INVESTIGATING U.S. CORONAVIRUS TESTING ACROSS 50 STATES IN MARCH-MAY 2020

Authors: Andriy Lunin, Jaisal Friedman, Mikołaj Debicki
New York University Abu Dhabi, Saadiyat Island, PO Box 129188, Abu Dhabi, UAE.

Advisors: Talal Rahwan, Bedoor AlShebli

Abstract: The modern world was certainly not ready to face the outburst of the current pandemic, which resulted in severe unnecessary casualties, an economic collapse, and irrational governmental responses all across the globe. In the scope of these atrocious events, we have decided to investigate the apparent as well as the not-so-obvious relationships between the epidemiological parameters in order to make predictions about the development of the virus, uncover the hidden biases, and evaluate which measures are genuinely effective for mitigating the disease. By investigating the spread of the N-COVID-19 pathogen in the U.S., we find corroborating evidence that most state authorities' effort to confine the disease was ineffective. Furthermore, we uncover the complex relationship between testing, positive cases, and deaths. By employing panel regression analysis of the empirical data, we build a model that is free of prevalent external biases between states and over time. In summary, we have built an interactive COVID-19 map, fitted the logistic function curve for the U.S. data, proposed a model for the future spread of the pandemic, and uncovered the significant correlations in the datasets.

1. Introduction and Literature Review

The emergence of the Novel Coronavirus (SARS-CoV-2) has already had a drastic impact on all of our lives. Motivated by the current havoc, we postulated that understanding the spatio-temporal dynamics of the pathogen is essential for devising a proper response and combatting the pandemic [1]. Therefore, in our exploratory research, we investigate the relationships between the epidemiological parameters, the effects of governmental lockdowns, and the spatial dynamics of the SARS-CoV-2 on the example of the United States testing data — due to the nature of the erratic spikes in the case numbers, inadequate governmental response, and the physical lack of tests, U.S. data was especially compelling for us to explore. Our primary aim was to gather vital, time-crucial information that would aid the process of the mitigation of the current crisis.

Coronavirus has already been perfunctorily described in the hodiernal scientific literature. Notably, the U.S. was slow to react to the propagation of the pandemic, and the first cases were discovered only when significant efforts were made. In the following sections, we are going to investigate the correlations between the epidemiological factors, and the effects of the testing spikes on the growth of new N-CoV-19 cases.

There has been some extensive research done on the topic of disease propagation in the past. For instance, in their paper, King and Zeng elaborated on how logistic regressions may be used for estimating and modeling low probability events [2]. This paper inspired us to fit a mechanistic logistic curve into our data and explore how the Coronavirus curve stands compared to the model. Similar research was performed on the H1N1 influenza pandemic of 2009 datasets by independent European and Mexican research teams [3].

At the same time, the data from the 2014 article about H1N1 influenza reinforces the two wave nature of pandemics and further strengthens our hypotheses for the behavior of the Novel Coronavirus. The current spread of the virus is generally geographically heterogeneous, with significant hubs being the places of most infections (New York, Wuhan). However, the seemingly inevitable coming of the second wave in the Fall could pose another prominent threat and lead to different, more severe developments of the infection [4].

Conclusively, comprehending the impact of governmental interventions is essential for preventing the atrocities of the second wave, as well as mitigating the impact of the current one. As the article suggests [5], the effects of correct measures have a notable impact on the short-time forecasting on the data obtained in Germany. In our study, we build upon that by focusing on the U.S. datasets and computing regressions regarding the spread of the Coronavirus with the primary goal of future forecasting and evaluating the efficiency and quality of conducted tests.

Ultimately, we hypothesized that the relationship between the ratio of tests to positive cases over time and between the ratio of deaths to positive cases with a 14-day lag will start out as an exponential relation, yet will fade out into logarithmic growth, and eventually converge to an asymptote. We predicted that this inflection point will not occur in the 8 weeks we are studying for either positive cases or deaths (March-May).

