Development of the Panel ARDL by Adding Space-Time effect to Modeling Monthly Paddy Producer Price in Java

M Ardiansyah1,2, A Djuraidah2*, I M Sumertajaya2, A H Wigena2, A Fitrianto2

1 BPS-Statistics of Kotawaringin Timur, Central Kalimantan, Indonesia
2 Department of statistics, IPB University, Bogor, Indonesia

*E-mail: anikdjuraidah@apps.ipb.ac.id

Abstract. Classic panel data modelling has large cross section data (N) and small time series data (T). The aim of the study was to develop a panel data type T > N by adding space-time effect to a panel ARDL model. The basic idea was to combine between the AR (Autoregressive), DL (Distributed Lag), and ST (space-time) effect. The model was applied to paddy producer price at the farmer level in Java from January 2016 to December 2019 where the explanatory variable was the Farmers’ Terms of Trade. Both variables were stationary in the first-difference I (1). The results showed that the ST-ARDL model was good for T > N panel data types. The ST-ARDL model with reparameterization of explanatory variables was able to overcome the problem of multicollinearity. The ST-ARDL model was able to improve the performance of the classic panel data model which was able to reduce the RMSE and increased R2-adj. The linear combination of this model was cointegrated or had a long-term equilibrium relationship. Another result of the study was the ST-ARDL model provided better estimation performance than the AR (p), ARDL (p, q) and GSTAR (p, 𝜆) models with the smaller MAPE values. For further research, the ST-ARDL model can be developed by adding the effect of space-time interaction.

1. Introduction
Java Island is a national rice granary in Indonesia. More than 54 percent of Indonesia's rice production is produced by farmers in Java. The provinces of Central Java, East Java and West Java are the three provinces as the largest rice producers in Indonesia with a total production of 9.66 million, 9.58 million and 9.08 million tons of dry milled unhulled rice in 2019 [3]. The highest rice production in 2019 occurred in March and the lowest production occurred in December.

The availability of paddy produced from Oryza Sativa Linaeus affects food security and national economic stability. The pattern of rice planting carried out simultaneously in the growing season caused an oversupply during the main harvest and a scarcity of supplies during famine. The pattern of rice planting in certain seasons has implications for excess supply at harvest and scarcity of supply when starving. The phenomenon of the harvest season causes the paddy price to plummet. As a result, the paddy price is relatively low throughout the harvest season and fluctuate from month to month.

The Government and the National Logistics Agency (Bulog) need information about the causes of the rise and fall of paddy prices. Panel data modeling is needed to provide this information. In addition, the government also requires paddy producer price forecast to adopt domestic pricing policies. An accurate paddy producer price estimate is needed as an anticipatory measure to prevent losses to farmers. Guaranteed stability in paddy purchase prices is expected to protect farmers from lower price. The rise
and fall of paddy prices at the farm level will be followed by an increase in the price of rice. Scarcity of rice causes high paddy prices, so the Bulog needs to get information about the estimated of paddy producer prices.

This research will answer two informations in one model. The first information is the factors that cause paddy producer prices fluctuations. The second information is the estimated paddy producer prices. This study aims to develop a panel data model type $T > N$ by combining between Autoregressive Distributed Lag (ARDL) and Space Time Autoregressive (STAR). Then, we evaluate the performance of the model compared to other models. This model is called the ST-ARDL. The ST-ARDL model combines cross section data and time series data. Merging between cross section data and time series will give more observations. Increasing the number of observations will increase information, so it be able to reduce colarity between variables and increase the degree of freedom which will be able to produce more efficient estimates. The ST-ARDL is a dynamic model which is a combination of the AR (Autoregressive) and DL (Distributed Lag) models by adding ST (space-time) effect. The traditional regression model ignores the time effects but the ST-ARDL model add information of time in the present and previous periods and enter ST effect into model.

The ST-ARDL model is special because it gives information of time, location, and lag time of the response and explanatory variables. We are interested in combining the two into the ST-ARDL model because there has been no research combining ARDL panel models by adding the effects of space time until 2020. The development of this model is expected to improve the performance of the model by increasing the accuracy of parameter estimation.

2. Literature Review

The Autoregressive (AR) model was first introduced by Yule [13]. Model AR (p) can be written as follows:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t$$  \hspace{1cm} (1)

where $y_t$ is the time series variable, $\mu$ is a constant, $\phi_i$ is the AR parameter in the lag of time $i$ with $i = 1, \ldots, p$ and $\epsilon_t$ is an error. Meanwhile, the Autocorrelation function (ACF) is calculated by the formula:

$$\hat{\rho}_k = \frac{\sum_{i=1}^{T} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^{T} (y_i - \bar{y})^2}$$

where $k = 0, 1, 2, \ldots$ and $\sum_{i=1}^{T} (y_i - \bar{y})^2$ is the number of observations.

The Partial Autocorrelation function (PACF) is calculated by the formula:

$$\hat{\phi}_{k+1,k+1} = \hat{\phi}_{k+1,k+1} = \hat{\phi}_{k} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-j} = \frac{\sum_{i=1}^{T} (y_i - \bar{y})(y_{i+k} - \bar{y})}{1 - \sum_{j=1}^{k} \hat{\phi}_j \hat{\phi}_{k+1-j}}$$

where $j = 1, 2, \ldots, k$.

The distributed lag (DL) model was first proposed by Alt at 1942 [12]:

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_k x_{t-k} + \epsilon_t$$  \hspace{1cm} (2)

The idea of combining AR and DL for cointegration was first developed by Pesaran and Shin [9]. The ARDL model is a regression model that combines the Autoregressive (AR) and Distributed Lag (DL) models. AR model is a model where the dependent variable $y_t$ is influenced by the variable itself in the past ($y_{t-j}$). The AR model explains the relationship between observations at a time with observations on the variable itself at previous times. The DL model is a model in which the dependent variable $y_t$ is influenced by the explanatory variable at the present time $x_t$ and the previous time $x_{t-j}$.

The ARDL model has long been discovered but it has only recently been the attention of econometrics experts because of its ability to long-run equilibrium. One of the advantages of the ARDL approach is that it produces parameter estimates that are consistent with good long-term coefficients, and regardless of whether the explanatory variables or regressors are I (0) or I (1). The advantage of the ARDL model is that it does not attach importance to the stationary level of data, although; this model cannot be used in the form of the second differences I (2). Among other advantages, ARDL performs well in small samples and the ARDL method of cointegration analysis is unbiased and efficient [8].

Based on the results of the study by Pesaran and Shin [9], the Ordinary Least Square (OLS) method for estimating short-term parameters in the ARDL framework is consistent $-\sqrt{T}$, and the ARDL-based
long-term coefficient estimator is very consistent in small sample sizes but asymptotically singular of covariance matrix. This model is applied to univariate time series data.

\[ y_t = \alpha_0 + \alpha_1 t + \sum_{j=1}^{p} \phi_j y_{t-j} + \beta' x_t + \sum_{j=0}^{q} \beta_j \Delta x_{t-j} + u_t \]  

(3)

where \( u_t \) is the random disturbance term.

