COVID-19 Data Analysis in Delaware
Summer Research Project

Tolga Atabas

July 18, 2020

Contents

1 Abstract

The COVID-19 pandemic has impacted this country's public health, education, and economic wellbeing in many ways. Schools were closed, the economy was shut down, and mask wearing was instituted. The purpose of this study was to quantitatively take a look at the changes in cases of COVID-19 patients as the pandemic progressed. Exponential models were fitted for both the cumulative cases and daily cases data for the State of Delaware. Important policy dates are also considered during data analysis. Thus, it is possible to relate any falls or rises in the exponential model parameters to policies or social activites occurring within that corresponding time period. Officials can thus learn the impact of policies and use the experiences as an aid for future actions. Figure 1 shows a sample of the data results.

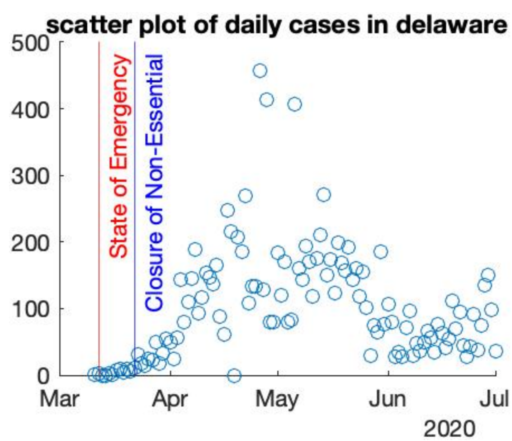


Figure 1: A scatterplot of the daily cases encountered in Delaware from its first cases on March 11, 2020 to the most recent day from the date of the creation of the figure which was June 23, 2020. This is intended to give a beginning understanding to what the project aims to do by showing that the data is visualized by looking at the real data and overlaying important dates of policies and also model calculations later on.

2 Introduction

Relating policy decisions in numerical standards allows for a more objective view on the progression of public health threats like the COVID-19-causing virus. Therefore, analyzing the data associated with COVID-19 by using quantitative methods opens windows to a better understanding of the relationship between the progression of the pandemic and policies taken by those in power. The COVID-19 pandemic has impacted this country's public health, education, and economic wellbeing in many ways. Schools were closed, the economy was shut down, and mask wearing was instituted. There was empirical evidence that such measures would have a desired impact, but there is quantification of what that impact might be. The purpose of this study was to take data from a small section of society and use this data from case counts to determine what, if any, impact these decisions might have. The data for the state of Delaware was taken for this study. It is a small self-contained community which experienced significant impact from the virus.

Several different models were applied to the data and the models were evaluated to determine which was most appropriate. The parameters in the model were indicative of a change in policy and its effectiveness as the values increased or decreased. Due to the highly infective nature of the virus, exponential models of forms $y = ae^{kt}$ and $y = ae^{kt} + b$ were used. The infective nature inspired the first model, an exponential model. The second model, with the $+b$ parameter was attempted in order to capture increases and decreases in the cases due to policies taking effect.

Using appropriate models, comparisons of policy decisions made by Delaware Governor John Carney can be determined with the corresponding effect on the cases and the model parameters. The State of Delaware keeps a record of all actions[1] that the State took in response to the pandemic. The data, which was acquired from a New York Times GITHUB Repository[2], was interpretted in the scope of these policies.

All code, figures, and functions can be found on my GITHUB Repository[4] at `https://github.com/tatabas/delaware-covid-analysis`.

3 Methods & The Model

The first step in starting the analysis was to obtain the data from the New York Times GITHUB Repository[2] which lists the cumulative cases in each state per day. After acquiring this data, it was simple enough to extract the data for Delaware only and then begin the analysis.

The structure of this section will be set as follows: the methods of analysis for the cumulative cases data will be explained first. This analysis was done with two models: $y = ae^{kt}$ and $y = ae^{kt} + b$. After an in depth review of the modeling methods employed in this project for the cumulative data, daily data will be explained in a similar manner. Daily data was only modeled using $y = ae^{kt}$.

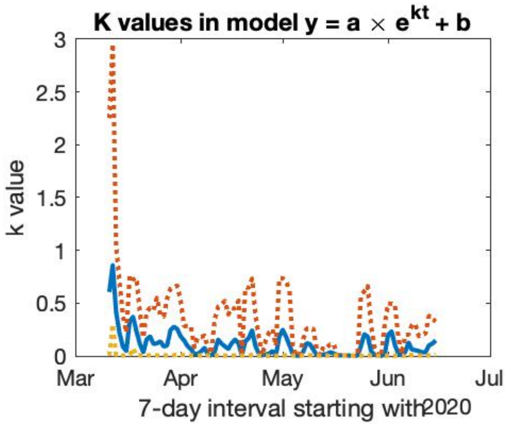
3.1 Cumulative Data: $\mathbf{y = ae}^{kt}$

It was appropriate to use a basic exponential function to begin a quantitative search into the present data set. Since the virus has a very infective nature, an exponential model would be able to capture the fast-paced compounded growth. Before going into the mathematics, it is worth taking note on the logic and assumptions behind the model.

