A Panel Data Analysis of Rice Production in Ngawi Regency, East Java

Dina Tri Utari¹,*, Tyas Yuliana¹, Andrie Pasca Hendradewa²

¹Department of Statistics, Universitas Islam Indonesia
²Department of Industrial Engineering, Universitas Islam Indonesia
* Corresponding author: dina.t.utari@uii.ac.id

ABSTRACT

Agriculture is one of the key sectors of the economy in Indonesia. Among some subsectors of agriculture, food crops is an important sub-sectors in the development of Indonesia. Widely consumed food crops are rice in Indonesian society. Based on data from the Ministry of Agriculture, East Java, Indonesia is the largest rice producers with rice production reached 13.13 million tons or 16.1% of total national production in 2016. Among the total 29 districts and 9 cities in East Java, the Ngawi district that ranks as the fifth largest rice producing areas in East Java. This study focused on identifying the influence of harvested area (X₁), productivity (X₂), machinery/agriculture (X₃), and extensive irrigation (X₄) on rice production in Ngawi. The analysis uses panel data regression. We obtained the Fixed Effect Model with R² values of 99.59% as the best model. The Fixed Effect Model interprets that every district has a different intercept without the effects of time. Additionally, the factors affecting rice production are harvested area and productivity.

Keywords: rice production, panel data regression, time series, cross section.

1. INTRODUCTION

As an agricultural country, the agricultural sector in Indonesia has become one of the largest economic contributors. Among some subsectors of agriculture, food crop is an important sub-sectors in the development of Indonesia. As stated in Law No. 18 the Year 2012 on Food, said that food is the most important basic human needs, and its fulfilment is a fundamental right of every people of Indonesia. Based on data from the Ministry of Agriculture, East Java is the largest rice producer with 13.13 million tons of rice or 16.1% of total national rice production. Among the total 29 districts and 9 cities in East Java, the Ngawi district that ranks as the fifth largest rice producing areas in East Java. This study focused on identifying the influence of harvested area (X₁), productivity (X₂), machinery/agriculture (X₃), and extensive irrigation (X₄) on rice production in Ngawi. The analysis uses panel data regression. We obtained the Fixed Effect Model with R² values of 99.59% as the best model. The Fixed Effect Model interprets that every district has a different intercept without the effects of time. Additionally, the factors affecting rice production are harvested area and productivity.

2. MATERIALS AND METHODOLOGY

The data used in this study is a combination of time series and cross section. Time series data used are annual data in Ngawi, while the number of cross section data is 19 as the number of sub-districts in Ngawi. The data is classified as secondary data that obtained from the publication entitled “Kabupaten Ngawi dalam Angka” from 2006 to 2014 [2].

2.1. Descriptive Statistics

Descriptive statistics are methods which related to the collection and presentation of a range of data to provide useful information [3]. Descriptive statistics only provides information overview on the data and does not obtain any inference or conclusion of a larger parent group. Examples of descriptive statistics are often arises as tables, charts, graphs, and other quantities. By using descriptive statistics, the collection of data obtained will be presented with a quick and tidy and can provide the core information from existing data set. The information can be obtained from descriptive statistics include measures of central tendency of data, size of data dissemination, as well as the tendency of a data range [4].
2.2. Panel Data Regression Models

According to Wanner & Pevalin [5], the panel data regression is a set of techniques for modeling the influence of the independent variable on the dependent variable in the data panel.

2.3. Common Effect Model

Common Effect Model is one of the simplest panel data approaches [6]. It assumed that there is no difference in the value of the intercept and slope of regression good results on the basis of differences between individuals and between time. Parameter estimation methods on a Common Effect Model require the technique of adding dummy variables or Least Square Dummy Variable (LSDV). There are three approaches contained in the Fixed Effects Model [7] as follows:

\[ Y_{it} = \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + \cdots + u_{it} \]

1. Differences intercept on individual stocks:

\[ Y_{it} = \alpha_1 + \alpha_2 D_{2it} + \alpha_3 D_{3it} + \alpha_4 D_{3it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \cdots + u_{it} \]

2. Differences intercept at the time effect:

\[ Y_{it} = \lambda_0 + \lambda_2 D_{2it} + \lambda_3 D_{3it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \cdots + u_{it} \]