Pesaran et al. [10] developed model (3) from the time series model into a panel data model by adding locations to the model. For example, \( t \) is the time period and \( i \) is the location/group, \( t = 1, \ldots, T; \ i = 1, \ldots, N \), then the ARDL\((p, q, \ldots, q)\) model is

\[ y_{it} = \sum_{j=1}^{p} \lambda_{ij} y_{i,t-j} + \sum_{j=0}^{q} \delta_{ij} x_{i,t-j} + \gamma_{ij} d_t + \varepsilon_{it} \]  

(4)

where \( y_{it} \) is the dependent vector at the \( t \)-th time of the \( i \)-th location with size \( nT \times 1 \), \( x_{it} \) is the explanatory variable at the \( t \)-time location of the \( i \)-th location with size \( nT \times 1 \), \( d_t \) is a design vector such as intercept, time trend, or dummy variable.

Chudik and Pesaran [5] developed model (2) to become Cross-Sectionally ARDL (CS-ARDL) as follows.

\[ y_{it} = c_{yi} + \sum_{j=1}^{p} \psi_{ij} y_{i,t-j} + \sum_{j=0}^{p} \beta_{ij} x_{i,t-j} + \sum_{j=0}^{q} \psi_{ij} z_{i,t-j} + \varepsilon_{it} \]  

(5)

where \( \bar{z}_t = (\bar{y}_t, \bar{x}'_t) \).

The use of the ARDL Panel has become popular since E-Views released a new feature called Pooled Mean Group (PMG) for ARDL models with individual effects. This model is good for panel data with a long amount of time series. Some recent studies using ARDL include: Bardi et al [4] and Magweva and Sibanda [6].

The combination between Space Time (ST) and AR (Autoregression) model is first introduced by Pfeifer and Deutrich [11] called STAR (Space Time Autoregression). STAR is the development of the univariate AR (Autoregression) time series model, into a combination of location and time models. The interrelationship between research locations in the STAR model is expressed by the weight matrix \( W \) which is a square matrix with entries in the form of weights between two corresponding locations. The Cross Correlation function (CCF) is calculated using the formula: 

\[ \hat{\rho}_{xy} = \frac{\sum_{t=1}^{T} (x_{t}-\bar{x})(y_{t+k}-\bar{y})}{\sqrt{\sum_{t=1}^{T} (x_{t}-\bar{x})^2 \sum_{t=1}^{T} (y_{t}-\bar{y})^2}} \]

3. Methods

The method used refers to the two targets of the study. The first target, a panel data model will be built to find out the factors that cause fluctuations in paddy prices at the farm level. The second target, a panel data model will be built for forecasting paddy prices at the farm level. The composition of the data in both are different. The first target uses the data structure of Table 1 and the second target uses the arrangement of data in Table 2.

| \( i \) | \( t \) | \( y_{i,t} \) | \( y_{i,t-1} \) | \( W y_{i,t-1} = \sum_{t=1}^{5} w_{it} y_{i,t-1} \) | \( x_{i,t} \) | \( x_{i,t-1} \) |
|---|---|---|---|---|---|---|
| 1 | 1 | \( y_{1,1} \) | \( y_{1,0} \) | \( 0, y_{1,0} + w_{12} y_{2,0} + \cdots + w_{15} y_{5,0} \) | \( x_{1,1} \) | \( x_{1,0} \) |
| 1 | 2 | \( y_{1,2} \) | \( y_{1,1} \) | \( 0, y_{1,1} + w_{12} y_{2,1} + \cdots + w_{15} y_{5,1} \) | \( x_{1,2} \) | \( x_{1,1} \) |
| 2 | 1 | \( y_{2,1} \) | \( y_{2,0} \) | \( w_{21} y_{1,0} + 0, y_{2,0} + \cdots + w_{25} y_{5,0} \) | \( x_{2,1} \) | \( x_{2,0} \) |
| 2 | 2 | \( y_{2,2} \) | \( y_{2,1} \) | \( w_{21} y_{1,1} + 0, y_{2,1} + \cdots + w_{25} y_{5,1} \) | \( x_{2,2} \) | \( x_{2,1} \) |
The arrangement of data into time series structure is needed to facilitate the forecasting computing. The data structure for the forecasting model can be seen in Table 2.

Table 2. Data structure for the ST-ARDL (1,1,1) forecasting model with N = 5 and T = 48

| t  | y<sub>1,t</sub> | ... | y<sub>5,t</sub> | y<sub>1,t-1</sub> | ... | y<sub>5,t-1</sub> | W<sub>y</sub><sub>1,t-1</sub> = \sum_{i=1}^{5} w_{i1}y_{i,t-1} | ... | x<sub>5,t-1</sub> |
|----|----------------|-----|----------------|-----------------|-----|----------------|-----------------|-----|----------------|
| (1) | (2) | ... | (6) | (7) | ... | (11) | (12) | ... | (26) |
| 1  | y<sub>1,1</sub> | ... | y<sub>5,1</sub> | y<sub>1,0</sub> | ... | y<sub>5,0</sub> | 0.y<sub>1,0</sub> + w<sub>12</sub>y<sub>2,0</sub> + ... + w<sub>15</sub>y<sub>5,0</sub> | ... | x<sub>5,0</sub> |
| 2  | y<sub>1,2</sub> | ... | y<sub>5,2</sub> | y<sub>1,1</sub> | ... | y<sub>5,1</sub> | 0.y<sub>1,1</sub> + w<sub>12</sub>y<sub>2,1</sub> + ... + w<sub>15</sub>y<sub>5,1</sub> | ... | x<sub>5,1</sub> |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 48 | y<sub>1,48</sub> | ... | y<sub>5,48</sub> | y<sub>1,47</sub> | ... | y<sub>5,47</sub> | 0.y<sub>1,47</sub> + w<sub>12</sub>y<sub>2,47</sub> + ... + w<sub>15</sub>y<sub>5,47</sub> | ... | x<sub>1,47</sub> |

3.1. Unit Root

Stationary time series data can be detected by looking at a chart or through a formal test using the Augmented Dickey-Fuller (ADF). The time series approach requires data that is stationary or does not contain random walks or does not have unit roots. If y<sub>t</sub> = \rho y<sub>t-1</sub> + \epsilon<sub>t</sub>, the hypothesis of ADF test is:

H<sub>0</sub>: \rho = 1 (Variable contains unit root (not stationary))

H<sub>1</sub>: \rho \neq 1 (Variable does not contain a root unit (stationary))

The following is ADF statistical value using the t-test statistics.

\[
ADF = \frac{\hat{\rho} - 1}{\sigma_{\hat{\rho}}}
\]

The null hypothesis is accepted when the probability value of the test results is more than the critical value (for example set at 5%). In contrast, the null hypothesis is rejected when the probability value of the test results is less than the critical value. The following is differentiated when a variable contains a root unit (the null hypothesis is accepted).

\[
\Delta y_t = y_t - y_{t-1}
\]

\[
y_t = \Delta y_t + y_{t-1}
\]

First difference is the difference between the observed values in month t and the observed values in previous month t - 1. The value of the difference obtained is checked again whether it is stationary or not. If it is not stationary, then one more differencing is made.
3.2. Cointegration
The idea of cointegration is a number of time series data that deviate from the average value in the short term and tend to move together towards equilibrium in the long run. In other words, if a number of variables have equilibrium in the long run and integrate with one another in the same order, these variables co-integrate. The cointegration approach is one of the time series data solutions which are not stationary. Cointegration testing is carried out on variables to assess whether the regression error has reached stationary or not. To check cointegration, you can use the unit root test (Augmentation test) using the Augmented Dickey-Fuller Test (ADF) for errors of the regression model.

