Stochastic Frontier Model Incorporating Spatial Effect to Measure Efficiency Component of Multifactor Productivity

Rizky Zulkarnain

BPS – Statistics Indonesia
Jl. Dr. Sutomo 6-8 Jakarta 10710, Indonesia
Email: zulqarnaen@bps.go.id

Abstract:
Accurate measures of productivity are required to monitor target 8.2 of Sustainable Development Goals (SDGs). The objectives of productivity measurement include technology, efficiency, real cost savings, benchmarking production processes and living standards. Various methods have been derived to measure efficiency and productivity, including parametric and nonparametric approach. In parametric approach, stochastic frontier analysis (SFA) is frequently preferred. The characteristic of SFA is that its error is decomposed into two components: one component to capture noise and another one to capture inefficiency. SFA assumes that production activities among units are independent. However, this assumption is often not realistic. Production units may be interdependent through many kinds of externalities and supply-chain networks. This paper considers spatial dependencies among regions to measure technical efficiency of rice farming in East Java province. Spatial relationships among regions are quantified using k-nearest neighbor (k-NN) spatial weights. Maximum likelihood method is used to estimate the parameters. The findings presented in this study are based primarily on data from Cost Structure of Paddy Cultivation Household Survey 2017 (SOUT2017-SPD) that is conducted by Badan Pusat Statistik (BPS). The result showed that there is positive spatial autocorrelation of paddy production in East Java. It is indicated by significant Moran's I statistic with p-value 0.0092. As standard SFA does not take into account the spatial lag into model, the residuals suffer from spatial autocorrelation. This implies that traditional efficiency measures using non-spatial stochastic frontier model generate biased and inconsistent results. Moreover, estimation result suggests that the influence of technical inefficiency in standard SFA is underestimated. It affects the technical efficiency distribution and its ranking. Considering spatial lag into model provides more reliable results. Spatial autocorrelation issues in residuals are resolved and better estimates for technical efficiencies are produced.
Keywords: efficiency, productivity, SDGs, spatial regression, stochastic frontier

1. Introduction:

Target 8.2 of Sustainable Development Goals (SDGs) mandates to achieve higher levels of economic productivity through diversification, technological upgrading and innovation. Traditionally, the productivity is measured using partial productivity such as labour productivity or capital productivity. However, partial productivity fails to capture any direct effects of technological change and does not take into account the elasticity of substitution among inputs. Multifactor productivity (MFP) is the more appropriate tool to measure productivity. MFP relates the output to a bundle of inputs. MFP growth can be decomposed into technical efficiency change, technical progress and scale economic effects. Therefore, an attempt to improve MFP could be undertaken via enhancing efficiency and adopting new technology.

Various methods have been derived to measure efficiency and MFP. The measurement of MFP can be classified into two main approaches: non-frontier approaches and frontier approaches. Non-frontier approaches assume that outputs are efficiently produced on production frontier, while frontier approaches allow for outputs being produced off the frontier (Suphannachart & Warr, 2010). Both the non-frontier and frontier approaches could be further grouped into nonparametric and parametric methods. Nonparametric methods do not use particular functional form, while parametric methods use specific functional form such as cobb-douglas or transcendental logarithmic (translog) production functions.

Stochastic frontier analysis (SFA) is the popular parametric frontier method for measuring efficiency and MFP. Efficiency measurement needs cross-section or panel data, while MFP measurement needs panel data. The characteristic of SFA is that its error is decomposed into two components: one component to capture noise and another one to capture inefficiency. SFA assumes that noise and inefficiency terms among units are independent (Coelli *et al.*, 2005). However, this assumption is often not realistic. Production units may be interdependent through many kinds of externalities and supply-chain networks. Knowledge spillover is a kind of spatial externalities that drives the imitation and innovation of production technology. While supply-chain networks can exist since production activities in a particular area impress the labor market and the intermediate goods market in its surrounding areas (Tsukamoto, 2019).

