# FORECASTING THE CONSUMER PRICE INDEX USING THE GARCH METHOD

**Anida Ulfiana, Wiwit Pura Nurmayanti*, Seni Eriani, Harista Almiatus Soleha, Basirun, Siti Hariati Hastuti**

*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Hamzanwadi (INDONESIA)*

**Corresponding author: wiwit.adiwinata3@gmail.com*

## Abstract

Consumer Price Index (CPI) is an index number that shows the level of prices of goods and services purchased by consumers in a certain period. Forecasting related to CPI data needs to be done to describe the price level of goods and services purchased by the public. There are several methods in statistics that can be used for forecasting, one of which is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) method. GARCH has advantages compared to other forecasting methods, namely that it can apply to data that has high volatility. Volatility occurs if the data variance is not constant and will certainly cause the data to be non-stationary so that it does not meet the assumptions in the time series analysis. The purpose of this study was to determine the best model of the GARCH method and to find out the prediction results of the CPI for the future period. Based on the results of the analysis, the best models used are GARCH (1,0). And the CPI value in January 2022 was 112,1116.
Keywords: Forecasting, Consumer Price Index (CPI), Generalized Autoregressive Conditional Heteroscedasticity, GARCH, Volatility.

## 1 INTRODUCTION

The Consumer Price Index (CPI) is one indicator that shows monetary success in controlling inflation, besides that the CPI calculation is also very important because it shows the price level of goods and services purchased by the public [1]. According to the Central Statistics Agency (BPS), the CPI is an index that calculates the average price change of a set of goods and services consumed by the population within a certain period. During the last three years, BPS recorded that the CPI value in each month tends to increase and was recorded in December 2019 of 139.07. Changes in the CPI value from time to time can describe the level of price increase (inflation) and the level of price decline (deflation) [2]. The higher the CPI value, the faster the inflation rate [3]. Therefore, information is needed that can describe the state of the CPI. One thing that can be done is forecasting or estimating the CPI figure for several periods in the future.

Forecasting is an accurate calculation in determining the future by using past data [1][4]. One of the forecasting methods that can be used is GARCH method. The GARCH model has been widely used to describe the volatility behaviour of a financial time series, especially on stock and currency exchange data [5].

There were several previous studies that used the GARCH method, including those conducted by Nella Angraeny [6], which conducted research on the value of exports in Indonesia from January 2009 to April 2019 and the results obtained were ARIMA (1,1,2) and GARCH(1,3) models. A similar study was also conducted by Aisyah Muhayani [7], namely by comparing APARCH, E-GARCH and T-GARCH in forecasting world gold prices, and the results of the study that the most optimal method is E-GARCH with a MAPE value of 4.66. another study was conducted by Anbiya, W., and Garin, F.C. [8]. This study discusses the Application of GARCH Forecasting Method in Predicting the Number of Rail Passengers (Thousands of People) in Jabodetabek Region. In this study, the best model was ARIMA (1,1,1) with an AIC value of 2,159.87 and the best model GARCH (1,1) with an AIC value of 18,314. Referring to these problems and several existing literature studies, the purpose of this study is to predict the value of the Consumer Price Index in the future using the GARCH method.

# 2 METHODOLOGY

The data used in this study is secondary data, namely Consumer Price Index data for the period January 2014 to December 2021. The data is sourced from the website of the Central Statistics Agency (BPS) at https://www.bps.go.id/indicator/3/2/1/indeks-harga-konsumen-umum-.html. This research uses time variable (month) and Consumer Price Index (CPI) variable. Data analysis in this study used the GARCH method using the R software.

## 2.1 Autoregressive Moving Average Model (ARIMA)

Some of the Jenkins Box models that can be used on time series data are as follows.

### 2.1.1 Autoregressive Process (AR)

The autoregressive process is used to describe a condition where the present value of a time series depends on the previous value plus the random stock. The general form of an autoregressive model of order p is in Equation (1) [9], where $Z_t$ = random variable at time t, $\omega_i$ = regression coefficient in the i-order AR process, i = 1, 2, ..., p, $\omega_0$ = average constant, $p =$ orde AR, $\varepsilon_t$ = error value at time t, $t =$ time.

$$Z_t = \omega_0 + \omega_1 Z_{t-1} + \cdots + \omega_p Z_{t-p} + \varepsilon_t \quad (1)$$