3.3. Identification of the Order of Space-Time
Order of spatial is generally limited to first order because the higher the order, the more difficult to interpret the model. Order lag time is determined by a tentative model based on ACF and PACF plots [7].

| ACF | PACF | Tentative Model |
|-----|------|-----------------|
| Cuts off after lag q | Tails off | MA(q) |
| Tails off | Cuts off after lag p | AR(p) |
| Cuts off after lag q | Cuts off after lag p | MA(q) or AR(p) then choose the best model |
| Tails off | Tails off | ARIMA (p,q) then check on various combinations of p, q |
| Tails off (slowly) | | The model is not stationary, so it needs a differentiation process until the data becomes stationary |

3.4. Location Weighting Matrix
The characteristic of model in spatial time data is the relationship in time and location. The weighting matrix is represented by W and the elements are represented by \( w_{ij} \) where \( i \) denotes the \( i \)-th row and \( j \) denotes the \( j \)-th column. The relationship between locations is represented in the form of a weight matrix. The sum of all elements in each row in the weighting matrix is equal to 1 and the weight of the location itself is zero. In the study determined using a distance inverse location weighting matrix which refers to the actual distance between the provincial capitals. The actual distance is obtained from google map. Distance inverse matrix will produce greater weight in the adjacent location. The farther the distance, the smaller the weight. The following is reverse distance matrix for 5 locations.

\[
W = \begin{bmatrix}
0 & w_{12} & w_{13} & w_{14} & w_{15} \\
w_{21} & 0 & w_{23} & w_{24} & w_{25} \\
w_{31} & w_{32} & 0 & w_{34} & w_{35} \\
w_{41} & w_{42} & w_{43} & 0 & w_{45} \\
w_{51} & w_{52} & w_{53} & w_{54} & 0
\end{bmatrix}
\]

where \( w_{ij} = 1/d_{ij} \) and \( d_{ij} \) are the actual distance between the capital cities of the \( i \)-th and \( j \)-th provinces.
After obtaining a weighting matrix, it is necessary to standardize each row, so that the weighting matrix element becomes \( W^* = \frac{w_{ij}}{\sum w_{ij}} \). Then normalized so that the number of rows is equal to 1.

3.5. Proposed Model (ST-ARDL)
The ST-ARDL model is the development of the ARDL panel model with the addition of space-time effects. The following is ST-ARDL formula (p, p, q).
\[ y_{it} = \alpha_0 + \sum_{k=1}^{p} \theta_k y_{i,t-k} + \sum_{k=1}^{q} \sum_{l=0}^{k} \phi_{kl} W^l y_{i,t-k} + \sum_{j=1}^{r} \sum_{h=1}^{q} \beta_j x_{j,i,t-h} + \varepsilon_{it} \]  

Model (6) is changed into a general form of a linear model to facilitate the computation of parameter estimates.

\[ Y = XB + \varepsilon \]

where \( Y \) is a \( nT \times 1 \) vector of dependent variable; \( X \) contains 1, \( y_{i,t-k} \), \( W^l y_{i,t-k} \), \( x_{j,i,t-h} \); \( \beta = (\alpha_0, \theta, \phi_{kl}, \beta_1, \ldots, \beta_r)^T \); and \( \varepsilon \) is an \( nT \times 1 \) error vector.

### 3.6 Estimating Parameters

Estimating parameters in the ST-ARDL model uses the Ordinary Least Squares (OLS) method. The principle of estimating the parameters is to minimize the sum of squares error. The ST-ARDL model is converted to a linear model

\[ y_{i} = X_i \hat{\beta} + e_i \]

If \( e \sim N(0, \sigma^2I) \).

\[ \hat{\beta}_{\text{ols}} = (X^TX)^{-1}X^Ty; \quad \text{E}(\hat{\beta}_{\text{ols}}) = \beta; \quad \text{Cov}(\hat{\beta}_{\text{ols}}) = \sigma^2(X^TX)^{-1}. \]  

The following are the assumptions in the OLS Estimator.

1. \( e_{it} \) follows the normal distribution because we need to test the parameters;
2. \( \text{Cov}(e_{it}, e_{i(t)}) = 0 \) (non autocorrelation)
3. \( \text{Var}(e_{it}) = \sigma^2 \) (non heteroscedasticity)
4. \( \text{Cor}(X_j, X_i) = 0 \) (non multicollinearity).

Flowchart about the method can be seen in Figure 1.
Figure 1. Flow chart estimating the model ST-ARDL
If the assumptions are not satisfied, it will produce spurious conclusions, so the OLS estimator cannot be used in parameter estimation. Reparameterization and removing the variable can handle multicollinearity. When heteroscedastic, the OLS method is inefficient. The solution is to use the Generalized Least Squares (GLS) method. If \( \varepsilon \sim N(0, \sigma^2 \Omega) \) with \( \Omega \) is a positive definite symmetric matrix of size \( nT \times nT \).

\[
\hat{\beta}_{\text{gls}} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y; \quad \text{E}(\hat{\beta}_{\text{gls}}) = \beta; \quad \text{Cov}(\hat{\beta}_{\text{gls}}) = \sigma^2 (X^T \Omega^{-1} X)^{-1}
\]  

(8)

3.7. Model Comparison

The dataset is divided into training and testing data. This model was built using training data. The time period used in training data is from January 2016 to December 2018. The time period used in testing dataset is from January to December 2019. Then, we forecast the price of paddy as much data in the testing dataset. Forecasting results are compared with the actual data in the testing data. The best model is determined using the minimum of Mean Absolute Percent Error (MAPE).

\[
MAPE = \sum_{t=1}^{n_{\text{testing}}} \left| \frac{y_{it} - \hat{y}_{it}}{y_{it}} \right| \times 100\%
\]  

(9)

The smaller the MAPE value indicates the better the performance of the model.

4. An Empirical Study

The variable of concern is the monthly paddy price at the farm level (Rp / kg) in harvested unhusked rice. The time period used in modeling is from January 2016 to December 2019. The locations in the study were all provinces in Java Island - Indonesia except DKI Jakarta because the scope of the survey of rice producer prices did not cover DKI Jakarta. Price recording is carried out by BPS-Statistics Indonesia officers regularly every month.

