Determining the Effectiveness of the Stay-At-Home Orders with Different Models

Tomo Bessho
NUID: 01340655

May 4, 2020

1 Introduction

The spread of COVID-19 has undoubtedly affected the daily lives of every American today. To counteract and slow the spread of this pandemic, each state and its governors were tasked to issue their own stay-at-home orders to reduce the number of people encountering one another daily.The governor of my home state of Nebraska, along with a handful of other states, has refused to enact a mandatory stay-at-home order (SAHO) to its state's citizens. Seeing that the majority of states have ordered people to stay at home during this pandemic, this has lead to me to ask the question of whether or not the SAHOs are helping to slow the spread of COVID-19 and to what extent these policies have on containing the spread of COVID-19 [1]. I will attempt to model the spread of COVID-19 in each state using different regression and epidemiological mathematical models before and after the issuing of the SAHO (for the states that have a mandatory stay-at-home order) and comparing these two models in each state to see if the spread of COVID-19 has been affected by these orders.

2 Methodology

In this section, I outlined where my data is coming from, some important notes/assumptions I made during this project, the derivation and the process of building the polynomial least square approximations of the data, and the background, derivation, and method of building the SIR epidemiological mathematical model.

2.1 Data Acquisition

The models in this project require up-to-date, accurate data about the spread of COVID-19 in each state in the Untied States. From the many sources available online, this project will incorporate the data from the 2019 Novel Coronavirus COVID-19 Data Repository by Johns Hopkins CSSE. This is a GitHub repo that the JHU CSSE updates regularly daily that provides data on the number of positive cases, number of deaths, number of recovered for every region in the world. This GitHub repo is being used by the Johns Hopkins visual dashboard that has been the

primary visual source for modelling the spread of COVID-19 in the world. In addition, its data sources include the World Health Organization, the CDC, and several other credible sources [2]. For this project, I will have a script that will pull the statistics from this repo, filter out the information for each state, and have this data available for modelling. The data taken from the repo contains the information of the total number of cases, total number of recovered people, and the total number of people that have died from the virus, which will be later used to create the models for each state. The total population for each state will be taken from a different site as well, as it will be needed for one of the parameters in the epidemiological mathematical models [3].

2.2   Important Notes

Before going into the method of data analysis for this project, I wanted to outline a few notes. Since the goal of this experiment is to determine whether a state's stay-at-home order has been effective or not, I only analyzed states that have already issued a stay-at-home order. This excludes a few states, such as Nebraska, Iowa, and a few other states. The reasoning behind this is that I wanted to compare the data for before and after the SAHO was issued in a state. If there was no data for after the SAHO was issued in a state, I cannot conclusively understand the effectiveness of the SAHOs.

A second note is that to fully determine the effectiveness of a SAHO in a state, instead of using the date the SAHO was issued in a state, I used 2 weeks after the date the SAHO was issued in a state. When infected, it can take a person from 2-14 days to show any symptoms of the virus [4]. In addition, it takes from a day to several days for test results to return [5]. Due to this, I am assuming that the data given on a certain date reflects the data two weeks prior to the given date. Thus, when I analyzed the effectiveness of the SAHO, I used the date two weeks after the SAHO was issued in the given state to separate the data for analysis. Whenever I use the phrase "separation date" in this paper, I am referring to 2 weeks in advance of the date the SAHO was issued in a state.

Lastly, there are some states that are not actively reporting the number of recovered people, which is an issue as the epidemiological model that I am using relies on the number of recovered individuals as a initial value (this will be explained more in detail later on). To combat this issue, I found that the average recovery rate of COVID-19 is approximately 97.5% [5]. Using this and the fact that it takes about 2 weeks for an individual to fully recover from the virus, I took the 97.5% of the number of cases from two weeks prior to estimate the number of recovered individuals for a certain date.

2.3 Discrete Polynomial Least Square Approximations

One method I will be using to try to understand the effectiveness of the SAHO is the use of discrete least square approximations. This approximation method fits a polynomial function – for this project, I fit a cubic function – to a set of data points to try to best model the data. This function can then be used to approximate both values inside and outside the given data range, although it is important that the further away the estimation is from the given data set, the less accurate the approximation becomes. For this project, I obtained the cubic function for the data before the separation date and another cubic function to the data after the separation date.

The reasoning behind the decision of using a least square cubic function to fit the data boils down to two reasons. Although most of the growth data of a virus is considered to be exponential, when I tried fitting the exponential function to the data, some of the exponential functions did not properly fit the functions as the cubic function did. Here is an example with the data in New York before the separation date.

