MonteCarlo Simulation's of the Covid-19 Pandemic

Austin Lackey, Chien Lin and Jin Peng

Tuesday, December 6th, 2022

Contents

1 Introduction (Should we do an Abstract?)

1.1 Background

In our MonteCarlo simulation analysis, we fit a gauss-error function over cumulaative Covid-19 data. In the study mentioned below, the authors stated that it is widely accepted that China's Covid-19 data follows a gauss-error function. We decided to test this hypothesis by fitting a gauss-error function over the data of China, Italy, and the United States using data from the World Health Organization (WHO).

We then used the fitted function to predict the time evolution of the Covid-19 Pandemic in each country. For context to our project, time evolution is the rate of change of the number of cases over time. By fitting a gauss-error function to partial data, we are able to prefict the day of flex point (which is the day of the highest cases/fatality occurrence) for that given country. Our primary motivation for this project is to see if the gauss-error function is a good fit for various different countries and use MonteCarlo simulation to predict for future trends where the data are “unknown”.

The results of this study are based on mathematical and statistic approaches alone and do not take into account any of the influential factors such as, stated in the study: number of daily nasopharyngeal swabs, medical, social distancing, virological and epidemiological or models of contamination diffusion. Consequently, the accuracy of this study in terms of real world application, may not be too useful.

1.2 Ciufolini and Paolozzi's MonteCarlo Study

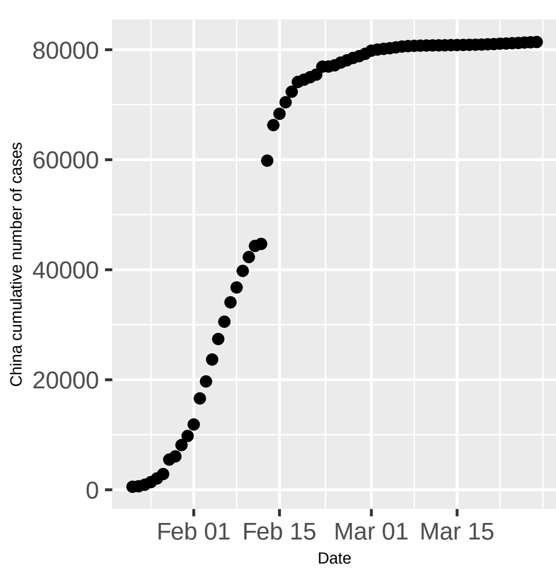
Our project is an extension and replication of the study done by Ciufolini and Paolozzi. *“Mathematical prediction of the time evolution of the COVID-19 pandemic in Italy by a Gauss error function and Monte Carlo simulations”*. The authors of this study used a MonteCarlo simulation to fit a gauss-error function over the data of China and extending it to Italy. In their study, it was widely accepted that China's Covid-19 data follows a gauss-error function.

$$Gauss - ErrorFunction = a + b * erf(c * x - d)$$

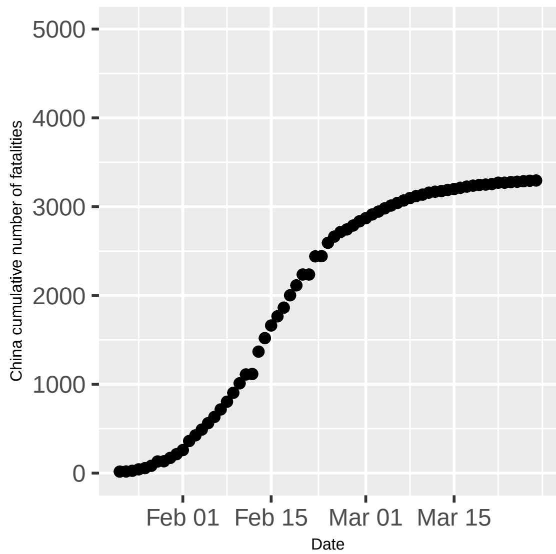
The authors perfomed a MonteCarlo simulation with 150 replications to approximate the day of flex point for China and Italy. The use of the MonteCarlo simulation was crucial to their study's success as reported cases and fatalities can differ up `to 10%` from the actual number of cases and fatalities.