Fusco (2017) explains that if spatial dependencies are significant, then traditional techniques used to estimate the SFA parameters generate biased results. If the errors are spatially correlated, then assumption of a spherical error covariance matrix is violated, leading to biased and inconsistent estimators. In order to resolve these problems, studies on stochastic frontier models incorporating spatial effects have developed rapidly since 2010. Fusco & Vidoli (2013) proposed a spatial stochastic frontier model with spatial error model (SEM) structure using maximum likelihood estimation. Adetutu *et al.* (2015) estimated stochastic frontier model that incorporates spatial lag of X (SLX) structure. Glass *et al.* (2016) developed a spatial autoregressive (SAR) stochastic frontier model using panel data. Tsukamoto (2019) proposed a spatial autoregressive stochastic frontier model for panel data incorporating a model of technical inefficiency. The model simultaneously estimates the stochastic frontier and determinants of technical inefficiency in single stage.

This paper uses spatial stochastic frontier model to measure technical efficiency of rice farming in East Java province. East Java is preferred as the second largest rice grower in Indonesia. Each district is considered as decision making unit since each district is an important policy maker for its rice farming. Therefore, aggregate technical efficiencies are estimated for 38 districts in East Java. The spatial relationships among districts are quantified using k-nearest neighbor (k-NN) spatial weights. For functional form of spatial stochastic frontier model, this paper uses cobb-douglas production function. Maximum likelihood method is used to estimate the parameters. The performance of spatial stochastic frontier model is compared to standard non-spatial stochastic frontier model using mean absolute prediction error (MAPE) criteria.

2. Methodology:

Data for this study is acquired from Cost Structure of Paddy Cultivation Household Survey 2017 (SOUT2017-SPD) that is conducted by Badan Pusat Statistik (BPS). Using the data, aggregate input-output variables are calculated for 38 districts in East Java. Table 1 presents input-output variables used in the analysis.

Table 1. Lists of input-output variables used in the analysis

No.	Input/output	Variable	Description	Unit
1	Output	Y	Quantity of production	Kg
2	Input	X_1	Land	m^2
3	Input	X_2	Seed	Kg
4	Input	X_3	Fertilizer	Kg
5	Input	X_4	Labour	Person-hours
6	Input	X_5	Capital	Thousand rupiahs

This study adopts model proposed by Fusco & Vidoli (2013) that incorporates spatial lag of inefficiency into standard stochastic frontier model. Output and input variables are linked using cobb-douglas production function. The model could be written in matrix formulation as:

$$\begin{aligned}\dot{\boldsymbol{y}} &= \dot{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{v} - \boldsymbol{u} \\ &= \dot{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{v} - (\boldsymbol{I} - \lambda\boldsymbol{W})^{-1}\tilde{\boldsymbol{u}}\end{aligned} \quad (1)$$

where $\dot{\boldsymbol{y}}$ is $(nx1)$ vector of output in logarithmic form, $\dot{\boldsymbol{X}}$ is (nxp) matrix of inputs in logarithmic form, $\boldsymbol{\beta}$ is $(px1)$ vector of parameters, $\boldsymbol{v}$ is $(nx1)$ vector of noise term, $\boldsymbol{v} \sim iid\ N(\boldsymbol{0}, \sigma_v^2\boldsymbol{I})$, $\boldsymbol{u}$ is $(nx1)$ vector of inefficiency term, $\boldsymbol{u} \sim N^{+}\left(\mathbf{0}, (\boldsymbol{I}-\lambda\boldsymbol{W})^{-1}(\boldsymbol{I}-\lambda\boldsymbol{W}^{T})^{-1}\sigma_{\tilde{u}}^{2}\boldsymbol{I}\right)$, $\tilde{\boldsymbol{u}}$ is $(nx1)$ random vector, $\tilde{\boldsymbol{u}} \sim N\left(\mathbf{0}, \sigma_{\tilde{u}}^{2}\boldsymbol{I}\right)$, $\lambda < |1|$ is a spatial autoregressive coefficient, and $\boldsymbol{W}$ is (nxn) spatial weights matrix.