Figure 1: Different Fitting Methods on the Data of New York
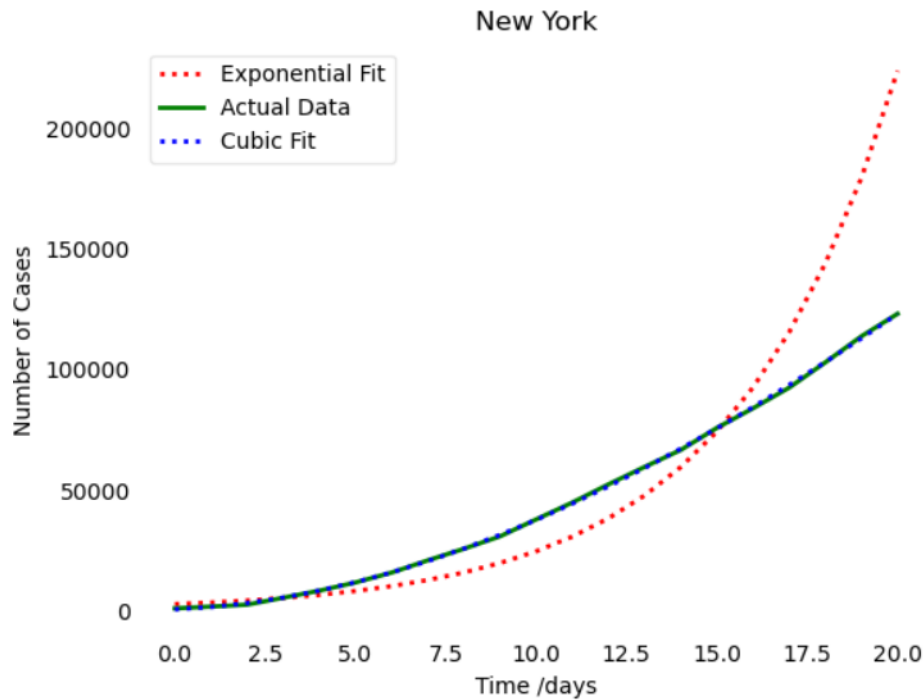
As you can see, the cubic function fits much better than the exponential function, especially as more time passes from the initial day of the outbreak in New York (Figure 1). This type of fitting was consistent with the other states as well. Due to this, I decided to fit a least square cubic function to all the data sets in this project.

To analyze the data from the cubic functions and to determine whether the SAHO are having any effect on the number of cases, I used some simple ideas from calculus. Since we are using a cubic function to model the growth of the number of cases, I proposed that we look at the

derivative and the second derivative of this function to understand the behavior of the growth of cases for each state. The derivative of the least square cubic function indicated to us how many cases of COVID-19 are increasing/decreasing in a certain time frame. The second derivative indicated the growth rate of these increases/decreases. I computed the average derivative and the second derivative at each point for both the data sets (before and after separation date) and compared these numbers to see if there was a relation.

In order to calculate and find the least square cubic function, I have to find the coefficients of each term in the cubic function. To do this, I first have to build a system of equations to solve for the coefficients. Deriving this system of equations stems from finding the error between the polynomial least square and the actual value. Let us start with a n degree polynomial we use for the approximation of the data [6].

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

To find the error of this approximation, we have to find the summation of the square differences between the approximation and the actual value [6].

$$E = \sum_{i=1}^{m} (y_i - P_n(x_i))^2$$

We then expand and simplify this equation and find the partial differential of the equation. Since we want to minimize this error, we are going to set the differential to be 0. For each coefficient $a_j, j \in [0, n]$, we get the following [6].

$$\frac{\partial E}{\partial a_j} = -2 \sum_{i=1}^{m} y_i x_i^j + 2 \sum_{k=0}^{n} a_k \sum_{i=1}^{m} x_i^{j+k} = 0$$

Simplifying each of the $n + 1$ equations, we are left with the following [6].

$$\sum_{i=1}^{m} y_i x_i^j = \sum_{k=0}^{n} a_k \sum_{i=1}^{m} x_i^{j+k}, j \in [0, n]$$

I then put all of these equations into a matrix that represented the system of equations. Doing that resulted in the following matrix equation ($Ax = b$).

$$\begin{bmatrix} \sum_{i=1}^{m} x_i^0 & \sum_{i=1}^{m} x_i^1 & \cdots & \sum_{i=1}^{m} x_i^n \\ \sum_{i=1}^{m} x_i^1 & \sum_{i=1}^{m} x_i^2 & \cdots & \sum_{i=1}^{m} x_i^{n+1} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{m} x_i^n & \sum_{i=1}^{m} x_i^{n+1} & \cdots & \sum_{i=1}^{m} x_i^{2n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y_i x_i^0 \\ \sum_{i=1}^{m} y_i x_i^1 \\ \vdots \\ \sum_{i=1}^{m} y_i x_i^n \end{bmatrix}$$

To build this system above, I noticed that the values in the A coefficient matrix are repeated, since the summations are brought down diagonally from the row above. To refrain from repeating

unnecessary computations, before building this system, I calculated the summations of all the x-values to the zeroth power, the first power, and so on, until I reached the $2n^{th}$ power. I stored all of these values in an array so that when building the equation, instead of computing the same summation over and over again, I could just pull the respective summation from this array to build the system of equation.

