Maximum Likelihood Estimation for the Policy-Guided Susceptible-Infected-Recovered Model of the COVID-19 Cases in Texas

David Han, Ph.D., Ethan Woods
The University of Texas at San Antonio, 1 UTSA Circle, San Antonio, TX 78249

Abstract
The goal of this research is to create a Susceptible-Infected-Recovered (SIR) model for the Texas COVID-19 cases based on the state data from March of 2020 through October of 2020, and to investigate the impact of public policies on the transmission of COVID-19. The data was pre-processed using Excel; some basic time series graphs were produced in Excel as well. All other data analysis, including the production of all graphs relating to the SIR model, was performed in the open-source statistical programming language R. Difficulty in estimating the model parameters by the maximum likelihood method was encountered due to the short durations between the implementation dates of various policies designed to curb the spread of COVID-19. Examining the estimate trends of beta, gamma, and R_0, a stabilizing pattern for R_0 was observed over time, which would require further investigations to understand the epidemiology of COVID-19 in Texas.

Key Words: COVID-19 pandemic, epidemiology, maximum likelihood estimation, susceptible-infected-recovered model

1. Introduction & Background

SARS-CoV-2 is the type of coronavirus that causes the COVID-19 disease. The airborne virus is thought to have emerged in Wuhan, China sometime near December of 2019. Rather than staying confined to Wuhan or the Hubei Province more generally, this virulent coronavirus soon spread to nearly every territory in the world. Nearly all countries have instituted lockdowns or other safety measures such as mandatory mask policies and social distancing requirements. The virus (and the response of governments to the virus) has dramatically impacted various sectors of society: food, entertainment, employment, education, transportation, and more. Yet the true impact on society transcends mere inconveniences. Millions of lives have been lost, including more than 500,000 in the U.S. and roughly 50,000 in Texas alone. The severity of COVID-19 cases can be quite unpredictable indeed; some who contract the virus have no symptoms while others suffer from acute respiratory distress syndrome (ARDS). Given the gravity of the COVID-19 pandemic and the importance of maintaining good public health, it is salient that we be able to model epidemiological trends as they relate to COVID and determine the efficacy of Texas policies designed to curb virus transmission. In this work, we aim to produce a Susceptible-Infected-Recovered (SIR) model of COVID-19 for the state of Texas and to examine the effectiveness of various state policies on the spread of COVID-19.

2. Methods

The data on COVID death and transmission statistics were gathered from https://covidtracking.com/data/#state-tx. The Texas policy datasets were obtained from https://www.openicpsr.org/openicpsr/project/119446 and combined with the transmission statistics into a single spreadsheet. The data were then pre-processed in Excel; columns were renamed for brevity, empty columns were deleted, and other columns were combined to avoid redundancy where possible; see Figure 1. Basic visualizations of the COVID-19 cases, hospitalizations, and deaths over time were also created. The remainder of the analysis was conducted in the open-source statistical programming language R.

	A	B	C	D	E	F	G	H
1	date	Column #	times	state	dataQualityGrade	death	deathConfirmed	deathIncrease
2	3/4/2020	1.00	0	TX				0
3	3/5/2020	2.00	1	TX				0
4	3/6/2020	3.00	2	TX				0
5	3/7/2020	4.00	3	TX				0

Figure 1: A data snapshot from the Excel file is shown above.
The entire set of data formed a 119 x 221 array.