The inefficiency term for each unit (u_i) could be denoted as $(1-\lambda\sum_i w_{i.})^{-1}\tilde{u}_i$ and by noting $(1-\lambda\sum_i w_{i.})$ as $\delta(\lambda)$, the density function of u_i could be written as:

$$f([\delta(\lambda)]^{-1}\tilde{u}) = \frac{2}{\sqrt{2\pi}[\delta(\lambda)]^{-1}\sigma_{\tilde{u}}} exp\left\{-\frac{[\delta(\lambda)]^{-2}\tilde{u}^2}{2[\delta(\lambda)]^{-2}\sigma_{\tilde{u}}^2}\right\} \quad (2)$$

The density function of noise term for each unit (v_i) could be written as follows:

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} exp\left\{-\frac{v^2}{2\sigma_v^2}\right\} \quad (3)$$

Given the independence assumption, the joint density function of $[\delta(\lambda)]^{-1}\tilde{u}$ and v is the product of equations (2) and (3):

$$f([\delta(\lambda)]^{-1}\tilde{u}, v) = \frac{1}{\pi[\delta(\lambda)]^{-1}\sigma_{\tilde{u}}\sigma_v} exp\left\{-\frac{v^2}{2\sigma_v^2} - \frac{[\delta(\lambda)]^{-2}\tilde{u}^2}{2[\delta(\lambda)]^{-2}\sigma_{\tilde{u}}^2}\right\} \quad (4)$$

Since $\varepsilon = v - [\delta(\lambda)]^{-1}\tilde{u}$, the joint density function for $[\delta(\lambda)]^{-1}\tilde{u}$ and ε becomes:

$$f([\delta(\lambda)]^{-1}\tilde{u}, \varepsilon) = \frac{1}{\pi[\delta(\lambda)]^{-1}\sigma_{\tilde{u}}\sigma_v} exp\left\{-\frac{(\varepsilon + [\delta(\lambda)]^{-1}\tilde{u})^2}{2\sigma_v^2} - \frac{[\delta(\lambda)]^{-2}\tilde{u}^2}{2[\delta(\lambda)]^{-2}\sigma_{\tilde{u}}^2}\right\} \quad (5)$$

Fusco (2017) shows that the marginal density function of ε is obtained by integrating $[\delta(\lambda)]^{-1}\tilde{u}$ out of $f([\delta(\lambda)]^{-1}\tilde{u}, \varepsilon)$, which yields:

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty f([\delta(\lambda)]^{-1}\tilde{u}, \varepsilon)\, du \\ &= \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\Phi\left(-\frac{\gamma\varepsilon}{\sigma}\right) \end{aligned} \quad (6)$$

where $\sigma = \sqrt{\sigma_v^2 + [\delta(\lambda)]^{-2}\sigma_{\tilde{u}}^2}$, $\gamma = \frac{[\delta(\lambda)]^{-1}\sigma_{\tilde{u}}}{\sigma_v}$, $\phi(.)$ is standard normal density function, and $\Phi(.)$ is standard normal distribution function.

The marginal density function $f(\varepsilon)$ is asymmetrically distributed with mean and variance:

$$\begin{aligned} E(\varepsilon) &= -E([\delta(\lambda)]^{-1}\tilde{u}) = -[\delta(\lambda)]^{-1}\sigma_{\tilde{u}}\sqrt{\frac{2}{\pi}} \\ V(\varepsilon) &= \sigma_v^2 + \frac{\pi-2}{\pi}[\delta(\lambda)]^{-2}\sigma_{\tilde{u}}^2 \end{aligned} \quad (7)$$

The log-likelihood function for a sample of n units is given by:

$$ln(L) = \sum_{i=1}^{n}\left\{\frac{1}{2}ln\left(\frac{2}{\pi}\right) - ln(\sigma) + ln\left[\Phi\left(-\frac{\gamma\varepsilon_i}{\sigma}\right)\right] - \frac{\varepsilon_i^2}{2\sigma^2}\right\} \quad (8)$$