3. Differences intercept on individual stocks and time:

\[ Y_{it} = \alpha_1 + \alpha_2 D_{2it} + \alpha_3 D_{3it} + \alpha_4 D_{3it} + \lambda_2 D_{2it} + \lambda_3 D_{3it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \cdots + u_{it} \]

with,

\( Y_{it} \) : Variable dependent on the individual units of the i-th and time to t
\( \beta_1 \) : Constants
\( \beta_2 \) : Variable coefficient \( X_{2it} \)
\( \beta_3 \) : Variable coefficient \( X_{3it} \)
\( X_{2it} \) : Variable 2nd in the individual all the time i and all t
\( X_{3it} \) : Variable 3rd in the individual all the time i and all t
\( u_{it} \) : Residual/error for individual i-th and time to t

2.4. Fixed Effect Model

Estimation methods of a panel data regression on Fixed Effect Model require the technique of adding dummy variables. There are three approaches contained in the Fixed Effects Models [7] as follows:

\[ Y_{it} = \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + \cdots + w_{it} \]

with,

\( Y_{it} \) : Variable dependent on the individual units of the i-th and time to t
\( \beta_1 \) : Constants
\( \beta_2 \) : Variable coefficient \( X_{2it} \)
\( \beta_3 \) : Variable coefficient \( X_{3it} \)
\( X_{2it} \) : Variable 2nd in the individual all the time i and all t
\( X_{3it} \) : Variable 3rd in the individual all the time i and all t
\( w_{it} \) : Residual/error for individual i-th and time to t

3. RESULTS AND DISCUSSION

3.1. Descriptive analysis

According to the result, it showed that the variable \( X_1 \) (harvest area) has an average of 5868.164 tons/ha, the total harvested area is 1003456 tons/ha, for the highest value of the harvested area is 14747 tons/ha and a low of 1120 tons/ha. For \( X_2 \) (productivity) is the average of 5.96 ku/ha, for the highest productivity value was 7.16 ku/ha and productivity low of 3.85 ku/ha. For \( X_3 \) (tools/agricultural machinery) on average amounted to 142 pieces, for the highest number is 497 pieces and the lowest 0. For variable \( X_4 \) (extensive irrigation) the average extent of extensive irrigation) the average extent of extensive irrigation is 35 ha, 98756 tons the highest production yield and lowest production yield 6038 tons.
3.2. Estimation Model

Table 1. Rated R-Square Each Model

<table>
<thead>
<tr>
<th>Model</th>
<th>R-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Effect</td>
<td>99.43%</td>
</tr>
<tr>
<td>Fixed Effect</td>
<td>99.64%</td>
</tr>
<tr>
<td>Random Effect</td>
<td>98.61%</td>
</tr>
</tbody>
</table>

On the Common Effect Model, independent variables are obtained that land crops, rice productivity, tools/farm machinery and irrigating vast influence on the dependent variable (rice production) amounted to 99.43%. Similarly, for the Fixed Effect Model and Random Effect Model to give effect to rice production variables, respectively 99.64% and 98.61%.

3.3. Selection of the Best Model

3.3.1. Chow test

Chow test is used to select the panel data regression models suitable for use between Common Effect Model or Fixed Effect Model. In the test the following hypotheses:

\[ H_0 : \text{Common Effect Model} \]

\[ H_1 : \text{Fixed Effect Model} \]

The Chow test calculation result is shown on Fig.1 below:

The basis for rejection of the above hypothesis is to compare the probability value F (p-value) with alpha (0.05). Based on the results of the calculation, the p-value is 0.000, so that the p-value < alpha (0.05) is obtained, which means the decision is to reject H_0, in the other word the chosen model is the Fixed Effect Model.

3.3.2. Hausman test

Hausman test is used to select the panel data regression models are suitable for use between Random Effect Model or Fixed Effect Model. In testing the hypothesis as follows:

\[ H_0 : \text{Random Effect Model} \]

\[ H_1 : \text{Fixed Effect Model} \]

The Hausman test calculation result is shown on Fig.2 below:

The basis for rejection of the above hypothesis is to compare the probability value F (p-value) with alpha (0.05). Based on the results of the calculation, the p-value is 0.000, so that the p-value < alpha (0.05) is obtained, which means the decision is to reject H_0, in the other word the chosen model is the Fixed Effect Model. Since the best model chosen in the Chow and Hausman test is the Fixed Effect Model, then the Lagrange Multiplier Test is not necessary.

