac{u^2}{2\sigma_u^2}\right\} du \tag{18}$$

Equation (18) is not a closed form, so that the solution could not be derived directly. As an alternative, the solution of Equation (18) could be approximated using simulation method (see Train, 2009). Using $f_v(v)$ and $f_u(u)$ in Equation (15), the Equation (18) can be expressed as follow:

$$f_\varepsilon(\varepsilon) = \int_0^\infty f_v(\varepsilon + u)\, f_u(u)\, du = E_u[f_v(\varepsilon + u)] \tag{19}$$

Thus, $f_\varepsilon(\varepsilon)$ is the expected value of $f_v(\varepsilon + u)$ with respect to u, where $u \geq 0$ is a half normal random variable, $u \sim N^+(0, \sigma_u^2)$. If we generate a half normal random variable (u_q) repeatedly, we can approximate the expected value in equation (19) as an average of simulation data as follow:

$$\widehat{f_\varepsilon(\varepsilon)} = \frac{1}{Q}\sum_{q=1}^{Q} f_v\left(\varepsilon + u_q\right) \; ; -\infty < \varepsilon < \infty, u_q > 0 \tag{20}$$

where Q is the number of simulations and u_q is generated from a half normal distribution, $u_q \sim N^+(0, \sigma_u^2)$. By replacing $f_v\left(\varepsilon + u_q\right)$ with density function of $f_v(v)$ in equation (15), the equation (20) can be expressed as:

$$\widehat{f_\varepsilon(\varepsilon)} = \frac{1}{Q}\frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v}\sum_{q=1}^{Q}\left[1+\frac{1}{a}\left(\frac{\varepsilon+u_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}} \tag{21}$$

Furthermore, the likelihood (L) and log-likelihood (l) functions can be derived as:

$$L = \prod_{i=1}^{n}\widehat{f_\varepsilon(\varepsilon_i)} = \left[\frac{1}{Q}\frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v}\right]^n \prod_{i=1}^{n}\left[\sum_{q=1}^{Q}\left[1+\frac{1}{a}\left(\frac{\varepsilon_i+u_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}\right] \tag{22}$$

$$\begin{aligned} l = \ln(L) &= -n\ln Q + n\ln\left[\Gamma\left(\frac{a+1}{2}\right)\right] - n\ln\left[\Gamma\left(\frac{a}{2}\right)\right] - \frac{n}{2}(\ln\pi + \ln a) \\ &\quad -n\ln\sigma_v + \sum_{i=1}^{n}\ln\left[\sum_{q=1}^{Q}\left[1+\frac{1}{a}\left(\frac{\varepsilon_i+u_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}\right] \end{aligned} \tag{23}$$

where n is the number of data, Q is number of simulations, a is the shape parameter, σ_v is the scale parameter, $\Gamma(.)$ is the gamma function, u_q is generated from a half normal distribution $u_q \sim N^+(0,\sigma_u^2)$, and $\varepsilon_i = \ln(y_i) - \ln[f(x_i;\beta)]$.

The parameters β, a, σ_u and σ_v are estimated by maximizing likelihood or log-likelihood functions in equation (22) or (23). Solution of equation (22) or (23) can be produced using iteration methods such as Newton-Raphson or Fisher scoring.

By following Jondrow et al. (1982) in equation (11), the formula of technical efficiency scores can be derived by firstly constructing $f(u|\varepsilon)$ as follow:

$$f(u|\varepsilon) = \frac{f_{u,\varepsilon}(u,\varepsilon)}{f_\varepsilon(\varepsilon)} = \frac{f_{u,\varepsilon}(u,\varepsilon)}{\widehat{f_\varepsilon(\varepsilon)}} = \frac{f_v(\varepsilon+u)f_u(u)}{\widehat{f_\varepsilon(\varepsilon)}} \quad ; u > 0 \tag{24}$$