With the following equation above, to solve for each coefficient a_j, I augmented A with b and used the reduced row echelon form of the resulting matrix. The coefficients produced from this system of equations were then used to build the polynomial least square approximation.

2.4 Epidemiological Mathematical Models

Although the discrete least square approximations can give you a best fit function for the data set, epidemiological mathematical models can offer more accurate models with their parameters. Unlike the discrete least square approximations, the epidemiological mathematical models have parameters that correspond directly to the growth of a virus. Parameters, such as the contact rate of the disease, the mean recovery rate of the virus, the reproductive rate of the virus, and the number of people already infected are all different parameters that are taken into consideration when building the epidemiological models [7]. Epidemiological models are useful to perform necessary theoretical experiments to try to predict the future spread of the virus and to compare the effects of prevention or control procedures. In fact, Hethcote and Yorke used these models to compare the different control methods of gonorrhea in 1984 [8].

Of the several different types of epidemiological models available to use, I chose to use the SIR model. When a virus spreads rapidly and is determined as an epidemic, the SIR model is an effective model [8]. The SIR model divides a general populous (for this project, a state's population) into three fixed populations, those who are susceptible but not yet infected $S(t)$, those that are currently infected $I(t)$, and those who have recovered from the disease and are considered immune to it $R(t)$. In addition, the SIR model also takes into consideration the following parameters: the total population N, the number of already infected people I0, the number of people who have already been recovered R0, the reproductive rate of the virus r0, the contact rate of the disease β, the time period of the spread of the virus t, and the recovery rate of the virus (in days) γ [7]. According to Smith on WebMD, COVID-19 takes on average 14 days to fully recover from the virus [4], which was the γ value for this project. Multiplying γ by the reproductive rate of the virus r0 resulted in the contact rate of the disease β.

Taking into consideration of all the parameters of the model, the SIR model then treats the virus as a differential equation with the following equations with the initial values of the number of infected people and the number of recovered people in the population [7, 8]

$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Using a script provided by the scipy package, the script will take all the parameters and the initial values, integrate the equations over the given time period, and return three lists [7]. The three lists correspond to the number of susceptible individuals, number of infected individuals, and the number of fully recovered individuals. Each list will be t elements long, where each element represents the corresponding information on the tth day. Although this script returns three lists, then only list that I used in this project was the estimated number of cases I.

I used the reproductive rate of the virus to determine how quickly the virus is spreading during a certain time frame in each state. This rate, also called the basic reproduction ratio, is defined as the average number of secondary infections that can occur when one infective is introduced into a completely susceptible population [9]. Previous studies have shown that when social distancing, this rate can be as low as 1.5 [10] but can be as high as 5.7 when there is no social distancing [11].

When separating the data of each state into two sets, one that holds the data prior to the separation date, and the other holding the data after the separation date, I made sure to set the correct parameters for both sets. It is important to note that the number of initially infected and recovered people will differ between these two sets, as they are taking data from different time points. Due to this, when modelling the number of cases with the SIR model, the last estimation before the separation date does not always match with the first estimation after the separation date, since the parameters are being updated. To compare the spread of the virus, I compared the calculated rate of reproduction between the two data sets to see if the SAHO had any effect on the virus' reproduction rate. If the reproduction rate decreased after the separation date, then I concluded that the SAHO could have had a positive effect on the virus' spread.

To determine the reproductive rate of the virus before and after the separation date, I used the following process. I first divided the data for before the separation date and after the separation date. For both data sets, I went through a for loop that iterated through every decimal to the thousandths place from 0 to 6 (the range of the reproductive rate of the virus). For each decimal, I tested that reproductive rate with all the other defined parameters into the SIR model. I took the returned list of infected individuals from the SIR model and compared each day's estimates from the model to the reported cases by finding the square difference between the two. I then found the average of all the squared differences for each day and compared this average with all the reproductive rates. The following process is illustrated in the algorithm below. I determined that reproductive rate that returned the lowest average squared difference was the best approximation for the given data.