2. Methodology

2.1. Methodological Approach

Due to the unavailability of reliable data, our research team was forced to employ both qualitative as well as

quantitative methodological approaches, with more attention towards the former. From the quantitative side, we strived to derive a generalizable formula for the proliferation of COVID-19. We made sure to filter and engineer the data (Section 3) in such a way that our results were reliable and could be replicated by other research teams that use different datasets. Consequently, we employed deductive reasoning and attempted to qualitatively explain the patterns that we came across while researching the data. This mixed approach allowed us to explore the patterns of how the coronavirus spreads, as well as derive appropriate formulae to model the further development of the pandemic.

2.2. Data Collection

In order to gain a better insight into the patterns of the pandemic proliferation, we began with gathering, parsing, and filtering the data. At first, we manually scraped the daily updates from such websites as John Hopkins COVID-19 [6] and NYTimes COVID-19 [7]; however, we soon found a reliable data source [8] that not only provides reputable state-specific data but also gives it a quality-grade. The Atlantic [8] project allowed us to reach our goal of gathering temporal state-by-state testing and case data and run a comprehensive analysis on it. More information can be found in Section 3 below.

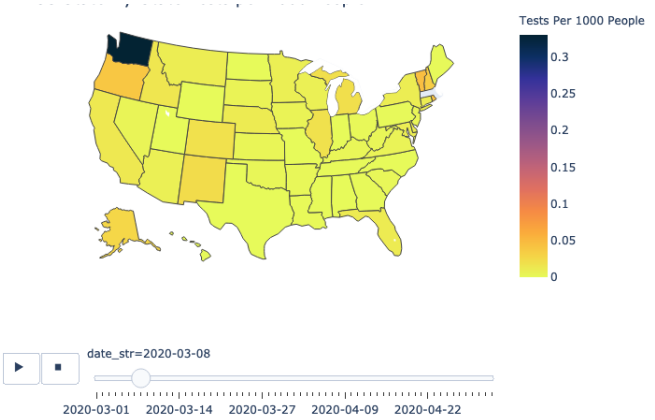

FIG. 1: March 8th Tests per 1000 People

2.3. Methods of Analysis

Data visualization is a significant part of our analysis. Firstly, we visualized the spread of the virus by building an interactive temporal map of the US state-by-state over tests per 1000 people performed Figure 1 and 2. Secondly, we plotted total tests and positive cases and have shown that the total number of positive cases grows in a logistic fashion (see Figure 3). We then performed a similar logistic analysis with fourteen-day lagged positive cases and total deaths. Having the two estimated asymptotes for the total cases and total deaths, we approximated the case fatality rate; comparing this with the estimated case fatality rate from the literature, we

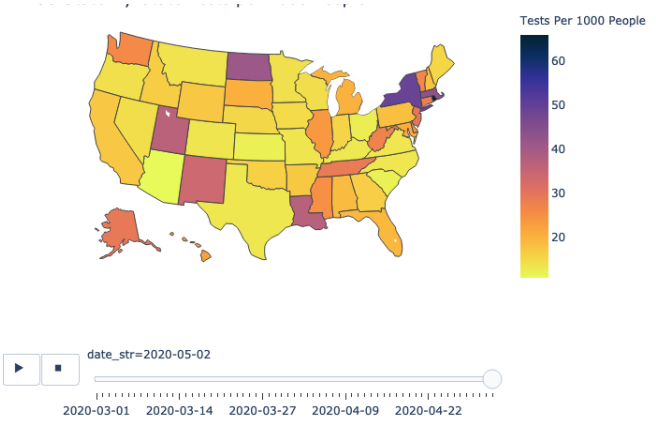

FIG. 2: May 2nd Tests per 1000 People

were able to evaluate how comprehensive the current testing is (see Figure 4 and 5. We did also provide an estimate of the total number of positive cases barring undetected cases. Next, we investigated the auto and cross-correlations between daily positive cases, daily deaths, and daily tests. Lastly, we performed a multivariate and panel regressions to investigate the significant factors correlated with daily increased positive cases. We employed the technique of panel regressions to remove unobserved time and state factors. Consequently, we observed that certain significant factors from the multivariate regression are absorbed in the panel regression. We could then assert that the remaining factors are truly significant regardless of the time-dependency and state differences in the outbreak.

