

2021

What effect do government policies have on the transmission of COVID-19?

Prestige, E.

Prestige, E. (2021) 'What effect do government policies have on the transmission of COVID-19?', *The Plymouth Student Scientist*, 14(2), pp. 356-428.

<http://hdl.handle.net/10026.1/18509>

The Plymouth Student Scientist
University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

What effect do government policies have on the transmission of COVID-19?

Emily Prestige

Project Advisor: Dr. Yinghui Wei, School of Engineering, Computing and Mathematics, University of Plymouth, Drake Circus, Plymouth, PL4 8AA

Abstract

This investigation aimed to determine the effect of government policies on the transmission of COVID-19 during the period 30/01/2020 – 03/01/2021. This investigation used qualitative data from policy legislation and journals and quantitative data from official government pandemic figures. It also used central factors of successful government approaches to assess the efficacy of the UK Government strategy. The measure used to monitor transmission rates is the instantaneous reproduction number R_t . I have used summary statistics, data visualisation, and time series representations to conduct exploratory data analyses. I also investigated the change in testing capacity and positivity rate over time to account for factors impacting the number of individuals testing positive for COVID-19. To estimate R_t , I used a deterministic Susceptible-Infected-Removed model and a stochastic epidemic model. I then used time series models to predict the positivity rate, transmission rate, and the number of new cases for the following week. From estimating R_t , we see that the rate of transmission fell during national lockdowns. Predictions showed that the positivity rate, transmission rate, and the number of new cases would increase if conditions had remained the same, i.e. if the government had not implemented additional mitigation strategies. Lastly, I compared the approaches taken in the UK with those in New Zealand and Brazil. From this, I was able to identify the possible influence of government policies on transmission. This investigation found that when governments communicated policies clearly, they had a more notable effect on reducing R_t . Furthermore, I have identified several areas in which the government could have improved to increase the effectiveness of policies on reducing transmission rates.

Keywords: mathematics, statistics, equations, COVID-19, transmission

1 Introduction

This investigation aimed to determine the effect government policies had on the transmission of COVID-19. COVID-19 is a disease caused by the virus “severe acute respiratory syndrome coronavirus 2”¹ (SARS-CoV-2). The name was derived as such: CO – corona, VI – virus, D – disease, and 19 – year of the first recorded case². COVID-19 is transmitted via droplet and airborne transmission. Droplet transmission occurs when a susceptible individual is near an infected individual through both large and small droplets that contain the virus. Airborne transmission occurs over larger distances and via smaller droplets and particles of the virus that can hang in the air for extended time periods³. The disease impacted life globally, and a widely discussed question was how to slow its spread. Throughout the pandemic, the scientific community found effective methods to minimise infection rates. However, the implementation of these methods was widely dependent on political leadership and public response. During many discussions throughout the pandemic, public figures used the “*R number*” to describe how quickly the disease was spreading. Other common references to transmission rates were to the value of R_0 , the basic reproduction number. However, they were usually describing R_t , the instantaneous reproduction number⁴. The measure R_0 describes the reproduction potential⁵ of a disease. This potential is the average number of individuals an infected individual would infect, under the assumption that everyone was susceptible to the disease⁶. Whereas the measure R_t describes reproduction at time t , that is, the average number of individuals an infected individual would go on to infect if conditions remained as they were at time t ⁷. Throughout this investigation, I have used R_t to illustrate the current speed at which the disease was spreading. When $R_t < 1$ spread was slowing, and when $R_t > 1$ spread was accelerating.

Governments all over the world implemented a variety of policies over the course of the pandemic. Because of this, it was possible to find commonalities in those that have had success. Researchers found that countries that successfully navigated the pandemic had certain characteristics in common when it came to their approaches⁸, which were:

1. A high level of confidence in leadership;
 - Messaging towards the public was clear and policies put in place were consistent, transparent, and their motivations were well communicated.
2. Accurate methods for monitoring transmission;
 - Inaccuracies which arose from lag were minimised.
3. Effective test, trace, and isolate systems;
 - Not only were infected people successfully traced, but they also adhered to isolation policies.
4. Adequate healthcare systems;
 - The settings were well staffed.
 - Staff were well protected from the virus.
 - There were enough facilities to treat patients.

5. Border controls;

- The flow of people into the country was being well managed.

I compared these approaches to those of the UK to assess the efficacy of COVID-19 mitigation strategies. Firstly, I outlined the policies which were implemented in the UK.

1.1 Key Government Policies

On the 16th March 2020, Prime Minister Boris Johnson gave his first speech regarding COVID-19, which detailed the social distancing measures that would be in place from the next day⁹. A week later, on the 23rd March 2020, a lockdown was implemented – where people in the UK were required to stay at home unless they were completing essential tasks¹⁰. The policy was clear and had “Stay home, Protect the NHS, Save Lives” as a slogan¹⁰, which effectively communicated its motivation. The next influential moment was when the lockdown was eased on the 15th June 2020, allowing all non-essential businesses to reopen. England was the only country that began easing restrictions, with Scotland, Wales, and Northern Ireland extending their lockdowns¹¹. England then eased its lockdown further on the 4th July 2020¹¹. However, we did not see restrictions eased any more during the study period.

Additionally, regional policies were implemented in areas seeing higher amounts of new infections¹². Following the implementation of regional restrictions, a three-tier system was introduced across the UK on the 14th October 2020¹³. These tiers determined which restrictions were in place, and the government assessed the tiers every two weeks¹⁴. Changes across the four countries in the UK then came when Northern Ireland implemented a “circuit breaker” lockdown on the 16th October 2020¹⁵, which ended on the 11th December 2020¹⁶. The lockdown was originally lifted on the 20th November 2020, but officials reimplemented it a week later¹⁷. Wales then implemented a “firebreak” lockdown on the 23rd October 2020¹⁸ which was lifted on the 9th November 2020. Whilst it did not enter a lockdown, Scotland introduced a separate five-tier system on the 2nd November 2020¹⁹, which remained in place during the study period. Lastly, England entered a four-week lockdown on the 5th November 2020²⁰ which was lifted on the 3rd December 2020. The final key policy seen was the fourth tier introduced across England, Wales, and Northern Ireland on the 20th December 2020, several areas across the three countries subsequently entered the new tier²¹.

1.1.1 Confidence in Leadership

Confidence in leadership plays a key role when it comes to slowing the spread of disease. According to Professor Duffy, director at the Policy Institute at King’s College London, “Trust in authority is key to maintaining compliance”²². When a population doesn’t trust their government, they will not engage with the policies they implement. A contributing factor towards public confidence is consistency in government policies. Policy consistency between the different countries in the UK became a problem when the countries began adopting different approaches. As a result, regulations varied depending on the country. Travel between the countries was hard to control due to there being no hard borders in the UK, and because of this, infections were harder to trace. These issues became more apparent when different countries implemented lockdowns at

separate times and implemented distinct restriction policies (see Section 1.1). Another area where consistency was lacking was the reopening of schools across the UK. In addition to schools reopening at different stages, they had independent policies in place regarding attendance. For example, schools in England were the last to reopen on the 1st September 2020, but had the strictest attendance policy. The policy stated that only those who lived with someone who tested positive for COVID-19 were exempt from attending in person²³. The schooling issues became worse when infections rapidly rose in December 2020, causing many to question whether schools should reopen after the Christmas holiday. A study subsequently found that 3% of secondary school children had the virus, with primary school children not far behind with 2% infected²⁴. This meant that children in school had the highest infection rates of any age group above 25 years, causing parents to become very concerned about safety. Despite advice from scientists that overall transmission would be lower if schools remained closed, most schools in England reopened²⁵. Aside from schools in areas of high transmission rates, which were allowed to remain closed. Notable examples of inconsistency were the policies implemented regarding Christmas celebrations. While the four countries originally planned to have one policy across the UK, this changed when the number of infections began to rise drastically. Each country took a different approach, announcing their new policies only two days before the holiday²⁶.

Clear messaging and policy transparency are huge influences on public confidence. Governments can achieve these by providing uncomplicated advice and explanations of what their strategies aim to achieve. However, as guidance began to change across the UK, many people expressed concern about messaging being unclear and undermining recovery efforts²⁷. An area in which clear messaging was flawed was the guidance on physical distancing. On the 17th March 2020, the government advised people to stay a “safe distance” from one another²⁸. This distance was not specified until almost a week later, with advice stating that people should stay two metres apart from one another²⁹. After stressing the importance of physical distancing, the government then relaxed the guidance to “one metre plus” – with the plus being additional measures. The intention was to increase the number of settings that could reopen, such as hairdressers, but the policy only encouraged the public to relax their attitudes. Some guidance that was missing transparency was that on face-coverings. After dismissing their efficacy, the government implemented a face-covering policy on the 24th July 2020. The policy made it mandatory to wear a face-covering in most indoor shops, public spaces, and public transport³⁰. Due to the change in messaging, many questioned why face-coverings were now required. Also, the existence of exemptions encouraged the public to ignore the guidance entirely. These exemptions specified that people who did not have to wear face-coverings were: children under eleven, those with a physical or mental illness and or disability, and those to whom it may cause distress³¹. The vague nature of these exemptions meant that many people were not wearing face-coverings and that those who were, did so inconsistently.

Independent polls run by The Health Foundation in July 2020 and King’s College London in November 2020 found that the British public had lost confidence in the government’s handling of the pandemic. The Health Foundation found that 56% of participants felt “the government has not handled COVID-19 well”³². King’s College London found that 57% of

participants “did not trust the government to control the spread of COVID-19”²². King’s College London also found that almost seven out of ten participants felt “the government response has been confused and inconsistent”²². Additionally, University College London found that the number of people in England who had “no confidence at all” in the governments handling of the pandemic rose from 6% to 27% from March 2020 to September 2020³³. Launched the week before the March 2020 lockdown began, this study followed over 70,000 participants throughout the pandemic. The study also found that when they compared responses from participants in England, Wales, and Scotland, the number of people with “no confidence” was lower in Wales and Scotland. They found that, in September 2020, 6% of Welsh participants had “no confidence” in the Welsh Government, and 10% of Scottish participants had “no confidence” in the Scottish Government³³.

1.1.2 Methods for Monitoring Transmission

Monitoring transmission is vital when it comes to successful mitigation strategies. If we can accurately know the transmission rate at a given time, we can assess what mitigation strategies are necessary. However, accurate monitoring requires extremely up-to-date data. Due to guidance given throughout the study period, individuals were often not tested until they were symptomatic, which prevented data from being up-to-date. The time delay between becoming infected and showing symptoms – and hence getting tested – meant that infected individuals would unknowingly spread the virus. Testing delays became more concerning when research into pre-symptomatic and asymptomatic transmission was considered. Namely, that asymptomatic³⁴ and pre-symptomatic³⁵ transmission could occur. Also, an individual’s infectiousness peaked before the onset of symptoms³⁵. The only way to mitigate the delay would be by testing everyone across the country regularly. Since this was difficult to implement, it was important to trace the contacts of infected individuals. Doing so meant that individuals exposed to the virus could isolate, preventing further transmission. Without mass testing and comprehensive tracing, it was impossible to monitor transmission accurately.

One policy that influenced the ability to monitor transmission was the introduction of lateral flow tests. With results found within one hour, lateral flow tests allowed communities to identify more cases than they otherwise would have been able to. When the scheme began to roll out in November 2020³⁶, the aim was to increase the ability to interrupt transmission chains. However, there was much contention as to whether these lateral flow tests were appropriate for mass testing schemes. Many cited the fact that the tests were not designed for use on asymptomatic individuals and that both false negatives – and positives – were found to occur³⁷. The main flaw with the scheme was the lack of clarity regarding how individuals should interpret test results. If an individual tested positive, they should have self-isolated and undertaken a PCR test (see Section 1.1.3). However, if they tested negative, they should have continued to follow guidelines and behaved as if they may have COVID-19. Understanding this was important due to findings suggesting that lateral flow tests could miss up to 50% of positive cases³⁸.

A crucial part of the UK’s monitoring of transmission was the genomic sequencing conducted by scientists across the nation³⁹. Genomic sequencing allowed us to monitor changes in the strains of COVID-19, eventually leading to the identification of a notable

new variant. In early December 2020, scientists found evidence of a variant spreading in Kent, causing case numbers to soar⁴⁰. After genomic sequencing, it was possible to discover that cases of this variant had been circulating as early as the 20th September 2020⁴¹. Because of this knowledge, additional measures were devised, such as the introduction of tier four (see Section 1.1). Researchers found that the “Kent variant” has a rate of transmission approximately 70% higher than the original strain of COVID-19⁴². This discovery provided some explanation for why cases were rapidly increasing, regardless of the measures in place.

1.1.3 Test, Trace, and Isolate System

As mentioned in Section 1.1.2 an important aspect of monitoring, and hence slowing, transmission is an effective find, test, trace, and isolate system. Contact tracing and isolation are vital when testing is not widespread. The NHS Test and Trace system was regularly criticised due to repeated evidence of design flaws. Research suggested that less than 20% of people in England self-isolated fully when instructed to do so⁴³. A study⁴⁴ corroborated this research, finding:

- There were 333,900 new infections in the first week of November 2020.
- Of these, 141,804 (42%) were identified and passed to the Test and Trace system.
- From the people identified; 99,212 (70%) provided details for 314,817 close contacts, 21,300 (15%) were reached but did not provide any contacts, and 21,292 (15%) were not reached.
- Of the close contacts provided, 190,129 (60%) were reached and asked to self-isolate.
- NHS Test and Trace estimated only around 50% of those contacted self-isolated fully.

It was possible to calculate that for every person who gave details, tracers identified three contacts. Therefore, for the 333,900 infected individuals, there were roughly 1,001,700 close contacts. If approximately 50% of the close contacts reached, self-isolated fully (95,065 people), then approximately only 9.5% of potentially infectious individuals were staying home. A factor contributing towards a lack of adherence to self-isolation policies was the financial ability to do so. Evidence suggested that the ability to self-isolate was three times lower in those who have a lower socio-economic status and that many were not eligible for the support which was on offer⁴³. The government made efforts to increase the number of people self-isolating by reducing quarantine to ten days⁴⁵. This change applied to those who: tested positive for COVID-19, developed symptoms of COVID-19, had been in close contact with someone who tested positive for COVID-19, or had returned from a country not on the travel corridor list (see Section 1.1.5).

A large part of the test, trace, and isolate system is the testing. Throughout the study period, it was possible to get a test either privately for a cost or via the NHS for free. Whilst testing capabilities had improved since the beginning of the pandemic, NHS testing capacity was still limited.

Testing availability increased with the introduction of lateral flow tests in November 2020³⁶. The mass testing scheme aimed to interrupt transmission chains but quickly became a

point of contention (see Section 1.1.2). Those involved in the trial of mass testing in Liverpool defended their use and felt they were an overall success⁴⁶. Professor Buchan, executive dean at the Institute of Population Health at the University of Liverpool, stated that the “fear of not having enough financial support in isolation was the biggest perceived barrier for testing”⁴⁶. He also added that “the main conclusion I would draw is trust local communities to self-organise around the smart testing approach”⁴⁶.

The two forms of testing which were on offer were Polymerase Chain Reaction (PCR) tests and lateral flow tests. The difference is that PCR tests search for the genetic material of the virus and lateral flow tests search for virus proteins⁴⁷. PCR tests are viewed as the ‘gold standard’ because they hold the most certainty, but they take up to three days to provide results⁴⁷. Lateral flow tests provide results in about 30 minutes, but they are less likely to detect infection and cannot confirm that an individual does not have COVID-19⁴⁷ (see Section 1.1.2).

The following individuals were eligible⁴⁸ for a free PCR test:

- People with COVID-19 symptoms; high temperature, new continuous cough, loss or change to sense of smell or taste.
- Care home residents or staff.
- People who are going into hospital.

1.1.4 Healthcare System

An adequate healthcare system can be the difference between having and losing control over the virus. Once a country’s healthcare system is overwhelmed, things become markedly harder to manage. Throughout the pandemic, the NHS had many struggles. These struggles included thousands of staff being brought back from retirement⁴⁹ and massive shortages of protective equipment at the start of the pandemic, causing staff to feel unsafe⁵⁰. An important thing to note is that hospital capacity in the UK was lower than that of many wealthy countries⁵¹. This hospital capacity meant that when the number of critical care patients rose, the healthcare system was quickly saturated. The motivation for many policies was to conserve NHS resources, and often policies were implemented when the system was close to being overwhelmed.

Throughout the pandemic, doctors across the world have tested treatment options for COVID-19 patients. The World Health Organisation ran one of the “largest international randomised trials” with almost 12,000 patients based in 500 hospitals in over 30 countries⁵². This trial found that corticosteroids had been the only effective treatment so far⁵². Other treatment discoveries have been the use of breathing support mechanisms, such as oxygen therapy, continuous positive airway pressure (CPAP), and invasive mechanical ventilation (IMV)⁵³. An example of how treatments have changed is how at the beginning of the pandemic, doctors were ventilating patients earlier, and only one-third of ventilated patients survived⁵³.

1.1.5 Border Control

Border control is paramount in controlling the spread of a virus as it can prevent additional cases from being imported. At the beginning of the pandemic, border control in the UK was minimal. However, there were several policies enacted which impacted travelling to and from the UK. The predominant impact on travel was the initial lockdown, which prevented people from travelling out of the UK unless essential. Another policy impacting travel specified that travellers entering the UK must self-isolate for two weeks⁵⁴ and was implemented on the 8th June 2020. On the 3rd July 2020, the UK Government introduced “travel corridors”. The government created travel corridors to stimulate the damaged economy. However, they allowed people into the UK without the need for self-isolation, given that they were entering from a pre-approved country⁵⁵. As of January 2021, individuals were not being tested upon arrival at airports, meaning that even those coming from countries deemed safe posed a sizeable risk of spreading the disease. When imported, cases were difficult to trace, and they presented a higher risk than local community transmission.

2 The Data

Throughout this investigation, I have used a combination of qualitative and quantitative data. The qualitative data is the government policy information, opinion pieces, and scientific journals. The quantitative data is the pandemic and testing data. Government policy information details the policies, when the policies began, and when the policies ended – or how long the policies were in place. The pandemic data is a combination of datasets available on the official government dashboard⁵⁶ and contains information from 30.01.2020 to 03.01.2021 – the period of study for this chapter. I defined the beginning of the study period as the date of the first recorded case of COVID-19 in the UK and the end of the study period as the date preceding the UK Government announcement of a third lockdown. The testing data is also a combination of datasets available on the official government dashboard⁵⁶ but contains information from 21.04.2020 to 03.01.2021 – as testing data was not available until the 21st April 2020.

2.1 Pandemic Data

For each date, the data contains the corresponding number of confirmed⁵⁷ and the number of dead⁵⁸. Where “confirmed” is someone who had tested positive with COVID-19 and the corresponding date is when the individual gave the sample. Here “dead” are those who had COVID-19 registered on their death certificate as one of the causes. Another important piece of information required for this investigation was the number of people susceptible to COVID-19 before the first infection in the UK. Throughout, “susceptible” describes those who had not yet contracted COVID-19 and are therefore vulnerable to it. To get the most accurate estimation of this figure before the first infection, I have used a mid-2019 population estimate from the Office for National Statistics⁵⁹. I then used this information within “for” loops in R. These loops calculate three useful variables, the current number of susceptible individuals, the number of currently infected, i.e. the number of infected individuals at a given time, and the number of recovered, i.e. people who tested positive more than 28 days ago. After I had this

information, I cumulated the number of recovered and edited the data set. This leaves the following, with 333 observations:

- Date
- Confirmed
- Dead
- Recovered
- Susceptible
- Currently infected
- Cumulative confirmed
- Cumulative dead
- Cumulative recovered

An infectious period is how long an infected person can infect others. For COVID-19, researchers estimated this to be up to 10 days from the onset of symptoms, which can develop up to 14 days after infection⁶⁰. Additionally, researchers found that transmission could occur before symptom onset³⁵ and from those who were not showing symptoms³⁴. The government guidance throughout the study period on estimating COVID-19 deaths states that a person had died of COVID-19 if they died within 28 days of their first positive test result⁶¹. Using this information about transmission and death, I defined the time taken to be recovered as 28 days – until which point we classed the individual as currently infected. For this investigation, I have assumed that recovered individuals were immune to catching COVID-19 again. It was still unknown how long immunity would last, and while there had been cases of reinfection⁶², it was unknown how frequently this occurred. It will likely take many years for the scientific community to understand the immune response to COVID-19. For these reasons, I have assumed that reinfection will be negligible – and hence can be ignored.

2.2 Testing Data

For each date, the data contains the corresponding number of tests conducted⁶³, number of positive results⁵⁷, and PCR testing capacity⁶⁴. The number of tests conducted is the “number of confirmed positive, negative or void COVID-19 virus test results”⁶⁵. This includes both PCR tests conducted in laboratories and lateral flow tests conducted outside of laboratories. Here the number of cases is the number of positive results by the date the sample was taken. The PCR testing capacity is “a projection based on reports from labs on how many lab-based tests they have capacity to carry out each day based on availability of staff and resources”⁶⁶. It does not reflect the capacity for lateral flow tests, which became more readily used since their national roll out on the 10th November 2020³⁶.

Using the number of cases and the number of tests conducted, I calculated the rate of test positivity for the associated date using the equation as follows:

$$\text{positivity rate} = \frac{\text{number of cases}}{\text{number of tests conducted}}$$

To contextualise this result I also expressed the positivity rate as a percentage.

3 Exploratory Data Analyses

For this investigation, I have conducted exploratory data analyses of the pandemic and testing data. I have used data visualisation for both the pandemic data (described in Section 2.1) and the testing data (detailed in Section 2.2). Additionally, I have also used summary statistics and time series representations for the pandemic data.

3.1 Summary Statistics – Pandemic Data

Summary statistics are produced using the "summary" function⁶⁷ and are shown in Table 1. These statistics are the: minimum, first quartile, median, mean, third quartile, and maximum. The minimum and maximum intuitively tell us the minimum and maximum values taken in the data. The more relevant values are the upper and lower quartiles, which exclude extreme data values, and can be used to calculate the inter-quartile range (IQR). The IQR can be used to describe the dispersion of data, also known as the spread. Finally, there is the mean and the median, describing the average values taken in the data. The mean is the sum of values in a sample, divided by the sample size. The median is the middle value of the data once it is in size order. For symmetrically distributed data, the mean is a good way to describe the average. For any other data, we should use the median⁶⁸.

Table 1 shows that the data had a large IQR, which implies it had a large dispersion. Large dispersion indicates that values were very high – or rising – at the end of the study period because as values fall, the IQR becomes smaller. The IQR gets smaller because high values at peaks become outliers and hence don't get included in the upper quartile. We can also see the mean and medians of the data were not close together, indicating an absence of symmetry.

Table 1: Summary statistics for daily pandemic data

Data Type	Min	Q1 ^a	Median	Mean	Q3 ^b	Max	IQR
Confirmed	0	772	2863	8249	13298	81542	12526
Dead	0	22	118.5	283.9	461.2	1457	439.2
Currently infected	2	72172	91562	214319	321172	1113615	249000

^a Q1 is the first (lower) quartile

^b Q3 is the third (upper) quartile

3.2 Data Visualisation – Pandemic Data

To identify patterns in the data and motivate the investigation, I have presented the pandemic data through data visualisation – which I computed in R. Figure 1 shows the number of confirmed cases each day, with each red point being a daily observation. We can see that after September 2020, the daily incidence showed an upward trend, which

peaked in November 2020. Following a brief fall in cases, the number of confirmed cases began to rise in December 2020 and continued to do so for the remainder of the study period.

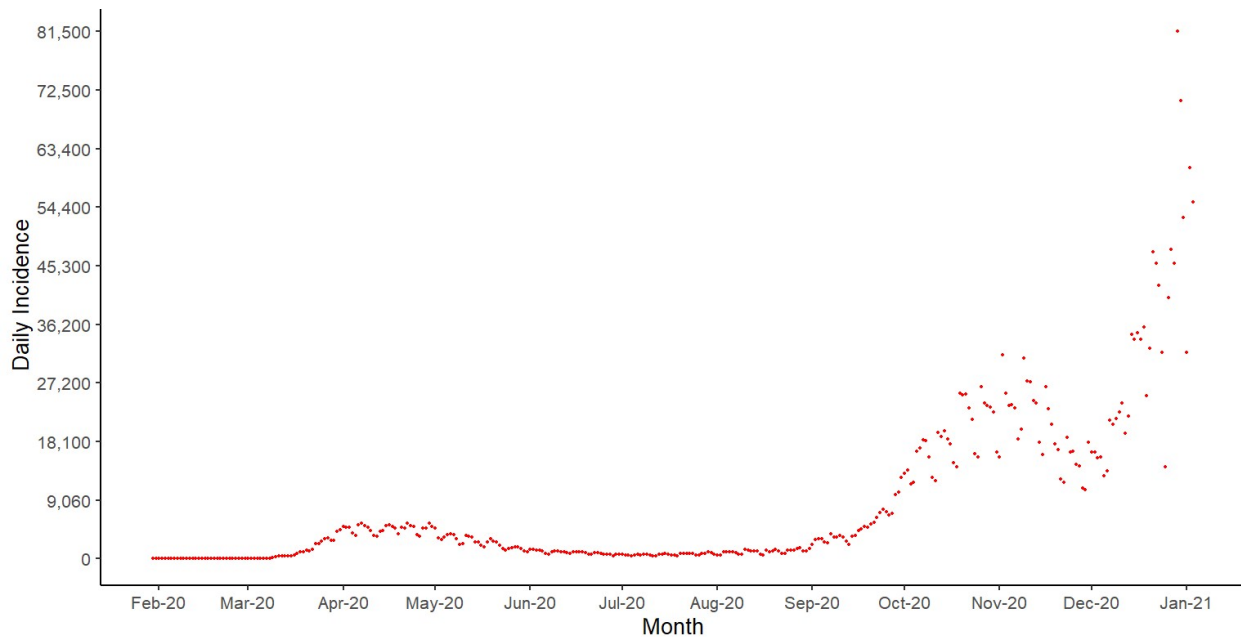


Figure 1: Scatter plot of the daily incidence of confirmed cases, where red points signify daily observations. Plotted using the "ggplot" function⁶⁹ from the "ggplot2" package⁷⁰.

Figure 2 used the same information as Figure 1 but presented it differently. Visualising the incidence this way can illustrate the difference in incidence each day more clearly. For example, we see that from December 2020, the daily incidence increased and had a comparatively steep upward trend.

Figure 3 shows the number of infected people at any given point in time, where red points signify daily observations. The number of infected people is an indication of whether the rate of transmission is growing or decaying. When the transmission rate is decaying, the number of currently infected will be falling. However, when the transmission rate is growing, the number of currently infected will be rising. We can see that the number of currently infected people did fall in both June 2020 and November 2020, meaning that the rate of transmission was decaying. We can also see that from July 2020 to September 2020, the number of currently infected remained steady at approximately 124,000. However, we see that the number began to rise in September 2020 and peaked in November 2020. We then see a dip in the number of currently infected at the end of November 2020, followed by a sharp increase from December 2020 into January 2021.

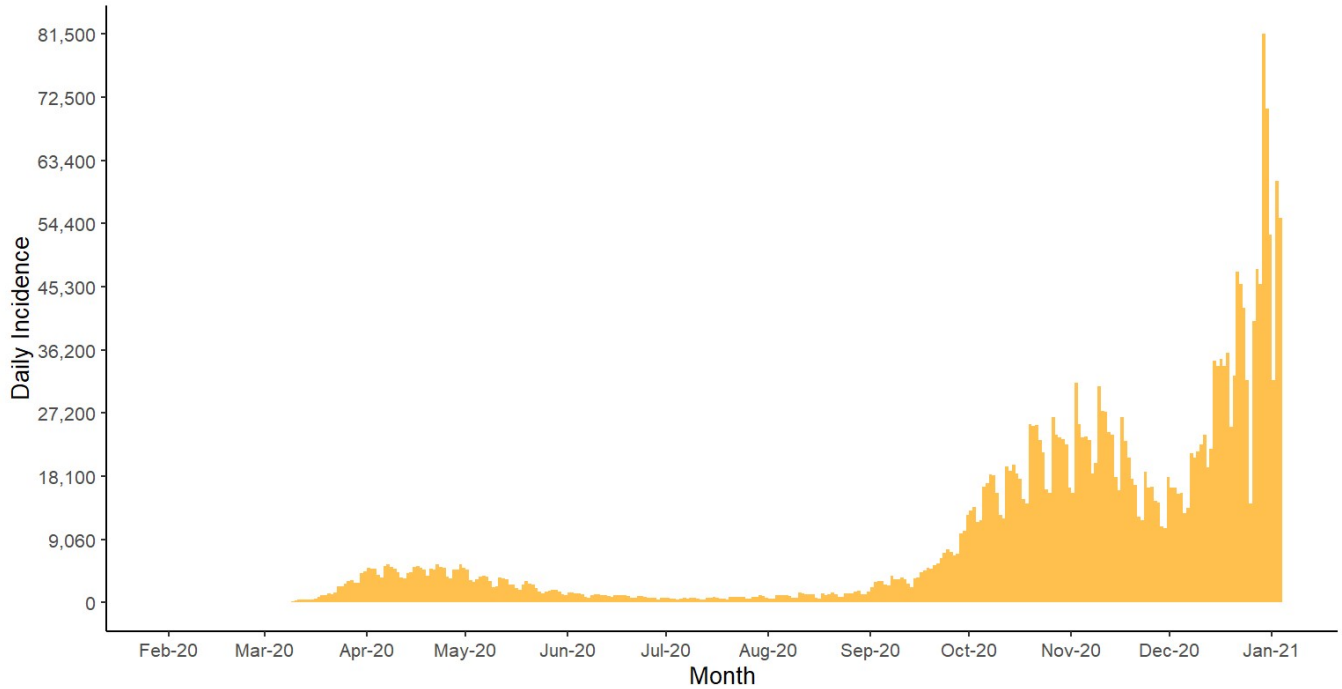


Figure 2: Incidence plot of the daily confirmed cases. Plotted using the "plot" function⁷¹ from the "graphics" package⁷². Data manipulated using the "uncount" function⁷³ from the "tidyr" package⁷⁴ with weights = 'confirmed' and the "incidence" function⁷⁵ from the "incidence" package⁷⁶.

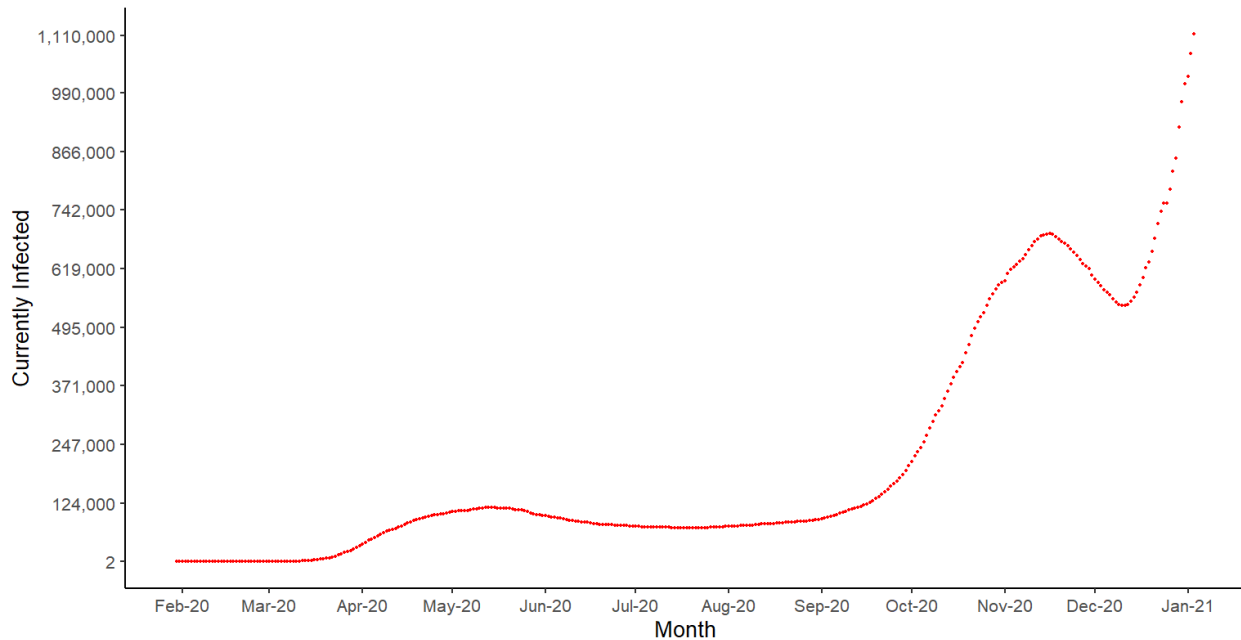


Figure 3: Scatter plot of the daily currently infected over time, where red points signify daily observations. Plotted using the "ggplot" function from the "ggplot2" package.

