ENTHUSIASTIC
INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE
Volume 3, Issue 1, April 2023, pp. 74-84

Predict Farmer Exchange Rate in the Food Crop Sector Using Principal Component Regression

Melody Effendi [a,1,*], Ricardo Ardhyatirta [a,2], Sheila Gracia Angelina [a,3], and Margaretha Ohyver [a,4]

[a] Statistics, BINUS University, Jl. Raya Kb. Jeruk No.27, RW.9, Kb. Jeruk, Kec. Kb. Jeruk, Kota Jakarta Barat, Special Capital Region of Jakarta, Indonesia 11530
[1] melody.effendi@binus.ac.id*; [2] Ricardo003@binus.ac.id, [3] sheila.angelina@binus.ac.id; [4] mohyver@binus.edu
*Corresponding author

ARTICLE INFO

Keywords
Agricultural
Food crop
Farmer exchange rate
Principal component regression

ABSTRACT

Farmer Exchange Rate (FER) in Indonesia is very concerning. According to BPS data, there are various regions that experience increases and decreases every year. The goal of this paper is to predict Farmer Exchange Rate in the food crop sector using Principal Component Regression (PCR) since there is multicollinearity in the data. Therefore, with the PCR method using data based on 33 different provinces in Indonesia can determine the Farmer Exchange Rate with supporting factors. The model used can help farmers to be able to improve the welfare and economic growth of Indonesia as it depends on farmers. Further analysis found that the Harvest Area, production, and Human Development Index had an effect on farmer exchange rate. By using this model, it is expected that farmers in Indonesia have an increasing level of welfare and solve multicollinearity problem.

1. Introduction

Indonesia is a country of agricultural accruals that are very famous for its abundant natural resources, one of which is agriculture. Agricultural food raw materials are very influential to our lives because it produces food raw materials. In addition, they are very influential to Indonesia's economic growth. However, there are still many farmers who does not live in prosperity, even many are classified as poor. A report in March 2021 from the Statistics Indonesia (Badan Pusat Statistik, BPS) shows that 51.33% of poor households in the country have the main source of income from the agricultural sector. Farmer Exchanger Rate (FER) is one measure in determining welfare of farmers in Indonesia. If the rate is high, then the relative level of the purchasing power of farmers will be stronger, meaning that the level of welfare of farmers will increase. Therefore, research to determine what factors affect the level of welfare of farmers in Indonesia was conducted.

Based on previous research, some have researched about FER in Sragen Regency, Central Java. The method used in the research was a simple random sampling method where the survey was conducted by collecting sample data from 34 farmers. Research data included primary data and secondary data that could be obtained from the BPS. The results of the research showed that, in 2015, farmers in Sragen Regency lived in prosperity. In this study, researchers used twelve independent variable factors to find out what factors affect the FER. The result of this research indicates there were five factors that had a significant effect: rice productivity, land area, grain prices, fertilizer costs, and non-food expenditures of farmers' households. So, based on previous research, the authors intended to examine the level of farmers' welfare in different provinces in Indonesia. Therefore, this research should still be conducted annually to see the level of welfare of farmers in a variety of different regions.

Factors that can affect the level of welfare of farmers in Indonesia by using the FER from 2010-2019 had been examined. The variables used were Consumer Price Index (CPI), the current price food GDP variable, and the rice price variable. Research data included secondary data which were obtained from BPS .The method used in the research was the time series method. A survey was conducted by collecting sample data from 2010-2019. The result of this research indicates the CPI and the current price food GDP has a significant effect to FER as one of the factors level of welfare of farmers in Indonesia.

This research used three independent variables, namely harvest area, production (ton), human development index (HDI) to calculate the FER. In this research, there were correlation between independent variables and multicollinearity. Multicollinearity causes the model hard to interpret and creates an overfitting problem. The multicollinearity causes large standard error, small changes in the data or variable will cause a drastic change in the value of partial regression coefficients. At the same time, the wide confidence interval value makes it difficult to reject the null hypothesis. Due to the multicollinearity, a regression model which can overcome the existence of this high correlation must be created.

Hence, in this research, the PCR method was used. The PCR method is a method that combines linear regression with the principal component analysis so as to transform correlated free change into orthogonal and non-collated changes. The reason underlying the use of the PCR method was due to the multicollinearity of the data tested. Multicollinearity occurs due to a correlation between free changes that can affect the variety of the smallest range of least squares estimators and the model estimates made.

