Forecasting the number of Australia daily cases by renewal equation

Author: Ruiqi Tang
Supervisor: Michael Lydeamore

Abstract

The COVID-19 pandemic has had a large impact on industry in Australia, and outcomes associated with the pandemic are highly related with decision-making and policy implementations. If the outbreak can be predicted as early and as precisely as possible, it becomes easier for decision makers to implement policy changes that will avoid significant loss. However, forecasting the outbreak is complex, because it is very difficult to understand the process of transmission of an infectious disease in the potential population. In this paper, the short-term forecasting of daily COVID-19 case numbers is the primary task, and reproduction number is used to do forecasting. The reproduction number should be estimated in advance by using a combination of dynamic disease models and time series models. Seasonally adjusted series will be used to estimate a suitable reproduction number, and the reproduction number can be used to do forecasting using the renewal equation. A Rolling-origin evaluation process is applied to test the performance of mode, and 11 rounds one-week-ahead out-of-sample prediction are produced. The result shows that our model perform very good at most of time. By comparing criterion value (e.g.: MAE and RMSE), the most suitable model is selected.

1.   Introduction

By the end of September 2022, WHO (World Health Organization) has reported that the global total number of reported confirmed cases for COVID-19 (novel coronavirus SARS-CoV-2) is more than 600 million, and the global total number of reported deaths excesses 6 million (WHO,2022). The situation in Australia is also very serious. The total number of infected cases is about ten million, and total number of deaths cases is 15 thousand (WHO, 2022). In the 2021 Census of population in Australia (ABS, 2022), it shows that the total population is 25,890,773 in Australia, which means that nearly half of population were infected by COVID-19. Since the earliest recorded pandemic, the COVID-19 is one of the worst pandemics (Yamamoto, 2021). We observed multiple policy implementations to fight COVID-19 for different countries. For instance, The number of confirmed cases in China increase very slowly. China has implemented a 'Zero Covid' strategy, including extensive lockdowns, which can be regarded as one of the most efficient methods for the government, as it stops the spread of the virus and reduces the number of confirmed and deaths cases. However, China is suffering from a large economic depression (Dyer, 2022).

Much previous research illustrates that the early intervention will reduce virus spreading efficiently (Bootsma & Ferguson, 2007). Pei et al (2020) estimated that the number of deaths will reduce 36,000, as some restrictions start one week earlier in the Unite of State at the beginning of May. However, to intervene early it is necessary to have a relatively accurate near-term forecasts of virus transmission. It is crucial to understand the transmission mechanism of the infectious diseases from the infective population to susceptive population through the chain of infections, and mathematical models of infection transmission may be very useful to determine the behavior of viruses (Kucharski, 2020). The earliest study on epidemiological models was by Bernoulli, who was the pioneer scientist in the field of epidemiological model (Yadav & Akhter, 2021). Now, more and more models are used to forecast future incidence (Chretien et al., 2015; Nsoesie et al., 2014). Some models rely on a purely statistical (non-mechanistic) framework (Goldstein et al., 2011), while some rely on biological models (Meltzer et al., 2014). In recent years, the accuracy of many models is very high, such as the model for Dengue (Dengue Forecasting, 2017) and Chikungunya (Chikungunya Forecasting, 2017). Each of these methods has merit. Statistical models tend to produce highly accurate near-term forecasts, while biological models have known mechanisms that allow for interrogation into transmission patterns. It is very important to balance the accuracy of their model and the underlying knowledge of the transmission disease. If researchers pay much attention to improving the accuracy of the statistics model and ignore the process of transmission, the model may not work very well for a long period.

In this paper, we will provide a model that combines time series methods and biological models. We will use a time decomposition method to remove the seasonal pattern and use seasonally adjusted daily new case numbers into disease models. Then, we will calculate one-week-ahead out-of-sample forecasting value by our model and compare the forecasting value with the observed value. Our model is able to provide a relatively accurate near-term forecasts, which may be much helpful for policy implementations. However, some problems are raised. The estimation depends on an unknown time parameter, called time interval, if different time intervals are chosen , the forecasting value will be much different. After obtaining an appropriate time interval, we will use the time parameter to estimate reproduction number, which is used to measure

transmission. Then, we can use the reproduction number to do forecasting. Therefore, we will mainly solve two questions for this paper. The first is to choose a suitable time interval to estimate reproductive number. The secondary task is to predict short term infected cases by the estimated production number.

The data of COVID-19 data in Australia will be applied into our model. The data contains information of all eight states in Australia, and we sum up the number of cases in all eight states. Then, the data start from 2021-07-01, and it ends at 2022-09-09. Our model shows a relatively accurate forecasting in most of situations. Some evaluation critical, such as MAE and RMSE, will be employed. The gap between the actual value and prediction value is very small. Then, a suitable time interval is one week for the estimation of reproduction number, and one-week interval provide the best prediction value compared with other time interval. Also, it avoids much statistical noise.

2.   Literature reviews
2.1 Reproduction number and serial interval

We will focus more on the reproduction number R, defined as the number of secondary infections that arise from a typical primary case in a completely susceptible population (Wallinga & Lipsitch, 2007). The reproduction number is one of the key parameters in the mathematical transmission models and it is usually used to measure the transmissibility during an epidemic (Anderson & May, 1991). If $R \leq 1$, it represents the number of infected cases will decline, usually at the end of epidemics. Otherwise (i.e. if $R \geq 1$), the number of infected cases increases, usually at the beginning of epidemics. By estimating the reproduction number, it is possible to forecast future incidence (Pierre et. al., 2018).