### 2.1.2 Moving Average Process (MA)

To estimating the value using the value in previous periods, the value can also be estimated using the residual value [9]. Moving Average (MA) model with order q is denoted MA(q). The general form of the MA(q) model is in Equation (2), where $Z_t =$ value of randomm variable at t, $\emptyset_i =$ regression coefficient on MA process of order i, i = 1, 2, ..., q, $q$ = orde MA, $\varepsilon_t$ = error value at time t, $t =$ time.

$$Z_t = \emptyset_0 + \varepsilon_t - \emptyset_1 \varepsilon_{t-1} - \cdots - \emptyset_q \varepsilon_{t-q} \quad (2)$$

### 2.1.3 Autoregressive Moving Average Process (ARIMA)

The ARIMA model is a combination of Autoregressive (AR) and Moving Average (MA) models as well as differencing processes (order d for non-seasonal, and D for seasonal) on time series data. In general, the non-seasonal ARIMA model can be written as ARIMA (p, d, q) with the following general form Equation (3) [9], where $\mathrm{Z_t}$ =

random variable at time t, $\omega_i$ = regression coefficient in the i-order AR process, i = 1, 2, ..., p, p = orde AR, $\emptyset_i$ = regression coefficient on MA process of order i, i = 1, 2, ..., q, q = orde MA, $å_t$ = error value at time t, t = time, d = orde differencing.

$$Z_t - Z_{t-d} = \emptyset_0 + \sum_{i=1}^{p} \omega_1 (Z_{t-1} - Z_{t-1-d}) + \sum_{j=1}^{q} \emptyset\, \varepsilon_{t-j} \quad (3)$$

## 2.2 Autoregressive Conditional Heteroscedastic Model (ARCH)

The Autoregressive Conditional Heteroscedasticity (ARCH) model is an autoregressive model that occurs in a state of non-constant variance. This model shows the instability of variance in the time series model so that it can be used as an alternative for calculating and modelling data [6]. The basic concept of the ARCH model is the variance of the squared residuals from several past periods. The ARCH model with order p denoted ARCH(p) is expressed in two equations, namely the average equation and the variance equation (4) and (5) [6], where $Y_t$ = dependent variable at time t, $X_t$ = independent variable at time t, $\beta_0$ = constant, $\beta_1$=multiple regression coefficient, $\varepsilon_t$ = residual.

$$Y_t = â_0 + â_1 X_t + å_I \quad (4)$$

$$ó^2{}_t = á \quad + á_1 å^2{}_{t-1} \quad (5)$$

## 2.3 Generalized Autoregressive Conditional Heteroscedasticity Model (GARCH)

The ARCH-GARCH model was developed primarily to address the issue of volatility in economic and business data, particularly in the financial sector. This causes the previous forecasting models to be less able to approach the actual conditions. This volatility is reflected in the residual variance that does not meet the assumption of homoscedasticity [10].

This model was developed as a generalization of the volatility model and in this model, the variance consists of three components [10]. GARCH is one approach to modelling time series with error conditions varying according to time (heteroscedasticity). GARCH is considered to provide simpler results because it uses fewer parameters, thereby reducing the error rate in calculations. The basic concept of GARCH is that variance is not only influenced by past residuals but also by the lag of the conditional variance itself [6].

Thus, the conditional variance in the GARCH model consists of two components, namely the past component of the squared residual (denoted by degree p) and the past component of the conditional variance (denoted by degree q), in Equation (6).

$$\text{ó}^2{}_t = \text{ù} + \sum_{i=1}^{p} \text{á}_i u^2{}_{t-1} + \sum_{j=1}^{q} \text{â}_j\, \text{ó}^2{}_{t-1} \qquad (6)$$

### 2.3.1 Stages of GARCH Analysis

The steps used to implement the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model on Indonesian export value data for the period January 2014-February 2022, are as follows:

1. Perform stationarity test and normality test
2. Identify the ARIMA model based on the ACF and PACF charts
3. Determine the ARIMA model estimate
4. Test the significance of the parameters on the estimation results of the ARIMA model
5. Test assumptions (normality, heteroscedasticity, and autocorrelation)
6. Identify the GARCH model with the ARIMA model
7. Define the GARCH model
8. Test the significance of the parameters on the estimation results of the GARCH model
9. Perform the LM-ARCH test
10. Doing the forecasting stage

## 2.4 Akaike dan Schwarz Information Criterion (AIC and SIC)

Akaike and Schwarz criteria (AIC and SIC) in model selection can also be done using Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) in Equation 7, and 8 [5], where $e$ = 2.718, $u$ = residual, $k$ = number of estimation parameter variables, $SSR =$ sum of squared residual, and $n$ = number of observations (sample). The model chosen is the model that has the smallest AIC and SIC.