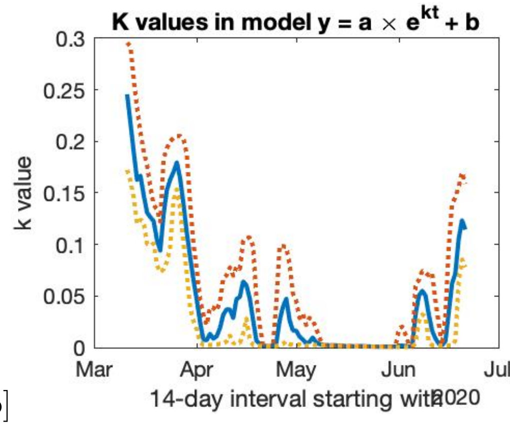
The key assumption in this model is that the model parameters do not change in a span of 14 days. Each modeling window consists of a constant set of time (t) values 0 to 13 days, [0:13], and the COVID-19 case (y) values for that window of time. Then, the next modeling window is created by shifting over one day. For example, the t vector for the first modeling window is [0:13] and the y vector is made up of the COVID-19 cases for days [1:14] of the pandemic in Delaware. Then the second modeling window has a t vector [0:13] again and the y vector is the case values for days [2:15]. This method allows for the viewing of changes in the infection rate as the virus progresses.

For the modling, a span of 14-days was chosen because it is a tradeoff between an accurate estimation of values and the ability to track those changes. A modeling window of 7 days was attempted initially, but the resulting parameter estimates were too noisy and made it difficult to interpret actual changes in infection rate to noise from a small data set in a short modeling window, as seen in Figure 2. Also, since the data are always modeled from times 0 to 13, the constant "a" provides an alternative estimate to the viral cases as opposed to the 7 day average. Also, in regards to the model's t vector, a constant set of [0:13] was used because a change with the day number (e.g [1:14] and then [2:15]) would result in an inconsistent model. Higher t values would decrease/change the values of the parameters being calculated. It was decided that 0 was a proper point to start the vector because it makes the system neater: $y = ae^{kt} = ae^0 = a$. This model is reflected in firstmodelCumulativeDelaware.m in my GITHUB Repository.

The model parameters for $y = ae^{kt}$, are estimated by linearization. Since this is a simple model, methods from linear algebra were utilized to begin the data fitting process. The natural log (base e) was taken to linearize the equation, resulting in $log(y) = log(a) + kt$. Using this model, the equations can be put into matrix form where there is a matrix A (a 14X2 matrix composed of a column of ones paired with the t vector, [0:13]), a vector $\hat{p}$ (made up of the parameters a and k), and a vector y (made by taking the log of cases data) $A\hat{p} = y$. Solving for p: $\hat{p} = (A^TA)^{-1}A^Ty$. With this method, it is possible to get the parameters for the specified window. At the time of this analysis, the available data consisted of cases up to and including July 18th. Thus, there were 117 modeling windows, producing 117 sets of parametes.



[a]



[b]

Figure 2: Plot A shows the k parameters in the $y = ae^{kt} + b$ model under the 7-day modeling window, whereas plot B shows it under the 14-day modeling window. The dashed lines represent the upper and lower bounds of the 95% confidence intervals for the model. Since the 7-day modeling window resulted in calculations that have wide ranges in the confidence interval and lots of suddent drops, the 14-day modeling window was deemed more appropriate. The 7-day modeling window does not provide much valuable information due to the noise created by sudden drops and rises. More will be explained as to how the data in these figures were calculated.

To get an estimate of the range of possible solutions that are consistent with the model, a technique called bootstrapping was used. Each modeling window was bootstrapped[5] to get an estimate of the probability distribution for the parameters of each modeling window. From this distribution an estimate of the 95% confidence interval is obtained. This presents a measure of uncertainty in the estimate. The larger the interval, the greater the uncertainty. The bootstrapping function works as follows: it takes a random sample of size 14 with replacement from the vector t and y 1000 times. Each randomized sample is then used to estimate the model paramters.

The result from adding a bootstrapping function to the 117 modeling windows produces two matrices of size 1000X117. One matrix is for the parameter a and the other for parameter k in the model $y = ae^{kt}$. Before accepting these results, the means and medians of the bootstrapped results for each modeling window were compared to the non-bootstrapped results of that window. All of these results were close in value, showing that the probability distribution was relatively normal and accurate. Therefore, the mean was taken as the estimated parameter value and the 25th and 975th points (after sorting from least to greatest) were taken as the lower and upper bounds, respectively, for the 95% confidence interval.

After going through these methods, it was possible to simply plot the values and acquire a visual representation of the changes in parameters for each modeling window with their corresponding 95% error bars, as seen in Figure 2b.

3.2 Cumulative Data: $\mathbf{y} = \mathbf{ae}^{kt} + b$

As should be expected, the policy changes enacted by the Governor's office in Delaware caused changes in cases. The model $y = ae^{kt}$ can not capture certain policy effects or sudden hot spots accurately. For example, closing businesses takes a large number of people out of the pool capable of spreading the virus or a newly detected burst of cases at a nursing home causes a sudden increase of cases. However, an additional $+b$ parameter allows for these changes to be reflected in the analysis. And it is for this reason that the $y = ae^{kt} + b$ model was introduced. The logic and assumptions from the previous model applies to this regarding the 14-day modeling windows.

