Study of Forecasting Method for Agricultural Products Using Hybrid ANN-GARCH Approach

Miftahul Jannah, Kusman Sadik*, and Farit Mochamad Afendi

Department of Statistics, IPB University, Indonesia

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

Abstract. The capital city of Jakarta as the center of socio-economic activities has to maintain the stability of its economy, including the stability of staple food prices such as rice, onion, and chili prices. To make a strategic plan for anticipatory action, one of the ways is by doing forecasting. The problem in forecasting staple food prices is the fact that those staple foods have high volatility prices. It makes the conditional variance of residual become inconstant. This research applied the Hybrid ANN-GARCH model that combined Artificial Neural Network (ANN) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. This method has used to predict rice, onion, red chili, and cayenne pepper prices in Jakarta. The data divided into two categories, training and testing data. The result shows that the Hybrid ANN-GARCH model produced smaller RMSE and MAPE values than ARIMA model. Based on the accuracy value of the model, it can be concluded that Hybrid ANN-GARCH better than ARIMA in forecasting the price of the four commodities.

1. Introduction

The price of food commodities always attracts the attention of various groups, especially when there is a national celebration because it directly related to the welfare of the wider community, both at the consumer and producer levels. Several food commodities whose price fluctuations are often in the spotlight, including rice, onion, red chilies, and cayenne pepper. The prices of these four commodities are more volatile than other foods included in the inflation calculation [1].

Rice is the staple food of Indonesian society that has a population that reaches 255.46 million people. It has growth rate of 1.31% and has a rice consumption rate reached 124.89 kg/capita per year [2]. Meanwhile, onion, red chilies, and cayenne pepper are commonly used commodities as ingredients in Indonesian’s cooking. Onion have the highest production value after cabbage and potatoes in the vegetable subsector [3]. On the other hand, red chili and cayenne pepper are two strategic vegetable commodities [4]. Complete information and the ability to predict or forecast future prices accurately are needed to produce the right food price stabilization policy.

The main problem in predicting food commodity prices is the high volatility of current food commodity prices, which makes the future price prediction process more complicated. A statistical method is needed to capture data fluctuation patterns well so that it can obtain accurate prediction results. Market players or observers use different approaches in predicting the value of economic variables. These methods divided into two categories, namely classical statistical and neural network methods. [5].
One of the problems in modeling time series data is the variety of residuals that are not constant (containing heteroscedasticity). The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model is a time series model. It can accommodate the inequality of the residual variants. Several previous studies that used the GARCH method include [6] predicting the volatility of rice prices in Indonesia and [7] modeling the national rice harvested land area.

Furthermore, the Artificial Neural Network (ANN) is a method that commonly used as a forecasting method. As was done by [8] who predicted rice productivity using ANN Backpropagation. In statistics, ANN is a broad class of nonlinear regression model, data reduction model, and nonlinear dynamic system. It is able to process a wide variety of data and make predictions that are sometimes very accurate [9].

A lot of studies have found benefits from combining ANN with GARCH in the form of a hybrid model. As was done by [10] using hybrid ANN and GARCH methods for modeling the volatility of the Shanghai stock price index and found that the EGARCH-ANN hybrid model gave better results in modeling the volatility of the log-return value of stocks on the Shanghai stock market. In this study, the researcher used the Hybrid ANN-GARCH model in forecasting several commodity prices in Jakarta, namely the prices of rice, onion, red chilies, and cayenne pepper.

2. Research Methods

2.1. Data

This study used secondary data on the prices of rice, onion, red chilies, and cayenne pepper in DKI Jakarta from the Central Statistics Agency. This data contained in the publication of the Weekly Development of Retail Price of Several Staple Food in The Capital Province of Indonesia. This data are weekly time series data taken from the first week of January 2010 to the fourth week of December 2018. The data divided into two parts that were training and testing data. 442 data of the first data as training data and 26 final data as testing data.

| No | Variables        | Unit              |
|----|------------------|-------------------|
| 1  | Rice             | Rupiah/kilogram   |
| 2  | Onion            | Rupiah/kilogram   |
| 3  | Red Chilies      | Rupiah/kilogram   |
| 4  | Cayenne Pepper   | Rupiah/kilogram   |

2.2. Method of Analysis

ANN-GARCH hybrid model combines the GARCH method and the ANN method in modeling the value of a time series data. The components that compiled the GARCH model (p, q) were used as input variables in the training process in the ANN. The following are the analysis steps.

