Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 9, 2022, page 157-174

Long-Term Dependent Analysis on Covid-19 Vaccination Using Log-Linear Poisson Autoregressive Model

Maryam Adibah Hanim Mohd Hazni, Norhaiza Ahmad*
Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia
*Corresponding author: norhaiza@utm.my

Abstract
The aim of this study is to apply the log-linear Poisson autoregressive in modelling the number of vaccinations of COVID-19 for different states in Malaysia from 24 February 2021 until 21 January 2022. The log-linear Poisson autoregressive model is suitable to model such long-term dependent time series count data and is known to handle overdispersion well. In the analysis, the vaccination data is described based on the calculation of the 7 days moving average, accumulated number of vaccinated and accumulated number of vaccination rate. The index of dispersion is also calculated in order to distinguish the true variation between the states. The number of daily vaccinations of coronavirus disease (COVID-19) in each state are found to be overdispersed. The model is estimated using the data of the number of daily vaccinations by comparing the parameters between the states. The model is validated between the value of the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Quasilikelihood under the Independence model Criterion (QIC). The model's performance is then evaluated using the sequence of cumulative Pearson residuals to check the adequacy of the model in each state in Malaysia. Results shows that the model is adequate in modelling the number of daily vaccinations in all states in Malaysia except for Perlis, Pulau Pinang, Negeri Sembilan, Melaka, Kelantan, Pahang, Terengganu, W.P. Putrajaya, and W.P. Labuan. The model is then applied to predict the number of daily vaccinations in March 2022 for the states in Malaysia. The model is capable of determine the COVID-19 trend and forecasting the number of daily vaccinations. Aside from that, the model's accuracy may be hampered by the significant and unpredictable quantity of daily vaccinations.

Keywords: Poisson autoregressive model; COVID-19; Vaccination; Count time series; Malaysia

1. Introduction

The coronavirus disease (COVID-19) pandemic produced by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused widespread devastation, particularly in Malaysia.

The comparing of statistics across states in Malaysia is needed to assess the scope and rate of vaccination distribution. In Malaysia, the government is promoting vaccination. Nevertheless, different states may have different rates. In Malaysia, each state has a different phase based on their total daily vaccination. Vaccination distribution in Malaysia relies on the availability of vaccines in the country. For example, all schools and educational institutions in the state that has already accomplished phase three are permitted to operate face to face with a fifty percent rotation of class capacity, whereas all schools, educational institutions, and private educational institutions in the state that has already accomplished phase four are permitted to reopen with the gradual student admission in three stages. The GitHub MOH Malaysia vaccination dataset is a public collected global dataset on vaccinations provided. It includes the entire period beginning on February 24, 2021, the date the first vaccination data were announced and has been constantly updated since then. The number of vaccinations between the states appear to be vary.

However, the variation between the trends in Malaysia tends to be overdispersed. One model that can handle the count data is Poisson regressive model but the equidispersed assumption must be achieved to apply this model. Therefore, a log-linear Poisson autoregressive model, which has the potential to determine the trend should be applied.

2. Literature Review

2.1. Time series analysis and hybrid models for COVID-19

2.1.1. Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) is an autoregressive statistical model that can predict future results based on past results. The ARIMA proposed model consists of three phases: model identification, parameter estimation, and model diagnostic checking [9]. Alternative mathematical and classical forecasting modelling techniques applied to COVID-19 data contradict the suggested hybrid approach [2].

2.1.2. Seasonal Autoregressive Integrated Moving Average (SARIMA)

Another method for calculating COVID-19 is to use the seasonal ARIMA forecasting package (SARIMA) in conjunction with the R statistical model. SARIMA, also known as Seasonal ARIMA, is a modification of ARIMA that explicitly handles univariate time series data with a seasonal component. ARIMA models can predict simple ups and downs and are more accurate than regressive models without affecting the general trend [3]. This is because ARIMA can only reflect at the data from dependent variables such as registered and recovered cases.

2.2. Compartmental models for COVID-19

2.2.1. Susceptible Infected Recovered (SIR) and Susceptible Exposed Infected Recovered (SEIR)

Susceptible Infected Recovered (SIR) is one of the basic models while Susceptible Exposed Infected Recovered (SEIR) is a derivative from the basic SIR model. Implying parameters of compartmental models such as susceptible-infectious-recovered (SIR) and susceptible-exposed infectious-recovered (SEIR), which are extensively employed in infectious disease predictions, is a basic problem when validating local models [8].

2.2.2. Stochastic Transmission Models

The model explicitly simulated a Poisson observed procedure of newly symptomatic cases, reported onsets of new cases, reported confirmation of cases, and a binomial observation method for infection frequency on evacuation flights to account for uncertainty in case observed [7].

2.3. Model for COVID-19 Probability Distribution (Handle Overdispersion)

2.3.1. Poisson Regression Model

The Poisson regression model is used to describe count data that is influenced by several known predictor variables [5]. The occurrence of excess zero data and lengthy right tails, both related to the Poisson assumption, is a property of many count data. Overdispersion in the data could explain for both phenomena. The additional zeros can be caused by grouping. Overdispersion tends to increase the fraction of zeros, and when there are too many zeros related to the Poisson assumption, a modified Poisson regression is suitable for the data fit. An overdispersed model with extra zeros should be used as an alternative for a better match.

Past researchers that using Poisson regression model stated that the main studies included vaccine effect estimates for total vaccine coverage and for each vaccine category [10]. One dose of each vaccination affected is examined to the hospital admissions related to laboratory-confirmed SARS-CoV-2 infection or diagnostic testing of COVID-19.

2.3.2. Zero-Inflated Negative Binomial (ZINB) model

Zero-Inflated Negative Binomial (ZINB) model is for count data with overdispersion and extra zeros. A zero-inflated negative binomial multivariate regressions is utilised to design the affiliations of time [4]. The past researchers wanted to see if there was a link between having a vaccine clinic including both resident and employee COVID-19 cases and deaths. A zero-inflated negative binomial is used to mixed effect regressions to handle the heavily distorted, continuous count measures with a high proportion of

zeros. The negative binomial model handles overdispersion by inserting a dispersion parameter that modifies the Poisson model's premise of equal mean and variance.

2.4. Poisson Autoregression model

Next, Agosto and Giudici [1] offer a mathematical model that can predict when the maximum of infection will occur, allowing preventive actions to be implemented and relaxed. The statistical distribution of new cases is presumed at time (day) t, conditional on the information up to $t-1$ is Poisson, with a log-linear autoregressive intensity, as follows :

$$y_t|F_{t-1} \sim Poisson(\lambda_t)$$
$$\log(\lambda_t) = \omega + \alpha \log(1 + y_{t-1}) + \beta \log(\lambda_{t-1}), \quad (1)$$

where,

- F_{t-1} denoted the α-field generated by $\{y_0,..., y_t\}$,
- ω is the intercept term,
- α and β express the dependence of the expected number of new infections,
- λ_t on the past counts of new infections.

The α component illustrates the short-term dependence on the prior time point. The β component represents a trend component, that is, the observed process's long-term dependence on all previous values [1]. The model is applicable to every country, location, or period. The application is demonstrated without sacrificing generality by utilizing data that is readily available.

3. Methodology

3.1. Log-linear Poisson autoregressive model

Count time series modelling has come a long way in the previous two decades. Let $Y = (Y_t)_{t\in z}$ be an integer-valued time series, with $F_t = (Y_s, s \leq t)$ indicating the σ-field produced by the entire past at time *t* as well as $L(\frac{Y_t}{F_{t-1}})$ indicating the conditional distribution of Y_t based on the past. The type of marginal distribution $L(\frac{Y_t}{F_{t-1}})$ and the dependence structure between $L(\frac{Y_t}{F_{t-1}})$ and the previous define a model. Models with a variety of marginal distributions and dependent architectures have been investigated. Consider the Poisson autoregression, where $L(\frac{Y_t}{F_{t-1}})$ is Poisson distributed with an intensity λ_t that varies as a function of λ_{t-1} and Y_{t-1}. They demonstrated the coherence and approximate normality of the maximum likelihood estimator of the regression parameter under linear autoregression by employing a perturbation strategy that allowed researchers to use the conventional Markov ergodic setup. The technique was expanded to nonlinear Poisson autoregression using $Y_t = f(\lambda_{t-1}) + b(Y_{t-1})$ for nonlinear quantifiable functions *f* and *b*.

