TENSOR Pure and Applied Mathematics Journal

April 2021 | Vol. 2 | No. 1 Halaman 33–44

p-ISSN: 2723-0325 e-ISSN: 2723-0333

Analysis of the Changes in 2019-nCov Before and After the Implementation of the "Required Swab PCR-Test for Entry into West Kalimantan via Air Transport" Policy

Nur'ainul Miftahul Huda[1], Nurfitri Imro'ah[2*]

[1] Mathematics Department, Universitas Tanjungpura, Pontianak, Indonesia
Email: nur'ainul@fmipa.untan.ac.id
[2] Statistics Department, Universitas Tanjungpura, Pontianak, Indonesia
*Email: nurfitriimroah@math.untan.ac.id

Manuscript submitted : February 2021;
Accepted for publication : April 2021.

Abstract: Implementing policies during the 2019-nCov pandemic are expected to reduce the number of cases added every day. West Kalimantan is one of the provinces that implements a policy of obliging to include negative results on the PCR-test swab every time they use air transportation to West Kalimantan. In this study, we wanted to know whether there were differences in data behavior before and after implementing the policy. These differences can be analyzed simply by looking at the descriptive statistics of the data. Furthermore, in this study, a time series analysis was also carried out, and the data patterns and the suitable models representing the data. Time series analysis is also needed to predict the next 5 days related to the addition of 2019-nCov cases in West Kalimantan. In modeling, modifications have been made by partitioning the data into two data, namely data before the policy is implemented and the rest is data after the policy is implemented. The result shows that the suitable model for before and after the policy is applied is ARIMA (1,0,0) and ARIMA (7,0,0)(1,0,0)7, respectively. This model shows a better performance in translating problems than using the entire data as input in modeling. The smaller MSE value indicates this than using the ARIMA model (1,0,0) for the entire data (without partition). Therefore, in the prediction stage, a model with partitioned data is used. The results showed that there was a decrease in daily cases in the next five days.

2010 Mathematical Subject Classification : *****(You can write more than one, separated by commas).

Keywords: 2019-nCov, policy, partition, autoregressive, seasonal.

1. Introduction

Coronavirus disease, 2019-nCoV, is an infectious disease caused by a newly discovered coronavirus (WHO, 2020). Since December 1st, 2019 (first detected in Wuhan) to date, as many as 117,166,984 people in the world have been infected with this virus, with total deaths caused by this virus reaching 2,601,465 and total infected people healed of 92,727,926 (last updated March 7th, 2021). USA, India, and Brazil are the three countries with the highest total cases reaching 29M, 11M, and 10M cases (respectively). Since being identified for the first time, namely March 2nd, 2020, Indonesian citizens have passed one year of living side by side with this virus. Indonesia is now in 18th position with total cases, deaths, and recoveries to date respectively 1,379,662; 37,266; 1,194,656. If viewed regionally by province, provinces in Java Island occupy the top positions in total cases. Table 1 shows the last update of 10 provinces with the total cases.

Table 1. List of 10 provinces with the most cases. The same color indicates being on the same island. The proportion is calculated from the total recovered / death divided by the total cases in each province.

Rank	Province	Total Cases	Recovered (Proportion %)	Death (Proportion %)
1	DKI Jakarta	350,425	337,371 (96%)	5,801 (2%)
2	West Java	222,400	182,991 (82%)	2,517 (1%)
3	Central Java	157,738	107,731 (68%)	6,744 (4%)
4	East Java	132,171	120,074 (91%)	9,295 (7%)
5	East Kalimantan	58,039	50,298 (87%)	1,353 (2%)
6	South Sulawesi	57,527	53,403 (93%)	874 (2%)
7	Bali	35,527	32,699 (92%)	973 (3%)
8	Riau	32,039	30,266 (94%)	781 (2%)
9	Banten	30,217	23,174 (77%)	628 (2%)
10	West Sumatera	29,706	28,022 (94%)	654 (2%)

Based on Table 1, it can be seen that the provinces in Java Island dominate the 2019-nCov cases in Indonesia. Meanwhile, on the island of Sumatra, there are two provinces with the most cases. In comparison, the islands of Sulawesi and Kalimantan each have only one province with the most cases of 2019-nCov. Transmission of the virus is difficult to stop because of the mobility of human movements. This transmission no longer only occurs in one region or the same province but can occur between provinces. Namely utilizing transportation, whether land, sea, or air. A person who moves from one place to another has a high chance of contracting and/or transmitting the virus.

The government has implemented various policies to stop the transmission of 2019-nCoV. Some of these policies are stay-at-home policies; Social distancing; Physical Limits; Use of Personal Protective Equipment; Maintain Personal Hygiene; Work and Study at home; Suspend all activities that gather large numbers of people; Large-scale social restrictions; until the implementation of the New Normal policy (Ministry of Foreign Affairs, 2020). From the transportation side, policies to limit the movement of people are also implemented. One of the provinces with tightened policies to suppress the increase in 2019-nCov cases is West Kalimantan Province. If seen in Table 1, West Kalimantan is not included in the 10 provinces with the highest cases. However, the local government took firm action, namely implementing a policy that everyone who enters West Kalimantan Province via air transportation (plane) must show a negative result on the PCR-test Swab. This policy will take effect on December 26th, 2020 (Governor Regulations, 2020). This policy is expected to shut down access to the virus's spread from outside to West Kalimantan residents. Thus it can help reduce the increase in daily cases. The total cases, deaths, and recoveries in West Kalimantan are 4,906, respectively; 4,354; 33 with a total of 519 active cases (last updated March 7th, 2021). In this study, the differences before and after the application of the PCR-test swab policy in West Kalimantan will be seen. Prediction of daily cases is also the aim of this study. Research on covid-19 cases has been carried out by many researchers. Caraka, et.al (2021) discuss about impact of COVID-19 large scale restriction on environment and economy in Indonesia. The impact to education is discussed by Abidah, et.al (2020). On poverty's side is also discussed by Asep, et.al (2020).