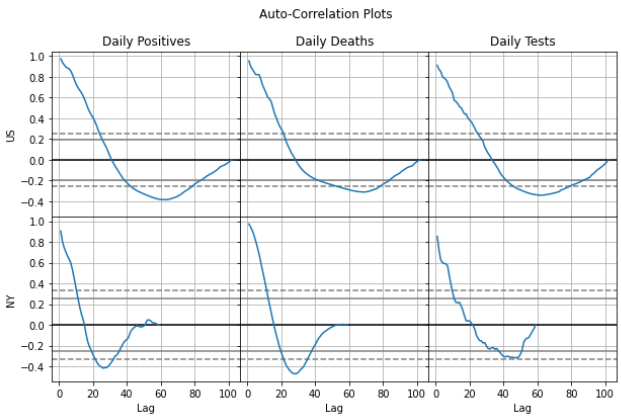

FIG. 3: Auto-Correlation of Daily Tests, Cases, and Deaths

2.4. Evaluation of Choices

For accurate estimations and process modeling, we required more extensive and rigorous datasets than the ones offered by the US Department of Healthcare of the Center for Disease Control [9]. The caveat of using public datasets lies in quality and reliability. The dataset

we picked, nevertheless, was deemed reputable by such sources as NYTimes, CNN, Wall Street Journal, etc. [8]. Additionally, we did spend extra time filtering and cleaning up the data tables. We have decided not to remove the outliers, as in our particular case, they provide valuable insight and allow for a better understanding of the growth patterns. Despite its disadvantages, this approach led us to textbook-precise estimations and curves that bolster our initial hypothesis as well as similar research conducted about other pandemics or analogous processes.

Nonetheless, it has to be admitted that our initial stance on the direction of the research — our hypothesis, as well as the subsequent analysis — operated under an assumption that the data is accurate. Undeniably, it was a strong postulate, yet we would have to reiterate that the primary goal of our research was to explore how accurate numbers and models can be predicted from biased and disposed data. That being said, the potential flaws in the data allowed us to pinpoint the testing defects, the inadequacy of the governmental response, and the practical impact from the lockdown.

Finally, we have chosen to visualize our analysis with various graphs and charts. Although it may be argued that graphed data loses its exactness, in our case, the visual tool does an excellent job of conveying the overall idea and letting the reader grasp the trends and tendencies before they are mentioned and discussed.

3. Data

3.1. Utilized Data

The scope of our analysis incorporated three data sources. The first and the most significant dataset came from The COVID Tracking Project spearheaded by The Atlantic. By the means of web-scraping, physical requests, and aggregations, The Atlantic computes daily online statistics for each state. They publish daily the *positive tests, negative tests, total tests, total deaths, total recovered, total pending tests, as well as the daily positive tests, negative tests, tests, deaths, and recoveries.* It includes 56 states and U.S. territories; however, some states' (e.g., South Dakota New Mexico) data was of rather poor quality and did not make it into our analysis.

The second dataset we used was the U.S. 2019 Census estimates for the state-level population. Here, we merged the population dataset to normalize testing by the state population. It is worth noting that the Census dataset did not include such U.S. territories as Gaum, Virgin Islands, Northern Mariana Islands, or American Samoa. These data points had to be collected individually from Google estimates.

The third data set we used was the New York times state-by-state stay-at-home orders published article, which specifies if and when each state announced stay-at-home orders. We compiled this article into a data-set with key-value pairs {state_abbreviation: stay_at_home_date}. We published this dataset through GitHub, allowing other researchers to take advantage of it [10]. For all of our analyses, we considered one positive test to correspond with one positive case. In reality, this may not be true as sometimes one patient may get multiple records and be administered two or more tests. However, given the shortage of tests in the U.S., it is a reasonable assumption that each patient that gets a positive test will not be re-tested and thus will not get double-counted.