Consider the following time series of counts $Y = (Y_t)_{t\in z}$ fulfilling :

$$\frac{Y_t}{F_{t-1}} \sim \text{Poisson}(\lambda_t) \text{ with } \lambda_t = F(\lambda_{t-1}, Y_{t-1}), (\lambda_{t-2}, Y_{t-2}) \dots) \quad (2)$$

where $F_t = \sigma(Y_s, s \leq t)$ and *F* is a quantifiable non-negative function. Next, we prove that the conditional mean λ_t can be represented as a function of previous observations under certain Lipschitz-type conditions on *F*. This leads to the model being considered as :

$$\frac{Y_t}{F_{t-1}} \sim \text{Poisson}(\lambda_t) \text{ with } \lambda_t = f(Y_{t-1}, Y_{t-2} \dots) \tag{3}$$

where *f* is a quantifiable non-negative function. After that, we presume that *f* is known up to a parameter θ_0, which belongs to a compact set $\Theta \subset \mathbb{R}^d$ with $d \in \mathbb{N} - \{0\}$. The equation is as follows :

$$\frac{Y_t}{F_{t-1}} \sim \text{Poisson}(\lambda_t) \text{ with } \lambda_t = f_{\theta_0}(Y_{t-1}, Y_{t-2} \dots) \text{ and } \theta_0 \in \Theta \tag{4}$$

Poisson processes can be used to illustrate equation (3.1), equation (3.2) and equation (3.3). We let $\{N_t(.)\,; t = 1,2, \dots\}$ be a sequence of independent Poisson processes of unit intensity. Y_t can be thought of as the number (say, $N_t(\lambda_t)$) of events of $N_t(.)$ occurrences in the period [0, λ_t]. Our equation now has a new representation as below :

$$Y_t = N_t(\lambda_t) \text{ with } \lambda_t = f_{\theta_0}(Y_{t-1}, Y_{t-2} \dots) \tag{5}$$

The Poisson autoregressive models are known to represent the overdispersion phenomena in counts data, which means that if the process $(Y_t)_{t \in z}$ is static, $\text{Var}(Y_t) \geq E(Y_t)$ usually happens. The approach model will be included in R to generate the GitHub data parameters. Let $\{Y_t\}$ be the time series of the number of daily vaccinations COVID-19, where $t \geq 0$ represents the time. The log-linear Poisson autoregressive model assumes the number of daily vaccinations follows the Poisson distribution with mean λ_t.

$$Y_t \sim Poisson(\lambda_t), \qquad In\lambda_t = \omega + \alpha In(1 + Y_{t-1}) + \beta In\lambda_{t-1} \tag{6}$$

Each parameter has their own meaning such as ω denotes the intercept term, α means the short-term dependence on yesterday's daily vaccinations and β expresses the long-term dependence on the daily vaccinations of the previous days. Studies shows that the trend prediction of each state can be evaluated by comparing the value of parameters between α and β. If $\alpha < \beta$, it indicates an increasing trend, while $\alpha > \beta$ indicates a decreasing trend.

3.2. Model Estimation and Validation

The model estimation can be done based on the values of ω, β, α, Akaike information criterion (AIC), Bayesian Information Criterion (BIC), Quasilikelihood under the Independence model Criterion (QIC) and the log-likelihood function. The log-linear Poisson autoregressive model is fitted to the data on each state from the period 24 February 2021 until 21 January 2022 is applied. The model equation is implemented using the *tsglm* package in R. The past observation and past mean used in this package is set to default.

Pearson residuals is also be used to check the model's adequacy. The cumulative periodogram of Pearson residuals shows the frequency of residuals in the relevant limits and it determines the model's adequacy. The dashed lines in the plot represent the approximate 95% confidence intervals on a continuous spectral density, which is used as a graphical check for uncorrelated residuals. Residuals

can be used to determine whether a model performs adequately. If there are correlations between residuals, then the residuals contain data should be employed in forecasting. This indicates that if the series of residuals is a sequence of white noise with constant variance, this model is adequate.

4. Results and discussion

4.1. About the vaccinated data

The data is taken from the Ministry of Health Malaysia's official website on GitHub[1]. The data covered the period from February 24, 2021, to January 21, 2022, with a total of 331 daily observations for all 16 states. In our study, 25,542,969 (78.2 %) adults aged 18 and above were vaccinated. The main analyses comprised vaccine effect estimates for overall total cumulative vaccination rate in states and for each vaccine type. The data is evaluated at how vaccine affected the shape of the graph in every states. Table 1 shows the variation between the states in Malaysia for various calculation with respect of vaccination starting from 24 February 2021 until 31 March 2022 before the period observed is cut into the present observed. This table contains (1) population size for all 16 states/ federal territories (PS). The population size considered the entire population of each state, including children under 18. (2) The calculated 7 days moving average until 31 March 2022 (CMA). (3) Accumulated number of people vaccinated (AV). (4) The accumulated number of vaccination rate (AVR) until 31 March 2022, this is calculated by taking the ratio AV against the population size.

Next, moving average statistics is used to describe the characteristics of every states. As shown in Figure 1 since the earliest date observed beginning on February 24, 2021, there are not many people who have taken the vaccination because COVID-19 vaccination had just been released at the time. However, as the government conducted a mass vaccination later, the number of daily vaccinations given in each state significantly grows after a month. The process proceeded until it began to decrease and then gradually grow to a reasonable number in the middle of the observation period. The observed trend came to a halt on the black line on 21 January 2022, because the graph shows that the active duration of the daily vaccination given after that date continued to decline until March 31, 2022. It shows that the number of daily vaccinations taken was getting smaller. Therefore, the data used to develop the models was changed from ended until March 31, 2022, to ended until January 21, 2022, since there were no obvious changes.

The characteristics number of vaccinated in East Malaysia are analysed in the graph with the new period from 24 February 2021 until 21 January 2022. The charts below include two lines which is an orange line and a purple line. The orange line represents the actual number of total daily vaccinations observed during the period, whereas the purple line represents the moving average based on total daily vaccinations observed during the same period. The purple line on the charts, which represents one-week moving average.

During the period observed, Kementerian Kesihatan Malaysia (KKM) prescribed four types of vaccination in Malaysia. There are Pfizer, AstraZeneca, Sinovac and CanSino. For each vaccine, people are encouraged to take three times of vaccination. The third vaccination are generally known as booster to the public. Figure 8 and 9 shows the distribution of the number of people vaccinated based on four types of vaccination in Malaysia.

[1] *GitHub - MoH-Malaysia/covid19-public: Official data on the COVID-19 epidemic in Malaysia. Powered by CPRC, CPRC Hospital System, MKAK, and MySejahtera.* (2022, April 3). GitHub. Retrieved April 4, 2022, from https://github.com/MoH-Malaysia/covid19-public

Table 1 : The descriptive statistics among states in Malaysia

States	Johor	W.P. Labuan	W.P. Kuala Lumpur	Perak	Perlis
PS	6921600	176000	3328100	4770300	471200
CMA	5100.429	145.0000	2131.143	3044.143	255.7143
AV	7217350	195602	7331987	4387821	458785
AVR	1.04	1.11	2.2	0.92	0.97

States	Kedah	Kelantan	Melaka	Negeri Sembilan	Pahang
PS	3998200	3337000	1728600	2088200	3044700
CMA	2473.429	1339.5714	1099.8571	1534.286	2041.286
AV	3686777	2604184	1828654	2331665	2776150
AVR	0.92	0.78	1.06	1.12	0.91

States	Pulau Pinang	Sabah	Sarawak	Selangor	W.P. Putrajaya	Terengganu
PS	3380000	6905800	5192000	11861700	83700	2182900
CMA	2774.286	4932.286	2501.143	11573.857	313.4286	1190.7143
AV	3741825	5123025	5711741	12168312	357053	2025207
AVR	1.11	0.74	1.1	1.03	4.27	0.93

