



GEORGIA INSTITUTE OF TECHNOLOGY  
SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING

# COVID-19 GLOBAL FORECAST

REGRESSION ANALYSIS — ISYE6414  
DECEMBER 11, 2022

Rahul Baid (782)rbaid3@gatech.edu  
Ogulcan Canbek (330) ogulcanbek@gatech.edu  
Omkar Kulkarni (583)okulkarni8@gatech.edu  
Benjamin Fuentes (531) bfuentes3@gatech.edu

## Abstract

The epidemic of COVID-19 has become a global threat and modeling forecasts provide important information for countries' policies and mitigation strategies. The aim of this study is to use different statistical models to predict the spreading of the disease over time, focusing in the number of confirmed cases and deaths for a given country. The research was based on daily confirmed cases of COVID-19 provided by Johns Hopkins University. A first statistical approach is given by using exponential and logistic growth models, where the exponential model describe the spreading in the short-term and the logistic has better prediction of the whole growth. Polynomial, Ridge and Lasso regressions are also proposed. The polynomial regression provide the best approximation of confirmed cases over time, being able to fit non-linear data closely, but with a high sensitivity to outliers. Between the two regularization models, ridge has a better performance on prediction of cases of corona virus. Finally, a SIR model estimate the peak and end of the pandemic.

## 1 INTRODUCTION

In December of 2019 the novel corona virus was recognized in Wuhan, China. The prevalence of this outbreak was unclear as the situation did rapidly evolved [1]. Several cases were soon reported outside of China in all over the world, with a total of 4515 cases including 106 deaths confirmed the 27<sup>th</sup> of January 2020 [2]. On March 12<sup>th</sup> 2020 was declared a global pandemic by the World Health Organization, and more than 2.4 million infections and over 160.000 deaths were confirmed for April 20<sup>th</sup> [3]. As the disease does not have any specific treatment, its prevention and the healthcare services preparation becomes really important.

This is why modeling and forecasting of confirmed cases of COVID-19 will help the treatment system and countries' policies. This study will apply different statistical approaches to describe the spreading growth of the virus. The models used will be the exponential and logistic growth models, Polynomial, Lasso and Ridge regression, regression trees and finally a SIR model to estimate the end of the epidemic.

This research will be answering the following questions:

- How will the COVID-19 virus spread over time and how many confirmed cases, deaths and recovered cases exist for a given date and region?
- Which statistical model will better describe the spreading growth of the virus?
- When will people recover and is it possible to find a confidence interval for the ending date of COVID-19 virus?

Studies have shown that the predictions of confirmed cases and deaths by the virus have a wide range of differing outcomes. The centers for Disease Control and Prevention suggested that between 160 million and 210 million confirmed cases will exist in the US in one year and the death toll ranged from 200,000 to 1.7 million. [4] This has been a world wide challenge and the main reason for these predicted results is the assumptions done to formulate the model. Most of the statistical models are highly sensitive to different assumptions.[5]

The report begins with a presentation of the data source. Methodology and models used are described in section 3 and a summary of the results and key findings of the proposed models is shown in section 4. Finally, conclusions are introduced.

## 2 DATA DESCRIPTION

The main data set used in this study was provided by Johns Hopkins University Center of Systems Science and Engineering. [6] This data set is daily updated from international health organizations around the world. Is possible to find data as a daily report of confirmed, deaths and recovered cases for a given country/region and province/state as the following image

Province_State	Country_Region	Last_Update	Lat	Long_	Confirmed	Deaths	Recovered
South Carolina	US	2020-04-08 22:51:58	34.22333378	-82.46170658	5	0	0
Louisiana	US	2020-04-08 22:51:58	30.295064899999996	-92.41419698	86	2	0
Virginia	US	2020-04-08 22:51:58	37.76707161	-75.63234615	11	0	0

Figure 1: COVID-19 daily reports data

## 3 PROPOSED METHODOLOGY

### 3.1 Exponential and Logistic growth models

The experience gained by previous epidemics drove researchers to implement exponential and logistic growth models to explain the outbreak of COVID-19 [7]. The reason for using an exponential growth model is to capture the initial outbreak of COVID-19 which is identified by below equation (1):

$$C(t) = C_0 \times g^t \quad (1)$$

Here,  $C_0$  and  $g^t$  represent the initial cases and growth factor (number of people infected by each patient), respectively, whereas  $C(t)$  gives the number of cases at time 't'. Therefore, these parameters that are calculated by a linear regression fit, can help comparing the outbreak at different countries during the exponential growth stage. On the other hand, logistic growth model can be represented with the following equation:

$$y(t) = \frac{Asym}{1 + \exp(\frac{X_{mid}}{scal})} \quad (2)$$

$Asym$ ,  $X_{mid}$  and  $scal$  are the parameters to be calculated with logistic growth modelling.  $Asym$  gives the asymptote of the logistic curve fitted, interpreted as the ultimate number of cases,  $X_{mid}$  shows the inflection point, in days, where the growth rate starts slowing down (concave downward curve), and  $scal$  represents a scaling parameter. A logistic growth model can be better applicable on overall trend in COVID-19 data mainly because of the mitigated number of cases with time that flattens out the time-dependent curve after a certain period of time. In other words, as people and governments start taking measures such as enhanced health caring units, stay-at-home orders, and wearing masks in public, it is expected to observe a deviation from an exponential curve towards a S-shaped curve due to the improvements on containing the virus in long term. However, there are some limitations that must be kept in mind for logistic growth models as well. For instance, premature relaxing or completely lifting the isolation and social distancing orders can cause a second wave of outbreak that results in underestimated number of cases [8]. With this in mind, we utilize exponential and logistic growth models to analyze the spreading of virus and

death tolls with time in some countries that have been suffering from COVID-19 heavily, such as China, Italy, South Korea, Spain, Germany and USA. Performance metrics used are: mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). Also, coefficient of determination, R-squared, was calculated for exponential models. The analyses was done with the data from January 22<sup>nd</sup> until April 13<sup>th</sup>. These countries are compared in terms of the number of cases (infected people) and death tolls from COVID-19. Parameters computed from fitted exponential and logistic growth models confer insights for the performances of countries on tackling COVID-19. A statistical software 'R' was used to compute logistic and exponential function parameters [9].

### 3.2 Polynomial Regression

In order to capture the relationship between the number of confirmed cases and the day(s) since the data was collected about COVID-19, a polynomial regression was used to fit the data and then predict the number of cases 7 days from the last date in the data. This will allow us to forecast the number of confirmed cases and help authorities to make informed decisions about policies and interventions to keep the spread of the coronavirus in check.