3.2. Data Processing & Engineering

We engineered the data to include the following additional columns: *number of days since the hundredth case in each respective state, number of days since stay-at-home orders (lockdown), the moving average of the previous seven days of daily tests, the moving average of the previous seven days of new positive tests, the fourteen-day lag in the total number of positive cases, the weekday, and is a weekday.* The number of days since the hundredth case was calculated by creating a set of tuples with the key-value pairs {state_abbreviation: date_of_100th_case}; the date_of_100th_case is the date when the number of COVID cases reached a hundred in each respective state. This method was employed as it allows for time-based comparisons of the pandemic spreading between states with outbreaks occurring at different points in time. For example, Washington and New York had their hundredth case three weeks before North and South Dakota. The number of days since the hundredth case can, therefore, be negative to allow for a continuous variable.

The number of days since stay-at-home orders was set to zero by default unless stay-at-home orders had been announced for that state. We did not consider the removal of stay-at-home orders since the scope of our Analysis was March through May, and no states reversed their orders by that time. The moving average for the past seven days for both: daily numbers of tests and positive cases proffers a good indication of the derivative of the total testing and positive cases by state. While we used a non-weighted average, we suggest that further analysis could explore the effects of weighted or exponential moving averages.

Lastly, since, on average, the virus takes fourteen to seventeen days to kill ([11]), the variable for the fourteen-day lag of positive cases was introduced to investigate the relationship between the respective daily deaths and the number of total positive cases fourteen-days before.

4. Results & Discussion

4.1. Interactive State Map

We built an interactive state-by-state map to illustrate the relationship between the number of tests per thousand people and daily positive cases between March and

May in the U.S. We have provided movie animations through GitHub ([10]) as well as the code in the IPython notebooks or on the shared Google Colab Notebooks [12]. Figures 1 and 2 illustrate the test rate per thousand people on March 8th and May 2nd respectively.

4.2. The Auto-correlation Analysis

We ran a comprehensive analysis and investigated the auto and cross-correlation of daily cases, testing, and deaths.

We used the auto-correlation plots to see whether there is a time-based relationship between a variable and itself in the time-series data, which provided us with a baseline for interpreting cross-correlation. At this point, we expected to observe a pattern of dependent functions since more cases often beget more cases, tests beget tests, etc.

Figure 3 illustrates the auto-correlation plots with the x-axis denoting the number of days lagged to calculate the correlation. The first row provides the auto-correlation graphs for the whole country, while the second one illustrates New York specifically. This comparison was made to emphasize the differences and similarities of the patterns across the country and the state with the worst outbreak (the dotted and solid grey lines in each plot represent the 95 and 97.5 percent confidence interval). Accordingly, we observe a common trend across all plots: auto-correlation begins at one and decreases until there is a negative correlation, then it trends back towards zero, which is rather peculiar for such a process. For reference, a typical auto-correlation plot with no time-series based correlation would look like a sine or cosine wave. Furthermore, since we have chosen to look at daily cases, tests, and deaths, there should be no cumulative trend effect that may be observed when dealing with cumulative or total values. Notably, the auto-correlations in New York seem to reach a negative value with less of a lag. We speculate that this may be due to the massive spike in cases and the strict quick measures taken by the New York officials; the rest of the U.S. was notoriously slower to react. Finally, it makes sense that deaths are the slowest to become negatively correlated with a time lag due to the extra time needed to observe deaths: the infection takes 14-17 days to result in a fatality.

It is worth noting that a 60-day lag would consider relatively few points for a data-set with an average of 70-80 of time-series data points, which could have contributed to the seemingly influential trends observed in the auto-correlation. Henceforth, we do not make definitive any statements regarding the Figure 3 except for its peculiar nature and prospect for further investigations.

4.3. Multi-variate and Panel Regressions

Next, we ran Multi-variate and Panel regressions on the data-set to predict the number of daily positive cases. We created dummy variables for the state and weekday variables for both regressions. The model has a high R-squared of 0.91 Table 1, which explains most of the variance in the daily cases. *The number of the daily tests, the state population, the total number of tests, the days since the hundredth case, the number of pending tests, the daily number of deaths, the number of recovered cases, the moving average of the previous seven days of daily tests, the moving average of the previous seven days of new positive tests, California, Alabama, Florida, Illinois, Louisiana, Massachusetts, North Carolina, New Jersey, New York, Oregon, Texas, Monday, Tuesday, Thursday, Friday, Saturday, and Sunday* variables turned out to be significant, corroborating our initial claim that testing would have a significant impact on the number of daily positive cases throughout the whole duration of the pandemic.

