



INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES



Forecasting the Spread of Daily Confirmed Covid-19 Cases in Malaysia

Siti Sarah Januri, Isnewati Ab Malek, Noryanti Nasir & Zaitul Anna Melisa Md Yasin

To Link this Article: <http://dx.doi.org/10.6007/IJARBSS/v12-i2/12114> DOI:10.6007/IJARBSS/v12-i2/12114

Received: 13 December 2021, **Revised:** 17 January 2022, **Accepted:** 05 February 2022

Published Online: 23 February 2022

In-Text Citation: (Januri et al., 2022)

To Cite this Article: Januri, S. S., Malek, I. A., Nasir, N., & Yasin, Z. A. M. M. (2022). Forecasting the Spread of Daily Confirmed Covid-19 Cases in Malaysia. *International Journal of Academic Research in Business and Social Sciences*, 12(2), 322–334.

Copyright: © 2022 The Author(s)

Published by Human Resource Management Academic Research Society (www.hrmars.com)

This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at: <http://creativecommons.org/licenses/by/4.0/legalcode>

Vol. 12, No. 2, 2022, Pg. 322 – 334

<http://hrmars.com/index.php/pages/detail/IJARBSS>

JOURNAL HOMEPAGE

Full Terms & Conditions of access and use can be found at
<http://hrmars.com/index.php/pages/detail/publication-ethics>



INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES



Forecasting the Spread of Daily Confirmed Covid-19 Cases in Malaysia

Siti Sarah Januri, Isnewati Ab Malek, Noryanti Nasir & Zaitul
Anna Melisa Md Yasin

Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Cawangan Negeri
Sembilan, Kampus Seremban, 70300 Seremban, Negeri Sembilan, Malaysia
Corresponding Author Email: isnewati@uitm.edu.my

Abstract

COVID-19 is rapidly expanding across the globe. As a Southeast Asian region, Malaysia has also been affected by COVID-19. Since the COVID-19 outbreak first emerged in China at the end of 2019, Malaysia has taken precautionary measures to prevent entering the nation. However, since COVID-19 is more than undoubtedly unstoppable, Malaysia eventually received the first case in early January 2020. The increase in the epidemic scale has led to the (preface of non-pharmaceutical countermeasures). Hence, it is of utmost importance to analyze the trends of the cases to develop a forecasting model that could anticipate the number of confirmed COVID-19 cases in Malaysia and select the best forecasting model based on forecast measure accuracy to forecast the future course of outcomes. For this purpose, the number of daily cases from 15 March 2020 to 31 March 2021 was retrieved from the Ministry of Health (MOH) website and estimated using the Box-Jenkins approach. There were five models developed such ARIMA (1,1,1), ARIMA (1,1,2), ARIMA (1,1,3), ARIMA (2,1,1) and ARIMA (2,1,2). The models' effectiveness is evaluated based on AIC, BIC and RMSE criteria. The findings indicate that ARIMA (1,1,3) is the preferred model for forecasting since it has better performance regarding adopted criteria than compared models. The forecasted values showed an upward trend of COVID-19 cases until January 2022. In conclusion, subsequent studies would yield more discoveries and a more systematic approach to have better and more accurate forecasting. In the instance of the COVID-19, the recommended model appears to be correct. More complex modelling methodologies and extensive information on the disease are required to forecast the pandemic.

Keywords: ARIMA, Forecast, COVID-19, Pandemic

Introduction

All around the world is experiencing a contagious disease pandemic called COVID-19 which stands for coronavirus disease 2019. Coronaviruses are a huge family of respiratory viruses believed to trigger diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS), ranging from the common cold to more serious diseases (Khounpaseuth, 2020). The epidemic was caused by a form of coronavirus that had never been discovered anywhere else in the world until it was uncovered in Wuhan, China in

December 2019 (Khounpaseuth, 2020). As a result, the World Health Organization (WHO) confirmed the COVID-19 epidemic as a Public Health Emergency of International Concern (PHEIC) (WHO 2020a) on 30 January 2020, indicating that it would become a global pandemic.

On 25 January 2020, the first case of COVID-19 was spotted and traced back to three locals who had previously been in contact with a contaminant in Singapore (Elengoe, 2020). On 18 March 2020, Malaysia's Prime Minister, enacted a Movement Control Order (MCO) after the number of positive cases surpassed 553 (Elengoe, 2020). The MCO, along with many other measures such as social distancing, isolation and quarantine was designed to break the chain of COVID-19 transmission in Malaysia (Singh et al., 2020). The Malaysian government have deployed various technologies such as MySejahtera contact device tracing, drive-through screening services, and a field hybrid Intensive Care Unit (ICU) (WHO, 2020). Malaysia reported zero cases of local COVID-19 transmissions on 1 July 2020, for the first time since March (Tee, 2020). Also, Malaysia has been acknowledged among the most promising countries in the COVID-19 control pandemic. However, at the end of July, multiple new local coronavirus transmissions have shown up, ending a streak of a month without such cases (BERNAMA, 2020).