Wallinga and Teunis developed a method to estimate the reproduction number (Wallinga &Teunis, 2004), using only the incidence number and the distribution of the serial interval. The serial interval is defined to be the time from the infection of a primary case to infection of the cases the individual generates (Cori et. al., 2013). An illustration of the serial interval is shown below, and the process can be explained by three time periods, denoted A_i, B_{ij}, A_j . At A_i period, individual i is infected, but the individual has no symptoms, and the individual is not infectious. At the second period B_{ij} The individual i is symptomatic and infectious. Then, individual j is infected by individual i after period B_{ij}. Individual j experiences an incubation period, which is A_j, then individual j will be infectious. Actually, the serial interval should be $A_i + B_{ij}$. However, it is impossible to determine the time when individual i is infected, so the time $A_i + B_{ij}$ may be found difficultly. Ferguson et. al (2005) prove that the $A_i + B_{ij}$ is equivalent to $B_{ij} + A_j$. Therefore, the serial interval can be defined as $A_i + B_{ij}$. It is very easy to identify the time of symptom onset for an individual.
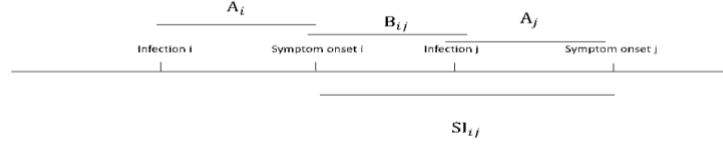
Figure 1: Graph of SI

2.2 Renewal equation

We will use the renewal equation to construct the distribution of incidence number and one-week-ahead forecasts. Mathematically, it is a Poisson transmission process, assume an individual is infected at time t. Then, the probability of others is infected is $\beta(t,\tau)\sigma$, $\beta(t,\tau)$ is function to describe the infectiousness, during time interval σ, where σ is a small time interval, t is a calendar date and τ is the period of observation. Fraser (2007) show the renewal equation is given by the formula

$$I(t) = \int_0^\infty \beta(t,\tau)I(t-\tau)\,d\tau$$

where $I(t)$ represents the incidence number at time t. This equation shows that the number of new cases is proportional to the number of past cases multiplying a parameter. This parameter relates to reproduction numbers (Fraser, 2007). it can be simply estimated as

$$\beta(t,\tau) = R(t)w(\tau)$$

It is the production of reproduction number and the serial interval distribution. It depends on the calendar data and the time interval. if time interval vary, the $\beta(t,\tau)$ will change as well, and it will relate with the prediction of incidence number.

Pierre et al. (2018) developed a simple express of renewal equation, and the equation given by

$$I_t \sim pois\left(R_t \sum_\tau^t I_{t-\tau} w(\tau)\right)$$

This equation illustrates that the average number of new infected cases, generated by someone infected at time $t-\tau$, is $R_t w_\tau$ at time t. The incidence number at time t is approximate to Poisson distribution (cori, et. al., 2013). By this equation, it is easy to construct the likelihood function given observation, and this likelihood function can be used to calculate a set of model parameters. Then, the posterior distribution of

reproduction number also can be obtained, given observations of incidence and prior distribution of the serial interval (cori, et. al. 2013)

2.3 Estimation of reproduction number

The incidence number follows Poisson distribution, some explanation is provided above. We can use a Bayesian framework to determine the posterior distribution of reproduction number after updated new information.

The prior distribution is the distribution of random variables with some previous knowledge. The posterior distribution can be defined as a distribution of a random variable that the information of sample observation data, which is likelihood function, is added in. Sometimes, researchers just choose a prior distribution through the experience, so it is a good start to assume a distribution, which is based on the features of the random variable, such as the range of parameter and the mean of the parameter. As more information is updated into the model, the posterior distribution will close the true underlying distribution of the interested parameter.

According to much research on the serial interval, the prior distribution of the serial interval is widely regarded as the gamma distribution (WHO Ebola Response Team, 2015a). Therefore, the incidence number at time t follows Poisson distribution, with mean $R_t \sum_{\tau}^{t} I_{t-\tau} w(\tau)$, the likelihood function is conditional all past value during the time period $[t - \tau + 1, \ t]$, where $t - \tau + 1$ is a time before t.

$$P(I_{t-\tau+1}, .., I_t | I_0, \ldots, I_{t-1}, w_\tau, R_t) = \prod_{t-\tau+1}^{t} \frac{(R_t \sum_{\tau=1}^{t} I_{t-\tau} w(\tau))^{I_t} e^{R_t \sum_{\tau=1}^{t} I_{t-\tau} w(\tau)}}{I_t!}$$

Then, the posterior shown as

$$P(I_{t-\tau+1}, .., I_t | I_0, \ldots, I_{t-1}, w_\tau) = \prod_{t-\tau+1}^{t} \frac{(R_t \sum_{\tau=1}^{t} I_{t-\tau} w(\tau))^{I_t} e^{R_t \sum_{\tau=1}^{t} I_{t-\tau} w(\tau)}}{I_t!} \times p(R_{t,\tau})$$

where $p(R_{t,\tau})$ is the prior distribution. Therefore, we can easily get the posterior distribution for the reproduction number by this method, and derive the parameter of the model. We can get the mean or median of reproduction number at time t, with the time interval τ. We have an assumption, reproduction number can be regarded as a constant in time interval τ.

3. Method and Martial

We decompose the daily notification data to obtain a seasonal adjusted series. Then, we use a Bayesian-based method to estimate the reproduction number for different time intervals. The estimated reproduction number will be used to forecast the future incidence by the renewal equation. Some evaluation critical will give an result about whether the prediction is well-structured.