Algorithm 1: Finding the Reproductive Rate

Input: Data set of reported cases RC
Result: Reproductive rate of data $r0$
Function testSIR($r0_{test}$):
   | $I \leftarrow SIR(N, I0, R0, \gamma, r0_{test})$
   | **for** $i \leftarrow 0$ **to** $len(I)$ **do**
   | | $sqDiff \leftarrow sqDiff + (I[i] - RC[i])^2$
   | **end**
   | **return** sqDiff
Function Main:
   | r0 = 0
   | **for** $i \leftarrow 0$ **to** *1200* **do**
   | | **if** *testSIR(i * .005) < testSIR(r0)* **then**
   | | | *r0 = i * .005*
   | | **else**
   | **end**
   | **return** *r0*
---

3 Results

In this section, I will outline the results I obtained from my polynomial least square approximations and the epidemiological mathematical models. There will be both tables and figures to help visualize the results for both sections.

3.1 Polynomial Least Square Results

To determine the effectiveness of the SAHO for each state, I took the separated data from before the separation date and after the separation date. For each data set, I generated the cubic least square polynomial to best fit the actual data. I plotted both the cubic least square polynomial and its derivative with the actual data to ensure that the equation was fitting the actual data correctly.

All the states' plots resulted in a similar pattern as the figure below (Figure 2). The cubic least square polynomial passes through or stays between the majority of the points, even if random spikes/noise is apparent in the actual data.

From the generated cubic least square polynomial, I found the average derivative and second derivative before the separation date and after the separation date. I recorded the results in the table below.

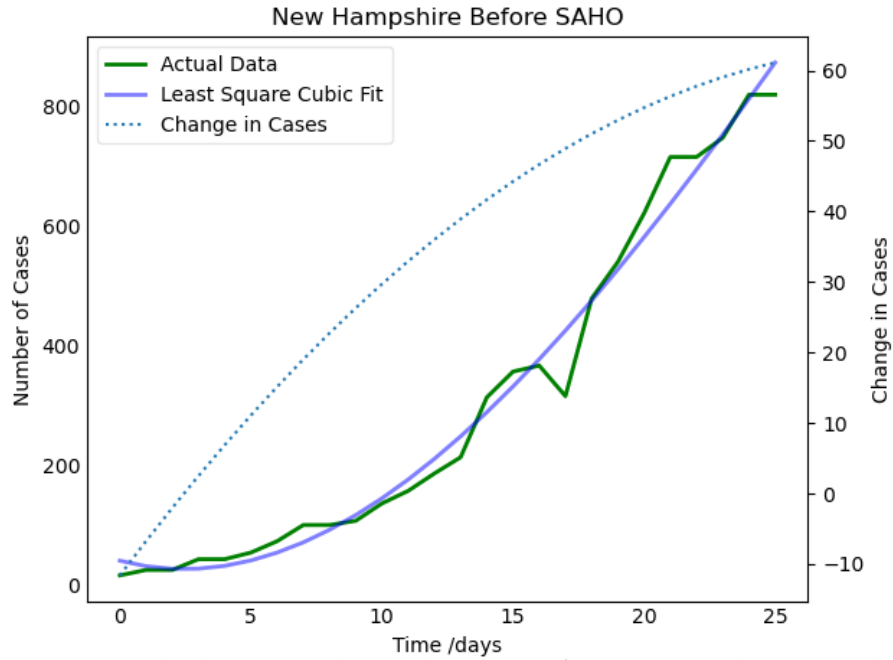# Figure 2: Accurate Fitting of Cubic Least Square Polynomial on New Hampshire

