FORECASTING COVID-19 CASES IN SELANGOR USING BOX-JENKINS METHOD

Norliana Mohd.Lip[1*], Siti Sarah Mohd. Nazir[2], Sharifah Wan Nurazreen Sayed Azri[3] & Puteri Nursakinah Mior Shahridan[4]

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban, Persiaran Seremban Tiga/1, 70300 Seremban, Negeri Sembilan.

*corresponding author: norliana7287@uitm.edu.my

Keywords: ARIMA, Box-Jenkins, COVID-19, forecasting, SARIMA

1. Introduction

An outbreak can be defined as when a large number of people become sick or have the same disease simultaneously, and it could remain in one area or spread out to other areas and become more extensive in the chain of the disease which can last in a few days or several years. While pandemic is a disease epidemic that affects several countries or continents simultaneously. It has an enormous impact and kills more people than an outbreak. When it became apparent that the disease was severe and spreading rapidly across a large region, the World Health Organization (WHO) declared COVID-19 a pandemic (Robinson, 2020).

People are currently amid an ongoing pandemic caused by the new strain of a novel coronavirus or COVID-19. It was reported at the end of 2019 by World Health Organizations and declared a global health emergency. The COVID-19's word had the meaning where "CO" stands for corona, 'VI' for virus, 'D' for disease, and 19 refer to the year the virus had been recognized. The virus began spreading in 2019, in Wuhan, Hubei, China (Ducharme, 2020).

Despite the inaccuracies of medical predictions, forecasting is still helpful for better understanding the present situation and planning for the future (Petropoulos & Makridakis, 2020). Therefore, this study conducted on the COVID-19 virus in Malaysia to predict future positives, recoveries, and deaths cases. It focused on cases in Selangor only because Selangor continues to mark the country's highest daily cases of COVID-19 (Ying, 2020). The Box-Jenkins method was used because it is one of the suitable methods for identifying, adapting, checking, and utilizing the result.

2. Methodology

The monthly data of positive, recovery and death COVID-19 cases (September 2020 - April 2021) obtained from the Ministry of Health Malaysia were used to forecast the COVID-19 cases in

Selangor. There are 5 main stages in the Box-Jenkins Method which involve data processing, model selection, parameter estimation, model checking and forecasting. The general models used in this study are:

ARIMA Model

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \quad (1)$$

SARIMA Model:

$$\phi(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D y_t = \delta + \Theta(B)\Theta_Q(B^s)\varepsilon_t \quad (2)$$

Where,

μ refers to constant.
$\phi_1 y_{t-1} + \ldots + \phi_p y_{t-p}$ refers to AR.
$\theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$ refers to MA.
ε_t refers to Gaussian white noise.
$\phi(B)$ refers to an ordinary autoregressive component.
$\Theta(B)$ refers to an ordinary moving average component.
$\Theta_Q(B^s)$ refers to the seasonal autoregressive components.
$\Phi_p(B^s)$ refers to the seasonal moving average components.
$(1-B)^d$ refers to an ordinary difference component
$(1-B^s)^D$ refers to the seasonal difference component.

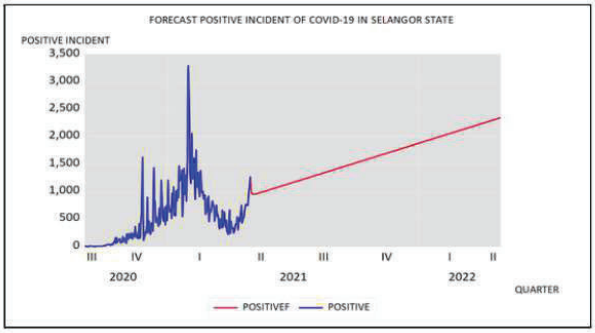
3. Results and Discussions

Durbin Watson (DW), Akaike Information Criterion (AIC) and Schwarz Criterion (BIC) were used to validate the appropriate ARIMA and SARIMA models. Table 1 shows the values of DW, AIC and BIC of the appropriate ARIMA and SARIMA models for three different cases. The future values for positive, recovery, and death cases in May 2021 until April 2022 were forecasted based on the best selected model for each case. Figure 1 illustrates the graphical representation of positive, recovery and death cases in May 2021 until April 2022.

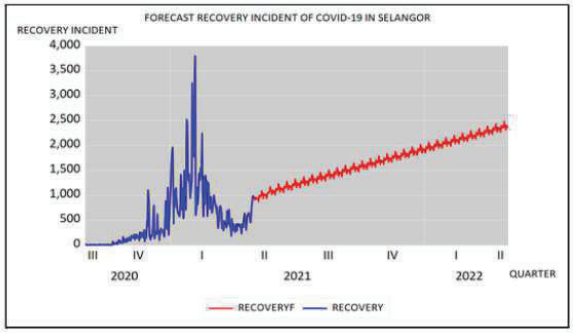
Table 1 The values of DW, AIC and BIC

Cases	Models	DW	AIC	BIC
Positive	ARIMA (1,1,1)	1.9146	14.0639	14.1218
Recovery	SARIMA $(1,0,3)(2,1,1)_{12}$	2.1553	14.8622	14.9519
Death	SARIMA $(1,0,1)(1,1,1)_{12}$	2.0147	4.3571	4.4468

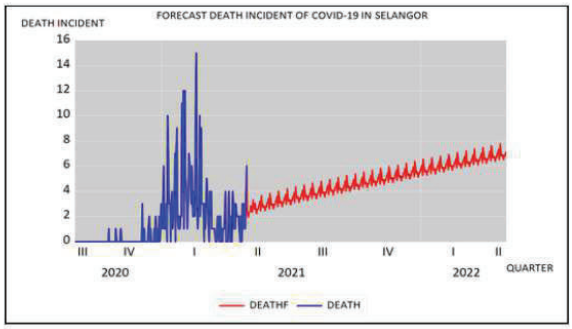
Figure 1 The graph of future values for (a) positive, (b) recovery and (c) death cases



(a)



(b)



(c)

4. Conclusion

In conclusion, this study can be used as a reference for future researchers on other severe and dangerous diseases. This is because this infectious disease will affect the human population and affect daily human life, and will disrupt the economic, transportation, and education sector. This study indicated that the Box Jenkins method was an easy-to-use modeling method for rapidly forecasting the spread of COVID-19 in Selangor. In addition, we recommend using another approach such as Artificial Neural Network. The development of ANN algorithm-generated COVID-19 outbreak curves are very close to the actual cases, and this method also attained future forecasts for both infections and deaths with high correlation coefficients with respect to reported cases in all studied countries.

References

Ducharme, J. (2020). *The World Health Organization Declares COVID-19 a 'Pandemic.' Here's What That Means.* Retrieved April 3, 2021 from TIME: https://time.com/5791661/who-coronavirus-pandemic-declaration/

Petropoulos, F., & Makridakis, S. (2020). *Forecasting the novel coronavirus COVID-19.* Retrieved April 4, 2021 from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231236

Robinson, J. (2020). *Pandemics.* Retrieved April 11, 2021 from WebMD: https://www.webmd.com/cold-and-flu/what-are-epidemics-pandemics-outbreaks

Ying, T. P. (2020). *COVID-19: Selangor records highest daily cases for 8th day running.* Retrieved April 2, 2021 from New Straits Times: https://www.nst.com.my/news/nation/2020/12/651771/COVID-19-selangor-records-highest-daily-cases-8th-day-running