3.3 Time Series Representations – Pandemic Data

To visualise the data separately, I have represented the pandemic data as time series. I have also used smoothing to extract the underlying trend in the confirmed cases.

3.3.1 Separated Time Series

Figure 4 shows the separated time series of the pandemic data. Visualising time series in this manner is a way to compare different patterns, especially between data on different scales. For example, if I plotted the currently infected or susceptible with the other data, it would be hard to compare them. This difficulty is because the number of currently infected was much higher than that of the other time series, and the number of susceptible was higher still. Figure 4 illustrates the similar patterns of the confirmed, dead, recovered, and currently infected. We see there was a dramatic spike in deaths in April 2020, which peaked at 1460, after which deaths remained low until November 2020. Notably, the proportion of deaths to confirmed cases in April 2020 was much higher than at any other point in the study period. The higher proportion of deaths may have been because we had fewer treatment options, PPE, and trained staff members at the beginning of the pandemic (see Section 1.1.4). We can also see that the number of susceptible individuals decreased as the number of confirmed, dead, recovered, and currently infected increase. We describe this as an inverse relationship.

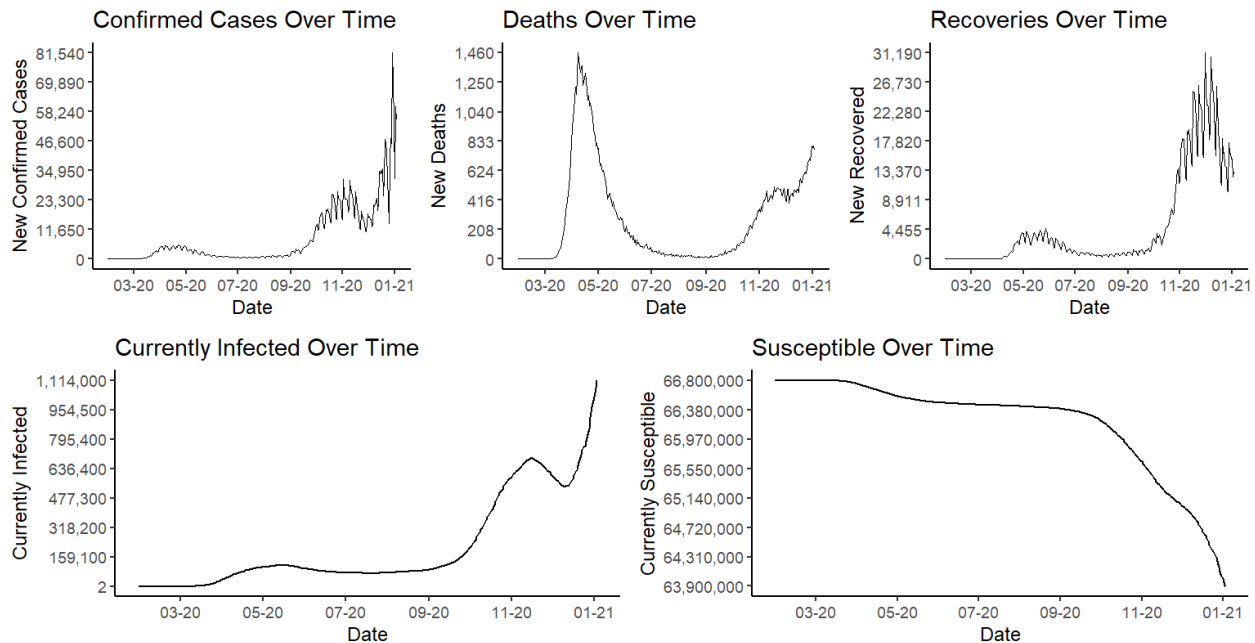


Figure 4: Plots of the individual time series of COVID-19 daily data. The top left plot shows the time series of confirmed cases, the top middle plot shows the time series of deaths, the top right plot shows the time series of recoveries, the bottom left plot shows the time series of currently infected, and the bottom right plot shows the time series of susceptible. Computed using the "ts" function⁷⁷ from the "stats" package⁷⁸, plotted using the "ggplot" function from the "ggplot2" package, and arranged using the "ggarrange" function⁷⁹ from the "ggpubr" package⁸⁰.

3.3.2 Smoothed Time Series of Confirmed Cases

Figure 5 shows the smoothed time series of the confirmed cases, computed by applying a simple moving average (SMA) to the data. A simple moving average is a type of linear filter that takes the arithmetic mean over the past n observations, where n is defined as the period⁸¹ – here in days. Different values of n apply different levels of smoothing to the time series, allowing trends to become more apparent. After smoothing, the upward trend of infections toward the end of the study period became more evident than before. When $n = 12$ (*days*), we have entirely smoothed out the volatile nature seen at the end of December 2020 to show the underlying upward trend of confirmed cases that was present.

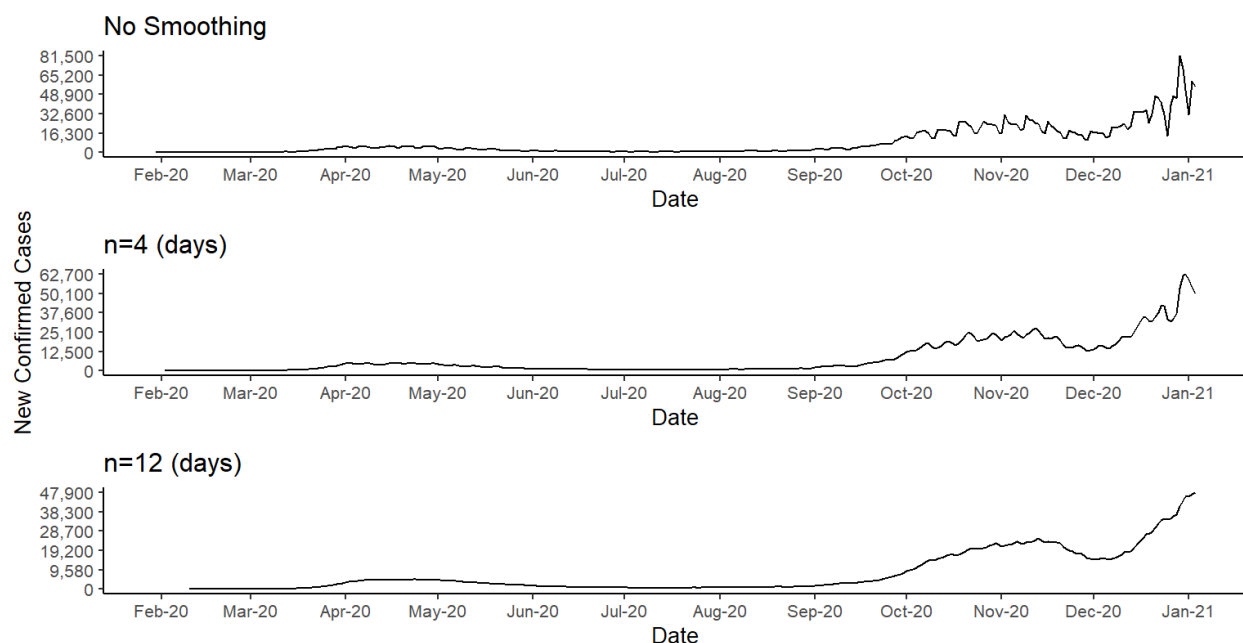


Figure 5: Plots of the smoothed time series of daily confirmed cases. The top plot shows the original time series of the daily confirmed cases, the middle plot shows the time series smoothed with $n = 4$ (*days*), and the bottom plot shows the time series smoothed with $n = 12$ (*days*). Computed using the "as.xts" function⁸² from the "xts" package⁸³, smoothed using the "SMA" function⁸¹ from the "TTR" package⁸⁴, plotted using the "ggplot" function from the "ggplot2" package, and arranged using the "ggarrange" function from the "ggpubr" package.

3.4 Data Visualisation – Testing Data

To provide a contextual understanding of the pandemic, I have used data visualisation for the testing data. From the testing data, we can understand three aspects of testing in the UK: testing capacity, tests conducted, and rate of positivity.

3.4.1 Testing Capacity

Figure 6 illustrates that testing capacity increased between May 2020 and January 2021. An increased capacity could account for the increase in positive test results we saw in November 2020 and December 2020 (see Figure 1 and Figure 2), as this was when there had been notable increases in capacity.

3.4.2 Tests Conducted

Figure 7 shows there was a general upward trend in the number of tests conducted daily. Whilst this increase could partially have been accounted for by the increased testing availability, it also suggests that there was an increase in those experiencing symptoms. We can also see that there was variability in the number of tests conducted each day, which suggests an element of seasonality. That is, the number of tests conducted depended on the day of the week.

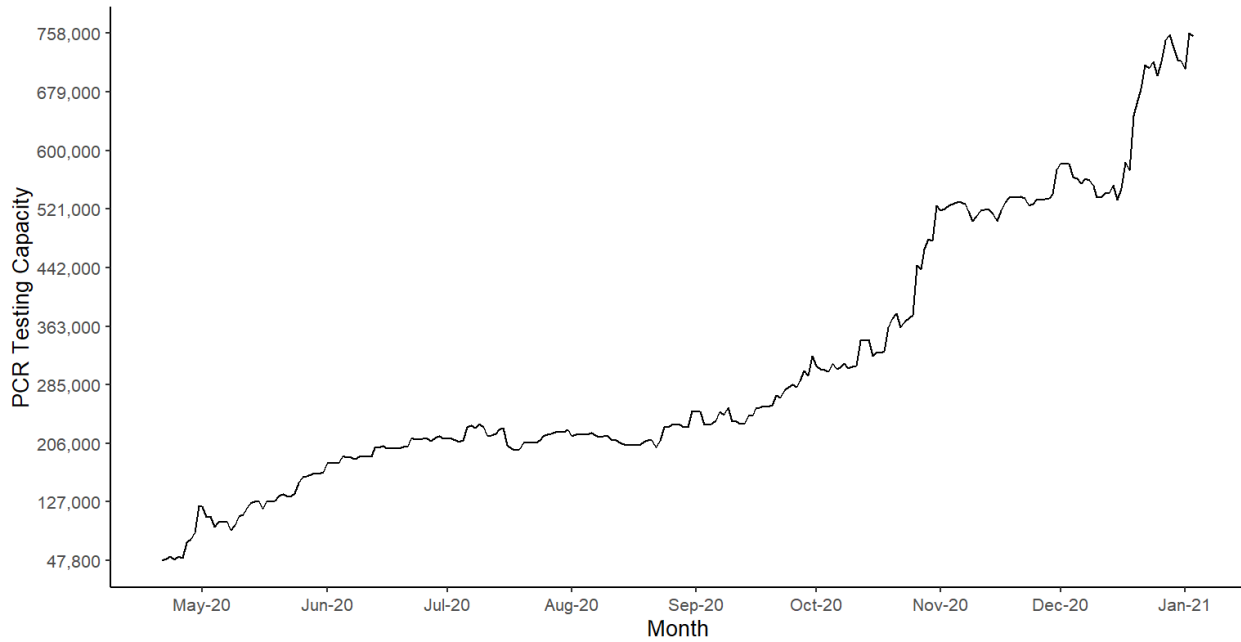


Figure 6: Plot of the daily PCR testing capacity over time. Plotted using the "ggplot" function from the "ggplot2" package.

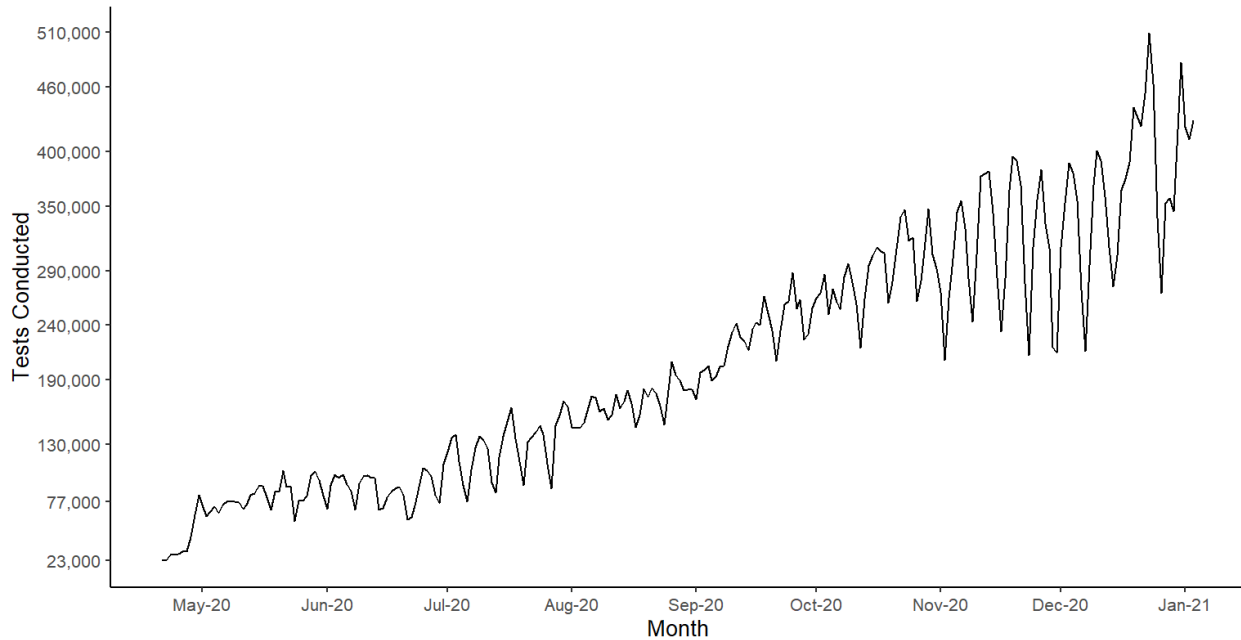


Figure 7: Plot of the daily number of tests conducted over time. Plotted using the "ggplot" function from the "ggplot2" package.

3.4.3 Rate of Positivity

Figure 8 shows that the positivity rate was high at the start of the data, reaching 0.24 in April 2020. When we compare Figure 8 to Figure 6 and Figure 7 we can see that the high rate related to the lower testing availability and the smaller number of tests conducted at the beginning of the pandemic. From this, we can assume that people who undertook tests suspected they were infected. Figure 8 shows that the positivity rate was falling until September 2020, where it began to rise. Throughout, we see there was volatility in the rate of positivity which became more pronounced from October 2020. We can see there was a positive correlation between the volatility and the variability described in Section 3.4.2. When we compare Figure 8 with Figure 2, we can see that from September 2020, both figures exhibited the same pattern. We conclude that the increasing rate of people who tested positive was not due to the increased availability of testing (see Figure 6). If the increased number of positive cases were related to an increase in testing capacity, the positivity rate would have decreased – as seen in May 2020.

newline

When we look at the test positivity rate as a percentage, we see that in May 2020, 24% of COVID-19 tests were positive. When percentage positivity was at its lowest, in July 2020 and August 2020, only 0.35% of tests were positive. Toward the end of the study period, the percentage of tests that were positive was approximately 10%.

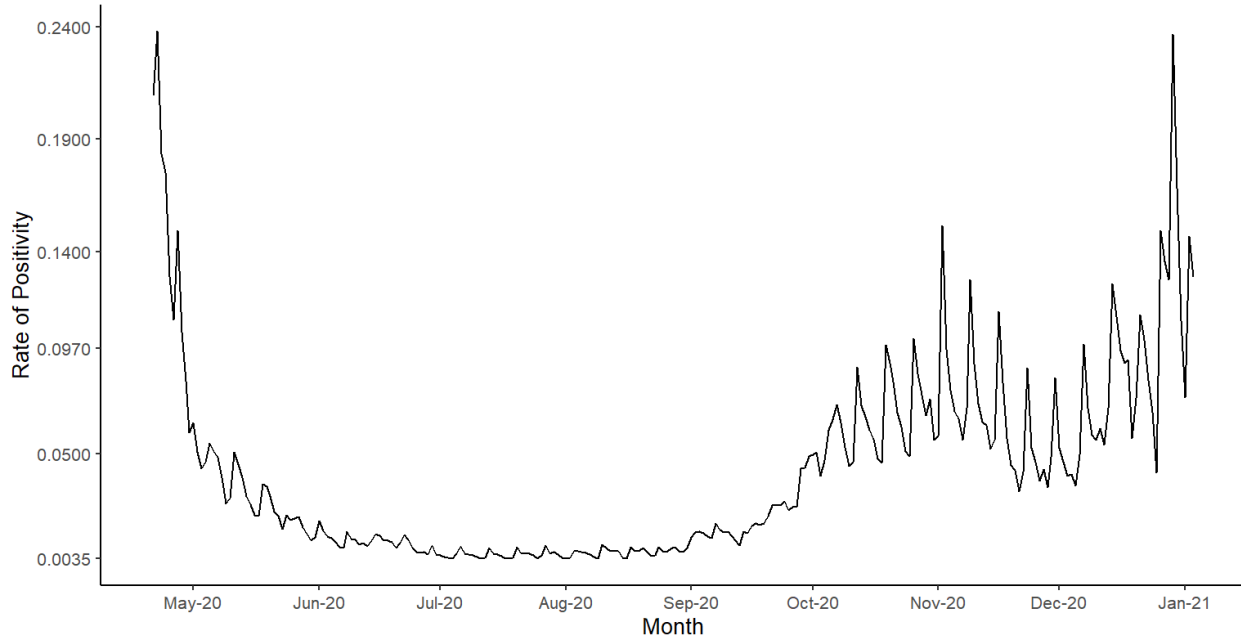


Figure 8: Plot of the daily rate of test positivity over time. Plotted using the "ggplot" function from the "ggplot2" package.

4 Methodology

For this investigation, I have used two methods to estimate transmission rates. These are a deterministic SIR model and a stochastic epidemic model. Deterministic models have a set of equations that describe the system inputs and outputs exactly and predict outcomes with 100% certainty. Comparatively, stochastic models account for an element of randomness and represent a situation where uncertainty is present⁸⁵. Following this, I have used time series modelling to predict the confirmed cases, rate of transmission, and rate of positivity.

4.1 Deterministic SIR Model

Three functions underpin the SIR model, describing the number of susceptible, $S(t)$, infected, $I(t)$, and removed, $R(t)$, people at time t . We define removed as the number of people who have either died or recovered. We can then use these functions to define the SIR model as a system of three Ordinary Differential Equations (ODEs)⁸⁶:

$$\frac{dS(t)}{dt} = -\frac{\beta I(t)S(t)}{N}, \tag{1}$$

$$\frac{dI(t)}{dt} = \frac{\beta I(t)S(t)}{N} - \gamma I(t), \tag{2}$$

$$\frac{dR(t)}{dt} = \gamma I(t). \tag{3}$$

The parameter β describes the total rate at which an infected person generates new infections. The parameter γ describes the rate at which people are removed from the population – either via death or recovery.

Equation (1) describes the rate at which the number of susceptible individuals changes. This value is always negative because susceptible individuals are becoming infected, and hence the number of them is constantly decreasing.

Equation (2) describes the rate at which the number of infected individuals changes and can be either positive or negative depending on the number of people removed.

Equation (3) describes the rate at which the number of removed individuals changes. This number is always positive as an individual cannot leave the removal bracket.

Figure 9 shows how individuals transition between states⁸⁷, related to the values of β and γ . From this, we can see that a higher value of β would increase the speed at which people become infected. And a higher value of γ would increase the speed at which people become removed. To solve this system of equations we need the initial values $S(0)$, $I(0)$, and $R(0)$ – the function values at $t = 0$. We also need the values of β and γ , which we will estimate.

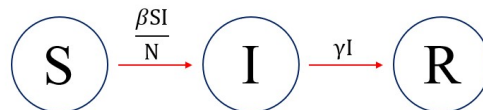


Figure 9: Diagram illustrating the SIR model system, where $S = S(t)$, $I = I(t)$, and $R = R(t)$.

To implement the deterministic SIR model, I used optimisation to estimate the values of β and γ . This optimisation was carried out using the "modFit" function⁸⁸ from the "FME" package⁸⁹, to find the best combination of β and γ for the model. For this optimisation, I used the "ode" function⁹⁰ from the "deSolve" package⁹¹ to solve the ODEs using initial values from the data and various combinations of values for β and γ . The process minimises the residual sum of squares, which is the difference between the data and the values predicted by the SIR model⁹². Once I have found the optimal combination of values, we can then use it to calculate R_0 , defined as:

$$R_0 = \frac{\beta}{\gamma}.$$

If we estimate β and γ over intervals, we can estimate R_t as:

$$R_t = \frac{\beta_t}{\gamma_t}$$

To do this, I first stored the relevant data in intervals. I then used optimisation to estimate β and γ using the previous 7 days of data. I carried out this optimisation process with two "for" loops, which model the data continuously. In the first loop, I defined a function for each 7-day interval, which can solve the ODEs given the initial values from the stored data.

It then calculates the residual sum of squares using the fitted values for each interval. In the second loop, I minimised the residual sum of squares to find the optimal combination of β and γ for each interval.

4.2 Stochastic Epidemic Model

The stochastic epidemic model⁹³ uses a Poisson process to model incidence over time, defined as:

$$I_t \sim \text{Poi}(\mu_t), \text{ with mean } \mu_t = R_t \sum_{s=1}^t I_{t-s} w_s,$$

where R_t is the instantaneous reproduction⁷ at time t , and w_s is a distribution approximated by a serial interval (SI). A serial interval is the number of days between symptom onset in a primary case and the onset of symptoms in secondary cases⁹⁴. The serial intervals in this model were assumed to follow a shifted Gamma distribution. If transmission is assumed to be constant for a time period, $[t - \tau + 1 : t]$ days, then the reproduction number can be denoted as $R_{t,\tau}$. This is the transmission in the period from day $t - \tau + 1$ to day t inclusive.

From this, we can show the likelihood function for the incidence during this period, $I_{t-\tau+1}, \dots, I_t$, is:

$$L(I_{t-\tau+1}, \dots, I_t | I_0, \dots, I_{t-\tau}, w_s, R_{t,\tau}) = \prod_{s=t-\tau+1}^t \frac{(R_{t,\tau} \Lambda_s)^{I_s} e^{-R_{t,\tau} \Lambda_s}}{I_s!}, \text{ where } \Lambda_t = \sum_{s=1}^t I_{t-s} w_s. \quad (4)$$

It is assumed that the prior of $R_{t,\tau}$ follows a $\text{Gamma}(a, b)$ distribution, and can be defined as:

$$P(R_{t,\tau}) = \frac{R_{t,\tau}^{a-1} e^{-\frac{R_{t,\tau}}{b}}}{\Gamma(a) b^a}, \quad (5)$$

where $\Gamma(a) = (1 - a)!$, for all positive integers. Note, this is a constant which can be ignored up to proportionality.

We can then use a Bayesian framework to estimate the parameters of the model. The Bayesian framework uses the prior distribution and the likelihood function to inform the posterior distribution. We define the posterior distribution of $R_{t,\tau}$ by the following proportionality:

$$P(I_{t-\tau+1}, \dots, I_t, R_{t,\tau} | I_0, \dots, I_{t-\tau}, w_s) \propto L(I_{t-\tau+1}, \dots, I_t | I_0, \dots, I_t, w_s, R_{t,\tau}) P(R_{t,\tau}). \quad (6)$$

By using Equation (4) and Equation (5) in Equation (6), we can show that the posterior distribution of $R_{t,\tau}$ is:

$$R_{t,\tau} \sim \text{Gamma} \left(a + \sum_{s=t-\tau+1}^t I_s, \left(\frac{1}{b} + \sum_{s=t-\tau+1}^t \Lambda_s \right)^{-1} \right). \quad (7)$$

Equation (7) shows that the posterior distribution of $R_{t,\tau}$ follows the same distribution as the prior. Hence, the Gamma prior is conjugate for the Poisson model – as it yields a Gamma posterior.

Given the fact that $R_{t,\tau}$ follows a Gamma distribution, we have:

$$\mu_t = \frac{\alpha_t}{\beta_t}, \quad (8)$$

$$\sigma_t^2 = \frac{\alpha_t}{\beta_t^2}, \quad (9)$$

$$CV_t = \frac{\sigma_t}{\mu_t} = \frac{\sqrt{\frac{\alpha_t}{\beta_t^2}}}{\frac{\alpha_t}{\beta_t}} = \sqrt{\frac{\alpha_t}{\beta_t^2}} \cdot \left(\frac{\beta_t}{\alpha_t} \right) = \sqrt{\frac{\alpha_t \beta_t^2}{\beta_t^2 \alpha_t}} = \sqrt{\frac{1}{\alpha_t}}, \quad (10)$$

where $\alpha_t = a + \sum_{s=t-\tau+1}^t I_s$ and $\beta_t = \left(\frac{1}{b} + \sum_{s=t-\tau+1}^t \Lambda_s \right)^{-1}$.

Equation (8) defines the posterior mean, μ_t , of $R_{t,\tau}$. Out of the values found from our Bayesian inference, the posterior mean is the one used most in this investigation.

Equation (9) defines the posterior variance, σ_t^2 , of $R_{t,\tau}$, and along with Equation (4) it is used to calculate the posterior coefficient of variation.

Equation (10) defines the posterior coefficient of variation, CV_t , of $R_{t,\tau}$, and is often used when defining the value of τ . It is important when defining τ that the value is small enough to be sensitive to change but not so small that it creates too much statistical noise.

To implement the stochastic epidemic model, I used the "estimate_R" function⁹⁵ from the "EpiEstim" package⁹⁶. The function defines τ as the number of days used for the sliding windows, which is by default set to 7 days. In other words, each date available marks the start of a 7-day interval until 7 days before the final date. I then used these intervals with the incidence data of the confirmed cases (see Figure 2). The "estimate_R" function uses a serial interval distribution to calculate the transmission rate, with a range of values for the mean and standard deviation of days between successive cases. Studies found these ranges using analyses of SARS, MERS and COVID-19⁹⁴. By default, the function uses *mean* = 5 and *standard deviation* = 5 for the prior distribution of $R_{t,\tau}$, giving $R_{t,\tau} \sim \text{Gamma}(1, 5)$. The use of these values describes the lack of information that we know about $R_{t,\tau}$ and research has shown they have little effect on the posterior distribution⁹³ of $R_{t,\tau}$.

4.3 Time Series Modelling

For the time series modelling section of this investigation, I have used two methods. The first method is univariate modelling, which means it involves only one time series. For the univariate modelling, I used seasonal Auto Regressive Integrated Moving Average (seasonal ARIMA) models. A seasonal ARIMA model is a simple stochastic time series model⁹⁷, which allows for seasonal influences. The second method is multivariate modelling, which means it involves multiple time series. For the multivariate modelling, I used a Vector Auto Regressive (VAR) model. A VAR model is a stochastic model of two or more time series which influence each other⁹⁸.

4.3.1 Univariate Time Series Modelling

For the univariate time series modelling, I fitted seasonal ARIMA models using three individual time series. These were the time series for the daily confirmed cases, positivity rate, and transmission rate. A seasonal ARIMA model is similar to an ARIMA model but with additional seasonal terms which involve backshifts of the seasonal period⁹⁹. A backshift, denoted by the backward shift operator B , shifts the data back by one period and is often used to signify differencing¹⁰⁰. Differencing is the process used to transform a non-stationary time series into a stationary one by finding the differences between consecutive observations¹⁰¹.

A seasonal ARIMA model is written as:

$$\text{ARIMA}(p, d, q)(P, D, Q)_m,$$

where (p, d, q) is the non-seasonal part of the model, $(P, D, Q)_m$ is the seasonal part of the model, and m is the frequency of the data.

The values of p and P denote the presence of autoregressive (AR) terms. These AR terms describe the influence that past observations have on current and new observations⁹⁷. The values of d and D denote the presence of integration (I) terms. These I terms describe the presence of trends in the data, that is, the number of differences taken to achieve stationarity⁹⁷. Another way of viewing these terms is that d and D denote the number of trends present in the season and non-seasonal data, respectively. The values of q and Q denote the presence of moving average (MA) terms. These MA terms describe the influence errors of past observations have on current and new observations⁹⁷. We then form the model by multiplying the non-seasonal and seasonal terms together⁹⁹.

For example, $\text{ARIMA}(1, 1, 1)(1, 1, 1)_4$ – a seasonal ARIMA model for quarterly data, can be written as:

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)x_t = (1 + \theta_1 B)(1 + \Theta_1 B^4)\epsilon_t,$$

where x_t is the data, ϵ_t is the error term, ϕ_1 is the non-seasonal AR term, Φ_1 is the seasonal AR term, θ_1 is the non-seasonal MA term, and Θ_1 is the seasonal MA term⁹⁹.

In this representation, we have the following:

- $(1 - \phi_1 B)$ which denotes the non-seasonal AR term,
- $(1 - \Phi_1 B^4)$ which denotes the seasonal AR term,
- $(1 - B)$ which denotes the non-seasonal I term,
- $(1 - B^4)$ which denotes the seasonal I term,
- $(1 + \theta_1 B)$ which denotes the non-seasonal MA term, and
- $(1 + \Theta_1 B^4)$ which denotes the seasonal MA term.

Note that when we have seasonal terms, the backshift is to the power of the frequency. As we include terms at higher lags, the powers of our backshifts will increase, that is, at lag 2 the backshift for a non-seasonal term would be B^2 but would be B^8 for a (quarterly) seasonal term. Here we have B^4 because the data is quarterly and we have only included terms at lag 1.

In this example, the non-seasonal behaviour can be modelled by an ARIMA(1,1,1) model. This shows that the non-seasonal data is non-stationary, as there is an integration term. It also shows that current observations are influenced by both the value and error term of the previous observation. The seasonal behaviour can be modelled by an ARIMA(1,1,1) model, which shows that the seasonal data is also non-stationary. Non-stationary seasonal data means that the seasonal trend is not the same for each season. The model for the seasonal behaviour also shows that the current observation is influenced by both the value and error term of the previous observation.

I have implemented the seasonal ARIMA models by creating time series using the relevant data, which I set to have a frequency of 7. This frequency is equivalent to a weekly seasonality, which accounts for the impact the day of the week had on the data. I then used the "auto.arima" function¹⁰² from the "forecast" package¹⁰³ to select the best seasonal ARIMA model for the data. The function compares various seasonal ARIMA models and determines the most suitable model based on a specified information criterion. The most suitable model from the output of the "auto.arima" function is given in the form:

$$\text{ARIMA}(p, d, q)(P, D, Q)[m],$$

where first bracket indicates the non-seasonal behaviour, the second bracket indicates the seasonal behaviour, and the square bracket indicates the frequency of the series.

