www.bircu-journal.com/index.php/birci
Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Humanities and Social Sciences
ISSN 2615-3076 Online)
ISSN 2615-1715 (Print)

Garch Modeling for Volatil Data: Air Pollution Standard Index in DKI Jakarta as A Parameter in Secure Smart City Development

Nurul Qomariasih[1], Herman Kabetta[2], I Komang Setia Buana[3], Yogha Restu Pramadi[4]
[1,2,3,4]*Politeknik Siber dan Sandi Negara (Poltek SSN), Bogor, Indonesia*
nurul.qomariasih@poltekssn.ac.id, herman.kabetta@poltekssn.ac.id, komang.setia@poltekssn.ac.id, yogha.restu@poltekssn.ac.id

Abstract

The DKI Jakarta government has the Jakarta Kini Application (JAKI) which was developed by the Jakarta smart city management unit. There is a JakISPU feature that displays the air pollution standard index (ISPU) in real-time, but the ISPU forecast value is not presented to prevent the ISPU from worsening in the future. This study presents a model for the concentration value of the ISPU parameter which has high volatility. The model obtained is expected to be used for forecasting the ISPU concentration value in the next few days. The resulting model is the SO2 arch garch model, namely $\sigma_t^2 = 3{,}235114 + 0{,}325440 e_{t-1}^2 + 0{,}611209 \sigma_{t-1}^2$ *and arch garch. Model NO2 that is* $\sigma_t^2 = 7{,}142879 + 0{,}122889 e_{t-1}^2 + 0{,}809007 \sigma_{t-1}^2$*. The model formed is free from the assumption of heteroscedasticity.*

Keywords
Garch modeling; violating data; air pollution

I. Introduction

The DKI Jakarta government currently has an application that is a center for information and services for its citizens. The application is called the Jakarta Kini Application (JAKI) which was developed through the Jakarta Smart City Management Unit. One of the features in the application is JakISPU, which is a facility for residents who are concerned about air quality in Jakarta. JakISPU displays the Air Pollution Standard Index (ISPU) which is integrated with sensors belonging to the Environment Agency in real-time. The sensors belonging to DLH are placed in all municipalities, so that every citizen can see the ISPU according to its location. Unfortunately, this application feature only displays the ISPU in real-time, and does not display how the ISPU forecast in the future. ISPU forecasting in the future can be very useful as a warning to citizens. Residents of DKI Jakarta and the city government can determine steps so that ISPU does not worsen in the future.

Development is a systematic and continuous effort made to realize something that is aspired. Development is a change towards improvement. Changes towards improvement require the mobilization of all human resources and reason to realize what is aspired. In addition, development is also very dependent on the availability of natural resource wealth. The availability of natural resources is one of the keys to economic growth in an area. (Shah, M. et al. 2020)

The purpose of this study is to model the concentration value of the ISPU parameter so that it can be forecasted the value of the ISPU parameter concentration in the next few days or months. The main focus of this research is modeling for parameters that have high volatility. Based on observations of the ISPU parameter concentration values from January 1

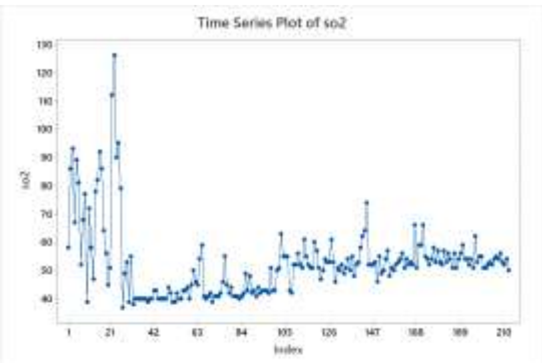
Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Volume 4, No. 4, November 2021, Page: 13786-13795
e-ISSN: 2615-3076 (Online), p-ISSN: 2615-1715 (Print)
www.bircu-journal.com/index.php/birci
email: birci.journal@gmail.com

