

Modelling of Cayenne Production in Central Java Using ARIMAGARCH

by Tarno Tarno

Submission date: 25-Jul-2022 08:36PM (UTC+0700)

Submission ID: 1875010455

File name: Modelling_of_cayenne_production_in_Central_Java_using_ARIMA.pdf (721.24K)

Word count: 2774

Character count: 14195

PAPER · OPEN ACCESS

Modelling of cayenne production in Central Java using ARIMA-GARCH

To cite this article: Tamo *et al* 2018 *J. Phys.: Conf. Ser.* **1025** 012120

View the [article online](#) for updates and enhancements.

You may also like

- [Starbursts Triggered by Central Overpressure. II. A Parameter Study](#)
Chanda J. Jog and Mousumi Das

- [A Study of Central Peaks and Central Pits on the Moon and Vesta](#)
Kailly Nocera, Michael F. Zeinhofer, Matthew Maritorana *et al.*

- [Roles of GARCH and ARCH effects on the stability in stock market crash](#)
Hai-Feng Li, Dun-Zhong Xing, Qian Huang *et al.*



ECS The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Early hotel & registration pricing ends September 12

Presenting more than 2,400 technical abstracts in 50 symposia

The meeting for industry & researchers in

BATTERIES
ENERGY TECHNOLOGY
SENSORS AND MORE!

 Register now!

 **ECS Plenary Lecture featuring M. Stanley Whittingham,**
Binghamton University
Nobel Laureate –
2019 Nobel Prize in Chemistry



Modelling of cayenne production in Central Java using ARIMA-GARCH

Tarno¹, Sudarno¹, Dwi Ispriyanti¹ and Suparti¹

¹Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro

Email: tarno.stat@gmail.com

Abstract.

Keywords: time series, cayenne production, ARIMA, GARCH

1. Introduction

Some regencies/cities in Central Java Province are the producers of horticultural crops in Indonesia, for example, Brebes which is the largest area of shallot producer in Central Java, while the others, such as Cilacap and Wonosobo are the areas of cayenne commodities production. Currently, cayenne is a strategic commodity and has broad impact to Indonesian economic development. Modelling the cayenne production is necessary to predict the commodity need for society. The price of this commodity fluctuates randomly and the market players should consider carefully about the possible market risks happening. Based on the reality, the decreasing of cayenne production will cause the increasing of society's basic needs price, and finally it will affect the inflation level at that area or even that country.

The cayenne production data are classified into time series data. To date, autoregressive integrated moving average (ARIMA) is the most popular model in time series analysis which proposed by Box-Jenkins. General procedure of ARIMA modelling covers identification, parameter estimation, model diagnostic and forecasting [1]. This procedure is very simple when implemented for linear and stationary data. When time series data have heteroscedasticity effect, generalized autoregressive conditional heteroscedasticity (GARCH) model can be implemented.

According to the statements above, this research focused on ARIMA modelling by considering the effect of autoregressive conditional heteroscedasticity (ARCH) to study about cayenne production in



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Published under licence by IOP Publishing Ltd

Central Java. The empirical study of ARIMA-GARCH is implemented for constructing model of cayenne production in Central Java based on data from January 2003 to November 2015.

2. Box-Jenkins ARIMA Model

Box-Jenkins methodology for ARIMA modelling of time series data includes some procedures that consist of identification, parameter estimation, model verification, and forecasting based on the best model which has been built [1]. Nonseasonal Box-Jenkins ARIMA model for stationary process consists of autoregressive order-p or AR(p), moving average order-q or MA(q) and mixed model which written as ARMA(p,q). While nonseasonal ARIMA model for nonstationary process consists of ARIMA(p,d,0), ARIMA(0,d,q) and ARIMA(p, d, q) where p: level of autoregressive, q: level of moving average and d: level of differencing.

2.1 Box-Jenkins Method

Box-Jenkins method for time series analysis uses backshift operator B that defined as $BZ_t = Z_{t-1}$ and difference operator ∇ defined as $\nabla Z_t = Z_t - Z_{t-1} = (1 - B)Z_t$. These operators have the relation of $\nabla = 1 - B$ and satisfy the rules of elementary algebra. The statistics process in the form of

$$\phi(B)Z_t = \theta(B)a_t$$

is often used in practice, where ϕ and θ as polynomials, and $\{a_t\}$ is a sequence which generated by white noise. Sequence $\{a_t\}$ are independent and distributed normally with mean 0 and constant variance σ^2 .