To verify the best model available, I ran the "auto.arima" function three times and used three criteria: AIC, AICc (the default criterion), and BIC. These criteria can be defined as follows:

1. Akaike Information Criterion¹⁰⁴: $\text{AIC} = -2\log\text{-likelihood} + 2K$

2. Akaike Information Criterion corrected¹⁰⁴: $\text{AICc} = -2\log\text{-likelihood} + 2K + \frac{2K(K+1)}{n-K-1}$

3. Bayesian Information Criterion¹⁰⁵: $\text{BIC} = K\log(n) - 2\log(L(\theta))$

Where,

- n is the sample size,
- K is the number of model parameters,
- *log-likelihood* is a measure of how well a model fits,
- $L(\theta)$, the maximised likelihood, is a measure of how likely it is we obtain the data we have, supposing the model being tested was a given.

Through the use of the "tsdiag" function¹⁰⁶ from the "stats" package, I then checked the suitability of the chosen model. This function prints information regarding the residuals of the model and the p-values for the Portmanteau test – which tests goodness of fit¹⁰⁶. Once I confirmed the most suitable model, I calculated 95% confidence intervals for the fitted values and visually compared the fitted and observed values. I calculated these intervals using the standard deviation given in the output of the "auto.arima" function, for a 95% confidence interval, the bounds are found using $estimate \pm 1.96 \times \sigma_{estimate}$ where the estimate is each fitted value in turn. Once the best model has been identified, confirmed to be suitable, and compared to the observations, the model can be used to make predictions. To do this I used the "forecast" function¹⁰⁷ from the "forecast" package, which constructs forecasts based on the results of predictions made using the "auto.arima" output¹⁰⁸.

4.3.2 Multivariate Time Series Modelling

For the multivariate time series modelling, I fitted a VAR model using a combination of the three time series described in Section 4.3.1. In a VAR model, each variable is modelled “as a linear combination of past values of itself and the past values of other variables in the system”⁹⁸. This linear combination means that the model assumes that the variables interact and influence future values.

A VAR(p) model is written as:

$$\mathbf{X}_t = \phi \mathbf{D}_t + \Theta_1 \mathbf{X}_{t-1} + \dots + \Theta_p \mathbf{X}_{t-p} + \epsilon_t,$$

where p is the number of lags included, \mathbf{X}_t is the vector of m time series, Θ_i is the matrix of VAR parameters, and ϵ_t is the vector of error terms¹⁰⁹. In addition, $\phi \mathbf{D}_t$ is a vector of deterministic terms which are restricted to the form:

$$\phi \mathbf{D}_t = \mu_t = \mu_0 + \mu_1 t,$$

where μ_0 is a vector of constants that can either be 0, restricted, or unrestricted. And $\mu_1 t$ is a vector of trends that can either be restricted or unrestricted¹⁰⁹.

For example, a trivariate VAR(1) model with an unrestricted constant and no trend would be:

$$\begin{aligned} x_{1,t} &= \mu_{0,1} + \theta_{11,1}x_{1,t-1} + \theta_{12,1}x_{2,t-1} + \theta_{13,1}x_{3,t-1} + \epsilon_{1,t} \\ x_{2,t} &= \mu_{0,2} + \theta_{21,1}x_{1,t-1} + \theta_{22,1}x_{2,t-1} + \theta_{23,1}x_{3,t-1} + \epsilon_{2,t} \\ x_{3,t} &= \mu_{0,3} + \theta_{31,1}x_{1,t-1} + \theta_{32,1}x_{2,t-1} + \theta_{33,1}x_{3,t-1} + \epsilon_{3,t}, \end{aligned}$$

where $x_{1,t}$, $x_{2,t}$, and $x_{3,t}$ are the time series, $\mu_{0,1}$, $\mu_{0,2}$, and $\mu_{0,3}$ are unrestricted constants, $\epsilon_{1,t}$, $\epsilon_{2,t}$, and $\epsilon_{3,t}$ are white noise processes, the coefficient $\theta_{ii,l}$ captures the influence of the l^{th} lag of variable x_i on itself, and the coefficient $\theta_{ij,l}$ captures the influence of the l^{th} lag of variable x_j on x_i ¹¹⁰.

In matrix form, this VAR(1) can be written as:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix} = \begin{bmatrix} \mu_{0,1} \\ \mu_{0,2} \\ \mu_{0,3} \end{bmatrix} + \begin{bmatrix} \theta_{11,1} & \theta_{12,1} & \theta_{13,1} \\ \theta_{21,1} & \theta_{22,1} & \theta_{23,1} \\ \theta_{31,1} & \theta_{32,1} & \theta_{33,1} \end{bmatrix} \times \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \\ x_{3,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}.$$

Suppose the time series in the model are non-stationary. In that case, we must determine whether cointegration exists – the presence of cointegration signifies long-run (or long-term) relationships between variables in a multivariate time series¹¹¹. Cointegration tells us that there exists a linear combination of the variables, which is stationary. If we find no cointegration, then the VAR model should not be used. However, if we find cointegration, then a Vector Error Correction Model (VECM) should be constructed. The VECM can then be used to find the cointegrating relationships. These relationships define the linear combination of the variables, which achieves stationarity.

The VECM for a VAR(p) model is given as:

$$\Delta \mathbf{X}_t = \phi \mathbf{D}_t + \Pi \mathbf{X}_{t-1} + \Gamma_1 \Delta \mathbf{X}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{X}_{t-p+1} + \epsilon_t,$$

where Δ signifies one difference having been taken, $\Pi = \Theta_1 + \dots + \Theta_p - \mathbf{1}_m$ is the long-run impact matrix, with $\mathbf{1}_m$ being the $m \times m$ dimension identity matrix, and Γ_k are the short-run impact matrices given in the following form¹⁰⁹:

$$\Gamma_k = - \sum_{j=k+1}^p \Theta_j, \quad k = 1, \dots, p-1.$$

The rank, r , of the matrix Π tells us the number of cointegrating relationships in the multivariate series. That is, if $r = 0$, then there is no cointegration present. For $0 < r < m$ the matrix Π can be written as the product:

$$\Pi = \alpha \beta',$$

where α and β are $(m \times r)$ dimension matrices with $rank(\alpha) = rank(\beta) = r$.

The VECM can then be written as:

$$\Delta \mathbf{X}_t = \phi \mathbf{D}_t + \alpha \beta' \mathbf{X}_{t-1} + \Gamma_1 \Delta \mathbf{X}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{X}_{t-p+1} + \epsilon_t,$$

where $\beta' \mathbf{X}_t \sim I(0)$, meaning stationary, since β' is a matrix of cointegrating vectors¹⁰⁹.

For the trivariate VAR(1) example, the VECM would be given as:

$$\begin{aligned}\Delta x_{1,t} &= \mu_{0,1} + \alpha_1(\beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \beta_3 x_{3,t-1}) + \epsilon_{1,t} \\ \Delta x_{2,t} &= \mu_{0,2} + \alpha_2(\beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \beta_3 x_{3,t-1}) + \epsilon_{2,t} \\ \Delta x_{3,t} &= \mu_{0,3} + \alpha_3(\beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \beta_3 x_{3,t-1}) + \epsilon_{3,t}.\end{aligned}$$

In matrix form, this VECM can be written as:

$$\begin{bmatrix} \Delta x_{1,t} \\ \Delta x_{2,t} \\ \Delta x_{3,t} \end{bmatrix} = \begin{bmatrix} \mu_{0,1} \\ \mu_{0,2} \\ \mu_{0,3} \end{bmatrix} + \begin{bmatrix} \alpha_1\beta_1 + \alpha_1\beta_2 + \alpha_1\beta_3 \\ \alpha_2\beta_1 + \alpha_2\beta_2 + \alpha_2\beta_3 \\ \alpha_3\beta_1 + \alpha_3\beta_2 + \alpha_3\beta_3 \end{bmatrix} \times \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \\ x_{3,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}.$$

From a VECM, it is possible to create a new VAR model which includes the interaction between the variables described by the cointegration relationships. Using this, we can model the data appropriately and predict the future behaviour of the system.

To implement the VAR model, I first subsetted the relevant data to span the same period. Subsetting the data was necessary because the positivity data begins later than the rest of the data. I then created separate time series out of this data, which had frequency 7, and then combined these to form one multivariate time series. Once this has been created, it is important to check for cointegration which is done using the "po.test" function¹¹² from the "tseries" package¹¹³. The "po.test" function computes the Phillips-Ouliaris test¹¹², which tests the following hypotheses whilst allowing for variability¹¹¹:

$$H_0: \text{no cointegration exists VS. } H_1: \text{cointegration exists}$$

Meaning that we assume there is no cointegration and prove otherwise.

Once cointegration has been confirmed, I then used the "VARselect" function¹¹⁴ from the "vars" package¹¹⁵, to select the best VAR model for the data. The function returns the best value of p for a VAR(p) model, with the maximum value being 10 by default for each of four available information criteria. The four criteria are the AIC, HQC, SC, and FPE, which can be defined as follows:

1. Akaike Information Criterion¹¹⁶: $AIC^{VAR} = \ln(\det(\hat{\Sigma}_p)) + \frac{2K}{n}$
2. Hannan-Quinn Criterion¹¹⁶: $HQC^{VAR} = \ln(\det(\hat{\Sigma}_p)) + \frac{2K \ln(\ln(n))}{n}$
3. Schwarz Criterion¹¹⁶: $SC^{VAR} = \ln(\det(\hat{\Sigma}_p)) + \frac{K \ln(n)}{n}$
4. Final Prediction Error¹¹⁴: $FPE^{VAR} = \left(\frac{n+p^*}{n-p^*}\right)^K \det(\hat{\Sigma}_p)$

Where,

- n is the sample size,
- K is the number of model parameters,

- $\hat{\Sigma}_p$ is the maximum likelihood estimator of the error variance Σ at lag p ,
- p is the lag order,
- p^* is the number of parameters in each equation.

I then chose the value of p which corresponded to the lowest value in most of the information criteria. Using this VAR(p) model, I then constructed the Vector Error Correction Model (VECM) using the "ca.jo" function¹¹⁷ from the "urca" package¹¹⁸, with the test type defined as 'eigen'. This function conducts Johansen's maximum eigenvalue test, which is a series of hypothesis tests to find the number of cointegration relationships¹¹⁹. The first hypothesis test is:

$H_0: \text{rank}(\Pi) = 0$, no cointegration VS. $H_1: \text{rank}(\Pi) = 1$, one cointegration relationship.

This test looks at the largest eigenvalue. If the largest eigenvalue $\lambda_1 = 0$, then there is no cointegration. If $\lambda_1 \neq 0$, then there is at least one cointegration relationship¹¹⁹ and we go on to the next hypothesis test:

$H_0: \text{rank}(\Pi) = 1$, one cointegration relationship VS. $H_1: \text{rank}(\Pi) = 2$, two cointegration relationships.

This test looks at the second largest eigenvalue. If the second largest eigenvalue $\lambda_2 = 0$, then there is exactly one cointegration relationship. If $\lambda_2 \neq 0$, then there is at least two cointegration relationships¹¹⁹. This test will be the last conducted, given that there are three variables the $\text{rank}(\Pi) \leq 2$ there can be a maximum of two cointegration relationships.

After finding the number of cointegration relationships present, I converted the VECM into a VAR model. I then fitted the subsequent VAR model to the data, which I then used to make predictions. For this, I used the "predict" function¹²⁰ from the "stats" package, which makes recursive forecasts for each variable¹²¹.

5 Results

This investigation aimed to determine the effect government policies have on the transmission of COVID-19. I have first estimated transmission (see Section 4.1 and Section 4.2), and I have then predicted the future values of key pieces of UK data (see Section 4.3).

5.1 Deterministic SIR Model Results

Before estimating transmission, I checked the suitability of the deterministic SIR model for the data.

Figure 10 shows that the SIR model provided a good fit for the data. Overall, the fitted line followed the observed values very closely, where the blue points signify daily observations. We see that the fitted values were slightly overestimated at the end of December 2020 and the beginning of January 2021, but were still a good fit.

Following this, I then estimated transmission over time using the deterministic SIR model.

Figure 11 shows that R_t significantly decreased during national lockdowns and the largest impact was the first lockdown. During the first lockdown, R_t fell from $R_t \approx 1.8$ to $R_t \approx 1$ and dropped as low as $R_t \approx 0.87$ at the beginning of June 2020. Whereas during the second lockdown, R_t fell from $R_t \approx 1.1$ to $R_t \approx 0.9$ just as the lockdown was lifted. We can also observe that R_t was stable at $R_t \approx 1$ throughout the warmer summer months and then became volatile toward the end of September 2020. Another observation we can make is that after the second lockdown ended, the rate of transmission dropped. However, transmission rates then rose again in the middle of December 2020. Something to note is that we see a high amount of volatility at the beginning of the pandemic.

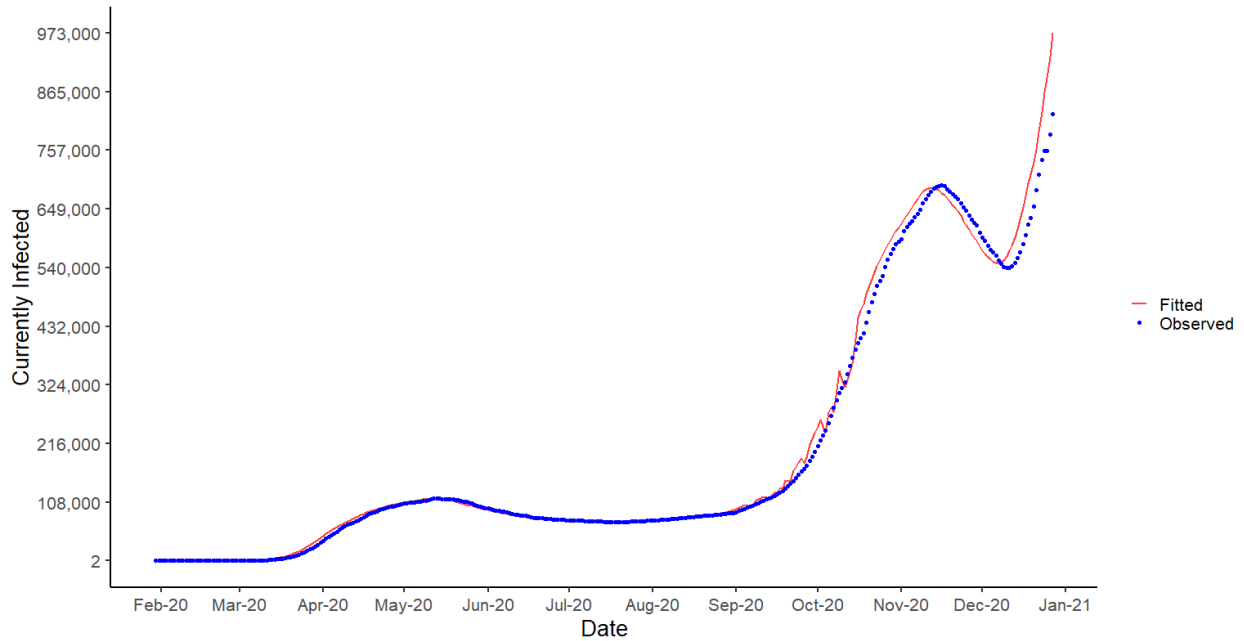


Figure 10: Plot of the fitted and observed data, using the SIR model, where blue points signify daily observed data points and the red line signifies the fitted data. Plotted using the "ggplot" function from the "ggplot2" package.

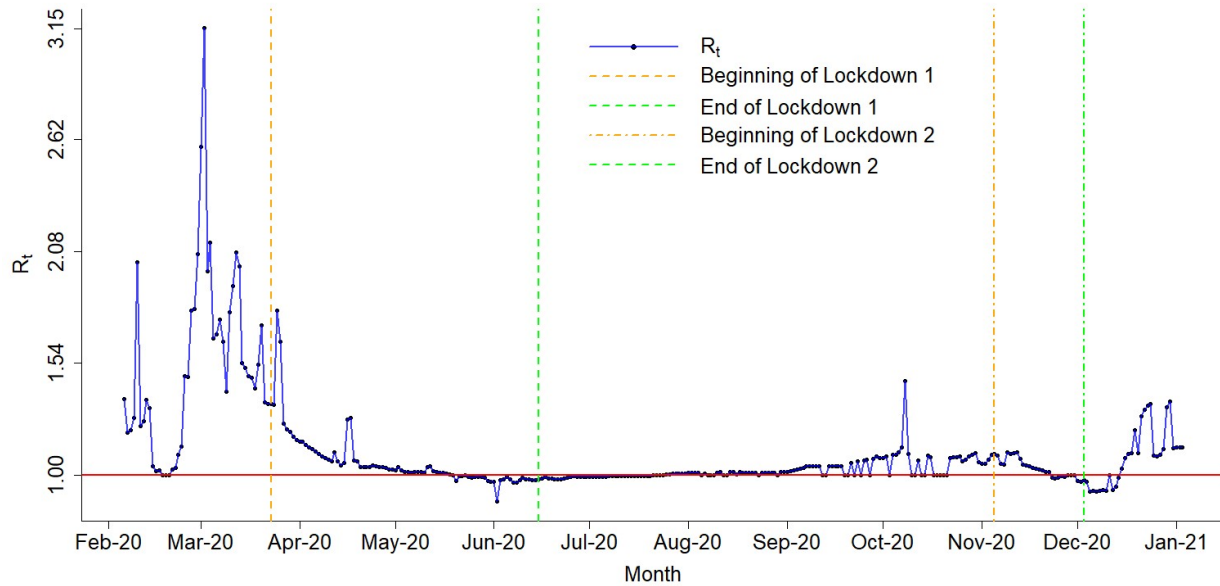


Figure 11: Plot of R_t over time. Plotted using the "plot" function from the "graphics" package, the blue line R_t was added using the "lines" function¹²² from the "graphics" package, and the vertical and horizontal lines were added using the "abline" function¹²³ from the "graphics" package. A red horizontal line was added at 1 to indicate when $R_t = 1$, to show whether the epidemic is growing or decaying. An orange vertical dashed line was added at March 23rd 2020 to indicate the beginning of the first lockdown and a green vertical dashed line was added at June 15th 2020 to indicate the end of the first lockdown. An orange vertical dot-dashed line was added at November 5th 2020 to indicate the beginning of the second lockdown and a green dot-dashed line was added at December 3rd 2020 to indicate the end of the second lockdown.

5.2 Stochastic Epidemic Model Results

From Section 4.2 we can approximate w_s by a serial interval distribution. The "estimate_R" function can implement a variety of these, the two I used were the parametric distribution (see Section 5.2.1) and the uncertain distribution (see Section 5.2.2). For estimates of the posterior mean of $R_{t,\tau}$ to be accurate, the function documentation recommends beginning estimations after there are 12 cases recorded in total⁹³. For the pandemic data, 12 cases have occurred 23 days after the initial case. The corresponding date will be on the 21st February 2020, and therefore, it is from 2020-02-21 that I began estimates.

5.2.1 Parametric Serial Interval

For the parametric SI, I assumed w_s followed a shifted Gamma distribution with a mean of 2.3 days and a standard deviation of 1.4 days⁹⁴. Notably, the serial interval is considerably shorter than that of similar viruses. The shorter interval is because of possible pre-symptomatic transmission³⁵ and undetected asymptomatic individuals³⁴.

Figure 12 shows a relatively wide credible interval for the posterior mean of $R_{t,\tau}$ at the beginning of estimation. The width suggests that the data was less informative at this time since only 12 cases had occurred when estimations began (see Section 5.2), and hence there were fewer data available.

Figure 13 shows the posterior mean of $R_{t,\tau}$ over time, in more detail. We can see that the estimations appeared to be quite volatile. However, given that the maximum value taken was $R_{t,\tau} \approx 2.5$, small changes in transmission will have been apparent. We can also see that transmission rates fell during both lockdowns and rates generally rose after the lockdowns ended. For example, during the first lockdown, transmission fell from $R_{t,\tau} \approx 1.5$ to $R_{t,\tau} \approx 1$ and fell as low as $R_{t,\tau} \approx 0.84$ in May 2020. After which, we see that rates rose to $R_{t,\tau} \approx 1.2$ by August 2020.

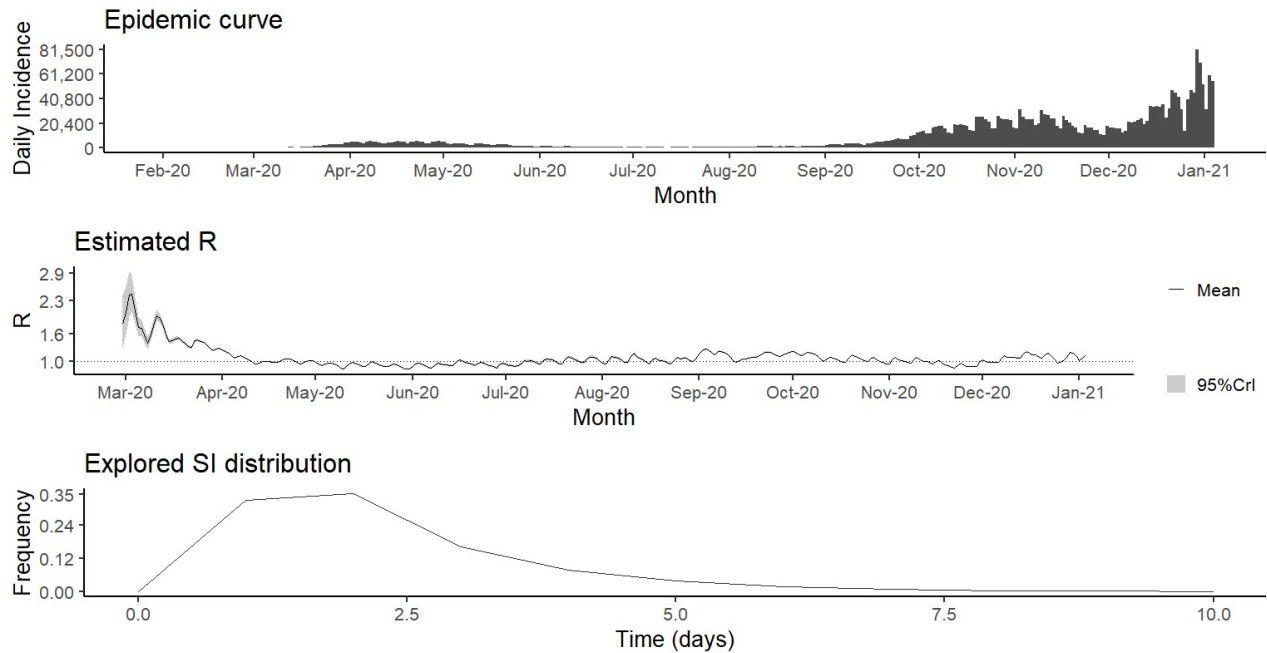


Figure 12: Plot of the output of the "estimate_R" function, using a parametric SI. Plotted using the "estimate_R_plots" function¹²⁴ from the "EpiEstim" package and arranged using the "ggarrange" function from the "ggpubr" package. The top plot shows the confirmed cases incidence data (see Figure 2). The middle plot shows the posterior mean of $R_{t,\tau}$ for each time window of τ days, with a 95% credible interval. The bottom plot shows the probability density function (PDF) of the shifted Gamma distribution which has been explored.

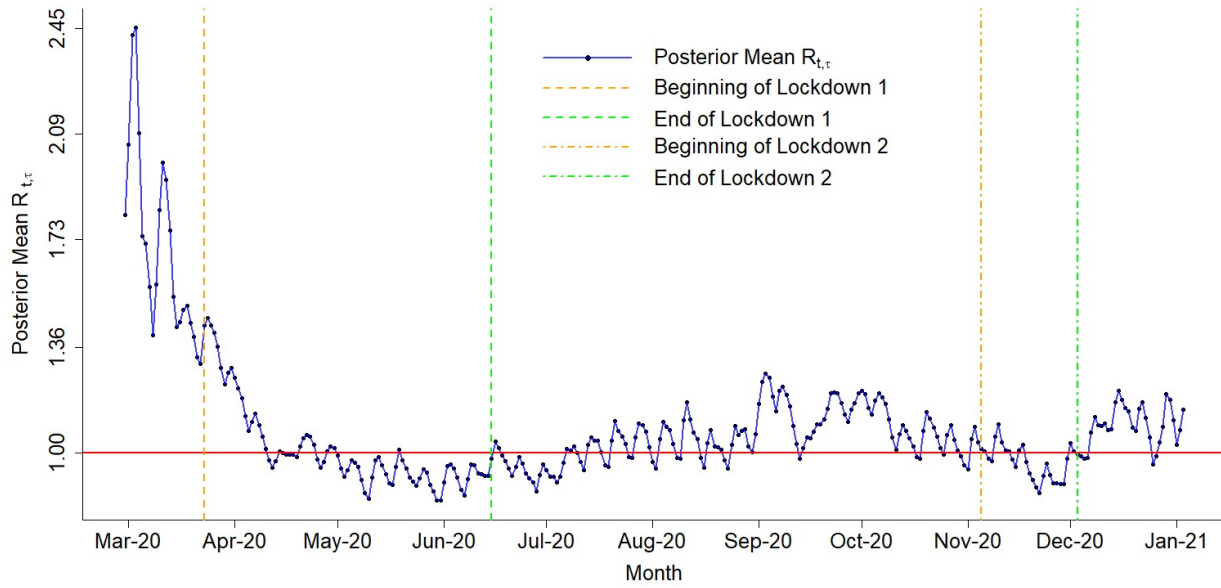


Figure 13: Plot of the posterior mean of $R_{t,\tau}$, for the parametric SI during each time window of τ days. Plotted using the "plot" function from the "graphics" package, the blue line was added using the "lines" function from the "graphics" package, and the vertical and horizontal lines were added using the "abline" function from the "graphics" package. A red horizontal line was added at 1 to indicate when the posterior mean $R_{t,\tau} = 1$, to show whether the epidemic is growing or decaying. An orange vertical dashed line was added at March 23rd 2020 to indicate the beginning of the first lockdown and a green vertical dashed line was added at June 15th 2020 to indicate the end of the first lockdown. An orange vertical dot-dashed line was added at November 5th 2020 to indicate the beginning of the second lockdown and a green dot-dashed line was added at December 3rd 2020 to indicate the end of the second lockdown.

Additionally, throughout the second lockdown, transmission fell from $R_{t,\tau} \approx 1.1$ to $R_{t,\tau} \approx 1$, and fell as low as $R_{t,\tau} \approx 0.9$ near the end of November 2020. After which rates rose to $R_{t,\tau} \approx 1.25$ within two weeks. From July 2020 until September 2020, we see that transmission rates were closer to 1, likely due, in part, to the warmer weather and consequently more people being outside. However, from September 2020 until the implementation of the second lockdown in November 2020, we see that transmission rates were higher.

5.2.2 Uncertain Serial Interval

For the uncertain SI, I assumed w_s followed a shifted Gamma distribution with unknown (uncertain) mean and standard deviation. The "estimate_R" function calculates the posterior distribution of $R_{t,\tau}$ for n_1 combinations of values for the mean and standard deviation. The mean, μ , and standard deviation, σ , are sampled from truncated (or restricted) normal distributions⁹⁵, with the mean number of days $\mu \sim N(7.5, 2^2)$ in the range (2.3, 8.4) days⁹⁴, and the standard deviation of days $\sigma \sim N(3.4, 1^2)$ in the range (0.5, 4) days⁹⁴. I set the number of distributions (combinations of μ and σ) as $n_1 = 750$, and the number of posterior samples to be drawn for each distribution as $n_2 = 750$. These were chosen to allow to balance accuracy and computation time.

Figure 14 shows higher values and a wider credible interval for $R_{t,\tau}$ than Figure 12. The wider interval means that the data provided less information about transmission. However,

we expect this given that the distribution was, by definition, uncertain. We can also see that many of the shifted Gamma distributions appear to have had a mean $\mu \approx 5$ days, which suggests that the mean defined in Section 5.2.1 was too low. However, simply changing the mean in the parametric SI will not improve accuracy since the standard deviation is still unknown and is more difficult to ascertain from Figure 14.

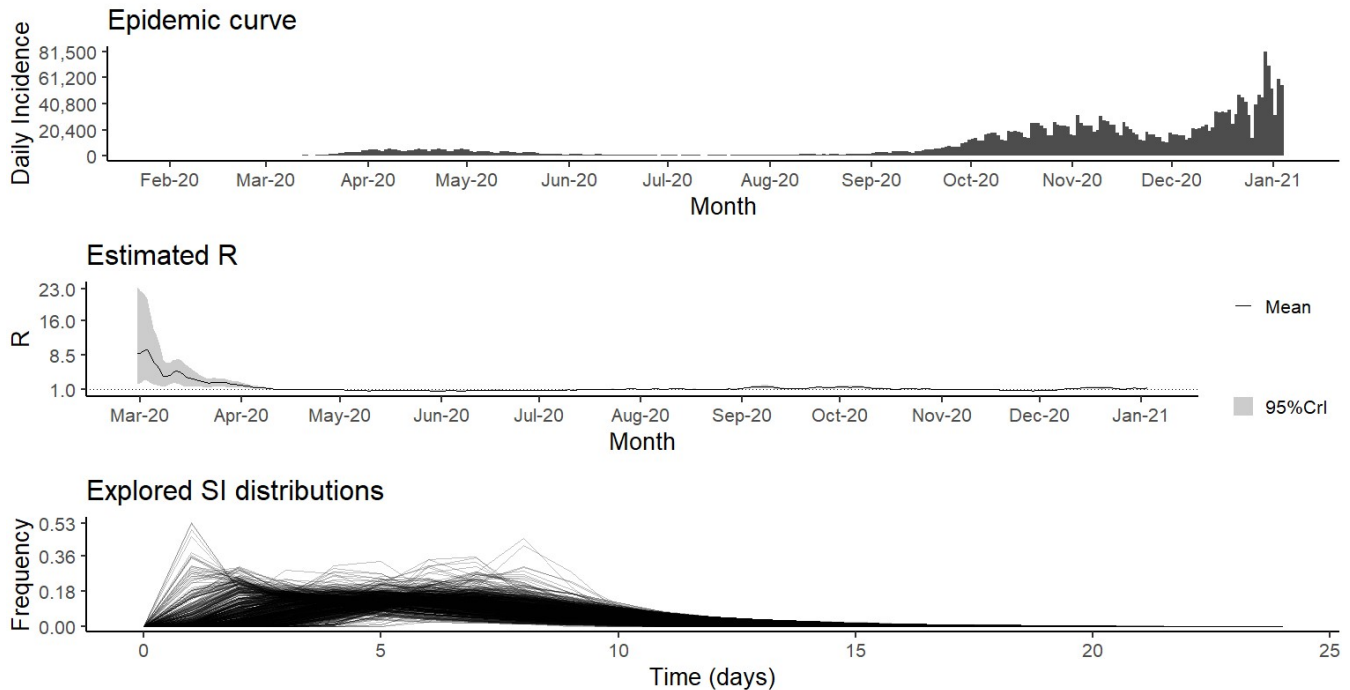


Figure 14: Plot of the output of the "estimate_R" function, using an uncertain SI. Plotted using the "estimate_R_plots" function from the "EpiEstim" package and arranged using the "ggarrange" function from the "ggpubr" package. The top plot shows the confirmed cases incidence data (see Figure 2). The middle plot shows the posterior mean of $R_{t,\tau}$ during each time window of τ days, with a 95% credible interval. The bottom plot shows the probability density functions (PDF) of the shifted Gamma distribution which have been explored.