Moreover, the moving average factors illustrate the significance of the change in the number of positive cases and tests in the past week. The days since the hundredth case is significant and has a positive coefficient. Hence, the moving averages and days since the hundredth case variables prove that the daily cases follow a time-framed outbreak. Additionally, it is worth noting that the pending tests and recovered cases, both having slightly negative coefficients, also turned out to be significant. Due to the coefficients being close to zero, we refrain from interpretations until further review of the panel regression results. We have observed that Monday, Tuesday, and Sunday report 43, 32, and 21 fewer cases, which can be explained by the testing lag due to the weekend. Since tests in the U.S. typically take 2-3 days to return results, the drop in the reported tests shifts to the drop in the tests over the weekend [13]. Analogously, the spike in the reported cases on Friday can be explained by the CDCs urge to report the cases before the weekend.

The states that are significant and have positive coefficients are correlated with states with bad outbreaks like New York, New Jersey, Massachusetts, Louisiana, and Illinois. States that are significant with negative coefficients are correlated with relatively few cases to the state population in Alabama, California, Florida, North Carolina, and Texas. Thus, this regression analysis acts as a proxy to understanding which states have successfully handled the pandemic and which did not.

Lastly, we observed that the days since the lockdown factor column is not significant. Though it is a striking revelation, we did expect it, as mentioned in Jonas Dehning's paper [5], the lockdown does not have any drastic impact on the spread of the pandemic unless it is inflicted at the exact moment of the first case identification, which is not the case for the U.S.

After executing the Multivariate regression, we ran a Panel regression to examine what is significant when time and state differences are removed. The panel regression accounts for unobserved factors across the States and time. For Table 2, the R-squared value is 0.7483, meaning that there are many unobserved factors between states that are not captured in the model. The weekday dummies and population variables are absorbed by the panel effects so that their impact is not observable across all of the states. The total number of daily tests, the

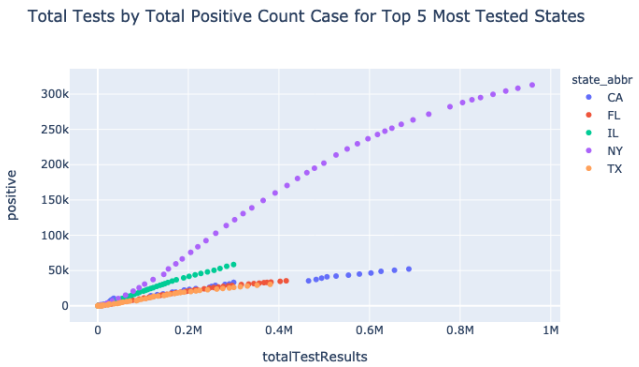

FIG. 4: Total Tests by Total Case for States with Most Tests as of May 2nd

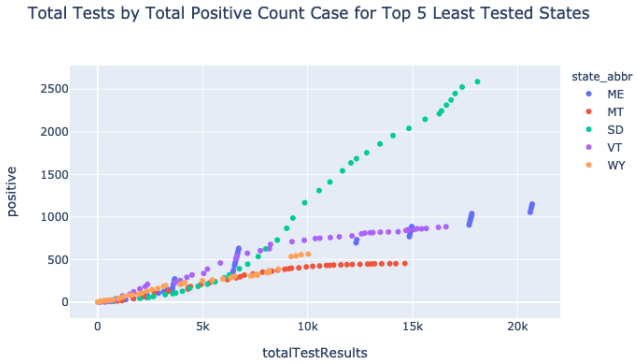

FIG. 5: Total Tests by Total Case for States with Least Tests as of May 2nd

total number of tests, the days since the hundredth case, pending cases, number of daily deaths, number of recovered cases, the moving average of the previous seven days of daily tests, and the moving average of the previous seven days of new positive tests were all significant factor columns. These results further prove the paper's hypothesis that testing and positive cases are correlated regardless of the time and state of the outbreak.

Once more, the days since the lockdown column was not significant. The absence of the correlation between the lockdown and the function of the positive tests further proves that the effects of the lockdown are not meaningful in the outbreak. We can further state that this may be because lockdowns are instigated proportionally to the number of cases and are thus too late per each state's outbreak.

