

## COMPARISON OF ARIMA AND GARMA'S PERFORMANCE ON DATA ON POSITIVE COVID-19 CASES IN INDONESIA

A'yunin Sofro<sup>1\*</sup>, Khusnia Nurul Khikmah<sup>2</sup>

<sup>1</sup>Department of Mathematics, Faculty of Mathematics and Natural Sciences, University State of Surabaya  
Ketintang Selatan St., Surabaya, East Java, 60231, Indonesia

<sup>2</sup>Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University  
Raya Dramaga St., Bogor, West Java, 16680, Indonesia

Corresponding author's e-mail: <sup>1\*</sup> [ayuninsofro@unesa.ac.id](mailto:ayuninsofro@unesa.ac.id)

---

**Abstract.** The development of methods in statistics, one of which is used for prediction, is overgrowing. So it requires further analysis related to the goodness of the method. One of the comparisons made to the goodness of this model can be seen by applying it to actual cases around us. The real case still being faced by people worldwide, including in Indonesia, is Covid-19. Therefore, research comparing the autoregressive integrated moving average (ARIMA) and the Gegenbauer autoregressive moving average (GARMA) method in positive confirmed cases of Covid-19 in Indonesia is essential. Based on the results of this research analysis, it was found that the best model with the Aikake's Information Criterion measure of goodness that was used to predict positive confirmed cases of Covid-19 in Indonesia was the Gegenbauer autoregressive moving average (GARMA) model.

**Keywords:** ARIMA, GARMA, AIC, Covid-19, Indonesia.

---

### Article info:

Submitted: 14<sup>th</sup> June 2022

Accepted: 29<sup>th</sup> August 2022

### How to cite this article:

A. Sofro and K. N. Khikmah, "COMPARISON OF ARIMA AND GARMA'S PERFORMANCE ON DATA ON POSITIVE COVID-19 CASES IN INDONESIA", *BAREKENG: J. Math. & App.*, vol. 16, iss. 3, pp. 919-926, September, 2022.

---



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).  
Copyright © 2022 Author(s)

## 1. INTRODUCTION

The development of statistical science is used for future data analysis to determine policy, one of which aims to assist in determining policy [1]. Analysis of future data is one of the most critical factors. Namely, the data used depends on past data that influence it, usually called forecasting. Forecasting is processing past data to get estimates of future data [2]. Some methods used to predict time series data are the autoregressive integrated moving average (ARIMA) [3] and the Gegenbauer autoregressive moving average (GARMA) [4].

The autoregressive integrated moving average (ARIMA) method is a univariate time series analysis method in which the model structure consists of Autoregressive (AR) and Moving Average (MA). Autoregressive (AR) models the autocorrelation of time series variables that depend linearly on the values of the previous variables [5]. Moving Average (MA) models the autocorrelation of previous errors in the time series [6]. While the Gegenbauer autoregressive moving average (GARMA) is a generalized model of the generalized ARFIMA model [4], which this model was introduced by Granger, Joyeux and Jonathan Hosking in 1981. This model also provides good accuracy in forecasting data [7].

The ARIMA and GARMA methods are used to analyze various fields, some of which are in the fields of [8], climate and weather [9], health [10], etc. In the health sector, one of these methods is used to analyze global urgency, namely predicting positive cases of Covid-19. *Covid-19* is caused by the acute respiratory syndrome coronavirus-2 and is classified as an infectious disease [11]. This disease was first reported in Wuhan, China, and spread worldwide, including in Indonesia. The first time this disease entered Indonesia was until 27 May 2022. This disease has infected more than six million Indonesians.

Therefore, this study aims to compare the best time series forecasting methods between the autoregressive integrated moving average (ARIMA) and the Gegenbauer autoregressive moving average (GARMA) with a case study of the number of COVID-19 cases in Indonesia.

## 2. RESEARCH METHODS

### 2.1 Autoregressive (AR)

The autoregressive model is a model which states that the data in the previous period affects the data for the current period. The model is often referred to as a model that processes the regression results by itself. If  $\phi_1, \phi_2, \dots, \phi_p$  is an autoregressive parameter coefficient with the order  $p$ , then mathematically, the  $AR(p)$  equation can be written [12]:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} \quad (1)$$