The method of forecasting the trend number of COVID-19 cases may be different and varies but they have similarities in the aim of recognizing and analyzing the trend of the COVID-19 confirmed cases (Konarasinghe, 2020a). A study conducted by Singh et al. (2020) used Auto-Regressive Integrated Moving Average (ARIMA) model to forecast daily confirmed COVID-19 cases in Malaysia between the period of 22 January 2020 and 1 May 2020. The overall conclusion is that ARIMA (0,1,0) models with their most suitable covariates selected are practical tools in Malaysia to track and forecast trends in COVID-19 cases. Edre et al. (2020) predicted the incidence of COVID-19 in Malaysia based on the Movement Control Order (MCO). The data of Malaysia COVID-19 new daily cases was used from the start of Movement Control Order (2020) until the end of phase 4 MCO (12 May 2020). The researchers utilized the Expert Modeler approach to predict COVID-19 cases in SPSS (Statistical Package for Social Sciences) and ARIMA model in R software around Malaysia. The result has been observed that both models used, ARIMA and Expert Modeler have appeared to maintain the rate of new cases.

The ARIMA model has been used by several researchers to predict the spread of COVID-19 in many countries. Auto-Regressive Integrated Moving Average (ARIMA) and Nonlinear Autoregressive Neural Networks (NAR) were used to generate forecasting models for COVID-19 cases in India using time series methods in the work of (Khan & Gupta, 2020). In both models, an increasing pattern was seen in the COVID-19 cases. On 30 January 2020, COVID-19 first case was reported in India and is anticipated to rise rapidly and exponentially by about 1500 cases each day throughout the next 50 days. In other study by Gupta and Pal (2020) forecasted the number of infected cases of COVID-19 outbreak in India. The primary results show that the number of infected cases in India is gradually increasing, with the mean number of infected cases per day growing from 10 to 73 between the first and 300th cases. According to the ARIMA model, infected cases could reach 700,000 in the next 30 days in the worst-case scenario, but the scenario could keep the number at 1000 to 1200. Furthermore, out of the existing 536 patients, the ARIMA model forecast for the next 30 days is around 7000.

Another study that used the ARIMA model in forecasting COVID-19 related issue done by (Kufel, 2020). The study introduced the ARIMA model for estimating the dynamics of confirmed cases of COVID-19 at various stages of COVID-19 of the pandemic that occurred in

32 selected European countries with the highest rate of infection. The various stages of the pandemic are the first stage of growth, occurring between the time when the highest number of everyday cases surpassed and the time when the epidemic eradicated. The finding showed that the ARIMA (1,2,0) model is very efficient in assessing cumulative case dynamics and analyzing parameters. The investigator concluded that ARIMA models can be used as an effective and direct epidemic monitoring system at national and regional levels (Kufel, 2020).

Konarasinghe (2020b) used the Autoregressive Integrated Moving Average (ARIMA), Double Exponential Smoothing (DES) and Autoregressive Distributed Lag Models (ADLM) to achieve their research objective for the research on Modelling COVID -19. Epidemic of USA, UK and Russia were conducted between 22 January to 28 May 2020 using the daily confirmed cases from WHO database. From the results, the ARIMA model did not meet the validation requirements in any of the countries, but ADLM and DES did. It is deduced that the ADLM is the most fitting model for the United State forecast and the UK's best model is the DES. Both versions are equally fine for Russia after all.

A study on Modelling and Forecasting for the number of COVID-19 cases with the Curve Estimation Models was conducted by (Yonar et al., 2020). The research included Germany, the United Kingdom, France, Italy, Russia, Canada, Japan, and Turkey. It was forecasted using several curve predictions models, Box-Jenkins (ARIMA), and Brown/Holt linear exponential smoothing methods between 22 January to 22 March 2020. As a result, the findings in this data collection from Japan (Holt Model), Germany (ARIMA (1,4,0)) and France (ARIMA (0,1,3)) are statistically significant, however, not technically eligible. The findings of the UK (Holt Model), Canada (Holt Model), Italy (Holt Model) and Turkey (ARIMA (1,4,0)) are much more accurate.

The next study about modelling and forecasting of the COVID-19 growth curve in India from 4 March 2020 until 11 July by Sharma and Nigam (2020) used regression analysis (exponential and polynomial), ARIMA, Exponential smoothing and Holt-Winter models. For the time series models fitting was compared using MAPE. The researchers obtained ARIMA (5,2,5) by having the lowest AIC value among all considered models. The exponential smoothing is not perfectly accurate for the data, but the researcher points out that Holt-Winter model is around 97.11 accurate. Both ARIMA (5,2,5) and Holt-Winter models suggest a rise in the number of cases in the coming days.