to July 31, 2021, there are two parameters whose values are highly volatile, namely SO2 and NO2. Nitrogen dioxide or abbreviated NO2 is one of the common air pollutant toxins that is often found as a mixture of nitrogen oxides (NOx) (Ashfaque et al, 2019), reddish brown in color and has serious effects on humans, such as inflammation of the respiratory tract, decreased lung function. , and increased response to allergens (Bauer et al, 1986) (Ehrlich et al, 1966). While Sulfur dioxide (SO2) is found mostly as a mixture of sulfur oxides (SOx) with a sharp and very unpleasant smell (Ashfaque et al, 2019), it can cause damage to the eyes, lungs and throat (Khan et al, 2014) (Nisar et al., 2014). al, 2013). In 1982, Engle proposed an Autoregressive Conditional Heteroscedasticity (ARCH) model for data that has high volatility, where the conditional variance is obtained from the current error as a function of the previous time. Meanwhile, GARCH is a generalize of the ARCH model which was developed by Bolerslev in 1986 as a more flexible model with respect to the number of parameters. GARCH has been widely carried out as a modeling and forecasting analysis as in 2017 by Chu Jeffrey et al for modeling Cryptocurrencies, in 2018 by Kim et al for forecasting stock price indexes on the money market, and in 2019 by Wang et al for forecasting stock price indexes.

II. Research Methods

2.1. Research Methods

The dataset needed in the study was obtained from the Open Data Jakarta website page https://data.jakarta.go.id/dataset/index-standard-air-pollution-ispu-year-2021. The dataset consists of data on the date of air quality measurement, measurement location at the station, particulates of one of the parameters measured, sulfides in the form of SO2, carbon monoxide, ozone, nitrogen dioxide, as well as the results of the calculation of the air pollution standard index in the period January 1, 2021 to July 31 2021 in DKI Jakarta, which is as much as 212 observation data. The data used for garch modeling is Sulfide data in the form of SO2 in DKI Jakarta per day in that time span. The SO2 variable has very high fluctuating data. The following is a time series graph for all parameters of the air pollution standard index:



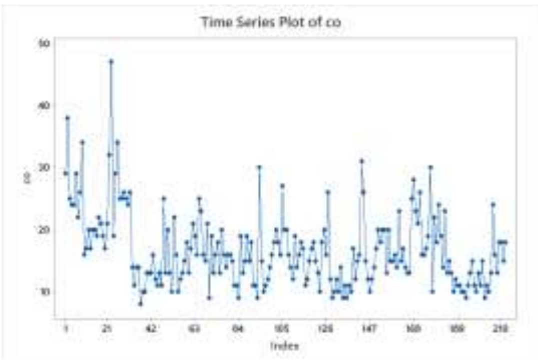
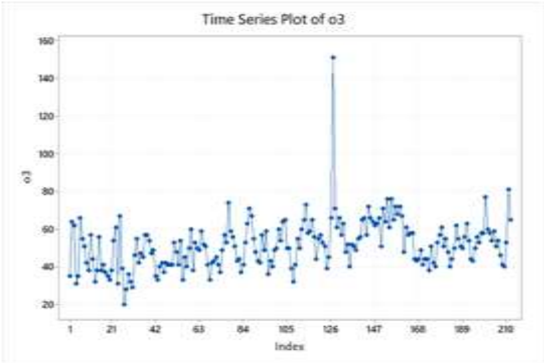


Figure 1. *SO2 Sulfide Time Series Graph Figure 2. Carbon Monoxide Time Series Graph*