1. Create a mean model in the form of the ARIMA model
2. Build residual variance model
   a. Testing of heteroscedasticity effect

   The building of the Variance model begins with a homogeneity examination of variance. Variance models formed when there is an inhomogeneity of the residual variance (heteroscedasticity). The testing of heteroscedasticity effects from residual data is by the ARCH-LM test. The remainder obtained from the ARIMA model is squared. Then proceed by regressing the residual squares using constants until the q-lag, thus forming a regression equation.

   \[ u_t^2 = k + a_1 u_{t-1}^2 + \cdots + a_q u_{t-q}^2 + e_t \]  

   If the estimated value \( a_1, \ldots, a_q \) is zero, it can be concluded that \( u_t^2 \) does not have real autocorrelation or there is no ARCH effect. The hypotheses used in this test are:

   - \( H_0: a_1 = \cdots = a_q = 0 \) (no ARCH effect)
   - \( H_1: a_1, \ldots, a_q \neq 0 \) (ARCH effect)

The test statistics is: 

\[ LM = \sum_{t=1}^{n-q} u_t^2 - \text{Mean} \]

If the estimated value \( a_1, \ldots, a_q \) is zero, it can be concluded that \( u_t^2 \) does not have real autocorrelation or there is no ARCH effect. The hypotheses used in this test are:

   - \( H_0: a_1 = \cdots = a_q = 0 \) (no ARCH effect)
   - \( H_1: a_1, \ldots, a_q \neq 0 \) (ARCH effect)
\( H_0: a_1 = a_2 = \cdots = a_q = 0 \) (there is no ARCH effect)
\( H_1: \) there is at least one \( a_i \neq 0 \), for \( i = 1,2, \ldots, q \)
with the LM test statistics as follows:

\[
LM = nR^2
\]

\( n \) is the number of observations and \( R^2 \) is the coefficient of determination of the residual square regression model in equation (1).

b. Build a residual variance model

The selection of variance model is by trying starting from the smallest ARCH / GARCH order until the ARCH / GARCH coefficient is no longer significant. Next, the best variance model is selected that is the GARCH model with the smallest Akaike Information Criterion (AIC) value.

3. Build a Hybrid ANN-GARCH model

The \( \hat{y} \) value calculated based on the mean model and the variance model from the previous steps. Then the value is used as an input variable in the training process in ANN. Then do the forecasting for the next 26 weeks for the testing process.

4. Evaluate model performance

The model evaluation was on training data and testing data by calculating its accuracy measures that are root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute deviation (MAD). Calculating three of model accuracy is for testing data on the four commodities in this study. Also, calculating the accuracy of the ARIMA model for comparison.

5. Do forecasting for the first period ahead.

3. Results and Discussion

Volatility in a time series refers to the condition where the conditional variance of the time series varies between times [11]. In other words, there was a heteroscedasticity problem in residual variance. The more volatile a price was, the higher the uncertainty of its price. It does mean that the more difficult it is to predict. The following is a plot of the price volatility of four food commodities in DKI Jakarta from 2010 to 2018.

Figure 1 The plot of volatility in retail prices for rice, onion, red chilies, and cayenne pepper in DKI Jakarta for January 2010 to December 2018

Figure 1 shows the volatility of the four staple food prices in DKI Jakarta from January 2010 to December 2018. It shows that the price of cayenne pepper has the highest level of volatility than red chilies, onion, and rice. The level of volatility in the price of cayenne pepper is quite different from the three other staple foods. The retail price of cayenne pepper has higher level
of uncertainty than the three other staple food prices. Meanwhile, the price volatility of red chilies and onion is not much different. While the price of rice shows lower price volatility compared to the other three commodities.

| Commodities    | ARIMA Model   | AIC    |
|----------------|---------------|--------|
| Rice           | ARIMA(1,1,0)  | 5649.66|
| Onion          | ARIMA(2,1,1)  | 8174.51|
| Red chilies    | ARIMA(0,1,1)  | 8606.87|
| Cayenne pepper | ARIMA(0,1,5)  | 8959.44|

The best ARIMA model is the tentative ARIMA model with the smallest AIC value. After obtaining the best ARIMA model for each commodity, the next step is to check the residual model to find out whether the ARIMA model is good or not for applying. The result summaries of the Lagrange Multiplier (LM) test for each ARIMA model are in the following table.

| lag | Rice   | Onion  | Red chilies | Cayenne pepper |
|-----|--------|--------|-------------|----------------|
|     | LM     | p-value| LM          | p-value        |
| 4   | 279.2  | <0.0001| 727.9       | <0.0001        |
| 8   | 128.5  | <0.0001| 318.6       | <0.0001        |
| 12  | 83.1   | <0.0001| 205.4       | <0.0001        |
| 16  | 61.0   | <0.0001| 148.6       | <0.0001        |
| 20  | 45.9   | 0.0005 | 91.7        | <0.0001        |
| 24  | 35.9   | 0.0424 | 65.7        | <0.0001        |