Subsequently, the objective of this study is to distinguish the best fit model and make forecasts for a future period. The expectation of the pace of disease of COVID-19 has become crucial for choices and policymakers in Malaysia. It is critical to gauge the rate as precisely as conceivable utilizing solid logical methods. Estimating the number of confirmed cases would help policymakers in a particular area to survey their present medical care limit and conclude which estimates should be taken to check and control the spread of COVID-19 (Alabdulrazzaq, 2021).

The remainder of the paper is organized as follows. In Section 2 emphasized the method used to forecast the number of COVID-19 confirmed cases and provide justification for a best-fit ARIMA model. The forecast generated by the ARIMA model is described in the section results and discussion. Finally, the accuracy of the ARIMA based prediction in conclusion.

Methodology

Description of Data

This study used secondary data that was extracted from Malaysia's Ministry of Health. The new daily confirmed cases of COVID-19 in Malaysia were obtained from Malaysia's Ministry of Health website from the period 15 March 2020 to 31 March 2021. A total of 382 observations were obtained with the daily new cases as the main variable analyzed.

Box-Jenkins Methodology

The Box-Jenkins method is associated with ARIMA's general modelling. It was first developed by George E. P. Box (University of Wisconsin, USA) with Gwilym M Jenkins (University of Lancaster, UK) in 1976. Box-Jenkins Analysis focused on the systematic way of identifying, fitting, checking, and using Autoregressive Integrated Moving Average (ARIMA) time series models. It is an ideal approach for mid to long-time series which uses at least 50 observations.

Mixed Autoregressive Integrated Moving Average (ARIMA) Model

If the variable's stationary assumption was not satisfied, then the modelling of ARIMA was generated. It was important to differentiate the data series to achieve stationary in this formulation. In general terms, the model obtained was defined as ARIMA (p, d, q) , was generated. It was important to differentiate the data series to achieve stationarity in where the symbol 'd' defined the number of times the y variable must be differenced to reach stationary. A simple model case ARIMA (1,1,1) is as shown below,

$$w_t = \mu + \phi_1 w_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (1)$$

where $w_t = y_t - y$, serve the first difference of the series and considered to be stationary. In this scenario, the values of $p = 1$, $d = 1$ and $q = 1$. Equation (1) could also be written as,

$$w_t - \phi_1 w_{t-1} = \mu - \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (2)$$

The first difference did not make the sequence stationary on some occasions. If a second difference was needed, then the second-order series would be integrated, where the value of d was two.

Assumption of Box-Jenkins

Under Box-Jenkins methodology, the ARIMA model behaves on the assumption which the data series is stationary. It has been a norm assumption among numerous practices and tools in time series analysis. Stationary means that the data series does not show growth or decline over time, and vice versa for the non-stationary data series (Palachy, 2019). If the assumption was not fulfilled, then before ARIMA could be used, it was important to perform the essential procedures to transform it to obtain stationarity. A simple procedure for eliminating non-stationary is by performing differencing.

The non-stationary of a series can be evaluated either by simple observation of the plotted data or more specifically by using statistical test procedures. Eagle and Granger (1987) proposed the Augmented Dickey-Fuller (ADF) because of its robust critical values. The first statistical test to be developed for testing the null hypothesis was the ADF test. The null hypothesis is that a unit root is in an autoregressive model of a given time series. Consequently, the process is non-stationary.

Stages in ARIMA Model Development

There are three major stages for the basis of the Box-Jenkins modelling approach: model identification, model estimation and validation and model application. The data series must be prepared due to the implementation of these stages, like trying to stabilize the variance through data transformation, monitoring inconsistent or missing values and satisfying the stationary state.

Model Identification

Box-Jenkins methodology assumed that a data series is stationary. For a series that did not fulfil the stationary condition, it would be called non-stationary series. The method of transforming non-stationary data series into a stationary data series is called differencing. By differencing, the mean of a time series was stabilized by eliminating changes in the level of a time series thus consequently dispensing trend and seasonality. When a data series appeared non-stationary, the first difference was performed. The first difference was defined as, $\Delta y_t = (y_t - y_{t-1})$ where y_t is the current value and y_{t-1} is the previous value.

If the data remained non-stationary, second order differencing was executed to obtain a stationary condition. After stationary condition has been accomplished, the parameters of the ARIMA model must be identified. ARIMA model has three parameters which are autoregressive (p), differencing (d) and moving average (q), (Chintalapudi, Battineni, & Amenta, 2020). Autocorrelation function (ACF) and partial correlation function (PACF) was plotted to identify autoregressive (p) and moving average (q) parameter. If the ACF decayed exponentially and PACF has spiked, the process was an Autoregressive (AR) model. It was then identified as AR (p) where p is the number of spikes in the PACF. The Moving Average (MA) was best used when PACF decayed and the ACF has spikes. The value of parameter q is equal to the number of significant spikes in ACF. The parameter d refers to the order of differencing required by the time series to get stationary.