The application's difference before and after the PCR-test swab policy can be made by analyzing descriptive statistics of the data used. One of the parameters is to see the average of the data, both before and after implementation and the whole data. Meanwhile, daily case predictions can be analyzed with a time series model. Some of these models are the Autoregressive (AR) and Seasonal Autoregressive (SAR) models. In this study, a modification was made to determine the best model, namely by partitioning the data into two parts (before and after the policy implementation). This aims to see the differences in the treatment of the two conditions. Section 2 explains the time series model used. Then the data analysis is given in section 3. Finally, the conclusion lies in section 4.

2. Time Series Model

2.1. Autoregressive (AR) model

The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term. Autoregressive processes are regressions on themselves. Specifically, a pth-order autoregressive process $\{Y_t\}$ satisfies the equation (Cryer and Chan, 2008)

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$$

where Y_t is a linear combination of the p most recent past values of itself plus an "innovation" term e_t that incorporates everything new in the series of time t that is not explained by the past values.
There are three main steps in modeling the data,

1) Identifying the order, how to choose appropriate values for p for a given series. By observing the general behavior of Autocorrelation Function (ACF) and Partial ACF (PACF), the order can be determined.
2) Estimating the parameter, how to estimate the parameters of a specific model. Some methods in estimating the parameters are Maximum Likelihood Estimation (MLE) and Ordinary Least Square (OLS).
3) Diagnostic checking, how to check on the appropriateness of the fitted model and improve it if needed. Model diagnostics is concerned with testing the goodness of fit of a model. The complementary approach is analysis of residuals from the fitted model. By concerning to the autocorrelation of residuals (independency) and the distribution of residuals (Normal distribution).

Those three main steps are developed by George E. P. Box and G. M. Jenkins (1976).

2.2. Seasonal AR (SAR) model

Seasonality in a time series is a regular pattern of repeated changes over S time periods, where S defines the number of time periods until the pattern repeats. The seasonal AR (SAR) incorporates both non-seasonal and seasonal factors in a multiplicative model. The notation is (Wei, 2006)

$$ARIMA(p,0,0) \times (P,0,0)S$$

where p is non-seasonal AR order, P is seasonal AR order, and S is time span of repeating seasonal pattern.
Let Z_t be the ARIMA$(p,0,0) \times (P,0,0)S$,

$$\Phi(B^S)\phi(B)(Z_t - \mu) = e_t \tag{1}$$

where $\Phi(B^S) = 1 - \Phi_1 B^S - \cdots - \Phi_P B^{PS}$,$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, and e_t is innovation term at time t.
For example in ARIMA$(1,0,0) \times (1,0,0)6$, then Eq. (1) can be written as

$$\Phi(B^6)\phi(B)(Z_t - \mu) = e_t$$

$$(1-\Phi_1 B^6)(1-\phi_1 B)(Z_t-\mu)=e_t$$

The seasonal model has the same structure as the non-seasonal. The procedure is same as like non-seasonal model. The addition is in identifying the order, that is identifying the order for the seasonal model (Wei, 2006).

3. Data Analysis

3.1. Descriptive Statistics Analysis

The data used in this paper are secondary data about daily new cases of 2019-nCov in West Kalimantan Province since October 24th, 2020 – February 26th, 2021. The sample size of data is 128 observations. The data are obtained from National Disaster Management Authority (NDMA). In modeling the data, two methods are used. The first is using full data to model using one of the time series model. In this paper, we called it as full data. The second is divided the data into two parts, those are the data before the policy implemented (October 24th, 2020 – December 25th, 2020) and the data after the policy implemented (December 26th, 2020 – February 26th, 2021). In this paper, we called it subsequently as 1st part of data and 2nd part of data. Table 1 describes the descriptive statistics of three kinds of data, those are full data, 1st part of data, and 2nd part of data.

Table 2. Descriptive Statistics of full data (128 samples), 1st part of data (64 samples), and 2nd part of data (64 samples). First and second part interpret the condition before and after the policy enforced

	Full Data	1st Part of Data (Before The Policy)	2nd Part of Data (After The Policy)
Minimum	0.00	0.00	0.00
1st Quartile	13.00	13.00	18.00
2nd Quartile	26.00	24.00	26.00
3rd Quartile	33.00	32.50	33.00
Maximum	83.00	83.00	56.00
Mean	24.31	24.23	24.37
Variance	234.37	143.77	196.59
Deviation Std.	15.31	11.99	14.02

The mean of all the types of data is around 24. It means that every day the increasing of daily new cases in West Kalimantan is about 24 even after the policy enforced. However, the mean after is greater than before the policy enforced (24.37 > 24.23). For the variance's side, after the policy enforced is also greater than before the policy enforced. While the maximum of daily new cases is 83, right before the policy enforced. Visually, the data can be seen in Figure 1. Fig 1.a shows the plot of full data. The red line is for daily new cases before the policy enforced and the blue line is for after the policy enforced. Visually, the daily new cases are not significantly different before and after the policy enforced.