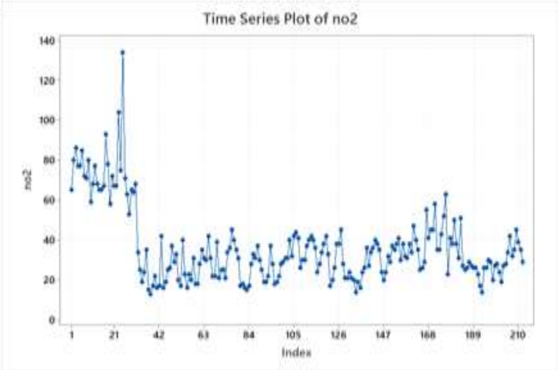


Figure 2. *Ozone Time Series Graph Figure 4. Nitrogen Dioxide Time Series Graph*

Based on Figures 1 to 4, among the four parameters, the SO2 and NO2 parameters have almost the same trend. The value is very high at the beginning of 2021 and decreases drastically around days 21-40. Meanwhile, the CO time series graph tends to show a seasonal trend but is not volatile and the O3 graph shows that there is an outlier that needs to be investigated further, whether the outlier occurs due to natural events or human error. So that the most suitable for Garch modeling are SO2 and NO2 data, because they are the most volatile compared to other data. Next, we will examine descriptive statistics for the values of these four parameters:

Table 1. Descriptive statistics of ISPU parameters

Descriptive statistics	***so2***	***co***	***o3***	***no2***
mean	52.52	16.83	51.59	36,48
Standard Error	0.88	0.42	0.9	1.29
median	52	16	51	31
Sample Variance	162.86	36.93	171.39	352.16
Range	89	39	131	121
Minimum	37	8	20	13
Maximum	126	47	151	134
Sum	11134	3567	10938	7734
Count	212	212	212	212

Based on table 1, it is shown that the largest variance is for the SO2 and NO2 parameters. The O3 range looks great but is caused by an outlier. So the next time series modeling using NO2 and SO2 data.

2.2. Related Research

Table 2. Several similar studies conducted to model and predict time series data using Garch:

Year	**Title**	**Substance**
2017	*GARCH Modeling of Cryptocurrencies*	Modeling 12 types of GARCH for 7 different cryptocurrencies
2018	*Regime changes in Bitcoin GARCH volatility dynamics*	Comparison of GARCH and MSGARCH
2018	*Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models*	Propose a new method for combining time series and neural network models

2.3. Air Pollution Standard Index (ISPU)

In the regulation of the minister of environment and forestry (Minister of Environment, 1997), ISPU is a number that does not have a unit that describes the condition of ambient air quality in a particular location, which is based on the impact on human health, aesthetic value and other living things. ISPU has several parameters, namely:

a. particulate (PM10)
b. particulate (PM2.5)
c. carbon monoxide (CO)
d. nitrogen dioxide (NO2)
e. sulfur dioxide (SO2)
f. ozone (O3)
g. hydrocarbon(HC)

The dataset provided from the Open Data Jakarta website is ISPU data from calculations, reports, and publications. The calculated parameter concentration values (I) (Apriawati et al, 2017) were converted using Table 3 and reported in the form of categories in Table 4.

$$I = \frac{I_a - I_b}{(X_a - X_b)}(X_x - X_b) + Ib$$

Where,
I = calculated ISPU
He= upper limit of ISPU
Ib= lower limit of ISPU
Xa = upper limit ambient level
Xb= lower limit ambient level
Xx= real ambient level measurement results

The results of the above calculations are then compared with the data in Table 3 contained in KEP-45/MENLH/10/1997 (State Minister for the Environment, 1997).

Table 3. Conversion Table of ISPU parameter Concentration Value

ISP	24 hours particulate (PM10) g/m3	24 hours particulate (PM2.5) g/m3	24 hours sulfur dioxide (SO2) g/m3	24 hours carbon monoxide (CO) g/m3	24 hours ozone (O3) g/m3	24 hours nitrogen dioxide (NO2) g/m3	24 hours hydrocarbon (HC) g/m3
0-50	50	15.5	52	4000	120	80	45
51-100	150	55.4	180	8000	235	200	100
101-200	350	150.4	400	15000	400	1140	215
201-300	420	250.4	800	30000	800	2260	432
>300	500	500	1200	45000	1000	3000	648

Table 4. Category of ISPU Range Score

Category	Color Status	Range Number
good	Green	Jan-50
Moderate	Blue	51 - 100
not healthy	Yellow	101 - 200
Very unhealthy	Red	201 - 300
Dangerous	Black	301