In order to obtain estimates of the technical efficiency of each unit, the conditional distribution of $[\delta(\lambda)]^{-1}\tilde{u}$ given ε is computed as:

$$\begin{aligned} f([\delta(\lambda)]^{-1}\tilde{u}|\varepsilon) &= \frac{f([\delta(\lambda)]^{-1}\tilde{u}, \varepsilon)}{f(\varepsilon)} \\ &= \frac{1}{\sqrt{2\pi}\sigma_*} exp\left\{-\frac{([\delta(\lambda)]^{-1}\tilde{u} - \mu_*)^2}{2\sigma_*^2}\right\} \Big/ \left[1 - \Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right] \end{aligned} \quad (9)$$

where $\mu_* = -\frac{\varepsilon[\delta(\lambda)]^{-2}\sigma_{\tilde{u}}^2}{\sigma^2}$ and $\sigma_*^2 = \frac{[\delta(\lambda)]^{-2}\sigma_{\tilde{u}}^2\sigma_v^2}{\sigma^2}$

Finally, technical efficiency (TE) score for each decision making unit can be calculated as below (Fusco, 2017):

$$\begin{aligned} TE_i &= E(exp\{-[\delta(\lambda)]^{-1}\tilde{u}_i\}|\varepsilon_i) \\ &= \left[\frac{1 - \Phi(\sigma_* - \mu_{*i}/\sigma_*)}{1 - \Phi(-\mu_{*i}/\sigma_*)}\right] exp\left\{-\mu_{*i} + \frac{1}{2}\sigma_*^2\right\} \end{aligned} \quad (10)$$

3. Result:

Figure 1 depicts spatial relationships of paddy production among districts in East Java. The figure shows that there is positive spatial autocorrelation of paddy production among districts. Moran's test shows that the spatial autocorrelation is significant with p-value 0.0092. As a result, the residuals of standard SFA are spatially correlated (Figure 2). This implies that traditional efficiency measures using non-spatial stochastic frontier model generate biased and inconsistent results.