.

$$AIC = e^{2\frac{k}{n}}\frac{\sum u_i{}^2}{n} = e^{2\frac{k}{n}}\frac{SSR}{n} \quad (7)$$

$$SIC = e^{\frac{k}{n}}\frac{\sum u_i{}^2}{n} = e^{\frac{k}{n}}\frac{SSR}{n} \quad (8)$$

## 3 RESULTS

The description of the data is done to find out the general picture of the data that will be used for further analysis. The data used is the Indonesian Consumer Price Index data for the monthly period starting from January 2014 to December 2021. The amount of data used in the formation of the model is 96 data.

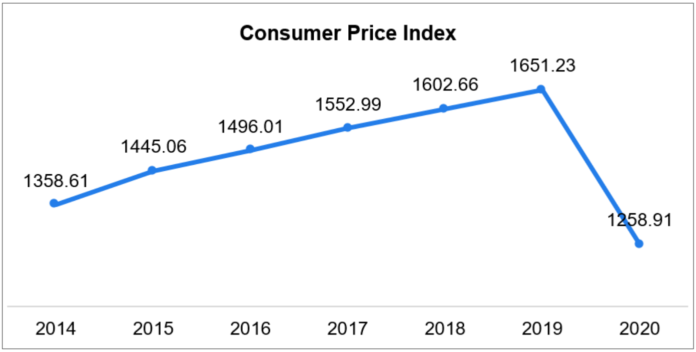


***Figure 1.*** *Graph of Consumer Price Index for the period 2014-2021.*

Based on ***Figure. 1***, the value of the Indonesian consumer price index has increased over the last 7 years. It was found that the highest value of the Indonesian consumer price index occurred in 2019 at 1651.23 and with the lowest value in 2020 at 1258.91.

### 3.1 Formation of the ARIMA Model

Determination of the ARIMA model begins by looking at the data pattern to determine the shape of the graph and the stationary of the data. The following is a plot of ACF and PACF before the data stationarity test is carried out.

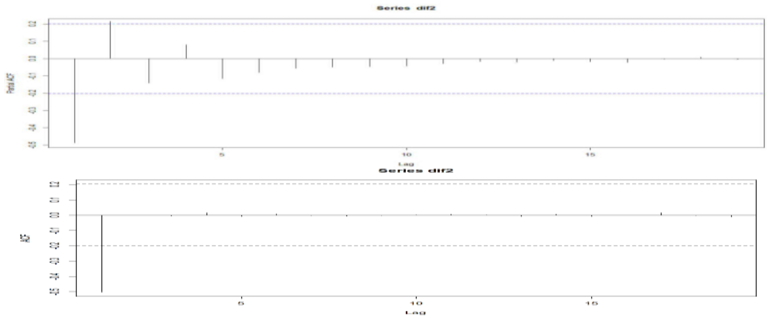


***Figure 2.*** *Plots of ACF and PACF ARIMA.*

For this reason, the stationarity test stage of the Consumer Price Index data is carried out next using the ADF test which is presented in ***Table. 1.***

***Table 1.*** *Data Stationary Test.*

| **Stationery Test (ADF)** | ***P-value*** |
|---|---|
| Differencing (1) | 0.69 |
| Differencing (2) | 0.01 |

After performing the ADF test, the data is stationary with a p-value of 0.01 which is less than 0.05 after differencing 2 times so that it can be continued at the model identification stage. Model identification is done by looking at the ACF and PACF plots in ***Figure. 1***, so that the model is obtained.

Based on ***Figure. 1,*** the ACF plot contains 1 lag that crosses the line which means it contains a moving average or MA (1). Meanwhile, the PACF plot shows that there are 2 lags that cross the line which means it contains autoregressive or AR (2). So that the main model is obtained, namely ARMA (2.1). Then overfitting the model around the main model to find out the best model. ***Table. 2*** shows the results of the model fit test around the main model.