2 Jin: China

2.1 Available Official Data (Positive Cases)



2.2 Available Official Data (Fatalities)

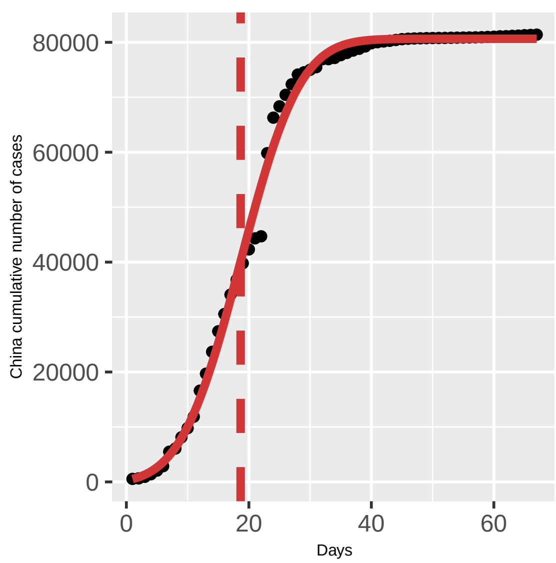


2.3 Fit by the Gauss Error Function

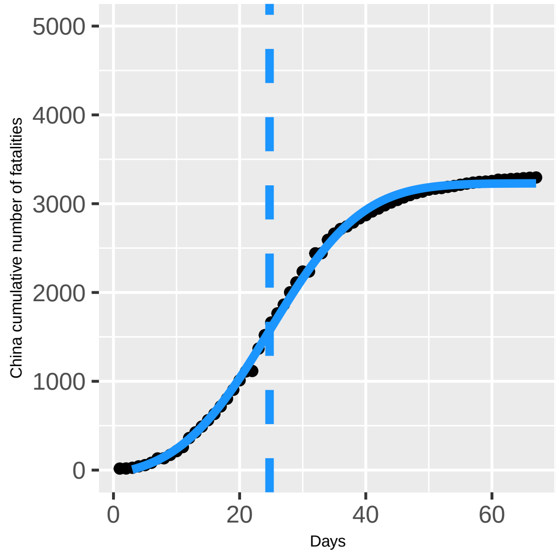
Apply an *nls* function to determine the nonlinear (weighted) least-squares estimates of the four parameters a, b, c, d of a Gauss error function model: $a + b * erf(c * x - d)$.

```
model <- nls(total_cases ~ gauss_fn(days,a,b,c,d),
             data=china,
             start=list(a=10000,b=10000,c=0.1,d=1))
```

2.4 Fit by the Gauss Error Function and Peak Day (Positive Cases)



2.5 Fit by the Gauss Error Function and Peak Day (Fatalities)



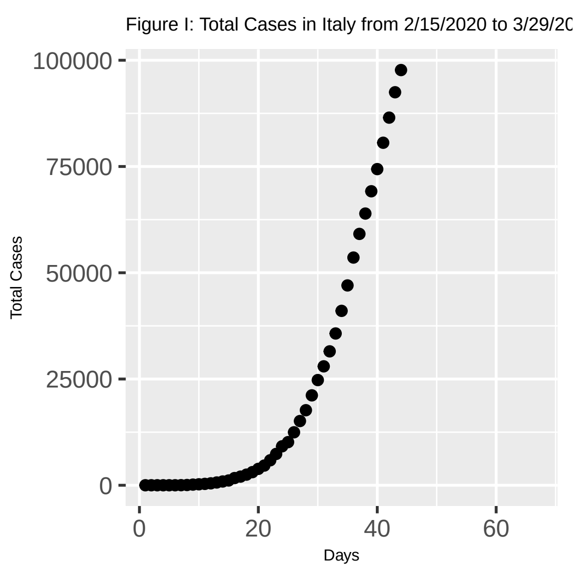
3 Italy

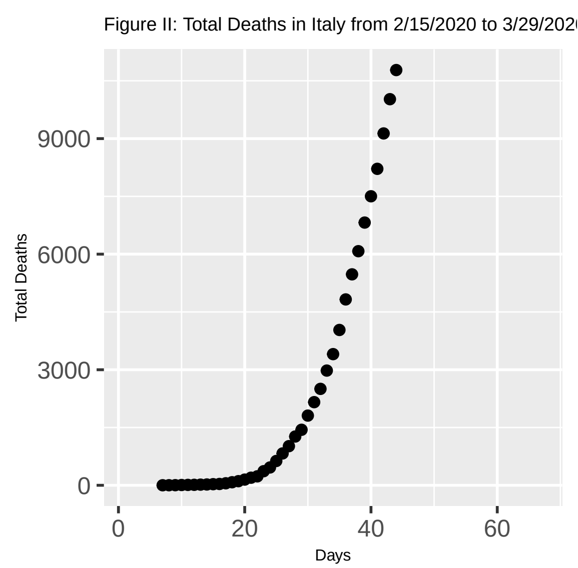
3.1 Preview of the Data

After the Gauss Error function is fitted to the China data, and verified that the function fits well with both the cases data and the fatality data, the study then proceeds to fit the function to the data from Italy.

In this part, we will fit the Gauss Error function to the Italy data and verify if the function is well fitted. For MonteCarlo simulation purposes we will only use Italy data from date 2/15/2020 to 3/29/2020, even though data for future dates are available.