2.2 Autoregressive (AR) Process

The general formula of autoregressive process order-p or AR(p) is:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (1)$$

where $a_t \sim N(0, \sigma_a^2)$. It can be seen that Z_t is regressed to p previous values Z_{t-k} , $k = 1, 2, \dots, p$.

The equation (1) can be written as:

$$\phi(B)Z_t = a_t$$

where $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ is called an operator of AR(p). The requirement for stationary AR(p) process is all the roots of $\phi(B) = 0$ are located outside the unit circle. AR(p) process is always invertible.

2.3 Moving Average (MA) Process

Moving average order q or MA(q) can be written as

$$Z_t = a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (2)$$

where a_t are independent and distributed normally with mean 0 and variance σ_a^2 . The equation (2) can be written as

$$Z_t = \theta(B)a_t \quad (3)$$

where $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$. Equation (3) can also be written as

$$\theta^{-1}(B)Z_t = a_t \quad \text{or} \quad \pi(B)Z_t = a_t \quad (4)$$

MA(q) process is invertible if the value of coefficient π is convergent series which is if and only if all the roots $\theta(B) = 0$ located outside the unit circle.

2.4 Mixed Process Autoregressive Moving Average (ARMA)

An equation gained from AR and MA model is a mixed model written as

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (5)$$

and called ARMA(p,q) model. Model (5) can be written as:

$$\phi(B)Z_t = \theta(B)a_t.$$

ARMA model is stationary and invertible, when roots of $\phi(B) = 0$ and $\theta(B) = 0$ are located outside the unit circle.

2.5 Nonstationary Time Series

Stationary time series is rarely found in real life, but stationarity is useful assumption while studying time series. There are a lot of things that cause nonstationarity of time series. Some examples of dataset which include in this category are stock exchange, number of sales in a corporation, exchange rate. Expenses time series like consumption expenses for old goods, government expenses, or corporate expenses for capital goods are also nonstationary time series.

Making the difference between the consecutive values of the nonstationary time series, homogeneous is a way to create time series data to be stationary. Furthermore, if it is defined into sequence of difference $W_t = Z_t - Z_{t-1}$, the general process of ARIMA can be written as

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (6)$$

Equation (6) can be seen as autoregressive integrated moving average (ARIMA). In many cases, the first difference of the time series is still nonstationary, may be the second difference is stationary. By writing the difference degree as d, the ARIMA process can be drawn as dimension p, d and q. So, ARIMA(p,d,q) means nonstationary time series difference is taken to d into stationary.

3. ARCH/GARCH Model

“Volatile” behaviour in financial market is usually referred as “volatility”. Volatility has been the important concept in theoretical and practical of finance, such as risk management, portfolio selection and others [2]. In statistical study, it is usually measured using variance or standard deviation. Engle in 1982 developed a volatility model for financial time series data which known as autoregressive conditional heteroscedasticity (ARCH) [3] and the more flexible model as known generalized autoregressive conditional heteroscedasticity (GARCH) has been developed by Bolerslev [4].

3.1 ARCH Model

If Z_t is a stationary time series, like financial return, Z_t can be stated as its mean added by a white noise if there is no significant autocorrelation in Z_t itself, which is:

$$Z_t = \mu_t + a_t \quad (7)$$

where μ_t as mean process of Z_t and $a_t = \sigma_t v_t$ with $v_t \sim N(0,1)$.

Seeing volatility clustering or conditional heteroscedasticity, assumed that $Var_{t-1}(a_t) = \sigma_t^2$ with $Var_{t-1}(\bullet)$ stated the conditional variance with information until (t-1) were given, and

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_p a_{t-p}^2 \quad (8)$$

Because a_t has mean 0 and $Var_{t-1}(a_t) = E_{t-1}(a_t^2) = \sigma_t^2$, hence, equation (7) becomes

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_p a_{t-p}^2 + u_t \quad (9)$$

where $u_t = a_t^2 - E_{t-1}(a_t^2)$ is a white noise with mean 0. Model (7) and (8) called as ARCH model [3].

3.2 ARCH Effect Test

Before estimating the ARCH model for time series data, testing for presence of ARCH effect in the residual should be done. If there is no ARCH effect in the residual, ARCH model is not conditional. The ARCH model can be expressed as AR model in the squared residual components as the eq. (9).

Lagrange Multiplier (LM) test is a simple procedure for testing the ARCH effect. The steps of testing can be constructed in the regression eq. (9) as follows.

Null hypothesis: $\alpha_1 = \alpha_2 = \dots = \alpha_p = 0$ (no effect of ARCH).

Statistical test: $LM = TR^2 \sim \chi^2(p)$ where T : sample size and R^2 is measured from regression of residual (9).