Model Estimation and Accuracy

It was worth considering a few potential models to limit the possibility of not picking the most suitable model structure in the model identification stage. To choose the best ARIMA model, some common statistical measures were applied. The statistical measures used to validate the best ARIMA models were the Akaike's Information Criteria (AIC), the Bayesian Information Criteria (BIC), Root Mean Square Error (RMSE) and Ljung-Box Statistic.

a) Akaike's Information Criteria (AIC)

The Akaike information criterion (AIC) is a mathematical method for assessing how efficient a model fits the data from which it was produced. AIC was implemented to compare distinct models and discover which one is the best fit for the data (Bevans, 2020). The AIC equation is given as

$$AIC = e^{2K/T} \sum_{t=1}^T e_t^2/T \quad (3)$$

where $k = p + q + P + Q$ depict the number of parameters estimated in the model, for p and q the usual respective terms of the AR and MA parts the P and Q the seasonality part of the ARIMA model and T is the total number of observations in the data series.

The test aimed to minimize the value of AIC by choosing the right p , q , P and Q . For instance, a model was deemed to be having a better fit than other models if the value of the AIC was the lowest.

b) Bayesian Information Criteria (BIC)

The BIC was developed by Gideon E. Schwarz and published in a 1978, where Bayesian argument for adopting it. The BIC aimed to choose a model that achieves the most accurate out-of-sample forecast by stabilizing between the models' complexity and goodness of fit (Hyndman, 2018). The BIC was calculated as

$$BIC = T^{k/T} \sum_{t=1}^T e_t^2 / T \quad (4)$$

where k is the number of parameters in the estimated model including the constant and T is the number of observations. The BIC was linked to AIC and one of the similarities was that the lower the value of BIC, the model was said to be the best ARIMA model.

c) Root Mean Square Error

The most used measure of forecast accuracy is the Root Mean Squared Error (RMSE) defined in equation (5).

$$RMSE = \sqrt{\frac{\sum_t^n e_t^2}{n}} \quad (5)$$

The results produced by any ARIMA model are measured in terms of forecast accuracy. Therefore, this study is considering the values of AIC, BIC and RMSE.

d) Ljung-Box Statistic

Examining the existence of correlation among the residuals, which is done by computing the Chi-Squared value of the error terms, is a straightforward approach for testing the misspecification. The portmanteau test is a frequent name for this type of method. Therefore, in addition to labelling this process as 'model validation', forecasters and modellers sometimes name it 'test for mis-specification'. The Box-Pierce Q statistic is written as follows:

$$Q = (T - d) \sum_{k=1}^h r_k^2 \quad (6)$$

where T is the number of observations throughout time series, h is the maximum lags of been tested, r_k is the residual term's k^{th} sample autocorrelation, p is the number of AR terms and q is the number of MA terms. The hypotheses to evaluate the Ljung-Box test can be written as,

H_0 : The errors are random (White noise)

H_1 : The errors are non-random (Not white noise)

If the p-value is less than 0.05, H_0 is rejected (the model is considered as mis-specified), then the best course of action is to try other model variations. Therefore, when choosing the best model under such condition, one must supplement the decision criteria with other statistics such as AIC, BIC or RMSE (Lazim, 2011).

Model Application

In the third stage, models that have been estimated were evaluated by comparing the forecast performance of the estimated models against each other. Finally, the best model was selected based on the results of comparing their respective measures in which the model that produced the smallest value of error measure calculated based on the out-of-sample

forecasts. To achieve the objectives, R-software was used to present the ARIMA model as a platform to perform the forecast of COVID-19 confirmed cases in Malaysia.

Results and Discussions

The trend number of COVID-19 daily new cases result in forecasting daily COVID-19 cases. The analysis of result and discussion are as followed.

Trend Number of Daily New Cases of Covid-19 In Malaysia

From 15 March 2020 to 31 March 2021, the pattern of COVID-19 daily confirmed cases in Malaysia was shown in Figure 3.