As for the total number of pending tests, the total number of test results, and the number of recovered cases, they are all deemed significant, having a positive coefficient for the first, and negative coefficients for the other two relations. It further proves that there is an inverse relationship between daily positive cases and total test results and recovered cases. Investigating this relationship could be the subject of a follow-up research paper.

4.4 OLS Logistic Function Fitting

To conclude, we fit a logistic function to the data using OLS to estimate the L value in the equation of the pandemic proliferation in the U.S., New York, and some other states. The general formula for logistic processes is as follows:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

EQ. 1: Logistic Relation.
x_0 = value of the midpoint
L = maximum value
k = growth rate

As suggested by our hypothesis, the case count across all states blew up exponentially (see Figure 4 and 5 but with time it developed into a logarithmic curve and began converging to the asymptote L – which is the numerical limit to the number of cases that can ever be registered for the particular trend. Comparing the real-life patterns to the model allows for evaluating the quality of testing and the efficiency of the governmental response. For instance, the propagation of the pandemic in some states, like New York, Arizona, Washington, or Rhode Island, uniformly converges into our hypothesized curve. This observation provides ample grounds to conclude that these states are sufficiently tested, the L-value is accurate, and the number of active cases will likely get stagnated. In other words, the respective states' authorities efficiently reacted to the outbreak. Though the L-value is not by any means representative of the true number of cases, it is generally a good indicator of how well the disease is confined. That being said, the consideration of the states of South Dakota, Maine, etc. Figure 5 reveals the erratic behavior of the case count growth patterns, which only signifies that these states are severely under-tested, while pandemic confinement measures are poorly implemented.

Further analysis of the death rate is indicative of the fact that the COVID-19 mortality in the NY state averages at 6.7% (see Figure 7). At the moment, there is insufficient data about the pandemic mortality rate to make any conclusive statements; however, two idiosyncrasies have to be noted. Primarily, the graph evolves as a logarithmic function. It may be explained by the fortnight delay between the moment one gets infected, and the time the disease becomes fatal, as well as the fact that the notion of death it's not contagious, ergo the graph is primarily a function of the case count: the logarithmic transformation of the underlying logistic function. Secondly, the death rate seems to be higher than that in other first-world nations. The explanation of this phenomenon is out of the scope of our exploratory research,

yet we speculate that the inflated mortality rate has to do with obesity and the national healthcare system. For the reasons mentioned above, the high death rate cannot be solely blamed on poor testing.

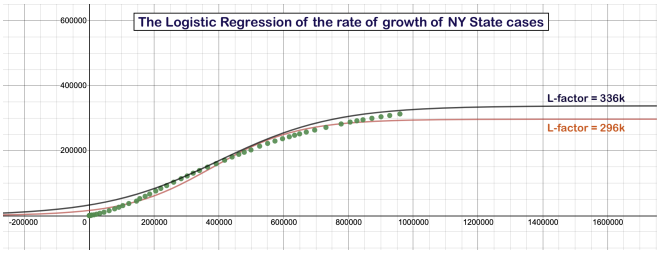

FIG. 6: OLS-Fitted Logistic Function of NY Cases

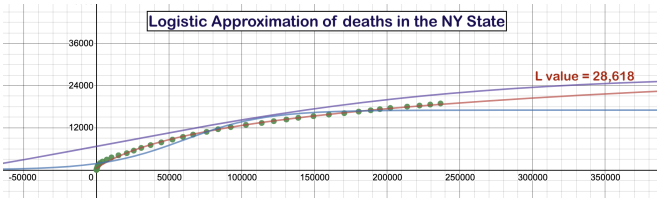

FIG. 7: OLS-Fitted Logistic Function of NY Deaths

5. Limitaitions

It shall be noted that our predictions and evaluations are based on expanding the premises that the current available N-COV-19 datasets are reliable and accurate, at least to some extent. Our findings are undoubtedly reproducible, but with marginal discrepancies and disparities. Additionally, we make no attempt to estimate how many individuals are sick: we only employ modeling and regression analysis to evaluate the testing quality and investigate the potential correlations in the researched data. We want to remind the readers that statistical and systematic errors are inevitable in any research of similar type due to the peculiarity of the datasets. Our work was further complicated by the ever-changing status quo, the uncertainty about the date of the first case, and the untimely governmental response that occurred after an apriori unknown delay.