***Table 2.*** *Arima. Parameter estimation results.*

| **Model** | **AR** | | **MA** | **Significant** |
|---|---|---|---|---|
| | **1** | **2** | **1** | |
| AR (1) | 1.951e-08 | | | Yes |
| AR (2) | 6.252e-12 | 0.0005781 | | Yes |

| | | | | |
|---|---|---|---|---|
| MA (1) | | | < 2.2e-16 | Yes |
| ARMA (1,1) | 0.9684 | | < 2e-16 | No |
| ARMA (2,1) | 0.9685 | 0.9993 | < 2e-16 | No |

The results of the ARIMA parameter estimation in ***Table. 2***. There are three significant models seen from the p-value smaller than alpha 0.05, namely AR (1), AR (2), and MA (1). To find out the best model that will be used for forecasting, a diagnostic test of residual data is carried out. There are three diagnostic tests for residual data, namely normality test, no autocorrelation test, and homoscedasticity test.

***Table 3.*** *Diagnostic test of ARIMA model.*

| | **Normality** | **No Autocorrelation** | **Homoscedasticity** |
|---|---|---|---|
| AR (1) | 2.2e+16 | 0.08568 | 0.009436 |
| **AR (2)** | 2.2e+16 | 0.3493 | 0.2523 |
| **MA (1)** | 2.2e+16 | 0.8349 | 0.9157 |

The results of the diagnostic test in ***Table. 3.*** there is one model, namely AR (1) which has the assumption that it meets the assumption of normality, no autocorrelation, and does not meet homoscedasticity. In other words, the model is heteroscedasticity. Therefore, it needs to be analysed further using the ARCH-GARCH method.

## 3.2 Formation of the GARCH Model

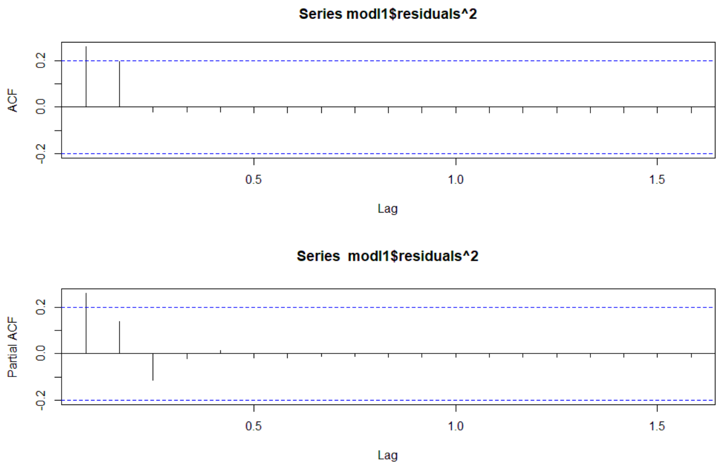


***Figure 3.*** *Plots of ACF and PACF model GARCH.*

After estimating the GARCH model, the best GARCH model is GARCH (1,0) because there are several significant parameters. Thus, the GARCH (1.0) model will be a more likely model to be used in forecasting the Indonesian Consumer Price Index for the next period, namely 2022 and the equation from the GARCH (1.0) model is obtained as follows:

$$\sigma_t{}^2 = 1.177e{+}01 + 1.000e{+}00\sigma_{t-1}{}^2$$

This means that the variance of the Indonesian Consumer Price Index in period t is found by a constant (1.177e+01) and a residual/square residual in the previous period. The results of the diagnostic test of the ARIMA (2,1) ~ GARCH (1,0) model that the model has been able to handle the symptoms of heteroscedasticity is 0.9999. The LM Arch Test has a p-value of more than 0.05, which means that the model does not contain the ARCH effect or is free from heteroscedasticity symptoms in the squared residuals. Therefore, the ARIMA-GARCH method can be used to predict the Indonesian Consumer Price Index. The following is a graph of forecasting results using the ARIMA-GARCH method.

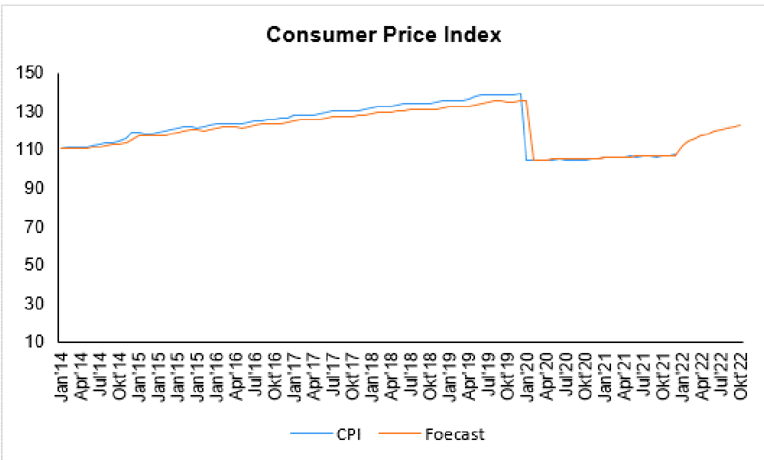


***Figure 4.** Graph of GARCH forecasting results.*

Based on the graph in ***Figure. 4***, the forecasted data is sufficient to follow the data pattern or does not differ much from the actual data pattern. If viewed from the graph, it can be seen that the Indonesian Consumer Price Index in the coming period tends to increase. For more details can be seen in ***Table. 4.***