First, we will take a look at the data:

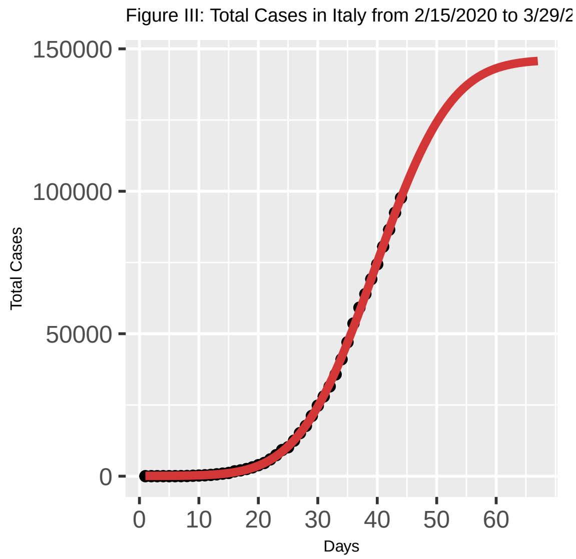


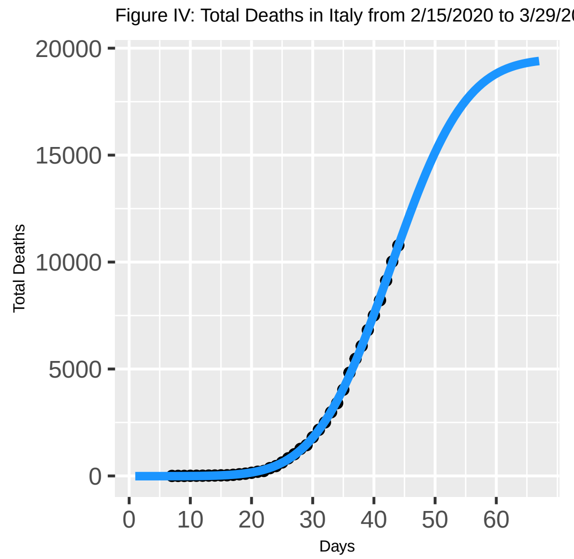


As you can see, the data points already follows a pattern that looks like the cumulative distribution of a Gauss Error function.

3.2 Fitting Gauss Error Function to Italy data

We will verify with the function fitted:





3.3 Monte Carlo Simulation

The Monte Carlo method we will simulate will be as follows:

- Create a random matrix : m X n
  - m = number of random outcomes (*chosen to be 150*)
  - n = number of observed days (n = *1,. . . ,j*)
- Each # in the matrix is a Gaussian distribution with mean = 1 and sigma = 0.1
  - Multiply each column (j) by the # of total cases (or deaths) that corresponds to each day j
- Each of the 150 simulations were fitted with the Gauss error function
- Then determined the date of the flex for each simulation

Basically, we want to simulate 150 of the function drawn according to the data (which means for different data, the function fitted will have different parameters) with a noise of a Normal random distribution with mean 1 and standard deviation 1.

3.4 MonteCarlo Simulation for Predicting Day of Flex for Italy

Then we run the simulation to get the mean day of the highest case/fatality:

3.4.1 MonteCarlo Simulation for Cumulative Positive Cases of COVID-19 in Italy

Figure V: Scatter Plot of MC Cases Day of Flex (Italy)

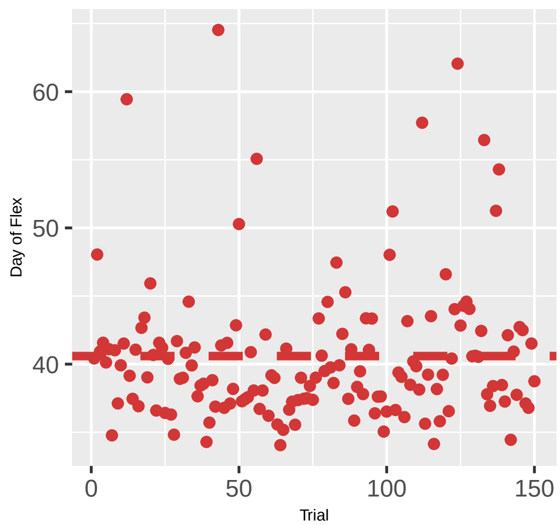
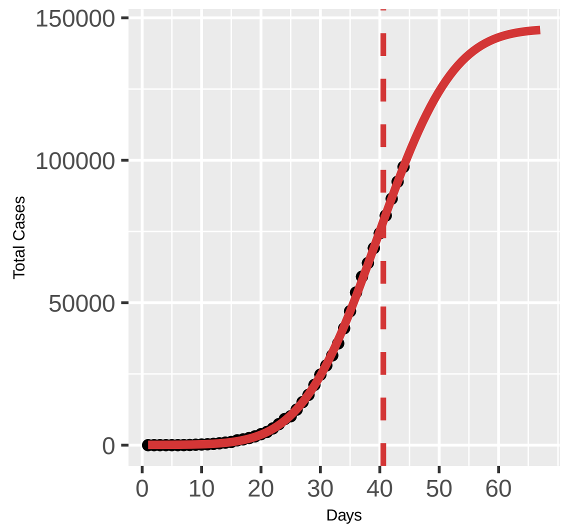