3.3 GARCH Model

If testing for ARCH effect is significant, the ARCH models could be estimated and the estimated volatility σ_t can also be obtained based on the past information. In practice, it often acquired large enough of lag p , the number of parameters to be estimated in the model are also quite large.

The proposed model that is more parsimonious to replace the AR model (8) with the following formulation [4].

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i a_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (10)$$

With all the coefficients $\alpha_i > 0$ ($i = 0, 1, \dots, p$) and $\beta_j > 0$ ($j = 1, \dots, q$) to ensure that the conditional variance σ_t^2 is always positive. Equation (10) together with (7) is known as generalized ARCH or GARCH (p, q). If $q=0$ then GARCH model becomes ARCH model.

4. Procedure of ARIMA Modelling

4.1 Model Identification

Testing for stationarity of time series data can be identified by time series plot or autocorrelation function (ACF) plot. The time series data are said stationary if, the plot of ACF shows dies down or exponential decays, whereas if the ACF drops slowly, the time series is said nonstationary [5]. If the time series is not stationary, the data should be stationary by doing differencing.

By changing the original time series to the differencing time series, ARMA model can be identified for the differencing data. Differencing data should be done continuously until the time series data at a specific level becomes stationary. When the stationary time series is achieved, the order of ARIMA model can be identified.

4.2 Parameter Estimation

Estimation of the parameters for tentative model can be done by using least squares (OLS) or maximum likelihood estimator (MLE).

4.3 Model Verification

The model estimated in step 4.2 can be verified by considering its residual properties, that its residual should be independent and distributed normally with mean 0 and its variance is constant. Ljung-Box statistics Q can be applied for testing appropriateness of this model. The test statistics Q is as follow [6]:

$$Q_m = n(n+2) \sum_{k=1}^m \frac{r_k^2(e)}{n-k} \sim \chi_{m-r}^2$$

where $r_k^2(e)$: autocorrelation of residual in the k -th lag,

n : number of residual,

m : number of time lag included in testing.

If at the residual analysis indicated that the residual is not independent and the variance is not constant (heteroscedasticity) so ARCH effect with LM test detection should be done. If ARCH effect is detected, ARCH/GARCH model estimation for that variance should be done.

4.4 Prediction/Forecasting

Prediction for one step ahead or k-steps ahead can be done by the constructed model. The performance of model can be evaluated by using AIC, BIC or RMSE criterion.

5. Results and Discussion

The data which implemented in this research are monthly data from January 2003 until November 2015 of cayenne production in Central Java. Data were collected from <https://jateng.bps.go.id>. The procedure of ARIMA modelling for the dataset are as follows.

5.1 Model Identification

Model of ARIMA is identified by plotting time series data, plotting autocorrelation function (ACF), and plotting partial autocorrelation function (PACF). The time series plot, ACF, and PACF are illustrated in Figure 3. Based on ACF plot can be said that data is stationary in mean because ACF plot form sinus wave. PACF plot shows that lag-1 and lag-3 are significant different to zero, so the subset ARIMA model in this case ARIMA([1,3],0,0) can be identified as an appropriate model. For determining order of subset ARIMA can be evaluated by using overfitting concept [7].