3.1 Data source and data transformation

Our data is the daily reported COVID-19 notifications in Australia. The data is collected from a third party website (COVID LIVE, 2022), which collates data reported from each state government website. This data set includes the daily new cases in eight states in Australia, and the data is collected from the beginning of 2020. We will forecast the national notifications, by adding together the notifications from each of the eight states. We focus on the time period 2020/01/25 to 2022/09/09.

The graph below shows the number of newly daily cases. Australia experienced a relatively low COVID-19 burden until early 2022, where cases increased rapidly, despite high vaccination rates in the population; approximately 97.1% of population have at least one dose (Vaccination, 2022). Also, all over the world was suffering a quick virus spreading at the same time. It may be caused by the Omicron variant (BA.1), BA.1 spreaded very quickly. After the peak at the beginning of 2022, the incidence showed a more regular pattern.
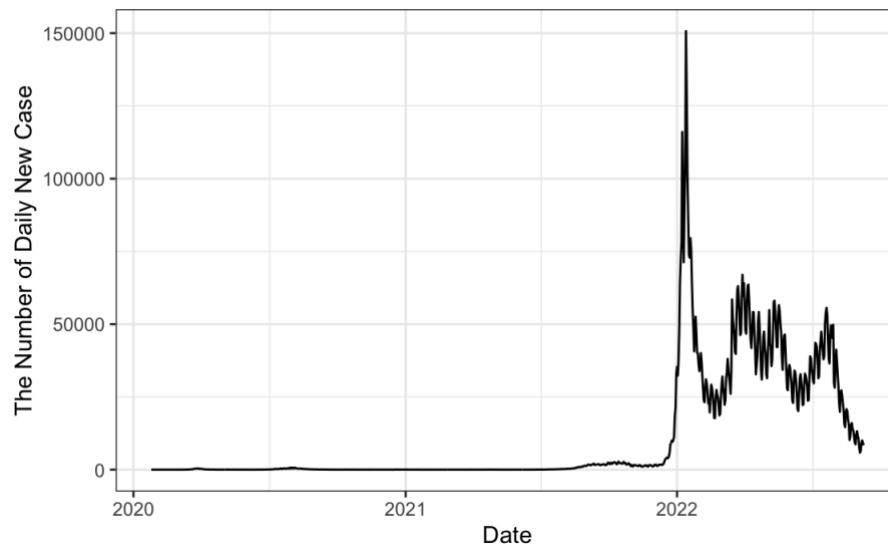

Figure 2: Raw data

From the above graph, the plot shows that variation looks unstable with the level of the series. For an unstable time series, a transformation must be applied, and a logarithmic transformation is often useful (Hyndman, 2021), because the logarithmic is interpretable, the logarithmic transformation usually represents the rate of change. We only select data after 2021-07-01, as the epidemiology of COVID-19 has changed over time, and so only recent observations are useful to benchmark our algorithm. This time period includes increase, peak and decreasing case numbers, allowing for robust testing.

From the graph, it is clear that there is a regular pattern. If we can estimate the time series without the regular pattern, the estimation may be much easier, because the pattern looks fixed for all periods and it does not have much effect on time series (Zhang et. al., 2014).

3.2 Time series decomposition and seasonal adjusted series

Time series decomposition methods try to extract the underlying pattern in the data series from randomness, and can be decompose three parts (Zhang, et al. 2014)

$$Time\ series = Trend + Seasonality + reminder$$

alternative, a multiplicative decomposition would be written as (Hyndman & Athanasopoulos, 2021)

$$Time\ series = Trend \times Seasonality \times reminder$$

equivalent to

$$log(Time\ series) = log(Trend) + log(Seasonality) + log(reminder)$$

The underlying pattern can be separated into two parts, shown as above formulae, which is trend and seasonality. Trend pattern represents a long term change, and seasonal pattern is the periodic fluctuations within a constant time interval (Zhang, et al. 2014). Then, the underlying pattern can be used to predict and forecast (Zhang, et al. 2014).

If the seasonal component is removed from the original data, the resulting values will be the "seasonally adjusted", which is

$$Time\ series/Seasonality = Trend \times reminder$$

or

$$Time\ series - Seasonality = Trend + reminder$$

or

$$log(Time\ series) - log(Seasonality) = log(Time\ series/Seasonality) = log(Trend) + log(reminder)$$

If the change due to seasonality is not the main problem, the seasonally adjusted will be much more useful (Hyndman & Athanasopoulos, 2021)

There are many time series decomposition methods, such as the X-11 method, SEATS method. We will choose the STL method (Cleveland et. al., 1990), which is Seasonal and Trend decomposition using Loess, and this method was developed by R. B. Cleveland et al in 1990. There are three main advantages for this method. The first one is that this decomposition allows researchers to handle any type of time frequency. Secondly, researchers can change the seasonal components and the rate of change. Then, it can be robust to outliers.

We have carefully investigated the number of daily new cases from the beginning of 2022 in all Australia. The graph shown above (figure 2), we find that there is a clearly fixed seasonal pattern, which is a weekly seasonal pattern. Therefore, we may ignore the seasonal component now, and focus on seasonally adjusted series. After we finish the forecasting of seasonally adjusted series, we can directly add the seasonal pattern back on the forecasting of seasonally adjusted series, because the seasonal pattern is fixed, and it does not have much effect on the prediction.