3.4. Classical Assumptions

3.4.1. Multicollinearity

Multicollinearity test is used to determine whether there is a perfect linear relationship between the variables that explain the regression model. His hypothesis was as follows:

\[ H_0 : \text{There is a multicollinearity problem} \]
H1: There is no multicollinearity problem
The results of this test indicated that there is a multicollinearity condition; therefore, it must be addressed by eliminating insignificant variables using the regression test. The insignificant variable from the regression test is \( X_4 \); consequently, the variable \( X_4 \) is omitted, so that the results of multicollinearity test is shown below:

<table>
<thead>
<tr>
<th>Table 2. Correlation between Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_1 ) &amp; 1 &amp; 0.357183 &amp; 0.494559</td>
</tr>
<tr>
<td>( X_2 ) &amp; 0.357183 &amp; 1 &amp; 0.276194</td>
</tr>
<tr>
<td>( X_3 ) &amp; 0.494559 &amp; 0.276194 &amp; 1</td>
</tr>
</tbody>
</table>

It is shown on Table 2 above that after excluded variable \( X_4 \), and there is no multicollinearity among all variables since the correlation among variables is less than 0.9 [8]. In other word, the condition already met Base Linear Unbiased Estimator (BLUE) assumption.

### 3.4.2. Heteroscedasticity

Heteroscedasticity test is used to determine the value of any residual variance. A good regression model is a residual value that appears in the regression model populations have the same variance or homoscedasticity. Testing heteroscedasticity on panel data analysis using Glejser test. The hypothesis is as follows:

- \( H_0 \): There was no trouble heteroscedasticity
- \( H_1 \): There was a problem heteroscedasticity

<table>
<thead>
<tr>
<th>Table 3. Heteroscedasticity Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Harvested Area</td>
</tr>
<tr>
<td>Productivity</td>
</tr>
<tr>
<td>Machinery/Agricultural Equipment</td>
</tr>
</tbody>
</table>

In this Glejser test, if the value prob < alpha (0.05) then there is heteroscedasticity. If instead, the residual free from violations heteroscedasticity assumptions.

### 3.5. Significance Test

#### 3.5.1. F-test

F-test or simultaneous test is performed to determine whether all the independent variables included in the model simultaneously have an influence on the dependent variable. The hypothesis is as follows:

- \( H_0 \): Variable area harvested, irrigation wide productivity and overall no effect on rice production in Ngawi district in 2006-2014
- \( H_1 \): At least one of the variables harvested area, productivity and extensive irrigation overall impact on rice production in Ngawi district in 2006-2014

According to statistical test results on the output Fixed Effect Model, it found that the probability value of 0.0000 < alpha (0.05), that means the independent variables jointly affect the dependent variable or variables \( X_1 \) (harvested area), \( X_2 \) (productivity) and \( X_3 \) (machinery/agriculture) effect overall against rice production in Ngawi years 2006-2014.

#### 3.5.2. t-test

The t-test or the so-called partial regression coefficient test to determine whether the independent variables partially affect the dependent variable. Here are the results of the t-test:

<table>
<thead>
<tr>
<th>Table 4. t-Test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Harvested Area</td>
</tr>
<tr>
<td>Productivity</td>
</tr>
<tr>
<td>Machinery/Agricultural Equipment</td>
</tr>
</tbody>
</table>

The t-test for each variable in the model corresponding output in Table 4. The views of value prob. Variable area harvested and productivity has prob value < alpha (0.05), meaning that these two variables significantly affect rice production in Ngawi. While the variable engine/agriculture has a value prob > alpha (0.05), so it can be concluded that the machinery/agriculture has no effect on rice production in Ngawi. Because there is a variable that does not significantly, then the interpretation of the model, the author will use only the significant variable. The regression results of models are used as follows:
Fixed Effects have been used to determine how much the independent variables harvest land area \(X_1\) and the productivity of the land \(X_2\) affect rice production dependent variable \(Y\). Determinant coefficients show how adaptable the independent variables explain the variation of the model. The coefficient of determination value is between 0-1 and its value will be better if the closer one.

Based on the test results of the best model (Fixed Effects Model), the Adjusted R² values is 99.59%. This means that the independent variable cropland area and productivity effect on the dependent variable rice production amounted to 99.59% and the rest is explained by variables outside the model.

### 3.5.3. Adjusted R-Squared

The coefficient of determination (R²) is the analysis used to determine how much the independent variables harvest land area \(X_1\) and the productivity of the land \(X_2\) affect rice production dependent variable \(Y\). Determinant coefficients show how adaptable the independent variables explain the variation of the model. The coefficient of determination value is between 0-1 and its value will be better if the closer one.

Based on the test results of the best model (Fixed Effects Model), the Adjusted R² values is 99.59%. This means that the independent variable cropland area and productivity effect on the dependent variable rice production amounted to 99.59% and the rest is explained by variables outside the model.

### 3.5.4. Model Interpretation

Based on the stage that has been done, in this case study found that the best model Fixed Effect Model. Fixed Effect Model in two models: a model with a constant slope, intercept varies among individual units and models with constant slope, intercept varies between individuals as well as time. Furthermore, the Lagrange Multiplier test also called Breusch-Pagan test. This test is used to test whether there are effects of time, the individual, or both [9].