While our model is meant to be employed for further studies of the N-COV-19, as well as similar pandemics, we urge the readers to be mindful of such varying factors as the ř (r_not - the contagion factor), the L-value, and the values, and adjust them to match their datasets.

6. Conclusion

In conclusion, we posit that our initial hypothesis proved itself correct. As expected, we did observe an exponential growth that soon converged to an asymptote. To our surprise, the inflection point did occur during the time of our research 4, and as a result, not only did we forecasted the development of the N-COV-19 pandemic in our proposal, but we also predicted its demise in our main paper.

Our regression analysis highlights the non importance of timely governmental lockdown irrespective of time, identifies the correlation between the number of conducted tests, pending tests, moving average of previous week's tests, moving average of previous week's positive cases, weekday effects, and days since the hundredth case and daily positive cases, and evaluates the efficacy of individual states response to the outbreak.

We also employed OLS-fitted Logistic Function and auto-correlation analysis to further explore and model the spatial-temporal dynamics of the novel pathogen. Our models let us estimate the L-values for the current outbreak, investigate the nature of the inflated death rate, algorithmically determine which States are poorly tested, and explain the erratic curve of the total cases in the United States.

7. Authors' Contributions

All three authors contributed to the original ideation and project proposal. Mikolaj Debicki performed the literature review and research needed to understand the topic at hand and the current trends in research. Andriy Lunin ideated the original hypothesis and the methodology for evaluating the hypothesis through OLS-fitted logistic functions on the data-set. Jaisal Friedman and Andriy Lunin performed the data collection. Friedman engineered the data and produced the interactive visualizations, auto-correlation plots, and the regression analysis. He also spent time performing exploratory analysis on the dataset. This process allowed the discovery of the results in the paper. Andriy Lunin discovered and proliferated the mathematical basis for logistic function analysis in Pandemics. He applied this mathematical basis to the data-set producing key findings in the asymptotic value approximations in the paper. The introduction and paper formatting was done by Debicki. The abstract, methodology, results, and conclusion were written by Lunin. The data section and parts of the methodology and results were written by Friedman. The latex formatting was written by Lunin's colleague, Tengiz Ibrayev, who we owe our gratitude to.

[1] C. J. Wang, C. Y. Ng, and R. H. Brook, Jama **323**, 1341 (2020).
[2] G. King and L. Zeng, Political analysis **9**, 137 (2001).
[3] A. Apolloni, C. Poletto, and V. Colizza, BMC infectious diseases **13**, 176 (2013).
[4] J. R. Gog, S. Ballesteros, C. Viboud, L. Simonsen, O. N. Bjornstad, J. Shaman, D. L. Chao, F. Khan, and B. T. Grenfell, PLoS computational biology **10** (2014).
[5] J. Dehning, J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek, and V. Priesemann, medRxiv (2020).
[6] *Covid-19 dashboard by the center for systems science and engineering (csse) at johns hopkins university (jhu)*, `https://coronavirus.jhu.edu/map.html`, accessed: 2020-04-30.
[7] *Coronavirus forces gulf states to reckon with their scores of migrant workers*, `https://www.nytimes.com/2020/05/09/world/live-coronavirus-news.html`, accessed: 2020-05-10.
[8] *The covid tracking project*, `https://covidtracking.com/api?fbclid=IwAR0J_lIh9A2e8Z-1gv3twJXH2ZTvf9_5C_I9H8kdA5uz8KBkc7I1ej_HE6Y`, accessed: 2020-04-30.
[9] *Center for disease control and prevention*, `https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html?fbclid=IwAR2wH_efRlGUI8czzBMIBew0cxYkFJ9GzmdJujymqX21L48ZZ3DZgCG_9TA`, accessed: 2020-04-30.
[10] *Github repository coronavirus us testing*, `https://github.com/jaisal1024/coronavirus-us-testing`, accessed: 2020-04-30.
[11] S. Knaption, The Telegraph (2020).
[12] *Google colaboratory v1*, `https://colab.research.google.com/drive/1qTz_nMK9lnh9_WSEpg1puValpNIlPujw#scrollTo=sZEyn7oNJlEI`, accessed: 2020-04-30.
[13] S. Kaplan and K. Thomas, The New York Times (2020).