Monthly paddy prices are recorded every 10th to 15th of every month. Data obtained from publications that have been released by BPS [1]. Total observations per variable are 5 provinces × 12 months × 4 years. The variables used in modeling and their sources can be seen in Table 4.

| Variables symbol | Variables name | Sources |
|------------------|----------------|---------|
| \( y_{it} \)    | Average Paddy Prices at Farmer Level in the form of unhusked dry rice ready for milling (Rp / kg) | Publication on Statistics of Paddy Producer Price in Indonesia in 2016 until 2019. |
| \( x_{it} \)    | Farmers’ Terms of Trade (FTT) | Publication on Farmers’ Terms of Trade (FTT) of each provinces on the Java Island in 2016 until 2019. |

Based on BPS [2], FTT is an indicator that describes the welfare of farmers which is a comparison between the price index received by farmers and the price index paid by farmers. FTT has a purpose to measure the ability to exchange products sold by farmers with products needed by farmers in household production and consumption. FTT figures indicate the level of competitiveness of agricultural products compared to other products. FTT > 100, means that farmers have a surplus. Production prices rose more than the increase in consumption prices. Farmers' income rises more than their expenses. FTT < 100, means farmers have a deficit.

Average paddy prices at the farm level during the January 2016 period to December 2019 in Java was Rp 4,727 per kg and the median was Rp 4,676 per kg. The median value which is almost the same as the average value indicates the paddy price data follows the normal distribution. The lowest price is Rp 3,670 per kg and the highest price is Rp 6,197 per kg. Fluctuations in paddy prices at the farm level in Java can be seen in Figure 2.
Figure 2. Fluctuations in Paddy Producer Price by provinces in Java for the period of January 2016 to December 2019

Figure 2 shows the fluctuations in paddy prices at the farm level by provinces in Java Island. In March-April, the price of paddy tends to decline in all provinces. This is due to the large harvest that takes place in March-April, so the rice supply is quite abundant which causes the price of paddy to be low. Different things happened at the end of the year, the price of paddy was high because of the lean season and peaked in February. The interesting thing from Figure 2, the price of paddy in 2019 actually tends to be lower than in 2018. This is a challenge in doing forecasting data. The month-to-month price fluctuations shown in Figure 2 show that paddy prices in all provinces on Java are not stationary.

4.1 Unit root test
Unit root is another term for random walk or nonstationary. Nonstationary data can be said to be data that contains random walks or have unit roots. Unit root test is performed on all data from January 2016 to December 2019. The test method used to examine the unit root of the data is the Augmented Dicky-Fuller (ADF) test using a 5% significance level. If the value of ADF is smaller than the critical value, then it can be concluded that the data used is stationary. Stationary test results can be seen in Table 5.

Table 5. Results of Augmented Dicky-Fuller unit root test (ADF)

| Provinces    | Variables | Level | First difference |
|--------------|-----------|-------|------------------|
|              |           | ADF   | p-value          | Inference | ADF   | p-value | Inference |
| Banten       | Y         | -2.9444 | 0.1970         | Not stationary | -3.5183 | 0.0498 | Stationary |
|              | X         | -2.7584 | 0.2712         | Not stationary | -4.2093 | 0.0100 | Stationary |
| West Java    | Y         | -3.2192 | 0.0951         | Not stationary | -3.9053 | 0.0216 | Stationary |
|              | X         | -2.6900 | 0.2986         | Not stationary | -3.5172 | 0.0499 | Stationary |
| Central Java | Y         | -3.4506 | 0.0595         | Not stationary | -3.5327 | 0.0486 | Stationary |
|              | X         | -3.0351 | 0.1607         | Not stationary | -3.5614 | 0.0463 | Stationary |
| DIY          | Y         | -3.4273 | 0.0630         | Not stationary | -3.9109 | 0.0214 | Stationary |
|              | X         | -1.4021 | 0.8129         | Not stationary | -4.6351 | 0.0100 | Stationary |
| East Java    | Y         | -3.0156 | 0.1685         | Not stationary | -3.5223 | 0.0495 | Stationary |
|              | X         | -3.4645 | 0.0573         | Not stationary | -3.5263 | 0.0492 | Stationary |

Table 5 shows that the condition of the paddy price data is not stationary at the level but is stationary at First difference I (1). The same condition also occurs for Farmers’ Terms of Trade which are not stationary at the level but are stationary at First difference I (1).
4.2 Time lag identification

AR model candidates for each location were built based on ACF and PACF plots. The technique of determining tentative models based on ACF and PACF plots refers to Montgomery et al [7]. The tentative model can be seen in Table 6.

**Table 6. Tentative models based on ACF and PACF plots for dependent variable**

| Provinces   | ACF          | PACF           | Tentative models |
|-------------|--------------|----------------|------------------|
| Banten      | Tails off    | Cuts off after lag 2 | AR (2)          |
| West Java   | Tails off    | Cuts off after lag 2 | AR (2)          |
| Central Java| Tails off    | Cuts off after lag 1 | AR (1)          |
| DIY         | Tails off    | Cuts off after lag 2 | AR (2)          |
| East Java   | Tails off    | Cuts off after lag 1 | AR (1)          |

In addition to the ACF and PACF correlogram methods, optimal lag determination can use the minimum Akaike Information Criterion (AIC) values presented in Table 7.

**Table 7. Determining the optimal lag of dependent variable with AIC**

| Lag | AIC       | Banten | West Java | Central Java | DIY | East Java |
|-----|-----------|--------|-----------|--------------|-----|-----------|
| 1   | 11.30771  | 11.66012 | 11.36604  | 11.38275     | 11.40246 |
| 2   | 11.21976* | 11.57843* | 11.33537* | 11.16386*     | 11.39633* |
| 3   | 11.25578  | 11.62778 | 11.37057  | 11.21151     | 11.42730 |
| 4   | 11.30027  | 11.64844 | 11.40370  | 11.19508     | 11.43502 |
| 5   | 11.34837  | 11.70081 | 11.44989  | 11.24713     | 11.48251 |
| 6   | 11.39709  | 11.74294 | 11.49888  | 11.29471     | 11.53015 |
| 7   | 11.26221  | 11.79439 | 11.54032  | 11.31031     | 11.53682 |

The lag selected is the lag model with the minimum AIC to minimize error specifications. We determine lag based on minimum AIC, which is lag 2 for each province. The tentative model for the FTT can be seen in Table 8. Based on Table 8, the tentative model for the FTT variable is AR (1).

**Table 8. Tentative models based on ACF and PACF plots for explanatory variable**

| Provinces   | ACF          | PACF           | Tentative models |
|-------------|--------------|----------------|------------------|
| Banten      | _Tails off_  | _Cuts off after lag 1_ | AR (1)          |
| West Java   | _Tails off_  | _Cuts off after lag 1_ | AR (1)          |
| Central Java| _Tails off_  | _Cuts off after lag 1_ | AR (1)          |
| DIY         | _Tails off_  | _Cuts off after lag 1_ | AR (1)          |
| East Java   | _Tails off_  | _Cuts off after lag 1_ | AR (1)          |

4.3 Location Weighting Matrix Formation

The weighting matrix used in the study is the inverse distance weighting matrix. Determination of inverse distance weighting is based on the actual distance between the provincial capitals in kilometers. The farther the distance from the provincial capital, the smaller the weight. The distance between the provincial capitals can be seen in Table 9.
Table 9. Distance between provincial capitals on Java Island (km)

| Provincial capitals | Serang | Bandung | Semarang | Jogja | Surabaya |
|---------------------|--------|---------|----------|-------|----------|
| Serang              | 0      | 250.8   | 521.4    | 641.3 | 861.4    |
| Bandung             | 250.8  | 0       | 439.8    | 398.67| 779.8    |
| Semarang            | 521.4  | 439.8   | 0        | 130.1 | 350.1    |
| Jogja               | 641.3  | 398.67  | 130.1    | 0     | 324.8    |
| Surabaya            | 861.4  | 779.8   | 350.1    | 324.8 | 0        |