More detail, it can be seen in Fig 1.b which describes plot per week, nine weeks before and after the policy enforced. The first week after the policy implemented shows the significant decreasing of the daily new cases' average, that is from 21.18 to 13.86. Also in the second week shows the same thing however it is increasing than the first week before that. It is caused by many passengers from out of West Kalimantan canceled their flight due to the implementation of that new policy. Some of them thinks that the policy is not fair, starting from the cost of the PCR test they have to pay and the PCR test validity period is short. But if we see the third week after the policy implemented, the average per week of daily new cases is highly increasing. One of the causes from the transportation side is many people choose another transportation, those are either using ship or bus from other Kalimantan's province. Because with this transportation, the policy is not implemented for the PCR test but only for the rapid antigen test. So that choosing transportation other than air allows them to save on travel costs, even though it will increase the length of the trip. However, in terms of the increase in daily cases, this can actually lead to an increase due to the effectiveness of the rapid test, which is the test's low accuracy level and fearing that a nonreactive result could provide a false sense of security (WHO, 2020). The fourth week was the peak of the increase in the average per week of cases, namely January 17-23, 2021. One of the reasons was the increase in visits by citizens from China to West Kalimantan. The increase occurred because of the Chinese New Year and Cap Go Meh celebrations which were centered in Singkawang City, West Kalimantan. Even though it has gone through a screening process using a swab test, it is possible that transmission of the virus may still occur on the way to an area. This fourth week is the highest average case increment compared to the other weeks before and after the policy was implemented. For the following weeks around the mean of the overall cases. Fig 1.c gives the boxplot for both of the 1st and 2nd data. No more outliers detected. It means there are not significant high daily cases or low daily cases.

3.2. Modelling Data

There are two model used, those are model with full data and partition data. Both of the data are analyzed using Autoregressive (AR) and Seasonal AR (SAR) model.

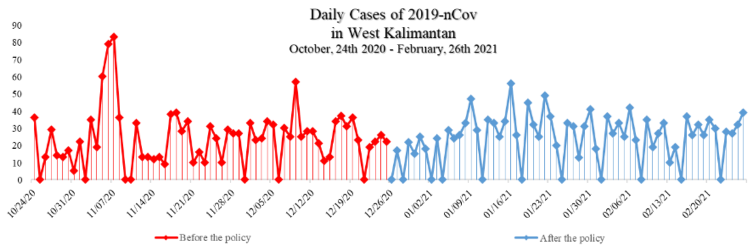
3.2.1. Full Data

Data plot of full data can be seen in Fig. 1.a. Visually, the data are already stationary in mean and variance. Therefore, the next step is identifying the order of the model based on ACF and PACF (see Fig. 2). Some probable order are ARIMA(1,0,0) and SARIMA(1,0,0)(2,0,0)6. After getting the order, estimate the parameters for each model. Table 3 gives the parameter gotten using Maximum Likelihood Estimation. Based on the Table 2, ARIMA(1,0,0) shows the smallest values of error measures (AIC and MAE). It means the ARIMA(1,0,0) is better for analyzing those data than SARIMA(1,0,0)(2,0,0)6 model. Let Y_t follows the ARIMA(1,0,0) model,

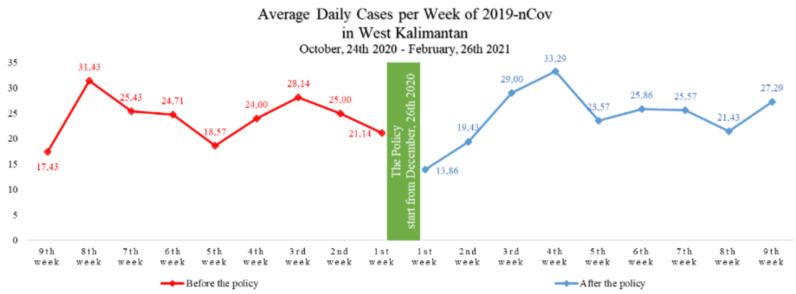
$$Y_t = 24.37 + 0.22Y_{t-1} + e_t$$

where Y_t is the daily case at time and e_t is noise process at time t for $t = 1,2, \ldots ,128$.

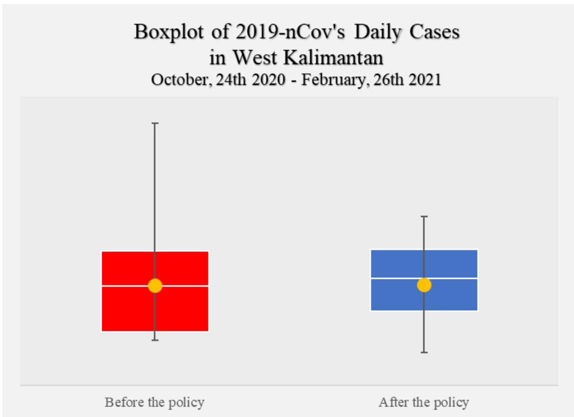
The last step of modeling the data is diagnostic checking. Figure 3 shows the visualization of residuals in terms of standardized residuals (Fig. 3.a), ACF of residuals for testing the independence (Fig. 3.b), and the distribution of the residuals (Fig. 3.c and 3.d). The residuals model should be independent among the lag of times. If the residual autocorrelation is within the 5% significance limit, then the correlation can be ignored, which means no residual correlation between time lags. On the other hand, if it is outside the limit of significance, there is still a correlation in the residuals. This indicates that the selected model needs to be reviewed starting from the selection of the model order. This shows that involving all observations using the ARIMA model (1,0,0) results in residuals that are still correlated between time lags (see Fig. 3.b). This means that this model is not good enough to represent the problem. This is reinforced by the residual distribution, which should be close to the normal distribution (see Fig. 3.c and 3.d).