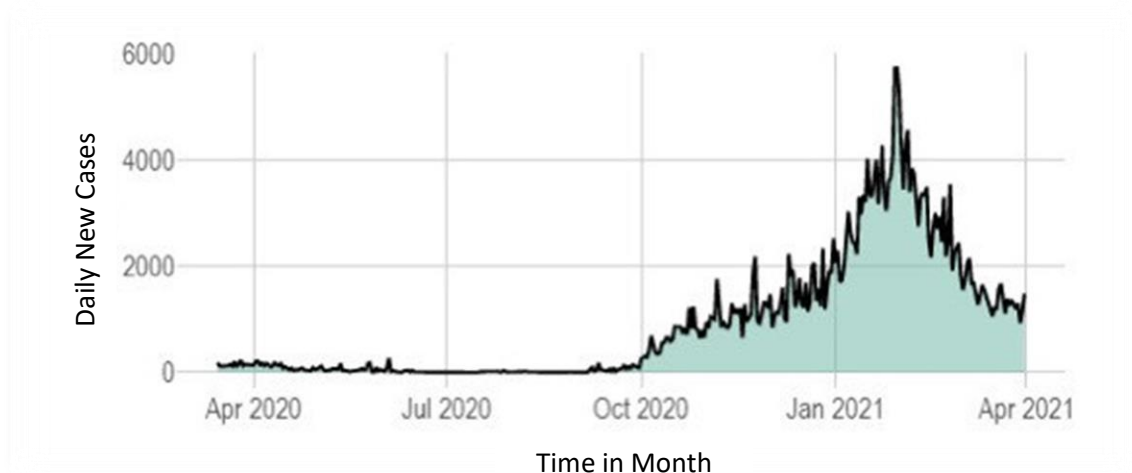


Figure 1 The Plot of Actual COVID-19 Daily New Cases in Malaysia

From the beginning, the number of incidents was quite high, with 190 cases on 15 March 2020. Following the declaration of the MCO, new cases managed to exhibit a consistent downward pattern, eventually reaching zero on 1 July 2020. However, the number of new cases steadily grew, starting on October 2020 with certain clusters had 260 new cases. Consequently, the number continues to climb again, reaching a high of 5728 new cases on 30 January 2021. Nevertheless, after several days of containment by the authorities, the cases began again to decrease until the end of March.

ARIMA Modelling Forecasting

The non-stationary conditions are illustrated in Figure 1, along with a trend. On the other hand, the slow decline in the values of the autocorrelation function (ACF) of the actual series, such as in Figure 2 indicates that the series of COVID-19 daily new cases in Malaysia is not stationary. Besides, Figure 2 shows the partial autocorrelation function (PACF) depicts there is a significant strong spike at lag 1 followed by the subsequent lags. Differencing is needed to render the series stationary because it does not fit the stationary required properties for the next stages.