Figure VI: Total Cases in Italy from 2/15/2020 to 3/29/2



The mean of the day of flex for cases is 40.59 days and the standard deviation is 5.222832 days.

In figure V,each red dot corresponds to the day of occurrence of the flex obtained with each of the 150 Monte Carlo simulations.

3.4.2 MonteCarlo Simulation for Cumulative Fatalities of COVID-19 in Italy

Figure VII: Scatter Plot of MC Fatality Day of Flex (Italy)

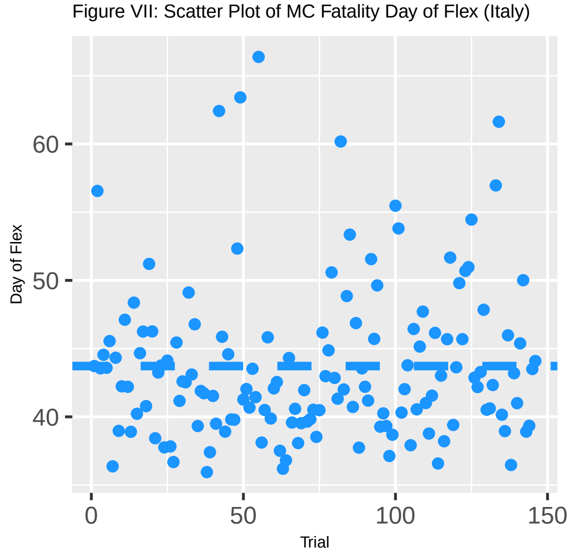
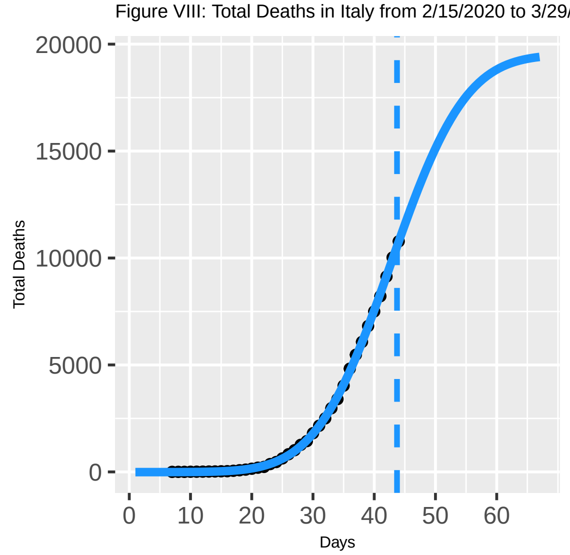


Figure VIII: Total Deaths in Italy from 2/15/2020 to 3/29/



The mean of the day of flex for fatalities is 43.46 days and the standard deviation is 5.228381 days. As you can see, the variance is high for both cases data and fatality data, an assumption might be because of smaller sample size.

4 Austin: Australia

4.1 Fitting a Gauss-Error Function to Australia's Data

We wanted to extend the findings from the study done by Ciufolini and Paolozzi to other countries. The first country that we decided to test was Australia. Since we have access to all of Australia's data at the time of writing this report, we decided to test the gauss-error function on a subset timeframe of the entire dataset. To begin, we limited our data to the timeframe of 6/2/2020 to 8/9/2020 which is 70 days. As you can see in Figure A and Figure B, the data begins to follow a gauss-error function shape. The days on the x-axis are the number of days following 6/2/2020.