The goal of this paper is to determine the exchange rate of farmers from various regions using the PCR. The data of farmers' exchange rates with supporting factors in 2020 obtained from the BPS were examined. The benefit of this includes determining factors that have a significant influence on the FER so that the welfare level of farmers in various regions of Indonesia can increase along with the increase of the economic growth in Indonesia. Therefore, the government can improve the welfare of farmers and economic growth in Indonesia.

1.1. The Structure

Indonesia is an agricultural country, meaning that agriculture plays a vital role in the national economy. There are various reasons why the agricultural sector is one of the important economic factors in Indonesia, one of which is to educate the nation's life achieved by having adequate nutritional intake, especially animal protein. Meeting this goal will ensure a sufficient amount of human resources. Therefore, special attention is needed to the future of the farming community because it is very closely related to the nation's future generation. The agricultural sector can be used as a steppingstone for developing rural areas and accelerating the pace of the economy, which is moving slowly. The context of this problem is the low level of farmers' welfare in the country, which is determined by the exchange rate of farmers. Therefore, the factors that affect the exchange rate of farmers were examined so that farmers in Indonesia can live more prosperously. This research can be done with the help of farmer data from the government every year. These data can be accessed by the BPS.

This research is very important because the majority of Indonesia's population relies on the agricultural sector as their income, yet many are categorized as poor. Everyone will be benefited from this research. From the government's point of view, they can increase national economic income so that development of disadvantaged areas can be carried out. From the farmer's perspective, they will be more prosperous and can increase production every year. From the community side, food

p-ISSN 2798-253X
e-ISSN 2798-3153

inflation will be controlled so that people do not need to experience inflation in buying foods. This research needs to be known to improve the economy in Indonesia so that the government can also increase the exchange rate of farmers so that the welfare level of the Indonesians can rise. This research must continue to be improved because technology experts in the field of industry that are currently happening are slowly shifting the agricultural sector. This research can be improved by leveraging technology that is constantly evolving.

Many factors can affect the farmers' welfare level in Indonesia. Therefore, this research aims to find the factors that play an important role in improving the farmers' welfare level. The gap in this research is the lack of indications that affect the exchange rate of farmers registered in BPS, such as the average age, experience, and education of farmers. From the government's point of view, what can be improved is to be more detailed when surveying farmers so that the data accessed from BPS can be more complete.

The researchers of this paper sought to carry out existing research with existing data from BPS and by considering several previous studies. Researchers tried to conduct research using different factors from previous studies to see which factors influence the exchange rate of farmers in Indonesia more.

In this study, there are several aspects that will not be discussed, for examples ways to overcome multicollinearity other than using the PCR method. This study aims to make the reader can understand easily. This research is limited to only 33 provinces because DKI Jakarta does not have agricultural sector data.

2. Related Works/Literature Review

The multicollinearity occurs when independent variables in a regression model are correlated. To overcome multicollinearity, a method called Principal Component Analysis (PCA) can be used. This method aims to observe the observed variables by reducing their dimensions [6]. Calculations using the PCA method require a regression analysis model namely PCR (PCR). PCR analysis is a regression analysis of the dependent variable on the main components that are not correlated with each other, where each principal component is a linear combination of all independent variables. Based on the research which examined data on infant mortality, it was found that the PCR method could be proven to overcome multicollinearity.

3. Material an Methodology

3.1. Dataset

The dataset in this paper obtained from the BPS with a total of 35 cases, including 34 provinces and 1 country. The dataset which including Harvest Area in hectares and production in tons was from 2018 to 2020. Meanwhile, the dataset involving Human Development Index was from 2019 to 2020. Then, the dataset including the dependent variable, FRE, was from 2019 to 2021.

Due to the different year range in each dataset, the data selected were each dataset in 2020. First, the source and link were removed. The country row was also removed because it means the summary of each dataset. Next, data with no value of the dependent variable, which is DKI Jakarta, was also removed. The reason for not including DKI Jakarta is because the data are incomplete so that it can interfere with the process of this research. This process resulted in 33 valid data.