(a)

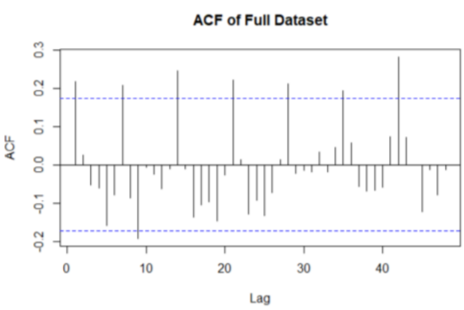


(b)

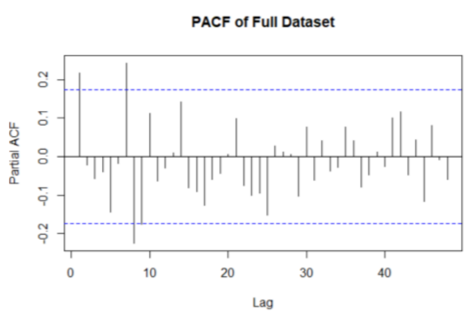


(c)

Fig. 1. (a) Plot of full data, (b) The average per week, and (c) Boxplot of 2019-nCov's daily cases. The red line and box interpret the data before the policy enforced, while the blue line and box interpret conversely.



(a)



(b)

Fig. 2. (a) ACF and (b) PACF of full data

Table 3. Parameter and error measures for AR and SAR model using full data. The red text shows the smallest value.

Model	Parameter				Error Measures		
	mean	ϕ	Φ_1	Φ_2	AIC	RMSE	MAE
ARIMA(1,0,0)	24.37	0.22	-	-	1060.49	14.88	11.41
SARIMA(1,0,0)(2,0,0)6	24.38	0.23	-0.11	-0.08	1062.54	14.76	11.42

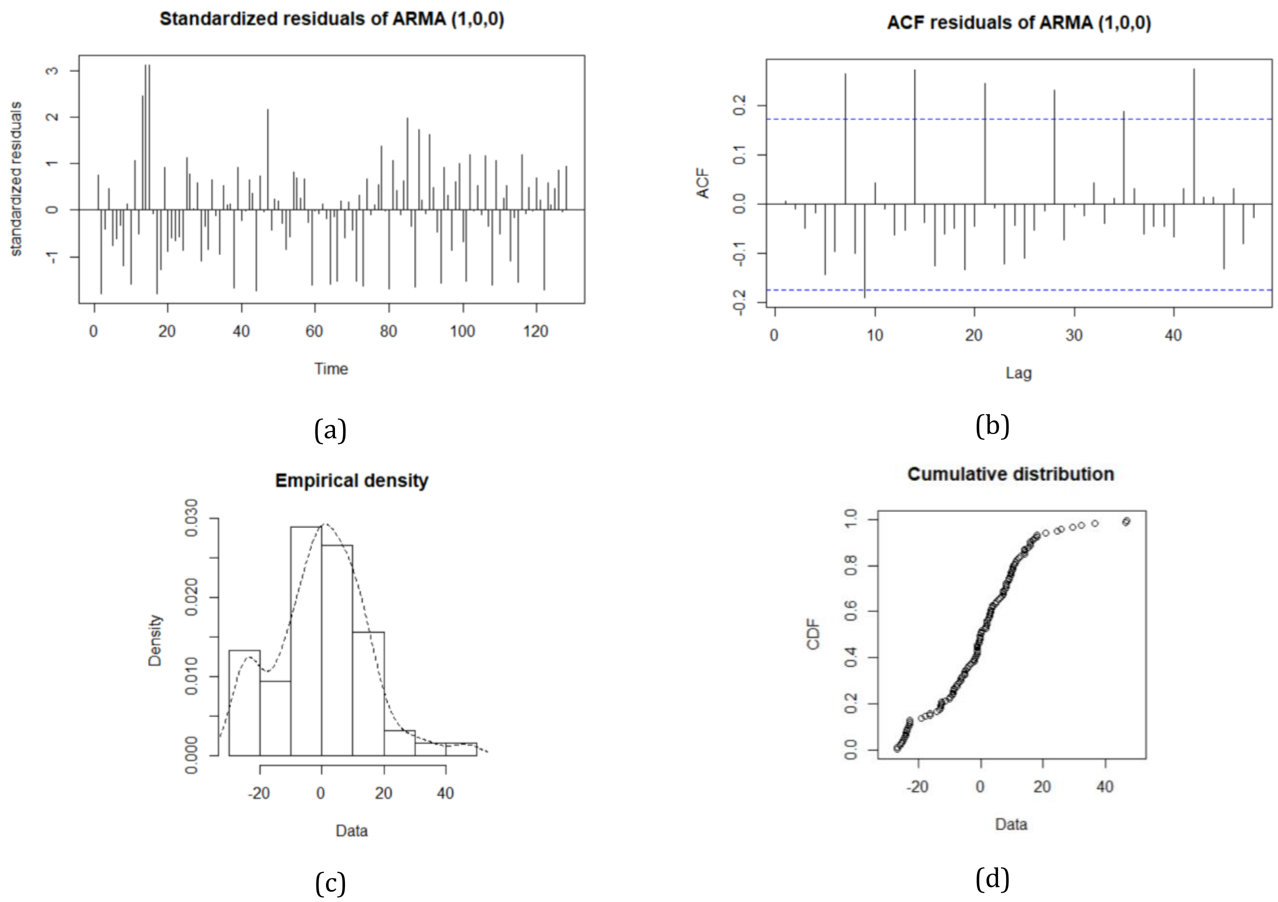


Fig. 3. Diagnostic checking comprises (a) standardized residuals, (b) ACF, (c) empirical density, and (d) cumulative distribution of residuals in ARIMA (1,0,0) model