Figure A: Total Cases in Australia from 6/2/2020 to 8/9/2

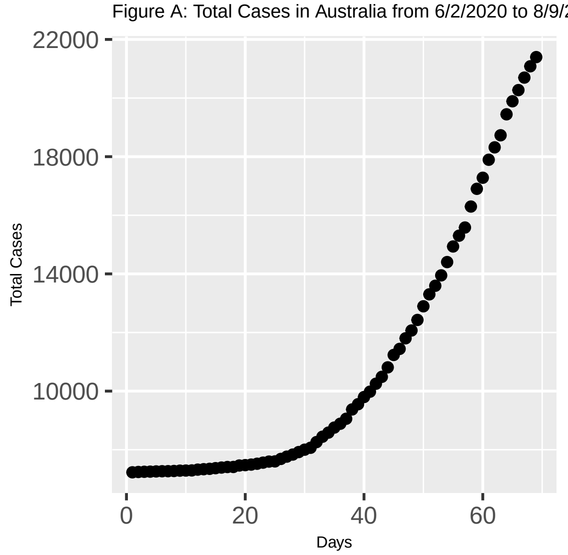
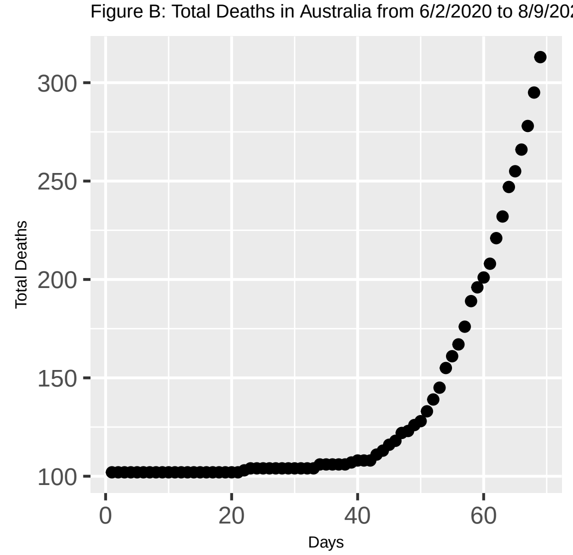


Figure B: Total Deaths in Australia from 6/2/2020 to 8/9/202



In order to fit a gauss-error function to the data, we used the `nlsLM` function from the `minpack.lm` library. This allows us to fit a gauss-error function to the data and find the parameters of the function. As you can see in Figures C and D, the fitted function follows the data very well. One important distinction to note is that the fatalities begin to follow the gauss-error function shape at a later time than the cases. This makes sense as death rates begin to increase after the number of cases begin to increase.

Figure C: Total Cases in Australia from 6/2/2020 to 8/9/2

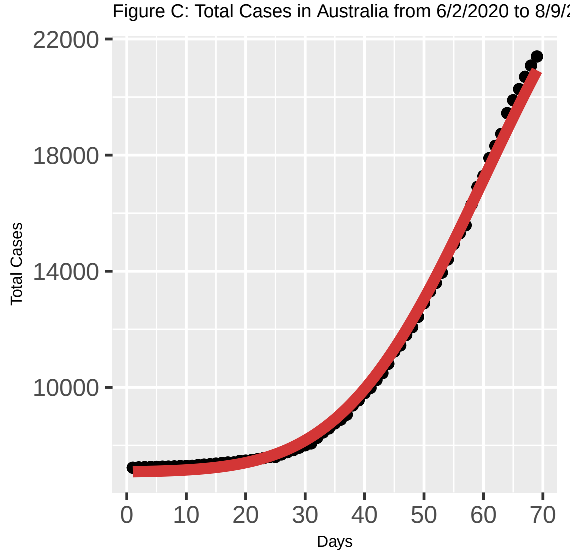
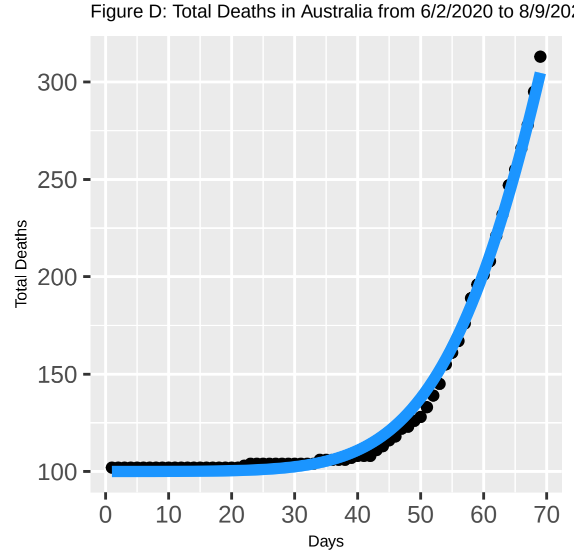


Figure D: Total Deaths in Australia from 6/2/2020 to 8/9/202



4.2 Extrapolating the Gauss-Error Function to Australia's Future Data

Once we have a fitted gauss-error function, we can extrapolate the function to the future. We decided to extrapolate the function to 200 days which is an increase from China's and Italy's extrapolation. The reasoning for this increase is that it takes longer for the number of cases to increase in Australia than it does in China and Italy. We are not quite sure why this is the case, but we believe it is due to the fact that Australia had different restrictions in place than China and Italy. As you can see in Figures E and F, the extrapolated function starts to show us an idea of what the future of Australia's data could look like if the current trend continues. The first thing we noticed was that the cases were getting close to their expected plateau and the fatalities were just beggining to increase.