***Table 4.*** *Forecasting Results.*

| Period (2022) | Consumer Price Index (CPI) |
|---|---|
| January | 112,1116 |
| February | 114,1788 |
| March | 115,7674 |
| April | 117,1077 |
| May | 118,2894 |

## 4 CONCLUSIONS

Based on the results and discussion, it can be concluded that: Based on the analysis of Indonesian consumer price index data, it can be concluded that the ARIMA (2,2,1) and GARCH (1,0) models are quite good models in modeling Indonesian consumer price index data. Forecasting results for the 2022 period are obtained sequentially, namely 112.1116, 114.1788, 115.7674, 117.1077, 118.2894, which means that the forecasted data is sufficient to follow the data pattern or does not differ much from the actual data pattern and tends to increase in the coming period. For further research, it is expected to be able to conduct research related to other financial data such as exchange rates, stocks, import values, gold prices and others. And can use data that would have high volatility or fluctuations in data that is not constant (heteroscedasticity).

## REFERENCES

[1] M. Hali Mukron, I. Susianti, F. Azzahra, Y. Nur Kumala, F. Risnita Widiyana, and M. Al Haris, "Peramalan Indeks Harga Konsumen Indonesia Menggunakan Autoregressive Integrated Moving Avarage," J. Stat. Ind. dan Komputasi, vol. 6, no. 1, pp. 20–25, 2021.

[2] I. S. Larasati and S. Sulasmiyati, "Pengaruh inflasi, ekspor dan tenaga kerja terhadap Produk Domestik Bruto," J. Adm. Bisnis, vol. 63, no. 1, pp. 8–16, 2018, [Online]. Available: http://administrasibisnis.studentjournal.ub.ac.id/index.php/jab/article/view/2694

[3] S. Syahdan and S. Aisyah, "Peramalan Indekx Harga Konsumen (IHK) Kota Tarakan dengan Metode Double Exponential Smoothing dari Brown Peramalan Indeks Harga Konsumen (Ihk) Kota Tarakan Dengan Metode Double Exponential Smoothing Dari Brown (Forecasting Of Consumer Price Index (Cpi) Ta," J. Mat. dan Pendidik. Mat., vol. 5, no. 1, p. 54, 2020.

[4] S. I. Wardani, W. P. Nurmayanti, S. A. Wulandya, and M. Gazali, "Implementation Of Exponential Smoothing And Arima On Collection And Distribution Data Of Zakat, Infaq And Shadaqah 2015-2020," 2021.

[5] P. Volatilitas, M. Garch, and B. O. X. Cox, "Dengan Volatilitas Lag-1 Ditransformasi," no. 1992, 2019.

[6] N. Angraeny, "Penerapan Metode Arch Garch Untuk Analisis Peramalan Nilai Ekspor Indonesia," p. 53, 2019, [Online]. Available: https://lib.unnes.ac.id/37556/1/4112316006.pdf

[7] A. Muhayani, Perbandingan metode aparch, egarch dan tgarch untuk prakiraan harga emas dunia. 2019. [Online]. Available: http://repository.uinjkt.ac.id/dspace/handle/123456789/55261

[8] F. C. Garini and W. Anbiya, "Application of GARCH Forecasting Method in Predicting The Number of Rail Passengers (Thousands of People) in Jabodetabek Region," J. Mat. Stat. dan Komputasi, vol. 18, no. 2, pp. 198–223, 2022, doi: 10.20956/j.v18i2.18382.

[9] Novi yulia Budiarti, "No Title," Sustain., vol. 4, no. 1, pp. 1–9, 2020, [Online].

[10] F. H. Syarofi, "Analisis Pengaruh Suku Bunga SBI, Kurs Rupiah/US$, Harga Minyak Dunia, Harga Emas Dunia, DJIA, Nikkei 225, dan Hang Seng Index Terhadap IHSG Dengan Metode Garch-M," Skripsi, pp. 1–109, 2014.