### 2.2 Moving Average (MA)

The moving average model is a model which states that the data in the  $t$ - $t$  period has a dependency relationship with the error values up to the  $t$ - $q$  period. If  $\theta_1, \theta_2, \dots, \theta_q$  is a moving average parameter coefficient with the order  $q$ , then mathematically, the equation  $MA(q)$  can be written [12]:

$$X_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (2)$$

### 2.3 Autoregressive Integrated Moving Average (ARIMA)

The autoregressive integrated moving average model is a combined model between the autoregressive (AR) and moving average (MA) models assuming that the resulting time-series data is not stationary, so there is a differencing process. Data that is stationary on a time series has two possibilities. Namely, it is not stationary concerning the mean so that a differencing process can handle it. Meanwhile, if the data is not stationary, the variance can be handled by performing data transformations [13]. If  $X_t$  is the value in the  $t$ -th observation,  $\phi_p$  is an autoregressive parameter of order  $p$ ,  $B$  is the backshift operator,  $d$  is the differencing value,  $\mu$  is a constant,  $\theta_q$  is a moving average parameter of order  $q$ ,  $e_t$  and is a residual. So, mathematically the  $ARIMA(p, d, q)$  equation can be written [14]:

$$\phi_p(B)(1 - B)^d X_t = \mu + \theta_q(B)e_t \quad (3)$$

## 2.4 Gegenbauer Autoregressive Moving Average (GARMA)

Gegenbauer Autoregressive Moving Average (GARMA) is one of the statistical forecasting methods introduced by Granger Joyeux and Hosking in 1980 and 1981 [15]. Based on previous research, this method is claimed to be able to predict with good results because it can capture the stochastic properties of the data. It was Complicated by fractional differentiation. If  $\phi_p(B)$  is the coefficient of the autoregressive parameter of order  $p$ ,  $(1 - 2uB + B^2)^d$  is the Gegenbauer component,  $\theta_q(B)$  is the coefficient of the parameter of order  $q$ ,  $e_t$  is the residual of hhg. So, mathematically the equation  $GARMA(p, d, u, q)$  can be written [16], [17]:

$$\phi_p(B)(1 - 2uB + B^2)^d X_t = \mu + \theta_q(B)e_t \quad (4)$$

## 2.5 Accuracy

### 1. Aikake's information criterion

Aikake's information criterion is one method of measuring goodness in model selection. This model has calculated the model's goodness based on the maximum likelihood estimation method, where the model with the smallest Aikake's information criterion value is the best model that can be continued for analysis. If  $k$  is the model parameter and  $L$  is the estimated value of the maximum likelihood estimation method. Then mathematically, the AIC equation can be written [18]:

$$AIC = 2k - 2 \ln(L) \quad (5)$$

### 2. Root Mean Square Error (RMSE)

Root mean square error is a method of measuring the accuracy of forecasting results, where this method measures the level of error from the results of analytical calculations and actual data. Forecasting results are good accuracy if the resulting RMSE value is getting smaller. If  $X_t$  is actual data and  $\hat{X}_t$  is analysis result data. Then mathematically, the RMSE equation can be written [19]:

$$RMSE = \sqrt{\sum_{t=1}^n (X_t - \hat{X}_t)^2} \quad (6)$$

### 3. Mean Absolute Percentage Error (MAPE)

Mean absolute percentage error is a method of measuring the accuracy of forecasting results, where this method measures the relative level of error from the results of analytical calculations with actual data. This accuracy value provides information on the percentage of an error where if the value is lower, the forecasting results have good accuracy. If  $X_t$  is actual data and  $\hat{X}_t$  is analysis result data. Then mathematically, the MAPE equation can be written [20]:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \hat{X}_t}{X_t} \right| \times 100\% \quad (7)$$