Figure E: Extrapolating Future Case Data with Gauss Er

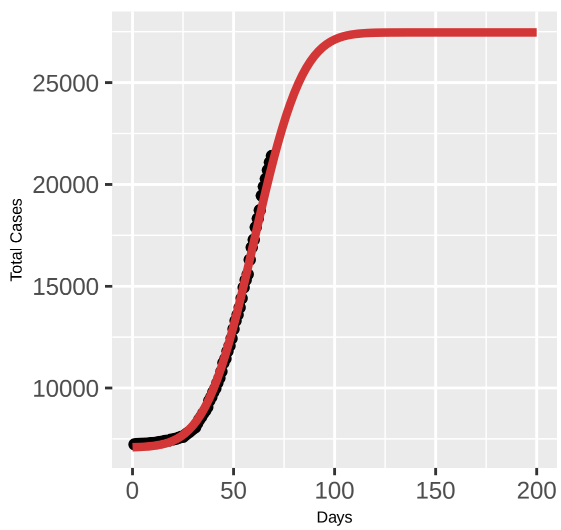
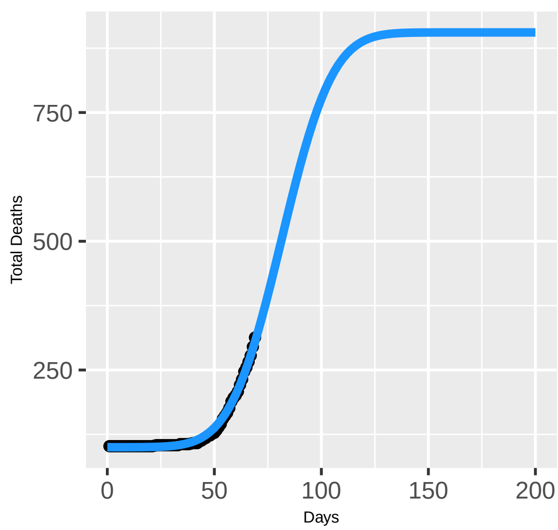


Figure F: Extrapolating Future Fatality Data with Gauss Err



4.3 MonteCarlo Simulation for Australia's Day of Flex

In order to account for the `10%` variance in the reported data, we need to run a MonteCarlo simulation. We decided to run a MonteCarlo simulation with a `150` iterations and a `0.1` standard deviation. The reason we chose a `150` iterations is because we wanted to have a large enough sample size to get a good idea of the variance in the data. This was also the same values that Ciufolini and Paolozzi used in their paper. As you can see in Figures G and H, the MonteCarlo simulation shows us that the variance in the data is not as large as it was for China and Italy. We believe this is due to the fact that since Australia required more days to reach their peak, the variance in the data is smaller since `n` is much larger. The day of flex for cases in Australia appears to reside around 60.22 days with a standard deviation of 0.77 days. The day of flex for fatalities in Australia appears to reside around 81.4 days with a standard deviation of 0.69 days.

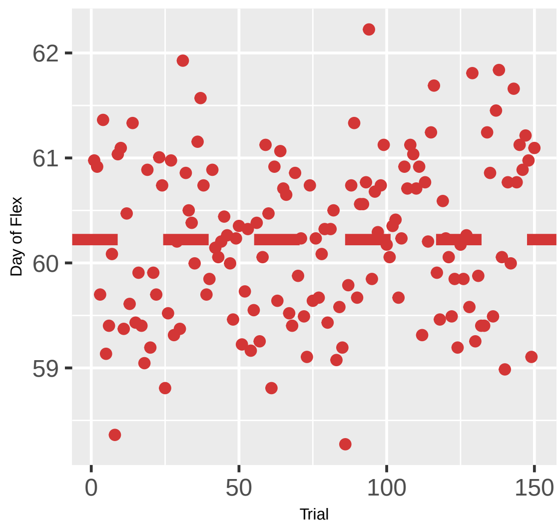


Figure G: MonteCarlo Simulation for Day of Flex (Cases)

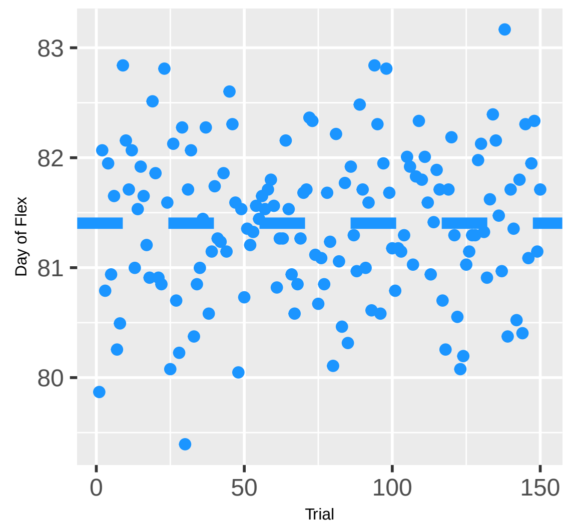


Figure G: MonteCarlo Simulation for Day of Flex (Fatalities)

4.4 Applying the day of flex indicator to the data

As you can see in Figures I and J, the day of flex indicator for cases and fatalities appears to sit right on the midpoint of the projected gauss-curve. This is because the day of flex indicator is the point at which the derivative of the gauss-curve is equal to 0. This is when the daily cases and fatalities are at their peak and begin to decline.