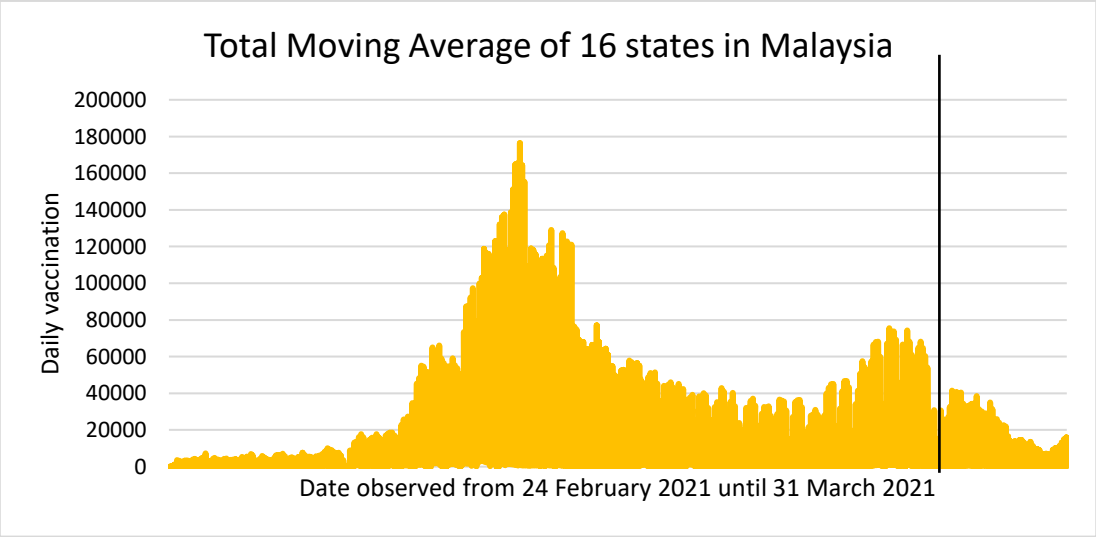


Figure 1 Total moving average daily vaccination of 16 states in Malaysia

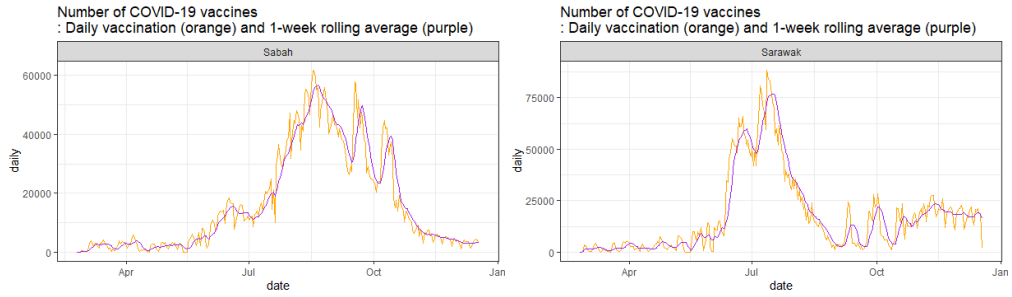


Figure 2 Daily vaccination and rolling moving average for 7 days in Sabah and Sarawak

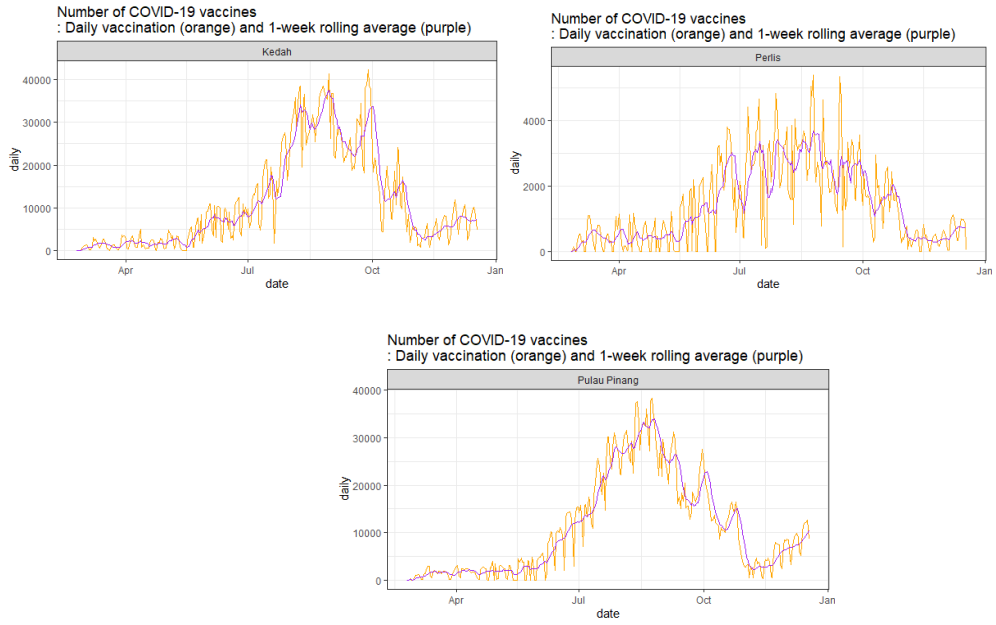


Figure 3 Daily vaccination and rolling moving average for 7 days in Kedah, Perlis and Pulau Pinang

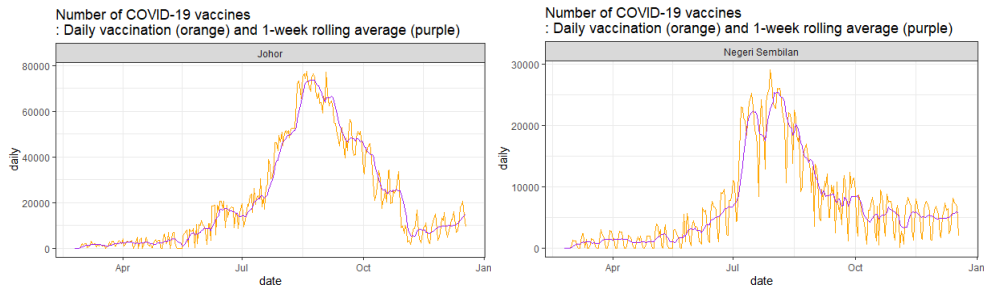


Figure 4 Daily vaccination and rolling moving average for 7 days in Johor and Negeri Sembilan

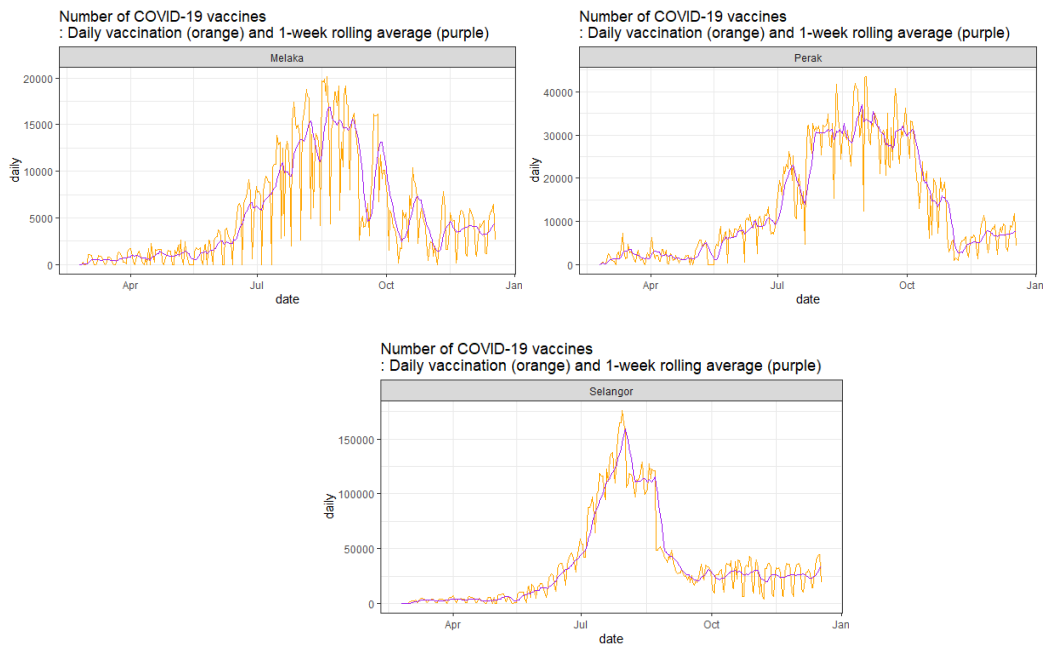


Figure 5 Daily vaccination and rolling moving average for 7 days in Melaka, Perak, and Selangor