:

TABLE 1: Multi-variate Regression

	positiveIncrease
const	3.2656
	(8.2185)
isWeekday	-2.1514
	(6.5393)
totalTestResultsIncrease	0.0214***
	(0.0016)
Population	0.0000***
	(0.0000)
totalTestResults	-0.0022***
	(0.0002)
days_since_100_cases	2.4080***
	(0.8091)
pending	0.0071***
	(0.0019)
deathIncrease	0.8971***
	(0.2290)
recovered	-0.0266***
	(0.0028)
MVA_totalTestResultsIncrease	0.0319***
	(0.0055)
MVA_positiveIncrease_yesterday	0.8057***
	(0.0228)
days_since_lockdown	-0.4522
	(1.1473)
state_abbr_AL	-77.3340*
	(40.4586)
state_abbr_CA	-465.4219***
	(39.6380)
state_abbr_FL	-160.3519***
	(39.4264)
state_abbr_IL	98.4816**
	(39.5861)
state_abbr_LA	95.0704**
	(40.5264)
state_abbr_MA	166.9494***
	(44.1790)
state_abbr_NC	-86.4923**
	(39.3162)
state_abbr_NJ	375.2192***
	(44.7317)
state_abbr_OR	-68.7500*
	(39.4798)
state_abbr_TX	-133.6022***
	(38.5866)
weekday_0	-42.9515***
	(13.4629)
weekday_1	-32.7659**
	(13.4112)
weekday_2	-14.2968
	(13.4181)
weekday_3	31.4361**
	(13.2054)
weekday_4	56.4267***
	(13.0785)
weekday_5	26.5930**
	(11.6090)
weekday_6	-21.1760*
	(11.9102)

* p<.1, ** p<.05, ***p<.01

:

TABLE 2: PanelOLS Estimation Summary

Dep. Variable:	positiveIncrease	**R-squared:**	0.7483
Estimator:	PanelOLS	**R-squared (Between):**	0.9421
No. Observations:	2617	**R-squared (Within):**	0.7483
Date:	Fri, May 08 2020	**R-squared (Overall):**	0.8878
Time:	13:58:30	**Log-likelihood**	-1.849e+04
Cov. Estimator:	Unadjusted		
		F-statistic:	844.70
Entities:	51	**P-value**	0.0000
Avg Obs:	51.314	**Distribution:**	F(9,2557)
Min Obs:	44.000		
Max Obs:	94.000	**F-statistic (robust):**	844.70
		P-value	0.0000
Time periods:	102	**Distribution:**	F(9,2557)
Avg Obs:	25.657		
Min Obs:	0.0000		
Max Obs:	51.000		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
const	53.620	9.0563	5.9208	0.0000	35.862	71.379
totalTestResultsIncrease	0.0218	0.0017	13.199	0.0000	0.0186	0.0251
totalTestResults	-0.0022	0.0002	-8.6213	0.0000	-0.0026	-0.0017
days_since_100_cases	2.4757	0.8137	3.0427	0.0024	0.8802	4.0712
pending	0.0072	0.0019	3.7141	0.0002	0.0034	0.0111
deathIncrease	0.9347	0.2279	4.1011	0.0000	0.4878	1.3816
recovered	-0.0265	0.0028	-9.3609	0.0000	-0.0320	-0.0209
MVA_totalTestResultsIncrease	0.0313	0.0056	5.6356	0.0000	0.0204	0.0422
MVA_positiveIncrease_yesterday	0.8008	0.0228	35.191	0.0000	0.7562	0.8455
days_since_lockdown	-0.3406	1.1543	-0.2951	0.7680	-2.6042	1.9229

F-test for Poolability: 8.4198
P-value: 0.0000
Distribution: F(50,2557)
Included effects: Entity