2.4. Unit Root Test

To determine the right type of analysis, in the time series data there is a stationarity test on the data. There are many stationarity tests, but the one used in this study is the Augmented Dickey Fuller Test (ADF) developed by statisticians David Dickey and Wayne Fuller (Dickey & Fuller, 1979). This test has Hypothesis 0 which indicates the unit root is in an autoregressive model or the data is not stationary and the alternative hypothesis is that the data is stationary. The ADF equation is written as follows:

$$\Delta Y_t = \alpha + \alpha_2 t + \delta Y_{t-1} + \gamma_i \Sigma_{i=1}^{m} \Delta Y_{(t-i)} + \epsilon_t$$

If the ADF result > McKinnon critical value, then H0 is accepted (Hanke & Wichern, 2005

2.5. Heteroscedasticity & Arch-Lm Test

In addition to testing the stationarity of the data, the variance in a regression model sometimes violates the assumption of homogeneity of variance, so it is necessary to test the homogeneity of variance on the time series data using the ARCH-LM test. Research conducted by Enders in 1995 stated that the variance of the residuals does not only come from the function of the independent variables but depends on the squared remainder of the previous period. The hypothesis is:

$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$ (Tidak terdapat efek ARCH)

$H_1: \exists \alpha_i \neq 0, i = 1,2,\ldots,p$ (Terdapat efek ARCH)

With the equation (Tsay, 2005):

$$\epsilon_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2 + \omega_t \; ; \;\; t = m+1, \ldots T$$

$\omega_t = error$

$m = integer$

$T = sampling\ number$

Hypothesis 0 will be rejected when: $F > X_p^2(\alpha)$

2.6. GARCH

According to Enders in 1995, GARCH is a model that was built to avoid orders that are too high in the ARCH model so that it can guarantee the variance is always positive. For example Xt . is a time series data, then

$$X_t = \mu_t + \sigma_t Z_t$$

Where,

μ_t : conditional means

σ_t : volatility process

This equation is included in the regression analysis with time series elements. In regression analysis, it is assumed that the variance is constant. Violation of assumptions causes the estimates to be biased and the estimates are invalid. Variation is not constant usually due to the high volatility between the data. In a study conducted in 1986, Bollerslev stated that the generalized model he had proposed named GARCH was a solution that could accommodate this problem. leftover models$\epsilon_t = X_t - \mu_t$ is said to follow the GARCH (p,q) model if

$$\sigma_t^2 = \sigma_0 + \sigma_1 \epsilon_{t-1}^2 + \sigma_2 \epsilon_{t-2}^2 + \cdots + \sigma_q \epsilon_{t-q}^2 + \gamma_1 \sigma_{t-1}^2 + \cdots + \gamma_p \sigma_{t-p}^2$$
$$= \sigma_0 + \Sigma_{i-1}^{q} \sigma_i \epsilon_{t-i}^2 + \Sigma_{j=1}^{p} \gamma_j \sigma_{t-j}^2$$

III. Results and Discussion

In the time series parametric analysis, there are assumptions of stationarity that must be met, namely stationary in the mean and variance. Stationary in the mean can be seen from the correlogram and the ADF method (Teguh Santoso, 2011). Both SO2 and NO2 data were not stationary in the mean, and became stationary when differencing was performed once. One-time differencing treatment is enough to change the p-value to less than =5%. A time series data is said to be stationary in the mean if the p-value of the Augmented Dickey-Fuller test < and TREND > . The p-value of NO2 and SO2 after one differencing is done is 0.0000 while the p-value of TREND is 0.6077 and 0.3419, respectively. The test results are shown in Table 5.