We remove the seasonal pattern by STL time decomposition and obtain seasonal-adjusted time series. From the graph, the seasonal pattern does not have much change after 2022-01-01. It moves in a constant range,

and it does not show too much randomness, so it is reasonable to drop this part. Then, there is a small gray bar in front of the graph, which shows the importance of this component. The smaller bar means less importance. otherwise, it is very important. The size of the bar behind the seasonal component is the largest, which means that the weight in the time series is very low. Therefore, we can drop this component temporarily.
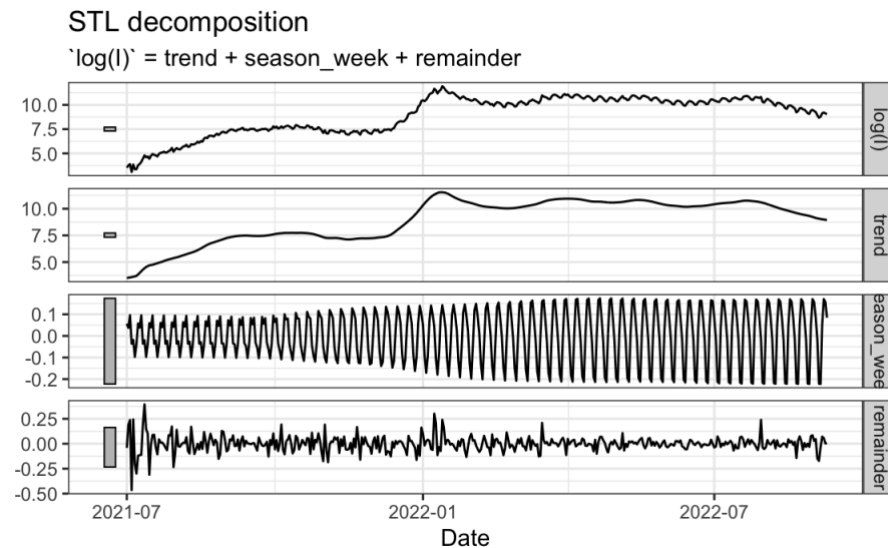

Figure 3: Time decomposition

The seasonally adjusted time serial will be easily obtained, after the seasonal pattern is dropped. The graph is given below. From the graph, there is less fluctuation, compared with the original data graph. Before we do the following estimation, the back transformation is a necessary process. Then, we will use the back transformation series to estimate the reproduction number. After we finish prediction of seasonally adjusted series, we will add back the seasonal component.
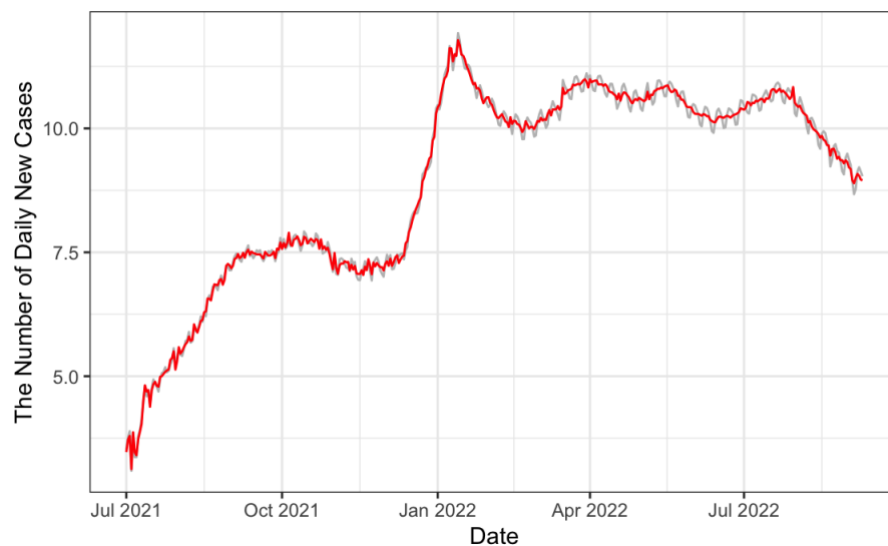

Figure 4: seasonally adjusted plot
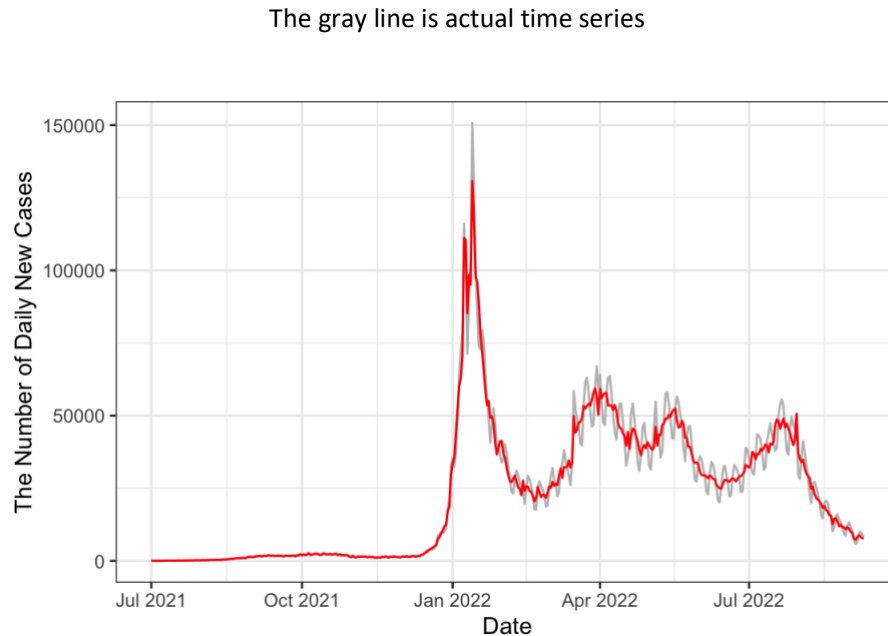
The red line is seasonally adjusted series