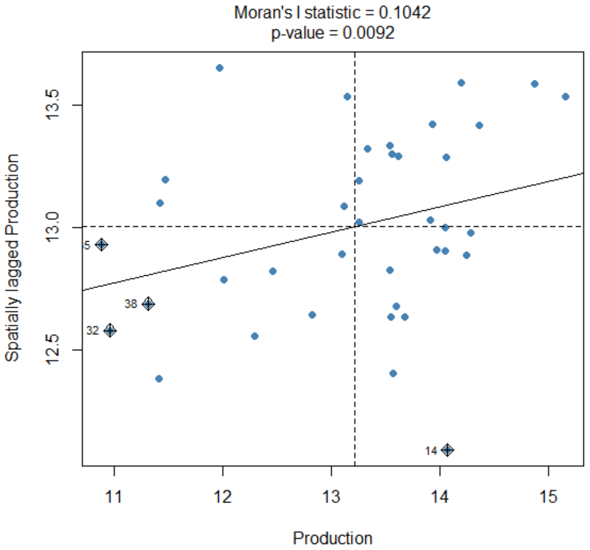

Figure 1. Moran's scatterplot of paddy production in East Java

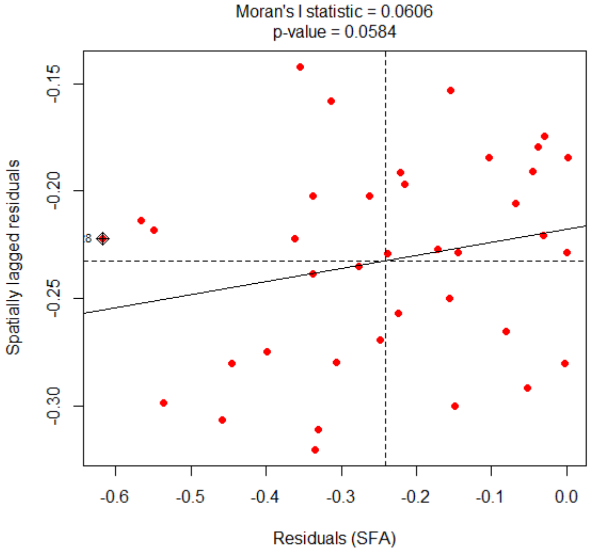

Figure 2. Moran's scatterplot of standard SFA residuals

The spatial autocorrelation issue in standard stochastic frontier analysis (SFA) is resolved by spatial stochastic frontier analysis (SSFA). Figure 3 shows that residuals of SSFA model are no longer dependent on the region. In other words, high or low residuals do not agglomerate in specific region. Figure 4 presents the Moran's scatterplot of SSFA residuals. Moran's test produce p-value 0.1946, thus indicate that there is no spatial autocorrelation in SSFA residuals. Therefore, SSFA model gives more reliable estimates for technical efficiency than standard SFA.

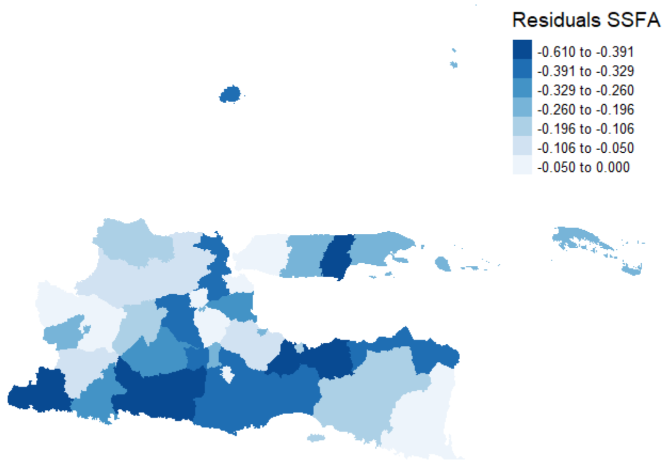

Figure 3. Spatial distribution of SSFA residuals

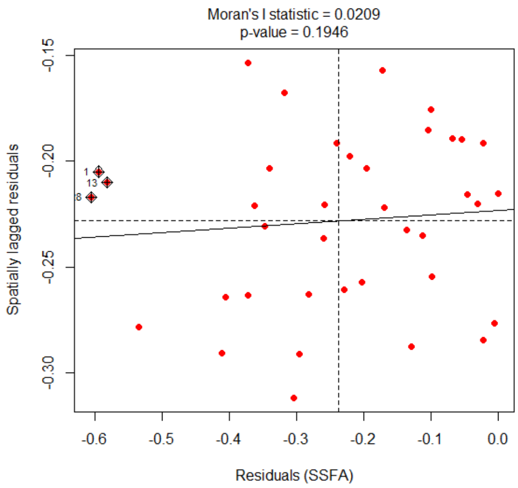

Figure 4. Moran's scatterplot of SSFA residuals

Table 2 presents the estimation results of SSFA compared to standard SFA. Estimation of production function in SSFA model is not substantially different with SFA model, but inefficiency parameter in SSFA model is much larger than that in SFA model. This suggests that the influence of technical inefficiency in SFA model is underestimated. Moreover, SSFA model gives better performance than SFA model, indicated by smaller mean absolute prediction error (MAPE).

Table 2. Estimation results

	SFA		SSFA	
	Coefficien t	Standard error	Coefficient	Standard error
$\boldsymbol{\beta_0}$	1.85033	1.16614	1.52048	0.55261
$\boldsymbol{\beta_1}$ (ln $\boldsymbol{Land}$)	0.46023	0.26741	0.56544	0.09603
$\boldsymbol{\beta_2}$ (ln $\boldsymbol{Seed}$)	0.20884	0.15034	0.09938	0.13321
$\boldsymbol{\beta_3}$ (ln $\boldsymbol{Fertilizer}$)	0.21865	0.10061	0.21332	0.07231
$\boldsymbol{\beta_4}$ (ln $\boldsymbol{Labour}$)	-0.04816	0.09121	-0.02520	0.10707
$\boldsymbol{\beta_5}$ (ln $\boldsymbol{Capital}$)	0.12661	0.04682	0.10925	0.09914
$\boldsymbol{\sigma_u^2}$	0.07055	0.01496	0.05614	0.00542
$\boldsymbol{\sigma_v^2}$	0.00012	0.00014	$8.85x10^{-7}$	$3.33x10^{-6}$
$\boldsymbol{\sigma^2}$	0.08732	-	0.08427	-
$\boldsymbol{\gamma}$ ($\boldsymbol{Inefficiency\ parameter}$)	24.56639	-	63413.1863 0	-
$\boldsymbol{\lambda}$ ($\boldsymbol{Spatial\ lag}$)	-	-	0.45086	-
Log Likelihood	16.20735	-	16.95964	-
MAPE	1.84805	-	1.81096	-