Table 1: Estimated Change and Acceleration of Cases in Relation with SAHOs

| State	Before SAHO f'(t)	Before SAHO f''(t)	After SAHO f'(t)	After SAHO f''(t)	State	Before SAHO f'(t)	Before SAHO f''(t)	After SAHO f'(t)	After SAHO f''(t)
AL	181.4620	7.8538	213.3505	-2.1195	MS	144.6009	6.2557	249.8299	3.0325
AK	12.9298	-0.0809	4.9572	-0.2796	MT	18.3372	-0.4417	3.5654	-0.3994
AZ	67.6626	4.0437	259.8588	8.1620	NV	112.3979	3.6881	131.4111	-1.4360
CA	203.8234	13.8004	1419.3478	26.6502	NH	32.9080	2.9084	74.1175	1.3490
CO	266.3411	13.8797	442.2707	13.8231	NJ	2076.6195	271.6442	3150.8873	-35.6184
CT	405.8781	44.2079	844.0956	-12.1129	NM	52.0202	7.0618	115.1020	4.7800
DE	51.4202	6.1160	165.9432	4.8553	NY	6091.9177	483.0776	7013.4108	-273.0060
FL	723.1250	38.0144	717.7023	-19.6739	NC	176.9250	11.2464	360.7474	10.9300
GA	527.4292	32.2547	795.1510	-6.0950	OH	298.6579	20.3489	532.3795	2.0792
HI	20.2673	1.3164	8.1184	-0.8480	OK	83.2848	8.0268	95.4107	-0.1911
ID	66.9105	3.0872	33.2570	-1.0409	OR	44.3616	3.8262	60.7899	-0.6954
IL	185.9866	12.0277	1796.6038	55.1085	PA	996.4760	75.1134	1384.8175	-2.9002
IN	298.4370	32.6033	563.0686	15.7342	RI	82.9309	8.8166	327.1861	0.0001
KS	60.4512	2.9423	190.9609	13.5586	SC	135.6944	1.8627	177.0519	1.3939
KY	61.0190	3.3154	146.1222	1.5411	TN	214.4467	4.9213	386.9085	21.3807
LA	979.5421	147.8623	514.8543	-19.8603	TX	364.8096	26.9538	879.6641	27.1618
ME	30.1259	0.0818	21.4884	0.8673	VT	33.5445	1.6496	10.9726	-0.7462
MD	377.1353	37.1091	834.6115	26.0994	VA	279.0997	22.7176	675.3872	22.5924
MA	319.3456	22.6734	2029.7162	25.4211	WA	145.0267	8.7862	244.5190	-1.7542
MI	1200.0598	101.6340	962.8497	-9.5896	WV	27.9902	2.0570	27.7139	-0.6932
MN	58.0850	2.3168	235.2772	21.1546	WI	60.4058	4.6459	205.9685	9.9698
US	17569.8919	1502.2259	28307.4458	-71.4150					

Looking at Table 1, the overall average change in cases in the US actually increased even after the SAHO were issued. States such as California, Illinois, Massachusetts, New Jersey, New York, Pennsylvania, Texas, and Virginia all experienced large increases in the change in cases statewide. A few states, such as Louisiana, Minnesota, Montana, and Vermont did experience an immediate decrease in the number of cases. In fact, most of the states still experienced the change in cases to increase after the SAHO was issued.

When comparing the rates of the changes of the number of cases, the US is experiencing a decrease in the case growth rate (as the negative second derivative indicates this); however, only 19 of the 42 states tested experienced this. The largest change in this measurement was New York, while states such as Illinois and California only experienced a quicker growth in the number of cases.

An example of a state that has experienced a heavy decrease in the rising of increased cases is New York. New York was the state that was hit the hardest in the nation by the COVID-19 outbreak, and as a countermeasure, instilled strict SAHOs. Even though New York is known to be one of the most densely populated areas in the world, its streets and attractions are currently empty. This effect can be seen in the data. The blue solid line represents the generated cubic least square polynomial and the green line (which happens to be under the blue line) represents the number of reported cases. The blue dotted line represents the change (or the derivative of the cubic polynomial) in the number of cases per day. The general trajectory of this dotted lines shows whether the state is experiences a slower rate of increase or not. As you can see, the change in cases is steadily increasing before the SAHO; however, after the state ordered its SAHO, the rise in cases has decreased dramatically. This is also illustrated in the table above, as New York's second derivative after the SAHO is negative (Figure 3).