The main reason a polynomial regression model was chosen among other regression models was because it is a simple, easy to implement and accurate model when we need to transform a linear model in order to fit non-linear data. Another reason was due to characteristics regarding number of active cases over time observed in China. There is a quasi-exponential growth after certain number of cases have reached. In order to visualise the growth of the cases and fatalities as the days pass, a plot has been shown below (Figure 2).

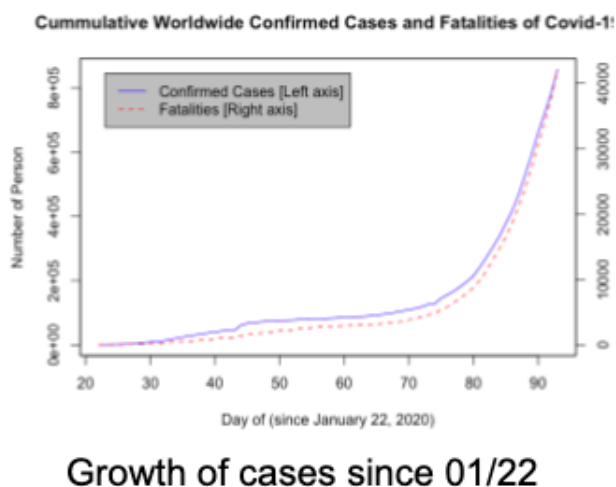


Figure 2: Growth of confirmed cases and fatalities since 01/22

The number of confirmed cases was evaluated as a function of the number of days. The following equation describes the model that was used to fit the data.

$$Y = X + X^2 + X^3 + X^5 \quad (3)$$

This polynomial was chosen on a trial basis. Here X variable is the number of days since . And the Y variable is number of confirmed cases. Also, number of fatalities was taken as the variable and the above equation was used to predict the number of deaths as well. The `lm` function in R was used for the model.

### 3.3 Ridge and Lasso Regressions

Ridge and Lasso regression are some of the simple techniques to reduce model complexity and prevent over-fitting which may result from simple linear regression.

**Ridge Regression:** In case of ridge regression a penalty is added to alter the cost function. The penalty is equivalent to square of the magnitude of coefficients.

$$\sum_{i=1}^M (y_i - \hat{y}_i)^2 = \sum_{i=1}^M (y_i - \sum_{j=0}^p w_j * x_{ij})^2 + \lambda \sum_{j=0}^p w_j^2 \quad (4)$$

$$\sum_{j=0}^p w_j^2 < c \dots (c > 0) \quad (5)$$

So ridge regression puts constraint on the coefficients (w). The penalty term  $\lambda$  regularizes the coefficients such that if the coefficients take large values the optimization function is penalized. So, ridge regression shrinks the coefficients and it helps to reduce the model complexity and multi-collinearity. Going back to the cost function of Ridge Regression one can see that when  $\lambda \rightarrow 0$ , the cost function becomes similar to the linear regression cost function. So lower the constraint (low  $\lambda$ ) on the features, the model will resemble linear regression model.

**Lasso Regression:** The cost function for Least Absolute Shrinkage and Selection Operator (LASSO) regression can be written as:

$$\sum_{i=1}^M (y_i - \hat{y}_i)^2 = \sum_{i=1}^M (y_i - \sum_{j=0}^p w_j * x_{ij})^2 + \lambda \sum_{j=0}^p |w_j| \quad (6)$$

$$\sum_{j=0}^p |w_j| < t \dots (t > 0) \quad (7)$$

Just like Ridge regression cost function, for  $\lambda = 0$ , the equation above reduces to the linear one. The only difference is instead of taking the square of the coefficients, magnitudes are taken into account. This type of regularization (L1) can lead to zero coefficients i.e. some of the features are completely neglected for the evaluation of output. So Lasso regression not only helps in reducing over-fitting but it can help us in feature selection.

The coding for this part was done completely in Python. For implementation of Ridge and Lasso regression we can use `Ridge()` and `Lasso()` libraries from `scikit-learn`. While implementing hyper parameter tuning for Ridge and Lasso we made use of Grid Search CV or Random Search CV technique. The overall procedure for implementation of models is exploratory data analysis, feature engineering, model implementation, hyper parameter tuning and model evaluation.

### 3.4 SIR Model

Now that we performed a comparative study on the statistics of infected, and recovered patients as well as fatalities using various regression-based models, the last model will focus on the giving us an estimate as to when the horrors of pandemic will end and give insights regarding the effects of public health interventions by the governments around the world. The model which will be used to address the above goal is the SIR model. This model is widely used for infectious diseases' statistics. The basic idea of the model is that there are 3 groups of people -

1. Susceptible - people who are vulnerable to exposure
2. Infectious - the infected people, who can pass on the illness and recover over time
3. Recovered - the people who have recovered

Despite the simplicity of the model, this model provides a reasonable estimate as to how number of the people in each group change over time. To build this model (Figure 3), we need the 3 following equations to describe the change in each group.

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \quad (8)$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \quad (9)$$

$$\frac{dR}{dt} = \gamma I \quad (10)$$

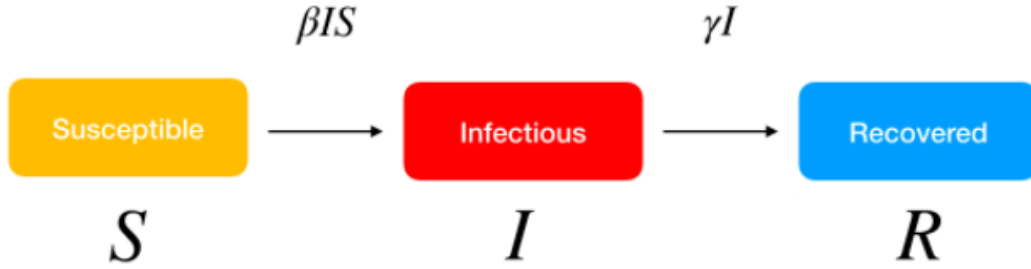


Figure 3: Schematic illustration of SIR model

Residual sum of squares (RSS) is used to find best values of  $\beta$  and  $\gamma$  based on the given observed data.