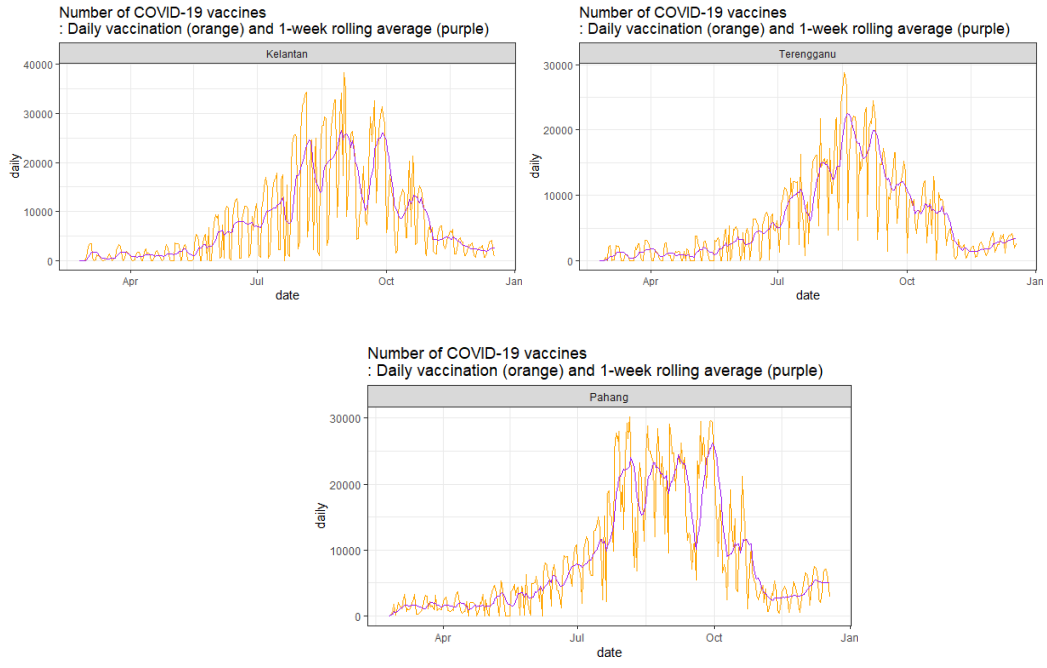


Figure 6 Daily vaccination and rolling moving average for 7 days in Kelantan, Pahang, and Terengganu

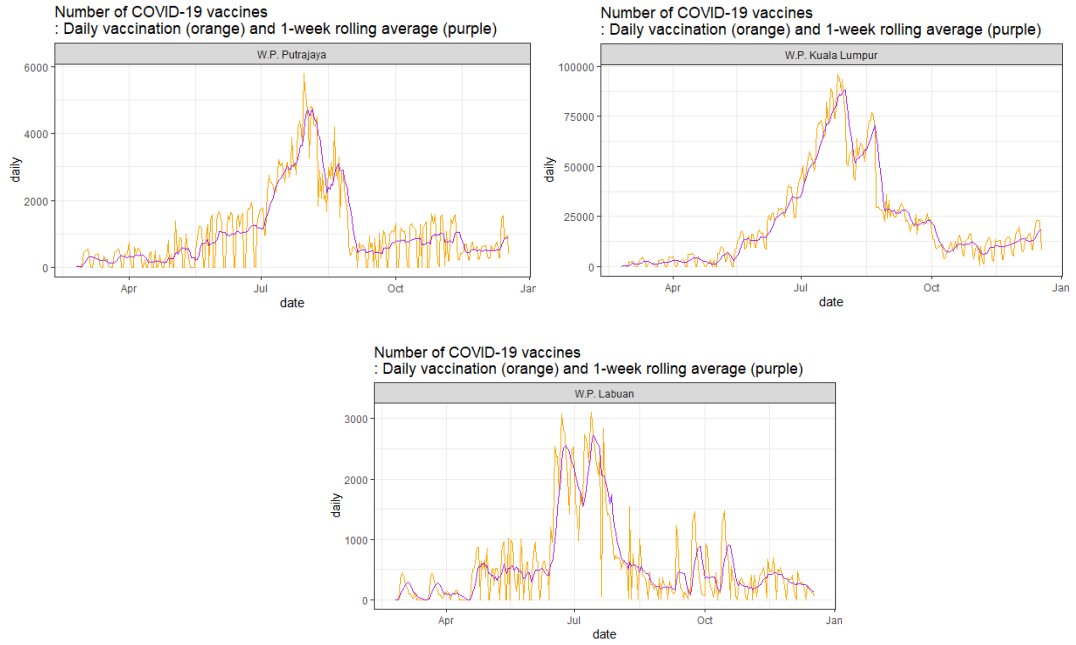


Figure 7 Daily vaccination and rolling moving average for 7 days in W.P. Putrajaya, W.P. Kuala Lumpur, and W.P. Labuan