Figure 5: Back transformation of seasonally adjusted
The red line is seasonally adjusted series
The gray line is actual time series

3.3 Estimate R and serial interval distribution

In order to produce a forecast, there are two required elements: estimation of reproduction number, and the distribution of serial interval distribution

When we estimate reproduction number, contact rates and transmissibility can change over time, which means the estimated reproduction number will vary over time. it highly depends on the time serial. However, over small time periods (say 1 month), it is feasible to assume that the reproduction number may be steady. The length of time interval is a problem. If a small-time interval is chosen, the estimated reproduction number changes very quickly and there will be much statistical noise (cori, et. all., 2013). Otherwise, if a large value of time interval is chosen, the estimated reproduction number will be more smoothing and lose much detail. In order to solve this problem, we set five different time intervals, and we called it a window, to estimate reproduction number, which are one-week, two-week, three-week, four-week and five-week. We will examine each window to obtain an appropriate reproduction number.

The serial interval usually cannot be obtained, because the reported cases usually report times of the result of PCR (polymerase chain reaction) test or the symptom onset, rather than times of infection. It is very hard to determine the time when an individual gets infected. Therefore, it is rare to get empirical data to estimate the distribution of the serial interval. However, Wallinga and Teunis developed an ideal situation, and they prove an alternative expression of serial interval, and more detail is discussed in literature reviews. Therefore, it is relatively easy to estimate serial interval distribution through the reported cases. Nishiura et. al. (2020) use

empirical research to get the mean of serial interval distribution and the standard deviation of serial interval distribution, which is 4.7 days and 2.9 days, they have tested the distribution of serial interval. Some research (WHO Ebola Response Team, 2015a) illustrates that the serial interval distribution should follow gamma distribution. Therefore, the gamma distribution with mean 4.7 days, and standard deviation is 2.9 days will be used in our model. Furthermore, there is no evidence shown that the serial interval has changed as the virus evolves, which means that the serial interval of the Omicron variant is no different with the Delta variant. We will not need to consider the change of the serial interval. The PDF (probability density function) of the serial interval is shown below. From the graph, we can see that the peak is located at around 10-th day, which means that the infected individual, primary case, are more likely to infect other, secondary case.
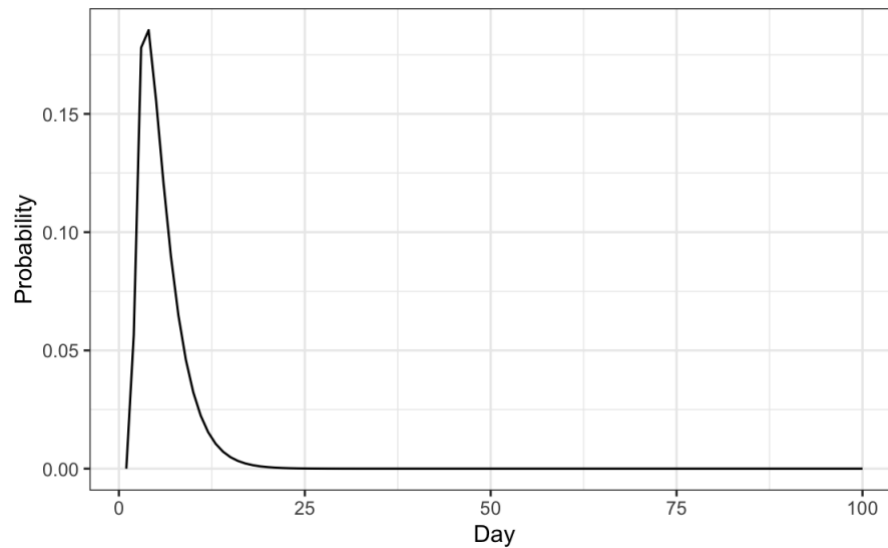

Figure 6: distribution of serial interval

After the two main problems are solved, we can estimate the reproduction number. We will implement a Bayesian-based method described in literature review part. We estimate the reproduction number by a R package called EpiEstim (Cori et al., 2013). This package only requires inputting the daily incidence number and parameters of serial interval. It will return estimated values, such as the mean, median and 95% quantile. The graph below shows the estimated mean of reproduction number.
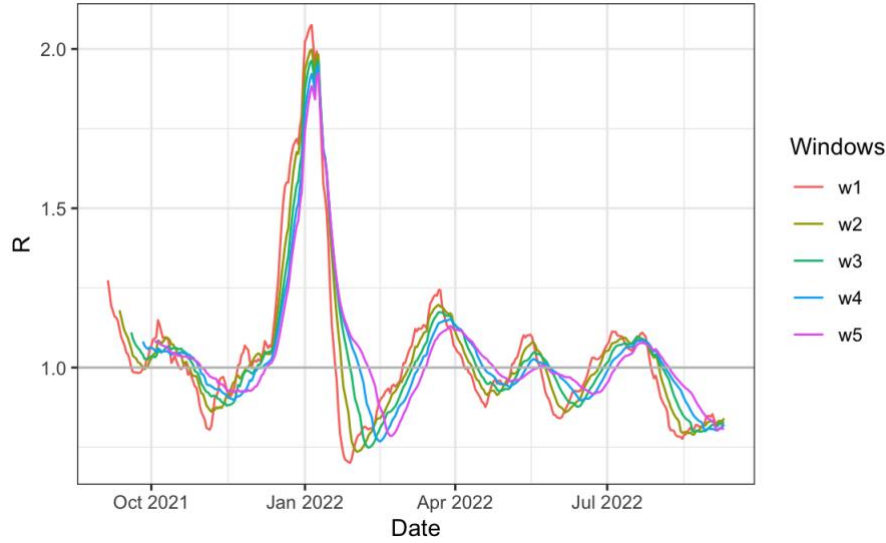
Figure 6: estimated R

An estimated reproduction number with a small value of window (such as one-week) changes rapidly. For a large value of window, such as a five-week window, estimated value will be more smoothing, and less statistical noise. Then, the estimated value is very large at the beginning of 2022, and a larger value means a larger number of secondary cases infected by primary cases, it also means a greater number of daily new cases. Compared with Figure 1, the daily new cases achieved a peak in January of 2022. Recently, the value is smaller than 1, which is a good indicator, the number of cases will decrease in the short term.