To be sure, Fig. 4 shows the plot of observations vs. fitted values obtained from the ARIMA model (1,0,0). The fitted values do not follow the fluctuations in actual observations. This, of course, impacts the large residual values, namely the difference between the observations and the fitted values.

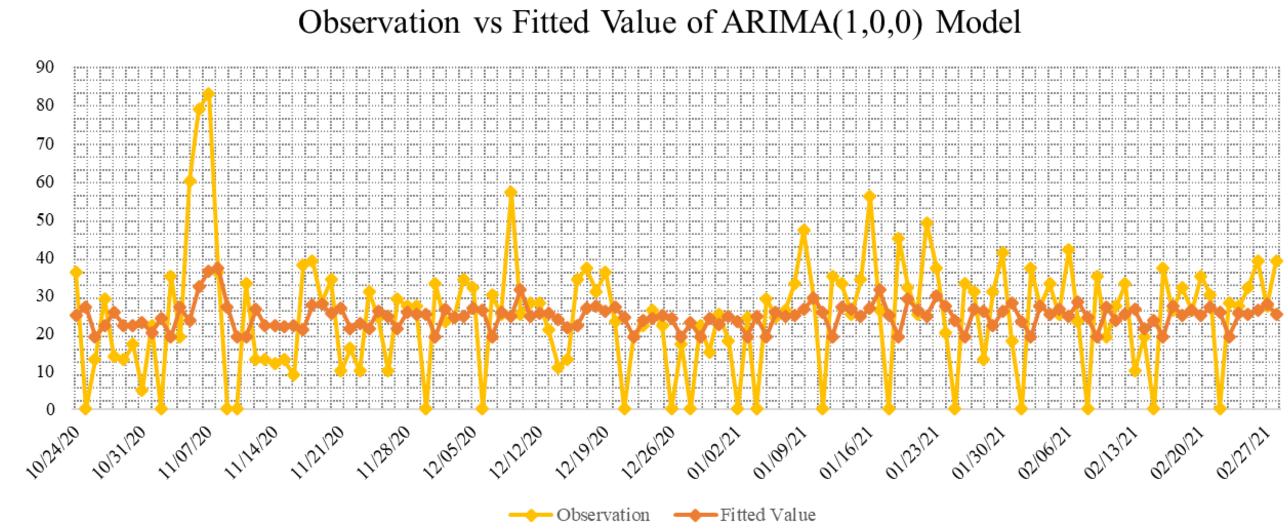


Fig. 4. Plot of observation vs. fitted values based on ARIMA(1,0,0) model.

3.2.2. Partition data

Modeling long-term fluctuating data using only one model poses a risk in choosing the model to interpret the results. In this study, the data is partitioned into two, considering the government's large policies. Namely, the addition of requirements to enter West Kalimantan via air transportation shows a negative result on the PCR-test swab. The initial assumption was that there should be differences in data patterns before and after the policy was implemented. Thus in modeling, it is also partitioned into two parts. The first is modeling the data before implementing policies. Then the second is modeling the data after the policy is applied. Order identification for each model can be seen in Figure 5.

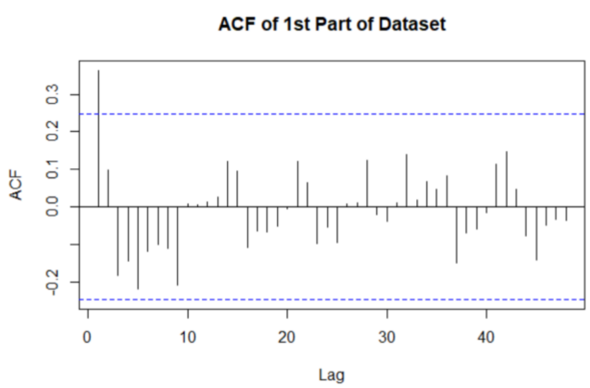
The 1st part of the data's model order is ARIMA(1,0,0) (autocorrelation cut off after lag 1 in PACF, see Fig. 5.b). The 2nd part of the data is ARIMA(7,0,0)(1,0,0)7 (autocorrelation of non-seasonal and seasonal cut off after lag 1 and a pattern repeating every 7 lags, see Fig 5.d). Several possible models other than those mentioned have been tested (not shown in this study). The result is that both models are models with the smallest error measures. After estimating the parameters using MLE, the following model is obtained. Let $Z_t^{(1)}$ follows the ARIMA(1,0,0) model. Superscript (1) in Z_t indicates the data used is the 1st part of data.