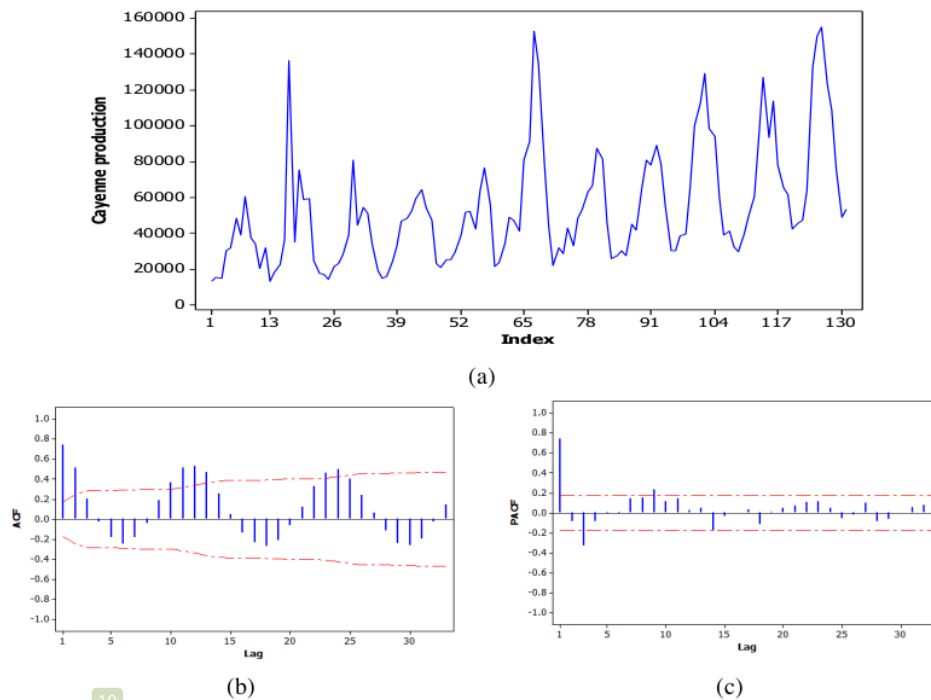


Figure 1. (a). Time series plot, (b). ACF plot, (c). PACF plot of cayenne production data

5.2 Parameter Estimation

The results of parameter estimation based on the monthly data for tentative model of ARIMA ([1,3],0,0) are shown in Table 1. From those results can be explained that the constant term,

coefficient of lag-1, and coefficient of lag-3 significantly influential to the data, because the p-value related to each parameter is less than level of significant 5%.

Table 1. Estimated parameter of ARIMA([1,3],0,0)

| Variables | Coefficient | Standard Error | t-statistic | Prob. |
|-----------|-------------|----------------|-------------|--------|
| Constant | 53065.860 | 6872.827 | 7.721 | 0.0000 |
| AR(1) | 0.974 | 0.073 | 13.374 | 0.0000 |
| AR(3) | -0.292 | 0.097 | -3.014 | 0.0026 |

5.3 Model verification

Residual analysis should be done when the tentative model has been estimated. The aim of residual analysis is to verify the appropriate model. Verification of model covers testing for independence of residual, normality and homoscedasticity. Testing for independence of residual using Durbin Watson (DW) yielded statistics DW equal to 2.428. So the residual satisfies the assumption of independency. According to normality test of residual using Jarque-Bera test shows that p-value is less than significance level 5% (p-value=0.00) (see the explanation of Figure 4). It can be concluded that residual is not distributed normally with mean zero and constant variance..

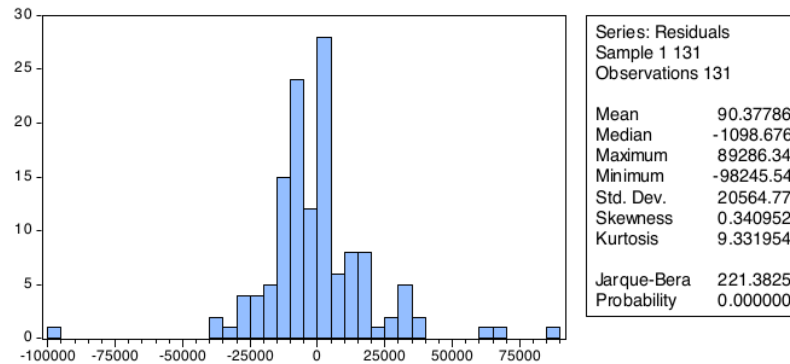


Figure 2. Normality test of residual

The results of testing for homoscedasticity using LM-test conclude that the variance of residual is not constant. So that the residual of model have ARCH effect (see Table 5).

Table 2. ARCH effect test for Residual

| Variable | Coefficient | Standard Error | t-statistic | Prob. |
|-------------|-------------|----------------|-------------|-------|
| Constant | 309000000 | 31643067 | 9.759 | 0.000 |
| a_{t-1}^2 | 0.201 | 0.099 | 2.033 | 0.042 |

The estimated model can be written as:

$$Z_t = 53065.86 + 0.974Z_{t-1} - 0.292Z_{t-3} + a_t \quad (11)$$

with $a_t \sim N(0, \sigma_t^2)$; $\sigma_t^2 = 309000000 + 0.201a_{t-1}^2$.

Equation (11) is ARIMA model with ARCH effect. Furthermore ARIMA([1,3],0,0)-GARCH(1,0) is an appropriate model for forecasting.

5.4 Prediction/Forecasting

Prediction results of model (11) based on *in sample* data shows value of root mean squares error (RMSE), Akaike Information Criterion (AIC), and Schwars Information Criterion (SIC) consecutively is equal to 20567; 28.37 and 28.44. The result of prediction using the model is shown in Figure 5.