Herd Immunity may be considered as a reason to explain this situation. Since more than half of the population are infected and 95% of the population get at least one dose of vaccine (COVID-19 vaccines, 2022), the portion of immune persons in the population is quickly large, which means that it is very hard for viruses to spread from one individual to another. It also leads to the protection of susceptible persons who are not immune.

3.4 Forecasting Daily new cases

We will rely on the renewal equation to predict future incidence number. If we have all data before $t + 1$, we are able to calculate average number the incidence number at $t + 1$ by the equation below.

$$I_{t+1} \sim pois(R_{t+1} \sum_{\tau}^{t} I_{t+1-\tau} w(\tau))$$

Since we have all data before $t + 1$, we can have $I_{t+1-\tau}$. Then, we have an assumption before, the reproduction number, R_t, is a fixed value in a very short time interval, so R_t is equal with R_{t+1}. The R_t can be estimated using the data in period $[t - \tau + 1, \ t]$. $w(\tau)$ is serial interval distribution. Therefore, we have all the parameters that we use to do forecasting. It is the same idea to predict $I_{t+2}, I_{t+3}, \dots, I_{t+n}, n \in Z$.

We use a method called rolling-origin evaluation process (Tashman, 2000). We start our perdition at 2022-07-01, and we use all information before 2022-07-01, and estimate R_{220701}. Then, we construct one-week-ahead

out-of-sample forecasting by the renewal equation, which is the period from 2022-07-01 to 2022-07-07. On 2022-07-07, we expanded the in-sample data to include 7 data points (before 2022-07-07). Then, we use all information before 2022-07-07 to estimate R_{220707}. We produce next one-week out-of-sample forecasting again, and we get the forecasting value from 2022-07-08 to 2022-07-14. We repeat this process for another 8 times. In total, we have 10 rounds of one-week-ahead out-of-sample forecasting, and the time period starts at 2022-07-01, and ends at 2022-09-09. We use this process to plot a graph, which combines the 10 rounds and the actual observation value. The graph shows below.
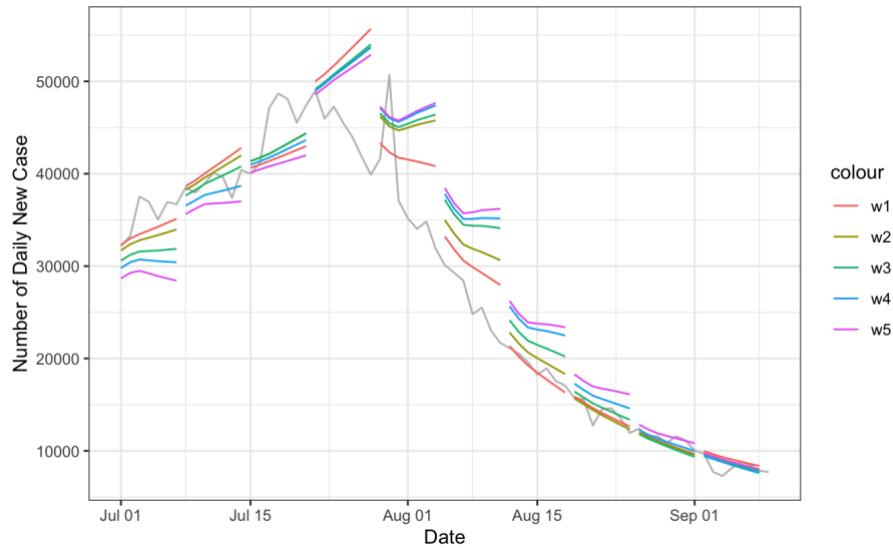

Figure 7: Forecasting for seasonal adjusted series
Colorful line is forecasting value
Gray line is actual value

The graph above shows the prediction value of seasonally adjusted series for five different window periods. For the last 4 rounds, the predicted value for different windows are very close, since the estimated R is very close from Figure 6. The difference between the forecasting value and the seasonally adjusted actual value is very small in the last 4 rounds as well. However, the prediction experiences a large increase during the end of July, it is much larger than the actual value, and the gap between the forecasting value and actual observation is very large. Then, the prediction value is smaller than the actual value at the first round for all five windows.