Figure 5 shows the comparison of technical efficiency distribution between SFA and SSFA. There is no extreme technical efficiency score for SFA and SSFA, but there is substantial gap of density for both methods in the range of technical efficiency from 0.75 to 1. SSFA predicts more decision making units having technical efficiency between 0.75 and 1.

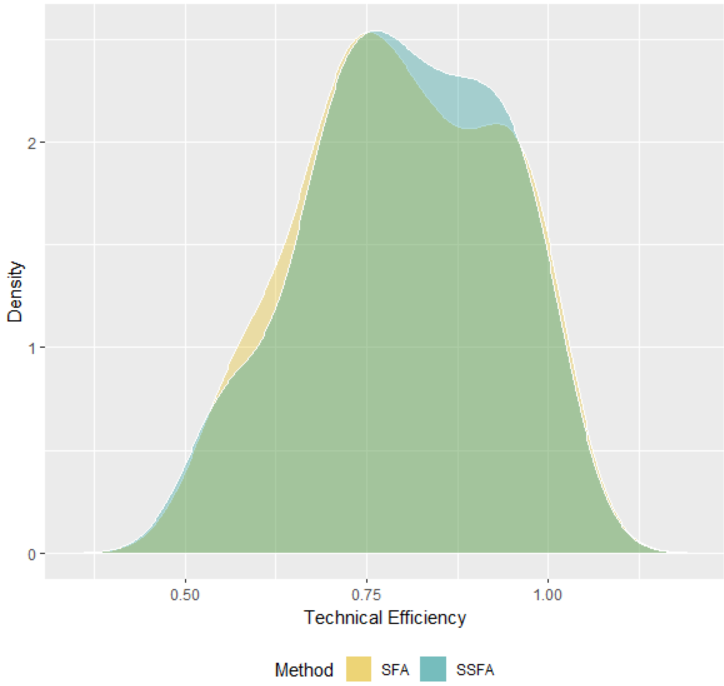

Figure 5. Technical efficiency distribution of SFA and SSFA

Considering spatial lag into stochastic frontier model affects technical efficiency ranking for several districts. Table 3 summaries the shifting of technical efficiency ranking before and after considering spatial effect in three affected districts: Kabupaten Banyuwangi, Kabupaten Lamongan and Kabupaten Bangkalan. These districts experience considerable shifting in their ranking. Rank of Kabupaten Banyuwangi goes up from 8 to 4 after spatial effect is considered. Kabupaten Lamongan experiences more substantial shifting, as its ranking falls from 4 to 10. While for Kabupaten Bangkalan, the rank increases from 9 to 6. A more complete picture of relative position shifting for each districts in East Java are presented as spatial distribution in Figure 6.