Weighting matrix uses inverse distance weighting. First, we inverse the distance between the provincial capitals. Then, the row is normalized so that the sum of each row is equal to one. The following is the inverse distance matrix.

\[
W = \begin{bmatrix}
0 & 0.462268 & 0.222357 & 0.180784 & 0.134591 \\
0.396672 & 0 & 0.226206 & 0.249543 & 0.127578 \\
0.130166 & 0.154317 & 0 & 0.193854 \\
0.105127 & 0.169107 & 0.521663 & 0 & 0.207567 \\
0.138558 & 0.153057 & 0.340915 & 0.36747 & 0
\end{bmatrix}
\]

4.4 Cointegration Test
Lag for dependent variable was AR(2) and explanatory variable was AR(1), so the cointegration test was performed on errors from the ST-ARDL panel model (2,2,1).

\[
y_{it} = \alpha_0 + \phi_1 y_{i,t-1} + \phi_2 W y_{i,t-2} + \phi_1 W y_{i,t-1} + \phi_1 W y_{i,t-2} + \beta_1 x_{1it} + \beta_2 x_{1i,t-1} + \epsilon_{it}
\]

The statistic of Augmented Dickey-Fuller test for errors from the ST-ARDL model (2,2,1) was -5.4561 with p-value = 0.01. Then \(H_0\) was rejected. It means that the error of the model is stationary, and there is cointegration or long-term relationship in the ST-ARDL panel model (2,2,1). The problem is a multicollinearity between \(x_{1it}\) and \(x_{1i,t-1}\), so we need to reparameterize as follows.

\[
y_{it} = \phi_1 y_{i,t-1} + \phi_2 W y_{i,t-1} + \beta_1 x_{1it} + \beta_2 x_{1i,t-1} + \epsilon_{it}
\]

The linear combination between variables in the model causes cointegration, so the model does not need to be differentiated.

4.5 Model comparison
The parameter estimator using the OLS method have to satisfy four assumptions. There are \(\epsilon_{it}\) is normally distributed because we need to test the parameters, \(\text{Cov}(\epsilon_{it}, \epsilon_{i(t-1)}) = 0\) (non-autocorrelation), \(\text{Var}(\epsilon_{it}) = \sigma^2\) (non heteroscedastic), \(\text{Cor}(X_j, X_j') = 0\) (non multicollinearity). If the assumptions are not satisfied, it will produce a spurious conclusion. The performance of the ST-ARDL Model in handling assumptions violations can be seen in Table 10.

Table 10. Comparison of OLS assumptions, RMSE, and \(R^2 - \text{adj}\) between models

| No | Model                | OLS assumptions | df | RMSE    | \(R^2 - \text{adj}\) |
|----|----------------------|-----------------|----|---------|----------------------|
|    |                      | Normality | Homoscedasticity | Non-multicollinearity | Non-autocorrelation |       |
| 1  | Fixed Effect Panel   | X          | ✓                | ✓                  | X              | 234   | 378.98 | 0.25  |
| 2  | ARDL                 | X          | X                | X                  | ✓              | 231   | 198.82 | 0.79  |
| 3  | ARDL-reparameterization | X       | X                | ✓                  | ✓              | 231   | 198.82 | 0.79  |
It can be seen from Table 10 that the problem of the ARDL model is the violation of the multicollinearity assumption. This needs to be handled by reparameterizing explanatory variables from $x_{it}$ to $\Delta x_{it} = x_{it} - x_{i,t-1}$. It can be seen from Table 10 that the problem of multicollinearity can be overcome after we parameterized explanatory variables.

The ST-ARDL model is able to improve the performance of the classic panel data Fixed Effect model which is able to reduce the RMSE from Rp. 378.98 to Rp.169.58 and increase $R^2_{adj}$ from 0.25 to 0.85. This model is able to handle violations of the normal, homoscedastic, non-multicollinearity, and non-autocorrelation assumptions, so the estimated parameters obtained are reliable. The variable $y_{t-2}$ needs to be eliminated from the model because it has no real effect on the improvement of the model. The non-multicollinear assumptions in model 5 can be seen in Table 11.

### Table 11. Variance Inflation Factor (VIF) values of explanatory variables in the ST-ARDL model with reparameterization and removing the variable $y_{t-2}$

| Explanatory variables | VIF | $\sqrt{\text{VIF}}$ |
|------------------------|-----|---------------------|
| $y_{t-1}$              | 6.64| 2.58                |
| $W y_{t-1}$            | 7.76| 2.79                |
| $W y_{t-2}$            | 3.64| 1.91                |
| $\Delta x_t$          | 1.14| 1.07                |
| $x_{t-1}$              | 2.86| 1.69                |

It can be seen in Table 11 that the ST-ARDL model with reparameterization and removing the variable $y_{t-2}$ no longer contains multicollinearity because the VIF value <10. Normality, homoscedastic, and non-autocorrelated assumptions are presented in Figure 3.
It can be seen in Figure 3 that the ST-ARDL model with reparameterization and removing variables $y_{t-2}$ satisfy assumptions of homoskedastic, non-multicollinearity, non-autocorrelated and normality.

### 4.6 Parameter estimation

After the normality, homoskedastic, non-multicollinearity, and non-autocorrelation is satisfied, estimating the parameters use the OLS method enough. Parameter estimation results can be seen in Table 12.

**Table 12. The results of estimating the parameters of the ST-ARDL model (1,2,1)**

| Variables | Parameter estimation | standard error | t-value | Pr(>|t|) |
|-----------|----------------------|----------------|---------|----------|
| $y_{t-1}$ | 0.493148             | 0.058739       | 8.3956  | 4.771e-15 *** |
| $Wy_{t-1}$ | 0.705832             | 0.071814       | 9.8286  | < 2.2e-16 *** |
| $Wy_{t-2}$ | -0.409787            | 0.048745       | -8.4067 | 4.437e-15 *** |
| $\Delta x_t$ | 120.512545           | 15.392544      | 7.8293  | 1.787e-13 *** |
| $x_{t-1}$  | 21.161629            | 5.491139       | 3.8538  | 0.0001509 *** |

It can be seen from Table 12 that all explanatory variables have a significant effect on paddy price fluctuations at the farm level. The general pattern of the ST-ARDL model for modeling paddy prices in Java from January 2016 to December 2019 can be seen from the equation.

$$\hat{y}_{it} = 0.493148 \, y_{1t-1} + 0.705832 \, \Delta y_{i1t-1} - 0.409787 \, \Delta y_{i2t-2} + 120.512545 \, \Delta x_{it} + 21.161629 \, x_{t-1}$$

### 4.7 Evaluate model performance

We divide the data into two datasets, namely 75% training dataset and 25% testing dataset. The model is built based on training data (January 2016 to December 2018). Forecasting the price of paddy 12 months later (January to December 2019) with the ST-ARDL model that has been formed requires the value of explanatory variables. If we don't have it, forecasting cannot be done with this model. Comparison of the forecast results for January to December 2019 for each province between AR (2), GSTAR (2,1), ARDL (2,1), ST-ARDL (2,2,1) and ST-ARDL models with optimal selection of lag by Eviews.

