8th International Seminar of Research Month 2023
Volume 2024
http://dx.doi.org/10.11594/nstp.2024.4110

NST Proceedings
OPEN ACCESS

Conference Paper

Modelling of Return of S&P 500 Using the Non Linear Generalized Autoregressive Conditional Heteroscedasticity (NGARCH) Model

Trimono Trimono*, Aviolla Terza Damaliana, Irma Amanda Putri

Data Science Department, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya 60294, Indonesia

*Corresponding author:
E-mail:
trimono.stat@upnajtim.ac.id

ABSTRACT

ARIMA Box-Jenkins is one of the most popular forecasting methods. ARIMA modeling requires a non-heteroskedastic care that shows constant residual variants. Awake, meaning residual residue from heteroscedastic ARIMA modeling (not constant). To overcome the problem of residual heteroskedasticity ARIMA used modeling volatility that is Generalized Autoregressive Conditional Heteroscedasticity (GARCH). GARCH is used to model the ARIMA residual variant which means symmetric. Some financial data has an asymmetric nature caused by the influence of good news and bad news. To accommodate these asymmetric properties, we use the Non-Linear Generalized Autoregressive Conditional Heteroscedasticity (NGARCH) volatility model which is the development of the GARCH model. This research applies NGARCH model using S & P 500 share price index data from January 1, 2019, until July 31, 2023 during active day (Monday-Friday). The purpose of this study, to find the best model NGARCH. The best model generated for S & P 500 stock price index data is ARIMA (1,0,1) NGARCH (1,1) because it has small AIC.

Keywords: Modelling, S&P 500, NGARCH

Introduction

Currently, the capital market is a market that is quite attractive for investors, especially in the financial sector. According to Anoraga and Pakarti (2001), an efficient market is a market that provides accurate, complete, relevant, and honest information. For investors, the speed and accuracy of information presented on securities developments is very necessary, especially to obtain profitable investment decisions. According to Darwanto (2011), investment is a positive net investment activity. In general, investment is categorized into two types, namely real investment and financial investment. Shares are one of the most traded securities on the capital market. Even now, with more and more issuers listing their shares on the stock exchange, trading is becoming more widespread and attracting investors to get involved in buying and selling shares (Anoraga & Pakarti, 2001).

Activities in the capital market are buying products (instruments) that are traded on the capital market with the hope of obtaining returns in the future (Sjahrial, 2009). Based on Campbell et al. (1997), by looking at the return value, investors can find out changes in the price of a share, the amount of profit or loss that will be received, so that it can be used as a guideline for deciding whether to invest in that share or not. In reality, stock prices often change at certain times and are difficult to predict. In connection with the uncertainty of changes in stock prices, a mathematical model is needed to predict future stock prices based on pre-existing stock prices.

Based on Brigo et al. (2007), the NGARCH model is a model that can be used to predict stock values based on historical data. In particular, this model predicts asymmetric volatility behavior or stock data

How to cite:
Trimono, T., Damaliana, A. T., & Putri, I. A. (2024). Modelling of return of S&P 500 using the non linear Generalized Autoregressive Conditional Heteroscedasticity (NGARCH) model. *8th International Seminar of Research Month 2023*. NST Proceedings. pages 45-51. doi: 10.11594/ nstp.2024.4110

that is too skewed at one point. Data under the influence of "good news" or positive returns will produce low volatility, while "bad news" or negative returns will produce high volatility. By adding a positive parameter, the influence on "good news" will be reduced and the influence on "bad news" will increase (Komang, 2015).

The stock price index is an indicator that shows stock movements within a period. The S&P 500 is an index consisting of 500 large-cap companies, mostly from the United States. This index is a fairly well-known index owned by Standard & Poor's, a division of McGraw-Hill. Based on this description, the author took the topic of "Modelling the return of the S&P 500 stock price index using Non-Linear Generalized Autoregressive Conditional Heteroscedasticity (NGARCH)". The NGARCH modeling uses S&P 500 stock price index return data for the period April 1 2019 to March 31 2023.