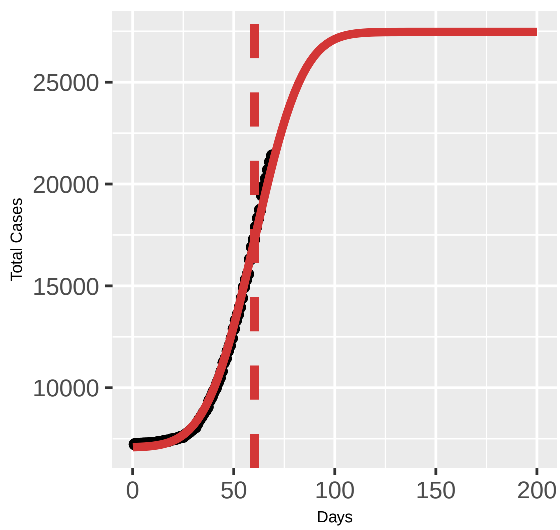


Figure I: Day of flex indicator for cases (60 days after)

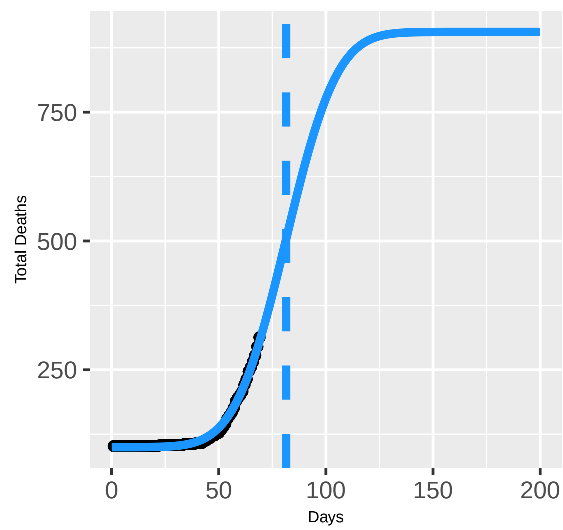
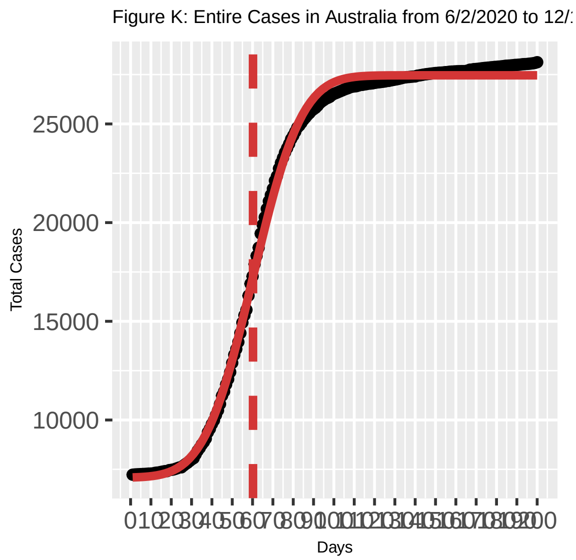
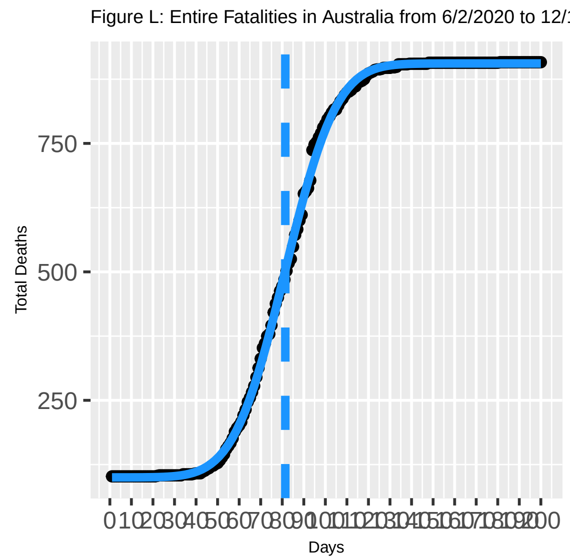