Table 5. Stationarity Test Results

SO2

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-12.40210	0.0000
Test critical values:	1% level	-4.003226	
	5% level	-3.431789	
	10% level	-3.139601	

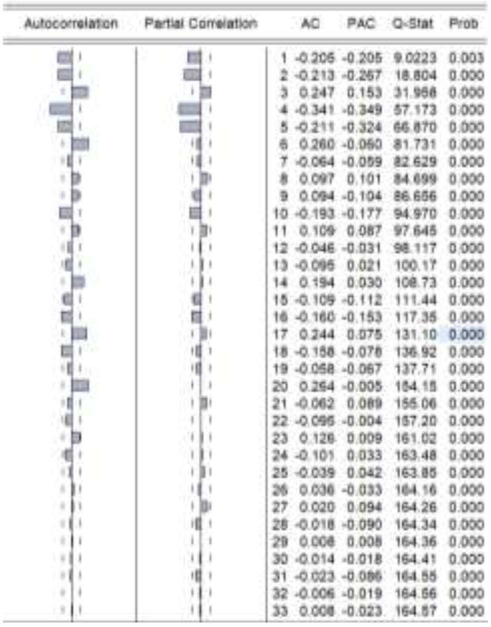
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(SO2(-1))	-2.358208	0.190146	-12.40210	0.0000
D(SO2(-1),2)	1.075962	0.158488	6.788916	0.0000
D(SO2(-2),2)	0.807921	0.135609	5.957711	0.0000
D(SO2(-3),2)	0.754107	0.096614	7.805387	0.0000
D(SO2(-4),2)	0.338354	0.063881	5.296671	0.0000
C	-1.252891	1.147202	-1.092128	0.2761
@TREND("1/01/2021")	0.008824	0.009263	0.952598	0.3419

NO2

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-21.68783	0.0000
Test critical values:	1% level	-4.002354	
	5% level	-3.431368	
	10% level	-3.139353	

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(NO2(-1))	-1.385270	0.063873	-21.68783	0.0000
C	-0.977771	1.521815	-0.642504	0.5213
@TREND("1/01/2021")	0.006384	0.012418	0.514095	0.6077

Descriptively, which means that decisions are made subjectively, data stationarity can be observed from the Autocorrelation and Partial Autocorrelation graphs. If the pattern is irregular, then the data is said to be stationary in the mean. Based on Figure 5 and Figure 6, the pattern is irregular, so the SO2 and NO2 time series data are stationary in the mean.



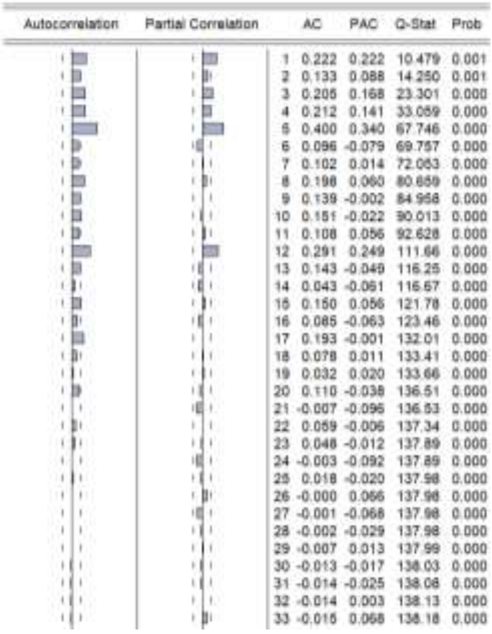


Figure 5. *Graph of ACF & PACF SO2 Figure 6. Graph of ACF & PACF SO2*

If the value of d for differencing is known, then the next step is to find the p value for AR and q for MA. For SO2 parameters, the significant ARIMA(p,d,q) model (Prob. <) is ARIMA(1,1,1,) (Table 6), while for the NO2 parameter the ARIMA(p,d,q) model is significant is ARIMA(0,1,1,) (Table 7).

Table 6. ARIMA SO2 . Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.086979	0.124990	-0.695889	0.4873
AR(1)	0.517892	0.078562	6.592109	0.0000
MA(1)	-0.906841	0.034808	-26.05239	0.0000

Table 7. ARIMA NO2 . Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.202612	0.331300	-0.611567	0.5415
MA(1)	-0.549413	0.057529	-9.550166	0.0000

If the time series data is stationary in the mean and variance, then the modeling is sufficient until the ARIMA model. However, because the SO2 and NO2 (residual) data were not stationary in variance, the modeling was carried out using GARCH analysis. The results of the stationary test of variance or commonly called heteroscedasticity on SO2 and NO2 are shown in Table 8 and Table 9. With Hypothesis 0: homogeneous data; and Hypothesis 1: heterogeneous data; and Prob.< , then the SO2 and NO2 time series data violate the law of homogeneity of variance/stationary variance.