Material and Methods

Time series analysis

Time series analysis was introduced in 1976 by George E. P. Box and Gwilym M. Jenkins in Makridakis et al. (1999). The basic idea of time series is that the current observation (Zt) depends on one or several previous observations (Zt-k). In other words, the time series model was created because statistically there is a correlation (dependency) between the observation series. The goals of time series analysis include understanding and explaining certain mechanisms, predicting future values, and optimizing control systems. In the time series method, there are several things that must be considered, namely data stationarity, autocorrelation function and partial autocorrelation function.

Stasioneritas

An observation series is said to be stationary if the process does not change along with changes in the time series. For example, Z1, Z2, … , Zt are stochastic processes for discrete time series. The process is called stationary if the mean and variance are constant for every point t and the covariance is constant for every time interval k, so it can be written:

$E(Z_t) = \mu$ for every *t* (1)

$Var(Z_t) = \sigma^2 = \gamma_0$ for every *t* (2)

$Cov(Z_t, Z_{t-k}) = \gamma_k$ for every *k,* where γ_k is autocovariance value at *k*-th lag. (3)

ARIMA model

The ARIMA (p,d,q) process is a non-stationary time series model. The general form of this model in Soejoeti (1987) is:

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B) a_t \quad (4)$$

where $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ are stationary AR operators and $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ are invertible MA operators. ARIMA model estimation can use FAK plots and FAKP plots, where the time series must have a stationary pattern (Soejoeti, 1987).

Subset ARIMA model

The ARIMA subset model is part of the generalized ARIMA model, so it cannot be expressed in general form. This ARIMA subset model is a subset of the ARIMA model. For example the ARIMA subset([1,5],0,[1,12]) can be written as:

$$(1 - \emptyset_1 B - \emptyset_5 B^5) Z_t = (1 - \theta_1 B - \theta_{12} B^{12}) a_t \quad (5)$$

Thus, the ARIMA subset model is an ARIMA model with several parameters equal to zero.

ARCH (Autoregressive Conditional Heteroscedasticity) model

The first model used to model residual data volatility was the ARCH model introduced by Engle and Ng (1993). The ARCH model assumes that the residual variance at one point in time is a function of the residual at another point in time. The order often used in ARCH models is denoted by *p*. According to Tsay (2002), the general form of the ARCH(p) model:

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_p a_{t-p}^2 \qquad (6)$$

Because it uses the concept of variance, the value must always be positive (non-negativity constraints), so that each squared residual parameter must be equal to or greater than zero or $\alpha_0 > 0$ and $\alpha_i \geq 0$ for each i = 1, ..., p. In the ARCH(p) model, the process a_t can be defined by:

$a_t : \sigma_t \varepsilon_t$
a_t: t-th residual value obtained from the ARIMA model
ε_t: t-th residual value from the ARCH/GARCH model

GARCH (Generalized autoregressive conditional heteroscedasticity) model

Bollerslev (1986) developed the ARCH model into a more general model known as Generalized Autoregressive Conditional Heteroscedasticity (GARCH). This model is used to overcome orders that are too large in the ARCH model. In the GARCH model, the conditional variance is not only influenced by past residuals but by the lag of the conditional variance itself (Ariefianto & Doddy, 2012). Thus, the conditional variance in the GARCH model consists of two components, namely the past component of the squared residual (denoted by degree q) and the past component of the conditional variance (denoted by degree of *p*). Mathematically the GARCH(p,q) model can be created in the following form:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \qquad (7)$$

In the GARCH(p,q) model, the process a_t can be defined by
$a_t : \varepsilon_t \sigma_t$
σ_t : square roots of σ_t^2 and $\varepsilon_t \sim Niid$ (0,1)

***Non Linier Generalized Autoregressive Conditional Heteroscedasticity* (NGARCH) Model**

The NGARCH model is a development of the GARCH model, which contains a parameter (γ) or asymmetric parameter which is an adjustment for stock returns. In general, the NGARCH model process is defined as:

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha (a_{t-1} + \gamma \sqrt{\sigma_{t-1}})^2 \qquad (8)$$

The parameters in NGARCH are ω which is a constant parameter, parameter α is an ARCH parameter, β is a GARCH parameter and γ is an asymmetric parameter. The NGARCH parameters consist of four parameters, namely ω, α, β, γ which are positive numbers with $\alpha + \beta(1 + \gamma^2) < 1$ (Komang, 2015).