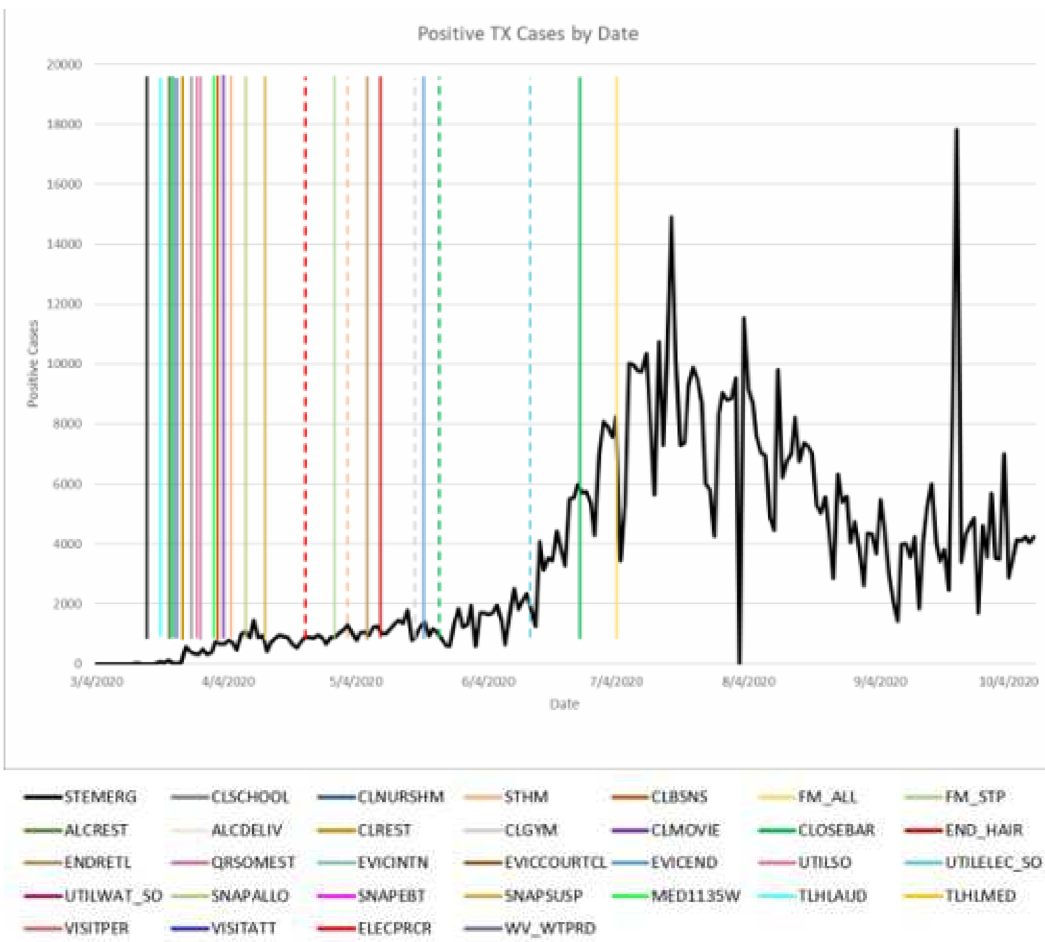

Figure 2: Cases of COVID-19 in Texas were plotted over a period of 7 months.

3. Computational Results

Figure 2 above shows the COVID-19 cases in Texas over a period of 7 months, from March through October of 2020. Vertical lines represent the implementation and expiration of various policies. Using the principle of the maximum likelihood estimation, the SIR models were fitted piecewise to understand the evolution of the model parameters after implementation of each policy, hoping to control the spread of COVID-19. Figures 3 to 6 below represent the evolution of the estimated model parameters.

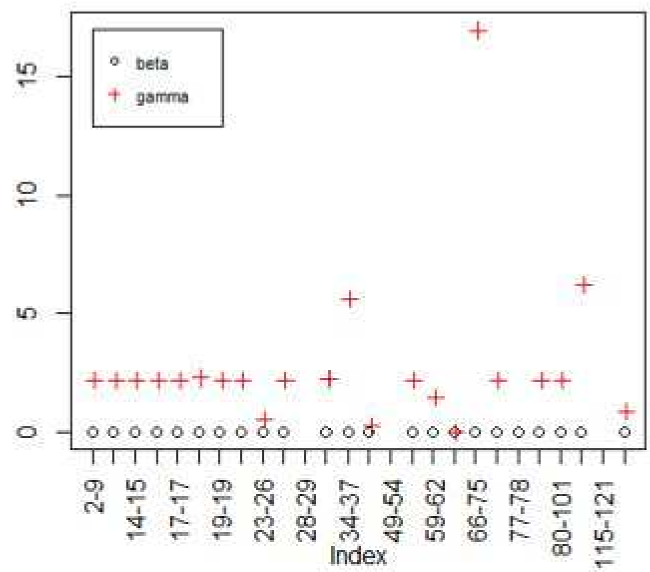

Figure 3: Point estimates of beta and gamma were plotted for each time period between Texas policy alterations.