Then, the expected value of $u|\varepsilon$ is given by:

$$\begin{aligned} E(u|\varepsilon) &= \int_0^\infty u.f(u|\varepsilon)\,du = \int_0^\infty u\,\frac{f_v(\varepsilon+u)f_u(u)}{\widehat{f_\varepsilon(\varepsilon)}}du \\ &= \frac{1}{\widehat{f_\varepsilon(\varepsilon)}}E_u[u.f_v(\varepsilon+u)] = \frac{\left[\frac{1}{Q}\sum_{q=1}^{Q}u_q.f_v(\varepsilon+u_q)\right]}{\widehat{f_\varepsilon(\varepsilon)}} \\ &= \frac{\sum_{q=1}^{Q}u_q\left[1+\frac{1}{a}\left(\frac{\varepsilon+u_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}}{\sum_{q=1}^{Q}\left[1+\frac{1}{a}\left(\frac{\varepsilon+u_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}} \end{aligned} \tag{25}$$

Finally, technical efficiency scores can be calculated as (Wheat et al., 2019):

$$TE_i = exp[-E(u_i|\varepsilon_i)] = exp\left\{-\frac{\sum_{q=1}^{Q}u_q\left[1+\frac{1}{a}\left(\frac{\varepsilon_i+u_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}}{\sum_{q=1}^{Q}\left[1+\frac{1}{a}\left(\frac{\varepsilon_i+u_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}}\right\} \tag{26}$$

where TE_i is technical efficiency score of ith production unit $(i = 1,2, \dots, n)$, n is the number of data, Q is the number of simulations, a is the shape parameter, σ_v is the scale parameter, $\varepsilon_i = \ln(y_i) - \ln[f(\boldsymbol{x}_i; \boldsymbol{\beta})]$ and u_q is generated from a half normal distribution $u_q \sim N^+(0, \sigma_u^2)$.

3. METHODOLOGY

3.1. Data

Data was acquired from the 2017 Cost Structure of Rice Farming Units (SOUT), a survey conducted by Statistics Indonesia (BPS). The data covers about fourteen thousand rice farming units in East Java. Quantity of production is used as output variable, while land, seed, fertilizer, labor and capital are used as input variables. To link the output and input variables, this study uses Cobb-Douglas (CD) and Translog production functions.

Model specification using CD:

$$\begin{aligned} \ln Production_i = {} & \beta_0 + \beta_1 \ln Land_i + \beta_2 \ln Seed_i + \beta_3 \ln Fertilizer_i + \\ & \beta_4 \ln Labour_i + \beta_5 \ln Capital_i + v_i - u_i \quad ; i = 1,2, \dots, n \end{aligned} \tag{27}$$

Model specification using Translog:

$$\begin{aligned} \ln Production_i = {} & \beta_0 + \beta_1 \ln Land_i + \beta_2 \ln Seed_i + \beta_3 \ln Fertilizer_i + \\ & \beta_4 \ln Labour_i + \beta_5 \ln Capital_i + \delta_1 (\ln Land_i)^2 + \delta_2 (\ln Seed_i)^2 + \\ & \delta_3 (\ln Fertilizer_i)^2 + \delta_4 (\ln Labour_i)^2 + \delta_5 (\ln Capital_i)^2 + \\ & \tau_{12} (\ln Land_i \, x \ln Seed_i) + \tau_{13} (\ln Land_i \, x \ln Fertilizer_i) + \\ & \tau_{14} (\ln Land_i \, x \ln Labour_i) + \tau_{15} (\ln Land_i \, x \ln Capital_i) + \\ & \tau_{23} (\ln Seed_i \, x \ln Fertilizer_i) + \tau_{24} (\ln Seed_i \, x \ln Labour_i) + \\ & \tau_{25} (\ln Seed_i \, x \ln Capital_i) + \tau_{34} (\ln Fertilizer_i \, x \ln Labour_i) + \\ & \tau_{35} (\ln Fertilizer_i \, x \ln Capital_i) + \tau_{45} (\ln Labour_i \, x \ln Capital_i) + \\ & v_i - u_i \quad ; i = 1,2, \dots, n \end{aligned} \tag{28}$$

where $Production_i$ is the production quantity of ith production unit (kg), $Land_i$ is the cultivated area of ith production unit (m^2), $Seed_i$ is the use of seed of ith production unit (kg), $Fertilizer_i$ is the use of fertilizer of ith production unit (kg), $Labour_i$ is the number of labor of ith production unit (person-hours), $Capital_i$ is the capital of ith production unit (thousand rupiahs), v_i is the noise component, u_i is the inefficiency component, v_i and u_i are independent.