Figure J: Day of flex indicator for fatalities (81 days after)

4.5 Confirming Results with entire dataset

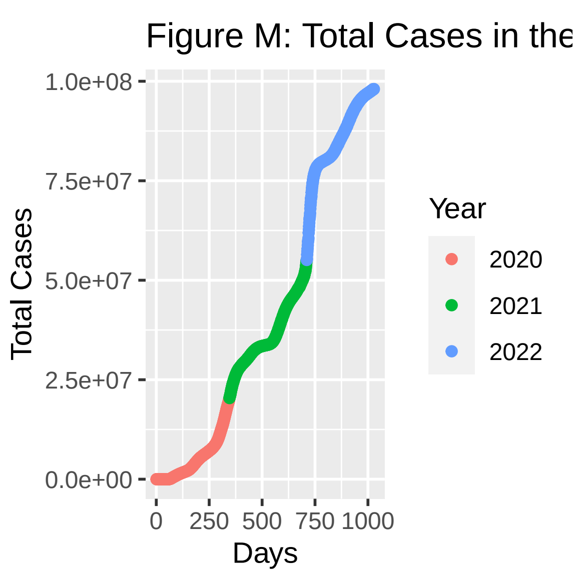
In order to confirm our results, we decided to overlay our findings with the entire dataset to see how well our model fits the actual data. As you can see in Figures K and L, our model did a decent job of fitting the actual data. It appears the model works the best for fatalities but the cases model starts to be a little innacurate once the data starts to plateau. Our theory for this inconsistency between fatalities and cases is that it is far easier to report fatalities than it is to report cases. It is not often that a person will die from the virus and not be tested or reported. However, it is very common for a person to have the virus and not be reported. This variance in the data once it begins to plateau could be due to the fact that testing procedures could have changed or the reporting procedures could have changed. Since the curve starts to overestimate and then starts to underestimate, Australia could have had a few weeks where they were not doing as much testing, and then they started doing more testing. This could explain the variance in the tail of the data. Regardless of this variance, our model did a good job at predicting and fitting the day of flex.

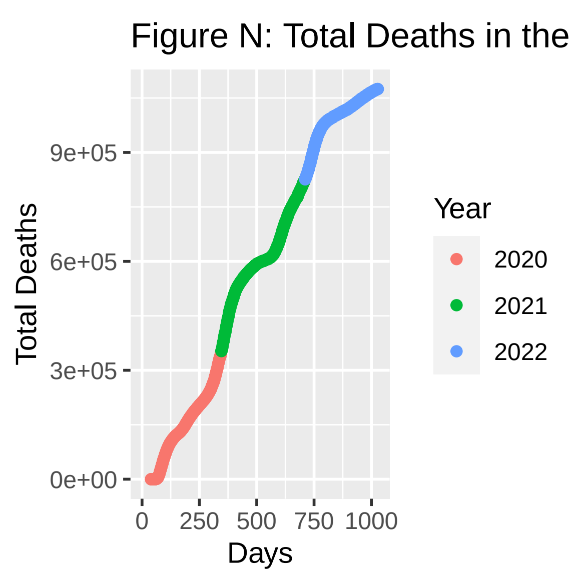




5 United States Anomaly

An interesting anomaly that we found in the data was that the United States had a hard time fitting the gauss-error funciton. As you can see in Figures M and N, the United States data does not ever begin to plateau. The data is always at an increasing rate. This makes it hard to predict when the day of flex will be. We believe that this is due to the fact that the United States had very lax restrictions when compared to other countries.





6 Simulation Analysis Summary

In our simulation analysis, we found that the day of flex indicator was a good predictor of when the day of flex would be for most countries. This model may not be the best model for countries who have different restrictions or have different testing procedures.

7 Potential Improvements

One idea that comes to mind in terms of improvements for this model is to see if we can account for the changes in testing procedures. This could be done by looking at the data and seeing when the testing procedures changed and then adjusting the model to account for this change. Although this would require advanced knowledge of the data as well as policy changes, but it could be a good way to improve the model. Another idea is to see if we could find other models that fit countries with laxer restrictions better. If they don't follow a gauss-error function, then we could try to explore other models that fit their data better.

8 Works Cited

- Link to the study:
  - I. Ciufolini, A. Paolozzi, Prediction of the time evolution of the Covid-19 Pandemic in Italy by a Gauss Error Function and Monte Carlo simulations. Submitted to BioRxiv on 03.26.2020 and transferred on 03.27.2020 to MedRxiv. https://doi.org/10.1101/2020.03.27.20045104
- Link to the covid data used:
  - https://covid19.who.int/
- Source to find the second derivative of fitted function:
  - https://proofwiki.org/wiki/Derivative_of_Error_Function