Because of the additional $+b$ parameter, it was no longer possible to use the linear algebra methods used in the previous model. Instead, a nonlinear optimization method was used to obtain the optimal estimates of the parameters using the fmincon[3] function.

As explained before, the calculation for modeling is conducted for each 14-day modeling window. There are 117 modeling windows for this method as of July 18,2020, the date for which this report is based upon. For each modeling window, the t vector was a constant [0:13] and the y vector was the cases data for the corresponding 14-day window.

Optimization minimizes the sum of square difference between the observed data and that calculated by the model. This is called least squares error estimation. By setting the objective function of fmincon as the least squares error, it is possible to find the least squares error. To illustrate how this works, it is best to show a code piece:

```
if day == 1
  %the initial point is in form [a;k;b]. 1.9 and 0.3 come from the
  %first model which was based upon y = ae^kt. It serves as an
  %approporiate starting point for this analysis.
p(day,:) = fmincon(@(x)lstSqrs(x,t,y),[1.9;0.3;0],[],[],[],[],[-inf;0;-inf],
   [inf;20;inf]); %[a k b]
%a and b can be negative due to policies that heed the growth of virus
%k is always positive because cumulative data cannot have negative growth
else
%the parameters for the previous window make an approporate guess
p(day,:) = fmincon(@(x)lstSqrs(x,t,y),[p(day-1,1);p(day-1,2);p(day-1,3)],
   [],[],[],[],[-inf;0;-inf],[inf;20;inf]);
end
```

This is a fairly complex piece of code, however the first note that should be made is regarding the structure of fmincon. Fmincon starts with a function as the first input. An implicit in-line function was created, as seen in the code, to create a variable for the guesses that the fmincon tries as it searches for the optimized value. The second input is the initial guessing point. For this program, the first guess came from the first modeling window in the previous model, $y = ae^{kt}$. Since the first modeling window in the previous model had $a = 1.9$ and $k = 0.3$, the initial guess for the first week in this model was $a = 1.9$, $k = 0.3$, and $b = 0$. The following brackets can be filled in with values for linear constraints, but no constraints were used in this calculation. The last two inputs are the limits for each parameter. Parameters a and b were left without a bound because they could be negative or positive depending on the cases of that modeling window. Since k did not exceed 5, it was appropriate to leave the upper bound at 20.

A good initial guess is very important in getting good estimates of the parameters. The approach adopted here is that the initial starting point for the fmincon function is the calculated parameters from the previous modeling window. This allowed for consistency and accuracy in the calculations since the parameters are not expected to change extremely between two adjacent modeling windows.

The lstSqrs function inputted to the fmincon function is a custom function that was created to take the fmincon inputs of x and model them into the equation $y = ae^{kt} + b$. Then it finds the sum of squared error between this equation and the real data. After finding the set of parameters that minimizes the sum of squared errors, this set of parameters is sent back as the answer, which is recorded.

As was done previously, bootstrapping was also added to this function to get a sense of the amount of error in the calculations. The bootstrapping was done in a similar way: random sample with replacement was taken from vectors t and y 1000 times, each time the randomized order was plugged into the fmincon function, then follwed the coding path to the lstSqrs function to find the best model to fit that set of data. In the same way, the mean of each modeling window's bootstrapped data was accepted as the estimate of the parameter, and the 25th and 975th points as the bounds for the 95% confidence interval, as illustrated in Figure 3. The code for this model can be found under the MATLAB code named finalCumulativeDelaware.m

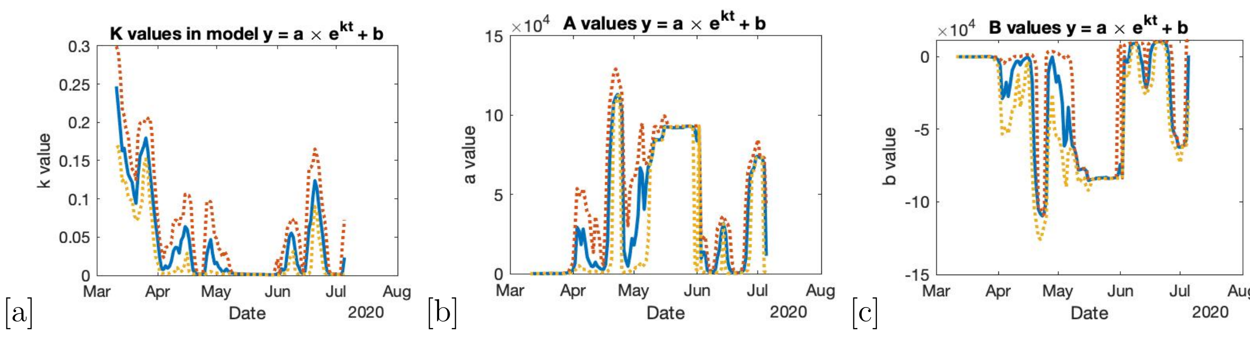
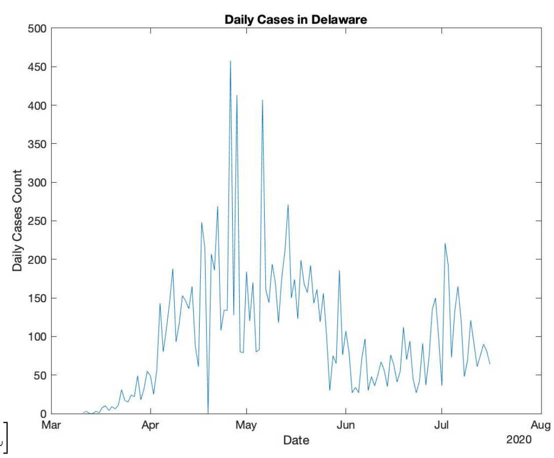