The `EpiModel` package in R was used to provide functions for stochastic, discrete-time, individual contact models (ICMs) that we will need for the SIR model. We then run the simulation on a population of 1000 people with some initial values to parameters such as exposure-to-infection rate, death rate, recovery rate, and probability of infection. We explore the public health interventions by quantifying parameters of the model affected by this (for example, reducing exposure rate due to social distancing).

Also, to compare against a more sophisticated model for this purpose (goal #3), we ran experiments on an another model called SEIQHRF (shown in Figure 4), built by Tim Churches. [10] It is a much more complex model and takes into multiple parameters like -

1. S - Susceptible individuals
2. E - Exposed and infected, not yet symptomatic but potentially infectious
3. I - Infected, symptomatic and infectious
4. Q - Infectious, but (self-)isolated
5. H - Requiring hospitalisation (would normally be hospitalised if capacity available)
6. R - Recovered, immune from further infection
7. F - Case fatality (death due to COVID-19, not other causes)

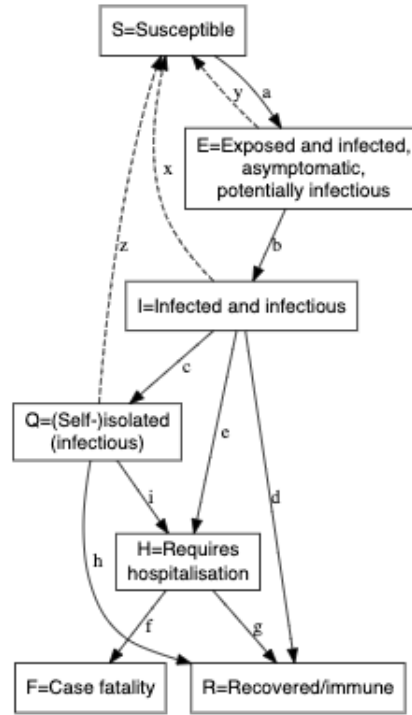


Figure 4: Model diagram of SEIQHRF model

The letters on the transition edges denote model parameters relating to that transition. We will not be discussing the intricate details of this model in the report (as this model is not built by us, we will only want to give an introduction to this model and compare its results against the SIR model).

## 4 ANALYSIS AND RESULTS

In this section, it is preferred to give analyses on number of confirmed cases which is also the main variable for death tolls for a given region or a country. Some of logistic growth model figures for death tolls and the number of confirmed cases that are not given in this section, are provided in Appendix.

Germany and USA are selected for data and modelling visualization in this section. Germany is given as an example of Europe where restrictions are being lifted gradually whereas USA is a country of interest that is still motivated by strict restrictions. Figure 5 shows the number of confirmed cases with time in Germany and USA. Exponential growth fitted curve and real data is represented as blue line and actual data is represented with red circles. Clearly, exponential fit underestimates up to a certain period of time for both of the countries then it starts overestimating the amount of spreading. R-squared of exponential models in both of the countries are nearly identical as 0.95 for Germany and 0.94 for USA. At the time of the analyses, Germany and USA had 127,854 and 555,313 confirmed COVID-19 cases in total, respectively. Although these countries have different populations and different number of confirmed cases by the time of this analysis, both of them showed an exponential increase in their number of confirmed cases between 40 and 50 days after the first case reported. This is a rather interesting observation indicating inadequate or zero interventions taken in such countries within that period to stop the spreading of virus.

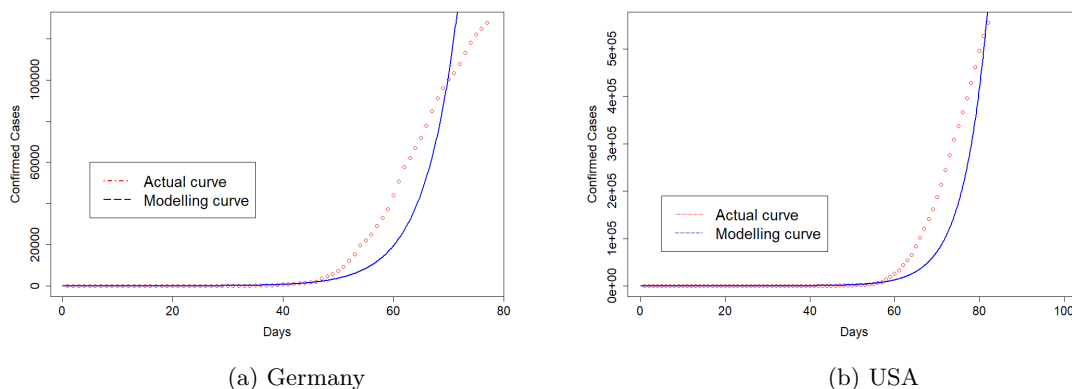


Figure 5: Number of confirmed cases vs Time with the exponential growth model. Red dots indicate the actual data whereas blue line shows the fitted curve.

With the motivation of obtaining a better fit, logistic function was implemented to same data and of results are presented in Figure 6. Both Germany and USA models show significantly



better fit with logistic growth models than exponential model. To illustrate this difference, MAE for Germany decreased from 13399.24 to 765.11 and RMSE was decreased from 35,157.66 to 1163.83. For USA, MAE was decreased from 25771.14 to 2428.55 while RMSE was decreased from 55165 to 3903.74 using logistic growth model.

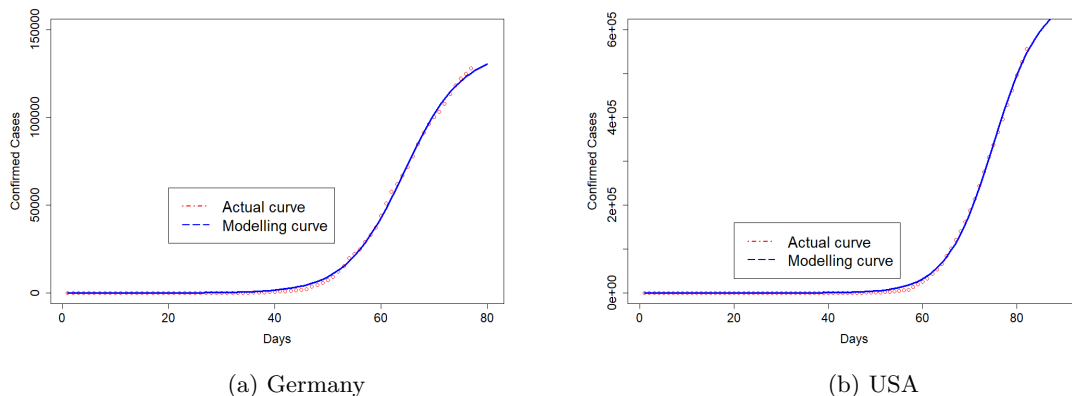


Figure 6: Number of confirmed cases vs Time with the logistic growth model. Red dots indicate the actual data whereas blue line shows the fitted curve.