3.2. Multicollinearity

The multicollinearity occurs when independent variables in a regression model are correlated. The problem appears because of independent variables that should be independent. The high degree of freedom of the correlation between variables can cause problems if we fit the model and interpret the results. Multicollinearity causes the estimator of coefficient to swing wildly based on which independent variables. The coefficients will be very sensitive to changes in the model even it is very small. Multicollinearity make the statistical power of the regression model weak since the precision of the estimated coefficients is reduced [1].

 p-ISSN 2798-253X
e-ISSN 2798-3153

One way to test multicollinearity is using Variance Inflation Factors (VIF) [2]. Each independent variable (X_j) needs to be regressed with every other independent variable and look for the coefficient of determination (R_j^2). After that, (1) can be used to find the respective VIF.

$$VIF = \frac{1}{1-R_j^2} \quad (1)$$

where R_j^2 is the coefficient of determination of the regression of X_j on all other independent variables in the model.

FER in Indonesia was the dependent variable that was investigated further using a regression method. Calculation and some tests were performed on the data. The result showed that there was multicollinearity between the two variables, the details of which are discussed in Section 4. The VIF test was used for determining whether multicollinearity is present. The value of VIF must be lower than 10 for multicollinearity to present the data.

3.3. Multiple Linear Regression

The Multiple linear regression is a generalization of simple linear regression that is used to predict the outcome of a variable based on the value of two or more variables [3]. The variable that was predicted is known as the dependent variable, while the variables that was used to predict the value of the dependent variable are known as independent variables. The following is the equation model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_j X_j + \varepsilon \quad (2)$$

where Y is the dependent variable; β_0 is the intercept; $\beta_1, \beta_2, \ldots, \beta_j$ is the regression coefficient (the slope) for each independent variable; $X_1, X_2, \ldots, X_j$ is the fixed value of the data; and ε is the model residual.

3.4. Principal Component Regression

PCR means transforming independent variables that have correlation in that model to the new variables which are not correlated [3].The new variables are orthogonal. The goal of this analysis is to simplify the observed variables by reducing the dimensions. It is done by removing correlation between variables through the transformation of the old variable that has correlation to the new variable that does not have correlation. This step is called principal component [4]. Principal components analysis is an effective tool for the detection of collinearity and as an alternative to the least squares method to obtain estimates of the regression coefficients in the presence of collinearity. The model can be written as:

$$Y = \theta_1 \tilde{X}_1 + \theta_2 \tilde{X}_2 + \cdots + \theta_j \tilde{X}_j + \varepsilon' \quad (3)$$

$$\tilde{Y} = \frac{y_i - \bar{y}}{s_y} \quad (4)$$

$$\tilde{X}_j = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (5)$$

where $\tilde{Y}$ is the standardized version of the dependent variable, $\tilde{X}_j$ is the standardized version of the jth independent variable, $\bar{y}$ is the means of Y, $\tilde{X}_j$ is the means of X_j, s_y is the standard deviations of Y, and s_j is the standard deviation of X_j.

As written in Section 3.1 and Section 3.2, the multicollinearity appears between two variables, harvest area, and production. PCR is one method that can solve multicollinearity problem, especially in this study.

p-ISSN 2798-253X
e-ISSN 2798-3153

4. Results and Discussion

4.1. Results

Statistical tests, calculations, and test assumptions using an application called R Studio were performed. Below is the statistics descriptive summary output for Harvest Area in Table 1, Production in Table 2, Human Development Index in Table 3, and FER in Table 4.

Table 1. Statistical Summary of Harvest Area

Mean	Median	1st Quartile	3rd Quartile	Minimum	Maximum
322,920	133,697.1	61,827.9	317,869.4	289.5	1,754,380.3

Based on Table 1, the average data of harvest area is 322,920 ha. The median or the middle value is 133,697.1 ha with 61,827.9 ha as the first quartile and 317,869.4 ha as the third quartile. The minimum value is 289.5 ha, while the maximum value is 1,754,380.3 ha.

Table 2. Statistical Summary of Production

Mean	Median	1st Quartile	3rd Quartile	Minimum	Maximum
1,655,899	532,168	243,685	1,655,170	853	9,944,538

Based on Table 2, the average data of production is 1,655,899 ton. The median or the middle value is 532,158 ton with 243,685 ton as the first quartile and 1,655,170 ton as the third quartile. The minimum value is 853 ton, while the maximum value is 9,944,538 ton.