3.2. Analysis Method

The steps of analysis are:

a. Transform the output and input variables into natural logarithm.

b. Choose the best production function for linking output and inputs. The likelihood ratio test in equation (3) is used as follow:

   H_0: Cobb-Douglas is better than Translog

   H_1: Translog is better than Cobb-Douglas

   $$LR = 2(l_{TR} - l_{CD})$$

   where LR is the log-likelihood ratio statistics, l_{CD} and l_{TR} are the maximum log-likelihood value of the Cobb-Douglas and Translog production functions, respectively. If $LR > \chi^2_{0.05;15}$ then Translog is better than Cobb-Douglas production functions. If $LR < \chi^2_{0.05;15}$ then Cobb-Douglas is better than Translog production functions.

c. Estimate the parameters using Normal – Half Normal and Student's t – Half Normal models. This paper uses *rfrontier* STATA package to derive the estimates (Wheat et al., 2019).

d. Calculate technical efficiency scores using equation (14) if Normal – Half Normal model is used and equation (26) if Student's t – Half Normal model is used.
e. Evaluate the performance of Normal – Half Normal and Student's t – Half Normal models using Akaike Information Criterion (AIC):

$$AIC = 2p - 2l$$

where p is the number of parameters in model and l is the log-likelihood value of the model.

4. RESULT

Table 1 presents the result of Likelihood Ratio (LR) test for choosing the best production function. Either Normal – Half Normal or Student's t – Half Normal model produces substantial LR statistics. The LR statistics are larger than $\chi^2_{0.05;15}$, thus indicate that Translog production function is better than Cobb-Douglas production function for linking output and input variables.

Table 1. The Result of Likelihood Ratio (LR) Test

	LR statistics	Degree of freedom	$\chi^2_{0.05;15}$	p-value	Decision
Normal – Half Normal	3430.07	15	25.00	0.00	Reject Ho
Student's t – Half Normal	3459.96	15	25.00	0.00	Reject Ho

Table 2. Estimation Results

	Normal – Half Normal			Student's t – Half Normal		
	Coefficient	Standard error	p-value	Coefficient	Standard error	p-value
β_0	4.822	0.077	0.000*	4.996	0.074	0.000*
β_1	-0.226	0.013	0.000*	-0.273	0.012	0.000*
β_2	-0.026	0.013	0.051	-0.038	0.014	0.006*
β_3	0.024	0.013	0.058	0.024	0.012	0.048*
β_4	0.015	0.013	0.224	0.013	0.012	0.286
β_5	0.011	0.012	0.368	0.005	0.012	0.702
δ_1	0.028	0.004	0.000*	0.032	0.004	0.000*
δ_2	-0.031	0.006	0.000*	-0.028	0.005	0.000*
δ_3	-0.030	0.005	0.000*	-0.037	0.005	0.000*
δ_4	0.016	0.004	0.000*	0.016	0.004	0.000*
δ_5	0.000	0.002	0.884	0.000	0.002	0.953
τ_{12}	0.038	0.007	0.000*	0.039	0.007	0.000*
τ_{13}	0.078	0.007	0.000*	0.083	0.008	0.000*
τ_{14}	-0.018	0.006	0.002*	-0.019	0.006	0.001*
τ_{15}	0.028	0.005	0.000*	0.022	0.005	0.000*
τ_{23}	0.000	0.008	0.958	-0.005	0.008	0.512
τ_{24}	-0.003	0.007	0.686	-0.002	0.007	0.724
τ_{25}	-0.026	0.005	0.000*	-0.023	0.005	0.000*
τ_{34}	-0.002	0.007	0.728	0.000	0.007	0.959
τ_{35}	-0.025	0.005	0.000*	-0.019	0.005	0.000*
τ_{45}	0.003	0.004	0.476	0.004	0.004	0.352
σ_v	0.244	0.005	0.000*	0.178	0.006	0.000*
σ_u	0.730	0.008	0.000*	0.710	0.008	0.000*
a	∞	-	-	3.408	0.228	0.000*

*Statistically significant at the 5% level

The estimation results of Normal – Half Normal and Student's t – Half Normal models based on Translog production function are presented in Table 2. Coefficient of β_2 and β_3 are insignificant in Normal – Half Normal model but become significant in Student's t – Half Normal model. Estimate of σ_v is smaller than the estimate of σ_u, either in Normal – Half Normal model or in Student's t – Half Normal model. It indicates that the deviance of production units to the frontier is more due to inefficiency effect.