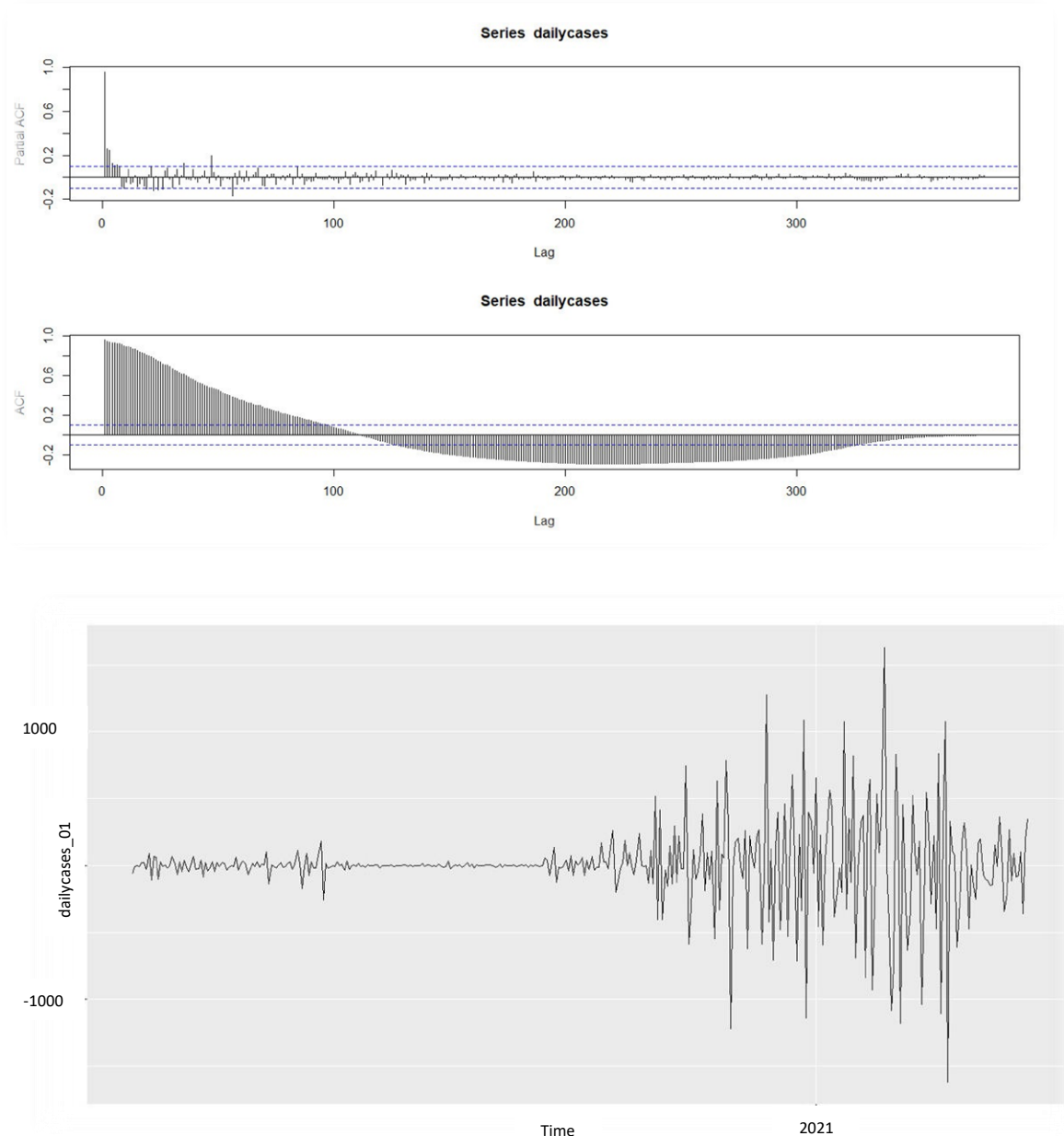


Figure 2: ACF and PACF of Covid-19 New Daily Cases in Malaysia

Figure 3: The Plot of COVID-19 Daily Cases in Malaysia After First Difference

Figure 3 shows the series was stationary and had a constant mean after differencing. The series fluctuates randomly about the fixed values, as can be observed, and hence is said to be stationary. Furthermore, as the data set at lag 7, the Augmented Dicky Fuller test (ADF) for

the first difference revealed that the series was stationary (ADF test statistic = -7.5737, $p = 0.01$), since the probability value is less than 0.05.

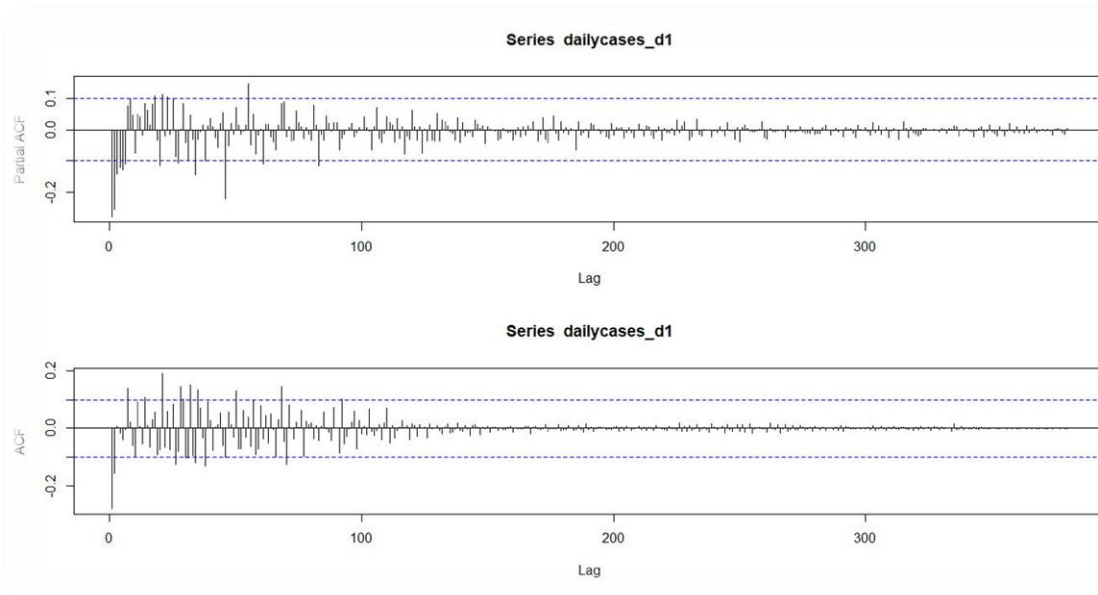


Figure 4: ACF and PACF of First Difference

After conducting the first difference, the ACF and PACF of the same series presented in Figure 4. In both cases, there is no discernible change in the trends. Spikes across multiple lags were observed for the ACF, reinforcing the previous assumption about the non-stationarity series. The rate of decay is much faster in which the values of the autocorrelation change from positive to negative at lag 1. The PACF of the series after the first difference, on the other hand, demonstrated a decaying pattern from lag 1 to lag 6. The PACF also exhibits many spikes, the most prominent of which occurs at lag one.

Utilizing different combinations of independent covariates, the following five models have been identified and estimated. ARIMA (1,1,1), ARIMA (1,1,2), ARIMA (1,1,3), ARIMA (2,1,1) and ARIMA (2,1,2) were chosen. Figure 4 indicates of correlograms ACF and PACF of first difference, to be matched for the best ARIMA model. The five models were chosen based on the number of significant spikes. However, as can be seen from the graph, there are many spikes that are significant. Thus, only the first five of the simplest models were selected from all existing significant spikes in the correlogram. Besides, to choose the best ARIMA model, the values of AIC, BIC, RMSE and Ljung-Box test statistic were compared. As for AIC, BIC and RMSE values, the lowest values were selected. As for the Ljung-Box test statistic, the values of p -value for every model were analysed to check if autocorrelation exists in a time series.