Table 3. Statistical Summary of Human Development Index

Mean	Median	1st Quartile	3rd Quartile	Minimum	Maximum
70.79	71.4	69.49	72.09	60.44	79.97

Based on Table 3, the average data of human development index is 70.79. The median or the middle value is 71.4 with the 69.49 as first quartile and 72.09 as third quartile. The minimum value is 60.44, while the maximum value is 79.97.

Table 4. Statistical Summary of FER

Mean	Median	1st Quartile	3rd Quartile	Minimum	Maximum
99.96	99.54	97.47	102.5	93.57	108.59

Based on Table 4, the average data of FER is 99.96. The median or the middle value is 99.54 with 97.47 as the first quartile and 102.5 as the third quartile. The minimum value is 93.57, while the maximum value is 108.59.

Based on Table 1, Table 2, and Table 4, it is right skewed since the mean is greater than the median. Meanwhile, based on Table 3, it is left skewed since the mean is lower than the median. The first quartile is the 25% data, and the third quartile is the 75% data. That is to say, in the harvest area based on Table 1, the 25% data were 61827.9 and the 75% data were 317869.4. The same interpretation applies for Table 2, Table 3, and Table 4.

For the next analyses, level of confidence was needs. The level of confidence used was 95%, meaning that an alpha of 0.05 was used for each result needing an alpha comparison. The result of regression model is in Table 5.

Table 5. Regression Model Coefficient

Model	Coefficient	Standard Error	t-Statistic	p-Value	F-Statistic	R-Squared	Adjusted R-Squared	p-Value Model
Constant	9.763×10^{1}	1.143×10^{1}	8.54	2.08×10^{-9}	2.139	0.1812	0.09647	0.1169
X_1	-3.451×10^{-5}	1.579×10^{-5}	-2.186	0.0371				
X_2	6.309×10^{-6}	2.78×10^{-6}	2.262	0.0314				
X_3	4.28×10^{-2}	1.608×10^{-1}	0.266	0.792				

After obtaining the regression model, some assumption tests were performed. Since the data obtained were in the form of multiple regression, the assumption test was carried out using four test

p-ISSN 2798-253X
e-ISSN 2798-3153

types, namely the Kolmogorov Smirnov test, Durbin Watson test, Breusch Pagan test, and VIF test, shown in Table 6 until Table 9, respectively. The details and results are as follows.

Table 6. Kolmogorov Smirnov Test

Distance	p-Value
0.10561	0.4601

Based on Table 6, the p-value obtained is 0.4601 which is greater than α of 0.05. It indicates that it fails to reject H_0, or errors are normally distributed. Therefore, the assumption for normally distributed errors is fulfilled.

Table 7. Durbin Watson Test

Durbin Watson Value	p-value
2.2528	0.7131

Based on Table 7, t, the p-value obtained is 0.7131 which is greater than α of 0.05. It indicates it fails to reject H_0 or errors are nonauto correlated. Therefore, the assumption for nonauto correlated errors is fulfilled.

Table 8. Breusch Pagan Test

Breusch Pagan Value	*df*	p-value
1.5078	3	0.6805

Based on Table 8, the p-value obtained is 0.6805 which is greater than α of 0.05. It indicates it fails to reject H_0 or errors are not homoscedastic. Therefore, the assumption for errors that is not homoscedastic is fulfilled. In addition, there is not spatial heterogeneity in data because p-value is greater than α.

Table 9. VIF Test

X_1	X_2	X_3
178.68713	179.061388	1.028251

Based on Table 9, X_1 and X_2 VIF value obtained are 178.68713 and 179.061388, which are greater than 10. It means there is multicollinearity between X_1 and X_2. Meanwhile, the X_3 VIF value is smaller than 10, meaning that X_3 is not multicollinear. The use of the PCR method is needed to overcome the problem of multicollinearity.

Principal component analysis needs the eigenvectors. The eigenvectors of a covariance or correlation matrix represent the core of a PCA. The eigenvectors of principal components determine the directions of the new feature space. Then, based on the calculation in R Studio, the eigen vectors as follows.

$$V = \begin{bmatrix} -0.6982275 & -0.1141307 & 0.70671956 \\ -0.6989264 & -0.1048678 & -0.70746350 \\ -0.1548555 & 0.9879154 & 0.00654739 \end{bmatrix} \quad (6)$$

The eigenvectors were calculated to obtain PCR. The result of PCR in RStudio is as follows.