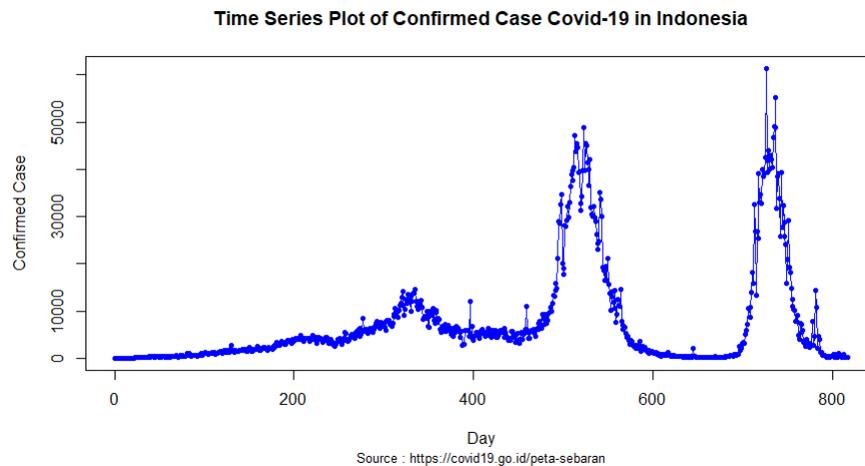
## 2.6 Data and Analysis Method

This study uses data on positive cases of Covid-19 in Indonesia. This data was taken from the official website of the Task Force for the Acceleration of Handling COVID-19 ([Peta Sebaran | Covid19.go.id](https://peta.sebaran-covid19.go.id)), where data was taken from positive cases of Covid-19 in Indonesia from March 2, 2020, to May 27, 2022. The data is then divided into two, namely, data training and data testing. The training data used to model the Covid-19 positive case data is 80% of the total data, while the testing data uses 20%.

In general, this study has two stages of modelling, namely the autoregressive integrated moving average (ARIMA) and gegenbauer autoregressive moving average (GARMA), namely: preprocessing data, testing data stability, identifying temporary models, estimating model parameters, diagnostic tests residual model, model selection, calculation of the best model forecasting accuracy.

### 3. RESULTS AND DISCUSSION

This study analyzes positive cases of Covid-19 in Indonesia using the autoregressive integrated moving average and generalized autoregressive moving average methods. Data on positive cases of Covid-19 in Indonesia is time-series data visually presented in Figure 1 and descriptively presented in Table 1.



**Figure 1.** Time-series plot of data on positive cases of Covid-19 in Indonesia

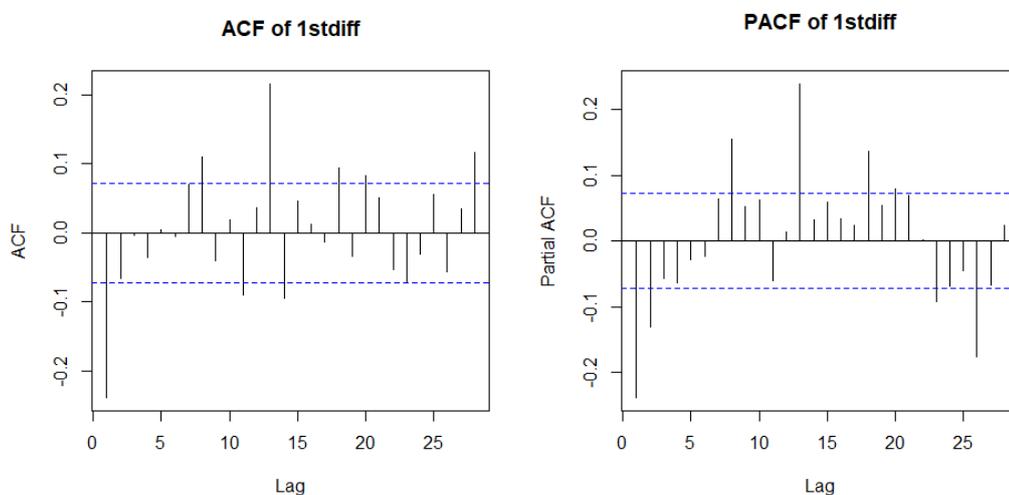
**Table 1.** Descriptive statistics of the data on positive cases of Covid-19 in Indonesia

Min	First Quartile	Median	Third Quartile	Max	Mean	Standard Deviation	Variance
0	837	3924	7514	61361	7390	10645	113337033

*Data source: R software Output, results of descriptive statistics of the data on positive cases of Covid-19 in Indonesia*

Modeling and predicting using the autoregressive integrated moving average model and generalized autoregressive moving average models for the first stage is splitting was carried out with 90% of the data used as training data and 10% of the data used as testing data. Furthermore, based on the time series plot in Figure 1, the data on positive cases of Covid-19 in Indonesia is a time-series data containing a trend pattern so that a stationary check will be carried out on the training data with the Augmented Dickey-Fuller test.

The Augmented Dickey-Fuller test carried out in this study used a significance level of 5%, and it was concluded that the data on positive cases of Covid-19 in Indonesia with a one-time differencing process concluded that the data on positive cases of Covid-19 was stationary, so that further it could be continued with the determination of candidates. The model based on the ACF and PACF plots is presented in Figure 2.



**Figure 2.** ACF and PACF plot of first difference training data on positive cases of Covid-19 in Indonesia

The candidates for the autoregressive integrated moving average model from data on positive cases of Covid-19 in Indonesia are presented in Table 2.