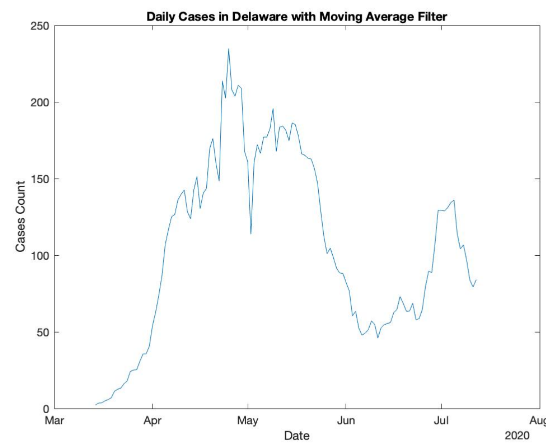
Figure 3: Cumulative COVID-19 cases data were modeled in 14-day windows to the equation $y = ae^{kt} + b$. The 95% Confidence are shown with the red and yellow dotted lines. The parameters of k, a, and b are shown here.

3.3 Daily Data: $\mathbf{y = ae}^{kt}$

After finishing calculations on the cumulative data with the two models, the focus was shifted towards the the daily data. To get the daily data from the cumulative data in the NYT GITHUB Repository, a simple y(i)-y(i-1) calculation was done for each day to get daily cases per day. However, there were large disparities from day to day. This caused problems in data modeling, so a moving average filter was implemented on a weekly, 7-day basis. This has a similar structure to the modeling windows in that the moving average filter first took the average of days [1:7], then [2:8] and so on. After doing this, everything was shifted 3 days since it was calculated that the middle of each moving average window would be the day for the data. For example, the first cases in Delaware were observed on March 11, but after the moving average filter the first day of data is on March 14. This is illustrated in Figure 4, where the daily moving average makes the data apt for modeling.



[a]



[b]

Figure 4: The plot a shows the extreme variance found in the daily counts of COVID-19 cases in Delaware. As can be expected, the moving average filter of 7-days creates a smoother plot and makes the data appropriate for fitting to a model, as shown in plot b.

After creating this dataset composed of daily data with a moving average filter, it was possible to begin the data fitting procedure. The procedure is similar to that of Cumulative Data: $y = ae^{kt}$. There are 14-day modeling windows, and the data is fitted by linearizing the function and solving for the parameters that result in the shortest distance between the real data and the model, a least squares fit. As was done before, this calculation also had a bootstrapping procedure where random samples with replacement were taken from the data and sent to data fitting 1000 times. The mean was accepted as the estimate of the parameter value and the 25th and 975th values as the lower and upper bounds, respectively.

The second model, $y = ae^{kt} + b$, was attempted on the daily data but found to be inappropriate for this dataset. The additional $+b$ parameter is not needed because of how the data was calculated. Since it was calculated by subtracting the one day from the previous, the following assumption results: if $y_{n+1} = ae^{k(t+1)} + b$ and $y_n = ae^{kt} + b$, then the result is a value that effectively nullifies b because there should not be extreme changes in b from one modeling window to the next. Because of this, daily data was only modeled with $y = ae^{kt}$. As previously stated, the methods for this are the same as the $y = ae^{kt}$ model for the cumulative data except for the addition of the moving average filter.

As of July 18th, 2020 there were 110 modeling windows in this analysis. These calculations are reflected in finalDailyDelaware.m.

4 Analysis & Application

4.1 The Daily Data

To begin applying the quantitative trends in the graphics to real events in Delaware, some markers were added to dates of importance. These plots in Figure 5 display the data for the daily cases data that were analyzed:

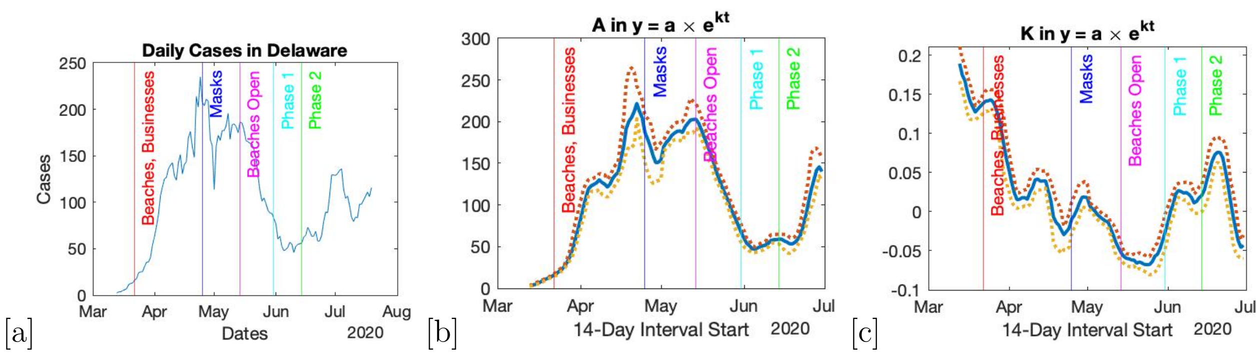


Figure 5: Daily cases data were modeled in 14-day windows to the equation $y = ae^{kt}$. The 95% Confidence are shown with the red and yellow dotted lines for each parameters k and a. The other lined markers display the following information:

- Red Line: "Beaches Businesses", March 22
- Blue Line: "Masks", April 25
- Purple Line: "Beaches Open", May 14
- Cyan Line: "Phase 1", May 31
- Green Line: "Phase 2", June 14

For the discussion on how these calculations relate to the real events in Delaware, most – if not all – the attention will be turned to the k parameter. The most important parameter is k in that it defines the exponential growth. If k is positive, there is exponential growth. The larger the k, the more severe the rate of infection is. If k is negative, then the virus is in decline. Because the 14-day window is moved down one day, it is possible to closely monitor k and relate it to events in the state.

The first incline in k values comes around March 20, making a parabolic shape, then dropping back down into a decline. This is an interesting point because it coincides with many activities in Delaware. First, during this time period there were outbreaks within nursing homes, which became very severe during March and early April. However, also worth noting, is the weekend starting with March 20th. During this weekend, there was a news article[6] that outlines the immense crowding of Delaware beaches on March 20th. Together, the nursing homes and crowding of beaches, caused rises in cases throughout Delware in the late March time period, which is reflected in the rises in the k values shown in Figure 6.

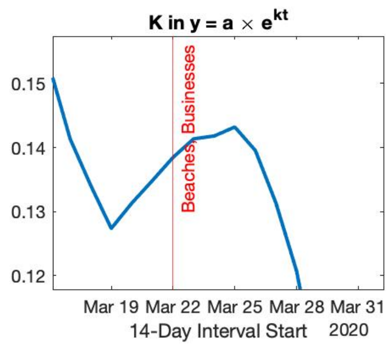


Figure 6: The K parameter from modeling the daily cases data of COVID-19 in Delware to $y = ae^{kt}$ in 14-day windows zoomed into the time period that relates to nursing homes and crowded beaches. The red line with beaches, businesses marks the date that beaches and non-essential businesses were forced to close.

In response to the nursing home outbreaks, the State of Delaware offered closely monitored policies specifically for nursing homes that were updated with frequency in the early days of the pandemic. These policies ranged from isolation to limiting visitors, to also defining how residents should be brought back to the facility after successful treatment. Since nursing homes had become one of the epicenters of the virus in Delaware, these precautions were necessary and rightfully taken. The policies aimed specifically at nursing homes seem to have worked in starting a decrease after a peak k value of 0.14 on March 25th. By March 28th, the k values were back to 0.12.

To measure the impact of the crowded beaches, it is important to note the incubation time of the virus. It was found in a study that it takes approximately 5 days[6] to develop the first symptoms of the virus. Therefore, adding 5 days after each major incidence should align with a logical conclusion in the rises or falls of the k parameter. If March 20th is taken as an exposure date for individuals that were in the crowd and possibly contracted the virus, adding 5 days results in March 25th, which is the peak of the parabolic increase in the k parameter. Since 5 days following exposure is located at the peak of the curve, it is reasonable to assume that the beaches were not a major factor in the parabolic increase seen in the plots. This may be due to the fact that the incident was for 1 day.

The measures taken in the nursing homes and the closure of beaches and businesses can be responsible for the steady decline following the peak of $k = 0.14$ on March 25th to $k = 0.015$ on April 4th. The fall can be attributed to Delaware's speedy responses to these situations. Governor Carney issued orders to close Non-Essential businesses and both public and private beaches just two days after the incidents on March 20th. With the closure of non-essential businesses, and regulations in essential business, the k parameter had a successful decrease of 89% before the next rise in k values.

This decline is then interrupted by another parabolic increase in k values in early- to mid-April, illustrated in Figure 7. Although other factors could have come into play, this outbreak can be attributed to the chicken factories in Sussex County. It was reported[8] on April 15th that the chicken factories were suffering from employee shortages. Since employees were already diagnosed and dismissed from the factories by this date, there is no need to add 5 days to calculate the exposure to symptomatic period.