Table 10. Principal Component Regression

Model	Coefficient	Standard Error	t-Statistic	p-value	F-Statistic	R-Squared	Adjusted R-Squared	p-Value Model
Constant	99.9609	0.5561	179.754	$<2\times10^{-16}$	2.139	0.1812	0.09647	0.1169
PC_1	0.4505	0.3971	1.134	0.266				
PC_2	0.2411	0.5719	0.422	0.676				
PC_3	23.7539	10.6739	2.225	0.034				

Based on Table 10, the regression model obtained for FER is as follows.

$$\hat{Y} = 99.9609 + 0.4505PC_1 + 0.2411PC_2 + 23.7539PC_3 \quad (7)$$

p-ISSN 2798-253X
e-ISSN 2798-3153

where,
Y = FER
PC_1 = Harvest area
PC_2 = Production (ton)
PC_3 = Human development index

After the results in Table 10 was obtained, it was performed calculations of the coefficient of final regression model for this FER data obtained from vector of regression coefficient for original independent variable or β. The regression model is as follows with the final is three components are used in this model.

$$\hat{Y} = 99.9609 + 16.445X_1 - 17.145X_2 + 0.324X_3 \quad (8)$$

After the final model was obtained, then below is the VIF test in Table 11.

Table 11. VIF Test after Principal Component Analysis

PC_1	PC_2	PC_3
1	1	1

Based on Table 11, all the VIF values obtained are below 10. It indicates that multicollinearity is not present between all independent variables. According to Table 9 and Table 11, it can be concluded that the multicollinearity between harvest area and production is solved by using the regression method named PCR.

4.2. Statement of Results

Statistical summary results in Section 4.1 must be completed with a graph giving a good indication of data distribution, which is boxplot. The boxplots were performed between a dependent variable which was FER and each independent variable, sequentially harvest area, production, and human development index. The results of three boxplots are presented in Figure 1, Figure 2, and Figure 3 as follow.

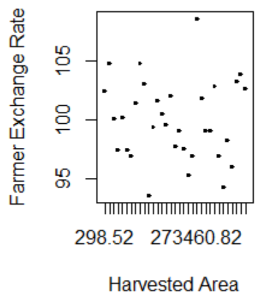

Fig. 1 FER and harvest area boxplot.

Based on Figure 1, it can be concluded that data distribution is evenly distributed and does not show an increase in the FER based on the increase and decrease in harvest area.

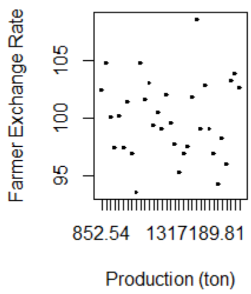

Fig. 2 FER and production boxplot.

p-ISSN 2798-253X
e-ISSN 2798-3153

From Figure 2, it can be concluded that the data distribution is evenly distributed and does not show an increase in the FER based on the increase and decrease in production (ton).

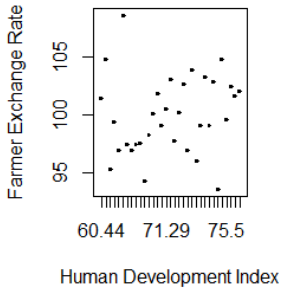

Fig. 3 FER and human development index boxplot.

From Figure 3, it can be concluded is that the data distribution is evenly distributed. Looking at the boxplot, there are some dots that show outlier for the FER, based on the human development index value. Some values of FER are increasing along with the increase of human development index.

4.3. Explanatory Text

Based on Table 5, the regression model obtained for FER is as follows.

$$\hat{Y} = 97.63 - (3.451 \times 10^{-5})X_1 + (6.309 \times 10^{-6})X_2 + (4.28 \times 10^{-2})X_3 \quad (9)$$

where,
Y = FER
X_1 = Harvest Area
X_2 = Production (ton)
X_3 = Human Development Index

Furthermore, the p-value of this model was 0.1169 which was greater than α of 0.05, meaning that there is at least one independent variable that has significant effect on dependent variable, FER. It can be supported by the result of each p-value. Each p-value of harvest area and production was smaller than α of 0.05, meaning that these variables have significant effects on dependent variable. Meanwhile, the p-value of human development index was greater than α of 0.05, meaning that this variable does not have effect on dependent variable. Then, the R-squared value obtained was approximately 18%.