$$Z_t^{(1)} = 24.07 + 0.36Z_{t-1}^{(1)} + e_t^{(1)}$$

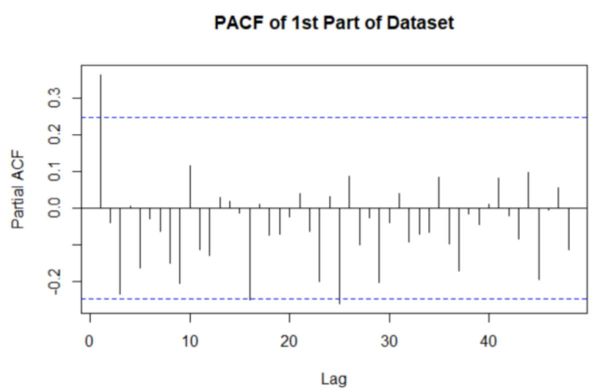
And also superscript (2) is for 2nd part of data used.

$$(1-\Phi_1 B^7)(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_7 B^7)(Z_t-\mu)=e_t$$

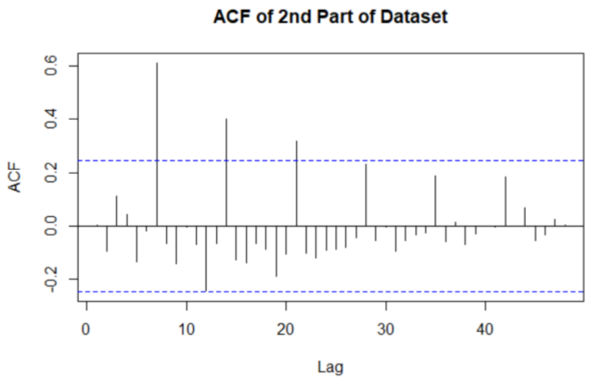
$$(1+0.08B^7)(1-0.04B-0.01B^2-0.08B^3+0.04B^4+0.03B^5+0.02B^6+0.72B^7)\left(Z_t^{(2)}-22.85\right)=e_t$$



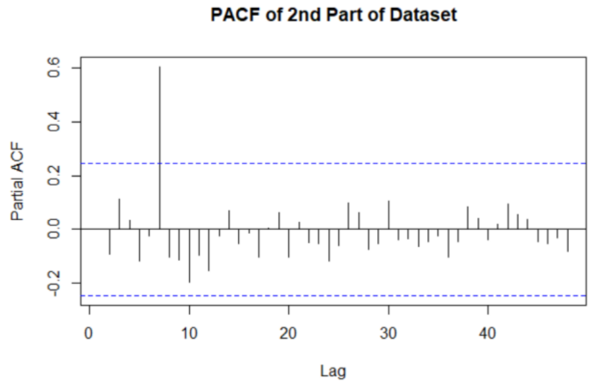
(a)



(b)



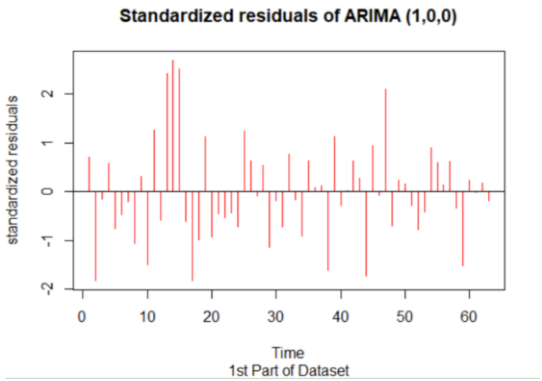
(c)



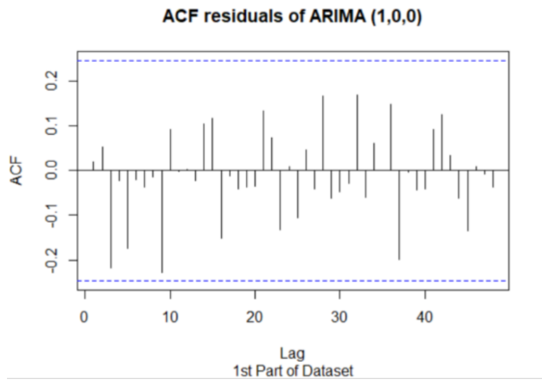
(d)

Fig. 5. ACF and PACF plot for (a);(b) 1st part of data, and (c);(d) 2nd part of data.

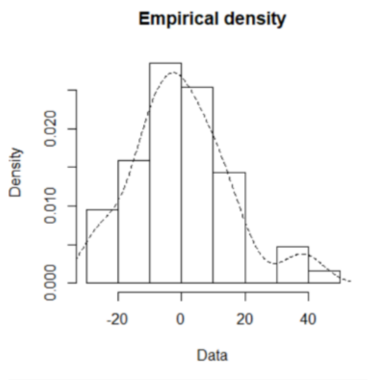
At the diagnostic checking stage (visually it can be seen in Figure 6), the residuals of both models showed independence between time lags as indicated by the correlation within the 5% significance limit (see Figs. 6.b and 6.f). Furthermore, in terms of the residual distribution, the two models also approach the normal distribution (see Fig. 6.c, 6.d, 6.g, and 6.h). Thus it can be said that this model is suitable for interpreting the problems in this study. When viewed from the observation plot vs. fitted values (see Figure 7), the ARIMA(1,0,0) and the ARIMA(7,0,0)(1,0,0)7 are slightly better at approaching the actual observation than using one model for analyzing long-term fluctuating data. The MSE value reinforces this between the model involving all data (using the ARIMA model (1,0,0) in section 3.2.1), and the data partition (using the ARIMA(1,0,0) and ARIMA(7,0, 0)(1,0,0)7, in section 3.2.2) are 221.39 and 167.21, respectively.