Table 3 summarizes the results of technical efficiency score for each production unit and its associated ranking. Normal – Half Normal and Student's t – Half Normal models produce considerable differences of technical efficiency score and ranking for several production units. For example, the ranking of sixth production unit is 676 in Normal – Half Normal model but becomes 52 in Student's t – Half Normal model.

Table 3. Technical Efficiency Score and Technical Efficiency Ranking of Each Production Units Based on Normal – Half Normal and Student's t – Half Normal models

Production Unit	Normal – Half Normal		Student's t – Half Normal	
	Technical Efficiency Score	Ranking	Technical Efficiency Score	Ranking
1	0.7952	2197	0.8202	2072
2	0.6909	5427	0.7142	5058
3	0.6685	6106	0.6899	5717
4	0.6570	6410	0.6717	6181
5	0.8109	1768	0.8362	1544
6	0.8616	676	0.8598	52
7	0.5881	8069	0.6038	7823
8	0.6852	5606	0.6988	5472
9	0.8179	1575	0.8391	1445
10	0.5134	9522	0.5172	9516
11	0.7850	2513	0.8004	2729
12	0.8469	916	0.8567	512
13	0.8285	1304	0.8440	1220
14	0.8386	1077	0.8541	679
15	0.6866	5558	0.7030	5345
16	0.6475	6653	0.6355	7134
⋮	⋮	⋮	⋮	⋮
13993	0.8840	360	0.8582	387
13994	0.5785	8248	0.5808	8306
13995	0.6770	5856	0.6937	5614
13996	0.7639	3271	0.7901	3052
13997	0.6580	6386	0.6669	6312
13998	0.7290	4376	0.7316	4656
13999	0.4279	10851	0.4367	10744
14000	0.7661	3190	0.7880	3121
14001	0.7070	4970	0.7335	4612
14002	0.9247	12	0.8086	2481
14003	0.5347	9142	0.5370	9159
14004	0.7433	3928	0.7692	3696
14005	0.7325	4267	0.7528	4114
14006	0.5654	8510	0.5682	8561
14007	0.3393	12421	0.3468	12398
14008	0.4210	10985	0.4198	11008

Figure 2 shows the boxplot of $E(u_i|\varepsilon_i)$, which is used to calculate technical efficiency scores. Normal – Half Normal model produces considerable outliers, which is represented as data points located outside the whiskers (Q1 – 1.5 IQR or Q3 + 1.5 IQR). The outliers are reduced in Student's t – Half Normal model, thus indicates that Student's t – Half Normal model can handle the outliers.

The scatterplot of technical efficiency scores from Normal – Half Normal and Student's t – Half Normal models is depicted in Figure 3. The use of Student's t distribution for noise component results in a shrinkage of technical efficiency scores at the tails. The extremely low technical efficiency scores in Normal – Half Normal model are revised upward by Student's t – Half Normal model. Conversely, the extremely high technical efficiency scores in Normal – Half Normal model are revised downward by Student's t – Half Normal model.