The regression model has other factors, such as the nature of each variable, the unit of measure of the variable, and the transformation of the data. A low R-squared value is generally a bad sign for predictive models. However, in some cases, a good model may show little value. Although the number obtained was less than 75%, but not all small R-squared means bad. It indicates that the X variable used in this model is not enough. The R-squared value of 18% indicates the variables used are not good for the model, but there are other significant variables, namely X_1 and X_2. Both variables are appropriate and can still be used. However, using the R-squared value of 18% means that another X variable is needed in this model.

After getting the regression model based on Table 10 and doing the multicollinearity test without multicollinearity result, the final regression model of estimated FER with the formula in (10) must be obtained.

$$\beta = V\alpha \quad (10)$$

where,
β = Vector of Regression Coefficient for original independent variable
V = EigenVector
α = Vector of Regression Coefficient for new independent variable

Then, the coefficient of final regression model for this FER data obtained from vector of regression coefficient for original independent variable or β. The regression model is as follows.

p-ISSN 2798-253X
e-ISSN 2798-3153

$$\hat{Y} = 99.9609 + 16.445X_1 - 17.145X_2 + 0.324X_3 \tag{11}$$

There are four values of coefficient, each belongs to constant, coefficient of X_1, X_2, and X_3. Constant is 99.9609 that means if all independent variable values are 0, then the value of dependent variable, FER, is equal to constant. Coefficient of X_1, X_2, and X_3 means if each X, independent variable, increase one unit, then the dependent variable increase or decrease by each coefficient, sequentially X_1 is harvest area increasing by 16.445, X_2 is production decreasing by 17.145, and X_3 is human development index increasing by 0.324.

4.4. Discussion

The statistical summary of each variable both independent and dependent variable is known. Mean is the average value of data. Then, when the data commonly is arranged from smallest to the highest, median was the middle value of data or the 2nd quartile, the lowest 25% was 1st quartile, and the 75% was 3rd quartile. The smallest value is the minimum, while the highest is the maximum.

Based on Section 4.1 and Section 4.3, the statistical summary from each variable is known. The independent variable is FER. The mean was 99.96; the median was 99.54; the first quartile was 97.47; the third quartile was 102.5; the minimum value was 93.57; and the maximum value was 108.59.

From the first independent variable is harvest area (ha), it can be concluded that the mean was 322,920; the median was 133,967; the first quartile was 61,827.9; the third quartile was 317,869.4; the minimum value is 289.5; and the maximum value is 1,754,380.3. From the second independent variable was production (ton), it can be concluded that the mean was 1,655,899; the median was 532,158; the first quartile was 243,685; the third quartile was 1,655,170; the minimum value was 853; and the maximum value was 9,944,538 ton. The last independent variable was human development index. The mean was 70.79; the median was 71.4; the first quartile was 69.49; the third quartile was 72.09; the minimum value was 60.44; and the maximum value was 79.97.

Based on Section 4.1 that the result of R-squared value is approximately 18% is discussed simply in Section 4.3. The larger the R-squared value, the more precisely the predictor variables are able to predict the value of the response variable, but how high an R-squared value needs to be considered as a good varies based on the field [5]. Although the number obtained was 18% which is less than 75%, but that means the X variable used in this model is not enough. The R-squared value of 18% indicates the variables used are not good for the model, but there are significant variables, namely X_1 and X_2 that concluded from the VIF value which both dependent variables have the high values of VIF. Both variables are appropriate and can still be used, but with the R-squared value of 18%, it means that another X variable is needed in this model. It can be said that farmer must think about the other factor in increasing and decreasing the FER.

One of many methods to predict FER is used in this study, namely linear regression method. Based on the result in Section 4.1, there is multicollinearity between harvest area and production that is shown from VIF test before PCR method. Because of the existing multicollinearity between X_1 and X_3, this study used the PCR method to predict the response variable. Then, the result of VIF test after the PCR method is small, representing that PCR can be the choice to solve the multicollinearity problem.

The analysis unit in this study was province, which is larger than the Sragen paper. The final result of this paper is not as specific as in other papers. It shows that smaller research units such as in Sragen will produce more specific and more accurate final results. This study can add other variables to strengthen and support the model that has been made so that later it can become a more focused and specific model. The other variables in the other paper can be used to model development that more accurate for FER.