Figure 7 displays the parameters computed by exponential growth model for different countries. Parameter 'g' (growth factor) is the highest for Spain, 1.24, and the lowest for South Korea with 1.13. In particular, among the countries studied, European countries displayed higher growth factors. On the other hand, calculated  $C_0$  values point out that South Korea was the country with the highest amount of confirmed cases initially while Spain has the lowest number of confirmed cases.

Country	$C_0$	g	$R^2$
Germany	0.92	1.18	0.95
Italy	0.81	1.22	0.91
Spain	0.15	1.24	0.95
South Korea	3.74	1.13	0.85
USA	0.35	1.19	0.94

Figure 7: Exponential growth model parameters

Parameters computed by logistic growth models vary significantly between countries (see Figure 8). However, 95% confidence intervals calculated, shown in Appendix, are narrow favoring the robustness of the model. We mentioned that South Korea has the highest number of initial confirmed cases that is deduced by exponential model. However, comparison between asymptotes (parameter Asym) of fitted logistic curves for different countries, indicates the lowest amount of

COVID-19 infections in South Korea, one and two magnitude of order lower than other countries compared. This success of South Korea in particular is not surprising due to their achievements in massive testing and countrywide quarantine measures [11].  $X_{mid}$ (inflection point) displays that China seems to be the most successful country that started reducing the spreading rate of virus. China achieved to reach inflection point 18 days after the first reported COVID-19 case, and South Korea reached after 42 days. Also, China has the second lowest 'Asym' after South Korea. One might expect that high population corresponds to the highest number of confirmed cases at the end. However, aforementioned lower ultimate infection prediction (Asym) for China underlines the importance of proactive measures against a pandemic discarding population as a variable. On the other hand, European countries such as Germany (64 days), Italy (55 days) and Spain (58 days), and especially USA (74 days) had relatively longer time periods until they can reduce the slope of their COVID-19 spreading curves. This difference between countries may be explained by the type and timing of interventions taken, and needs to be further analyzed.

Country	Asym	Xmid	scal	MAE	MAPE	RMSE
China	8.16E+04	18.87	4.59	1271.39	0.13	1868.2
Germany	1.38E+05	64.33	5.45	765.11	1.78	1163.83
Italy	1.61E+05	56.27	6.58	1457.46	25.28	1958.3
Spain	1.76E+05	59.06	5.12	875.95	9.09	1318.15
South Korea	9.65E+03	42.68	4.62	343.59	1.03	441.79
USA	6.80E+05	75.02	5.01	2428.55	1.9	3903.74

Country	Date of first case reported	Date of Xmid (Inflection point)	Days passed between first case and Xmid
China	1/22/2020	2/9/2020	18
Germany	1/27/2020	3/30/2020	63
Italy	1/31/2020	3/26/2020	55
Spain	2/1/2020	3/30/2020	58
South Korea	1/22/2020	3/4/2020	42
USA	1/22/2020	4/5/2020	74

Figure 8: Logistic growth model parameters

After training the polynomial regression model on the data, we achieved an accuracy of 0.924 on the number of cases and 0.937 on the number of fatalities. Following figure (9) shows how well the model fits the data in Italy, Hubei province (China) and New York (US) respectively.

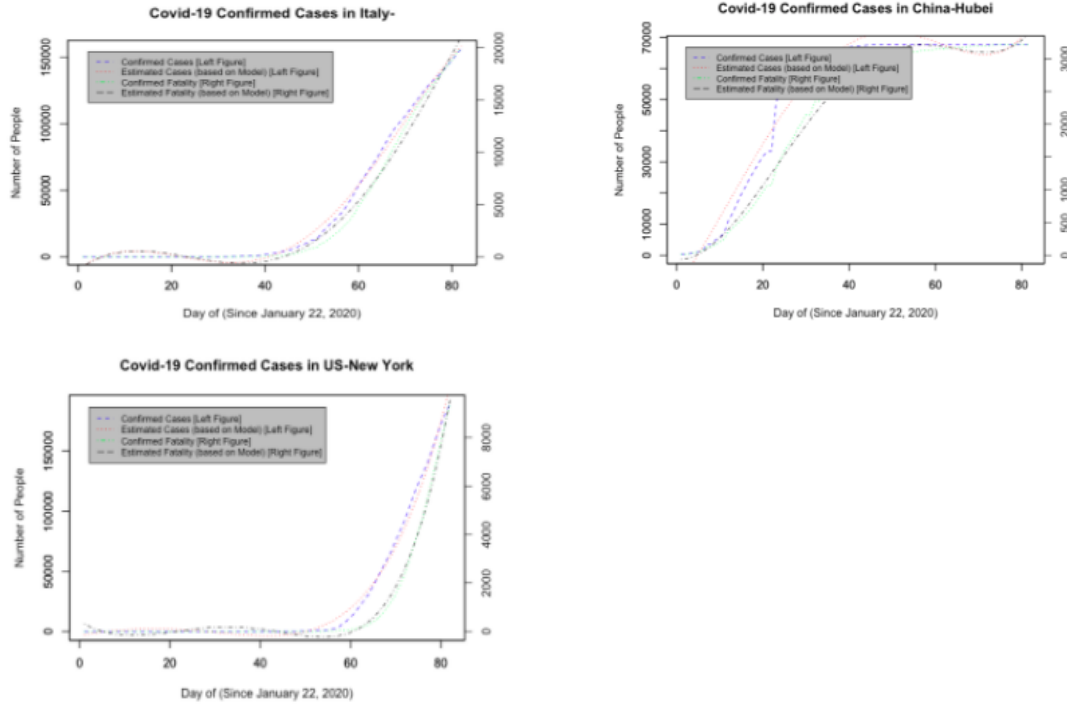
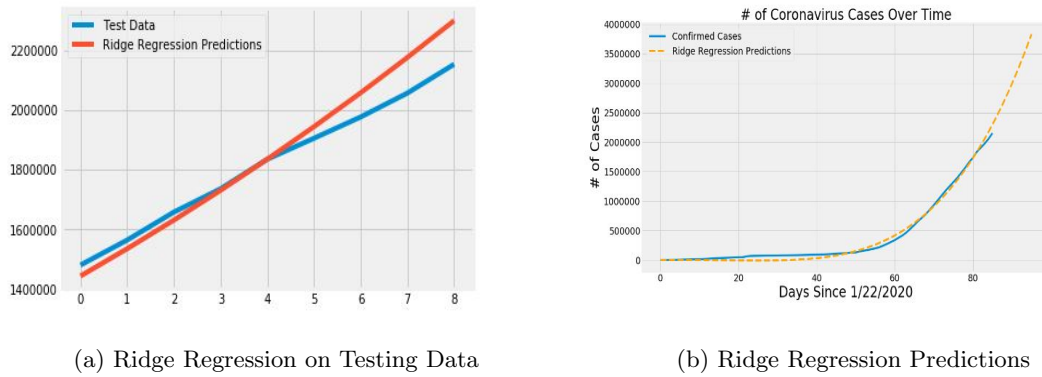