Figure 3: Changes in Increase in Cases Before and After SAHO in New York
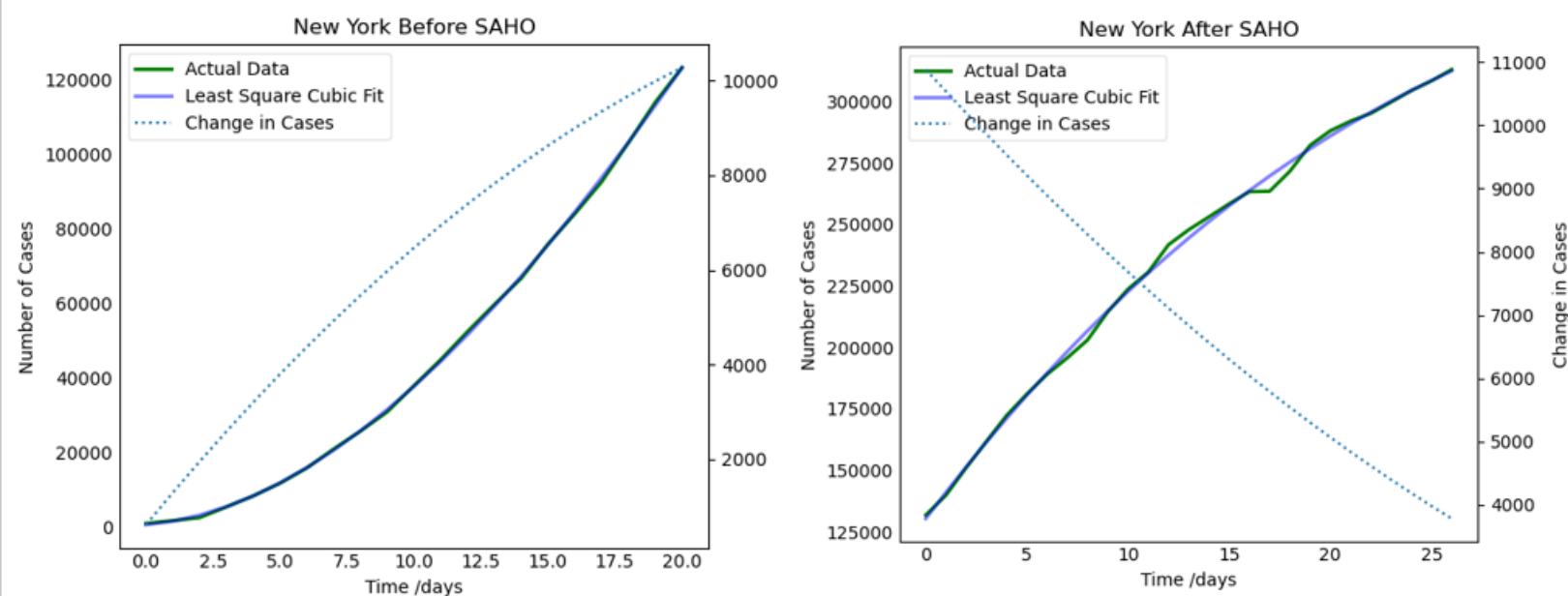
Although New York has shown the largest positive effect of the SAHO, other states, such as Illinois, seem to show the opposite effect (Figure 4). From Table 1, we can see that both the number of increased cases and the rate at which these cases are increasing has increased dramatically. Illinois' rapid case growth rate has only increased more rapidly, even after the SAHO was issued.

Figure 4: Changes in Increase in Cases Before and After SAHO in Illinois
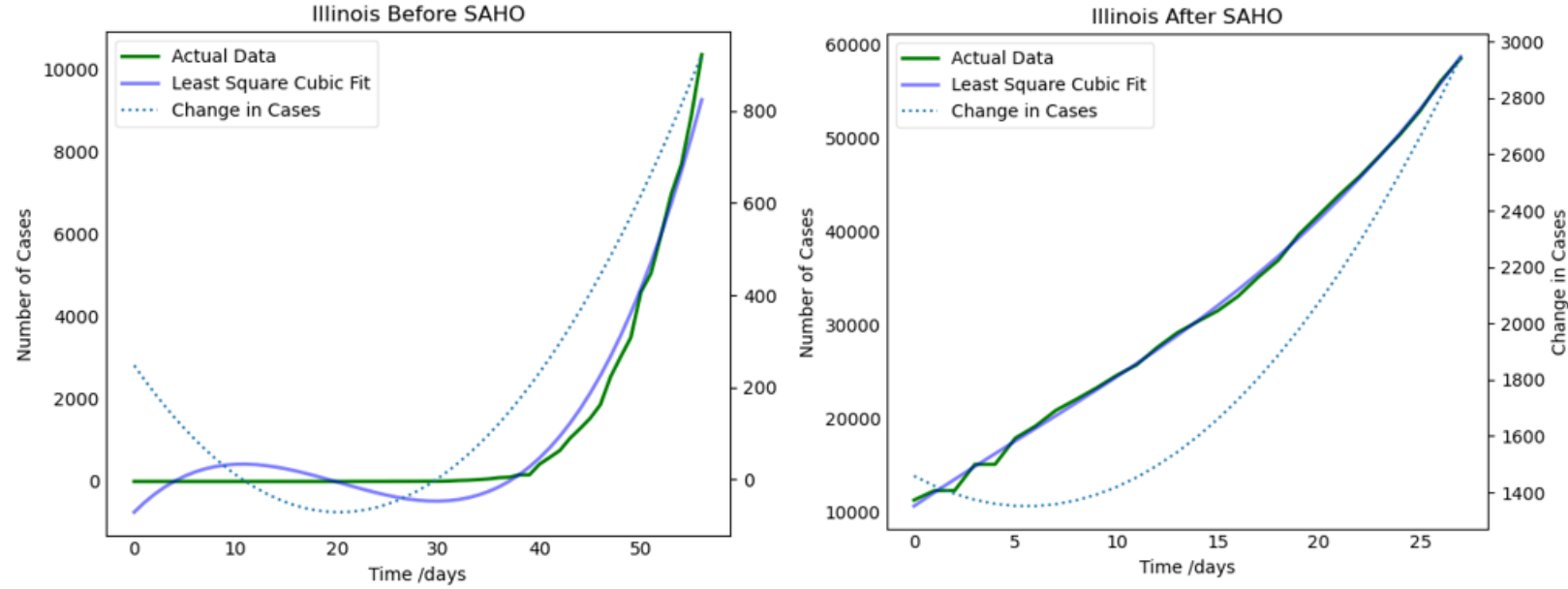
Although some states showed a positive impact from the SAHO, such as New York, there are some states that showed a conflicting effect, such as Illinois. I then turned to the SIR epidemiological model to determine if this kind of finding was consistent or not, as the SIR model is geared more towards these kinds of outbreak data and takes parameters outside of the number of cases into consideration.