We consider adding seasonal patterns back into predicted seasonally adjusted series. Since the decomposition plot (Figure 3) shows a relatively fixed fluctuation of seasonality after 2022 Jan, it is forecast by simply taking the last period of each round (Hyndman & Athanasopoulos, 2021). In other words, it is the same as the seasonal naïve method. The graph below shows the forecasting with seasonal patterns.
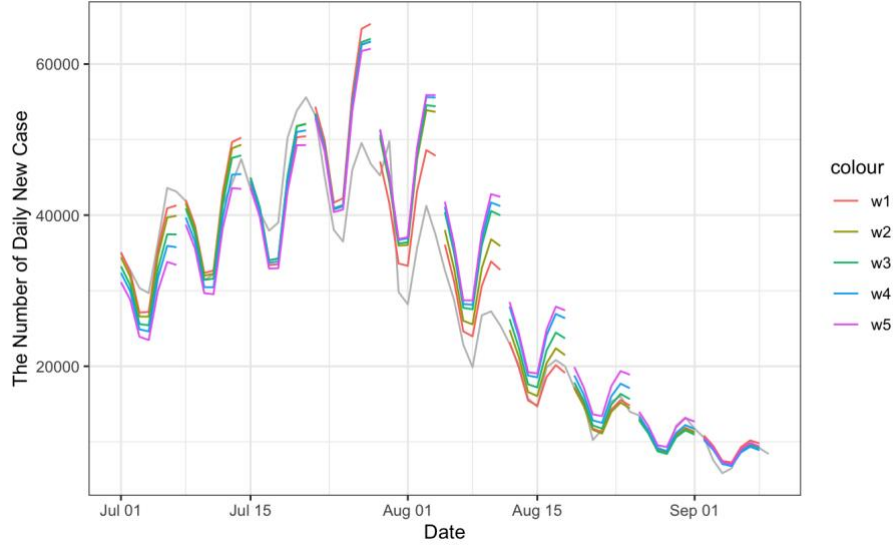
Figure 8: Forecasting with seasonal pattern
Colorful line is forecasting value
Grey line is actual value

From the graph, it is shown that most of rounds fit very well for all windows. However, there is a large increase at the end of July, and the forecasting value is about 10000 more than the observation value. For the last four rounds, they are almost the same as the actual value.

3.5 The accuracy of forecasting

It is not very easy to compare the accuracy between different windows by visualization, or plots. We have to find a quantitative method. Since we want to compare the several time series with the same units, the two most commonly measures used scale-dependent, which mean that the unit of time series has no effect on comparison, are based on the absolute errors or squared errors (Hyndman & Athanasopoulos, 2021). These two measures are easy to understand and compute. The formula given as below

$$Mean\ absolute\ error: MAE = mean(|e_t|)$$

$$Root\ Mean\ squared\ error: RMSE = \sqrt{mean(e_t^2)}$$

where e_t can be defined as the difference between the predicted value and the actual value.

$$e_t = y_t - \hat{y}_t$$

A smaller value of these two indicators represents a high accuracy, which means a high performance of the model. We have calculated the MAE and RMSE for each window, and we can find which model is the best.

Windows	MAE	RMSE
One	2841	4516
Two	3238	5012
Three	3845	5712
Four	4390	6199
Five	4956	6654

Table 1: Accuracy table

From the table, we find that the one week is the most suitable time interval, since both MAE and RMSE return a smallest value for one-week window. Therefore, it is better to do forecasting by using a one-week time interval. The plot of one-week window forecasting shown below. The graph shows that the forecasting with one-week window is very close the observed value
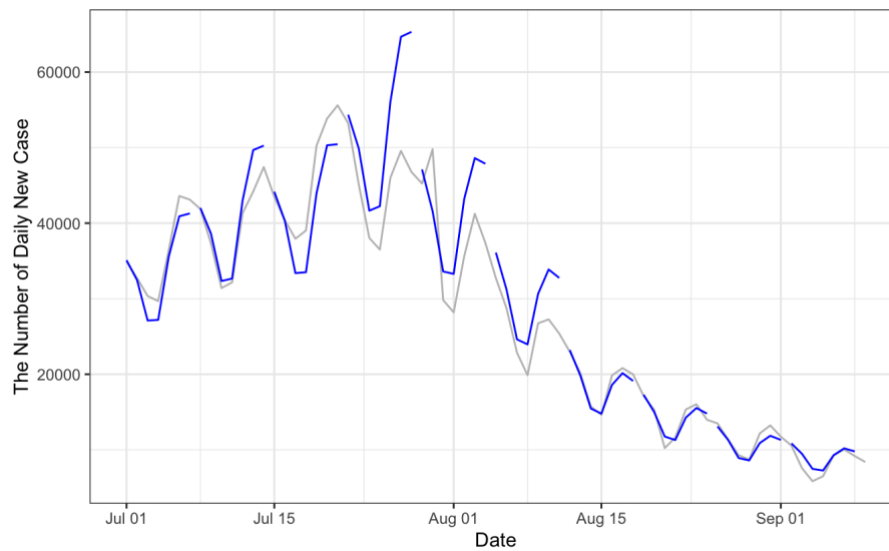

Figure 9: Forecasting with one week time interval

Blue line is forecasting value

Gray line is actual value

4.   Conclusion and discussion

Forecasting of the outbreak is a very challenging project, and the forecasting of disease is useful to monitor the development of virus spreading and help policy makers consider appropriate measures to decrease the negative impacts of the pandemic. It is valuable for decision and policy makers to take action to strengthen or relax interventions. In this paper, our task is to develop a method to do short-term forecasting. We use STL time series decomposition method to obtain seasonally adjusted series and remove the impact of seasonality. Then, we have estimated the reproduction number by five different time intervals, called windows, by a Bayesian framework. After getting an estimated reproduction number, we use the renewal equation to construct one-week-ahead out-of-sample forecasting, and we also apply a rolling-origin evaluation process to produce 10 rounds prediction. We use the data of daily new cases in Australia from 2021-07-01 to 2022-09-09

into our model. The most suitable time interval, windows, is solved by the data. By comparing MAE and RMSE, the most suitable windows are selected, which is one week.

Overall, our forecasting works very well in most situations with small mean absolute error (MAE) and root mean squared error (RMSE) . Then, our model also captures the seasonality pattern of the actual value. However, the forecasted values increase too quickly at the end of July. This is possibly because of changes in behaviour or policy at that time, which can have a large impact on the spreading of COVID-19. Our model cannot detect other effects on the COVID-19, such as policy implementation. If some strict policies are published at a certain time, the number of daily new cases will decrease quickly in the following days. Our model only considers that COVID-19 spreads in the natural environment without any restriction. Therefore, if some restrictions are imposed, our model will not return a relatively reasonable result.