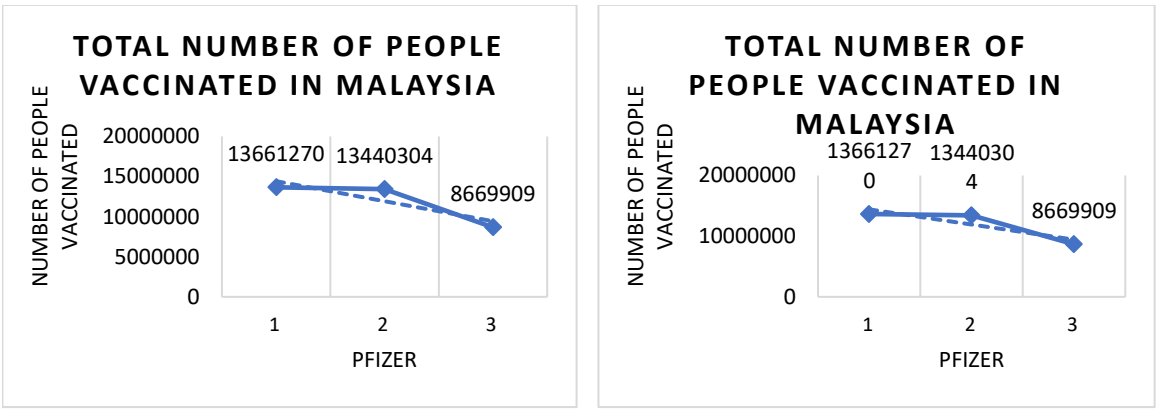


Figure 8 Statistics on daily dose uptake of Pfizer and Sinovac

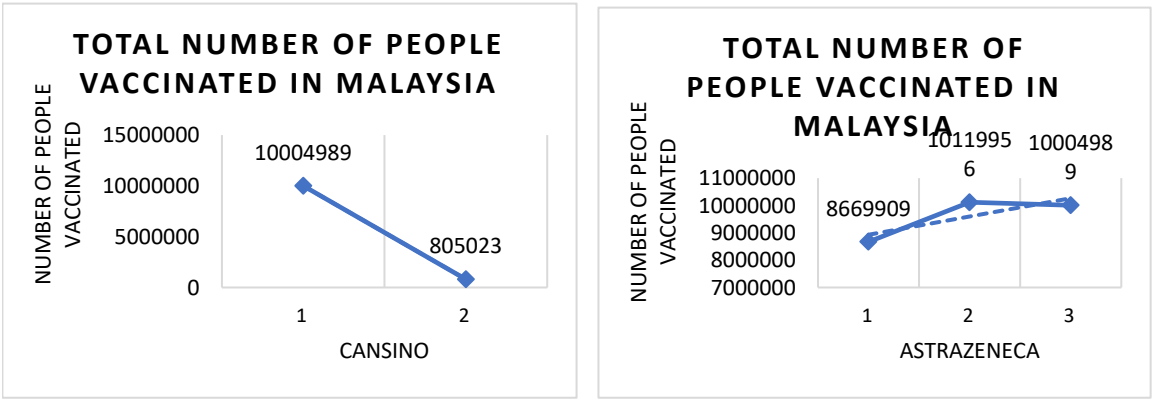


Figure 9 Statistics on daily dose uptake of CanSino and AstraZeneca

4.2. Data Description

Table 2 shows a descriptive statistic of the number of daily vaccinations in each state. The information in the table is sorted in descending order based on the biggest number of the mean, variance, and index of dispersion until the smallest number. If the index of dispersion, D = 1, then the data is equidispersed. Meanwhile, if $D > 1$, the data is overdispersed and the data is underdispersed if $0 \leq D < 1$. In this case, all states have big number of mean and variance that results to get bigger value index of dispersion. It is observed that the index of dispersion for all states is greater than 1. Clearly from the table, the number of vaccination data for all states are overdispersed. Therefore, a fitted model that does not rely on the assumption of equidispersion need to be used. So, one of the models that is suitable for this kind of data is log-linear Poisson autoregressive model to estimate the trend of the vaccination rate in each state.

Table 2 : Summary of descriptive statistics by state

States	Maximum daily vaccination	Mean	Variance	Index of Dispersion
Selangor	176320	36706.17	1481603042	40363.87
W.P. Kuala Lumpur	96120	22109.63	490682400	22193.15
Johor	77513	21739.00	446609740	20544.17
Sarawak	88457	17225.35	350754376	20362.68
Sabah	61924	15458.42	273532603	17694.73
Perak	43535	13236.67	136651226	10323.69
Pulau Pinang	38340	11280.40	100235512	8885.81
Kedah	42388	11117.13	118242481	10636.06
Pahang	30210	8380.23	66117706	7889.72
Kelantan	38335	7846.68	83997479	10704.85
Negeri Sembilan	29106	7035.32	48075902	6833.51
Terengganu	28818	6105.12	40391175	6615.95
Melaka	20158	5514.70	28215787	5116.47
Perlis	5384	1383.20	1569974	1135.03
W.P. Putrajaya	5805	1075.84	1290635	1199.66
W.P. Labuan	3110	590.95	503510	852.04

4.3. Model Estimation

In this section, log-linear Poisson autoregressive model is fitted to the data from the period 24 February 2021 until 21 January 2022 for each state is applied. Figure 10 shows the time series graph representing the number of daily vaccinations at the observed period. It is observed that there is state variation with respect to the number of daily vaccinations. Meanwhile, Table 3 shows the summary result for the estimated parameters in each state. The past observation and past mean used in this package is set to default. All the estimated parameters ranged between -1 until 800000. All the parameters between the states have different value and there exists a variation between these states.

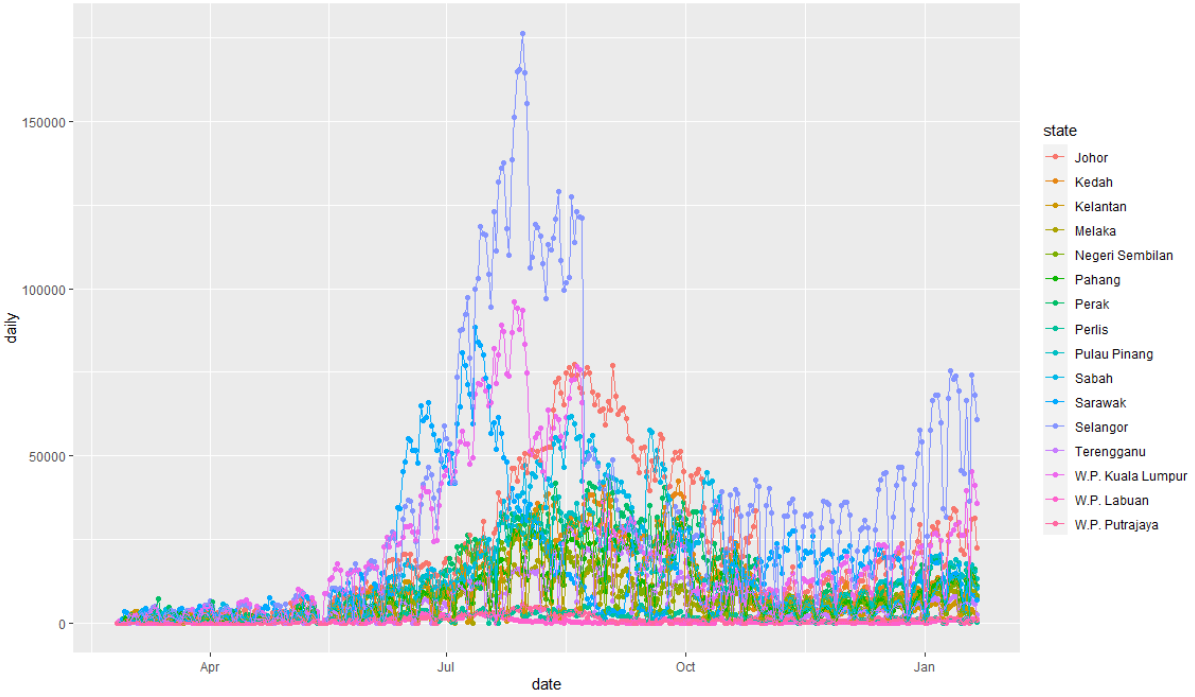


Figure 10 The observed daily number of vaccinations for all states in Malaysia

Table 3 : The summary estimated parameters of each state in Malaysia

States	ω	β	α	AIC	BIC	QIC	Log-likelihood function
Johor	0.806	0.492	0.431	821390.400	821401.800	821388.300	-410692.200
Kedah	1.530	0.871	-0.029	615122.600	615134.100	615123.800	-307558.300
Kelantan	3.966	0.675	-0.099	1461513.000	1461525.000	1461513.000	-730753.600
Melaka	0.766	0.119	0.799	699090.400	699101.800	699085.100	-349542.200
Negeri Sembilan	0.669	0.127	0.803	715105.500	715117.000	715097.800	-357549.800
Pahang	0.725	0.215	0.710	790189.700	790201.100	790182.800	-395091.800
Perak	0.998	0.386	0.514	797213.400	797224.800	797211.400	-398603.700
Pulau Pinang	0.816	0.217	0.700	712684.700	712696.100	712677.500	-356339.300
Sabah	0.553	0.928	0.016	389791.100	389802.600	389799.000	-194892.600
Sarawak	0.955	0.937	-0.031	750396.800	750408.200	750402.000	-375195.400
Selangor	1.092	0.762	0.138	1766062.000	1766074.000	1766064.000	-883028.200
W.P. Labuan	0.900	0.156	0.720	134803.000	134814.400	134801.500	-67398.500
W.P. Putrajaya	0.795	0.101	0.795	207944.300	207955.700	207940.300	-103969.200
W.P. Kuala Lumpur	1.153	0.553	0.337	1310264.000	1310276.000	1310264.000	-655129.200
Perlis	0.851	0.085	0.805	256344.000	256355.400	256340.000	-128169.000
Terengganu	0.752	0.110	0.810	814428.600	814440.000	814422.500	-407211.300

4.3. Model Validation

The fitted model applied to each state is validated using the Pearson residuals. Figure 11-Figure 13 shows the cumulative periodogram in Perlis, Pulau Pinang, Negeri Sembilan, Melaka, Kelantan, Pahang, Terengganu, W.P. Putrajaya, and W.P. Labuan. The rest of the cumulative periodograms is shown in Figure 14-Figure 21. Based on Figure 11-Figure 13, part of the cumulative periodogram do lies outside the 95% confidence interval. It indicates that the model is not adequate in these states. While the rest of the cumulative periodograms in Figure 14-Figure 21 do lies inside the 95% confidence interval indicates that the model is adequate in the states observed.