Selection of the best model

The AIC (Akaike's Information Criterion) value can be used to determine the selection of the best model. The best model is the model that has the minimum AIC value. The formula for obtaining the AIC value is written as follows (Rosadi, 2012):

$$AIC = n\log\left(\frac{SSR}{n}\right) + 2k \quad (9)$$

where,

n : sample size k : number of parameters in the model

SSR : $\sum_{i=1}^{n} \varepsilon_i^2$

Results and Discussion

Data description

The data used in this research is the closing data of the S&P 500 stock price index for the period 01 January 2019 until 31 July 2023 during active days (Monday to Friday). The S&P 500 stock price index data is secondary data taken from www.finance.yahoo.com (Appendix 1). The S&P 500 stock price index return data used in this analysis amounts to 1009 data. The characteristics of the S&P 500 return data can be seen based on the following Table 1.

Table 1. Descriptive statistics of S&P 500 return value

Statistics	Value
Number of Observations	1009
Mean	0.0004101
Median	0.0003969
Maximum	0.0382913
Minimum	-0.0402114

The average or mean value of 1009 S&P 500 return data is 0.0004101. The maximum value of 0.0382913 means that the largest profit value of the S&P 500 during the period 01 January 2019 until 31 July 2023 was 0.0382913. Meanwhile, the minimum return value for the S&P 500 stock price index is -0.0402114, meaning that the maximum loss resulting from the S&P 500 is 0.0402114.

ARIMA model identification

Determining the AR order and MA order in model identification uses the autocorrelation function (FAK) and partial autocorrelation function (FAKP) plots which can be seen in the following figure 1.

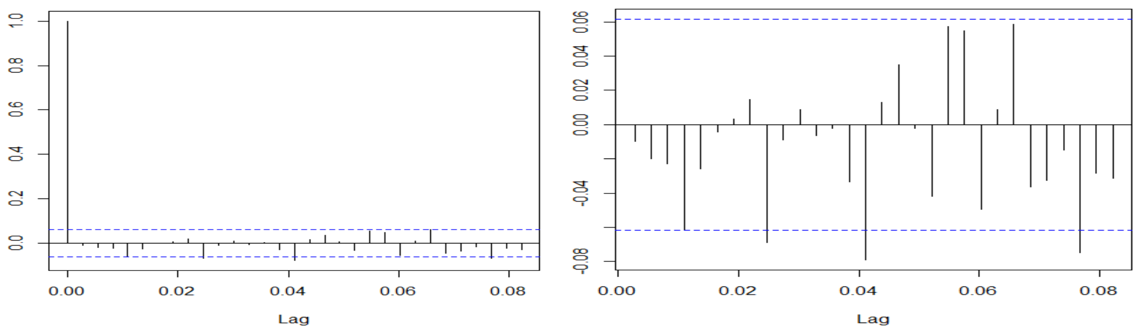

Figure 1. Plots function of: (a) ACF and (b) PACF

Based on the results of hypothesis testing at a significance level of 5%, of the 30 ARIMA models, both ARIMA and ARIMA subsets, only 8 models have significant parameters. The ARIMA models will be used for further analysis, namely the verification stage. The ARIMA model can be seen in the following Table 2:

Table 2. Estimation and significance testing of ARIMA parameters

Model	Parameter	Estimated Value	|t-hit|	$t_{(0,025;n-p)}$	Keputusan
Subset ARIMA ([9],0,0)	ϕ_9	-0,0632	-2,0127	1,962320	Rejected H_0
Subset ARIMA ([15],0,0)	ϕ_{15}	-0,0754	-2,3861	1,962320	Rejected H_0
Subset ARIMA (0,0,[9])	θ_9	-0,0623	-1,9968	1,962320	Rejected H_0
Subset ARIMA (0,0,[15])	θ_{15}	-0,0799	-2,4665	1,962320	Rejected H_0
Subset ARIMA ([9],0,[9,15])	ϕ_9	-0,4693	-2,3547	1,962325	Rejected H_0
	θ_9	0,4171	2,0446	1,962325	Rejected H_0
	θ_{15}	-0,0862	-2,8355	1,962325	Rejected H_0
Subset ARIMA ([9,15],0,0)	ϕ_9	-0,0631	-2,0160	1,962323	Rejected H_0
	ϕ_{15}	-0,0753	-2,3905	1,962323	Rejected H_0
ARIMA (1,0,1)	ϕ_1	-0,9147	-4,5417	1,962323	Rejected H_0
	θ_1	0,9118	-4,5363	1,962323	Rejected H_0
ARIMA (0,0,0)	-	-	-	-	-