3.2 Epidemiological Model Results

When running the algorithm illustrated in the section before for each state, I get the following results. As one can see from Table 2, it is quite evident that the reproduction rate of the virus has decreased for all states after the SAHO was issued.

Table 2: Estimated Reproduction Rates of Virus in Relation with SAHOs

State	Before SAHO	After SAHO	State	Before SAHO	After SAHO
AL	2.365	1.465	MS	2.465	1.62
AK	2.565	1.26	MT	2.63	1.155
AZ	2.81	1.585	NV	2.9	1.425
CA	2.935	1.785	NH	3.225	1.68
CO	2.995	1.66	NJ	4.275	1.725
CT	4.055	1.86	NM	3.04	1.86
DE	3.38	2.005	NY	4.405	1.57
FL	3.285	1.335	NC	3.29	1.62
GA	3.2	1.455	OH	3.175	1.84
HI	2.935	1.295	OK	3.73	1.555
ID	3.585	1.335	OR	3.17	1.515
IL	3.105	1.885	PA	3.47	1.54
IN	4.205	1.715	RI	3.395	1.98
KS	2.995	1.875	SC	2.565	1.465
KY	3.35	1.85	TN	3.035	1.525
LA	3.46	1.415	TX	3.82	1.5
ME	2.2	1.3	VT	3.05	1.265
MD	3.39	1.73	VA	3.11	1.79
MA	3.59	1.885	WA	3.09	1.335
MI	3.325	1.495	WV	3.855	1.65
MN	2.65	1.875	WI	3.285	1.575
US	3.222857	1.60131			

When looking at a graphical representation of the growth of the virus (Figure 5), we can see the difference in growth of the number of cases after the separation date. The figure on the left indicates the cases before the SAHO, and the figure on the right displays the cases after the SAHO. The figure on the left plots all the number of cases (in blue) before the separation date along with the predicted SIR model (in red). For Louisiana, the estimated reproduction rate was 3.46 before the separation date. On the right, the figure plots the number of cases after the separation date (in blue) with the SIR model with updated parameters (in green). This SIR model's parameters were updated to match the initial values after the separation date and a calculated estimated reproductive rate of 1.415.

The red line in the figure on the right is the SIR model with the updated parameters while keeping the reproduction rate prior to the separation date to estimate the growth of cases had the SAHO not been issued. As one can see, had the reproduction rate stayed the same, the number of cases would have been dramatically higher than the actual number of reported cases.