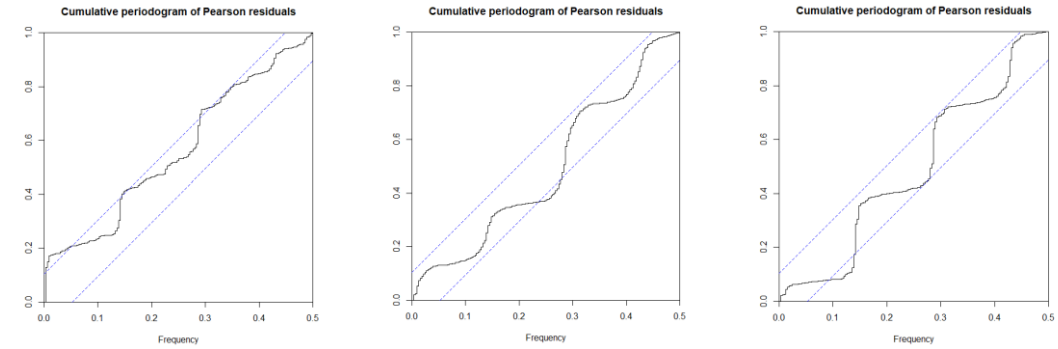


Figure 11 : Cumulative periodograms in Perlis, Pulau Pinang and Negeri Sembilan

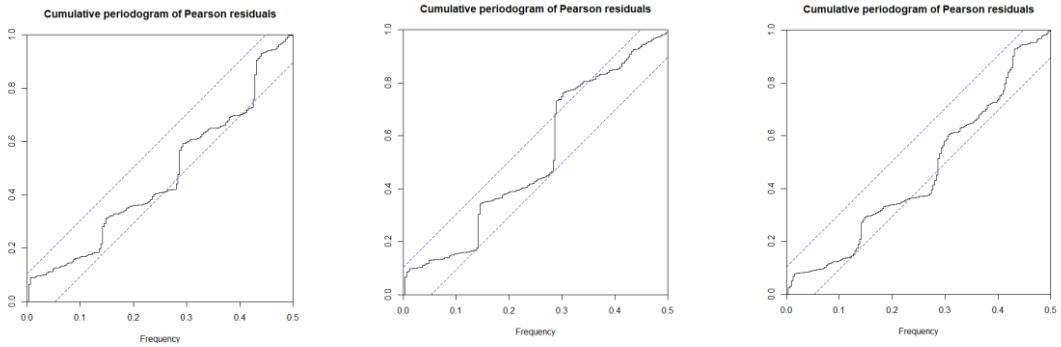


Figure 12 : Cumulative periodograms in Melaka, Kelantan, Pahang

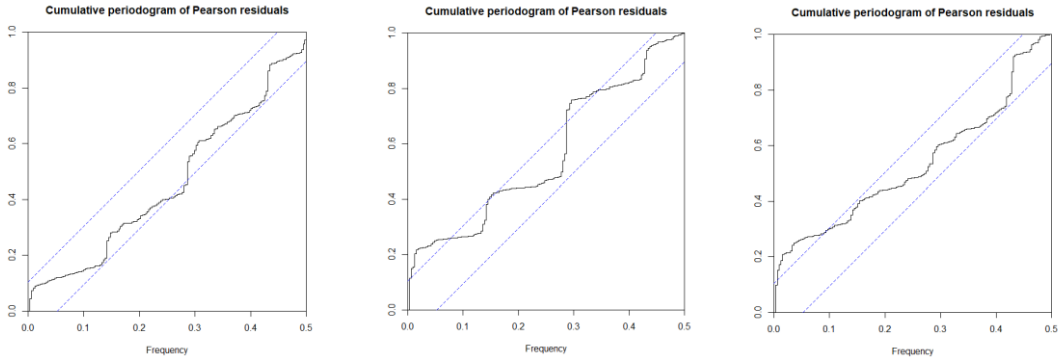


Figure 13 : Cumulative periodograms in Terengganu, W.P. Putrajaya, and W.P. Labuan

4.4. Model Prediction

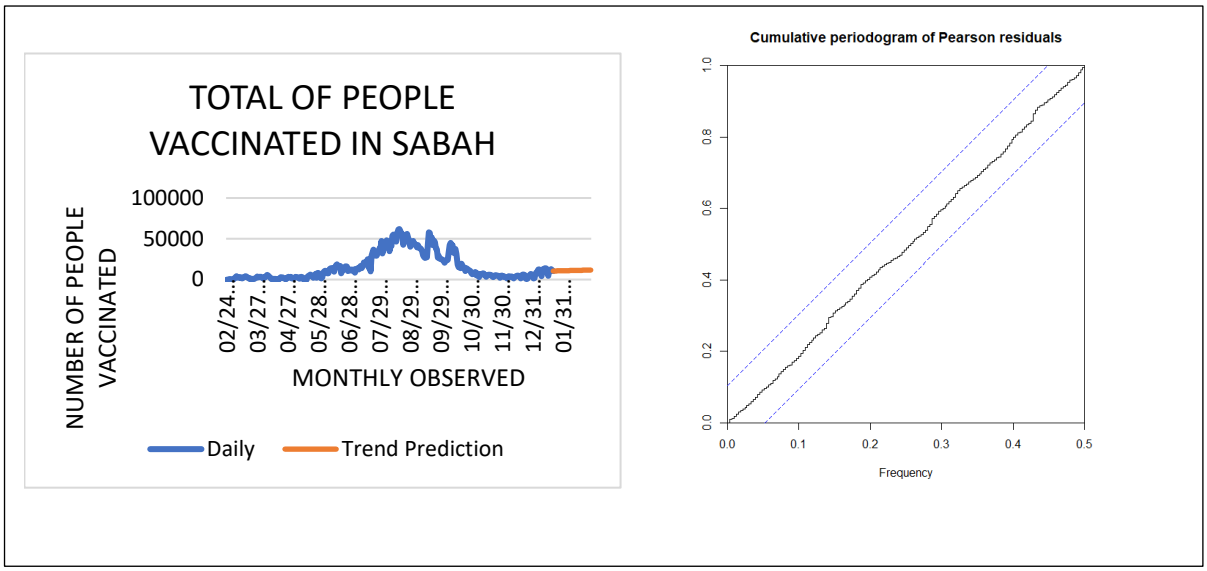
The model proceeds to predict the trend of daily vaccination for COVID-19 in February 2022. The trend prediction for all states except Perlis, Pulau Pinang, Negeri Sembilan, Melaka, Kelantan, Pahang, Terengganu, W.P. Putrajaya, and W.P. Labuan are illustrated in Figure 14–Figure 21. The actual number of daily vaccination (blue lines) from February 2021 until January 2022 are included to observe the trend. The parameters for respective states are presented in Table 4.

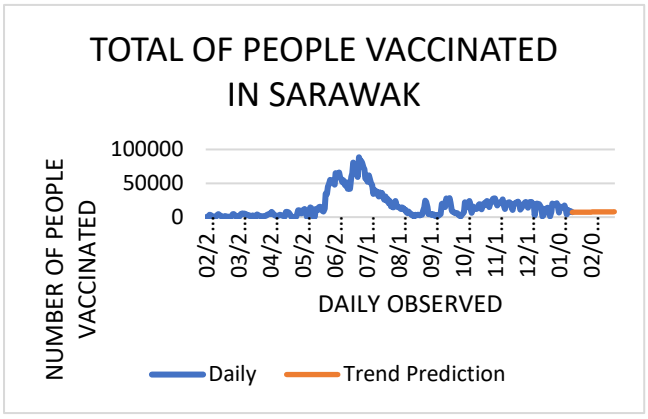
Generally, the prediction of the model is close to the actual number of daily vaccinations. The trend prediction in all states observed which are Sabah, Sarawak, Selangor and W.P. Kuala Lumpur almost accurate and close to the actual number of daily vaccinations. On the other hand, Kedah and Johor seems to have a stable trend despite having fluctuation in the earlier timeline. Specifically, the model prediction in Perak does not give the accurate trend as it is assumed to have a decreasing trend. It seems to have unstable line. Hence, the model prediction might not be accurate by comparing the values of parameters for COVID-19 in Malaysia.