In this test, reject \(H_0\) if prob value < 0.05. From the testing that has been done, the results obtained as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harvested Area</td>
<td>6.3660</td>
<td>0.1461</td>
<td>43.5722</td>
<td>0.000</td>
</tr>
<tr>
<td>Productivity</td>
<td>5311.91</td>
<td>218.865</td>
<td>24.2702</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### Table 5. Regression Results

By analyzing the slope value for the two independent variables listed in Table 5, the next model is obtained as follows:

\[
\hat{Y}_{it} = \hat{\beta}_0 + 6,366066X_{1it} + 5311,910X_{2it}
\]

According to Fixed Effect Model, intercept value is different for each individual. Therefore, in this case there will be different models for each district. Among the 19 existing models, the following have 6 models to represent the whole of the model:

1. Sine sub-district:
   \[
   \hat{Y}_{it} = -33534,96 + 6,366066X_{1it} + 5311,910X_{2it}
   \]
2. Jogorogo sub-district:
   \[
   \hat{Y}_{it} = -33026,18 + 6,366066X_{1it} + 5311,910X_{2it}
   \]
3. Karangjati sub-district:
   \[
   \hat{Y}_{it} = -33965,83 + 6,366066X_{1it} + 5311,910X_{2it}
   \]
4. Ngawi sub-district:
   \[
   \hat{Y}_{it} = -33951,88 + 6,366066X_{1it} + 5311,910X_{2it}
   \]
5. Kedunggalar sub-district:
   \[
   \hat{Y}_{it} = -33029,20 + 6,366066X_{1it} + 5311,910X_{2it}
   \]
6. Karanganyar sub-district:
   \[
   \hat{Y}_{it} = -30095,99 + 6,366066X_{1it} + 5311,910X_{2it}
   \]

Information:

- \(X_{1it}\): harvest area the i-th sub-district and in all t
- \(X_{2it}\): Productivity of rice on the i-th sub-district and in all t

Explanation of the above models are assuming other variables remain so in every increase of 1 ha of harvested area will increase rice production amounted to 6.366066 tons. While the increase of 1 kuha productivity will increase rice production as much as 5311.910 tons. The result of Mean Absolute Percentage Error (MAPE) of the model above is shown on Table 7 below.

### Table 6. Breusch-Pagan Test Results

<table>
<thead>
<tr>
<th>Model</th>
<th>p-Value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-way direction</td>
<td>5.766 x 10-14</td>
<td>Reject H0</td>
</tr>
<tr>
<td>Individual</td>
<td>1.386 x 10-14</td>
<td>Reject H0</td>
</tr>
<tr>
<td>Time</td>
<td>0.1904</td>
<td>Fail Reject H0</td>
</tr>
</tbody>
</table>

Based on the results in Table 6. seen that the Fixed Effect Model, there are no time effects, but there are individual effects and effects in both directions. After finding out that there is an individual effect, then the author will show the value of the coefficient (slope) for each variable.

### Table 7. MAPE value

<table>
<thead>
<tr>
<th>Model</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sine sub-district</td>
<td>1.3%</td>
</tr>
<tr>
<td>Jogorogo sub-district</td>
<td>3.4%</td>
</tr>
<tr>
<td>Karangjati sub-district</td>
<td>0.6%</td>
</tr>
<tr>
<td>Ngawi sub-district</td>
<td>0.8%</td>
</tr>
<tr>
<td>Kedunggalar sub-district</td>
<td>2.1%</td>
</tr>
<tr>
<td>Karanganyar sub-district</td>
<td>12%</td>
</tr>
</tbody>
</table>

According to Table 7 above, it is shown that the model for the District Sine, Jogorogo, Karangjati, Ngawi, and...
Kedunggalar has been very good because MAPE has a value of less than 10%. As for the model Karanganyar has good performance because it is in the range between 10% and 20%. The criteria used for selecting the model is that the smaller MAPE value, the better the model. In other word, Karangjati sub-district is the best model among others.

4. CONCLUSION

The study reveals that the variable land area harvested individually has a significant influence on rice production. Assuming other variables remain, safely said that for 1 ha increment in the harvested area it would increase the rice production as 6,367 tons. Meanwhile, productivity variables also affect rice production. Assuming other variables remain so in every increase of 1 ku/ha productivity will increase rice production as much as 5311.910 tons. As for the widely variable irrigation and tools/agricultural machinery not significant. A panel regression model that can best describe rice production factor is the Fixed Effects Model with different intercepts for each district.

REFERENCES