### 4.8 Forecasting of Paddy Producer Price

#### 4.8.1 Forecasting of Paddy Producer Price in Banten Province

The following is the results of estimating the parameters of the five models in Banten Province.

| Model          | Equation                                                                 |
|----------------|--------------------------------------------------------------------------|
| AR(2)          | $\hat{y}_{it} = 1.216605 \, y_{1t-1} - 0.218688 \, y_{1t-2}$            |
| GSTAR(2,1)     | $\hat{y}_{it} = 0.391409 \, \Delta y_{i1t-1} - 0.108129 \, y_{t-2} + 0.846777 \sum_{s=1}^{5} w_{i} y_{i,t-1} - 0.156543 \sum_{s=1}^{5} w_{i} y_{i,t-2} + 120.512545 \Delta x_{it} + 21.161629 x_{t-1}$ |
| ARDL(2,1)      | $\hat{y}_{it} = 0.945799 \, y_{1t-1} - 0.053027 \, y_{1t-2} + 215.5668 \Delta x_{1t} + 5.215958 x_{t-1}$ |
| ST-ARDL (2,2,1)| $\hat{y}_{it} = 0.694778 \, y_{1t-1} + 0.007815 \, y_{1t-2} + 0.345461 \sum_{s=1}^{5} w_{i} y_{i,t-1} - 0.102193 \sum_{s=1}^{5} w_{i} y_{i,t-2} + 178.7594 \Delta x_{1t} + 2.333097 x_{t-1}$ |
| ST-ARDL optimal lag | $\hat{y}_{it} = -76.35339 + 0.557748 \, y_{1t-1} + 0.494549 \sum_{s=1}^{5} w_{i} y_{i,t-1} - 0.597063 \sum_{s=1}^{5} w_{i} y_{i,t-2} + 0.711676 \sum_{s=1}^{5} w_{i} y_{i,t-3} - 0.169656 \sum_{s=1}^{5} w_{i} y_{i,t-4} + 184.0379 x_{1t} - 108.4533 x_{1t-1} - 75.02866 x_{1t-2}$ |
After estimating the parameters, forecasting is carried out in the next 12 months and compared with the actual paddy price data for January to December 2019. Comparison of MAPE between models in Banten can be seen in Table 13.

Table 13. Comparison of MAPE between models in Banten Province

| Months      | y actual | AR(2) | GSTAR(2,1) | ARDL(2,1) | ST-ARDL (2,2,1) | ST-ARDL optimal lag |
|-------------|----------|-------|------------|-----------|-----------------|---------------------|
| January 2019| 5600.00  | 5369.13| 5303.52    | 5262.75   | 5295.13         | 5251.32             |
| February 2019| 5100.00  | 5359.95| 5356.15    | 5293.95   | 5385.69         | 5339.51             |
| March 2019   | 4477.14  | 5346.78| 5363.175   | 5145.09   | 5282.17         | 5295.17             |
| April 2019   | 3977.88  | 5332.76| 4806.01    | 5062.56   | 5035.47         | 4999.76             |
| May 2019     | 3900.00  | 5318.59| 4171.88    | 5036.57   | 4755.96         | 4962.07             |
| June 2019    | 3969.74  | 5304.41| 4028.39    | 4772.77   | 4399.42         | 4618.12             |
| July 2019    | 4047.37  | 5290.26| 4137.59    | 4863.66   | 4474.71         | 4398.36             |
| August 2019  | 4552.27  | 5276.15| 4471.85    | 5140.68   | 4778.77         | 4714.87             |
| September 2019| 4896.43 | 5262.08| 4763.49    | 5464.89   | 5100.85         | 5014.46             |
| October 2019 | 4817.65  | 5248.04| 4921.98    | 5575.29   | 5195.91         | 5181.18             |
| November 2019| 4875.00  | 5234.04| 4967.29    | 5465.57   | 5100.31         | 5104.88             |
| December 2019| 5069.57  | 5220.08| 5034.39    | 5529.26   | 5204.29         | 5215.41             |
| MAPE        | 17.17    | 5.90  | 15.25      | 10.27     | 10.69           |                     |

4.8.2 Forecasting of Paddy Producer Price in West Java

The following is the results of estimating the parameters of the five models in West Java Province.

AR(2) : \[ \hat{y}_{1t} = 1.229613y_{1t-1} - 0.231544y_{1t-2} \]

GSTAR(2,1) : \[ \hat{y}_{1t} = 0.231321y_{1t-1} + 0.155258y_{1t-2} + 1.203257 \sum_{i=1}^{5} w_iy_{1t-i} - 0.574176 \sum_{i=1}^{5} w_iy_{1t-i} \]

ARDL(2,1) : \[ \hat{y}_{1t} = 0.862294y_{1t-1} - 0.090100y_{1t-2} + 245.5285\Delta x_{1t} + 9.972464x_{1t-1} \]

ST-ARDL (2,2,1) : \[ \hat{y}_{1t} = 0.293655y_{1t-1} + 0.084690y_{1t-2} + 0.801492 \sum_{i=1}^{5} w_iy_{1t-i} - 0.290583 \sum_{i=1}^{5} w_iy_{1t-i} + 163.6895\Delta x_{1t} + 5.415207x_{1t-1} \]

ST-ARDL optimal lag : \[ \hat{y}_{1t} = -895.9562 + 0.007648y_{1t-1} - 0.059584y_{1t-2} + 0.185488y_{1t-3} - 0.244285y_{1t-4} + 0.863964 \sum_{i=1}^{5} w_iy_{1t-i} - 187.2228x_{1t} - 166.9067x_{1t-1} \]

Forecasting is carried out in the next 12 months after estimating the parameters and compared with the actual paddy price data for January to December 2019. Comparison of MAPE between models in West Java can be seen in Table 14.

Table 14. Comparison of MAPE between models in West Java Province

| Months      | y actual | AR(2) | GSTAR(2,1) | ARDL(2,1) | ST-ARDL (2,2,1) | ST-ARDL optimal lag |
|-------------|----------|-------|------------|-----------|-----------------|---------------------|
| January 2019| 5789.2   | 5355.78| 5576.83    | 5399.07   | 5556.84         | 5655.34             |
| February 2019| 5725.45  | 5349.18| 5700.76    | 5254.99   | 5554.89         | 5607.33             |
| March 2019   | 5092.68  | 5337.33| 5349.55    | 4785.27   | 5079.42         | 5080.10             |
| April 2019   | 4330.71  | 5324.28| 4719.12    | 4694.89   | 4737.35         | 4792.79             |
| May 2019     | 4199.79  | 5310.98| 4256.95    | 4706.01   | 4397.68         | 4329.19             |
| June 2019    | 4507.83  | 5297.65| 4287.15    | 4691.96   | 4357.33         | 4190.83             |
The following is the results of estimating the parameters of the five models in Central Java Province.