Table 3. Residual independence test using the Ljung Box Test

Model	Conditions	Decision
subset ARIMA ([9],0,0)	*p-value* > α	Accepted H_0
subset ARIMA ([15],0,0)	*p-value* > α	Accepted H_0
subset ARIMA (0,0,[9])	*p-value* > α	Accepted H_0
subset ARIMA (0,0,[15])	*p-value* > α	Accepted H_0
subset ARIMA ([9],0,[9,15])	*p-value* > α	Accepted H_0
subset ARIMA ([9,15],0,0)	*p-value* > α	Accepted H_0
ARIMA (1,0,1)	*p-value* > α	Accepted H_0
ARIMA (0,0,0)	*p-value* > α	Accepted H_0

Table 4. The result of Jarque Bera test

Model	Jarque Bera	Prob	Decision
Subset ARIMA ([9],0,0)	274,93	0,0000	Rejected H_0
Subset ARIMA ([15],0,0)	276,72	0,0000	Rejected H_0
Subset ARIMA (0,0,[9])	274,58	0,0000	Rejected H_0
Subset ARIMA (0,0,[15])	276,11	0,0000	Rejected H_0
Subset ARIMA ([9],0,[9,15])	291,24	0,0000	Rejected H_0
Subset ARIMA ([9,15],0,0)	290,71	0,0000	Rejected H_0
ARIMA (1,0,1)	262,92	0,0000	Rejected H_0
ARIMA (0,0,0)	260,92	0,0000	Rejected H_0

Table 5. The result of *Lagrange Multiplier* test

Model	LM	Prob	Keputusan
subset ARIMA ([9],0,0)	163,4	0,0000	Rejected H_0
subset ARIMA ([15],0,0)	163,01	0,0000	Rejected H_0
subset ARIMA (0,0,[9])	163,34	0,0000	Rejected H_0
subset ARIMA (0,0,[15])	162,93	0,0000	Rejected H_0
subset ARIMA ([9],0,[9,15])	163,56	0,0000	Rejected H_0
subset ARIMA ([9,15],0,0)	163,15	0,0000	Rejected H_0
ARIMA (1,0,1)	163,96	0,0000	Rejected H_0
ARIMA (0,0,0)	163,04	0,0000	Rejected H_0

Based on Table 4, it can be concluded that all ARIMA model residuals are not normally distributed. Based on the decisions in Table 5, it can be concluded that there is a heteroscedasticity effect on the residuals of each ARIMA model. Once it is known that each ARIMA model has a heteroscedasticity effect on the residuals, it will be modeled using ARCH/GARCH modeling. Then, only the ARIMA (1,0,1) GARCH (1,1) and ARIMA (0,0,0) GARCH (1,1) models have significant parameters. Therefore, the ARIMA (1,0,1) GARCH (1,1) and ARIMA (0,0,0) GARCH (1,1) models will be used for further analysis.

Sign bias test

In this section, we will test the existence of asymmetric effects in the residuals of the GARCH model using the sign bias test.

Table 6. *Sign bias test result*

Model	LM	Probability	Decision
ARIMA (1,0,1) GARCH (1,1)	15,40269	0,00150	Rejected H_0
ARIMA (0,0,0) GARCH (1,1)	11,60472	0,00886	Rejected H_0

Based on the test criteria, the probability value is smaller than α, so it can be concluded that there is an asymmetric effect on the two residuals of the GARCH model so that it can be modeled using the NGARCH model.

NGARCH model

Based on the conclusions obtained from the sign bias test, an NGARCH model can be formed from the ARIMA (1,0,1) and ARIMA (0,0,0) models.

The results of parameter significance testing in Table 10 show that the ARIMA (1,0,1) NGARCH (1,1) and ARIMA (0,0,0) NGARCH (1,1) models have significant parameters and meet the white noise assumption. The selection of the best model in this research uses the AIC value as a comparison, then the model with the smallest AIC is selected as the best model.