Figure 15 shows the posterior mean of $R_{t,\tau}$ was much higher than when the parametric SI was used (see Figure 13). For example, at the beginning of March 2020, the uncertain distribution estimated $R_{t,\tau} \approx 9.5$, when the parametric distribution estimated $R_{t,\tau} \approx 2.5$. As we saw in previous estimates of transmission (see Figure 11 and Figure 13), the rate of transmission fell during lockdowns. From this result, we can suggest that lockdowns may have effectively reduced transmission. During the first lockdown, rates fell from $R_{t,\tau} \approx 2.8$ to $R_{t,\tau} \approx 0.8$, and during the second lockdown, rates fell from $R_{t,\tau} \approx 1$ to $R_{t,\tau} \approx 0.8$. At the end of both lockdowns, just before they ended, the rate of transmission fell as low as $R_{t,\tau} \approx 0.68$. We can also observe that after both lockdowns, the rate of transmission increased. After the first lockdown, rates rose above 1 within a month, and after the second lockdown, rates rose above 1 within a week – suggesting that the lockdown was not in place long enough. In general, we can see that transmission rates were only below 1 during, or not long after, a lockdown.

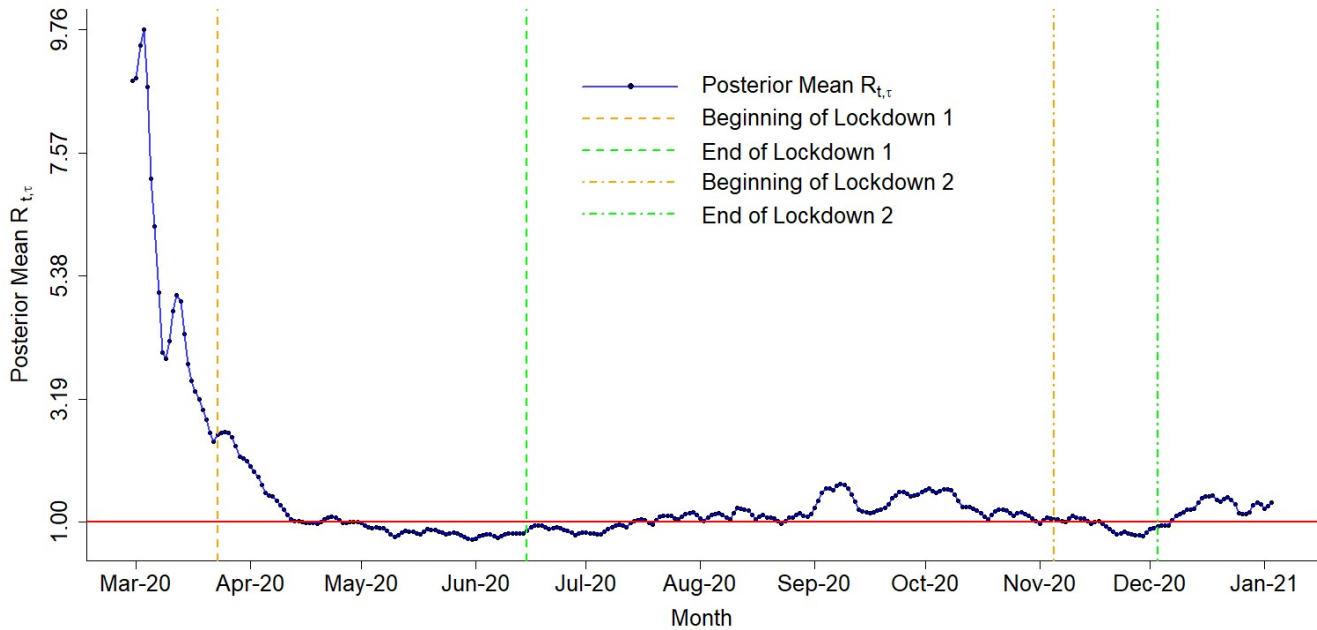


Figure 15: Plot of the posterior mean of $R_{t,\tau}$, for the uncertain SI during each time window of τ days. Plotted using the "plot" function from the "graphics" package, the blue line was added using the "lines" function from the "graphics" package, and the vertical and horizontal lines were added using the "abline" function from the "graphics" package. A red horizontal line was added at 1 to indicate when the posterior mean $R_{t,\tau} = 1$, to show whether the epidemic is growing or decaying. An orange vertical dashed line was added at March 23rd 2020 to indicate the beginning of the first lockdown and a green vertical dashed line was added at June 15th 2020 to indicate the end of the first lockdown. An orange vertical dot-dashed line was added at November 5th 2020 to indicate the beginning of the second lockdown and a green dot-dashed line was added at December 3rd 2020 to indicate the end of the second lockdown.

5.2.3 Comparison of Serial Intervals

To ascertain the impact of the distribution of w_s on the estimates of transmission, I have compared the findings from the parametric serial interval (detailed in Section 5.2.1) and the uncertain serial interval (specified in Section 5.2.2).

Figure 16 shows that the posterior means of $R_{t,\tau}$ were very similar between the two distributions. Whilst we see different posterior means at the beginning of estimations, we expect this as there was less information at this point. We can see that from mid-April 2020, the values followed very closely with each other and that even the credible intervals overlapped. We can also see that the apparent volatility in Figure 13 no longer appeared, due to the maximum value being $R_{t,\tau} \approx 22$ from the uncertain SI credible interval.

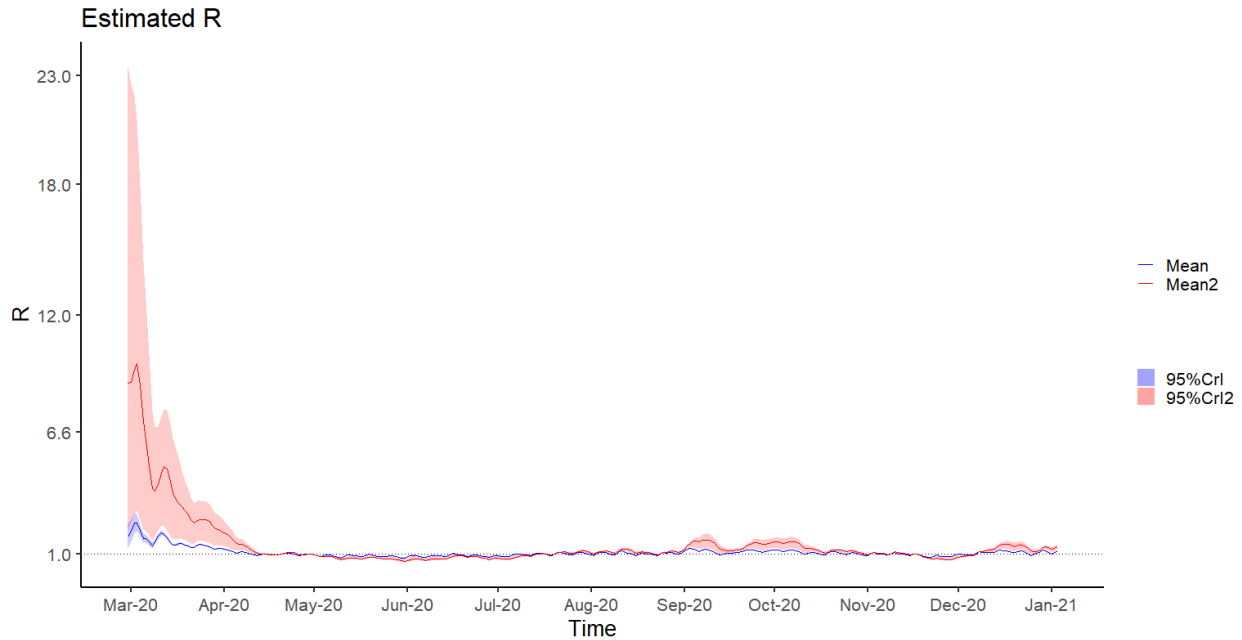


Figure 16: Plot of the posterior means of $R_{t,\tau}$ from both the parametric and uncertain distributions. Plotted using the "estimate_R_plots" function from the "EpiEstim" package. The parametric SI is indicated by the blue line and the uncertain SI is indicated by the red line. The associated credible intervals are indicated by the respective coloured fillings.

5.3 Time Series Modelling Results

Using seasonal ARIMA and VAR models, I have predicted the future values of confirmed cases, positivity rate, and transmission rate under the assumption that the government introduced no further mitigation measures.

5.3.1 Univariate Time Series Modelling – Confirmed Cases

The first variable modelled was the daily confirmed cases, with weekly seasonality. The most suitable model identified through the "auto.arima" function from the "forecast" package was a SARIMA(0, 1, 4)(1, 0, 0)₇ model. For the confirmed cases, the non-seasonal behaviour can be modelled by an ARIMA(0, 1, 4) model. This information shows that the data was non-stationary, as there is an integration term. It also shows that the error terms of the previous four observations influenced current observations. The seasonal behaviour can be modelled by an ARIMA(1, 0, 0) model. This information shows that the value of the previous observation influenced the current observation.

This model can be written as:

$$(1 - \Phi_1 B^7)(1 - B^7)x_t = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4)\epsilon_t.$$

Using the results from the R code, I found these values to three decimal places:

$$(1 - 0.879B^7)(1 - B^7)x_t = (1 - 0.466B - 0.157B^2 + 0.415B^3 - 0.535B^4)\epsilon_t.$$

I then compared the fitted model to the daily confirmed cases data to determine the suitability of the model.

Figure 17 shows that the fitted values followed the observed values very closely. We can see that the confidence interval for the fitted values was seemingly quite wide, which means that the model was not very well informed. We can see that as the number of daily confirmed cases rose, the confidence interval narrowed. This narrowing occurred because as the fitted values increased, the standard deviation influenced the estimates less.

Following this, I used the model to make predictions about the daily number of confirmed cases for the next 7 days.

Figure 18 shows that if the government introduced no new policies, then the daily number of confirmed cases would increase over the next week. The predictions show that confirmed cases may have reached as high as approximately 96,200 without new mitigation strategies. We see that whilst there wouldn't be a consistent increase, the number of daily confirmed cases would reach higher than in the previous week. In general, we see there was an upward trend in the number of daily confirmed cases towards the end of the study period, this is reflected by the rise in the predictions.

To determine the accuracy of these predictions, I then compared the predicted values to the observed values.

Figure 19 shows us that the predicted number of new confirmed cases was higher than the actual number. However, we can see that the observed values followed a similar pattern to that of the predicted. At the end of the predictions, the trend was upward, whereas the actual number of cases was trending downwards. A possible impact on the observed number of cases was that the UK Government announced a third lockdown on the 4th January 2021¹²⁵. The implementation of this lockdown may have reduced the number of people requiring coronavirus tests. Additionally, we see there was a spike in the number of confirmed cases towards the end of the data, which may have influenced the predictions.

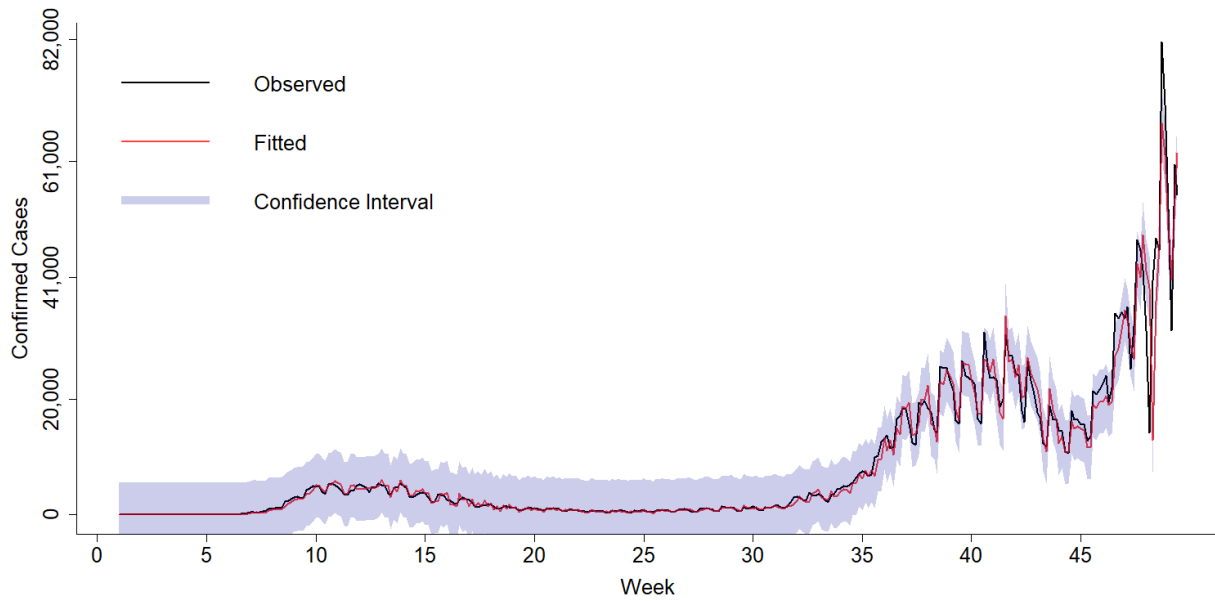


Figure 17: Plot of the univariate time series of daily confirmed cases, with fitted and observed values. Plotted using the "plot" function from the "graphics" package, the fitted values were added using the "lines" function from the "graphics" package, and the confidence interval was added using the "polygon" function¹²⁶ from the "graphics" package.

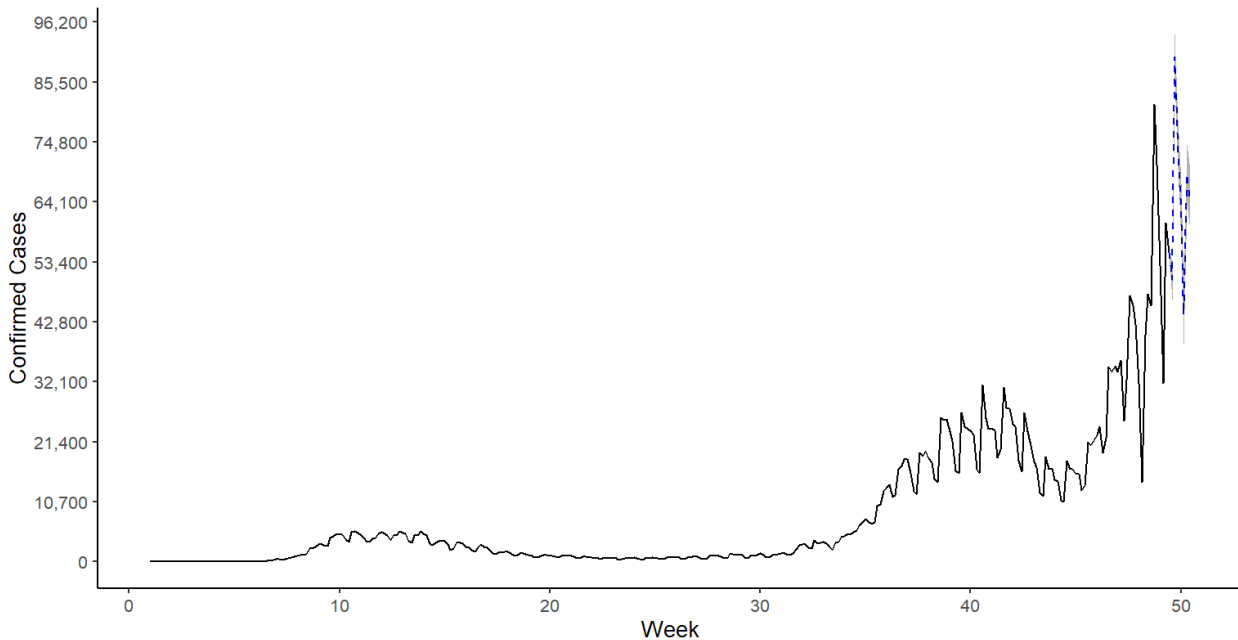


Figure 18: Plot of the predictions of daily confirmed cases from the univariate time series model. The blue dashed line indicates predictions, with the prediction interval in grey. Plotted using the "autoplot" function¹²⁷ from the "forecast" and "ggfortify" packages¹²⁸.

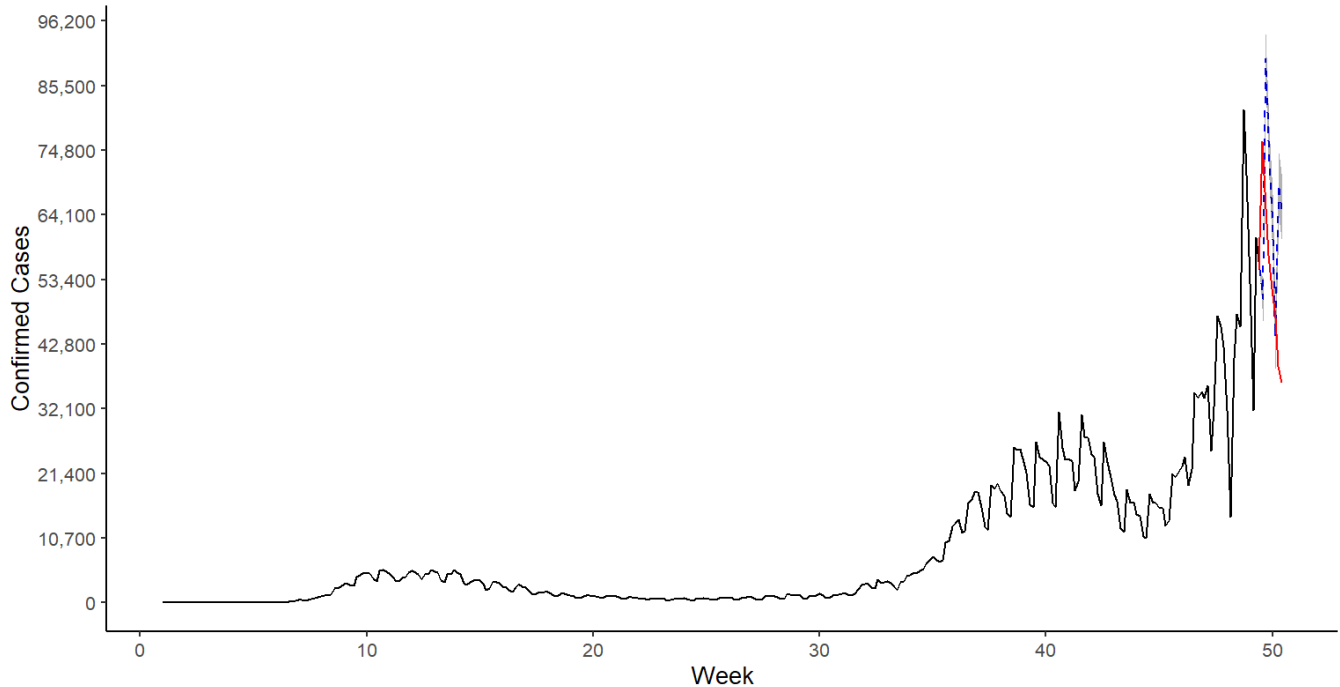


Figure 19: Plot of the comparison between the predicted and observed confirmed cases from the univariate time series model. The blue dashed line indicates predictions, with the prediction interval in grey, and the red line indicates observed values. Plotted using the "autoplot" function from the "forecast" and "ggfortify" packages and the "geom_line" function¹²⁹ from the "ggplot2" package.

5.3.2 Univariate Time Series Modelling – Positivity Rate

The second variable modelled was the daily positivity rate, with weekly seasonality. The most suitable model identified through the "auto.arima" function from the "forecast" package was a SARIMA(2, 1, 0)(1, 0, 2)₇ model. For the positivity rate, the non-seasonal behaviour can be modelled by an ARIMA(2, 1, 0) model. This information shows that the data was non-stationary, as there is an integration term. It also shows that the values of the previous two observations influenced the current value. The seasonal behaviour can be modelled by an ARIMA(1, 0, 2) model, which shows that the previous observation and the error terms of the previous two observations influenced the current value.

This model can be written as:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_1 B^7)(1 - B)x_t = (1 + \Theta_1 B^7 + \Theta_2 B^{14})\epsilon_t.$$

Using the results from the R code, I found these values to three decimal places:

$$(1 + 0.424B + 0.373B^2)(1 - 0.834B^7)(1 - B)x_t = (1 - 0.273B^7 + 0.449B^{14})\epsilon_t.$$

I then compared the fitted model to the daily positivity rate data to determine the suitability of the model.

Figure 20 shows that the fitted values followed the observed values quite closely. We can see that the confidence interval for the fitted values was quite wide, with it becoming narrower when the positivity rate was higher. When the fitted value was small, the standard deviation had more influence on the width of the confidence interval.

Following this, I used the model to make predictions about the daily positivity rate over the next 7 days.

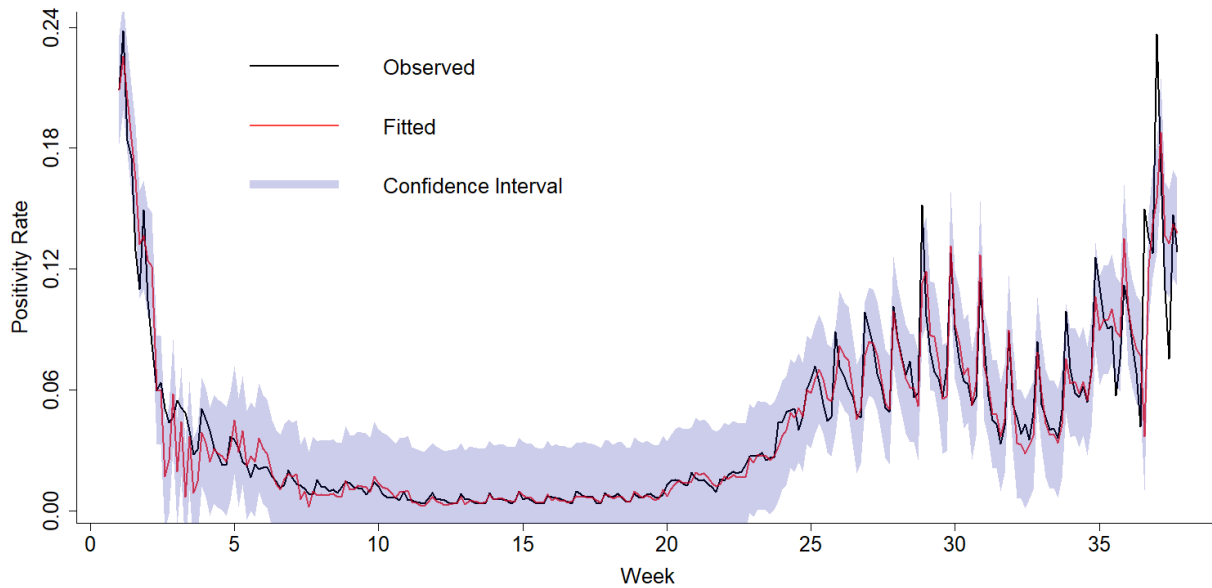


Figure 20: Plot of the univariate time series of positivity rate, with fitted and observed values. Plotted using the "plot" function from the "graphics" package, the fitted values were added using the "lines" function, and the confidence interval was added using the "polygon" function from the "graphics" package.

Figure 21 shows that if the government introduced no new policies, then the daily positivity rate would vary over the next week. The predictions show that the rate may have reached approximately 0.2 without new mitigation strategies. We see that the volatile nature of the positivity rate would continue. We also see that, in general, the positivity rate would be high.

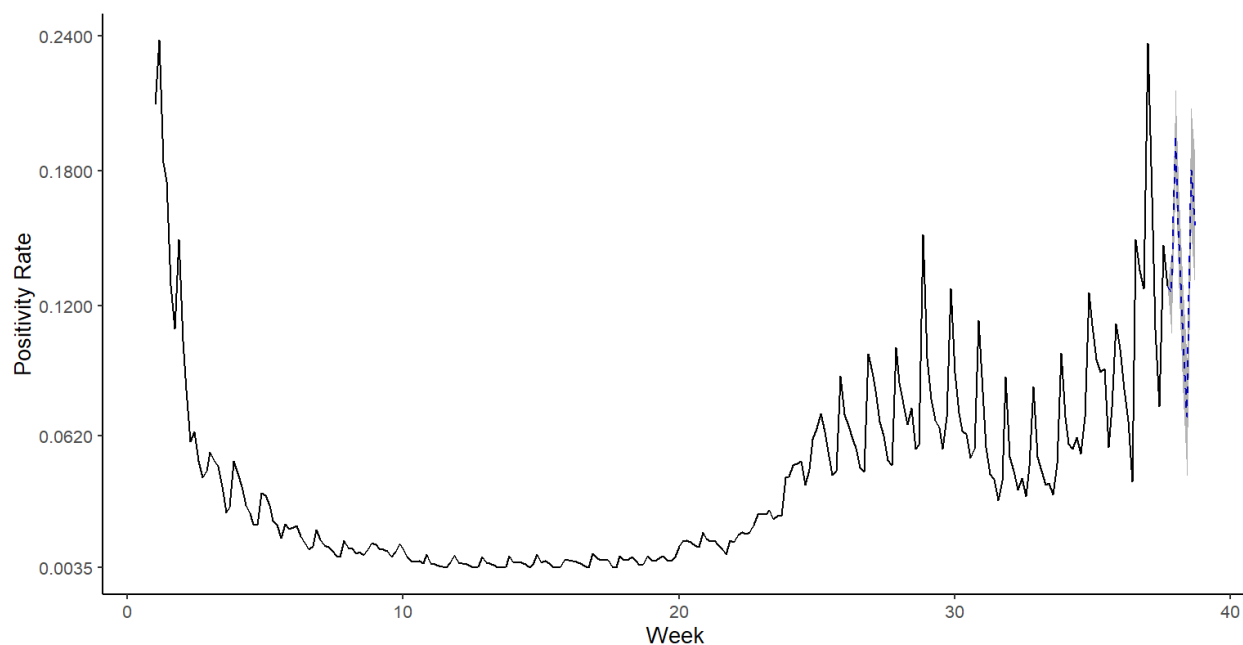


Figure 21: Plot of the predictions of positivity rate from the univariate time series model. The blue dashed line indicates predictions, with the prediction interval in grey. Plotted using the "autoplot" function from the "forecast" and "ggfortify" packages¹²⁸.

To determine the accuracy of these predictions, I then compared the predicted values to the observed values.

Figure 22 shows us that the predicted positivity rate was higher than the actual positivity rate. However, we can see that the observed values followed a similar pattern to that of the predictions. The predicted values exhibited more volatility than the true values, which appear to have been at a turning point. The observed values remained lower when compared to the predicted values at the end of the predictions. As suggested in Section 5.3.1, this may have been due to the lockdown, which came into effect on the 5th January 2021¹²⁵, as it may have limited the number of people getting tested. The lockdown may have lowered the number of symptomatic individuals getting tested as they were staying home anyway and hence felt a test was unnecessary.

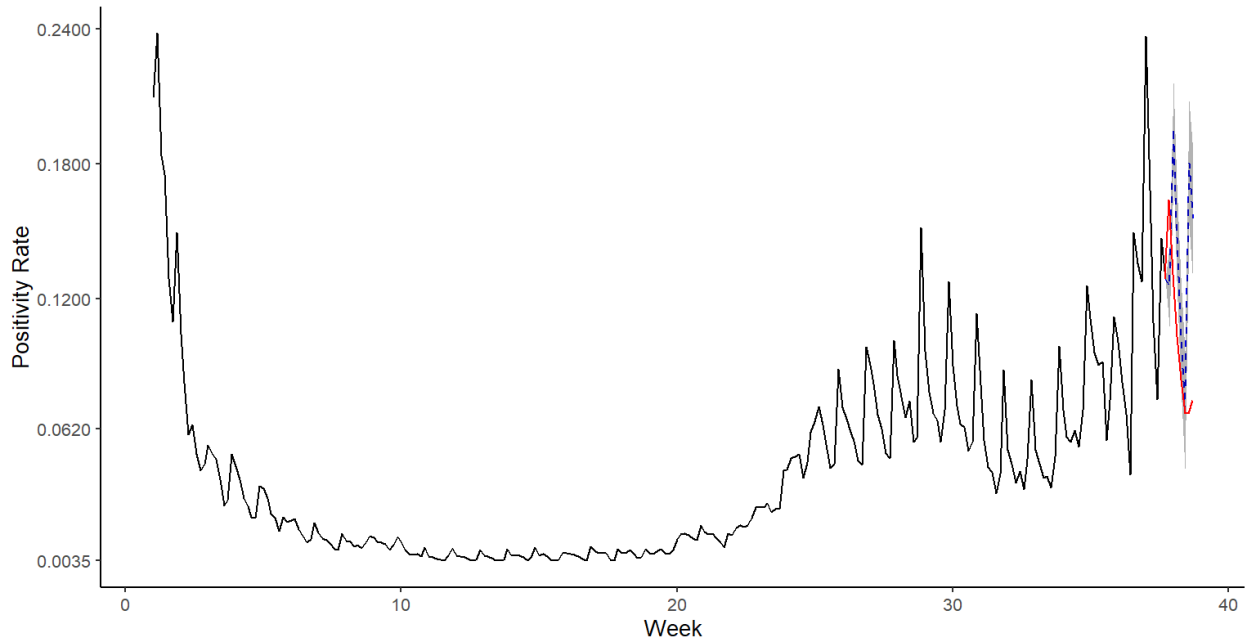


Figure 22: Plot of the comparison between the predicted and observed positivity rate from the univariate time series model. The blue dashed line indicates predictions, with the prediction interval in grey, and the red line indicates observed values. Plotted using the "autoplot" function from the "forecast" and "ggfortify" packages and the "geom_line" function¹²⁹ from the "ggplot2" package.

5.3.3 Univariate Time Series Modelling – Transmission Rate

The third variable modelled was daily transmission rate, with weekly seasonality. The most suitable model identified through the "auto.arima" function from the "forecast" package was a SARIMA(1, 1, 1)(2, 0, 1)₇ model. For the transmission rate, the non-seasonal behaviour can be modelled by an ARIMA(1, 1, 1) model. This information shows that the data was non-stationary, as there is an integration term. It also shows that both the value and error term of the previous observation influence the current value. The seasonal behaviour can be modelled by an ARIMA(2, 0, 1) model, which shows that the values of the previous two observations and the error term of the previous observation influenced the current value.

This model can be written as:

$$(1 - \phi_1 B)(1 - \Phi_1 B^7 - \Phi_2 B^{14})(1 - B)x_t = (1 + \theta_1 B + \Theta_1 B^7)\epsilon_t.$$

Using the results from the R code, I found these values to three decimal places:

$$(1 - 0.550B)(1 + 1.111B^7 + 0.386B^{14})(1 - B)x_t = (1 - 0.757B + 0.864B^7)\epsilon_t.$$

I then compared the fitted model to the daily transmission rate data to determine the suitability of the model.

Figure 23 shows that the fitted values followed the observed values quite closely. We can see that the confidence interval for the fitted values was quite wide, with it being widest when transmission rates were lower. The interval became wider because when the fitted value was larger, the influence of the standard deviation was smaller.