Table 8. ARIMA SO2 . Model

Heteroskedasticity Test: ARCH

F-statistic	10.71399	Prob. F(1,207)	0.0012
Obs*R-squared	10.28516	Prob. Chi-Square(1)	0.0013

Table 9. ARIMA NO2 . Model

Heteroskedasticity Test: ARCH

F-statistic	15.44872	Prob. F(1,208)	0.0001
Obs*R-squared	14.51891	Prob. Chi-Square(1)	0.0001

The next modeling uses ARCH GARCH analysis and the following results are obtained:

Table 10. Model GARCH SO2

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	3.235114	0.963211	3.358677	0.0008
RESID(-1)^2	0.325440	0.102308	3.180987	0.0015
GARCH(-1)	0.611209	0.081345	7.513802	0.0000
R-squared	0.079423	Mean dependent var	-0.171429	
Adjusted R-squared	0.070528	S.D. dependent var	9.718484	
S.E. of regression	9.369506	Akaike info criterion	6.376184	
Sum squared resid	18172.04	Schwarz criterion	6.471816	
Log likelihood	-663.4993	Hannan-Quinn criter.	6.414844	
Durbin-Watson stat	1.516208			

Table 11. Model GARCH NO2

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	7.142879	4.210360	1.696501	0.0898
RESID(-1)^2	0.122889	0.037146	3.308239	0.0009
GARCH(-1)	0.809007	0.065105	12.42620	0.0000
R-squared	0.189231	Mean dependent var	-0.170616	
Adjusted R-squared	0.185352	S.D. dependent var	11.79222	
S.E. of regression	10.64341	Akaike info criterion	7.423120	
Sum squared resid	23675.95	Schwarz criterion	7.502547	
Log likelihood	-778.1391	Hannan-Quinn criter.	7.455226	
Durbin-Watson stat	1.933280			

Based on Table 10, the probability value of ARCH and GARCH is less than = 5%, so the correct model for the SO2 time series data is the ARCH GARCH model with the variance equation:

$$\sigma_t^2 = 3{,}235114 + 0{,}325440 e_{t-1}^2 + 0{,}611209 \sigma_{t-1}^2$$

While Table 11 shows the same thing for the NO2 parameter, namely the suitable model is ARCH GARCH with the variance equation as follows:

$$\sigma_t^2 = 7{,}142879 + 0{,}122889 e_{t-1}^2 + 0{,}809007 \sigma_{t-1}^2$$

Then, to test that the GARCH model formed is free from the heteroscedasticity problem, it is to do a test using the ARCH-LM test procedure as has been done previously.

Table 12. GARCH LM SO2 . test

Heteroskedasticity Test: ARCH

F-statistic	0.288286	Prob. F(1,207)	0.5919
Obs*R-squared	0.290667	Prob. Chi-Square(1)	0.5898

Table 13. GARCH LM NO2 . test

Heteroskedasticity Test: ARCH

F-statistic	1.653905	Prob. F(14,182)	0.0689
Obs*R-squared	22.23430	Prob. Chi-Square(14)	0.0739

Table 12 and Table 13 are proof that the estimation results for GARCH SO2 and NO2 are free from heteroscedasticity. This is indicated by the value of Prob. FSO2 and Prob. F NO2 is greater than =5% where the probability is not statistically significant. This causes H0 to be rejected and the GARCH model no longer contains elements of heteroscedasticity.