Figure 9: Polynomial regression fitting on different countries

In this sub - section, we shall discuss the evaluation metrics of Ridge and Logistic Regression. The theoretical details regarding these regression techniques have been given in the proposed methodology section. In this case, we have trained the model on the global data and both the models then predict global growth. The plots for ridge regression can be seen as below in Figure 10:



(a) Ridge Regression on Testing Data

(b) Ridge Regression Predictions

Figure 10: Predictions using Ridge Regression: On Testing data and Actual Forecast

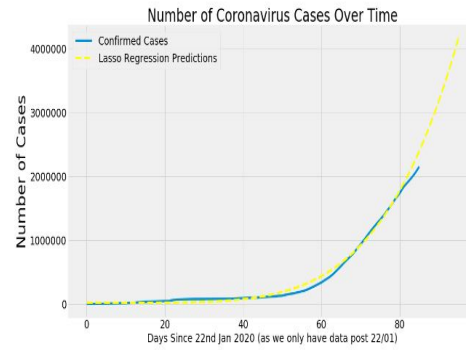
The above figure show the performance of Ridge Regression on the testing data set and the predictions of Ridge Regression.

The Mean Absolute Error in this case is 53659.44 and the Mean Squared Error is 5.173E+09. Now let us compare the results of Ridge Regression with Lasso.

The plots of Lasso Regression are as below in Figure 11:



(a) Lasso Regression on Testing Data



(b) Lasso Regression Predictions

Figure 11: Predictions using Lasso Regression: On Testing data and Actual Forecast

The above figures show the performance of Lasso Regression on the testing data set and the predictions of Lasso Regression on the global data set.

The Mean Absolute Error in this case is 90745.43 and the Mean Squared Error is 15.5E+09. Now let us compare the results of Ridge Regression with Lasso.

The MAE (Mean Absolute Error) measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables. The MAE is a linear score which means that all the individual differences are weighted equally in the average. The MSE (Mean Squared Error) is a quadratic scoring rule which measures the average magnitude of the error. The equation for the MSE is given in both of the references. Expressing the formula in words, the difference between forecast and corresponding observed values are each squared and then averaged over the sample. This means the MSE is most useful when large errors are particularly undesirable.

Understanding the metric definitions and looking at the MAE and MSE values, we can say that Ridge Regression performs better than Lasso Regression. It has better predictive power at least for the data set under consideration.

To further substantiate our conclusions, please find the below two tables forecasting next 10 days of Coronavirus cases globally using Ridge and Lasso Regressions. The table forecasts cases from 16th April to 26th April. After looking at the actual values for the dates and comparing it with the forecast values we can say that Ridge Regression performs better as compared to Lasso.

Date	Ridge Regression Predicted number of Confirmed Cases Worldwide
0 04/17/2020	2426935.0
1 04/18/2020	2560420.0
2 04/19/2020	2699234.0
3 04/20/2020	2843517.0
4 04/21/2020	2993406.0
5 04/22/2020	3149045.0
6 04/23/2020	3310576.0
7 04/24/2020	3478145.0
8 04/25/2020	3651897.0
9 04/26/2020	3831982.0

(a) 10 Days Forecast: Ridge Regression

Date	Lasso Regression Predicted no of Confirmed Cases Worldwide
0 04/17/2020	2539482.0
1 04/18/2020	2689425.0
2 04/19/2020	2846422.0
3 04/20/2020	3010721.0
4 04/21/2020	3182572.0
5 04/22/2020	3362233.0
6 04/23/2020	3549967.0
7 04/24/2020	3746043.0
8 04/25/2020	3950736.0
9 04/26/2020	4164328.0

(b) 10 days Forecast: Lasso Regression

Figure 12: Comparison between 10 days forecasts of Lasso and Ridge Regression

We shall now end the evaluation of Ridge and Lasso Regression algorithms. Moving further we shall now concentrate on SIR modeling, which shall provide us a rough estimate of the peak and also the end of the outbreak ,assuming the current health interventions are still in place.

On simulating the SIR model on a population of a 1000 people, the following results were obtained:

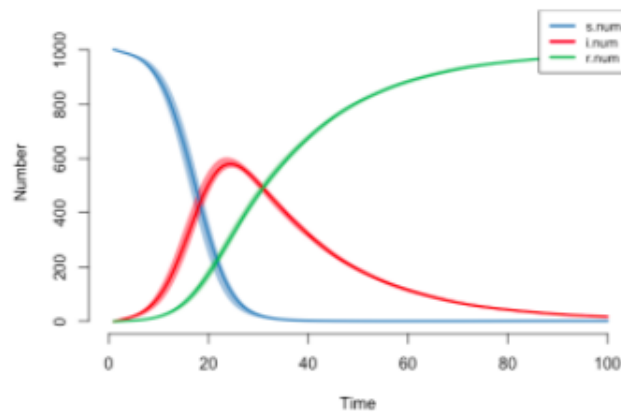


Figure 13: The peak of infection reveals  $R_0$  of 2.2. This is how the timeline would have been in a population of 1000 during COVID-19