Table 3. Technical efficiency ranking for several districts before and after considering spatial effect

Code	Districts	Ranking before considering spatial effect (SFA)	Rangking after considering spatial effect (SSFA)
3510	Kabupaten Banyuwangi	8	4
3524	Kabupaten Lamongan	4	10
3526	Kabupaten Bangkalan	9	6

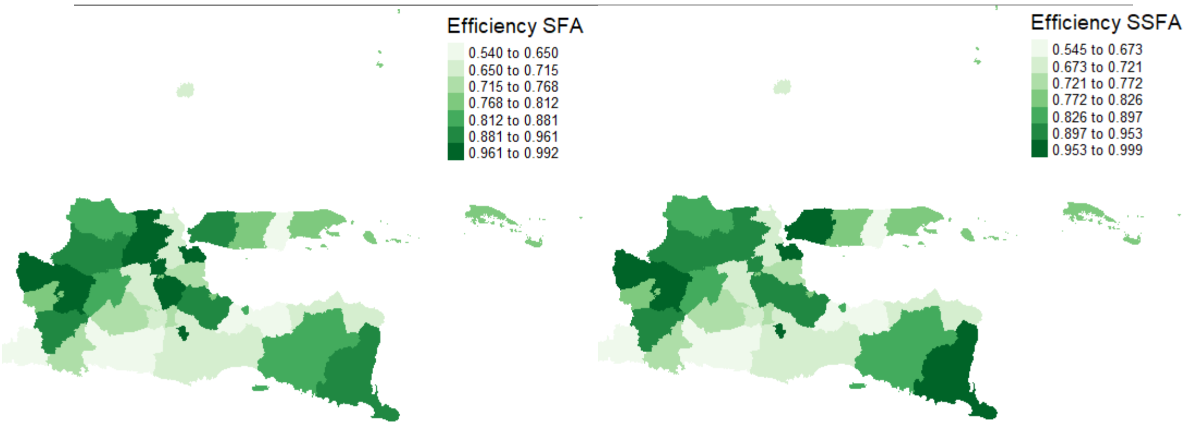

Figure 6. Spatial distribution of technical efficiency for SFA and SSFA

4. Discussion, Conclusion and Recommendations:

The result showed that there is significant spatial dependency of rice farming in East Java. Thus, traditional efficiency measures using stochastic frontier analysis (SFA) generate biased and inconsistent results. Moreover, estimation result suggests that the influence of technical inefficiency in SFA is underestimated. These spatial autocorrelation issues are effectively handled by spatial stochastic frontier analysis (SSFA). Residuals of SSFA model are no longer dependent on the region, indicated by insignificant Moran's test with p-value 0.1946. Therefore, SSFA model gives more reliable estimates for technical efficiency than SFA model. In addition, SSFA model gives better performance than SFA model, indicated by smaller mean absolute prediction error (MAPE).

Considering spatial effect into stochastic frontier model affects the distribution of technical efficiency and its ranking. There are three districts in East Java that experience considerable shifting in their technical efficiency ranking: Kabupaten Banyuwangi, Kabupaten Lamongan and Kabupaten Bangkalan. Other districts are relatively persistent with their technical efficiency ranking even though the spatial effect is considered. Finally, this paper recommends that spatial stochastic frontier model is better used to calculate the technical efficiency since production activities are probable to correlate depending on their geographical proximities. Accurate measures of technical efficiency implies the accurate measures of multifactor productivity.

References:

1. Adetutu M *et al.* (2015). The effects of efficiency and TFP growth on pollution in Europe: a multi stage spatial analysis. J. Product. Anal. 43(3), 307–326. https://doi.org/10.1007/s11123-014-0426-7.
2. Coelli TJ *et al.* (2005). An Introduction to Efficiency and Productivity Analysis. New York: Springer.
3. Fusco E, Vidoli F. (2013). Spatial stochastic frontier models: controlling spatial global and local heterogeneity. Int. Rev. Appl. Econ. 27(5), 679–694. https://doi.org/10.1080/02692171.2013.804493.
4. Fusco E. (2017). New developments in frontier models for objective assessments [Ph.D. thesis]. Rome: Sapienza-University of Rome.
5. Glass AJ *et al.* (2016). A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers. Journal of Econometrics 190(2), 289 – 300.

6. Suphannachart W and Warr P. (2010). Total Factor Productivity in Thai Agriculture. ARE Working paper No. 2553/1 (April 2010).
7. Tsukamoto T. (2019). A spatial autoregressive stochastic frontier model for panel data incorporating a model of technical inefficiency. Japan & The World Economy 50 (2019) 66-77.