**Tabel 2. Candidates models of ARIMA model**

Candidates Model	Coefficient of parameter	AIC
ARIMA(0,1,1)	$\theta_1=0.30$ $\mu=65.33$	13400.34
ARIMA(2,1,0)	$\phi_1=-0.27$ $\phi_2=-0.13$ $\mu=65.46$	13401.66
ARIMA(2,1,1)	$\phi_1=0.18$ $\phi_2=-0.02$ $\theta_1=0.46$ $\mu=64.46$	13399.19
ARIMA(0,1,2)	$\theta_1=0.28$ $\theta_2=0.08$ $\mu=64.48$	13396.78

*Data source: R software Output, results of ARIMA Model of the data on positive cases of Covid-19 in Indonesia*

The best ARIMA model chosen to be continued in the analysis based on Table 2 is ARIMA(0,1,2) with the smallest AIC value. This best model is then continued with an analysis of the residuals with a diagnostic test of residuals, and the remainders meet the white noise assumption. So that the best model formed mathematically can be written as follows:

$$\phi_p(B)(1-B)^d Y_t = \mu + \theta_q(B)e_t$$

$$Y_t = 64.48 + Y_{t-1} + e_t + 0.28e_{t-1} + 0.08e_{t-2}$$

The ARIMA model is compared with the generalized autoregressive moving average in forecasting accuracy to determine the suitability of the data with the method used. The data is modeled using the generalized autoregressive moving average, the candidate model is determined based on the ACF and PACF plots, and the generalized autoregressive moving average model candidate is shown in Table 3.

**Tabel 3. Candidates models of GARMA model**

Candidates Model	Coefficient of parameter	AIC
GARMA(0,1,1,1)	$\theta_1=0.30$ $\mu=65.19$	13400.24
GARMA(2,1,1,0)	$\phi_1=-0.27$ $\phi_2=-0.13$ $\mu=65.18$	13401.57
GARMA(2,1,1,1)	$\phi_1=0.18$ $\phi_2=-0.02$ $\theta_1=0.46$ $\mu=64.18$	13399.08
GARMA(0,1,1,2)	$\theta_1=0.28$ $\theta_2=0.08$ $\mu=64.48$	13397.13

*Data source: R software Output, results of GARMA Model of the data on positive cases of Covid-19 in Indonesia*

The best generalized autoregressive moving average model for further research is the GARMA(0,1,1,2) model. This model has a diagnostic test on the remainder and fulfills the assumption of white noise. So that the best GARMA model formed mathematically can be written as follows:

$$\phi_p(B)(1 - 2uB + B^2)^d X_t = \mu + \theta_q(B)e_t$$

$$X_t = 64.48 + 2X_{t-1} - X_{t-2} + e_t - 0.28e_{t-1} - 0.08e_{t-2}$$

Based on the two best models, ARIMA(0,1,2) and GARMA(0,1,1,2), which compare their accuracy and goodness in modeling positive cases of Covid-19 in Indonesia, we will compare their accuracy in predicting positive cases of Covid-19 in Indonesia. The results of the prediction accuracy of the two best models are presented in Table 4.

**Table 4. Accuracy models between ARIMA and GARMA model**

Accuracy	ARIMA(0,1,2)	GARMA(0,1,1,2)
MAE	9482.94	9381.69
MASE	3.77	3.73
RMSE	15976.03	15850.72
MAPE	0.93	0.85

*Data source: R software Output, results of accuracy Model of ARIMA and GARMA models*

Based on the results of the accuracy of the ARIMA(0,1,2) and GARMA(0,1,1,2) models in predicting positive cases of Covid-19 in Indonesia, it can be obtained that the GARMA(0,2) model has accuracy in predicting data on positive cases of Covid-19 in Indonesia, which is better than the ARIMA(0,1,2) model. This can be seen based on the MAE, MASE, RMSE, and MAPE values of the GARMA(0,1,1,2) model, which are smaller than the ARIMA(0,1,2) model.

#### 4. CONCLUSIONS

This study provides several results based on the analysis using the autoregressive integrated moving average (ARIMA) and the Gegenbauer autoregressive moving average (GARMA) that has been carried out above. The best model for predicting positive cases of Covid-19 in Indonesia is the GARMA(0,1,1,2) model. Where the GARMA(0,1,1,2) model has better accuracy than ARIMA, these results are seen based on MAE, MASE, RMSE, and MAPE values. This analysis of the comparison of the autoregressive integrated moving average (ARIMA) and the Gegenbauer autoregressive moving average (GARMA) model can be used to select the best forecasting method. For further research, it is hoped that this best method can be compared with methods that have recently emerged, such as using deep learning.