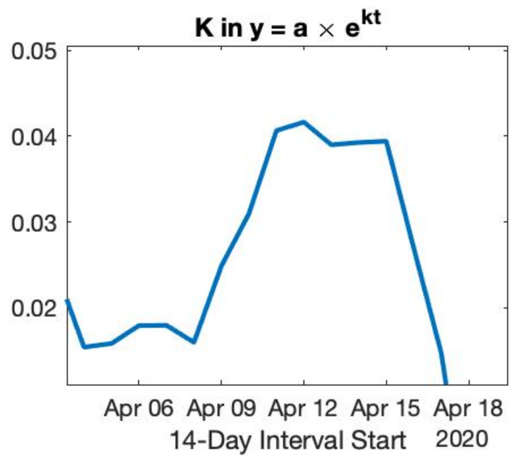
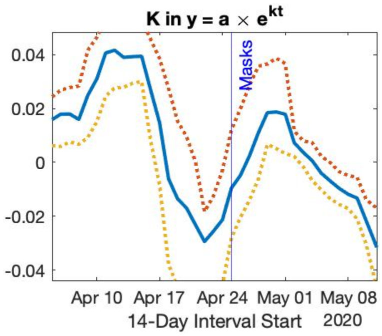
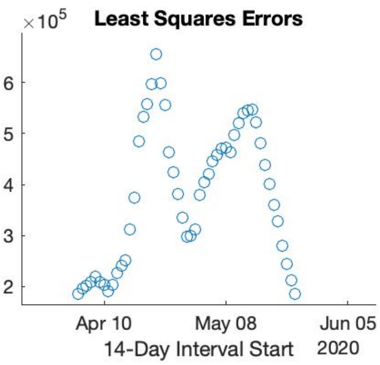


Figure 7: The K parameter from modeling the daily cases data of COVID-19 in Delaware to $y = ae^{kt}$ in 14-day windows zoomed into the time period that relates to chicken factory outbreaks.

After the chicken factory outbreak period the k value changes in many unusual ways, as seen in the rises and falls of k in Figure 8a. However, since this portion of the data exists in the crossing between a positive and negative k value, it should be taken with a grain of salt. Especially in some points within this period of mid-April, the 95% confidence interval, marked with the red and yellow dotted lines, include the point 0. If the confidence interval includes 0, there is a chance that the virus is not growing in infection rate at all, which is good. This shows that the collective effects of the policies helped curb the spread of the virus and control the pandemic. This critical interval is an important transition period that the model used here may not capture with accuracy. Since it is transitioning from an infective rate to a declining rate, the transition may cause higher rates of uncertainty. This uncertainty can be seen with the width of the confidence intervals but also with the least squares errors, as shown in Figure 8.



[a]



[b]

Figure 8: The K parameter from modeling the daily cases data of COVID-19 in Delware to $y = ae^{kt}$ in 14-day windows zoomed into the time period that relates to chicken factory outbreaks in 8a. and the corresponding least squares errors in 8b. The blue line in 8a labeled Masks is the date, April 25th, which a universal requirment of masks were required for people in Delaware.

During the period of uncertainty in the model, it is unknown in this study whether the quick increase is a continuation of the chicken factory outbreak or another source. Whatever the cause, the k values were showing a quick and rapid rise that continued through April 29th. During this rise, the k values marked with blue are negative, but the confidence interval includes positive values. At this point, the virus is transition between an increasing and declining nature. However, the real cases do keep on rising. There were 161 more cases reported on April 23rd. Governor Carney introduced a universal masking requirement for people in Delaware on April 25th. This fast response helped with preventing a further increase in infectivity. 5 days after this requirment, the k values starts to reach a local maxima and begin to fall.

To further explore the instances in this period, a simple rate of change calculation between each k value was performed (a simple slope calculation). This is illustrated in Figure 9 after being applied to all k values. Even though the calculations are very rapidly changing and do not appear to show a smooth transitions, it portrays the idea that the k parameter changes from a decreasing change to positive change up until April 24th since the rate of change is negative on April 17th and rises to a positive until the 24th. Then the rate of change drops until April 29th, where the rate of change of k is close to 0. By April 30th, the rate of change of the k value is back to negative values, indicating a decrease in infectivity. This aligns with the effective period of the masking policy. From these calculations, the universal, required masking helped control the spread of the virus.

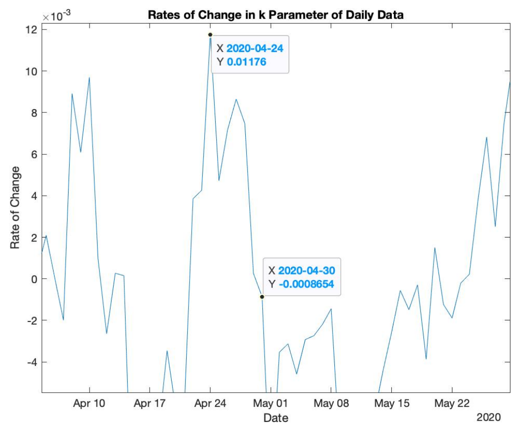


Figure 9: A simple rate of change calculation between each k value in the form of $\frac{(k+1)-k)}{((x+1)-x)}$. This was applied to the k values obtained from the 14-window modeling on daily data for the $y = ae^{kt}$ equation.

To compare the the effects of closing beaches and non-essential businesses with a masking requirement, these intervals will be split into phases. This method was inspired by another study conducted in the MassGen Boston hospitals[9]. The date of closures were taken as the beginning of the phase, then there were 5 days added on to wait for the virus to show up in the data as cases. The effective phase for the closing of beaches is taken between March 27th and April 4th, as shown in Figure 10. This interval expresses 5 days after the closure policy, up to the beginning of the chicken factory outbreak. To calculate the slope line here, a simple slope formula of $y = mx+b$ was used. The slope for this period was -0.0145, as seen in Figure 10a. The same process was applied to masking and the slope was -0.0051 (interval was April 30 to May 14) also shown in Figure 10b. The rate of decrease for the period following beach and business closures was almost three-fold that of the period following masks. Even though this calculation is by no means complex and comprehensive, it shows that logical relation that keeping people in their homes has a greater impact on controling the spread of the virus. If a greater percentage of people are sheltering in their homes, then the susceptible population does not get the virus because susceptible and infectious individuals are all in their homes, limiting interaction.