However, there are some limitations in our research. we have not produced an uncertainty interval on the forecast. If we directly use the distribution of incidence number, which is Poisson distribution, the prediction interval will be very narrow. Since the rate of Poisson, which is the mean of Poisson distribution, is too large, the distribution will be more centralized, and the confidence level will be narrower, and the confidence level is very reasonable. The quantiles from a Poisson distribution are smaller than expected from the epidemiological process. Propagation of uncertainty from the estimate of the effective reproduction number could lead to more sensible uncertainty estimates. This is a focus of future work.

Our model is able to provide a simple and precise forecasting for the decision makers. The forecasting could give decision makers some insights on the policy of intervention, and they can change the policy by estimating future incidence number. If our forecasting shows a large increase in the following days, the restriction will be very strong. Otherwise, the restriction will be released. It may be easier for the government to set the policy, and avoid a significant loss on COVID-19. The forecasting may be easier to make decisions.

reference list

Who coronavirus (COVID-19) dashboard (no date) World Health Organization. World Health Organization. Available at: https://covid19.who.int/

Australian Bureau of Statistics. (2022). National, state and territory population. ABS. https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/mar-2022.

Yamamoto, Vicky, Bolanos, Joe F., Fiallos, John, Strand, Susanne E., Morris, Kevin, Shahrokhinia, Sanam ,. et al., 2020. COVID-19: Review of a 21st Century Pandemic from Etiology to Neuro-psychiatric Implications. Journal of Alzheimer's Disease, 77(2), 459-504.

Dyer, O,. (2022) "Covid-19: Lockdowns spread in China as omicron tests 'zero covid' strategy," BMJ (Online), 376, p. o859.

Bootsma, M. C. J., & Ferguson, N. M. (2007). The effect of public health measures on the 1918 influenza pandemic in U.S. cities. Proceedings of the National Academy of Sciences of the United States of America, 104(18), 7588–7593.

Yamana, T., Pei, S., Kandula, S., & Shaman, J. (2020). Projection of COVID-19 cases and deaths in the US as individual states re-open may 4, 2020. medRxiv, 2020.05.04.20090670.

Kucharski A, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID- 19: a mathematical modelling study. Lancet Infec Dis. (2020) 20:553–
8. doi: 10.1016/S1473-3099(20)30144-4

Yadav, Subhash Kumar, & Akhter, Yusuf. (2021). Statistical Modeling for the Prediction of Infectious Disease Dissemination With Special Reference to COVID-19 Spread. Frontiers in Public Health, 9, 645405.

Chretien, J.-P., Riley, S., & George, D. B. (2015). Mathematical modeling of the West Africa Ebola epidemic. eLife, 4, http://dx.doi.org/10. 7554/eLife.09186.

Nsoesie, E.O., Brownstein, J.S., Ramakrishnan, N., Marathe, M.V., (2014). A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza Other Respir. Viruses 8 (3), 309–316.

Dengue Forecasting. (2017). http://dengueforecasting.noaa.gov/.

Chikungunya Forecasting. (2017). http://www.darpa.mil/news-events/2015-05-27.

Goldstein, E., Cobey, S., Takahashi, S., Miller, J.C., Lipsitch, M., (2011). Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: a statistical method. PLoS Med. 8 (7), e1001051.

Wallinga, J, & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings of the Royal Society. B, Biological Sciences, 274(1609), 599-604.

Anderson, R., & May, Robert M. (1991). Infectious diseases of humans : Dynamics and control (Oxford science publications). Oxford ; New York: Oxford University Press.

Nouvellet, Pierre, Cori, Anne, Garske, Tini, Blake, Isobel M, Dorigatti, Ilaria, Hinsley, Wes, . . . Riley, Steven. (2017). A simple approach to measure transmissibility and forecast incidence. Epidemics, 22, 29-35.

Wallinga, Jacco, & Teunis, Peter. (2004). Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. American Journal of Epidemiology, 160(6), 509-516.

Fraser, Christophe, Ferguson, Neil M, Iamsirithaworn, Sopon, Riley, Steven, Cauchemez, Simon, Burke, Donald S, . . . Meeyai, Aronrag. (2005). Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature, 437(7056), 209-214.

Fraser, Christophe. (2007). Estimating individual and household reproduction numbers in an emerging

epidemic. PloS One, 2(8), E758.

CORI, Anne, FERGUSON, Neil M, FRASER, Christophe, & CAUCHEMEZ, Simon. (2013). A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. American Journal of Epidemiology, 178(9), 1505-1512.

WHO Ebola Response Team, 2015a. West African Ebola epidemic after one year—slowing but not yet under control. N. Engl. J. Med. 372 (6), 584–587.

Coronavirus cases in Australia - covid live. https://covidlive.com.au/

Zhang, Xingyu, Zhang, Tao, Young, Alistair A, & Li, Xiaosong. (2014). Applications and comparisons of four time series models in epidemiological surveillance data. PloS One, 9(2), E88075.

Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. (1990). STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics, 6(1), 3–33. http://bit.ly/stl1990

Nishiura, Hiroshi, Linton, Natalie M., & Akhmetzhanov, Andrei R. (2020). Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases, 93, 284-286.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting, 16(4), 437–450.

Vaccination numbers and statistics. Department of Health and Aged Care. Australian Government. https://www.health.gov.au/initiatives-and-programs/covid-19-vaccines/numbers-statistics#