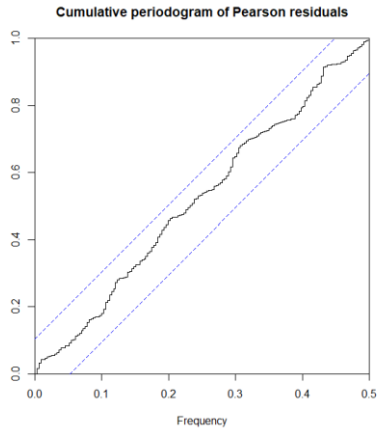
Table 4 : Parameters of the model and trend prediction for the respective state

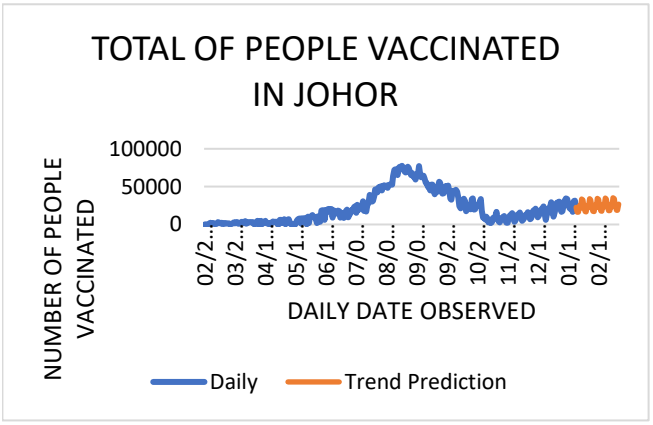
	Sabah	Sarawak	Johor	Kedah
ω	0.553	0.955	0.806	1.53
α	0.016	-0.031	0.431	-0.029
β	0.928	0.937	0.492	0.871
Trend	Increase	Increase	Increase	Increase

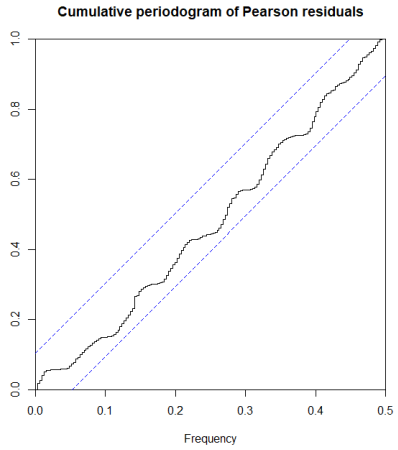
	Perak	Selangor	W.P. Kuala Lumpur
ω	0.998	1.092	1.153
α	0.514	0.138	0.337
β	0.386	0.762	0.553
Trend	Decrease	Increase	Increase

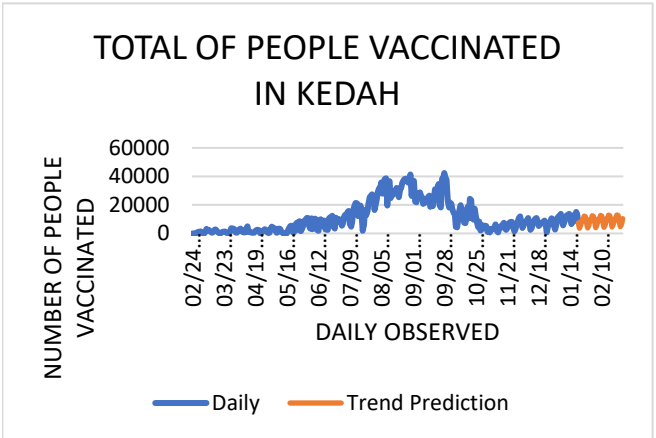


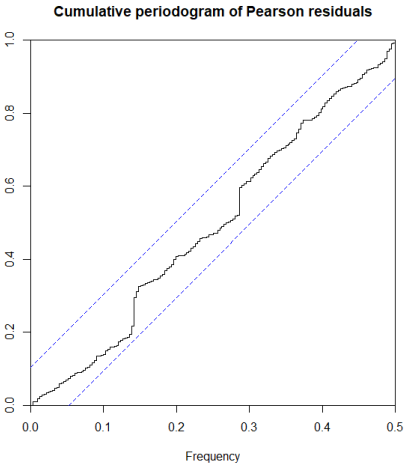

TOTAL OF PEOPLE VACCINATED IN SARAWAK
NUMBER OF PEOPLE VACCINATED
100000
50000
0
02/2...
03/2...
04/2...
05/2...
06/2...
07/1...
08/1...
09/1...
10/1...
11/1...
12/1...
01/0...
02/0...
DAILY OBSERVED
Daily
Trend Prediction

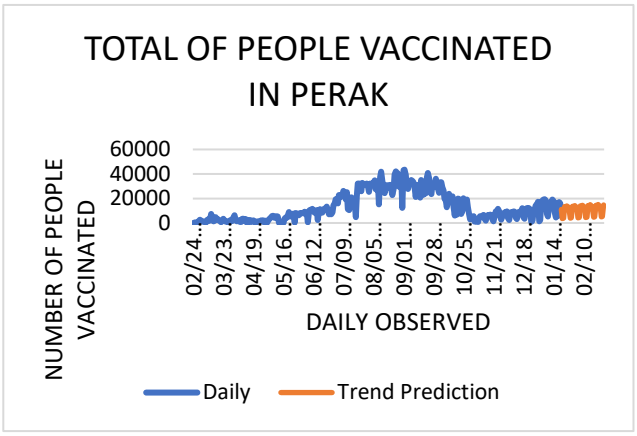

Cumulative periodogram of Pearson residuals
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Frequency

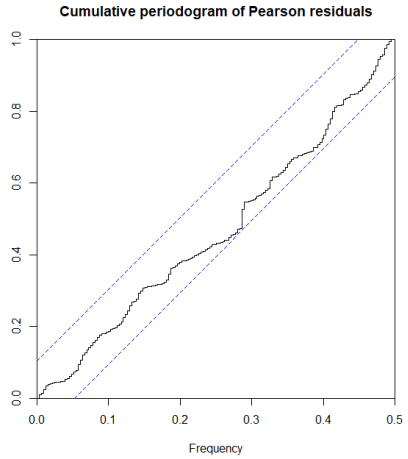

TOTAL OF PEOPLE VACCINATED IN JOHOR
NUMBER OF PEOPLE VACCINATED
100000
50000
0
02/2...
03/2...
04/1...
05/1...
06/1...
07/0...
08/0...
09/0...
09/2...
10/2...
11/2...
12/1...
01/1...
02/1...
DAILY DATE OBSERVED
Daily
Trend Prediction

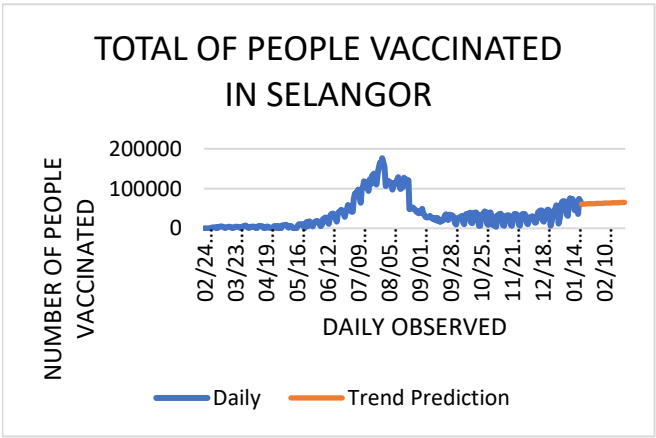

Cumulative periodogram of Pearson residuals
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Frequency