#### AKNOWLEDGEMENT

If so, write an Acknowledgment or appreciation in this section. Acknowledgments can be addressed to funders (sponsors) who contributed to the article. It could be also to people who contributed to the article or data in the article.

#### REFERENCES

- [1] M. A. Hernán, J. Hsu, and B. Healy, "A second chance to get causal inference right: a classification of data science tasks," *Chance*, vol. 32, no. 1, pp. 42–49, 2019.
- [2] T. Chakraborty and I. Ghosh, "Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis," *Chaos, Solitons & Fractals*, vol. 135, p. 109850, 2020.
- [3] A. Hernandez-Matamoros, H. Fujita, T. Hayashi, and H. Perez-Meana, "Forecasting of COVID19 per regions using ARIMA models and polynomial functions," *Appl. Soft Comput.*, vol. 96, p. 106610, 2020.

- [4] G. S. Dissanayake, M. S. Peiris, and T. Proietti, "Fractionally differenced Gegenbauer processes with long memory: A review," *Stat. Sci.*, vol. 33, no. 3, pp. 413–426, 2018.
- [5] E. Beard *et al.*, "Understanding and using time series analyses in addiction research," *Addiction*, vol. 114, no. 10, pp. 1866–1884, 2019.
- [6] M. Fernandes, A. Canito, J. M. Corchado, and G. Marreiros, "Fault detection mechanism of a predictive maintenance system based on autoregressive integrated moving average models," in *International Symposium on Distributed Computing and Artificial Intelligence*, 2019, pp. 171–180.
- [7] H. Yan, G. W. Peters, and J. Chan, "Mortality models incorporating long memory for life table estimation: a comprehensive analysis," *Ann. Actuar. Sci.*, vol. 15, no. 3, pp. 567–604, 2021.
- [8] S. Ben Amor, H. Boubaker, and L. Belkacem, "Forecasting electricity spot price for Nord Pool market with a hybrid k-factor GARMA–LLWNN model," *J. Forecast.*, vol. 37, no. 8, pp. 832–851, 2018.
- [9] F. Maddanu and T. Proietti, "Modelling Persistent Cycles in Solar Activity," *Sol. Phys.*, vol. 297, no. 1, pp. 1–22, 2022.
- [10] H. Yan, G. W. Peters, and J. S. K. Chan, "Multivariate long-memory cohort mortality models," *ASTIN Bull. J. IAA*, vol. 50, no. 1, pp. 223–263, 2020.
- [11] Y. Cao, Q. Deng, and S. Dai, "Remdesivir for severe acute respiratory syndrome coronavirus 2 causing COVID-19: An evaluation of the evidence," *Travel Med. Infect. Dis.*, vol. 35, p. 101647, 2020.
- [12] J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, and A. Lachhab, "Forecasting of demand using ARIMA model," *Int. J. Eng. Bus. Manag.*, vol. 10, p. 1847979018808673, 2018.
- [13] M. B. Shrestha and G. R. Bhatta, "Selecting appropriate methodological framework for time series data analysis," *J. Financ. Data Sci.*, vol. 4, no. 2, pp. 71–89, 2018.
- [14] K. E. ArunKumar, D. V Kalaga, C. M. S. Kumar, G. Chilkoor, M. Kawaji, and T. M. Brenza, "Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Averag," *Appl. Soft Comput.*, vol. 103, p. 107161, 2021.
- [15] C. W. J. Granger and R. Joyeux, "An introduction to long-memory time series models and fractional differencing," *J. time Ser. Anal.*, vol. 1, no. 1, pp. 15–29, 1980.
- [16] A. K. Diongue and M. Ndong, "The k-factor GARMA process with infinite variance innovations," *Commun. Stat. Comput.*, vol. 45, no. 2, pp. 420–437, 2016.
- [17] P. Beaumont and A. Smallwood, "Conditional Sum of Squares Estimation of Multiple Frequency Long Memory Models," 2019.
- [18] J. E. Cavanaugh and A. A. Neath, "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 11, no. 3, p. e1460, 2019.
- [19] W. Wang and Y. Lu, "Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model," in *IOP conference series: materials science and engineering*, 2018, vol. 324, no. 1, p. 12049.
- [20] S. Theocharides, G. Makrides, A. Livera, M. Theristis, P. Kaimakis, and G. E. Georghiou, "Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing," *Appl. Energy*, vol. 268, p. 115023, 2020.