Table 7. Model evaluation test

Model	Parameters	Estimated value	t-hit	Prob	Decision
ARIMA (1,0,1) NGARCH (1,1)	ϕ_1	-0,883259	-12,675	0,0000	Rejected H_0
	θ_1	0,848229	10,866	0,0000	Rejected H_0
	ω	0,000003	38,216	0,0000	Rejected H_0
	α_1	0,057292	11,856	0,0000	Rejected H_0
	β_1	0,520205	23,250	0,0000	Rejected H_0
	γ	2,646217	43,814	0,0000	Rejected H_0
ARIMA (0,0,0) NGARCH (1,1)	ω	0,000003	21.952	0,0000	Rejected H_0
	α_1	0,055622	10.647	0,0000	Rejected H_0
	β_1	0,508818	21.713	0,0000	Rejected H_0
	γ	2,730551	37.495	0,0000	Rejected H_0

The results of parameter significance testing in Table 10 show that the ARIMA (1,0,1) NGARCH (1,1) and ARIMA (0,0,0) NGARCH (1,1) models have significant parameters and meet the white noise assumption. The selection of the best model in this research uses the AIC value as a comparison, then the model with the smallest AIC is selected as the best model.

Table 8. AIC value

ARIMA models	AIC value
ARIMA (1,0,1) NGARCH (1,1)	-7.0761
ARIMA (0,0,0) NGARCH (1,1)	-7,0748

Based on Table 8, it can be concluded that ARIMA (1,0,1) NGARCH (1,1) is the best model because it has the smallest AIC value compared to ARIMA (0,0,0) GARCH (1,1). So the resulting model is:

$$Z_t = -0{,}883259\, Z_{t-1} + a_t - \; 0{,}848229\, a_{t-1}$$
$$\sigma_t^2 = 0{,}000003 + 0{,}520205\sigma_{t-1}^2 + 0{,}057292\, (a_{t-1} + 2{,}646217 \sqrt{\sigma_{t-1}^2})^2$$

Conclusion

The conclusions obtained from this research are:

1. Modeling S&P 500 return data using the ARIMA Box-Jenkins model has a residual variance that changes over time (heteroscedasticity) so ARCH/GARCH volatility modeling is carried out.
2. GARCH modeling was carried out on the 8 best ARIMA models and 2 selected GARCH models were obtained, namely ARIMA(1,0,1) GARCH (1,1) and ARIMA (0,0,0) GARCH (1,1).
3. Based on the sign bias test, it can be concluded that there is an asymmetric effect in the ARIMA(1,0,1) GARCH (1,1) and ARIMA (0,0,0) GARCH (1,1) models so that they can be modeled with NGARCH.
4. The best S&P 500 return model produced is ARIMA(1,0,1) NGARCH (1,1) because it has the smallest AIC value, with the equation:

$$Z_t = -0{,}883259\, Z_{t-1} + a_t - \; 0{,}848229\, a_{t-1}$$
$$\sigma_t^2 = 0{,}000003 + 0{,}520205\sigma_{t-1}^2 + 0{,}057292\, (a_{t-1} + 2{,}646217 \sqrt{\sigma_{t-1}^2})^2$$

Acknowledgment

This work was financially supported by Research Center for Biomaterials through “DIPA 2017”. Therefore, we are grateful for this funding and support of this research.

References

Anoraga, P., & Pakarti, P. (2001). *Pengantar pasar modal*. Jakarta: Rineka Cipta.
Ariefianto & Doddy, M. (2012). *Ekonometrika esensi dan aplikasi dengan menggunakan e-views*. Jakarta: PT. Gelora Aksara Pratama.
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307-327
Brigo et al. (2008). A stochastic processes toolkit for risk management. *Journal of Risk Management in Financial Institutions, 1*(4), 5-13.
Campbell, J. Y., Lo, A., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press, Princeton, N. J.
Darwanto. (2011). Membangun daya saing UKM dalam Perekonomian Nasional. *Jurnal Ilmu Manajemen dan Akuntasi Terapan, 2*(1), 21-30.
Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance, 48*(5), 1749-78.
Komang, M. (2015). Model non linier GARCH untuk mengestimasi nilai value at risk pada IHSG. *E-Jurnal Matematika. 4*(2), 59-66.
Sjahrial, D. (2009). *Manajemen keuangan lanjutan*. Jakarta: Mitra Wacana Media.