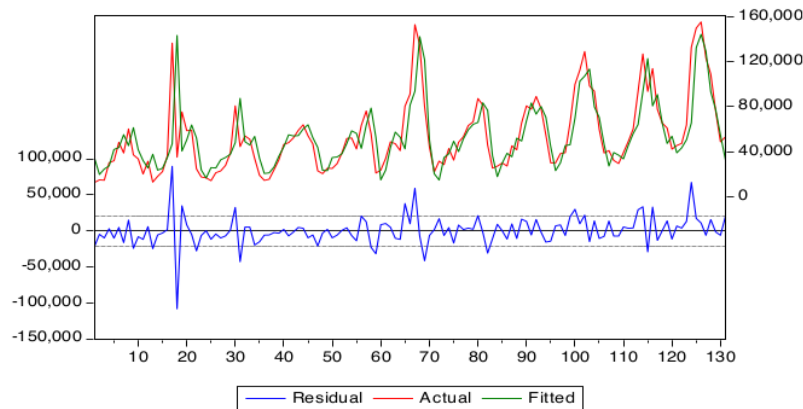


Figure 3. Plot *in sample* data and prediction result using AR([1,3],0,0)-ARCH(1)

6. Conclusion

Based on the results and discussion, the best model for monthly data from January 2009 up to November 2015 of cayenne production in Central Java is ARIMA([1,3,0,0]-GARCH(1,0)). The formula of the model is as follow:

$$Z_t = 53065.86 + 0.974Z_{t-1} - 0.292Z_{t-3} + a_t$$

with $a_t \sim N(0, \sigma_t^2)$; $\sigma_t^2 = 309000000 + 0.201a_{t-1}^2$. Performance of the model was evaluated using RMSE criterion. The RMSE value of the prediction is 20567.

Acknowledgments

We would like to thank to Rector of Diponegoro University, Chief of Research and Public Services Diponegoro University for their support to this research.

References

Modelling of Cayenne Production in Central Java Using ARIMAGARCH

ORIGINALITY REPORT

15%

SIMILARITY INDEX

7%

INTERNET SOURCES

12%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

- 1 Tarno Tarno, Suparti Suparti, Dwi Ispriyanti. "Modeling cayenne production data in Central Java using adaptive neuro fuzzy inference system (ANFIS)", Model Assisted Statistics and Applications, 2018 2%
Publication
- 2 Lee, Suri. "International Journal of Business and Management, Vol. 5, No. 2, February 2010, all in one PDF file", International Journal of Business and Management, 2010. 2%
Publication
- 3 Panik, . "Issues in Time Series Modeling and Estimation", Regression Modeling Methods Theory and Computation with SAS, 2009. 2%
Publication
- 4 Tarno, , Subanar, Dedi Rosadi, and Suhartono. "New procedure for determining order of subset autoregressive integrated moving average (ARIMA) based on over-fitting concept", 2012 International Conference on 1%

Statistics in Science Business and Engineering (ICSSBE), 2012.

Publication

| | | |
|----|--|------|
| 5 | www.ripublication.com Internet Source | 1 % |
| 6 | Zivot, Eric, and Jiahui Wang. "Univariate GARCH Modeling", Modeling Financial Time Series with S-Plus®, 2003. Publication | 1 % |
| 7 | www.civeng.adelaide.edu.au Internet Source | 1 % |
| 8 | Submitted to Higher Education Commission Pakistan Student Paper | 1 % |
| 9 | Ameet V Joshi. "Chapter 11 Time Series Models", Springer Science and Business Media LLC, 2020 Publication | <1 % |
| 10 | Olive R. Cawiding, Gina May R. Natividad, Crisostomo V. Bato, Rizavel C. Addawe. "Forecasting typhoid fever incidence in the Cordillera administrative region in the Philippines using seasonal ARIMA models", AIP Publishing, 2017 Publication | <1 % |
| 11 | Submitted to University of Durham Student Paper | <1 % |

12 "Econometrics for Financial Applications",
Springer Science and Business Media LLC,
2018
Publication <1 %

13 [dokumen.pub](#)
Internet Source <1 %

14 [maryclare.github.io](#)
Internet Source <1 %

15 [www.sel.eesc.sc.usp.br](#)
Internet Source <1 %

16 [folk.uio.no](#)
Internet Source <1 %

17 [hdl.handle.net](#)
Internet Source <1 %

18 [www.iorcf.unisg.ch](#)
Internet Source <1 %

19 [mpra.ub.uni-muenchen.de](#)
Internet Source <1 %

20 [onlinelibrary.wiley.com](#)
Internet Source <1 %

21 [shura.shu.ac.uk](#)
Internet Source <1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off

Modelling of Cayenne Production in Central Java Using ARIMAGARCH

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8