Figure 5: Number of Cases in Louisiana in Relation to the SAHO
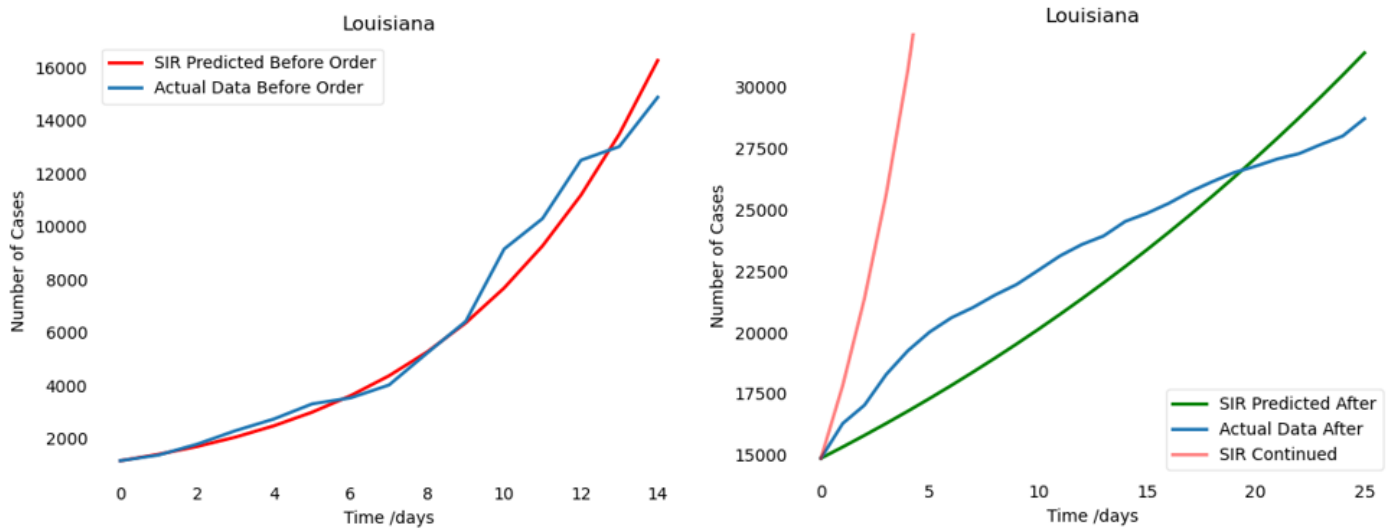
Applying the SIR model to all the states that have issued a SAHO have all manifested similar results to that of Louisiana. The reproductive rate was shown to have seen a decrease after the SAHO was issued (Table 2). Although the magnitude of cases will differ between states, the overall structure and case growth rate that have issued SAHO are similar. Before the SAHO was issued, each state was shown to have a exponential growth in the number of cases. After two weeks had passed from the issuing of the SAHO, each state saw a decrease in growth of the number of cases (although it is important to note that the number of cases were still increasing). Each state's graph indicated that had the reproductive rate was kept from prior to the issuing of the SAHO, the number of predicted cases would have been greatly larger than the actual number of reported cases.

4 Discussion and Conclusions

Overall, I cannot decisively conclude that there is a positive effect from the issuing of the SAHO just by looking at the cubic least square polynomial. The results from the cubic least square polynomial indicate that the majority of the states are still experiencing rising numbers of cases, and in some states, the case growth rate is still increasing. The increasing cases and the rate of increases in states such as Illinois do counter the intended purpose of the SAHO, which was to minimize the case growth rate. The nation as a whole is still seeing an increase in the number of cases, despite the issuing of SAHO in the majority of the states.

Even though the cubic least square polynomial can raise some doubt in the effectiveness of the SAHO, I do still believe it is important to highlight the findings of the SIR model. The cubic model only takes into consideration the number of cases and no other parameter. This is certainly not the case in the real world, as the virus depends on many more factors than the just

the number of infected people. The SIR model is more accurate in predicting the impact of the SAHO in each state due to its ability to utilize different parameters a virus' spread depends on in reality. Parameters such as recovery rates, number of people already recovered, the contact rate, and the reproductive rate of the virus are all important to consider, not just the number of cases.

When analyzing the results of the SIR model, it is clear that all the states have seen a positive impact from the SAHO. The reproductive rate of the virus have all gone down, and in some states significantly, after the SAHO was issued. With the reproductive rate of the virus decreasing, even though the virus still has the ability to infect people, its ability to reproduce and keep infecting people will decrease. This is important to consider, as the number of interactions can affect this reproductive rate. The SAHO are meant to control and lower the number of interactions we have to minimize and lower this reproductive rate. Due to this, even though states are still experiencing an increase in cases (as illustrated by the cubic model), the unanimous decrease in the rate of reproduction of the virus is enough for me to conclude that the SAHO are effective in controlling the increase of the spread of COVID-19.

References

[1] A. Lee, "These states have implemented stay-at-home orders. here's what that means for you," Apr 2020.

[2] CSSEGISandData, "Cssegisanddata/covid-19," May 2020.

[3] "United states population (live)."

[4] A. Cassoobhoy, "Coronavirus and covid-19: What you should know," May 2020.

[5] B. Nazario, "Coronavirus (covid-19) testing," May 2020.

[6] R. L. Burden, J. D. Faires, and A. M. Burden, *Numerical analysis*. Cengage Learning, 9th ed., 2016.

[7] "The sir epidemic model."

[8] H. W. Hethcote, "Three basic epidemiological models," *Applied Mathematical Ecology Biomathematics*, p. 119–144, 1989.

[9] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, p. 599–653, Aug 2006.

[10] J. R. Koo, A. R. Cook, M. Park, Y. Sun, H. Sun, J. T. Lim, C. Tam, and B. L. Dickens, "Interventions to mitigate early spread of sars-cov-2 in singapore: a modelling study," *The Lancet Infectious Diseases*, Mar 2020.

[11] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, "High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2," *Emerging Infectious Diseases*, vol. 26, no. 7, 2020.