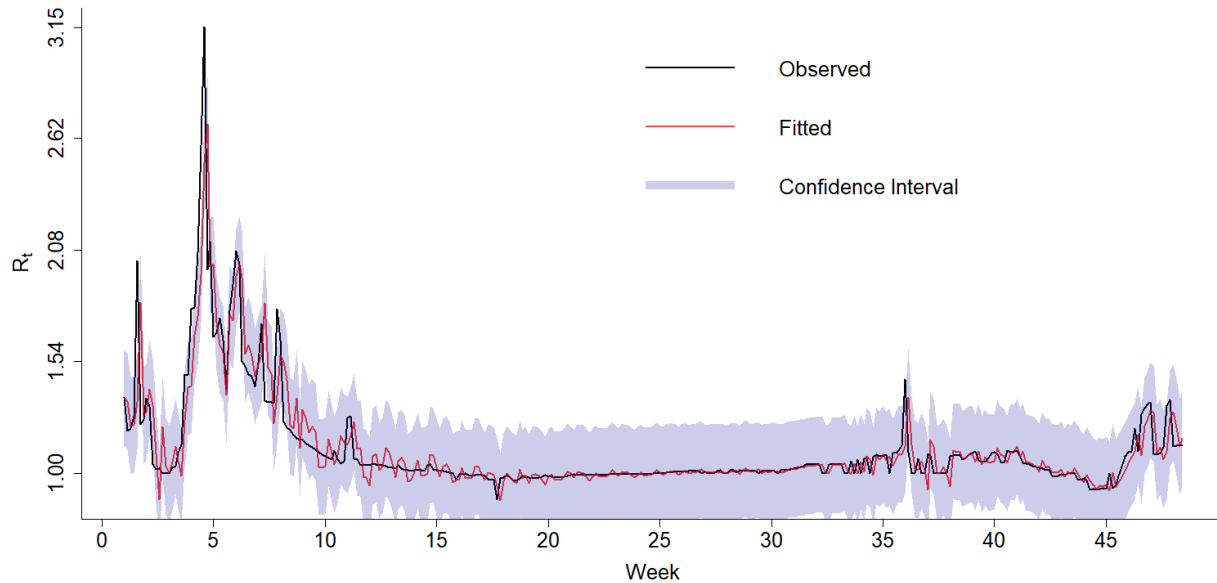


Figure 23: Plot of the univariate time series of transmission rate R_t , with fitted and observed values. Plotted using the "plot" function from the "graphics" package, the fitted values were added using the "lines" function from the "graphics" package, and the confidence interval was added using the "polygon" function from the "graphics" package.

Following this, I then used the model to make predictions about the daily transmission rate over the next 7 days.

Figure 24 shows that if the government introduced no new policies, then the daily transmission rate would increase over the next week. We see that whilst predictions show there would be an initial decrease, the transmission rate would increase overall. Predictions show that the transmission rate may have reached as high as $R_t \approx 1.25$ without new mitigation strategies. We can suggest that the prediction interval was large because the confidence interval was wider towards the end of observations. From the prediction interval, we see that transmission could have fallen to $R_t \approx 0.9$, or risen to $R_T \approx 1.4$.

To determine the accuracy of these predictions, I then compared the predicted values to the observed values.

Figure 25 shows the predicted transmission rate was lower than the actual transmission rate. It is worth noting that the observed rate of transmission was almost always within the prediction interval. While in the predicted values, we see a dip and then an upward trend, we see the reverse in the observed values. We expect transmission rates to fall during a lockdown (as seen in Figure 11, Figure 13, and Figure 15) which explains the fall in the observed transmission rate toward the end of the prediction interval.

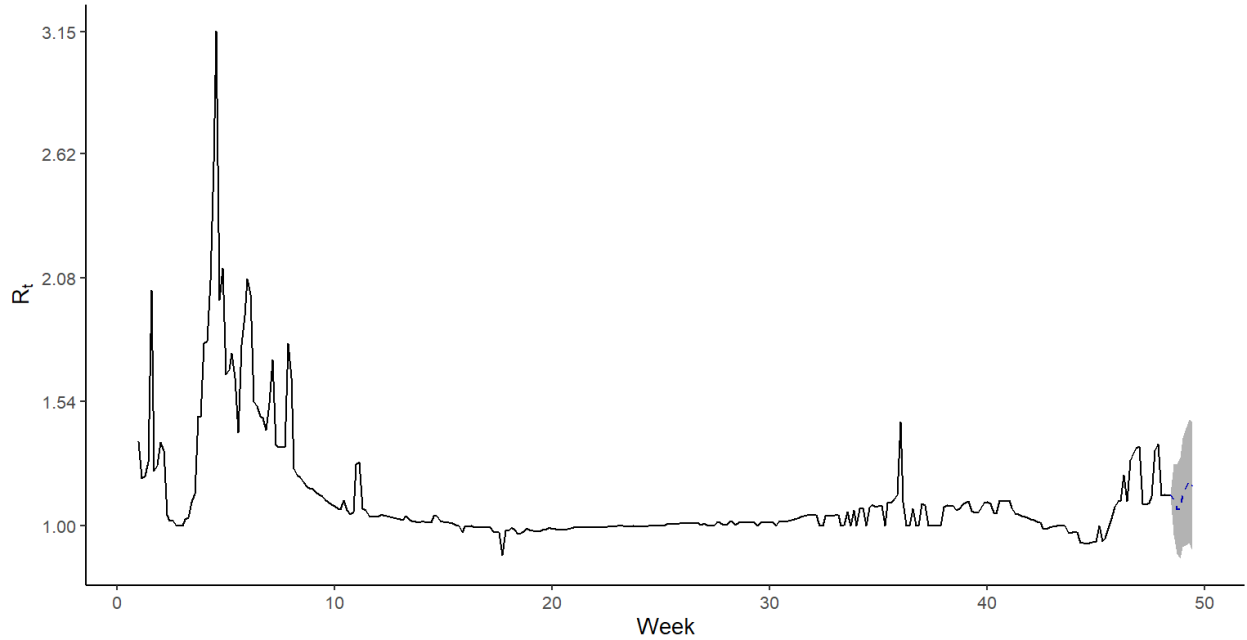


Figure 24: Plot of the predictions of transmission rate R_t from the univariate time series model. The blue dashed line indicates predictions, with the prediction interval in grey. Plotted using the "autoplot" function from the "forecast" and "ggfortify" packages¹²⁸.

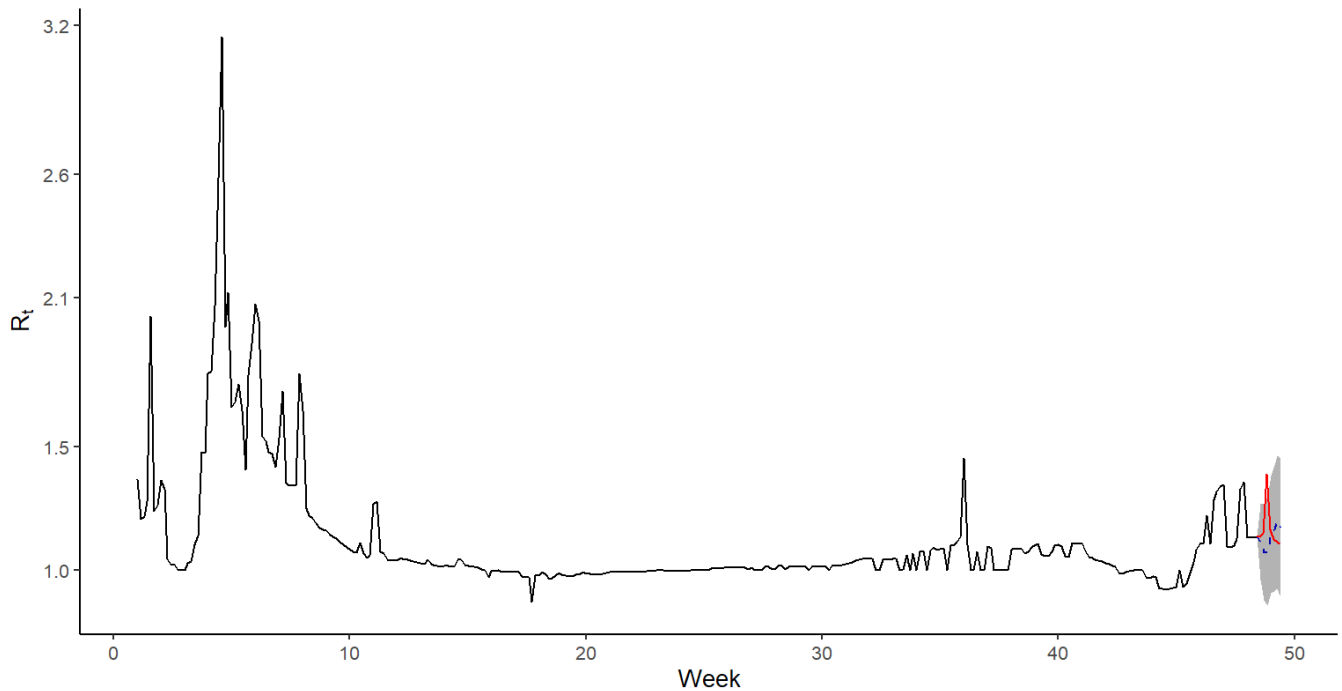


Figure 25: Plot of the comparison between the predicted and observed transmission rate from the univariate time series model. The blue dashed line indicates predictions, with the prediction interval in grey, and the red line indicates observed values. Plotted using the "autoplot" function from the "forecast" and "ggfortify" packages and the "geom_line" function¹²⁹ from the "ggplot2" package.

5.3.4 Multivariate Time Series Modelling

Finally, I modelled the three variables (daily confirmed cases, positivity rate, and transmission rate) together as a multivariate time series. The most suitable model identified by the "VARselect" function from the "vars" package was a VAR(10) model. A VAR(10) model means that the previous ten observations of each variable influenced the current value. I then ran the Phillips-Ouliaris test and found there was cointegration present. This result meant that the variables had a "long-run statistically significant relationship"⁹⁸, i.e. the variables influenced each other. Following this, I ran Johansen's maximum eigenvalue test using the "ca.jo" function from the "urca" package, and found one cointegrating relationship. This function also gave the VECM, which implements the cointegrating relationship found, that I then converted back into a new VAR model.

This final VAR model can be written as:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix} = \begin{bmatrix} \mu_{0,1} \\ \mu_{0,2} \\ \mu_{0,3} \end{bmatrix} + \begin{bmatrix} \theta_{11,1} & \theta_{12,1} & \theta_{13,1} \\ \theta_{21,1} & \theta_{22,1} & \theta_{23,1} \\ \theta_{31,1} & \theta_{32,1} & \theta_{33,1} \end{bmatrix} \times \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \\ x_{3,t-1} \end{bmatrix} + \dots + \begin{bmatrix} \theta_{11,10} & \theta_{12,10} & \theta_{13,10} \\ \theta_{21,10} & \theta_{22,10} & \theta_{23,10} \\ \theta_{31,10} & \theta_{32,10} & \theta_{33,10} \end{bmatrix} \times \begin{bmatrix} x_{1,t-10} \\ x_{2,t-10} \\ x_{3,t-10} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}.$$

Using the results from the R code, I found these values to three decimal places:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix} = \begin{bmatrix} 0.258 \\ -1734.773 \\ -0.001 \end{bmatrix} + \begin{bmatrix} 0.418 & 0 & -0.559 \\ -14092.736 & 0.163 & 111065.201 \\ -0.049 & 0 & 1.104 \end{bmatrix} \times \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \\ x_{3,t-1} \end{bmatrix} \\ + \begin{bmatrix} 0.166 & 0 & 0.362 \\ -552.441 & 0.618 & -137795.519 \\ -0.003 & 0 & -0.225 \end{bmatrix} \times \begin{bmatrix} x_{1,t-2} \\ x_{2,t-2} \\ x_{3,t-2} \end{bmatrix} + \begin{bmatrix} 0.089 & 0 & -1.544 \\ -7305.020 & 0.096 & 43291.538 \\ 0.005 & 0 & -0.137 \end{bmatrix} \times \begin{bmatrix} x_{1,t-3} \\ x_{2,t-3} \\ x_{3,t-3} \end{bmatrix} \\ + \begin{bmatrix} 0.097 & 0 & 1.686 \\ -4587.210 & -0.184 & 54903.166 \\ -0.002 & 0 & 0.355 \end{bmatrix} \times \begin{bmatrix} x_{1,t-4} \\ x_{2,t-4} \\ x_{3,t-4} \end{bmatrix} + \begin{bmatrix} -0.177 & 0 & 0.436 \\ 15285.315 & 0.014 & 30022.742 \\ 0.033 & 0 & 0.011 \end{bmatrix} \times \begin{bmatrix} x_{1,t-5} \\ x_{2,t-5} \\ x_{3,t-5} \end{bmatrix} \\ + \begin{bmatrix} -0.019 & 0 & -0.756 \\ -14031.214 & 0.641 & -97765.222 \\ -0.045 & 0 & -0.006 \end{bmatrix} \times \begin{bmatrix} x_{1,t-6} \\ x_{2,t-6} \\ x_{3,t-6} \end{bmatrix} + \begin{bmatrix} 0.229 & 0 & 0.074 \\ 24571.438 & 0.553 & 65966.612 \\ 0.077 & 0 & 0.396 \end{bmatrix} \times \begin{bmatrix} x_{1,t-7} \\ x_{2,t-7} \\ x_{3,t-7} \end{bmatrix} \\ + \begin{bmatrix} 0.074 & 0 & -0.477 \\ 1974.084 & -0.176 & -101419.311 \\ 0.034 & 0 & -0.477 \end{bmatrix} \times \begin{bmatrix} x_{1,t-8} \\ x_{2,t-8} \\ x_{3,t-8} \end{bmatrix} + \begin{bmatrix} -0.063 & 0 & 0.050 \\ -10223.236 & 0.180 & -5913.753 \\ -0.045 & 0 & -0.153 \end{bmatrix} \times \begin{bmatrix} x_{1,t-9} \\ x_{2,t-9} \\ x_{3,t-9} \end{bmatrix} \\ + \begin{bmatrix} -0.067 & 0 & 0.327 \\ 10672.601 & -0.914 & 40338.8256 \\ -0.004 & 0 & 0.133 \end{bmatrix} \times \begin{bmatrix} x_{1,t-10} \\ x_{2,t-10} \\ x_{3,t-10} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}.$$

I then compared the fitted model to the daily confirmed cases, positivity rate, and transmission rate data to determine the suitability of the model.

Figure 26 shows that the fitted values followed the observed values quite closely for each variable. We see that fitted values started after the beginning of the observations, due to the large value of p in the $\text{VAR}(p)$ model.

Following this, I used the model to make predictions about the daily number of confirmed cases, positivity rate, and transmission rate over the next 7 days.

Figure 27 shows that if the government introduced no new policies, then the daily number of new confirmed cases would increase over the next week. We see that whilst the increase would not be constant, the predictions had a general upward trend. Predictions show that the daily confirmed cases may have reached as high as $\approx 120,000$ without new mitigation strategies. We also see that the daily positivity rate would continue to vary, and in the process, it could have reached a higher value than previously seen – up to ≈ 0.3 . Lastly, we can see that the daily transmission rate may have drastically spiked over the next week, and may have reached as high as $R_t \approx 1.8$.

To determine the accuracy of these predictions, I then compared the predicted values to the observed values.

Figure 28 shows us that the predicted values were higher than the actual values for all variables. However, we can see that the pattern of the observed values was very similar to that of the predicted values. We should consider the possible influences of the lockdown when determining the accuracy of the predictions, i.e. we made predictions under the assumption that nothing would change and the implementation of policies likely influenced the observed data.

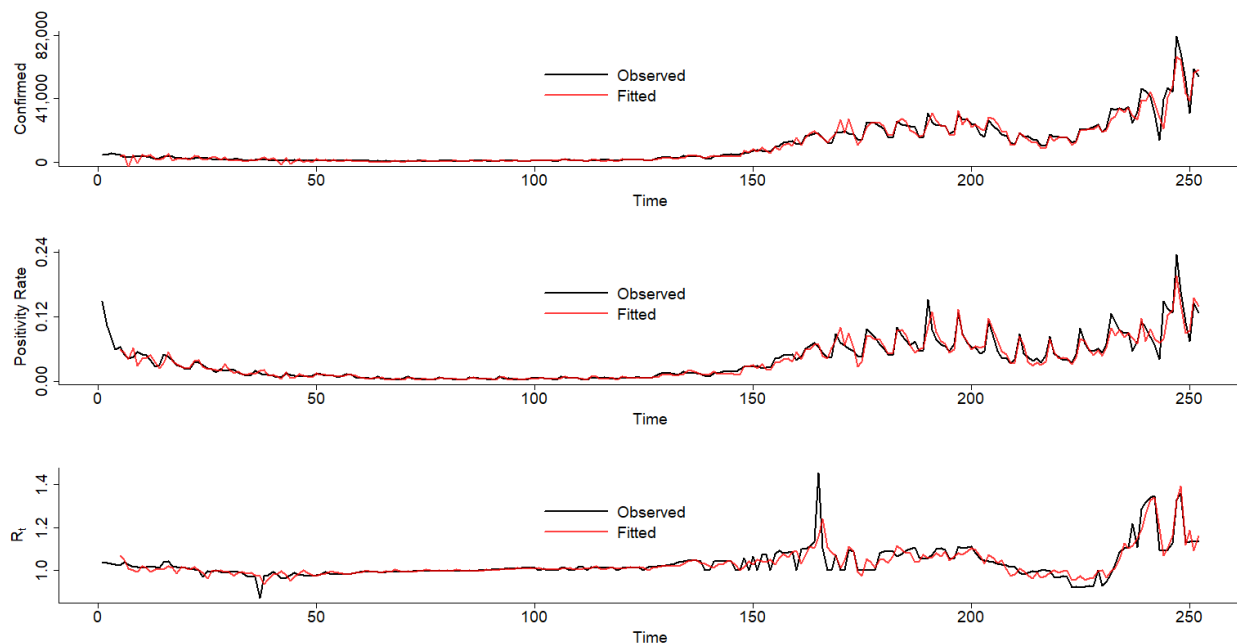


Figure 26: Plot of the multivariate time series of confirmed cases, positivity rate, and transmission rate, with fitted and observed values. Plotted separately using the "plot" function from the "graphics" package, the fitted values were added using the "lines" function from the "graphics" package, and were arranged using graphical parameters with the "par" function¹³⁰ from the "graphics" package.

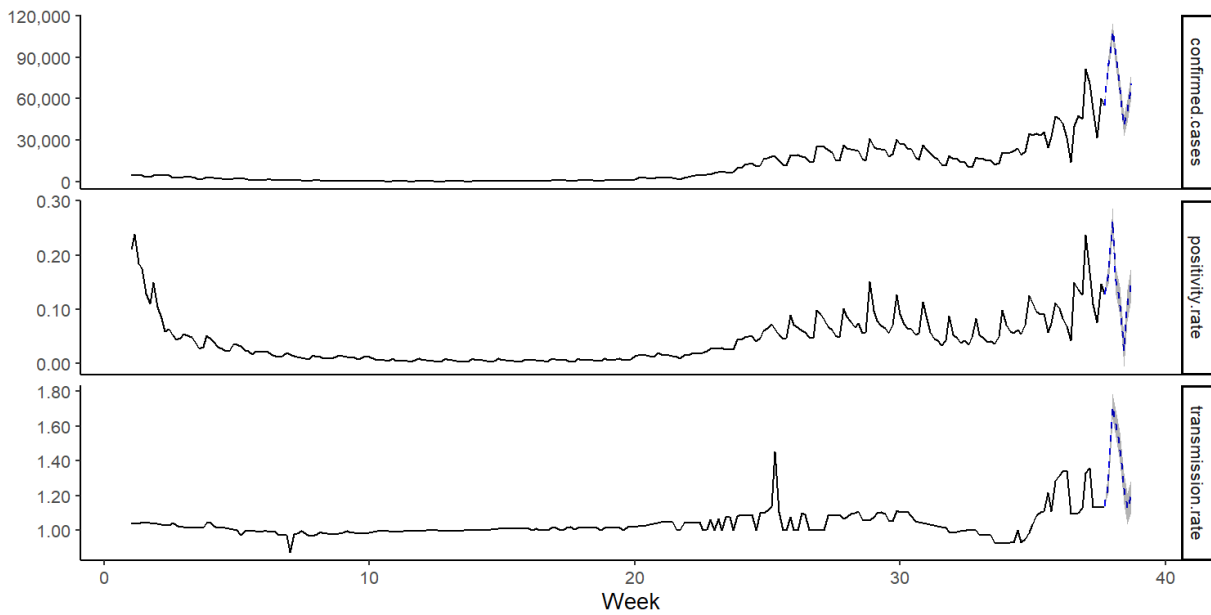


Figure 27: Plot of the predictions of time series of confirmed cases, positivity rate, and transmission rate, from the multivariate time series model. The blue dashed line indicates predictions, with the prediction interval in grey. Plotted using the "autoplot" function from the "forecast" and "ggfortify" packages¹²⁸.

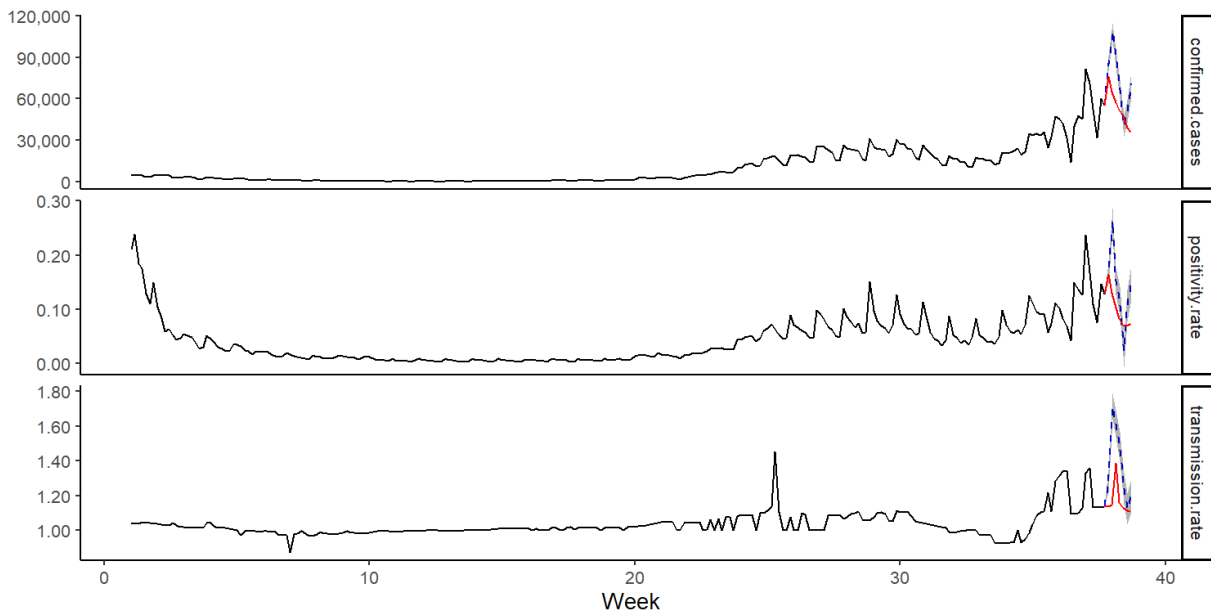


Figure 28: Plot of the comparison between the predicted and observed confirmed cases, positivity rate, and transmission rate from the multivariate time series model. The blue dashed line indicates predictions, with the prediction interval in grey, and the red line indicates observed values. Plotted using the "autoplot" function from the "forecast" and "ggfortify" packages and the "geom_line" function¹²⁹ from the "ggplot2" package.

5.3.5 Comparison of Time Series Models

To discover which model better predicted the behaviour of the data, I have compared Figure 27 with Figure 18, Figure 21, and Figure 24. By comparing these figures, we see that predictions from both model types were generally similar. While the predictions for transmission rates appear to have been quite different, the implications were the same – transmission rates would increase overall.

To compare the accuracy of the predictions, I have looked at the forecast errors¹³¹:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T},$$

where T is the time of the last observation, h is the number of steps ahead, e_{T+h} is the error at time $T + h$, y_{T+h} is the observed value at time $T + h$, and \hat{y}_{T+h} is the predicted value at time $T + h$.

Table 2 shows that the residuals were negative when the predicted values were higher than the observed values. We see that the majority of the residuals found were negative, regardless of the model type. The sign of the values may have been due to the implementation of the national lockdown on the 5th January 2021¹²⁵. We can also see that the residuals were quite variable throughout the predictions, and therefore so was the difference between them. We can infer that, in general, no one model type had smaller residuals in magnitude. However, those from the multivariate model were often larger than from the univariate models. This finding implies that the predictions from the multivariate model were less accurate.

Table 2: Forecast errors for both the univariate and multivariate models

Day	Confirmed Cases		Positivity Rate		Transmission Rate	
	Univariate	Multivariate	Univariate	Multivariate	Univariate	Multivariate
1	26,089.008	-9,161.406	0.039	-0.002	0.021	-0.149
2	-24,781.470	-43,188.695	-0.065	-0.132	0.076	-0.565
3	-16,922.802	-31,589.298	-0.046	-0.049	0.312	-0.226
4	-10,412.021	-14,662.154	-0.022	-0.030	0.018	-0.317
5	3,081.766	6,325.316	-0.001	0.047	-0.044	0.117
6	-29,970.074	-11,813.359	-0.111	-0.038	-0.080	-0.007
7	-28,133.588	-34,379.264	-0.081	-0.075	-0.062	-0.095

To verify this finding, I compared Figure 19, Figure 22, and Figure 25 with Figure 28. By comparing these figures, we see that the multivariate time series appeared to have predicted future behaviour more accurately. This conclusion is logical because a higher transmission rate would lead to higher confirmed cases and hence positivity rate. Therefore, modelling the three variables together should more accurately predict the behaviour of the variables. Whilst the residuals were larger for the multivariate model, the predicted behaviour appears to have been more accurate. The multivariate model may have overestimated the predictions due to the new mitigations strategies reducing the true values.

6 Contrasting Approaches

By looking at other approaches taken by different nations, I was able to look closer at the impact of government policy on transmission rates. From research based at The Oxford Martin Programme¹³², it was possible to retrieve data regarding confirmed cases and deaths from around the world¹³³. This investigation focused on New Zealand and Brazil as comparisons with the UK approach.

6.1 New Zealand

The first country that provided a good comparison was New Zealand. To make this comparison, I first identified differences between approaches. Then, through data visualisation, I explored infections and transmission rates.

6.1.1 Key Policy Differences

The first notable difference between the policies implemented in New Zealand, and those implemented in the UK, was the border controls placed on the nation. On the 3rd February, New Zealand placed flight bans on foreigners entering from China. Additionally, they placed self-isolation orders on residents re-entering the country from China¹³⁴. After extending these policies to other affected countries, it was on the 14th March 2020 that a policy required everyone entering New Zealand to self-isolate for 14 days¹³⁵. Following this, on the 20th March 2020, borders were officially closed to all but New Zealand citizens and residents¹³⁶. These border controls gave the country the ability to monitor where cases were – and their sources. For example, over 50% of their cases were imported¹³⁷ and identified at the border. Another noteworthy difference was the lockdown policy which New Zealand introduced, which aimed for elimination from the start¹³⁴. The four-tier system was introduced on the 21st March 2020, and the country entered Alert Level Four (the highest level) on the 25th March 2020¹³⁸. On the 4th May 2020, after moving to Alert Level Three, New Zealand saw no new cases. Subsequently, the country moved to Alert Level Two on the 14th May 2020¹³⁸. By the 8th June 2020, New Zealand had no ‘active’ cases of COVID-19 and moved down to Alert Level One the next day¹³⁸. Note that in New Zealand, ‘recovered’ individuals were “people who had the virus, where at least 10 days have passed since their symptoms started and they have not had symptoms for 72 hours, and they have been cleared by the health professional responsible for their monitoring”¹³⁹. We saw the differences in approach clearly, in the actions taken by the New Zealand Government when they found four new cases in the community. After discovering the cases in Auckland, the government responded by raising the lockdown level for the whole country to Alert Level Two and Auckland was raised to Alert Level Three¹³⁸. It was then on the 21st September 2020 that the country was moved back down to Alert Level One, with Auckland being moved to Alert Level Two on the 23rd September 2020 and then joined the rest of the country at Alert Level One on the 1st October 2020¹³⁸. Finally, a large difference between approaches was the public’s confidence in them. Not only did the World Health Organisation praise the country’s response, but polls also found that more than 80% of New Zealanders agreed with the measures the government were taking¹³⁴.

6.1.2 Data Visualisation

To build an understanding of infections in New Zealand, I first explored the confirmed cases.

Figure 29 shows most of the confirmed cases were in April 2020 and May 2020, exhibiting a clear spike in infections. We can see that throughout most of June 2020, there were no new cases and that from July 2020, there was a small number of new infections daily. We can also see in September 2020, there was a jump in the number of daily confirmed cases, and that numbers remained slightly elevated for the rest of the study period. Even at the peak of infections, New Zealand saw no more than 89 cases a day.

Following this, I modelled transmission rates to provide a deeper comprehension of the pandemic. To calculate transmission, I used a deterministic SIR model (see Section 4.1) and a stochastic epidemic model (see Section 4.2).

Figure 30 shows that the fitted values followed the observed values closely. We can see that the number of currently infected peaked in May 2020 at 340, and there was a small peak in September 2020. Due to the different definitions of 'recovered' (see Section 6.1.1), the number of currently infected does not reach zero even when there were no 'active' cases, according to the New Zealand Government.

After I confirmed that the deterministic SIR model was appropriate, I then estimated transmission over time.

Figure 31 shows there was a drop in R_t in mid-March 2020, likely due to the low number of new cases being reported after the initial one (see Figure 29). We can see that the rate of transmission peaked in April 2020 at $R_t \approx 4.20$ and fell sharply afterwards, to below 1 in May 2020. Throughout June 2020, we see an upward trend in R_t with it having reached $R_t \approx 1.4$ in July 2020, after which it remained at ≈ 1 . Following this, in September 2020, there was a spike in transmission rates which peaked at $R_t \approx 1.75$. We then see that rates fell and went below 1 in October 2020. Subsequently, R_t remained close to 1 for most of the remainder of the study period.

I then used the stochastic epidemic model to further estimate transmission rates.

Figure 32 shows a wide credible interval around June 2020 to July 2020, which indicates that the posterior was less informed at this time. This lack of information was because when new cases started to appear in late June 2020, there had been no new cases for most of that month. Due to this, the posterior had little information when the new cases occurred.

Figure 33 shows the posterior mean of $R_{t,\tau}$ fell from April 2020 until June 2020, where we see there was a sharp rise. At the end of June 2020, rates peaked, with a dramatic spike at $R_{t,\tau} \approx 15$, and then fell throughout July 2020. Subsequently, in August 2020, we see there was a slight rise in the rate of transmission. Following this rise, we see a spike, which peaked at $R_{t,\tau} \approx 4$, returning to 1 in September 2020. From this point, the rate of transmission remained around 1 on the whole, other than a spike in November 2020 with

it reaching $R_{t,\tau} \approx 2.5$.

Figure 34 shows credible interval was wider at the beginning of estimations, from June 2020 to July 2020, and at the end of August 2020. The interval widened when there was a drastic increase in the number of confirmed cases. The first occurrence of this was when COVID-19 was beginning to circulate. The second occurrence was when there had been no new infections for a while. The third and final occurrence was when there was a spike in the number of cases. These occurrences suggest that when there were low amounts of infections in the population, estimations were less informed.