[a] [b]

Figure 10: For Figure 10a a line of form $y = -0.0145x + 0.33$ is overlayed on the k values calculated from the daily cases COVID-19 data through the 14-day modeling window for the equation $y = ae^{kt}$. Figure 10a expresses the effect of beach and business closures. The same was done for the Figure 10b that shows the effects of required masking. The line is of form $y = -0.0051x + 0.264$

After beaches start to open up again, the rate of decrease goes down, as can be seen visually in Figure 11 by the rounding of the decline and the change to an positve rate of change occurring after the beaches open up. The portion of data following the opening of beaches has many activities that can cause an increase in the spread of the virus. The State of Delaware had very specific guidelines for opening up the beaches to ensure distancing and safety, but there is only so much the State can do. Of course, it cannot be taken out of consideration that if it were not for these precautions, the spread of the virus could have been worse. However, it is impossible to ignore the fact that even with the precautions there was an increase in the spread of the virus. The opening of beaches and the subsequent phased opening are illustrated in Figure 11.

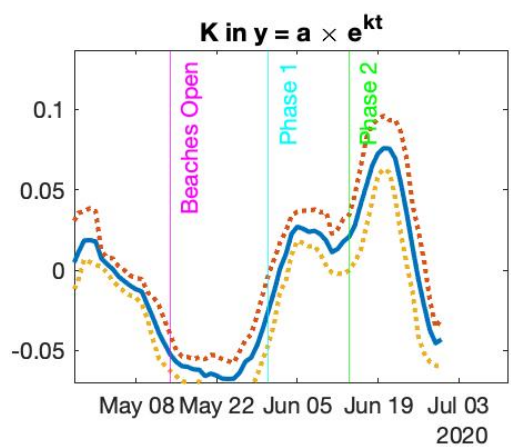


Figure 11: The K parameter from modeling the daily cases data of COVID-19 in Delware to $y = ae^{kt}$ in 14-day windows zoomed into the time period that relates to final portions of the analyzed data.

With the introduction of phased openings, many business sectors saw increased legislation on their business operations. Many were for good reason, helping businesses open and restart the economy in a responsible and safe way. The effects and specificity in the documents pertaining to each type of business is truly commendable. Also, Phase 1 openings do not seem to play a big role in an increase or decrease of k. By applying the 5-day gap, it is possible to see that the Phase 1 effective period (the period when its effects can be measured) correlates with the peak of the previous incline, as seen in Figure 11. This could possibly indicate that with more experience in applying distancing measures on beaches, the infection rate was decreasing. The enforcement of controlled social interactions do make an impact.

The story is a bit different with Phase 2. With more establishments opening up, and the susceptible population is more prone to interactions with a member of the infectious population, causing rises in the exponential growth, also reflected in Figure 11. This is inevitable. The important part is controlling the growth, which the State of Delaware did – and continues to do – very well. Each outbreak was followed by timely and proper policies that helped combat the threat of increased transmission.

The good news from the visual is that the k value peaks and then drops. This drop may be associated with the fact that Delawre enacted further measures that modified the Phase 2 opening plans. For example, it was seen that bars were acting as a major hotspot for transmission, so further regulations were put in place for bars and their taprooms. Further limitations and sanitary requirements were placed for indoor and outdoor dining. All of these in concert help create an environment where transmission is decreased and the spread of the virus is controlled.

4.2 The Cumulative Data

After getting a sense of the trends in the State with the daily date, it is worth taking a note at the cumulative data. In this analysis, it is again important to look at the k values since this is the exponential parameter that reflects the extent of the virus spreading. Moreover, since the cumulative data was based on the model $y = ae^{kt} + b$, it is important to study the impact in b, since this should express another view on the extent of the increases or decreases in the cases as policies change.

- Red Line: "Beaches Businesses", March 22
- Blue Line: "Masks", April 25
- Purple Line: "Beaches Open", May 14
- Cyan Line: "Phase 1", May 31
- Green Line: "Phase 2", June 14