IV. Conclusion

Based on the results of research on several parameters of the Air Pollution Standard Index in DKI Jakarta, the measurement data looks very volatile and very high volatility, especially on the SO2 and NO2 parameters. Although the differencing method has been carried out and resulted in an ARIMA model with the order of d=1, there is a regression assumption that is still violated, namely the variance of the remainder is not homogeneous. So that the ARCH GARCH analysis was carried out on the data and a significant probability value was generated for both ARCH and GARCH. This means that the conditional variance residual is not only influenced by the square of the residual, but also by the conditional variance of the previous year. So the model is more appropriate to use ARCH GARCH.

References

Apriawati, Eka, and Abadi Agung Kiswandono.(2017). Kajian IndeksStandar Polusi Udara (ISPU) Nitrogen Dioksida (NO2) di Tiga Lokasi Kota Bandar Lampung. Analit, E-ISSN 2540-8267.

Ardia, David., Bluteau, Keven., Ruede, Maxime. (2018). Regime changes in Bitcoin Garch volatility dynaics. https://doi.org/10.1016/j.frl.2018.08.009.

Ashfaque, M.D. Hossain Khan, Rao, Mulpuri V., Li, Qiliang. (20190. Recent Advances in Electrochemical Sensors Detecting Toxic Gases: NO2, SO2 and H2S. Sensors 2019, 19, 905, DOI: 10.3390/s19040905.

Bauer, M.A., Utell, M.J., Morrow, P.E., Speers, D.M., Gibb, F.R. (1986). Inhalation of 0.30 ppm nitrogen dioxide potentiates exercise-induces bronchospasm in asthmatics. Pubmed: Am. Respir. Dis. 1203-1208.

Bollerslev, T. (1987). A conditionally Heteroskedastic Time Seies Model for Speculative Prices and Rates of Return. The Review of Economics and Statistics, 542-547.

Chu, Jeffrey., Chan, Stephen., Nadarajah, Saraless., Osterrieder, Joerg. (2017). GARCH Modelling of Cryptocurrencies. J. Risk Manag 2017, 10, 17; doi: 10.3390/jrfm 10040017.

Dickey, David., and W. Fuller. (1979). Distribution of Estimators for Autoregressive Time Series With an Unit Root. Journal of The American Statistical Association. 74:366.

Ehrlich, R. (1966). Effect of nitrogen dioxide on resistance to respiratory infection. Bacteriol. Pubmed: 604–614.

Enders, W. (1995). Applied Econometric Time Series. John Wiley and Sons Inc, New York.
Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica: Journal of the Econometric Society, 987-1007.
Hanke, J.E. and Wichern, D. (2005). Business Forecasting 8 Edition. Pearson Prentice Hall. New Jersey.
Khan, R.R., Siddiqui, M.J.A. (2014). Review on effects of Particulates; Sulfur Dioxide and Nitrogen Dioxide on Human Health. Int. Res. J. Environment Sci. Vol 3, 70–73
Kim,, Ha Young., Won, Chang, Hyun. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. Elsevier Ltd.
Menteri Negara Lingkungan Hidup. (1997). Keputusan Menteri Negara Lingkungan Hidup No.45/MENLH/10/1997 Tentang : Indeks Standar Pencemaran Udara. Jakarta.
Nisar, J., Topalian, Z.; Sarkar, A.D., Osterlund, L., Ahuja, R. (2013). TiO2 based gas sensor: A possible application to SO2. CrossRef: ACS Appl. Mater. Interfaces Vol 5, 8516–8522.
Santoso, Teguh. (2011). Aplikasi Model Garch Pada Data Inflasi Bahan makanan Indonesia Periode 2005.1-2010.6. Jurusan IESP FF UNDIP Semarang.
Shah, M. et al. (2020). The Development Impact of PT. Medco E & P Malaka on Economic Aspects in East Aceh Regency. Budapest International Research and Critics Institute-Journal (BIRCI-Journal). P. 276-286.
Tsay, R. S. (2005). Analysis of Financial Time Series. Second Ediion. John Wiley and Sons Inc, New York.
Wang, Lu., Ma, Feng., Liu, Jing., Yang, Lin. (2019). Forecasting stock price volatility: New evidence from the GARCH-MIDAS model. https://doi.org/10.1016/j.ijforecast.2019.08.005