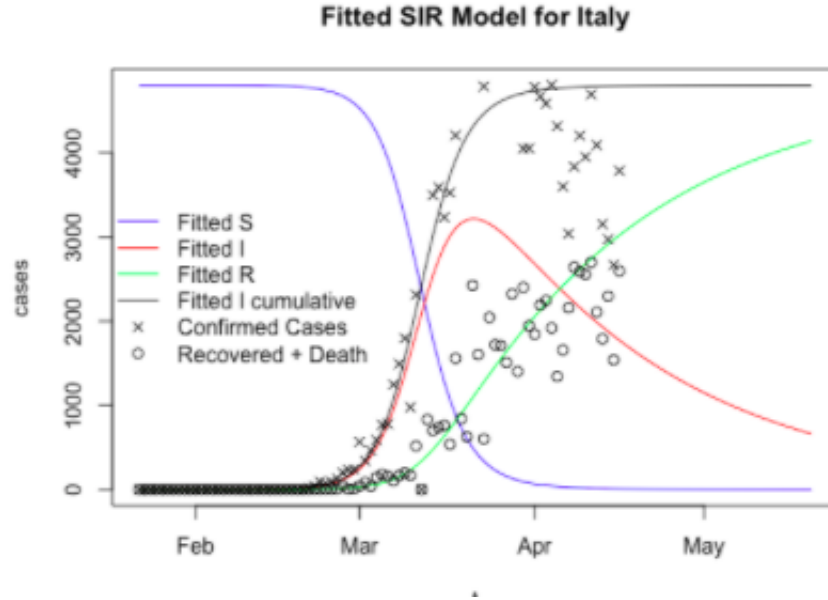


Figure 14: SIR model on Italy data. Y axis represents daily cases.

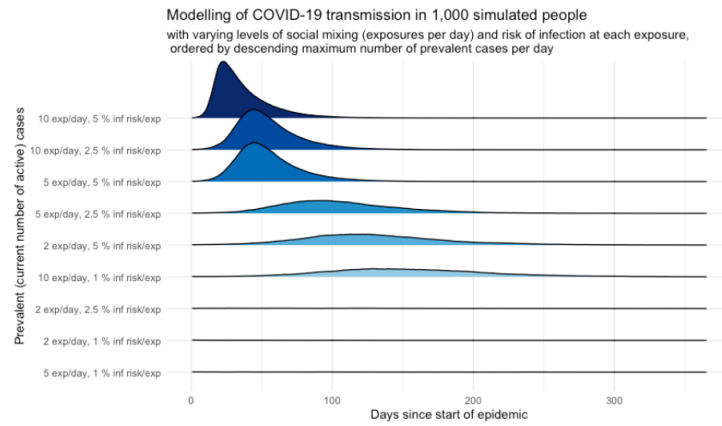


Figure 15: In order to flatten the curve, social distancing is necessary and better if done early

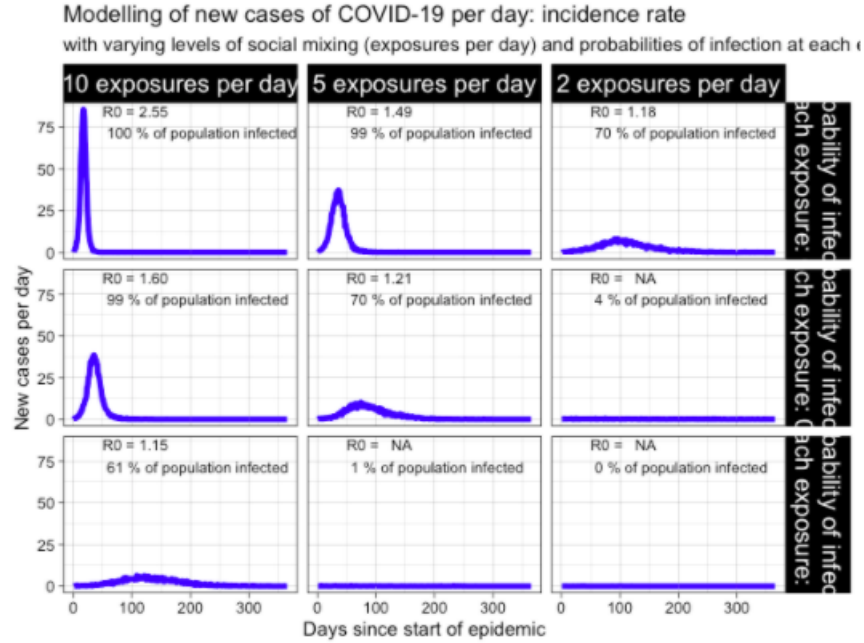


Figure 16: Modelling new cases of COVID-19

As we decrease the average number of exposures per day to infected people, from left-to-right, and decrease the risk of becoming infected at each of those exposures, from top-to-bottom, the epidemic curve is flattened. Not only is it spread out over a longer period of time, the total area under the curve, which gives the total number of people in our population who are infected, is reduced.

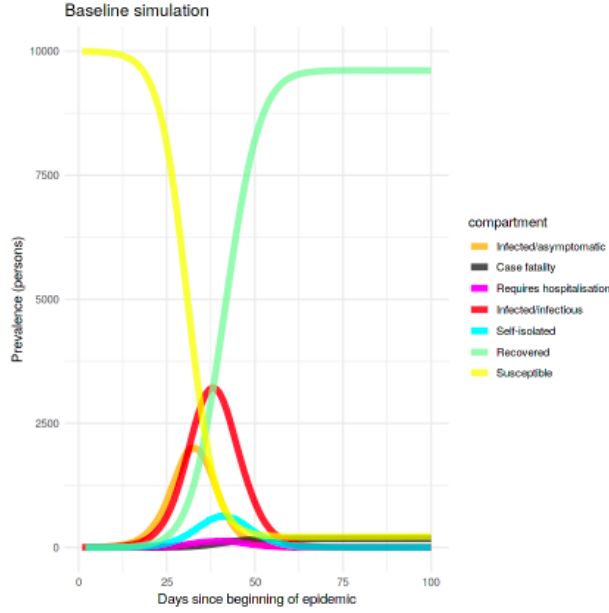


Figure 17: Result from SEIQHRF model.

Number of people in the I and Q compartments are similar but lag behind the E compartment. Number requiring hospitalisation seems reasonable. Number in the case fatality compartment is monotonically increasing.

## 5 CONCLUSIONS

- Exponential growth modelling is useful to explain initial spreading of virus. However, it overestimates after certain period of time and loses its prediction accuracy.
- Logistic growth model is promising to capture the overall trend, make comparison between countries and make predictions for future. It also enables gaining insights about the time when spreading of virus levels off due to interventions taken by governments.
- Polynomial Regression seemed one of the most accurate models till now. It provides the best approximation of the relationship between number of cases and time, as it is able to fit the non-linear data more closely, which means it can fit a broad range of functions. However, it is too sensitive to outliers.
- Comparing the regularization models specifically, the ridge regression stands out in predicting Coronavirus cases when we look at the model evaluation metrics. The main reason here could be that ridge regression shrinks the coefficients and it helps reduce the model complexity and multi-collinearity.