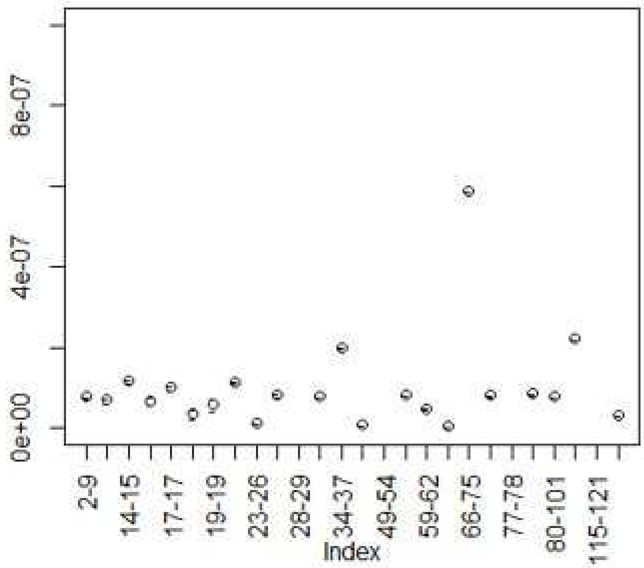

Figure 4: Point estimates of beta over time are given, along with 95% confidence intervals.

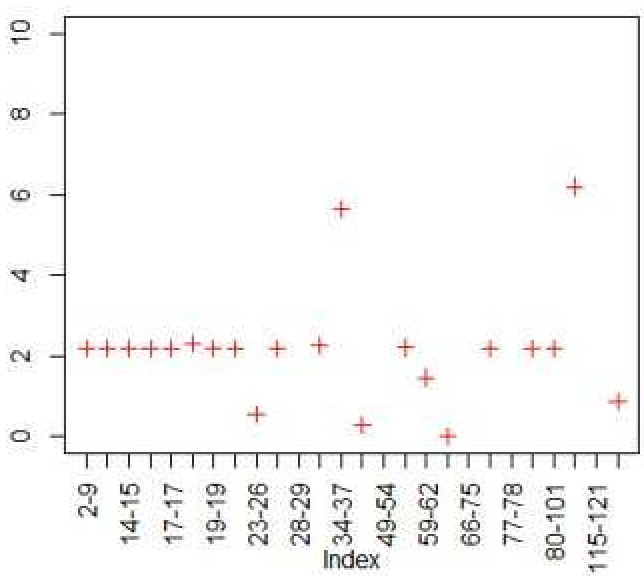

Figure 5: Point estimates of gamma over time are given. It was not possible to estimate the confidence limits for gamma.

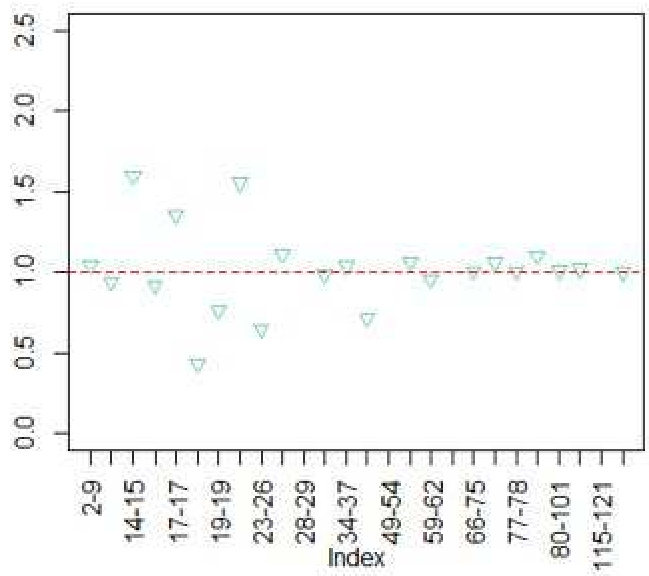

Figure 6: Point estimates of R_0 were produced for each time period. Note the stabilization of R_0 values over time.

4. Concluding Remarks

There were limitations in sample sizes in each policy period that prevented the estimation of confidence intervals for the model parameter gamma. The global estimate had to be used as an initial point in each period. It was difficult to determine the effect of any given policy due to the small number of days between changes in Texas state policy. Still, the values of the basic reproductive rate R_0 stabilized over time to near 1.0 after roughly one month had passed; this phenomenon is worthy of further investigation.

References

Raifman, Julia, Nocka, Kristen, Jones, David, Bor, Jacob, Lipson, Sarah, et al. COVID-19 US State Policy Database. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-09-28. https://doi.org/10.3886/E119446V33

Shi, Y., Wang, G., Cai, Xp. et al. An overview of COVID-19. J. Zhejiang Univ. Sci. B 21, 343–360 (2020). https://doi.org/10.1631/jzus.B2000083

The Atlantic. (n.d.). The Data. The COVID Tracking Project. Retrieved October 11, 2020, from https://covidtracking.com/data/#state-tx