Table 3 shows the p-value less than 0.05, starting from lag 4 to lag 24 for each commodity. It means that ARCH / GARCH affects the mean model error. The p-value in the LM test can also be an indication of the right model for the variety of the mean model. Based on table 3, all commodities have a significant p-value up to a lag of 24 that can be said to be quite a long lag, so this indicates that the GARCH model is more appropriate for the variance model. Table 4 below shows a summary of the results of selecting the GARCH model for the residual.

| Commodities    | Model GARCH | AIC   |
|----------------|-------------|-------|
| Rice           | GARCH(1,0)  | 12.550|
| Onion          | GARCH(1,1)  | 18.004|
| Red chilies    | GARCH(1,1)  | 19.347|
| Cayenne pepper | GARCH(1,1)  | 19.945|

The selection of variance models is by trying to start from the smallest ARCH / GARCH order until the ARCH / GARCH coefficient is no longer significant. Furthermore, selecting the best variance model is the GARCH model with the smallest AIC value. Table 4 shows the best GARCH model for each commodity.

After obtaining the mean and variance models, calculating the estimated value of each commodity price ($\hat{y}$) for the 442 of the first data as input variables in the ANN process. ANN
model is using the net function available in the package forecast in R. Forecasting for test data and calculating the accuracy value of each commodity as below. As a comparison, done forecasting using the ARIMA model that has been obtained in the previous step and calculating its accuracy value.

| Commodity      | Model          | RMSE  | MAPE  |
|----------------|----------------|-------|-------|
| Rice           | ARIMA          | 212.679 | 0.018 |
|                | Hybrid ANN-GARCH | 87.764 | 0.007 |
| Onion          | ARIMA          | 8806.956 | 0.274 |
|                | Hybrid ANN-GARCH | 2103.660 | 0.048 |
| Red chilies    | ARIMA          | 3708.270 | 0.084 |
|                | Hybrid ANN-GARCH | 1788.076 | 0.070 |
| Cayenne pepper | ARIMA          | 12106.138 | 0.282 |
|                | Hybrid ANN-GARCH | 4624.848 | 0.070 |

4. Conclusion
The commodities of rice, onion, red chilies, and cayenne pepper in Jakarta are food commodities that have fairly high volatility. It caused the inhomogeneity of the residual variance of the mean model. Handling the inhomogeneity of the residual variance with the GARCH model and the combination of the GARCH and ANN models into a hybrid ANN-GARCH produced better accuracy values seen from the smaller RMSE and MAPE values than ARIMA.

References
[1] Pradana RS 2019 Kajian Perubahan dan Volatilitas Harga Komoditas Pangan Strategis serta Pengaruhnya Terhadap Inflasi di Kota Banda Aceh JIEP-Vol. 19, No. 2, 85-100
[2] Ministry of Agriculture 2016 Outlook Komoditas Pertanian Subsektor Tanaman Pangan Padi Pusat Data dan Sistem Informasi Pertanian Kementerian Pertanian
[3] Sahara, Utari MH, and Azijah Z 2019 Volatilitas Harga Bawang Merah di Indonesia Buletin Ilmiah Litbang Perdagangan, Vol. 13, No. 2, 309-335
[4] Anwarudin MJS, Sayekti AL, Marendra AK, and Hilman Y 2015 Dinamika Produksi dan Volatilitas Harga Bawang Merah di Indonesia Buletin Ilmiah Litbang Perdagangan, Vol. 13, No. 2, 309-335
[5] Nazarian R, Alikhani NG, Naderi E, and Amiri A 2013 Forcasting Stock Market Volatility: A Forecast Combination Approach MPRA Paper
[6] Busnita SS, Oktaviani R, and Novianti T 2017 How far climate change affects the Indonesian paddy production and rice price volatility? International Journal of Agricultural Sciences 1(1):1-11
[7] Iqbal TA, Sadik K, and Sumertajaya IM 2014 Pemodelan Pengukuran Luas Panen Padi Nasional Menggunakan Generalized Autoregressive Conditional Heteroscedastic Model (GARCH) Penelitian Pertanian Tanaman Pangan Vol. 33 No. 1, 17-26
[8] Ramadhona G, Setiawan BD, and Bachtiar FA 2018 Prediksi Produktivitas Padi Menggunakan Jaringan Syaraf Tiruan Backpropagation Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer; Vol.2. No.12, 6048-6057
[9] Djuraidah A 2008 Penggunaan Jaringan Syaraf Tiruan untuk Pendugaan Model Linear Terampat dengan Koefisien Keragaman Konstan Forum Statistika dan Komputasi; p:7-10
[10] Xunfa Lu, Danfeng Que, and Guangxi Cao 2016 Volatility Forecast Based on the Hybrid Artificial Neural Network and GARCH-type Models Procedia Computer Science 91: 1044-1049

[11] Cryer JD and Chan KS 2008 Timer Series Analysis with Applications in R Second Edition New York: Springer