The null hypotheses of the Ljung-Box test assume that the residuals obtained have the property of white noise, and therefore, it is expected that no significant serial correlation exists and meets the stationarity condition. In addition, a model indicates to have no serial correlation if the probability value of Ljung-Box test is more than 0.05. Hence, a model might be chosen as the best model since there is no serial correlation exists while at the same time considering other factors. The following table shows the AIC, BIC, RMSE and p -values of Ljung-Box test for the chosen model:

Table 1: Summary of ARIMA Models

Model	AIC	BIC	RMSE	Ljung-Box (p-value)
ARIMA (1,1,1)	3778.43	3789.4	176.46	0.0437
ARIMA (1,1,2)	3779.26	3793.88	176.09	0.0818
ARIMA (1,1,3)	3762.45	3780.73	170.03	0.4087
ARIMA (2,1,1)	3777.11	3791.74	175.41	0.1711
ARIMA (2,1,2)	3762.96	3781.24	170.20	0.3143

Table 1 shows a comparison of accuracy results between the five ARIMA models with different value for parameter p , d , q for confirmed cases. Based on the p -value of the Ljung-Box test, all probability values of the selected models are more than 0.05 except for ARIMA (1,1,1) which is 0.0437. ARIMA (1,1,3) produced the p -value of 0.4087 which is more than 0.05 that has led to the null hypothesis to be accepted. Hence, the four models are well specified and adequate. ARIMA (1,1,3) was proposed as the best model because it has the lowest AIC, BIC and RMSE. In addition, even though values of ARIMA (1,1,3) and ARIMA (2,1,2) are not far apart, however, by utilizing the principle of parsimony where the simplest possible model should be chosen, it was decided that ARIMA (1,1,3) is the most appropriate model to be selected as best fit model and suit for forecasting purposes. Accordingly, there is no serial correlation in the model. Therefore, model ARIMA (1,1,3) can be proposed for forecasting the daily confirmed cases of COVID-19 in Malaysia.

Forecast the Number of Covid-19 Confirm Cases

By referring to Table 1, ARIMA (1,1,3) model was proposed as the best model in forecasting daily COVID-19 confirmed cases. Therefore, the model was used to forecast the data from April 2021 to January 2022. The future trend forecast is depicted in Figure 5, which appears to be concerning at a 95 percent confidence interval.

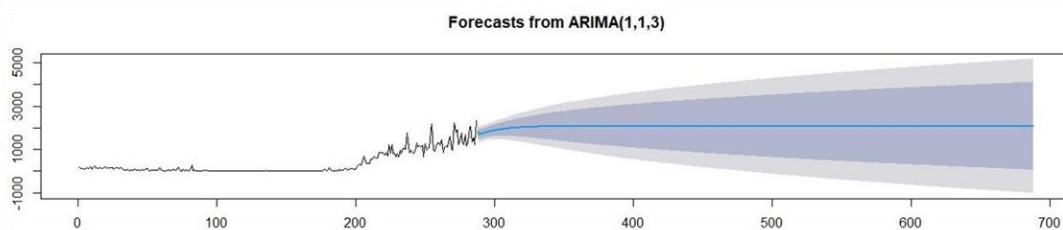


Figure 5: Forecast value of COVID-19 Daily Cases in Malaysia

Figure 5 displayed the forecasted daily new cases by using the evaluated data from day 288 and above are shows the good outcomes to be in steady for the next 401 days (about 1 year) until 31 January 2022. The horizontal axis stands for the number of days from 15 March 2020 and onward while the vertical axis represents the daily cases. The number of daily new cases continues to rise steadily at an average of 0.00021 percent each day up until the first month of 2022. Even though daily new cases are growing, the positive news is that the rate of growth is not sudden. Thereby, authorities should be able to remedy the circumstance to avoid a recurrence of events, albeit slowly.

Conclusion

The number of COVID-19 daily confirmed cases starting from 15 March 2020 shows an inconsistent pattern with the cases reaching their highest peak in January 2021. There was also a decline in the new cases afterwards, but it was still in an alarming situation because the daily cases were still at 4-digit numbers every day. There are several ways taken by the government to control the spread of the virus such as the implementation of MCO and the emphasis on Standard Operating Procedure (SOP) in the community.

Next, this study demonstrated the effectiveness of ARIMA models as an early warning strategy that can provide accurate COVID-19 forecast despite limited data points. The ARIMA model is not only effective, but it is a simple and easy method by which COVID-19 trends can be predicted based on free access data. Using the selected ARIMA model which is ARIMA (1,1,3), the forecasted values show that the daily new cases still increase steadily from April 2021 up until January 2022. Nevertheless, the increasing values do not increase abruptly, which means that the authorities still can take the required action to contain and control the confirmed new cases in the coming days.