Forecasting of Paddy Producer Price in Central Java

The following is the results of estimating the parameters of the five models in Central Java Province.

| Months      | y actual | AR(2) | GSTAR(2,1) | ARDL(2,1) | ST-ARDL (2,2,1) | ST-ARDL optimal lag |
|-------------|----------|-------|------------|-----------|----------------|---------------------|
| January 2019| 5480.39  | 5526.51| 5526.80    | 5418.26   | 5422.62        | 5442.57             |
| February 2019| 5001.42 | 5558.63| 5583.07    | 5145.92   | 5226.12        | 5240.49             |
| March 2019   | 4550.00 | 5586.75| 5304.31    | 5057.99   | 5013.74        | 4962.82             |
| April 2019   | 4138.35 | 5614.27| 4938.15    | 4895.05   | 4625.39        | 4866.89             |
| May 2019     | 4191.88 | 5641.80| 4352.96    | 5049.04   | 4359.08        | 4615.30             |
| June 2019    | 4345.33 | 5669.45| 4106.07    | 4972.72   | 4093.60        | 4628.43             |
| July 2019    | 4482.22 | 5697.22| 4225.90    | 5017.02   | 4380.15        | 4628.75             |
| August 2019  | 4750.81 | 5725.14| 4442.05    | 5170.77   | 4813.96        | 4879.07             |
| September 2019| 5048.47| 5753.18| 4685.57    | 5303.37   | 5069.89        | 5323.08             |
| October 2019 | 5096.65 | 5781.37| 4864.42    | 5312.07   | 5038.66        | 5540.60             |
| November 2019| 5111.11| 5809.69| 4988.81    | 5204.57   | 4915.92        | 5526.26             |
| December 2019| 5302.22| 5838.15| 5075.71    | 5138.75   | 4966.33        | 5425.60             |

**Table 15. Comparison of MAPE between models in Central Java Province**

| Months      | MAPE   | ST-ARDL (2,2,1) | ST-ARDL optimal lag |
|-------------|--------|----------------|---------------------|
| June 2019   | 4.09   | 4.32           | 6.61                |
| July 2019   | 5.05   | 5.05           | 4.78                |
| August 2019 | 4.85   | 4.78           | 4.32                |

Forecasting of Paddy Producer Price in DI Yogyakarta

The following is the results of estimating the parameters of the five models in DI Yogyakarta Province.
After estimating the parameters, forecasting is carried out in the next 12 months and compared with the actual paddy data for January to December 2019. Then MAPE was calculated and compared between five models. The smallest MAPE value indicates better model performance. The comparison of MAPE between models in DIY can be seen in Table 16.

**Table 16. Comparison of MAPE between models in DI Yogyakarta Province**

| Months      | y actual | AR(2) | GSTAR(2,1) | ARDL(2,1) | ST-ARDL (2,2,1) | ST-ARDL optimal lag |
|-------------|----------|-------|------------|-----------|-----------------|---------------------|
| January 2019 | 5742.00  | 5792.12 | 5865.28   | 5757.25   | 5844.24         | 5710.41             |
| February 2019 | 5366.67  | 5829.90 | 5961.90   | 5655.80   | 5825.74         | 5679.54             |
| March 2019   | 5334.21  | 5858.77 | 5933.80   | 5561.99   | 5770.90         | 5547.55             |
| April 2019   | 4580.23  | 5884.69 | 5121.20   | 5411.33   | 4817.71         | 4731.55             |
| May 2019     | 4356.00  | 5909.69 | 4821.34   | 5397.97   | 4567.94         | 4421.95             |
| June 2019    | 4653.13  | 5934.45 | 4841.37   | 5350.85   | 4625.84         | 4537.95             |
| July 2019    | 4746.15  | 5959.19 | 4879.12   | 5357.99   | 4813.98         | 4670.26             |
| August 2019  | 4837.14  | 5983.99 | 5010.94   | 5404.16   | 5060.12         | 4841.44             |
| September 2019 | 4900.00 | 6008.88 | 5232.05   | 5418.86   | 5305.39         | 5063.73             |
| October 2019 | 5147.92  | 6033.87 | 5425.30   | 5425.73   | 5515.22         | 5224.71             |
| November 2019 | 5262.50 | 6058.96 | 5418.23   | 5380.50   | 5433.95         | 5174.76             |
| December 2019 | 4931.03 | 6084.16 | 5466.95   | 5336.17   | 5429.14         | 5163.59             |

MAPE: 19.89 6.95 9.84 5.33 2.55

4.8.5 Forecasting of Paddy Producer Price in East Java

The following is the results of estimating the parameters of the five models in East Java Province.

AR(2) : \( \hat{y}_{1t} = 1.342020y_{1t-1} - 0.339267y_{1t-2} \)
GSTAR(2,1) : \( \hat{y}_{1t} = 0.663197y_{1t-1} + 0.150511y_{1t-2} + 0.888825\sum_{i=1}^{5}w_{i}y_{1t-i} - 0.686832\sum_{i=1}^{5}w_{i}y_{1t-2} \)
ARDL(2,1) : \( \hat{y}_{1t} = 1.316613y_{1t-1} - 0.455903y_{1t-2} + 53.23857\Delta x_{1t} + 6.820867x_{1t-1} \)
ST-ARDL (2,2,1) : \( \hat{y}_{1t} = 0.606471y_{1t-1} - 0.018131y_{1t-2} + 0.914662\sum_{i=1}^{5}w_{i}y_{1t-i} - 0.611473\sum_{i=1}^{5}w_{i}y_{1t-2} + 61.47344\Delta x_{1t} + 6.240253x_{1t-1} \)
ST-ARDL optimal lag : \( \hat{y}_{1t} = 1.561.519 + 0.696140y_{1t-1} + 0.805443\sum_{i=1}^{5}w_{i}y_{1t-i} - 0.653394\sum_{i=1}^{5}w_{i}y_{1t-2} - 6.176097x_{1t} \)