5. Conclusion

The PCR method is one of the effective methods to solve multicollinearity problems. PCR produces key component components that have less than ten VIFs. PCR can be assisted by software

p-ISSN 2798-253X
e-ISSN 2798-3153

R. In this study, there were three independent variables and one dependent variable. The response or dependent variable was the FER as Y and the independent variables were the harvest area as X_1, production as X_2, and the human development index as X_3. In this case, it is proven that PCR can solve the multicollinearity problem and tested that VIF value is below 10.

Because R-squared value was about 18.12%, the solution was by adding the independent variables that could affect the increasing effects of FER. Based on the result of R-squared value, it can be concluded that all the factors which are independent variables have significant effects in this regression model. The final model for this data on (5) in Section 3 was calculated from the procedure of PCR method that can solve multicollinearity in data.

Variables that affect FER were X_1 and X_2. If this model is used to predict the FER, those two variables can be used, which are harvest area and production. Some other variables also can be added based on the other study.

References

[1] J. Frost, *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models.* Statistics by Jim Publishing: 2020.

[2] R.J. Freund, W.J. Wilson, and P. Sa, *Regression Analysis: Statistical Modeling of a Response Variable,* 2nd ed. MA, USA: Academic Press, 2006.

[3] S. Chatterjee and A.S. Hadi, *Regression Analysis by Example*, 5th ed. NJ, USA: Wiley, 2012.

[4] V. Gasperz, *Analysis Techniques in Experimental Research 2*. Bandung, Indonesia: Tarsito, 1995.

[5] "What is a Good R-squared Value?" Statology. https://www.statology.org/good-r-squared-value/ (accessed Feb. 24, 2019).

[6] P. Pendi, "Analisis Regresi dengan Metode Komponen Utama dalam Mengatasi Masalah Multikolinearitas," *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 10, no. 1, pp. 131–138, 2021, doi: 10.26418/bbimst.v10i1.44750.

Appendix

Appendix 1. Dataset

Province	Harvest Area	Production (ton)	Farmer Exchange Rate	Human Development Index
Aceh	317,869.41	1,757,313.07	99.08	71.99
North Sumatra	388,591.22	2,040,500.19	96.9	71.77
West Sumatra	295,664.47	1,387,269.29	99.06	72.38
Riau	64,733.13	243,685.04	93.57	72.71
Jambi	84,772.93	386,413.49	100.48	71.29
South Sumatera	551,320.76	2,743,059.68	98.27	70.01
Bengkulu	64,137.28	292,834.04	103.02	71.4
Lampung	545,149.05	2,650,289.64	94.26	69.69
Bangka Belitung Islands	17,840.55	57,324.32	100.2	71.47
Riau Islands	298.52	852.54	102.5	75.59
West Java	1,586,888.63	9,016,772.58	103.28	72.09
Central Java	1,666,931.49	9,489,164.62	103.93	71.87
Special Region of Yogyakarta	110,548.12	523,395.95	102.04	79.97
East Java	1,754,380.3	9,944,538.26	102.66	71.71
Banten	325,333.24	1,655,170.09	102.91	72.45
Bali	90,980.69	532,168.45	99.54	75.5

 p-ISSN 2798-253X
e-ISSN 2798-3153

Province	Harvest Area	Production (ton)	Farmer Exchange Rate	Human Development Index
West Nusa Tenggara	273,460.82	1,317,189.81	108.59	68.25
East Nusa Tenggara	181,690.63	725,024.3	95.28	65.19
West Kalimantan	256,575.43	778,170.36	96.91	67.66
Central Kalimantan	143,275.05	457,952	99.08	71.05
South Kalimantan	289,836.35	1,150,306.66	101.85	70.91
East Kalimantan	73,568.44	262,434.52	101.6	76.24
North Kalimantan	9,883.05	33,574.28	100.09	70.63
North Sulawesi	61,827.86	248,879.48	104.82	72.93
Central Sulawesi	178,066.94	792,248.84	97.58	69.55
South Sulawesi	976,258.14	4,708,464.97	95.96	71.93
Southeast Sulawesi	133,697.15	532,773.49	97.77	71.45
Gorontalo	48,686.34	227,627.2	96.93	68.68
West Sulawesi	64,826.18	345,050.37	99.35	66.11
Maluku	28,668.22	110,447.3	97.44	69.49
North Maluku	10,301.91	43,382.85	97.47	68.49
West Papua	7,570.63	24,378.33	104.83	65.09
Papua	52,727.52	166,002.3	101.46	60.44

p-ISSN 2798-253X
e-ISSN 2798-3153