In addition, this research gives a clear picture regarding the importance of the roles of the government and the authorities to make sure that COVID-19 handling is efficient and systematic. Also, it is extremely crucial to ensure that the same precaution steps taken will not be repeated continuously or else, many parties will be affected by this. For instance, after 4 June 2020, Malaysia successfully turns down the number of new cases below 50 until one point. In this situation, the responsibilities to tackle the COVID-19 cases should be done by all Malaysians regardless of social strata. Hence, even though it was a tough journey, nonetheless, with this mature mindset, all steps that have been taken to reduce the case can be run smoothly and successfully.

Acknowledgment

This research was supported by Universiti Teknologi MARA (UiTM) and funded under UiTM internal Grant No. 600-RMC/LESTARI COVID/5/3 (042/2020). This support is gratefully acknowledged. The authors would like to thank other researchers, lecturers and friends for ideas and discussion.

References

- Alabdulrazzaq, H., Alenezi, M. N., Rawajfih, Y., Alghannam, B. A., Al-Hassan, A. A., & Al-Anzi, F. S. (2021). On the accuracy of Arima based prediction of COVID-19 spread. *Results in Physics*, 27, 104509. <https://doi.org/10.1016/j.rinp.2021.104509>
- BERNAMA. (2020). *Malaysia recognised as one of most successful in handling covid-19*.
- Bevans, R. (2020). *Akaike information criterion: When amp; how to use it*. Retrieved September 18, 2021, from <https://www.scribbr.com/statistics/akaike-information-criterion/>.
- Chintalapudi, N., Battineni, G., & Amenta, F. (2020). Covid-19 virus outbreak forecasting of registered and recovered cases after sixty-day lockdown in Italy: A data driven model approach. *Journal of Microbiology, Immunology and Infection*, 54(3), 396-403. <https://doi.org/10.1016/j.jmii.2020.04.004>
- Elengoe, A. (2020). Covid-19 outbreak in Malaysia. *Osong Public Health and Research Perspectives*, 11(3), 93.
- Gupta, R., & Pal, S. K. (2020). Trend analysis and forecasting of covid-19 outbreak in India. *MedRxiv*. <https://doi.org/10.1101/2020.03.26.20044511>

- Hyndman, R. J. (2018). *Forecasting: Principles and practice, 2nd edition* (G. Athanasopoulos, Ed.). OTexts. Retrieved 2020-12-29, from <https://otexts.com/fpp2/stationarity.html#fn14>.
- Khan, F. M., & Gupta, R. (2020). Arima and Nar based prediction model for time series analysis of covid-19 cases in India. *Journal of Safety Science and Resilience, 1*(1), 12–18.
- Khounpaseuth, S. (2020). *Ministry of health and who respond to first case of covid-19 in Laos*.
- Konarasinghe, K. M. U. B. (2020a). Forecasting covid-19 spread in Malaysia, Thailand, and Singapore. doi: 10.5281/ZENODO.3935923
- Konarasinghe, K. M. U. B. (2020b). Modeling covid-19 epidemic of USA, UK and Russia. *Journal of New Frontiers in Healthcare and Biological Sciences, 1*(1), 1–14.
- Kufel, T. (2020). Arima-based forecasting of the dynamics of confirmed covid-19 cases for selected European countries. *Equilibrium, 15*, 181-204. doi: 10.24136/eq.2020.009.
- Lazim, M. A. (2011). *Introductory business forecasting: A practical approach 3rd edition* UiTM Press.
- Palachy, S. (2019). *Stationarity in time series analysis - towards data science*. Towards Data Science. Retrieved 2020-12-29, from <https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322>.
- Sharma, V. K., & Nigam, U. (2020). Modeling and forecasting of covid-19 growth curve in India. *Transactions of the Indian National Academy of Engineering, 5*(4), 697–710.
- Singh, S., Sundram, B. M., Rajendran, K., Law, K. B., Aris, T., Ibrahim, H., . . . Gill, B. S. (2020). Forecasting daily confirmed covid-19 cases in Malaysia using Arima models. *The Journal of Infection in Developing Countries, 14*(09), 971–976.
- Tee, K. (2020). *Dr Noor Hisham: Malaysia records zero local covid-19 transmissions for the second time, three imported cases*.
- WHO. (2020). *Universal health coverage and covid-19 preparedness response in Malaysia*.
- Yonar, H., Yonar, A., Tekindal, M. A., & Tekindal, M. (2020). Modeling and forecasting for the number of cases of the covid-19 pandemic with the curve estimation models, the Box-Jenkins and exponential smoothing methods, *EJMO, 4*(2), 160-165.