Figure 35 shows the posterior mean $R_{t,\tau}$ was ≈ 6.75 in April 2020, after which it began to fall. From the end of April 2020 we see $R_{t,\tau}$ below 1, where it remained until mid-June 2020. At the end of June 2020 the rate rose quickly, and peaked at $R_{t,\tau} \approx 13$ at the end of the month. We then see that rates fell throughout July 2020 to below 1, having spiked twice in August 2020 at $R_{t,\tau} \approx 3$ and $R_{t,\tau} \approx 5$ respectively. From September 2020, we can observe that $R_{t,\tau} \approx 1$ apart from a spike in November 2020 where it reached $R_{t,\tau} \approx 3.25$.

Figure 36 shows that the parametric and uncertain distributions followed a similar pattern. We see that the credible interval for the uncertain distribution was wider than the parametric, which we expect due to the nature of the distribution. We also see that a wide credible interval accompanied the spike at the end of June 2020. This spike was likely due to there being no new cases for 28 days before. This width means that the model had very little information to form the posterior distribution, and hence the credible interval became wider.

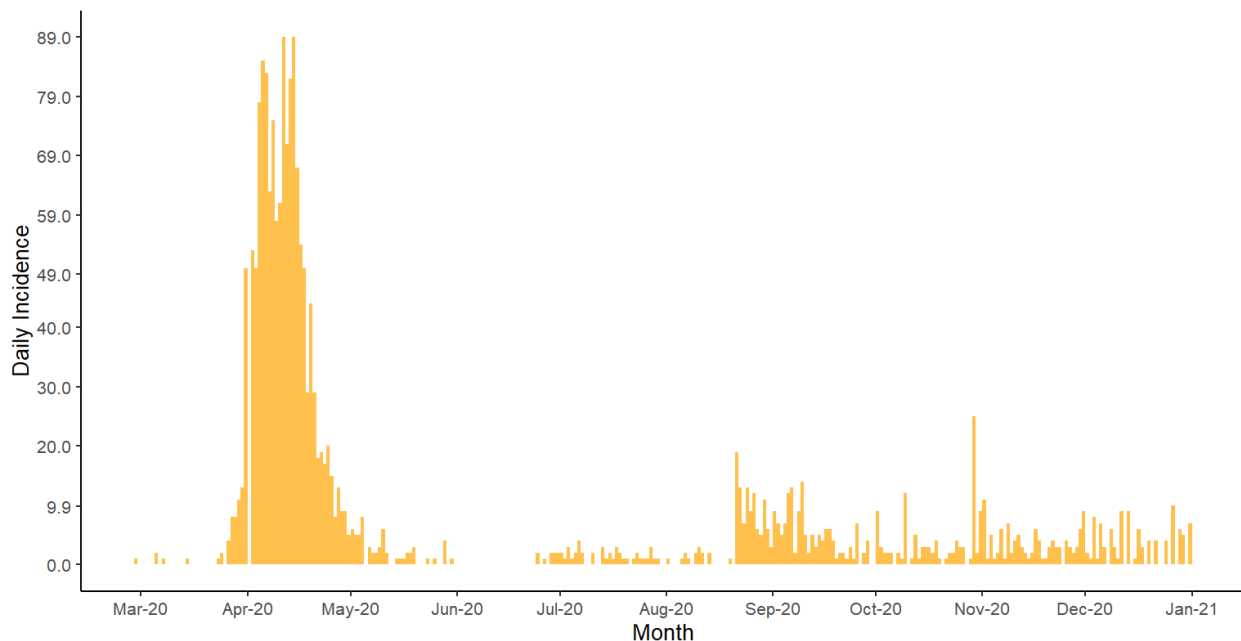


Figure 29: Incidence plot of daily confirmed cases in New Zealand. Plotted using the "plot" function from the "graphics" package. The data was manipulated using the "uncount" function from the "tidyr" package with weights = 'confirmed' and the "incidence" function from the "incidence" package.

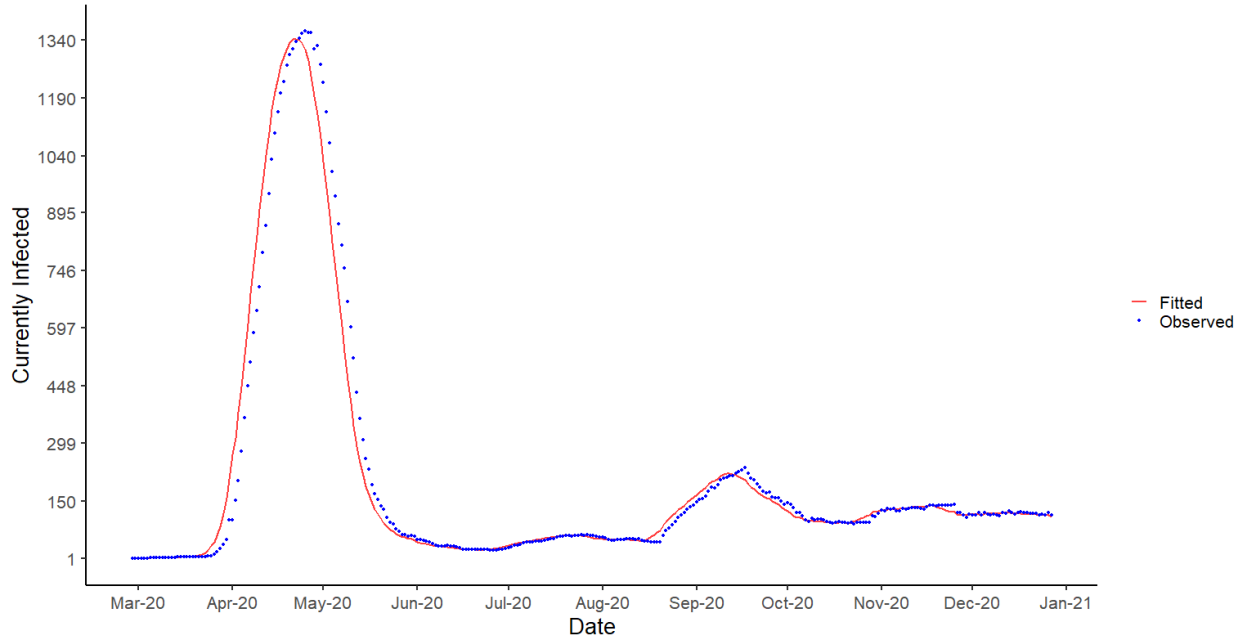


Figure 30: Plot of the fitted and observed data for New Zealand, using the SIR model, where blue points signify daily observed data points and the red line signifies the fitted data. Plotted using the "ggplot" function from the "ggplot2" package.

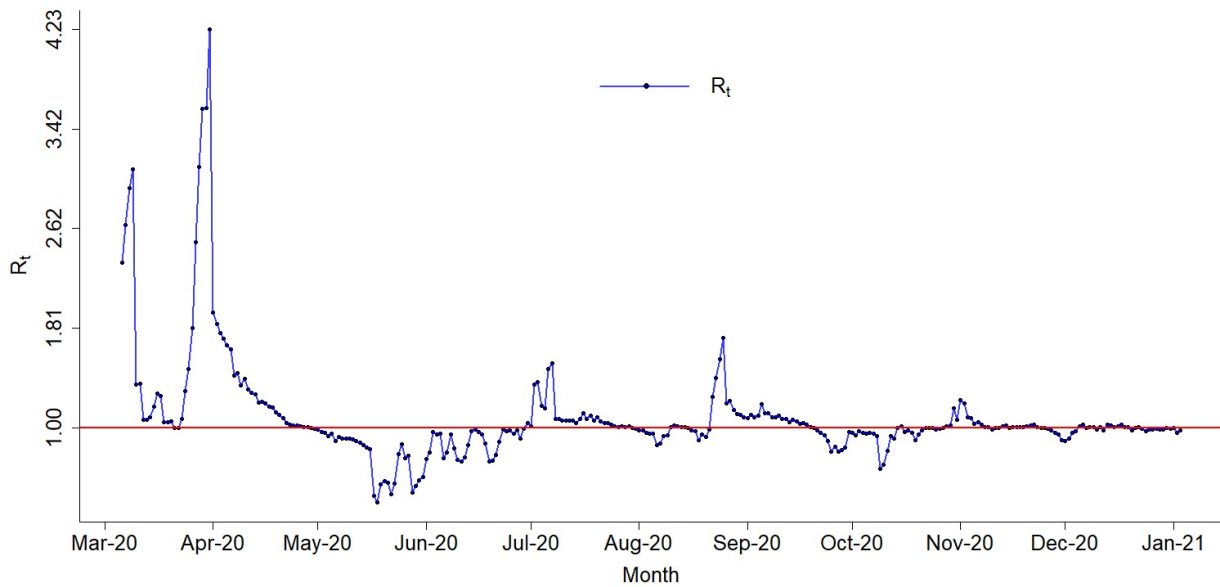


Figure 31: Plot of R_t over time for New Zealand. Plotted using the "plot" function from the "graphics" package, the blue line was added using the "lines" function from the "graphics" package, and the vertical lines were added using the "abline" function from the "graphics" package. A red horizontal line was added at 1 to indicate when $R_t = 1$, to show whether the epidemic is growing or decaying.

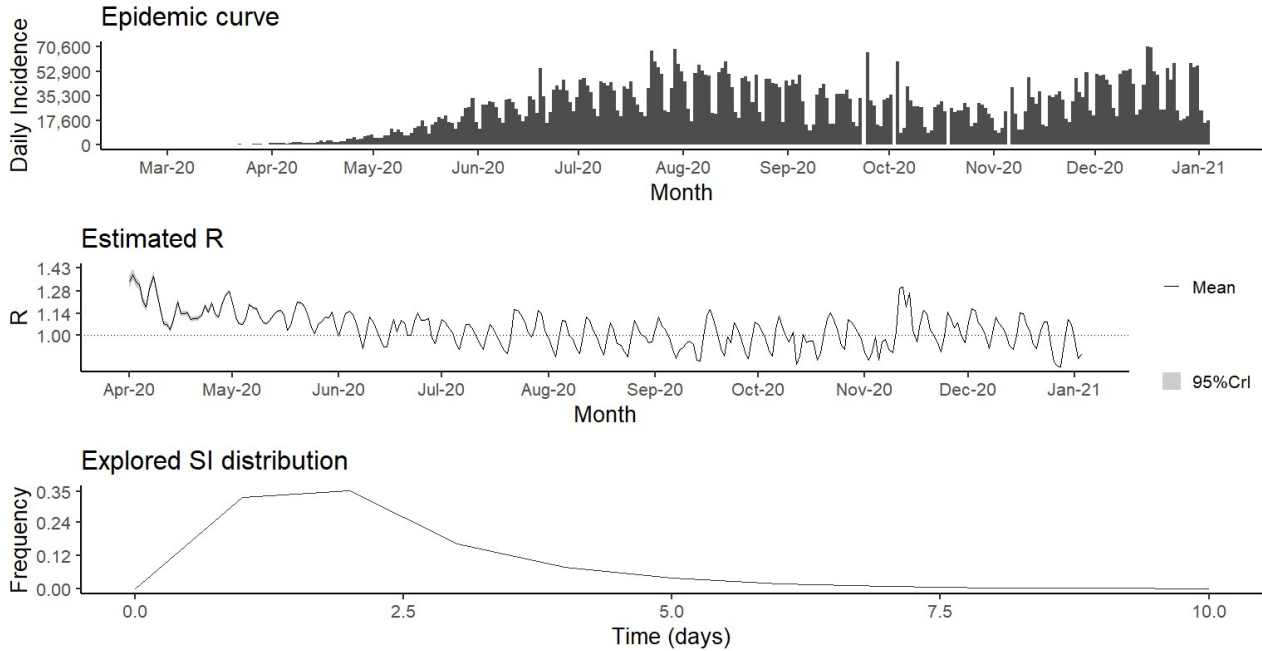


Figure 32: Plot of the output of the "estimate_R" function, using a parametric SI for New Zealand. Plotted using the "estimate_R_plots" function¹²⁴ from the "EpiEstim" package and arranged using the "ggarrange" function from the "ggpubr" package. The top plot shows the confirmed cases incidence data (see Figure 29). The middle plot shows the posterior mean of $R_{t,\tau}$ during each time window of τ days, with a 95% credible interval. The bottom plot shows the probability density function (PDF) of the shifted Gamma distribution which has been explored.

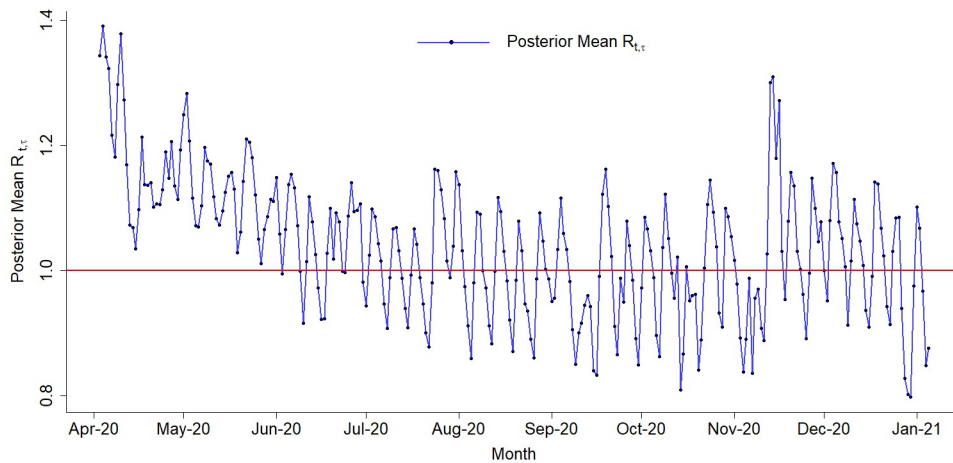


Figure 33: Plot of the posterior mean of $R_{t,\tau}$ for New Zealand, for the parametric SI during each time window of τ days. Plotted using the "plot" function from the "graphics" package, the blue line was added using the "lines" function from the "graphics" package, and the horizontal line was added using the "abline" function from the "graphics" package. A red horizontal line was added at 1 to indicate when the posterior mean $R_{t,\tau} = 1$, to show whether the epidemic is growing or decaying.

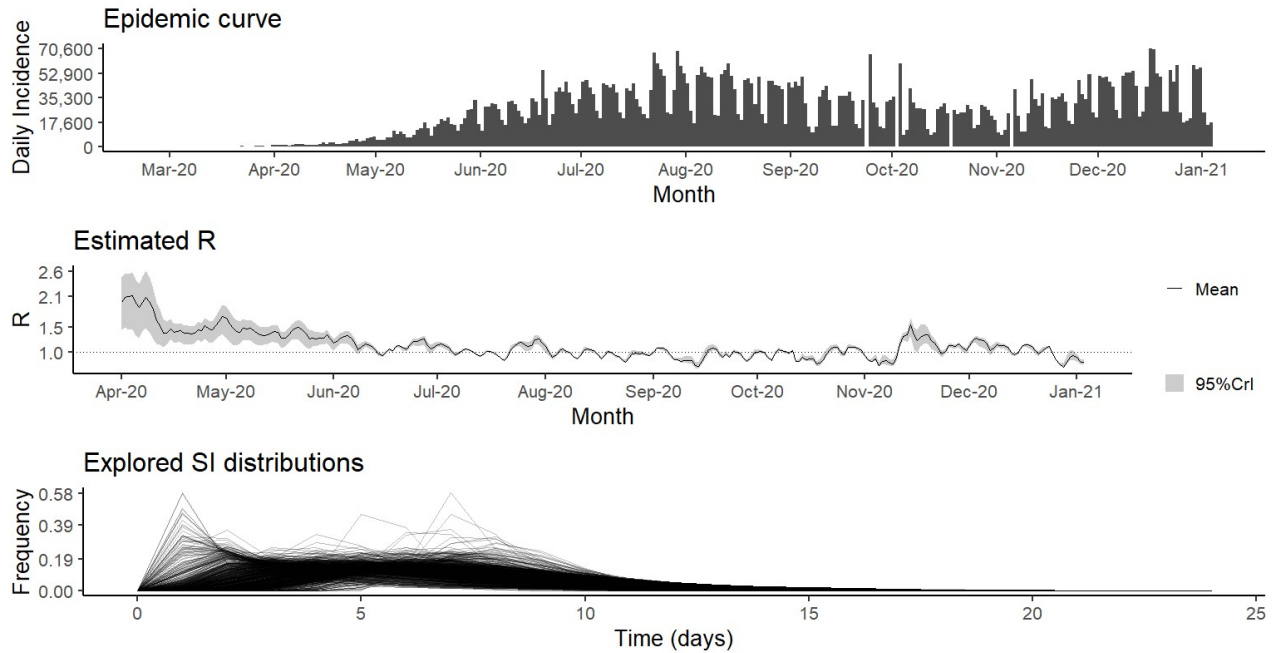


Figure 34: Plot of the output of the "estimate_R" function, using an uncertain SI for New Zealand. Plotted using the "estimate_R_plots" function from the "EpiEstim" package and arranged using the "ggarrange" function from the "ggpubr" package. The top plot shows the confirmed cases incidence data (see Figure 29). The middle plot shows the posterior mean of $R_{t,\tau}$ during each time window of τ days, with a 95% credible interval. The bottom plot shows the probability density functions (PDF) of the shifted Gamma distribution which have been explored.

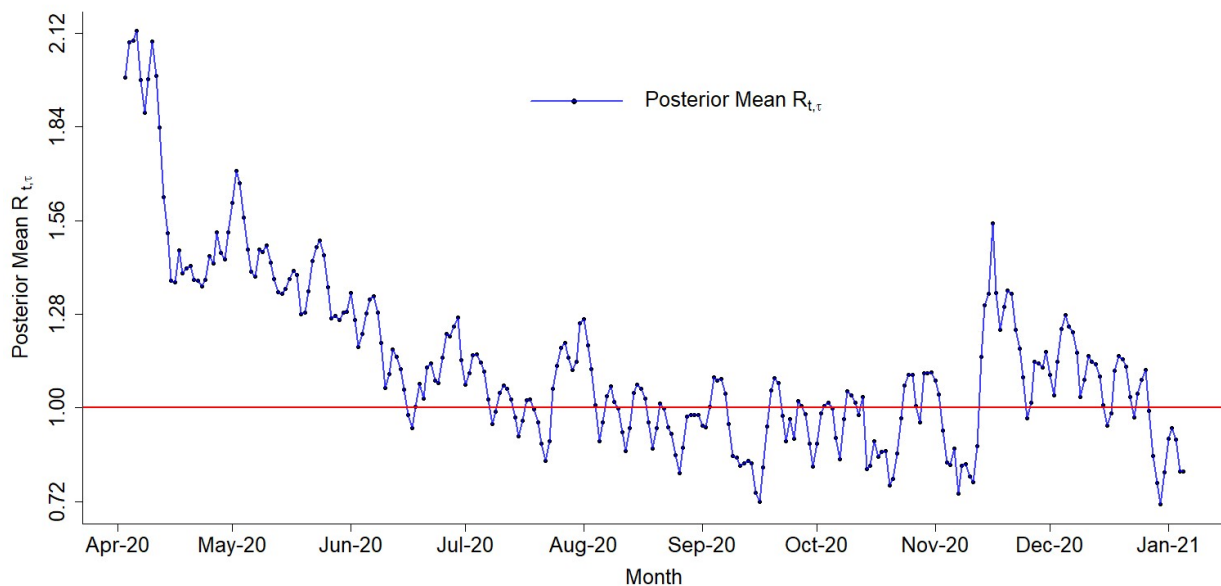


Figure 35: Plot of the posterior mean of $R_{t,\tau}$ for New Zealand, for the uncertain SI during each time window of τ days. Plotted using the "plot" function from the "graphics" package, the blue line was added using the "lines" function from the "graphics" package, and the horizontal line was added using the "abline" function from the "graphics" package. A red horizontal line was added at 1 to indicate when the posterior mean $R_{t,\tau} = 1$, to show whether the epidemic is growing or decaying.

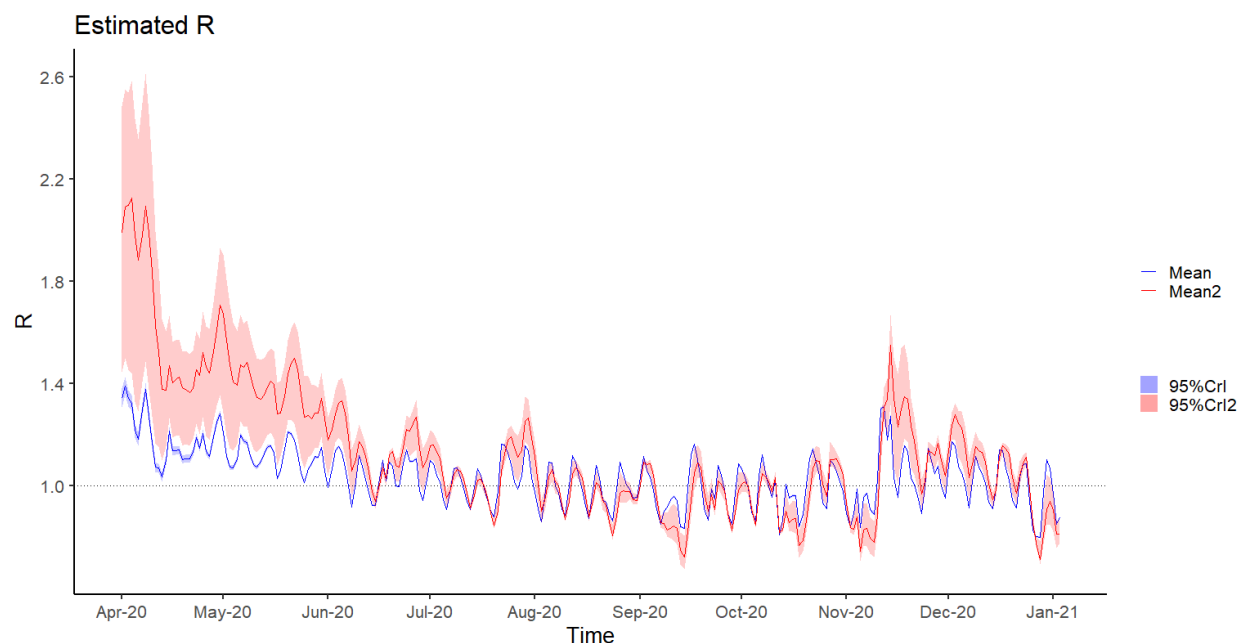


Figure 36: Plot of the posterior means of $R_{t,\tau}$ from both the parametric and uncertain distributions for New Zealand. Plotted using the "estimate_R_plots" function from the "EpiEstim" package. The parametric SI is indicated by the blue line and the uncertain SI is indicated by the red line. The associated credible intervals are indicated by the respective coloured fillings.

6.2 Brazil

The second country which provided a good comparison was Brazil. To make this comparison, I identified differences in approach. After this, I used data visualisation to explore infections and transmission rates.

6.2.1 Key Policy Differences

The principal difference between the approach taken in Brazil, and that taken in the UK, was the lockdown policies. Brazil did not have a national lockdown at any point, and only certain municipalities held local lockdowns, which residents protested¹⁴⁰. Not only this, but President Bolsonaro repeatedly downplayed the risks posed by COVID-19 and pushed municipalities to lift their restrictions¹⁴¹. By the 20th June 2020, Brazil was the second country in the world to surpass one million cumulative cases¹⁴⁰, and on the 7th July 2020 President Bolsonaro tested positive for COVID-19¹⁴². While Brazil saw record high case numbers, experts questioned whether the actual number of cases was higher due to a lack of testing¹⁴⁰. Researchers found in April 2020 that the actual number of cases could have been up to eight times higher than published¹⁴³ and studies in November 2020 found that the number could have been ten to twelve times higher than reported¹⁴⁴. Testing strategies marked another difference between approaches. A report published in September 2020, when Brazil had recorded more than four million cumulative cases, found that officials had distributed less than one-third of the available PCR tests¹⁴⁵. Another difference between approaches was the evolution of border controls. In March 2020, Brazil restricted the entry of foreigners by any means. However, from the 29th July 2020 restrictions no longer applied to air travel¹⁴⁶. This change was

influential because of the proportion of travel that occurs via air and the fact that Brazil did not impose quarantine measures on those entering the country¹⁴⁶. Finally, an important difference between approaches was the public's confidence in them. In April 2020, whilst individual municipalities were trying to implement safety measures, President Bolsonaro joined anti-lockdown protests, describing the participants as "patriots"¹⁴⁷. A month later, in May 2020, officials estimated that less than 50% of people were following self-isolation requests by São Paulo officials, one of the country's worst-hit areas¹⁴⁸.

6.2.2 Data Visualisation

To build an outline of the pandemic in Brazil, I initially inspected the confirmed cases.

Figure 37 shows that the number of daily confirmed cases did not fall below ≈ 8000 from mid-May 2020. We can see four dates with 0 confirmed cases reported and a high number of cases the following day, which was likely due to issues with reporting. We also see a downward trend in infections from August 2020 until it trended upward in November 2020. High levels of immunity gained throughout the beginning of the pandemic (see Section 6.2.1) and the implementation of some preventative measures¹⁴⁴ may have influenced this downward trend. Local authorities cited relaxed restrictions and mass gatherings as the cause of the rise¹⁴⁹ in November 2020.

Figure 38 shows that the fitted values followed the observed values closely. We can see the number of currently infected peaked in mid-August 2020. We then see that the number fell until November 2020, where it rose again – and reached higher than before at 1,350,000.

After I confirmed that the deterministic SIR model was appropriate, I then estimated transmission over time.

Subsequently, I modelled transmission rates to gain a more detailed understanding of the pandemic. To calculate transmission, I used a deterministic SIR model (see Section 4.1) and a stochastic epidemic model (see Section 4.2).

Figure 39 shows that transmission rates were highest in late February 2020 at $R_t \approx 4.6$, after which R_t decreased until April 2020. We see that the transmission rate was volatile throughout April 2020 and the beginning of May 2020 but then settled at $R_t \approx 1$ where it remained the rest of the study period. This result was likely due to the sustained high level of infections throughout the pandemic. We can observe that transmission rates were slightly below 1 from August 2020 until the end of October 2020 and slightly above 1 from November 2020. This slight change in the transmission rate reflects the change from the downward and upward trend in confirmed cases shown in Figure 37.

I then used the stochastic epidemic model to further estimate transmission rates.

Figure 40 shows that the credible interval was wider at the start of estimations. However, we see that the interval was very narrow from April 2020. The interval was narrow due to the high levels of infections informing the posterior, and the interval remained so throughout the study period.

Figure 41 shows the posterior mean $R_{t,\tau}$ fell in March 2020 from $R_{t,\tau} \approx 2.2$ and plateaued in June 2020 around $R_{t,\tau} \approx 1.1$. We can see that the transmission rate appeared to be

quite volatile throughout. However, the maximum value taken was quite small, meaning minor changes in transmission rates were apparent. We can also see that posterior mean $R_{t,\tau}$ was below 1 more often from July 2020 until mid-November 2020, where it rose above 1 more frequently.

Figure 42 shows a wide credible interval throughout March 2020, which narrowed in April 2020. By May 2020, the interval was very narrow. This narrowing shows us that the posterior became better informed throughout the study, which is unsurprising as the number of confirmed cases remained quite high for the extent of the study period.

Figure 43 shows the rate of transmission was highest in March 2020 at posterior mean $R_{t,\tau} \approx 7.8$ and fell consistently from April 2020 until mid June 2020, where it stabilised at $R_{t,\tau} \approx 1$. We can see that from August 2020 until mid-November 2020, transmission rates were mostly below 1. Subsequently, $R_{t,\tau}$ was almost entirely above 1 until the end of December 2020.

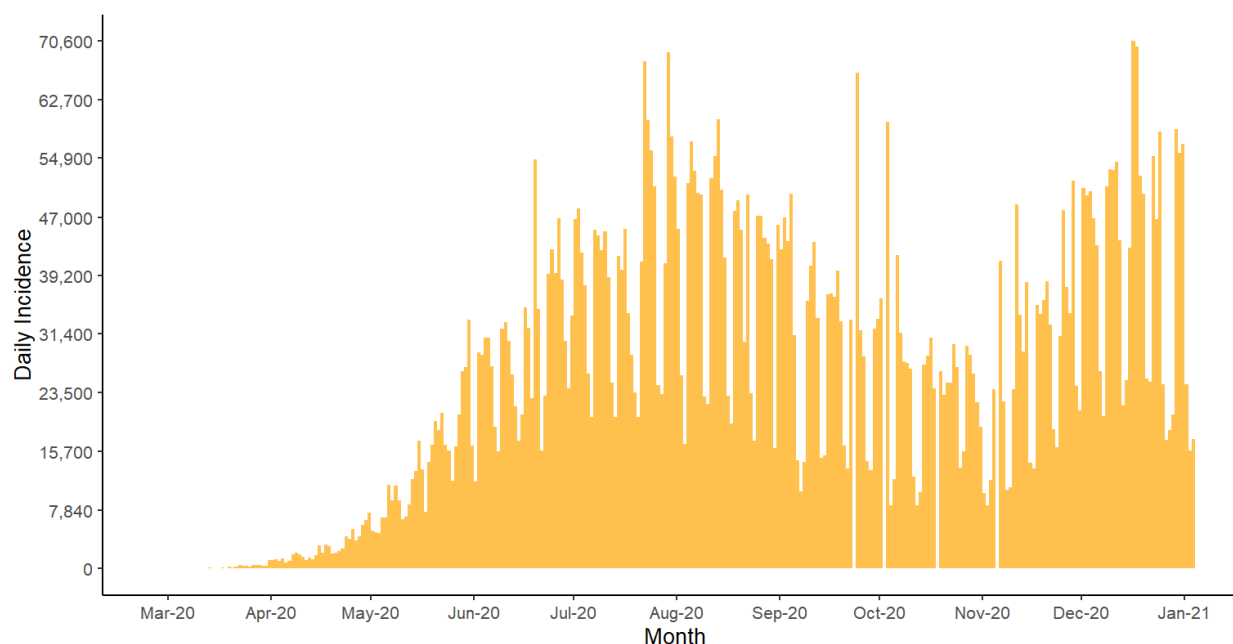


Figure 37: Incidence plot of daily confirmed cases in Brazil, plotted using the "plot" function from the "graphics" package. The data was manipulated using the "uncount" function from the "tidyr" package with weights = 'confirmed' and the "incidence" function from the "incidence" package.