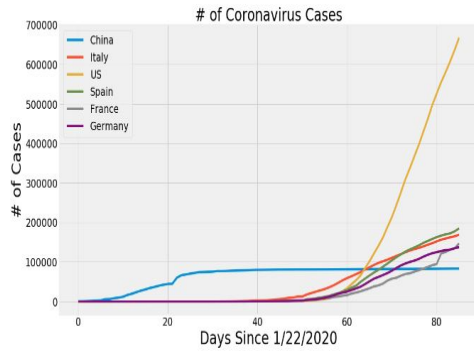
- If we look at figure 11 of our report, it further substantiates our conclusion that Ridge regression better predicts the Coronavirus cases as compared to Lasso. We are only comparing the regularization models here. The tables are 10 day forecast values using Ridge and Lasso. Comparing it with the actual values during the period, Ridge stands out.
- Concluding from SIR model, a rough estimate of the peak can be generated and also the end of the outbreak, assuming the current health interventions are still in place.

## 5.1 Learnings

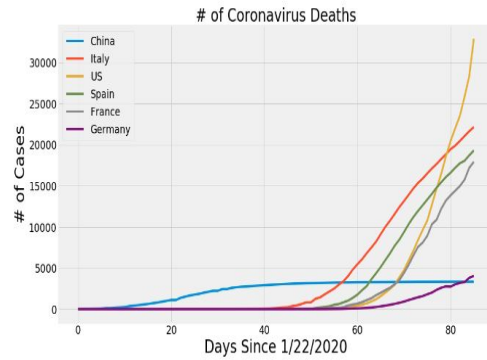
- The class project was valuable to get experience on how to approach real data, and how to make meaningful conclusions by performing exploratory data analysis. By focusing on intricate details in the data, we learnt how to decide which models that we should implement to address our goals in this project.
- Understanding the spreading trend of a pandemic was of paramount importance on real-time data. Systematical comparisons between exponential and logistic growth models provided clear insights about the performances of different countries in addressing the COVID-19 outbreak. In addition, experience with 'R' was enhanced throughout this project by using new packages for the computational purposes.
- We also learned how to build models for polynomial regression and SIR, and how to visualise the results, using various packages and functions of R, like `caret`, `DEoptim`, `tidyr`, `ggplot` and many more.
- In conjunction with the above regression models, we also learned how to implement regularization models i.e. Lasso and Ridge Regression for predictive modeling. As these regularization models were implemented using scikit-learn machine learning library, we got good exposure to implementing models using scikit-learn. The model implementation did involve hyper parameter tuning and this exemplified us techniques like Random Search CV and Grid Search CV to tune our hyper parameters for model optimization. The overall project further provided us with learnings in Data Visualization using matplotlib and seaborn libraries. In was an end to end regression and predictive modeling project.
- Apart from the technical knowledge, we also improved in other aptitudes like teamwork and communication. Most importantly, we learnt how to collaborate remotely, in this unusual situation.

## 6 APPENDIX

### 6.1 Exploratory Data Analysis of our dataset

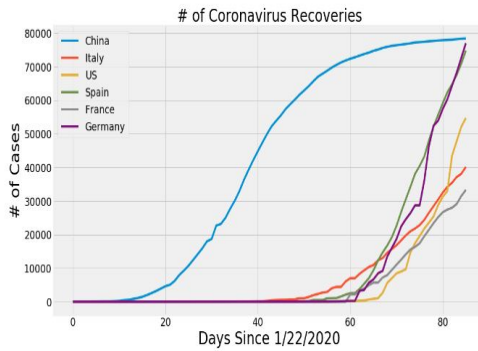


(a) No of World Cases

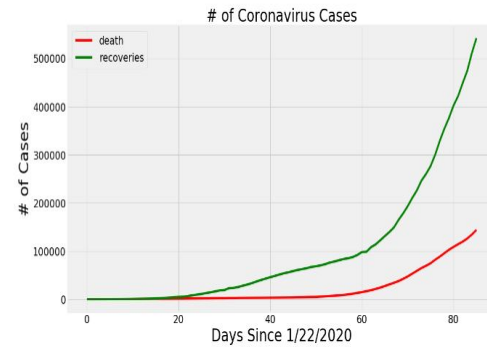


(b) Number of Deaths in World

The above figures show the total number of confirmed cases and the total number of deaths in the world as of 16th April.

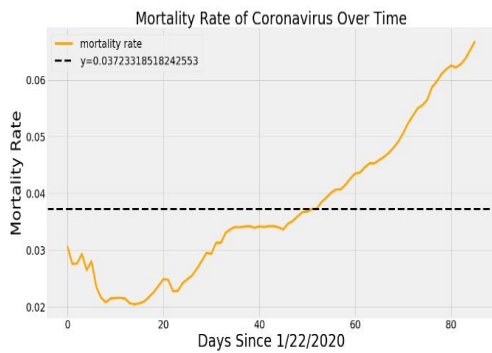


(a) World Recoveries

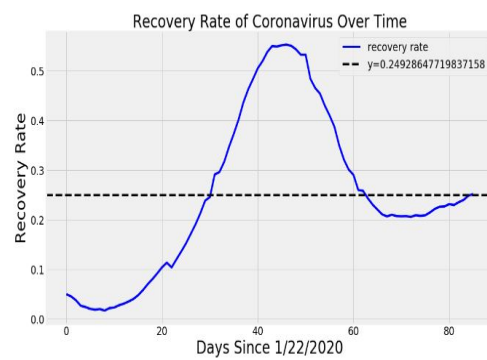


(b) Death vs Recoveries

The figures show the recovery and death statistics across the globe.