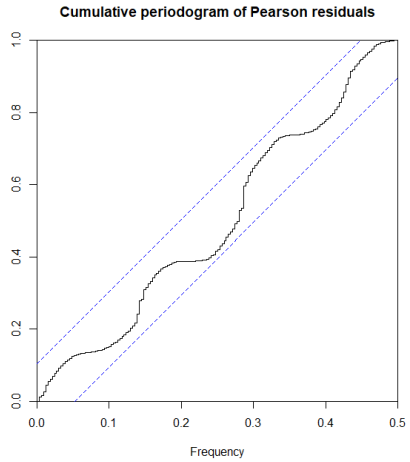

TOTAL OF PEOPLE VACCINATED IN KEDAH
NUMBER OF PEOPLE VACCINATED
60000
40000
20000
0
02/24...
03/23...
04/19...
05/16...
06/12...
07/09...
08/05...
09/01...
09/28...
10/25...
11/21...
12/18...
01/14...
02/10...
DAILY OBSERVED
Daily
Trend Prediction

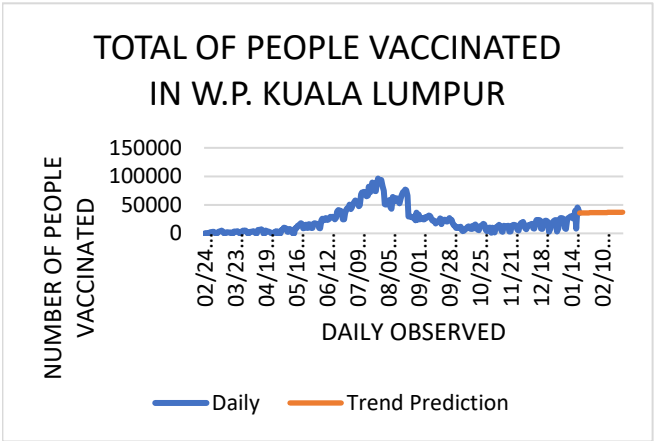

Cumulative periodogram of Pearson residuals
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Frequency

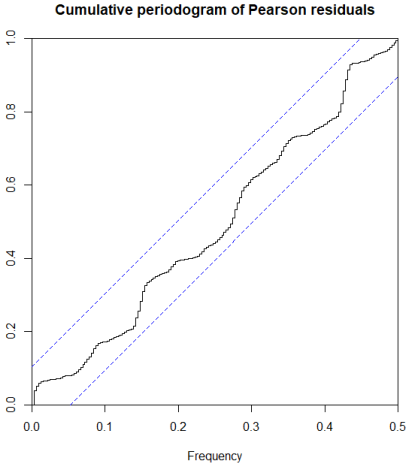

TOTAL OF PEOPLE VACCINATED IN PERAK
NUMBER OF PEOPLE VACCINATED
60000
40000
20000
0
02/24...
03/23...
04/19...
05/16...
06/12...
07/09...
08/05...
09/01...
09/28...
10/25...
11/21...
12/18...
01/14...
02/10...
DAILY OBSERVED
Daily
Trend Prediction


Cumulative periodogram of Pearson residuals
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Frequency


TOTAL OF PEOPLE VACCINATED IN SELANGOR
NUMBER OF PEOPLE VACCINATED
200000
100000
0
02/24...
03/23...
04/19...
05/16...
06/12...
07/09...
08/05...
09/01...
09/28...
10/25...
11/21...
12/18...
01/14...
02/10...
DAILY OBSERVED
Daily
Trend Prediction


Cumulative periodogram of Pearson residuals
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Frequency


TOTAL OF PEOPLE VACCINATED IN W.P. KUALA LUMPUR
NUMBER OF PEOPLE VACCINATED
150000
100000
50000
0
02/24...
03/23...
04/19...
05/16...
06/12...
07/09...
08/05...
09/01...
09/28...
10/25...
11/21...
12/18...
01/14...
02/10...
DAILY OBSERVED
Daily
Trend Prediction


Cumulative periodogram of Pearson residuals
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Frequency

Conclusion

This study has satisfied the objectives outlined in Chapter 1. To summarize, the daily vaccination data observed in all states in Malaysia are overdispersed. Then, the suitable model used for the data is by using log-linear Poisson autoregressive model.

The difference between the greatest and lowest values could be attributed to environmental factors such as the requirement for working in the city. For example, W.P. Putrajaya has the highest rate of vaccination, possibly because the state is in a governmental region and many people who work there are obliged to get vaccinated to go to work.

Results find that the model is suitable for modelling the daily vaccination of COVID-19 in all states in Malaysia except for Perlis, Pulau Pinang, Negeri Sembilan, Melaka, Kelantan, Pahang, Terengganu, W.P. Putrajaya, and W.P. Labuan for the period from 24 February 2021 until 21 January 2022. Therefore, the model is used in predicting the number of daily vaccinations in all states in February 2022 except Perlis, Pulau Pinang, Negeri Sembilan, Melaka, Kelantan, Pahang, Terengganu, W.P. Putrajaya, and W.P. Labuan as the model is not suitable to use on these states.

Lastly, the log-linear Poisson autoregressive model used fit the COVID-19 data as it can interpret the data into produce the output that consists of the parameters. From the model's parameter, there exists a variation and trend between the states.

Acknowledgement

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Dr NorHaiza Binti Ahmad, for encouragement, guidance, critics and friendship. I am also very thankful to my family for their guidance, advice and motivation. Without their continued support and interest, this thesis would not have been the same as presented here. I am also indebted to Universiti Teknologi Malaysia (UTM) for giving me the opportunity in this study. Librarians at UTM, friends and lecturers also deserve special thanks for their assistance in supplying the relevant literatures. My fellow undergraduate student should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all people around me.

References

[1] Agosto, A. and Giudici, P. (2020) 'A poisson autoregressive model to understand covid-19 contagion dynamics', *Risks*, 8(3).

[2] Chakraborty, T. and Ghosh, I. (2020) 'Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis', *Chaos, Solitons and Fractals*, 135.

[3] Chintalapudi, N., Battineni, G. and 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*, 53(3).

[4] Domi, M., Leitson, M., Gifford, D., Nicolaou, A., Sreenivas, K. and Bishnoi, C. (2021) 'The BNT162b2 vaccine is associated with lower new COVID-19 cases in nursing home residents and staff', *Journal of the American Geriatrics Society*, 69(8).

[5] Famoye, F. and Singh, K. P. (2021) 'Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data', *Journal of Data Science*, 4(1).

[6] Jamaludin, 'Aaishah Radziah, Yusof, F. and Suhartono, S. (2020) 'Modelling asthma cases by count analysis approach: Poisson INGARCH and Negative Binomial INGARCH', *MATEMATIKA*, 36(1).

[7] Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., Sun, F., Jit, M., Munday, J. D., Davies, N., Gimma, A., van Zandvoort, K., Gibbs, H., Hellewell, J., Jarvis, C. I., Clifford, S., Quilty, B. J., Bosse, N. I., Abbott, S., Klepac, P. and Flasche, S. (2020) 'Early dynamics of transmission and control of COVID-19: a mathematical modelling study', *The Lancet Infectious Diseases*, 20(5).

[8] Mbuvha, R. and Marwala, T. (2020) 'Bayesian inference of COVID-19 spreading rates in South Africa', *PLoS ONE*, 15(8).

[9] Singh, S., Parmar, K. S., Kumar, J. and Makkhan, S. J. S. (2020) 'Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19', *Chaos, Solitons and Fractals*, 135.

[10] Vasileiou, E., Simpson, C. R., Shi, T., Kerr, S., Agrawal, U., Akbari, A., Bedston, S., Beggs, J., Bradley, D., Chuter, A., de Lusignan, S., Docherty, A. B., Ford, D., Hobbs, F. R., Joy, M., Katikireddi, S. V., Marple, J., McCowan, C., McGagh, D., McMenamin, J., Moore, E., Murray, J.L., Pan, J., Ritchie, L., Shah, S. A., Stock, S., Torabi, F., Tsang, R. S., Wood, R., Woolhouse, M.,Robertson, C. and Sheikh, A. (2021) 'Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study', *The Lancet*, 397(10285).