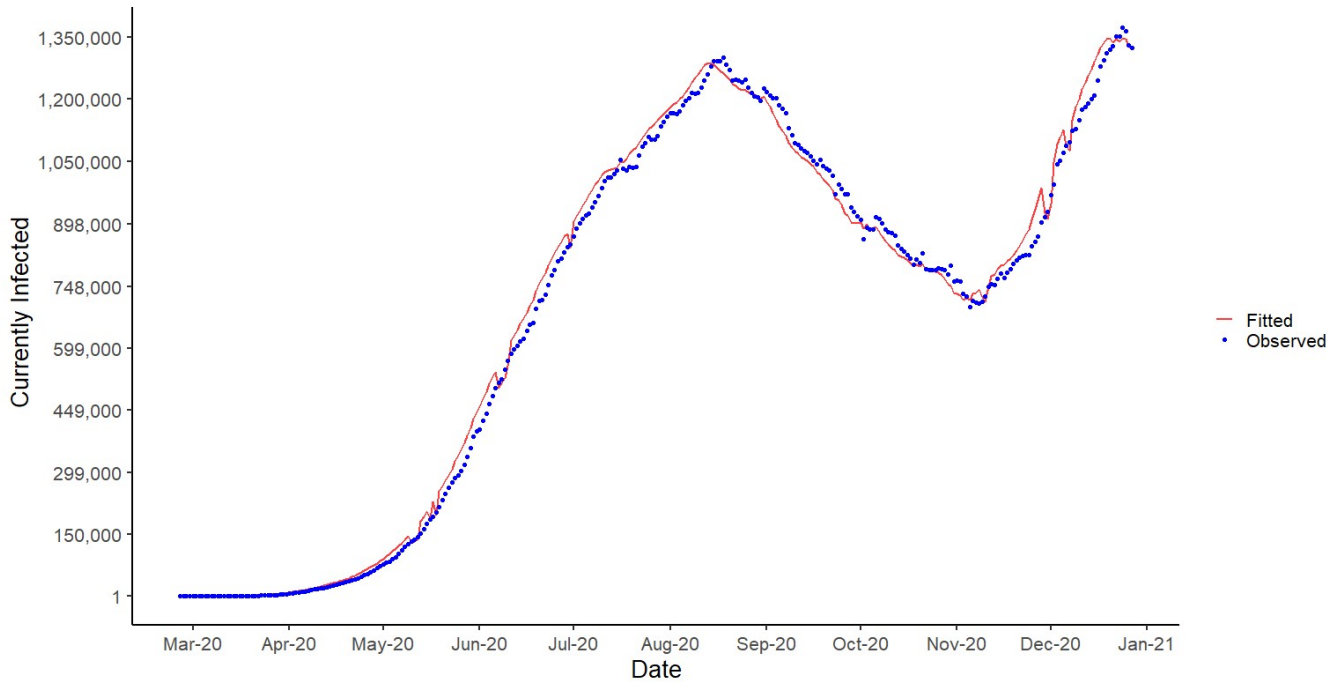


Figure 38: Plot of the fitted and observed data for Brazil, using the SIR model, where blue points signify daily observed data points and the red line signifies the fitted data. Plotted using the "ggplot" function from the "ggplot2" package.

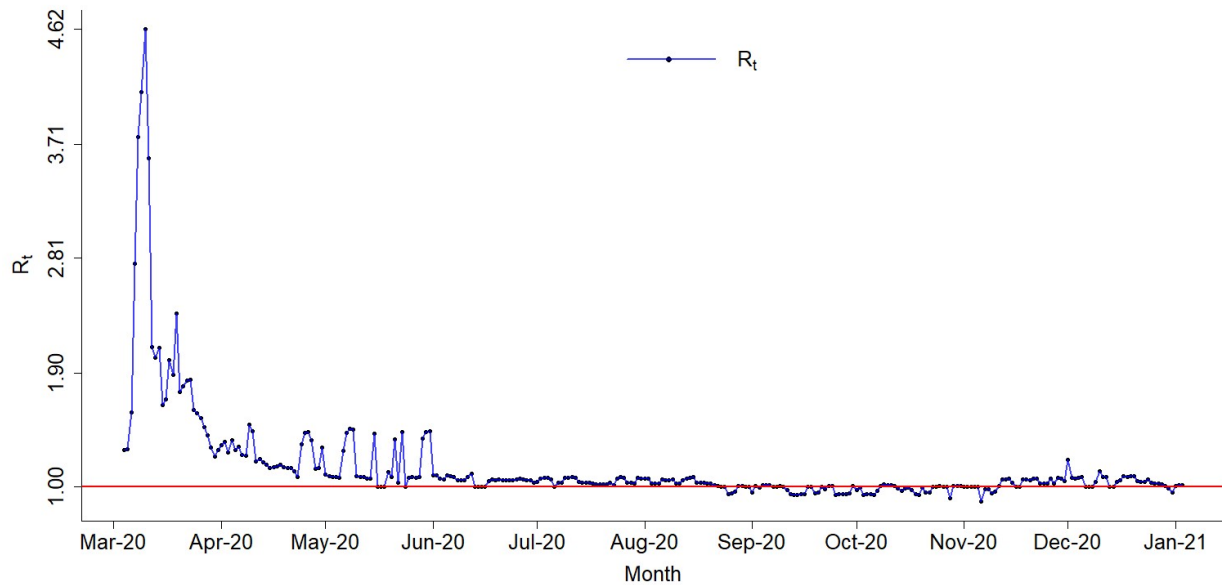


Figure 39: Plot of R_t for Brazil. Plotted using the "plot" function from the "graphics" package, the blue line was added using the "lines" function from the "graphics" package, and the horizontal line was added using the "abline" function from the "graphics" package. A red horizontal line was added at 1 to indicate when $R_t = 1$, to show whether the epidemic is growing or decaying.

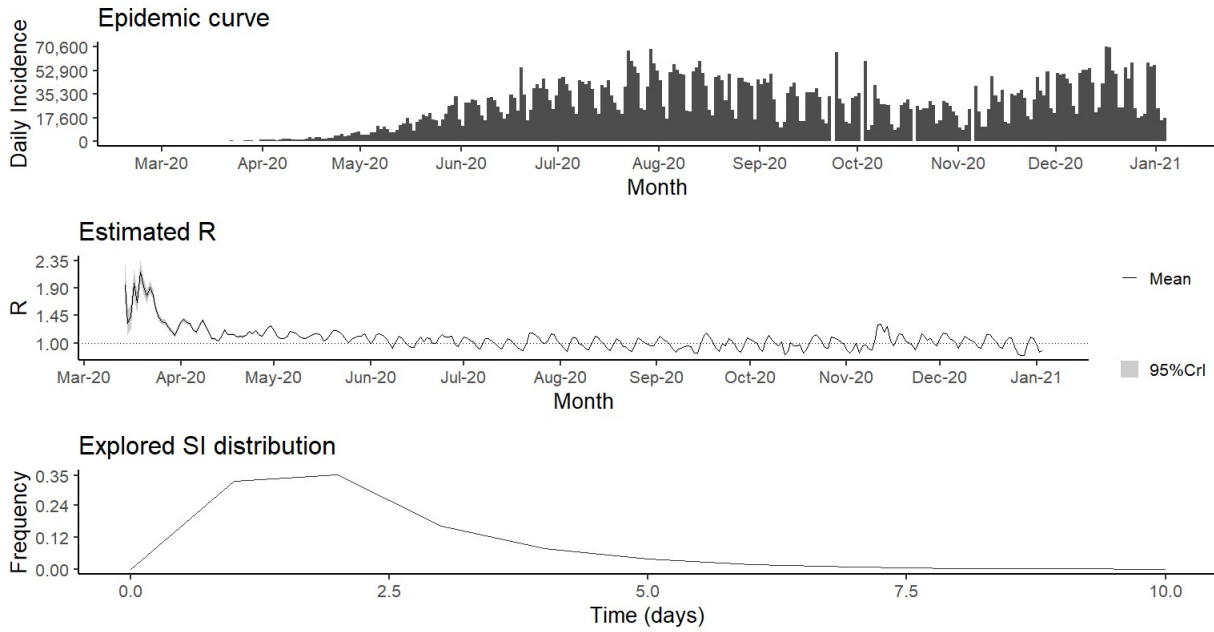


Figure 40: Plot of the output of the "estimate_R" function, using a parametric SI for Brazil. Plotted using the "estimate_R_plots" function¹²⁴ from the "EpiEstim" package, and arranged using the "ggarrange" function from the "ggpubr" package. The top plot shows the confirmed cases incidence data (see Figure 37). The middle plot shows the posterior mean of $R_{t,\tau}$ during each time window of τ days, with a 95% credible interval. The bottom plot shows the probability density function (PDF) of the shifted Gamma distribution which has been explored.

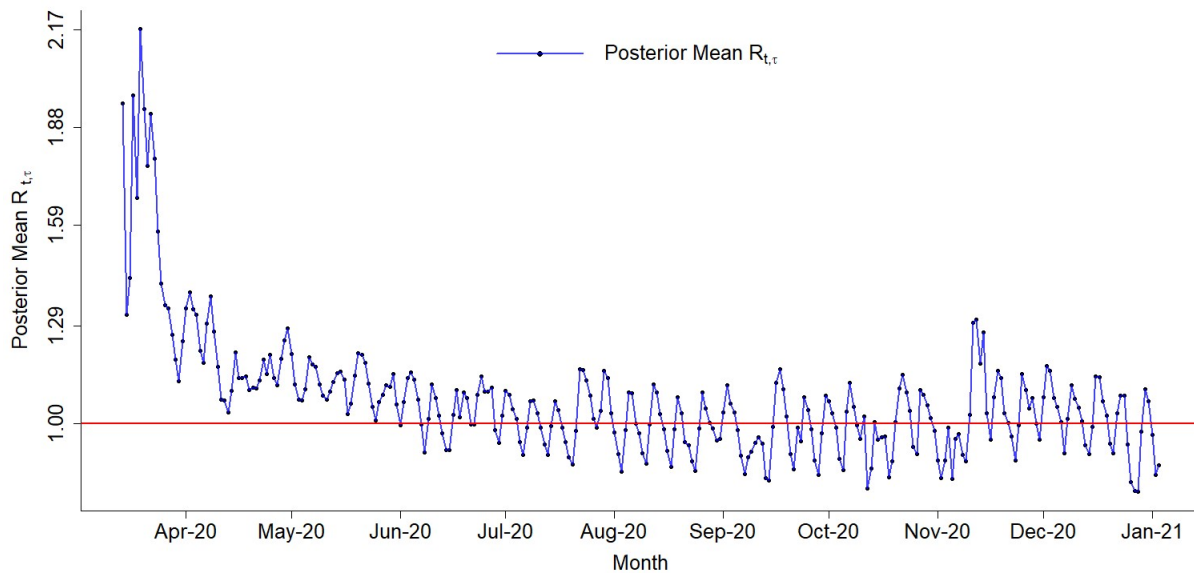


Figure 41: Plot of the posterior mean of $R_{t,\tau}$ for Brazil, for the parametric SI during each time window of τ days. Plotted using the "plot" function from the "graphics" package, the blue line was added using the "lines" function from the "graphics" package, and the horizontal line added using the "abline" function from the "graphics" package. A red horizontal line was added at 1 to indicate when the posterior mean $R_{t,\tau} = 1$, to show whether the epidemic is growing or decaying.

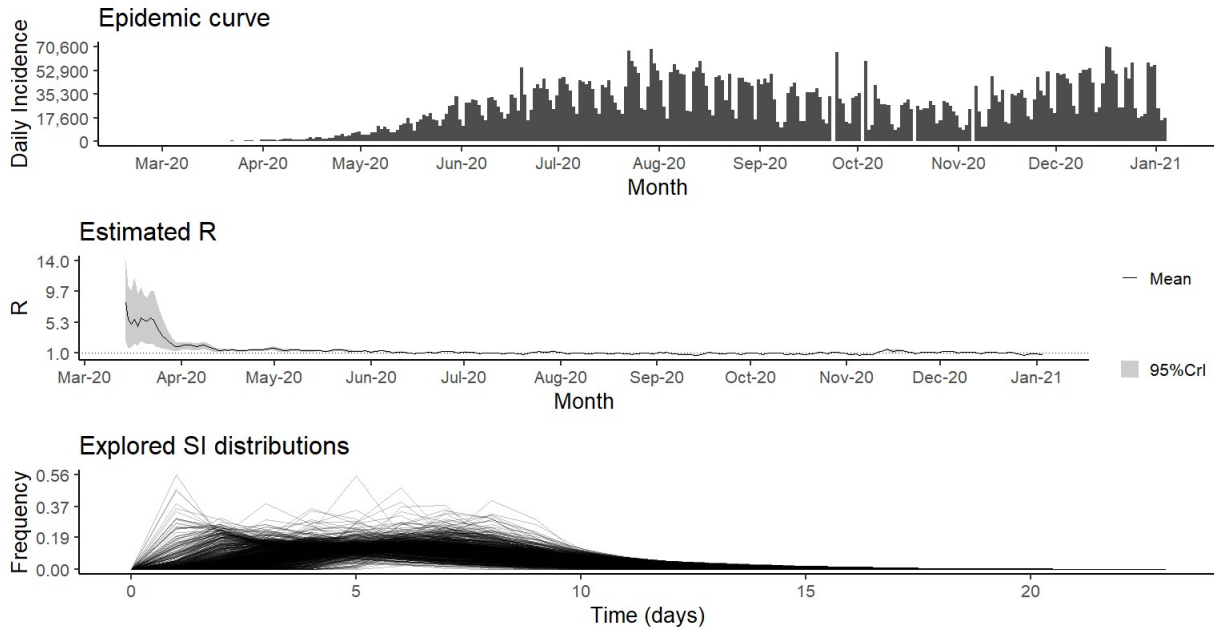


Figure 42: Plot of the output of the "estimate_R" function, using an uncertain SI for Brazil. Plotted using the "estimate_R_plots" function from the "EpiEstim" package, and arranged using the "ggarrange" function from the "ggpubr" package. The top plot shows the confirmed cases incidence data (see Figure 37). The middle plot shows the posterior mean of $R_{t,\tau}$ during each time window of τ days, with a 95% credible interval. The bottom plot shows the probability density functions (PDF) of the shifted Gamma distribution which have been explored.

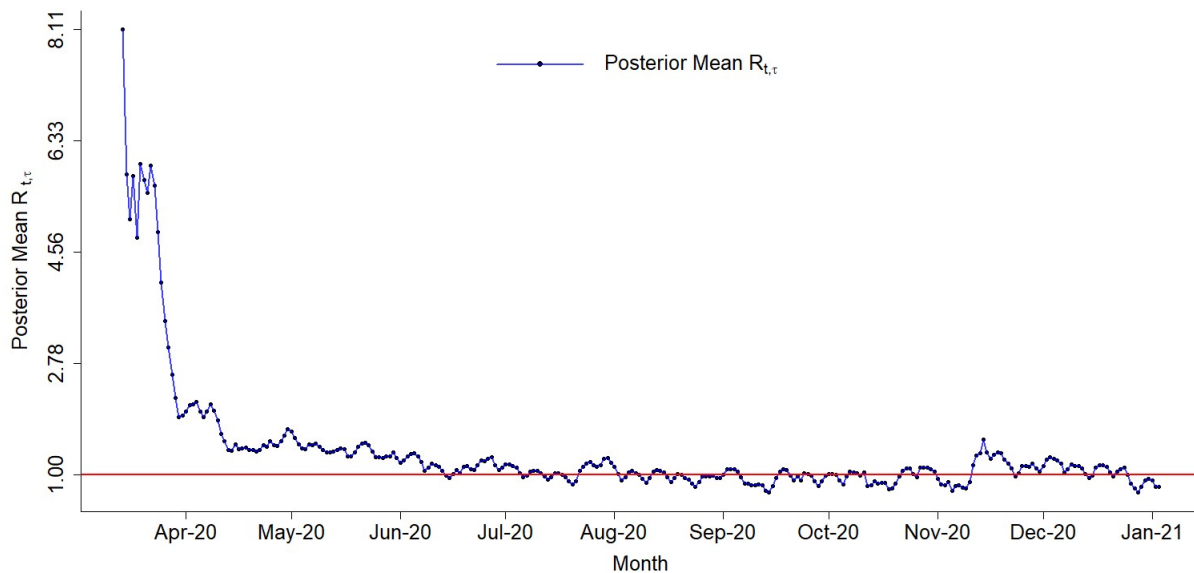


Figure 43: Plot of the posterior mean of $R_{t,\tau}$ for Brazil, for the uncertain SI during each time window of τ days. Plotted using the "plot" function from the "graphics" package, the blue line was added using the "lines" function from the "graphics" package, and the horizontal line was added using the "abline" function from the "graphics" package. A red horizontal line was added at 1 to indicate when the posterior mean $R_{t,\tau} = 1$, to show whether the epidemic is growing or decaying.

Figure 44 shows that at first, the uncertain distribution estimated higher values. We can see that the credible interval was much wider for the uncertain distribution, which we expect from its definition (see Section 5.2.2). We also see that from mid-June 2020, estimations of $R_{t,\tau}$ from both distributions followed the same pattern.

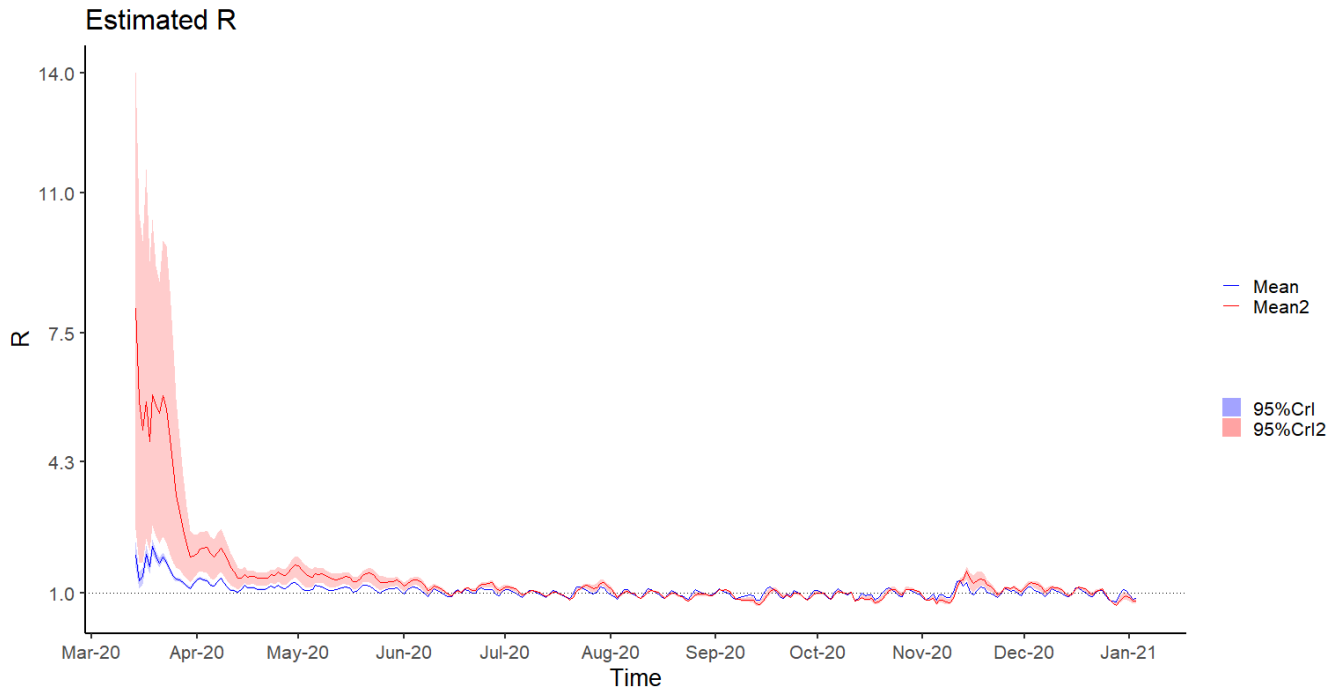


Figure 44: Plot of the posterior mean of $R_{t,\tau}$ from both the parametric and uncertain distributions for Brazil. Plotted using the "estimate_R_plots" function from the "EpiEstim" package. The parametric SI is indicated by the blue line and the uncertain SI is indicated by the red line. The associated credible intervals are indicated by the respective coloured fillings.

7 Discussion

7.1 Principle Findings

Section 3.2 shows there was a rapid rise in infections from December 2020, and Table 1 shows that the maximum number of confirmed cases was much higher than the upper quartile. This difference suggests that some data was being considered anomalous and was not included in the upper quartile. If infections continued at the trend seen in January 2021, we could expect the statistics to shift towards larger values. Figure 3 shows that the number of currently infected individuals in January 2021 was drastically increasing and had been since December 2020. This rise further demonstrates the growing number of infections seen in both Figure 1 and Figure 2. From Section 3.4 we can see that testing increased throughout the study period. If this increase accounted for the higher number of confirmed cases, we would expect the positivity rate to have either fallen or remained the same. However, from Figure 8 we can see that this was not the case and that from December 2020, the rate of positivity had an upward trend. Figure 11 shows that at the beginning of the pandemic, there was a lot of volatility in the value of R_t , and the transmission rate was much higher. We can explain this volatility by examining the limited

number of tests conducted at the beginning of the pandemic, the number of which was not recorded until May 2020 (see Section 3.4.2).

Figure 11, Figure 13, and Figure 15 show that the rate of transmission was highest in March 2020, after which we see there was a dramatic increase in the number of new confirmed cases in Figure 2. The higher transmission rates can be explained by the lack of measures in place at the start of the year, with many policies not being implemented until late March 2020. These figures suggest that lockdowns successfully reduced the transmission rate, especially in the first lockdown. This indicates there was a high level of community engagement with the lockdown policies. For good community engagement, there must have been a high level of confidence in leadership. Here, confidence likely stemmed from the clarity of the lockdown policy. The lockdowns also meant that people who could not afford to self-isolate (see Section 1.1.3) were more likely to be at home, meaning fewer people were exposed to infected individuals. Figure 11, Figure 13, and Figure 15 also show that transmission rates rose in September 2020, and around the same time, reports found that confidence in leadership had fallen (see Section 1.1.1). From this, we can suggest that confidence in leadership may have had a notable impact on transmission rates.

We can also consider the border control policies in place (see Section 1.1.5) as one of the contributing factors toward why rates have risen. Notably, lockdowns included harsher restrictions on international travel, which otherwise were relatively relaxed. Figure 2 shows that a rise in confirmed cases followed a rise in transmission. This delay appeared to be more pronounced at the beginning of the study period and was likely due to the methods used to monitor transmission (see Section 1.1.2), which have been improved throughout the pandemic. The delay in confirmed case numbers may also have delayed the implementation of policies, and without the correct timing, their influence was reduced. Another impact on this delay was likely the problems seen with the test and trace system (see Section 1.1.3), which may have also influenced the rate of transmission. The problems with the healthcare system (see Section 1.1.4) may have impacted the implementation of policies. It is possible that during the beginning of the pandemic, the struggling system contributed to the large numbers of deaths (see Figure 4), which occurred. Since then, officials have provided additional resources in the hope that the NHS can avoid further problems. Furthermore, the treatment of COVID-19 patients has evolved, meaning that more people are surviving. From Section 5.3 we can see that the mitigation strategies in place at the end of the study period were not enough as predictions suggested that the number of confirmed cases, rate of positivity, and rate of transmission, would follow an upward trend if the government implemented nothing new.

When we compare the UK with other countries, we further see the effect of government policies. If we compare Figure 2 with Figure 29, we can see a clear difference in the pattern of infections. Firstly, we can see that the peak of infections in New Zealand was drastically lower than that seen in the UK. The peak of confirmed cases in New Zealand was lower than even the first quartile of confirmed cases in the UK (as shown in Table 1). Secondly, we can see that after the peak of infections, New Zealand was able to successfully reduce case numbers and even when they had new cases discovered,

isolation at the border was implemented (see Section 6.1.1). When comparing the policies and their impact, we can see that border controls may have contributed to the success seen in New Zealand. If we look at the difference in public confidence, we see there were very different attitudes. Whilst 80% of people in New Zealand agreed with measures being taken (see Section 6.1.1), which suggests a high level of confidence in leadership, 57% of people in the UK did not believe the government could control the spread of COVID-19 (see Section 1.1.1). From this, we can infer that public confidence may have had an impact on the number of infections. If we compare Figure 11 with Figure 31, we can see that R_t was more regularly below 1 in New Zealand. This result means that the epidemic spent more time decaying. Whereas, in the UK, R_t was above 1 more regularly. This result tells us that the epidemic spent more time growing. We can see, when looking at Figure 16 and Figure 36 that transmission rates were much more volatile in New Zealand. However, here that indicated the low presence of COVID-19 before the identification of new cases.

If we compare Figure 2 with Figure 37, we can see that Brazil sustained a high number of daily confirmed cases throughout the study period. Whereas in the UK, there was evidence of infection waves. We also see that it was during lockdowns that cases in the UK began to fall. From this, we can suggest that lockdown policies may have had an impact on infections. When we compare the number of cases seen in Brazil regularly with the upper quartile of confirmed cases in the UK (as shown in Table 1), we see Brazil regularly exceeded the upper quartile. This difference implies that until recently, the UK saw fewer confirmed cases than Brazil. When making these comparisons, it is also important to recall that reports suggested the true number of cases in Brazil could be up to twelve times higher than reported (see Section 6.2.1). By comparing Figure 11 and Figure 39, we can see that R_t was more often above 1 in Brazil than in the UK, especially at the beginning of the pandemic. We also see that R_t went below 1 more regularly in Brazil but fell further in the UK, meaning that the epidemic decayed faster in the UK. This difference was likely down to the lockdown policies implemented in the two countries. In the UK, there were national lockdowns that caused the rate of transmission to fall dramatically. Whereas in Brazil, there was no national lockdown, and local lockdowns ended quickly. If we compare Figure 16 with Figure 44, we see similar patterns in the UK and Brazil. We also see in the UK that the posterior means of $R_{t,\tau}$ settled down to similar values earlier than in Brazil and that in Brazil, the values were more variable. This variability was likely down to the inconsistent measures that were in place. This variability suggests that consistent messaging and policies may have impacted transmission rates. The similarity we see between transmission rates in the UK and Brazil was likely due to the high number of cases seen in both countries.

7.2 Strengths and Limitations

By exploring a variety of sources for the qualitative data (information on government policies), I was able to gain different insights. For example, by looking directly at government publications, it was possible to look at the details of policies. However, by looking at news articles, we could get a sense of public opinion. Additionally, by comparing the approach in the UK with that of other countries, it was possible to strengthen my conclusions. With the use of two methods to estimate transmission rates, I

have corroborated the findings of this investigation, and the effect of government policy was easier to identify. Finally, the use of time series modelling to make predictions about key measures of spread has allowed me to suggest whether the measures in place at the time were sufficient.

However, owing to a lack of data on recovery, recovery was defined using the range of estimates of the infectious period and the methods used to estimate a COVID-19 death (see Section 2.1). Furthermore, calculations for the positivity rate were only ever approximations due to the lack of information available on the exact number of tests conducted, which were positive, per day. Because of the possible delay between when an individual takes a sample and the time spent testing it, there will have been an error in the proportion of positive results to tests conducted. An unforeseen limitation was the inability to account for the influence of new variants on the transmission of COVID-19. Given that experts believed these variants are more transmissible, they will have impacted how effective preventative measures were (see Section 1.1.2). Estimations and predictions depended largely on the set up of the models used. As a result of multivariate model estimates relating to all variables, it wasn't possible to provide individual confidence intervals for the fitted model of each variable. The stochastic epidemic model (see Section 4.2) was limited in the accuracy of estimations of $R_{t,\tau}$ due to the choice of time interval and serial interval distribution¹⁵⁰. Whilst there exist methods to select the optimum value of τ ⁹³, this investigation did not use them due to the complexity of the methods required. Also, the choice of serial interval distribution was highly variable between literature sources. For the deterministic SIR model (see Section 4.1), the choice of the interval impacted the estimations of R_t . Finally, due to the inaccuracy of detection rates, uncoordinated data systems, and human error, the data used throughout this investigation was unreliable. Consequently, I was limited in how confidently I could draw conclusions.

8 Conclusion

This investigation has indicated that unclear and poorly communicated government policies may have had a reduced impact on transmission rates and caused public confidence to fall. Moreover, the implementation of policies may have been too late because of outdated estimates of R_t . I have concluded that it is possible that when correctly communicated, government policy could dramatically slow the spread of COVID-19. Not only this, but the infrastructure created by government policy, such as NHS Test & Trace, may make or break an approach towards managing the pandemic. Also, the implementation of effective border control may have a drastic impact on the rate of transmission.

There are several things that the UK Government could have changed, which may have reduced transmission rates more effectively. First, officials could have clearly outlined policies to the public and stated the motivations behind them. Second, the state could have expanded testing to include close contacts of individuals who tested positive for COVID-19 to improve methods of monitoring transmission. Third, officials could have overhauled the tracing system to better the contact tracing to increase the effectiveness of mass testing. Fourth, officials could have improved adherence to self-isolation policies.

The only feasible way to have done this was by increasing motivation to comply. The state could have provided more comprehensive financial support to those isolating. Also, officials could have explained the benefits of self-isolation to the general public. Fifth, to minimise the impact of having relaxed border controls, inbound travellers could have been required to test negative for COVID-19 upon arrival. Additionally, they could have been required to self-isolate and subsequently test negative again, regardless of their country of origin.

Acknowledgements

I wish to express my gratitude to my supervisor, Dr. Yinghui Wei, whose support allowed me to conduct this investigation. I would also like to thank the lecturers in the Mathematics department; without their encouragement, I would never have begun this investigation. Their help over the last four years has encouraged me to carve out a new path for myself. Additionally, I would like to give a special mention to Dr. Jenny Sharp, who has supported me since I started my foundation year four years ago.

References

1. World Health Organization. *Naming the coronavirus disease (COVID-19) and the virus that causes it*. URL: [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it). (Last Accessed: 04.02.2021).
2. Centers for Disease Control and Prevention. *About COVID-19*. URL: <https://www.cdc.gov/coronavirus/2019-ncov/cdcreponse/about-COVID-19.html>. (Last Accessed: 04.02.2021).
3. The Lancet Respiratory Medicine. "COVID-19 transmission — up in the air". In: *The Lancet* (2020). URL: [https://www.thelancet.com/journals/lanres/article/PIIS2213-2600\(20\)30514-2/fulltext](https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(20)30514-2/fulltext). (Last Accessed: 08.02.2021).
4. David Adam. "A guide to R — the pandemic's misunderstood metric". In: *Nature* (2020). URL: <https://www.nature.com/articles/d41586-020-02009-w>. (Last Accessed: 29.11.2020).
5. Mary Hui. "Rt: The number that can guide how societies ease coronavirus lockdowns". In: *Quartz* (2020). URL: <https://qz.com/1834700/rt-the-real-time-r0-guiding-how-to-lift-coronavirus-lockdowns/>. (Last Accessed: 18.11.2020).
6. James Holland Jones. *Notes on R_0* . URL: <https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf>. (Last Accessed: 09.02.2021).
7. Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics". In: *American Journal of Epidemiology* (2013). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3816335/>. (Last Accessed: 15.11.2020).
8. Melisa Mei Jin Tan, Emeline Han, et al. "Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe". In: *The Lancet* (2020). URL: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(20\)32007-9/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)32007-9/fulltext). (Last Accessed: 20.11.2020).