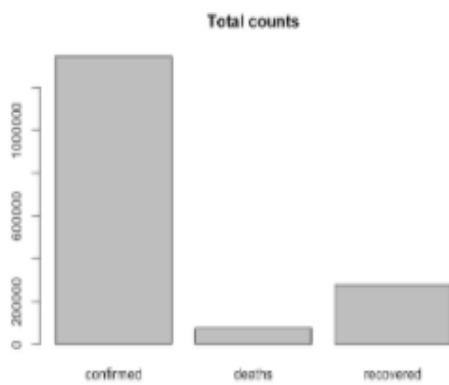


(a) Recovery Rate

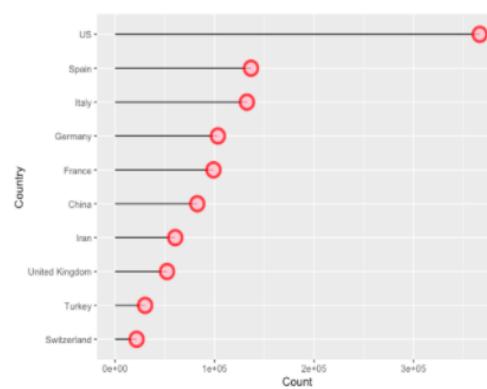


(b) True Recovery Rate

Here, we can see a comparison between mortality rate and recovery rate.

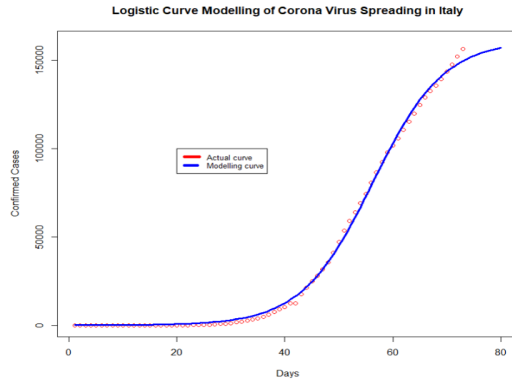


(a) Confirmed vs Death vs Recovered cases

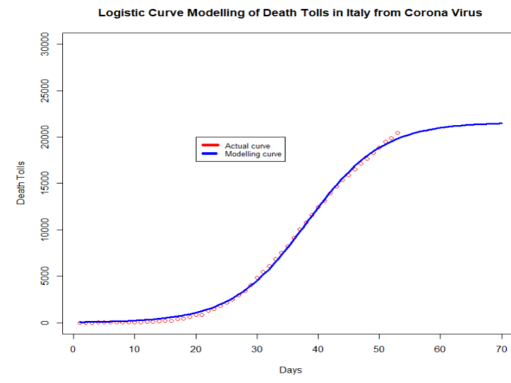


(b) Top 10 affected Countries

## 6.2 Logistic Growth Model Curves and Confidence Intervals

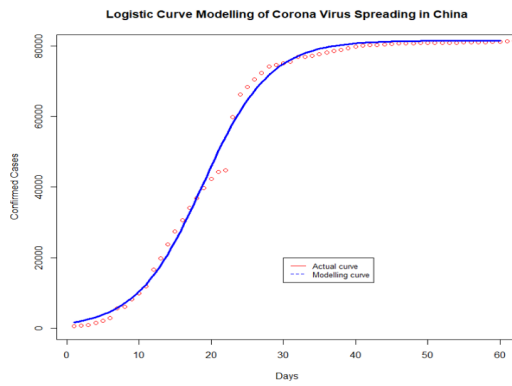


(a) Number of confirmed cases

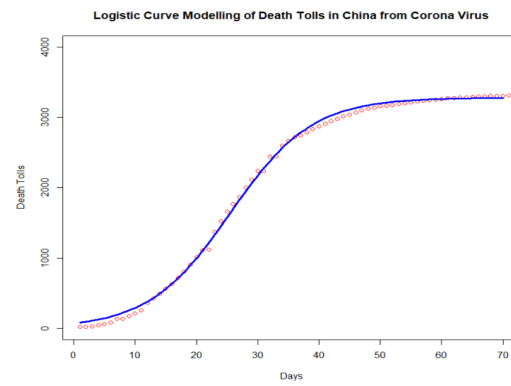


(b) Death tolls

Logistic growth fitting for Italy



(a) Confirmed cases



(b) Death tolls

Logistic growth fitting for China

Country	95% Confidence Interval		
	Asym	Xmid	scal
<b>China</b>	[8.10E+04, 8.2E+04]	[18.61, 19.13]	[4.36, 4.81]
<b>Germany</b>	[1.35E+05, 1.4E+05]	[64.1, 64.6]	[5.3, 5.6]
<b>Italy</b>	[1.58E+05, 1.65E+05]	[55.9, 56.7]	[6.3, 6.8]
<b>Spain</b>	[1.73E+05, 1.78E+05]	[58.8, 59.3]	[5.0, 5.3]
<b>South Korea</b>	[9.45E+03, 9.86E+03]	[42.1, 43.3]	[4.1, 5.2]
<b>USA</b>	[6.61E+05, 7.00E+05]	[74.7, 75.4]	[4.87, 5.15]

Confidence intervals of parameters obtained by logistic growth fitting

## References

- [1] Anthony Fauci Catherine Paules, Hilary Martson. Coronavirus infections-more than just the common cold. *JAMA*, 323(8):707–708, 2020.
- [2] K. Hayashi et al. S. Jung, A. Akhmetzhanov. Real-time estimation of the risk of death from novel coronavirus (covid-19) infection: Inference using exported cases. *Clinical Medicine*, 2020.
- [3] Jasmin Collier. Covid-19 live updates: Total number of cases passes 2.4 million. *Medical News Today*, 2020.
- [4] Reid Wilson. Worst-case coronavirus models show massive us toll. *The Hill*, 2020.
- [5] N. Michel K. Dayaralna. The challenges of forecasting the spread and mortality of covid-19. *The Heritage Foundation*, 2020.
- [6] Johns Hopkins University Center of Systems Science and Engineering. Novel coronavirus (covid-19) cases. Available from <https://github.com/CSSEGISandData/COVID-19>, 2020.
- [7] Pehong Chen. Gnu emacs bibtex-mode. Technical report, Technical Report 87/317, Computer Science Division, University of California . . . , 1986.
- [8] Raj Dandekar and George Barbastathis. Quantifying the effect of quarantine control in covid-19 infectious spread using machine learning. *medRxiv*, 2020.
- [9] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [10] Tim Churches. Modelling the effects of public health interventions on covid-19 transmission using r, 2020.
- [11] Kanga Kong. Seoul’s full cafes, apple store lines show mass testing success, 2020. <https://www.bloomberg.com/news/articles/2020-04-18/seoul-s-full-cafes-apple-store-lines-show-mass-testing-success>, Last accessed on 2020-04-20.