After estimating the parameters, forecasting of Paddy Producer Price is carried out in the next 12 months and compared with the actual paddy data for January to December 2019. Then MAPE was calculated and compared between five models. The minimum of MAPE value indicates better model performance. The comparison of MAPE between models in East Java can be seen in Table 17.
Table 17. Comparison of MAPE between models in East Java Province

| Months       | y actual | AR(2)  | GSTAR(2,1) | ARDL(2,1) | ST-ARDL (2,2,1) | ST-ARDL optimal lag |
|--------------|----------|--------|------------|-----------|-----------------|---------------------|
| January 2019 | 5421.08  | 5261.14| 5393.38    | 5043.54   | 5043.54         | 5043.54             |
| February 2019| 5003.32  | 5271.47| 5434.25    | 4651.70   | 5043.54         | 5043.54             |
| March 2019   | 4300.57  | 5278.64| 5393.38    | 5043.54   | 5043.54         | 5043.54             |
| April 2019   | 4067.79  | 5286.12| 4491.14    | 4543.09   | 5043.54         | 5043.54             |
| May          | 4280.54  | 5293.59| 4725.64    | 4625.71   | 5043.54         | 5043.54             |
| June         | 4394.19  | 5301.06| 4725.64    | 4625.71   | 5043.54         | 5043.54             |
| July         | 4483.01  | 5308.55| 4725.64    | 4625.71   | 5043.54         | 5043.54             |
| August       | 4722.19  | 5316.04| 4725.64    | 4625.71   | 5043.54         | 5043.54             |
| September    | 5019.96  | 5323.55| 4725.64    | 4625.71   | 5043.54         | 5043.54             |
| October      | 5022.43  | 5331.07| 4725.64    | 4625.71   | 5043.54         | 5043.54             |
| November     | 5027.94  | 5338.60| 4725.64    | 4625.71   | 5043.54         | 5043.54             |
| December     | 4997.74  | 5346.13| 4725.64    | 4625.71   | 5043.54         | 5043.54             |

| MAPE         | 13.47    | 7.07   | 6.43       | 4.93      | 4.04            |

From Tables 13 to 17, a brief comparison of MAPE values between provinces in Java Island can be seen in Table 18.

Table 18. Comparison of MAPE values between provinces in Java

| Provinces    | AR(2) | GSTAR(2,1) | ARDL(2,1) | ST-ARDL (2,2,1) | ST-ARDL optimal lag |
|--------------|-------|------------|-----------|-----------------|---------------------|
| Banten       | 17.17 | 5.90       | 15.25     | 10.27           | 10.69               |
| Jawa Barat   | 8.50  | 5.05       | 4.78      | 4.32            | 4.09                |
| Jawa Tengah  | 19.52 | 7.36       | 8.61      | 4.38            | 6.61                |
| DIY          | 19.89 | 6.95       | 9.84      | 5.33            | 2.55                |
| Jawa Timur   | 13.47 | 7.07       | 6.43      | 4.93            | 4.04                |
| Average of MAPE | 15.71 | 6.47       | 8.98      | 5.85            | 5.60                |

Table 18 shows that the ST-ARDL model has better performance than the AR (2), GSTAR (2,1), and ARDL (2,1). It can be seen that the average of MAPE for the ST-ARDL model is lower than other models. Therefore, the ST-ARDL model can be said to be good for modeling paddy prices at farmer level in Java.

5. Conclusion

The ST-ARDL Panel Model is a dynamic model which is a combination of the AR (Autoregressive) and DL (Distributed Lag) models with adding the effect of ST (space-time). The ST-ARDL model is applied to modeling paddy prices in Java. The condition of the paddy price data is not stationary at the level, but it is stationary at the first difference I(1). The same condition also occurs for Farmers' Terms of Trade (FTT) which are not stationary at the level but are stationary at the first difference I(1). Determination of the lag of paddy price variable is based on the smallest AIC, which follows AR (2) for each provinces while the FTT variable is AR (1). Therefore, the appropriate model is ST-ARDL (2,2,1).

The weighting matrix used is the inverse distance weighting matrix. The farther the distance from the provincial capital, the less influence. Based on the results of the Augmented Dickey-Fuller test, errors from the ST-ARDL model (2,2,1) are stationary, so there is cointegration or long-term relationships. The problem is a multicollinearity between $x_{i,t}$ and $x_{i,t-1}$, so it needs to be handled by reparameterizing the explanatory variable from $x_{i,t} \to \Delta x_{i,t} = x_{i,t} - x_{i,t-1}$. The ST-ARDL model is able to handle the violations of OLS assumptions and provide a smaller RMSE, so the model is good for $T > N$ type panel data.
The ST-ARDL model can improve the performance of the classic panel data model by reducing the RMSE from Rp. 378.98 to Rp.169.58 and increase R^2 – adj from 0.25 to 0.85. The linear combination of the model is cointegrated or has a long-term equilibrium relationship. The ST-ARDL model provides better forecasting performance than the AR (p), ARDL (p,q), and GSTAR (p, λ) models as evidenced by the smaller MAPE values. The average MAPE model AR (2) is 15.71%, the GSTAR model (2,1) is 6.47%, the ARDL model (2,1) is 8.98%, and the ST-ARDL (2,2,1) is 5.85%. The weakness of this model is that it requires explanatory variable data for forecasting. For further research development, the effect of space-time interaction can be added to the ST-ARDL model and examining seasonal patterns of rice prices.

Acknowledgments

The authors thank the Editor and the Associate Editor for their helpful comments and constructive suggestions, which lead to improvement in the quality of the paper. This research was supported by the BPS-Statistics Indonesia Doctoral Scholarship Program and Department of Statistics, IPB University.

References

[1] BPS 2016-2019a Publikasi Statistik Harga Produsen Gabah di Indonesia tahun 2016, 2017, 2018, and 2019.
[2] BPS 2016-2019b Nilai Tukar Petani Provinsi Banten, Jawa Barat, Jawa Tengah, DIY, and Jawa Timur tahun 2016-2019.
[3] BPS 2020 Luas Panen dan Produksi Padi di Indonesia 2019 (Hasil Survei Kerangka Sampel Area) Berita Resmi Statistik. No. 16/02/Th. XXIII, 4 Februari 2020.
[4] Bardi W, Ayouni SE, Hamdaou M 2019 Are structural policies in countries bordering mediterranean appropriate to economic convergence: Panel ARDL application. Cogent Economics & Finance (Taylor and Francis Group) 7(163649). pp 2-20. doi.org/10.1080/23322039.2019.1636496.
[5] Chudik A. and Pesaran MH 2015 Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors. Journal of Econometrics 188 (2). pp 393-420.
[6] Magweva R and Sibanda M 2020 Inflation and infrastructure sector returns in emerging markets: panel ARDL approach. Cogent Economics and Finance (2020) 8(730078). pp 1-17 https://doi.org/10.1080/23322039.2020.1730078.
[7] Montgomery DC, Jennings CL, Kulahci M 2008 Introduction to Time Series Analysis and Forecasting. Canada (CA): Wiley Interscience.
[8] Narayan PK and Narayan S 2006 Savings behaviour in Fiji: An empirical assessment using the ARDL approach to cointegration International Journal of Social Economics. pp 1-28. DOI: 10.1108/03068290610673243
[9] Pesaran MH and Shin Y 1995 An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis. Cambridge Working Papers in Economics 9514. Faculty of Economics. University of Cambridge.
[10] Pesaran MH, Shin Y and Smith RP 1999 Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. Journal of the American Statistical Association. 94(446). pp 621-634. DOI: 10.1080/01621459.1999.10474156
[11] Pfeifer PE and Deutsch S J 1980 A Three-Stage Iterative Procedure for Space-Time Modeling. Technometrics. 22(1). pp 35-47. DOI: 10.1080/00401706.1980.10486099
[12] Pollock S 2014 Econometrics: An Historical Guide for the Uninitiated. Departemen of Economics. University of Leicester. Working Paper No 14/05.
[13] Yule, G.U 1927 On a method of investigating the periodicities of disturbed series, with special reference to Wolfers sunspot numbers. *Philosophical Transactions of the Royal Society (A).* 226. pp 267-298