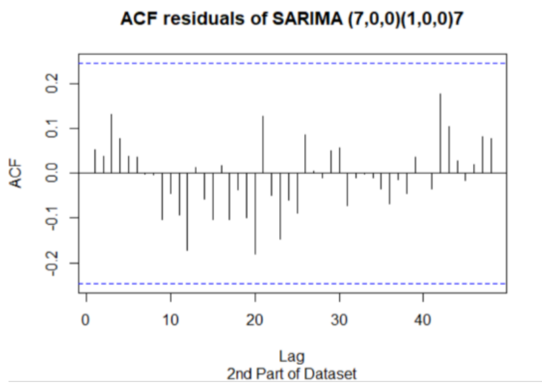
(a)



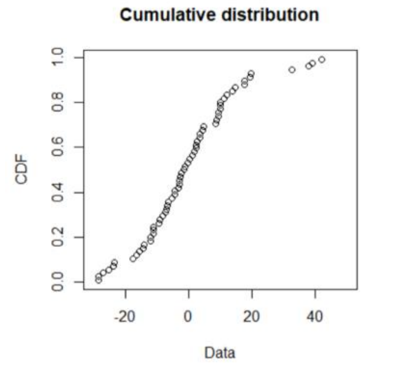
(b)



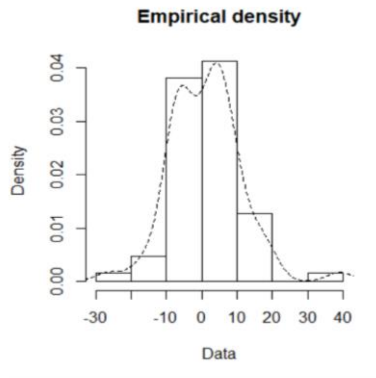
(c)



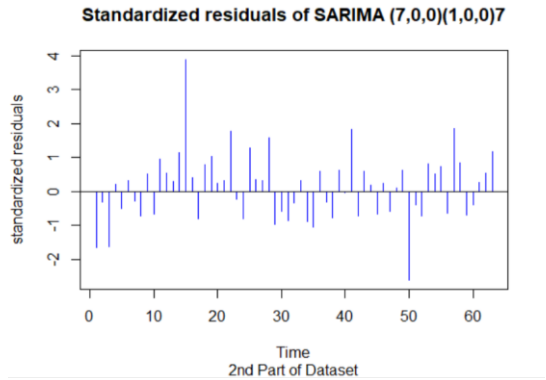
(f)



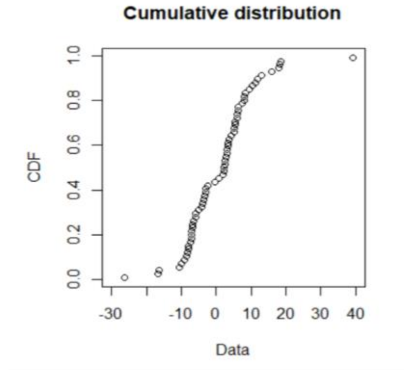
(d)



(g)



(e)



(h)

Fig. 6. (a) and (e) Standardized residuals, (b) and (f) ACF residuals, (c) and (g) Empirical density, (d) and (h) Cumulative distribution. All is respectively for 1st and 2nd part of data.

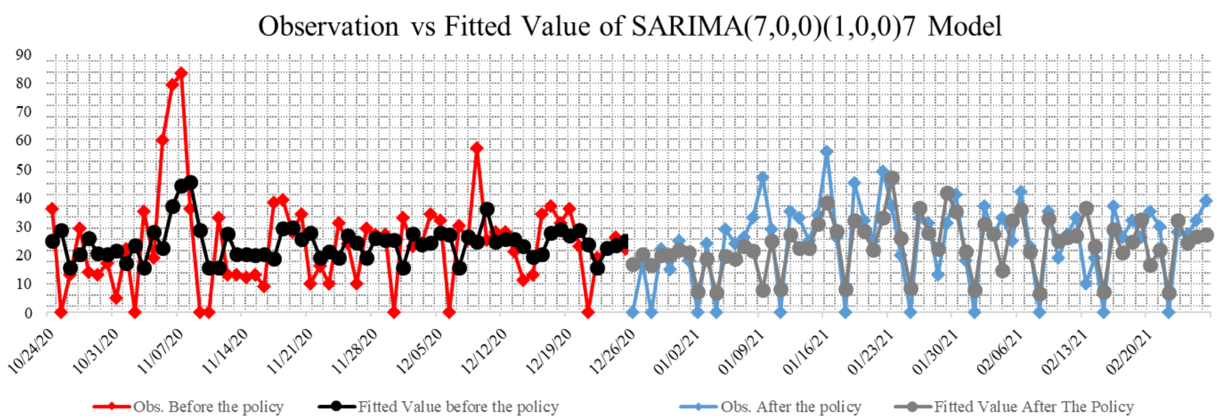


Fig. 7. Plot of observation vs. fitted values based on ARIMA(1,0,0) and ARIMA(7,0,0)(1,0,0)7 model.