9. Boris Johnson. *Prime Minister's statement on coronavirus (COVID-19): 16 March 2020*. URL: <https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-16-march-2020>. (Last Accessed: 20.11.2020).
10. Boris Johnson. *Prime Minister's statement on coronavirus (COVID-19): 23 March 2020*. URL: <https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020>. (Last Accessed: 14.02.2021).
11. Richard Vaughan. "Lockdown rules easing: a timeline of how coronavirus measures have gradually lifted since UK lockdown started". In: *The i* (2020). URL: <https://inews.co.uk/news/politics/lockdown-rules-easing-coronavirus-uk-measures-lifted-when-start-453497>. (Last Accessed: 18.09.2020).
12. Kate Devlin. "Coronavirus: First local lockdown imposed in Leicester as government orders schools and shops to close". In: *The Independent* (2020). URL: <https://www.independent.co.uk/news/uk/politics/coronavirus-leicester-lockdown-announcement-today-covid-19-cases-a9592406.html>. (Last Accessed: 18.09.2020).
13. Gareth Lacobuci. "Covid-19: Three tier alert system takes effect across England". In: *The British Medical Journal* (2020). URL: <https://www.bmj.com/content/371/bmj.m3961>. (Last Accessed: 13.02.2021).
14. David Hughes. "When is the next tier review? Date England's Covid restrictions will be reviewed again after latest update". In: *The i* (2020). URL: <https://inews.co.uk/news/uk/tier-review-when-next-covid-restrictions-england-rules-update-date-announcement-lockdown-map-812424>. (Last Accessed: 13.02.2021).
15. Grace C. Roberts. "Northern Ireland's circuit breaker lockdown: why now and will it work?". In: *The Conversation* (2020). URL: <https://theconversation.com/northern-irelands-circuit-breaker-lockdown-why-now-and-will-it-work-148216>. (Last Accessed: 13.02.2021).
16. BBC News. *Covid-19: Lockdown restrictions lift across NI*. URL: <https://www.bbc.co.uk/news/uk-northern-ireland-55251973>. (Last Accessed: 13.02.2021).
17. Stephen McDermott. "Covid-19: Northern Ireland re-enters 'circuit-breaker' lockdown after seven-day easing of restrictions". In: *The Journal* (2020). URL: <https://www.thejournal.ie/northern-ireland-circuit-breaker-covid-19-lockdown-december-5280550-Nov2020/>. (Last Accessed: 13.02.2021).
18. Welsh Government. *National coronavirus firebreak to be introduced in Wales on Friday*. URL: <https://gov.wales/national-coronavirus-fire-break-to-be-introduced-in-wales-on-friday>. (Last Accessed: 13.02.2021).
19. Tom Gillespie. "Coronavirus: Scotland heads for new tiered system as COVID-19 restrictions extended". In: *Sky News* (2020). URL: <https://news.sky.com/story/coronavirus-scotland-extends-covid-19-restrictions-as-nation-heads-for-new-tiered-system-12110159>. (Last Accessed: 13.02.2021).
20. Molly Blackall. "Boris Johnson announces four-week national Covid lockdown in England". In: *The Guardian* (2020). URL: <https://www.theguardian.com/world/2020/oct/31/boris-johnson-announces-national-lockdown-england-coronavirus>. (Last Accessed: 13.02.2021).
21. Prime Minister's Office. *Prime Minister announces Tier 4: 'Stay At Home' Alert Level in response to new COVID variant*. URL: <https://www.gov.uk/government/news/prime-minister-announces-tier-4-stay-at-home-alert-level-in-response-to-new-covid-variant>. (Last Accessed: 13.02.2021).

22. King's College London and Ipsos Mori.
"COVID-19: Majority of British public do not trust government to manage pandemic - survey".
In: *Sky News* (2020). URL: <https://news.sky.com/story/covid-19-majority-of-british-public-do-not-trust-government-to-manage-pandemic-survey-12152793>.
(Last Accessed: 14.02.2021).
23. Georgina Littlejohn. "When do UK schools go back? How Scotland, England, Wales and Northern Ireland's primary and secondary schools are reopening". In: *The i* (2020).
URL: <https://inews.co.uk/news/education/schools-uk-when-go-back-reopening-scotland-england-wales-northern-ireland-primary-secondary-date-575536>.
(Last Accessed: 18.09.2020).
24. James McAsh. "The government has pitted England's schools against health. It didn't have to".
In: *The Guardian* (2020).
URL: <https://www.theguardian.com/commentisfree/2020/dec/31/government-england-schools-health-teachers-pupils-closing>. (Last Accessed: 13.02.2021).
25. Robert Booth, Richard Adams, and Helen Pidd.
"Parents face week of uncertainty over school reopenings in England". In: *The Guardian* (2021).
URL: <https://www.theguardian.com/education/2021/jan/03/parents-face-week-of-uncertainty-over-school-reopenings-in-england>. (Last Accessed: 16.04.2021).
26. Cabinet Office. *Rules on travel over the Christmas Period*.
URL: <https://www.gov.uk/government/publications/rules-on-travel-over-the-christmas-period/rules-on-travel-over-the-christmas-period>. (Last Accessed: 14.02.2021).
27. Patrick Butler. "'Hyperbolic messaging' eroding public trust in UK's Covid response".
In: *The Guardian* (2020). URL: <https://www.theguardian.com/world/2020/oct/09/hyperbolic-messaging-eroding-public-trust-uk-covid-response>. (Last Accessed: 21.11.2020).
28. Boris Johnson. *Prime Minister's statement on coronavirus (COVID-19):17 March 2020*. URL:
<https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-17-march-2020>.
(Last Accessed: 16.12.2020).
29. Boris Johnson. *Prime Minister's statement on coronavirus (COVID-19):22 March 2020*. URL:
<https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-22-march-2020>.
(Last Accessed: 16.12.2020).
30. Department of Health and Social Care.
Face coverings to be mandatory in shops and supermarkets from 24 July.
URL: <https://www.gov.uk/government/speeches/face-coverings-to-be-mandatory-in-shops-and-supermarkets-from-24-july#:~:text=statement%20to%20Parliament,Face%20coverings%20to%20be%20mandatory%20in%20shops%20and%20supermarkets%20from,supermarkets%20from%2024%20July%202020>. (Last Accessed: 14.02.2020).
31. Department of Health and Social Care.
Face coverings: when to wear one, exemptions, and how to make your own. URL: <https://www.gov.uk/government/publications/face-coverings-when-to-wear-one-and-how-to-make-your-own/face-coverings-when-to-wear-one-and-how-to-make-your-own#exemptions>.
(Last Accessed: 19.04.2021).
32. The Health Foundation and Ipsos Mori.
"Coronavirus: Public confidence in government's handling of COVID-19 pandemic drops - survey".
In: *Sky News* (2020). URL: <https://news.sky.com/story/public-confidence-in-governments-handling-of-coronavirus-pandemic-drops-survey-12064867>. (Last Accessed: 14.02.2021).
33. Daisy Fancout. "Significant rise of 'no confidence' in government's handling of Covid-19".
In: *UCL News* (2020). URL: <https://www.ucl.ac.uk/news/2020/oct/significant-rise-no-confidence-governments-handling-covid-19>. (Last Accessed: 14.02.2021).

34. Sten Vermund and Virginia Pitzer. "Asymptomatic Transmission and the Infection Fatality Risk for COVID-19: Implications for School Reopening". In: *Oxford Academic* (2020).
URL: <https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciaa855/5862668>.
(Last Accessed: 21.11.2020).
35. Xi He, Eric Lau, et al. "Temporal dynamics in viral shedding and transmissibility of COVID-19". In: *Nature Medicine* (2020). URL: <https://www.nature.com/articles/s41591-020-0869-5>.
(Last Accessed: 21.11.2020).
36. Sarah Bunn. "Mass testing for COVID-19 using lateral flow tests". In: *The Parliamentary Office of Science and Technology* (2020).
URL: <https://post.parliament.uk/mass-testing-for-covid-19-using-lateral-flow-tests/>.
(Last Accessed: 11.02.2021).
37. Mike Gill and Muir Gray. "Mass testing for covid-19 in the UK". In: *The British Medical Journal* (2020).
URL: <https://www.bmj.com/content/371/bmj.m4436>. (Last Accessed; 14.02.2021).
38. Jonathan Deeks and Angela Raffle. "Lateral flow tests cannot rule out SARS-CoV-2 infection". In: *The British Medical Journal* (2020). URL: <https://www.bmj.com/content/371/bmj.m4787>.
(Last Accessed: 14.02.2021).
39. COVID-19 Genomics UK Consortium. *Welcome from the Director, Professor Sharon Peacock*.
URL: <https://www.cogconsortium.uk/about/>. (Last Accessed: 14.02.2021).
40. Public Health England. *PHE investigating a novel variant of COVID-19*. URL:
<https://www.gov.uk/government/news/phe-investigating-a-novel-variant-of-covid-19>.
(Last Accessed: 29.01.2021).
41. Maya Wei-Haas. "Why new coronavirus variants 'suddenly arose' in the U.K. and South Africa". In: *National Geographic* (2020).
URL: <https://www.nationalgeographic.com/science/2020/12/why-new-coronavirus-variants-suddenly-arose-in-uk-and-south-africa/>. (Last Accessed: 14.02.2021).
42. Elisabeth Mahase. "Covid-19: What have we learnt about the new variant in the UK?". In: *The British Medical Journal* (2020). URL: <https://www.bmj.com/content/371/bmj.m4944>.
(Last Accessed: 16.02.2021).
43. Sarah Boseley. "Less than 20% of people in England self-isolate fully, Sage says". In: *The Guardian* (2020). URL: <https://www.theguardian.com/world/2020/sep/11/less-than-20-of-people-in-england-self-isolate-fully-sage-says>. (Last Accessed: 17.12.2020).
44. Nick Triggle, Rachel Schraer, and Phil Kemp. "Coronavirus: Inside test-and-trace - how the 'world beater' went wrong". In: *BBC News* (2020).
URL: <https://www.bbc.co.uk/news/health-55008133>. (Last Accessed: 21.11.2020).
45. Department of Health and Social Care. *UK Chief Medical Officers' statement on the self-isolation period: 11 December 2020*.
URL: <https://www.gov.uk/government/news/uk-chief-medical-officers-statement-on-the-self-isolation-period-11-december-2020>. (Last Accessed: 14.02.2021).
46. Alexandra Topping. "Scientists in Liverpool mass Covid testing trial defend rapid tests". In: *The Guardian* (2020). URL: <https://www.theguardian.com/world/2020/dec/23/scientists-in-liverpool-mass-covid-testing-trial-defend-rapid-tests>. (Last Accessed: 14.02.2021).
47. Tara Parker-Pope and Katherine Wu. "What You Need to Know About Getting Tested for Coronavirus". In: *The New York Times* (2020).
URL: <https://www.nytimes.com/2020/12/09/well/live/covid-coronavirus-testing.html>.
(Last Accessed: 14.02.2021).

48. UK Government. *Getting tested for coronavirus (COVID-19)*. URL: <https://www.gov.uk/getting-tested-for-coronavirus>. (Last Accessed: 20.11.2020).
49. NHS England. *Thousands of former NHS staff are back on the front line in the NHS fight against coronavirus*. URL: <https://www.england.nhs.uk/2020/04/thousands-of-former-nhs-staff-are-back-on-the-front-line-in-the-nhs-fight-against-coronavirus/>. (Last Accessed: 21.11.2020).
50. British Medical Association. *BMA survey reveals almost half of doctors have relied upon donated or self-bought PPE and two thirds still don't feel fully protected*. URL: <https://www.bma.org.uk/bma-media-centre/bma-survey-reveals-almost-half-of-doctors-have-relied-upon-donated-or-self-bought-ppe-and-two-thirds-still-don-t-feel-fully-protected>. (Last Accessed: 21.11.2020).
51. Stephen Rocks and Omar Idriss. "Did hospital capacity affect mortality during the pandemic's first wave?" In: *The Health Foundation* (2020). URL: <https://www.health.org.uk/news-and-comment/charts-and-infographics/did-hospital-capacity-affect-mortality-during-the-pandemic>. (Last Accessed: 21.11.2020).
52. The World Health Organisation. *Solidarity" clinical trial for COVID-19 treatments*. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments>. (Last Accessed: 14.02.2021).
53. UK Research and Innovation. *How is coronavirus disease treated in hospital?* URL: <https://coronavirusexplained.ukri.org/en/article/vdt0008/>. (Last Accessed: 14.02.2021).
54. Sky News. *Coronavirus: How the UK's 14-day travel quarantine will work*. URL: <https://news.sky.com/story/coronavirus-how-the-uks-14-day-travel-quarantine-will-work-11992551>. (Last Accessed: 14.02.2021).
55. Department for Transport. *Coronavirus (COVID-19): Travel Corridors*. URL: <https://www.gov.uk/guidance/coronavirus-covid-19-travel-corridors>. (Last Accessed: 12.11.2020).
56. Public Health England. *Coronavirus Dashboard*. URL: <https://coronavirus.data.gov.uk/>. (Last Accessed: 06.10.2020).
57. Public Health England. *Coronavirus Dashboard: Cases - Cases by specimen date: Data*. URL: https://coronavirus.data.gov.uk/details/cases#card-cases_by_specimen_date. (Last Accessed: 19.03.2021).
58. Public Health England. *Coronavirus Dashboard: Deaths - Daily deaths with COVID-19 on the death certificate by date of death: Data*. URL: https://coronavirus.data.gov.uk/details/deaths#card-daily_deaths_with_covid-19_on_the_death_certificate_by_date_of_death. (Last Accessed: 19.03.2021).
59. Office For National Statistics. *Population Estimates*. URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/>. (Last Accessed: 25.08.2020).
60. Harvard Medical School. "If you've been exposed to the coronavirus". In: *Harvard Health Publishing* (2020). URL: <https://www.health.harvard.edu/diseases-and-conditions/if-youve-been-exposed-to-the-coronavirus>. (Last Accessed: 29.11.2020).
61. Public Health England. *Coronavirus Dashboard: About the data - Daily and cumulative deaths within 28 days of a positive test*. URL: <https://coronavirus.data.gov.uk/details/about-data#daily-and-cumulative-deaths-within-28-days-of-a-positive-test>. (Last Accessed: 06.02.2021).

62. Nisreen A. Alwan, Rochelle Ann Burgess, et al.
“Scientific consensus on the COVID-19 pandemic: we need to act now”. In: *The Lancet* (2020).
URL: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(20\)32153-X/fulltext#bib4](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)32153-X/fulltext#bib4). (Last Accessed: 06.02.2021).
63. Public Health England. *Coronavirus Dashboard: Testing - Virus tests conducted: Data*.
URL: https://coronavirus.data.gov.uk/details/testing#card-virus_tests_conducted.
(Last Accessed: 19.03.2021).
64. Public Health England.
Coronavirus Dashboard: Testing - Lab-based testing and capacity, by test type: Data.
URL: https://coronavirus.data.gov.uk/details/testing#card-lab-based_testing_and_capacity_by_test_type. (Last Accessed: 19.03.2021).
65. Public Health England. *Coronavirus Dashboard: Testing - Virus tests conducted*.
URL: <https://coronavirus.data.gov.uk/details/testing>. (Last Accessed: 11.02.2021).
66. Public Health England.
Coronavirus Dashboard: Testing - Lab-based testing and capacity, by test type.
URL: <https://coronavirus.data.gov.uk/details/testing>. (Last Accessed: 11.02.2021).
67. R documentation. *summary function*.
URL: <https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/summary>.
(Last Accessed: 04.12.2020).
68. Mark Gardener. *Summarising Data*. URL: <https://www.dataanalytics.org.uk/data-analytics-knowledge-base-tips-tricks-r-excel/statistics-guide/data-summary/>.
(Last Accessed: 04.12.2020).
69. R documentation. *ggplot function*.
URL: <https://www.rdocumentation.org/packages/ggplot2/versions/3.3.2/topics/ggplot>.
(Last Accessed: 19.11.2020).
70. R documentation. *ggplot2 package*.
URL: <https://www.rdocumentation.org/packages/ggplot2/versions/3.3.2>.
(Last Accessed: 19.11.2020).
71. R documentation. *plot function*.
URL: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/plot>.
(Last Accessed: 20.11.2020).
72. R documentation. *graphics package*.
URL: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2>.
(Last Accessed: 27.02.2021).
73. R documentation. *uncount function*.
URL: <https://www.rdocumentation.org/packages/tidyft/versions/0.4.5/topics/uncount>.
(Last Accessed: 29.11.2020).
74. R documentation. *tidyr package*.
URL: <https://www.rdocumentation.org/packages/tidyr/versions/0.8.3>.
(Last Accessed: 29.11.2020).
75. R documentation. *incidence function*. URL:
<https://www.rdocumentation.org/packages/incidence/versions/1.7.3/topics/incidence>.
(Last Accessed: 29.11.2020).
76. R documentation. *incidence package*.
URL: <https://www.rdocumentation.org/packages/incidence/versions/1.7.3>.
(Last Accessed: 29.11.2020).

77. R documentation. *ts function*.
URL: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/ts>.
(Last Accessed: 23.12.2020).
78. R documentation. *stats package*.
URL: <https://www.rdocumentation.org/packages/stats/versions/3.6.2>.
(Last Accessed: 23.12.2020).
79. R documentation. *ggarrange function*.
URL: <https://www.rdocumentation.org/packages/ggpubr/versions/0.4.0/topics/ggarrange>.
(Last Accessed: 04.12.2020).
80. R documentation. *ggpubr package*. URL: <https://www.rdocumentation.org/packages/ggpubr>.
(Last Accessed: 23.12.2020).
81. R documentation. *SMA function*.
URL: <https://www.rdocumentation.org/packages/TTR/versions/0.24.2/topics/SMA>.
(Last Accessed: 12.11.2020).
82. R documentation. *as.xts function*.
URL: <https://www.rdocumentation.org/packages/xts/versions/0.12.1/topics/as.xts>.
(Last Accessed: 22.03.2021).
83. R documentation. *xts package*.
URL: <https://www.rdocumentation.org/packages/xts/versions/0.12.1>.
(Last Accessed: 22.03.2021).
84. R documentation. *TTR package*.
URL: <https://www.rdocumentation.org/packages/TTR/versions/0.24.2>.
(Last Accessed: 19.11.2020).
85. Stephanie Glen. "Stochastic Model/Process: Definition and Examples". In: *Statistics How To* (2016).
URL: <https://www.statisticshowto.com/stochastic-model/>. (Last Accessed: 22.04.2021).
86. Herbert W. Hethcote. "Three Basic Epidemiological Models". In: *Applied Mathematical Ecology*. Springer: Berlin Heidelberg, 1989, pp. 119–144. ISBN: 978-3-642-61317-3.
DOI: [10.1007/978-3-642-61317-3_5](https://doi.org/10.1007/978-3-642-61317-3_5). URL: https://doi.org/10.1007/978-3-642-61317-3_5.
(Last Accessed: 29.11.2020).
87. Kai Sasaki. *COVID-19 dynamics with SIR model*.
URL: <https://www.lewuathe.com/covid-19-dynamics-with-sir-model.html>.
88. R documentation. *modFit function*.
URL: <https://www.rdocumentation.org/packages/FME/versions/1.3.6.1/topics/modFit>.
(Last Accessed: 22.12.2020).
89. R documentation. *FME package*.
URL: <https://www.rdocumentation.org/packages/FME/versions/1.3.6.1>.
(Last Accessed: 22.12.2020).
90. R documentation. *ode function*.
URL: <https://www.rdocumentation.org/packages/deSolve/versions/1.27.1/topics/ode>.
(Last Accessed: 19.11.2020).
91. R documentation. *deSolve package*.
URL: <https://www.rdocumentation.org/packages/deSolve/versions/1.28>.
(Last Accessed: 19.11.2020).
92. Stephanie Glen. *Sum of Squares: Residual Sum, Total Sum, Explained Sum, Within*. 2014.
URL: <https://www.statisticshowto.com/residual-sum-squares/>. (Last Accessed: 18.02.2021).

93. Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics - Supplementary Material". In: *American Journal of Epidemiology* (2013). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3816335/>. (Last Accessed: 15.11.2020).
94. Tim Churches. *An exploration of COVID-19 incidence data using R tools and packages*. URL: <https://timchurches.github.io/blog/posts/2020-02-18-analysing-covid-19-2019-ncov-outbreak-data-with-r-part-1/>. (Last Accessed: 01.11.2020).
95. R documentation. *estimate_R function*. URL: https://www.rdocumentation.org/packages/EpiEstim/versions/2.2-3/topics/estimate_R. (Last Accessed: 19.11.2020).
96. R documentation. *EpiEstim package*. URL: <https://www.rdocumentation.org/packages/EpiEstim/versions/2.2-3>. (Last Accessed: 19.11.2020).
97. Farhad Malik. "Understanding Auto Regressive Moving Average Model — ARIMA". In: *Medium* (2018). URL: <https://medium.com/fintechexplained/understanding-auto-regressive-model-arima-4bd463b7a1bb#:~:text=ARIMA%20stands%20for%20Auto%20Regressive,and%20observations%20of%20lagged%20terms>. (Last Accessed: 16.02.2021).
98. Selva Prabhakaran. "Vector Autoregression (VAR) – Comprehensive Guide with Examples in Python". In: *Machine Learning Plus* (2019). URL: <https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/>. (Last Accessed: 16.02.2021).
99. Rob J. Hyndman and George Athanasopoulos. "Forecasting: principles and practice, 2nd edition". In: OTexts: Melbourne, Australia, 2018. Chap. 8.9 Seasonal ARIMA Models. URL: <https://otexts.com/fpp2/seasonal-arima.html>. (Last Accessed: 21.02.2021).
100. Rob J. Hyndman and George Athanasopoulos. "Forecasting: principles and practice, 2nd edition". In: OTexts: Melbourne, Australia, 2018. Chap. 8.2 Backshift Notation. URL: <https://otexts.com/fpp2/backshift.html>. (Last Accessed: 28.02.2021).
101. Rob J. Hyndman and George Athanasopoulos. "Forecasting: principles and practice, 2nd edition". In: OTexts: Melbourne, Australia, 2018. Chap. 8.1 Stationarity and Differencing. URL: <https://otexts.com/fpp2/stationarity.html>. (Last Accessed: 28.02.2021).
102. R documentation. *auto.arima function*. URL: <https://www.rdocumentation.org/packages/forecast/versions/8.13/topics/auto.arima>. (Last Accessed: 17.02.2021).
103. R documentation. *forecast package*. URL: <https://www.rdocumentation.org/packages/forecast/versions/8.13>. (Last Accessed: 17.02.2021).
104. Stephanie Glen. *Akaike's Information Criterion: Definition, Formulas*. 2015. URL: <https://www.statisticshowto.com/akaikes-information-criterion/>. (Last Accessed: 18.02.2021).
105. Stephanie Glen. *Bayesian Information Criterion (BIC) / Schwarz Criterion*. 2018. URL: <https://www.statisticshowto.com/bayesian-information-criterion/>. (Last Accessed: 18.02.2021).
106. R documentation. *tsdiag function*. URL: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/tsdiag>. (Last Accessed: 21.02.2021).

107. R documentation. *forecast function*.
URL: <https://www.rdocumentation.org/packages/forecast/versions/8.13/topics/forecast>.
(Last Accessed: 18.02.2021).
108. R documentation. *forecast.fracdiff function*. URL: <https://www.rdocumentation.org/packages/forecast/versions/8.13/topics/forecast.fracdiff>.
(Last Accessed: 18.02.2021).
109. Alessandro Cardinali. "MATH3623: Financial Statistics". In: University of Plymouth, 2020. Chap. 9 Cointegration, pp. 152–155. (Last Accessed: 02.03.2021).
110. Rob J. Hyndman and George Athanasopoulos. "Forecasting: principles and practice, 2nd edition". In: OTexts: Melbourne, Australia, 2018. Chap. 11.2 Vector Autoregressions.
URL: <https://otexts.com/fpp2/stationarity.html>. (Last Accessed: 28.02.2021).
111. Stephanie Glen. *Cointegration: Definition, Examples, Tests*. 2016.
URL: <https://www.statisticshowto.com/cointegration/>. (Last Accessed: 18.02.2021).
112. R documentation. *po.test function*.
URL: <https://www.rdocumentation.org/packages/tseries/versions/0.10-48/topics/po.test>.
(Last Accessed: 18.02.2021).
113. R documentation. *tseries package*.
URL: <https://www.rdocumentation.org/packages/tseries/versions/0.10-47>.
(Last Accessed: 18.02.2021).
114. R documentation. *VARselect function*.
URL: <https://www.rdocumentation.org/packages/vars/versions/1.5-3/topics/VARselect>.
(Last Accessed: 18.02.2021).
115. R documentation. *vars package*.
URL: <https://www.rdocumentation.org/packages/vars/versions/1.5-3>.
(Last Accessed: 18.02.2021).
116. Herman J. Bierens. "Information Criteria and Model Selection".
In: *Pennsylvania State University* (2004), p. 3.
URL: https://www.researchgate.net/profile/Herman-Bierens/publication/267374496_Information_Criteria_and_Model_Selection/links/54b043ec0cf220c63cceb5fb/Information-Criteria-and-Model-Selection.pdf. (Last Accessed: 18.02.2021).
117. R documentation. *ca.jo function*.
URL: <https://www.rdocumentation.org/packages/urca/versions/1.2-9/topics/ca.jo>.
(Last Accessed: 18.02.2021).
118. R documentation. *urca package*.
URL: <https://www.rdocumentation.org/packages/urca/versions/1.2-9>.
(Last Accessed: 18.02.2021).
119. Gerald P. Dwyer. "The Johansen Tests for Cointegration". In: *Clemson University* (2015), p. 5.
URL: <http://www.jerrydwyer.com/pdf/Clemson/Cointegration.pdf>. (Last Accessed: 18.02.2021).
120. R documentation. *predict function*.
URL: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/predict>.
(Last Accessed: 18.02.2021).
121. R documentation. *predict.VAR function*.
URL: <https://www.rdocumentation.org/packages/tsDyn/versions/0.9-44/topics/predict.VAR>.
(Last Accessed: 18.02.2021).

122. R documentation. *lines function*.
URL: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/lines>.
(Last Accessed: 21.02.2021).
123. R documentation. *abline function*.
URL: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/abline>.
(Last Accessed: 21.02.2021).
124. R documentation. *estimate_R_plots function*. URL: https://www.rdocumentation.org/packages/EpiEstim/versions/2.2-4/topics/estimate_R_plots.
(Last Accessed: 13.02.2021).
125. Boris Johnson. *Prime Minister announces national lockdown*.
URL: <https://www.gov.uk/government/news/prime-minister-announces-national-lockdown>.
(Last Accessed: 23.03.2021).
126. R documentation. *polygon function*.
URL: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/polygon>.
(Last Accessed: 21.02.2021).
127. R documentation. *autoplot function*.
URL: <https://www.rdocumentation.org/packages/ggplot2/versions/3.3.3/topics/autoplot>.
(Last Accessed: 21.02.2021).
128. R documentation. *ggfortify package*.
URL: <https://www.rdocumentation.org/packages/ggfortify/versions/0.4.11>.
(Last Accessed: 22.03.2021).
129. R documentation. *geom_line function*.
URL: https://www.rdocumentation.org/packages/ggplot2/versions/0.9.0/topics/geom_line.
(Last Accessed: 22.03.2021).
130. R documentation. *par function*.
URL: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/par>.
(Last Accessed: 21.02.2021).
131. Rob J. Hyndman and George Athanasopoulos. "Forecasting: principles and practice, 2nd edition". In: *Forecast Errors*. OTexts: Melbourne, Australia, 2018. Chap. 3.4 Evaluating Forecast Accuracy.
URL: <https://otexts.com/fpp2/accuracy.html>. (Last Accessed: 26.03.2021).
132. Max Roser, Hannah Ritchie, et al. "Coronavirus Pandemic (COVID-19)". In: *Our World in Data* (2020).
URL: <https://ourworldindata.org/coronavirus>. (Last Accessed: 24.02.2021).
133. Our World In Data. *COVID-19 data - xlsx*. URL: <https://github.com/owid/covid-19-data/raw/master/public/data/owid-covid-data.xlsx>.
(Last Accessed: 24.02.2021).
134. Anna Jones. "How did New Zealand become Covid-19 free?" In: *BBC News* (2020).
URL: <https://www.bbc.co.uk/news/world-asia-53274085>. (Last Accessed: 24.02.2021).
135. Sharon Marris.
"Coronavirus: New Zealand brings in 'world's toughest border restrictions' to fight outbreak".
In: *Sky News* (2020). URL: <https://news.sky.com/story/coronavirus-new-zealand-brings-in-worlds-toughest-border-restrictions-to-fight-outbreak-11957379>.
(Last Accessed: 24.02.2021).
136. Eleanor Ainge Roy. "New Zealand and Australia close borders to foreigners amid coronavirus crisis".
In: *The Guardian* (2020). URL: <https://www.theguardian.com/world/2020/mar/19/new-zealand-closes-borders-to-foreigners-amid-coronavirus-crisis>. (Last Accessed: 24.02.2021).

137. New Zealand Government. *COVID-19: Source of cases*.
URL: <https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-data-and-statistics/covid-19-source-cases>.
(Last Accessed: 03.03.2021).
138. New Zealand Government. *History of the COVID-19 Alert System*.
URL: <https://covid19.govt.nz/alert-system/history-of-the-covid-19-alert-system/>.
(Last Accessed: 24.02.2021).
139. New Zealand Government. *COVID-19: Current Cases - Case Details*.
URL: <https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-data-and-statistics/covid-19-current-cases>.
(Last Accessed: 27.02.2021).
140. BBC News. *In pictures: How coronavirus swept through Brazil*.
URL: <https://www.bbc.co.uk/news/world-latin-america-53429430>.
(Last Accessed: 25.02.2021).
141. Editorial. "The Guardian view on Covid-19 and cults of strength: the weakest response".
In: *The Guardian* (2020). URL: <https://www.theguardian.com/commentisfree/2020/may/28/the-guardian-view-on-covid-19-and-cults-of-strength-the-weakest-response>.
(Last Accessed: 25.02.2021).
142. BBC News. *Coronavirus: Brazil's President Bolsonaro tests positive*.
URL: <https://www.bbc.co.uk/news/world-latin-america-53319517>.
(Last Accessed: 25.02.2021).
143. Terrence McCoy and Heloísa Traiano. "Limits on coronavirus testing in Brazil are hiding the true dimensions of Latin America's largest outbreak". In: *The Washington Post* (2020). URL: https://www.washingtonpost.com/world/the_americas/coronavirus-brazil-testing-bolsonaro-cemetery-gravedigger/2020/04/22/fe757ee4-83cc-11ea-878a-86477a724bdb_story.html.
(Last Accessed: 26.02.2021).
144. Pedro Fonseca.
"Uncounted infections might help explain recent fall in Brazil COVID-19 cases, experts say".
In: *Reuters* (2020).
URL: <https://www.reuters.com/article/uk-health-coronavirus-brazil-idUKKBN27P2HV>.
(Last Accessed: 27.02.2021).
145. Ana Mano. "Brazil uses less than a third of available coronavirus tests, newspaper says".
In: *Reuters* (2020).
URL: <https://www.reuters.com/article/us-health-coronavirus-brazil-idUSKBN25V10Y>.
(Last Accessed: 25.02.2021).
146. Gabriela Lessa. "The challenge of COVID-19 for business travel and mobility: Brazil and beyond".
In: *Ius Laboris* (2020). URL: <https://iuslaboris.com/insights/the-challenge-of-covid-19-for-business-travel-and-mobility-brazil-and-beyond/>. (Last Accessed: 26.02.2021).
147. BBC News. *Coronavirus: Brazil's Bolsonaro joins anti-lockdown protests*.
URL: <https://www.bbc.co.uk/news/world-latin-america-52351636>.
(Last Accessed: 26.02.2021).
148. Terrence McCoy. "While other countries look to open up, Brazil can't find a way to shut down".
In: *The Washington Post* (2020). URL: https://www.washingtonpost.com/world/while-other-countries-look-to-open-up-brazil-cant-find-a-way-to-shut-down/2020/05/10/2c02abea-92c1-11ea-87a3-22d324235636_story.html. (Last Accessed: 26.02.2021).

149. Charlotte Peet. "COVID cases skyrocket in Brazil with no vaccination start date". In: *Aljazeera* (2020).
URL: <https://www.aljazeera.com/news/2020/12/15/covid-cases-skyrocket-in-brazil-with-no-vaccine-start-date>. (Last Accessed: 27.02.2021).
150. Antonello Maruotti, Massimo Ciccozzi, and Fabio Divino.
"On the misuse of the reproduction number in the COVID-19 surveillance system in Italy".
In: *Journal of Medical Virology* (2021).
URL: https://onlinelibrary.wiley.com/doi/10.1002/jmv.26881?fbclid=IwAR11ZfV1Q_YNEt4-zkv5IZoc0ffPPpA86mxKUCvdCCPdEtKirQ1IQCrM2GU. (Last Accessed: 15.03.2021).