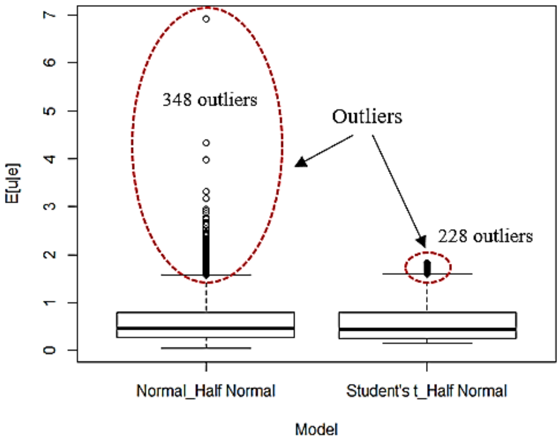

Figure 2. Boxplot of $E(u_i|\varepsilon_i)$

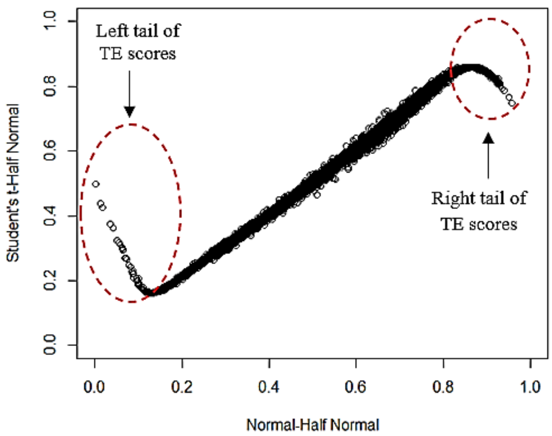

Figure 3. Scatterplot of Technical Efficiency Scores

The performance of Normal – Half Normal and Student's t – Half Normal models are compared using Akaike Information Criterion (AIC) in Table 4. The AIC value of Student's t – Half Normal model is smaller than the Normal – Half Normal model, thus indicates that Student's t – Half Normal model has better performance than the Normal – Half Normal model. In other words, the performance of stochastic frontier model is improved after the outliers are handled.

Table 4. Comparison of AIC Value

Model	AIC
Normal – Half Normal	19406.17
Student's t – Half Normal	19312.66

5. CONCLUSION

Utilization of Student's t distribution in noise component of stochastic frontier model can affect the technical efficiency scores. Student's t – Half Normal model produces less outliers than the Normal – Half Normal model. The performance of Student's t – Half Normal model is better than the Normal – Half Normal model if the outliers are present.

REFERENCES

Fusco, E. (2017). *New Developments in Frontier Models for Objective Assessments*. Sapienza. University of Rome.

Heriqbaldi, U., Purwono, R., Haryanto, T., Primanthi, M.R. (2015). An Analysis of Technical Efficiency of Rice Production in Indonesia, *Asian Soc. Sci.*, Vol. 11, No.3, pp 91–102.

Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P. (1982). On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model, *J. Econom.*, Vol. 19, pp 233–238.

Liu, J., Zhang, C., Hu, R., Zhu, X., Cai, J. (2019). Aging of Agricultural Labor Force and Technical Efficiency in Tea Production: Evidence from Meitan County, China, *Sustainability*, Vol. 11, No.22, pp 6246.

Mariyono, J. (2018). Decomposition Total Factor Productivity of Indonesian Rice Production, *Econ. J. Emerg. Mark.*, Vol. 10, No.2, pp 121–127.

O'Donnell, C.J. (2018). *Productivity and Efficiency Analysis: An Economic Approach to Measuring and Explaining Managerial Performance*. New York. Springer.

Rasmussen, S. (2013). *Production Economics: The Basic Theory of Production Optimisation, Second. ed*. New York. Springer.

Ray, S. and Kumbhakar, S. (2015). *Benchmarking for Performance Evaluation: A Production Frontier Approach*. New Delhi. Springer India.

Septianto, H. and Widiharih, T. (2010). Analisis Efisiensi Bank Perkreditan Rakyat di Kota Semarang dengan Pendekatan Data Envelopment Analysis, *Media Statistika.*, Vol. 3, No.1, pp 41–48.

Siagian, R.A. and Soetjipto, W. (2020). Cost Efficiency of Rice Farming in Indonesia: Stochastic Frontier Approach, *Agricultural Socio-Economics Journal.*, Vol. 20, No.1, pp 7–14.

Sholikah, S. and Kadarmanto. (2020). The Analysis of Technical Efficiency of Inbred and Hybrid Lowland Rice Farming Business, *SOCA J. Sos. Ekon. Pertan.*, Vol. 14, No.3, pp 381–397.

Train, K. (2009). *Discrete Choice Methods with Simulation, Second. ed*. Cambridge. Cambridge University Press.

Wheat, P. (2017). Scale, Quality and Efficiency in Road Maintenance: Evidence for English Local Authorities, *Transp. Policy*, Vol. 59, pp 46–53.

Wheat, P., Stead, A.D., Greene, W.H. (2019). Robust Stochastic Frontier Analysis: A Student’s t-Half Normal Model with Application to Highway Maintenance Costs in England, *J. Product. Anal.*, Vol. 51, pp 21–38.

Zulkarnain, R. and Indahwati. (2021). Robust Stochastic Frontier Using Cauchy Distribution for Noise Component to Measure Efficiency of Rice Farming in East Java, *J. Phys. Conf. Ser.*, Vol. 1863, pp 012031.