3.3. Forecasting

Prediction is the main objective in modeling time series data. After getting the right model to analyze a problem, the next step is to predict future time. In this study, predictions were made for the next 5 days based on the ARIMA(7,0,0)(1,0,0)7 model. Table 4 details the predicted numbers, both point prediction and interval prediction for the 80% and 95% confidence intervals. Visually, the prediction for the next five days is depicted in Figure 8. The results of the prediction show that there will be a decrease in daily cases within five days. There was a big decline on March 1, 2021, which is only predicted to increase by 8 cases. The results of this prediction can be confirmed when the latest data is updated. Note that the MSE of this model is 167.21, which means an error of ± 13 cases. So there may be a difference of ± 13 cases in prediction accuracy.

Table 4. Forecasting for five days ahead using ARIMA(7,0,0)(1,0,0)7 model

Time	Point Forecast	Confidence Interval			
		Lo 80	Hi 80	Lo 95	Hi 95
Feb, 27th 2021	31.21	17.29	45.13	9.92	52.50
Feb, 28th 2021	28.68	14.74	42.61	7.37	49.99
March, 01st 2021	8.09	-5.84	22.03	-13.22	29.41
March, 02nd 2021	26.06	12.07	40.04	4.67	47.45
March, 03rd 2021	25.11	11.12	39.11	3.71	46.51

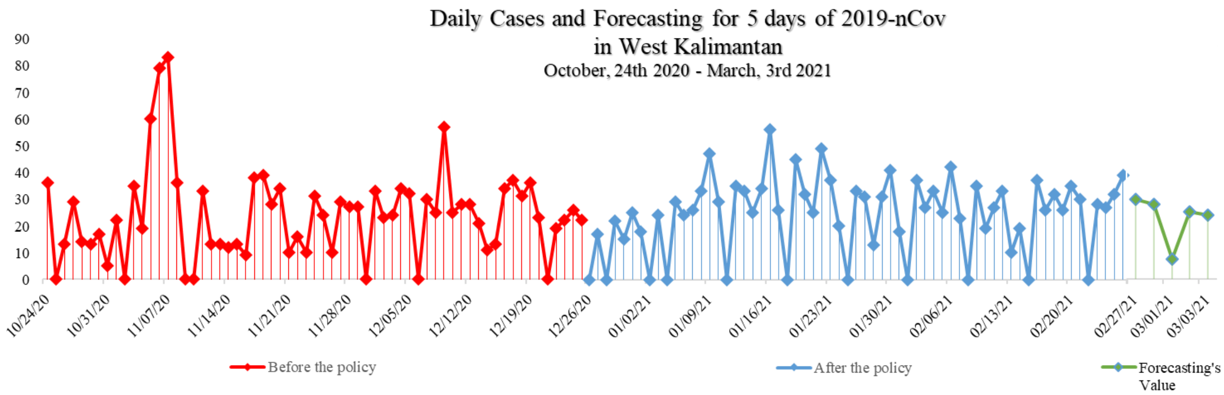


Fig. 8. Forecasting plot for five days ahead.

4. Conclusion

The policy's application in adding the requirements to use air transportation, namely including evidence of negative results on the PCR-test swab, gave results that were not much different before the policy was implemented. This can be seen from the average before and after the policy is implemented, which does not significantly differ. Also, compared with the overall average data for 4 months, there is no significant difference in the mean. Changes are only visible in the first and second weeks after the policy is implemented. After that, the number of additional cases returned to their original condition before the policy was implemented. However, this does not mean that this policy has not been successfully implemented; many factors cause cases to continue to increase.

Dividing data into two parts, before and after the policy is implemented, can help find the right model in interpreting the problem. Before the policy was implemented, the data followed a pattern in the autoregressive model. Meanwhile, after the policy is implemented, the data is more suitable to be analyzed using the seasonal autoregressive model. Compared to using the entire data for modeling, this model shows better results.

Acknowledgment

References

[1] World Health Organization (WHO). (2020). *"Coronavirus disease (COVID-19) pandemic".* Retrieved February 25th, 2021.

[2] Ministry of Foreign Affairs. (2020). "*The government policy of the republic of indonesia related to the plague of covid-19"*. Retrieved March 2nd, 2021.

[3] Governor Regulations. (2020). West Kalimantan governor regulation number 149 in 2020. Indonesia: West Kalimantan.

[4] Caraka, R. E., Lee, Y., Kurniawan, R., *et al.* (2020). Impact of COVID-19 large scale restriction on environment and economy in Indonesia. *Global Journal of Environmental Science and Management, 6(SI),* 65-84*.*

[5] Abidah, A., Hidaayatullaah, H. N., Simamora, R. M., *et al.* (2020). The Impact of Covid-19 to Indonesian Education and Its Relation to the Philosophy of "Merdeka Belajar". *Studies in Philosophy of Science and Education, 1L* 38-49.

[6] Asep, S., Izzati, R. A., and Suryadarma D. (2020). The Impact of COVID-19 Outbreak on Poverty: An Estimation for Indonesia. *The SMERU Research Institute.*

[7] Cryer, J.D. and Chan, KS., (2008). *Time Series Analysis: With Apllication in R*, Second Edition. USA: Spinger Science and Businiess Media

[8] George E. P. Box and G. M. Jenkins. (1976). *Time Series Analysis: Forecasting and Control.* San-Fransisco: Holden-Day.

[9] Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods Second Edition.* New Jersey: Pearson Prentice Hall.

[10] (Wei, 2006)

[11] National Disaster Management Authority (NDMA). (2021). *"Covid-19 situation in Indonesia".* Retrieved March 5th, 2021.