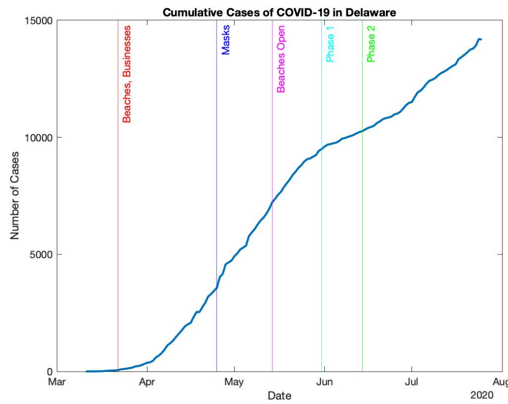
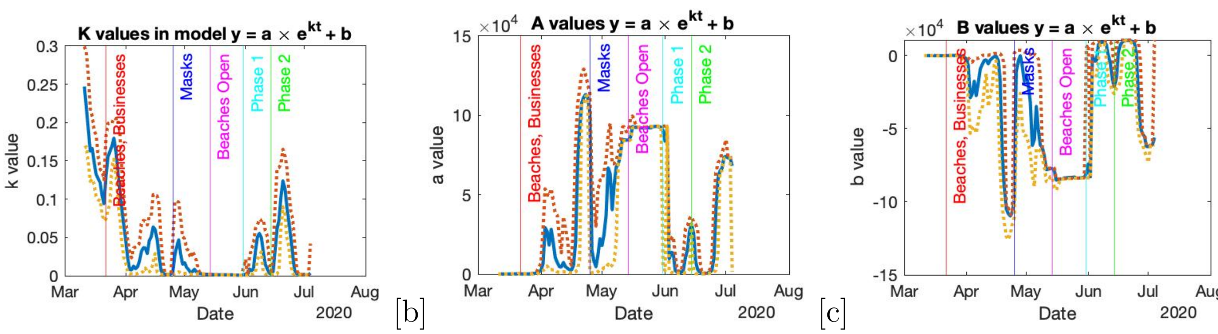


Figure 12: The untouched data of cumulative COVID-19 cases in Delaware. The same markings from before are used to show when policies were enacted.

Of course, because the data in this set is of cumulative cases, there will never be a negative exponential growth. A decrease of positivity in k will be accepted as a decrease in the spread of the virus. In many instances, the conclusions that were made with the previous model can also be made with this model on cumulative data. For example, the first outbreak that was associated with nursing homes and beaches, and the second outbreak associated with chicken factories also show up in this model, as shown in Figure 13.

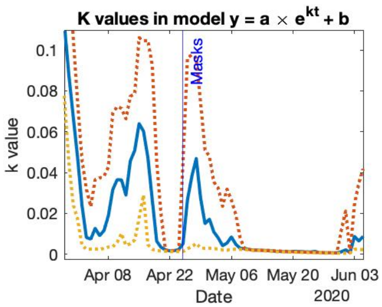


[a] [b] [c]

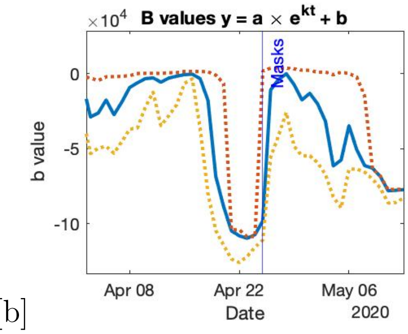
Figure 13: Cumulative COVID-19 cases data were modeled in 14-day windows to the equation $y = ae^{kt} + b$. The 95% Confidence are shown with the red and yellow dotted lines. The parameters of k, a, and b are shown here. The other line markers display the same information as in the caption of Figure 12, with each line a date of the enactment of a policy.

The same conclusions can be made about these outbreaks since the increasing parabolic trends and subsequent decreases are all shaped in a similar fashion. Even though this is true, it is worth taking a look at the late April interval where the virus seemed to have a transition period and there was uncertainty in the previous model.

This portion of the model in the cumulative model shows the same trend: following the chicken factories outbreak in the April 8-17 time frame, there is a sudden drop, as shown in Figure 14. This is probably due to the fact that this was a localized outbreak and did not ripple throughout the rest of Delaware. Then, there is a sudden rise, which is curtailed by the masking requirement. The masking requirements effects are seen once again as the k values and b values fall following April 30th in Figure 14. The b value solidifies the chicken factory outbreak as a more isolated event because the b value is high during the outbreak due to rapid increases in the number of cases. Then, after the outbreak the b value falls rapidly. Since the cases in the factories were controlled, the rapid decrease in the number of cases caused an large change in the model.



[a]



[b]

Figure 14: Cumulative COVID-19 cases data were modeled in 14-day windows to the equation $y = ae^{kt} + b$. The 95% Confidence are shown with the red and yellow dotted lines. This figure focuses on the interval during the chicken factory outbreak and after.

The b parameter is also appropriate to further commend the usage of masking in Delaware. As was explained in the daily data, masking did cause a decrease in the k parameter. The same pattern is reflected in the cumulative data, so this will not be explained again. Moreover, masking's effects can also be shown in the b parameter, illustrated in Figure 14b with a declining trend following masking. 5 days after masking, the b parameter falls to negative values. A decline in the b parameter indicates that a number of the spreaders of the virus has dropped out of the pool of those who can infect. This indicates that masks were working.

In efforts of comparing the last few policy intervals, the daily data graphic is reintroduced with the cumulative data:

The decline in the exponential parameters after masking results in a flat-line for the k and b values in the cumulative data examples. Whereas, the daily data example is a gradual fall. This could mean that the $+b$ parameter is not appropriate for this time interval. The decrease in decline, but still decay of the virus growth rate after beaches open, as observed in the daily data example, is more logically sound. Comparing the two models is important because one model may capture one event while the other does not.

The rest of the events follow the same patterns as the daily data. An increase between Phase 1 and Phase 2 that later falls. Then an increase between after Phase 2 that also falls.

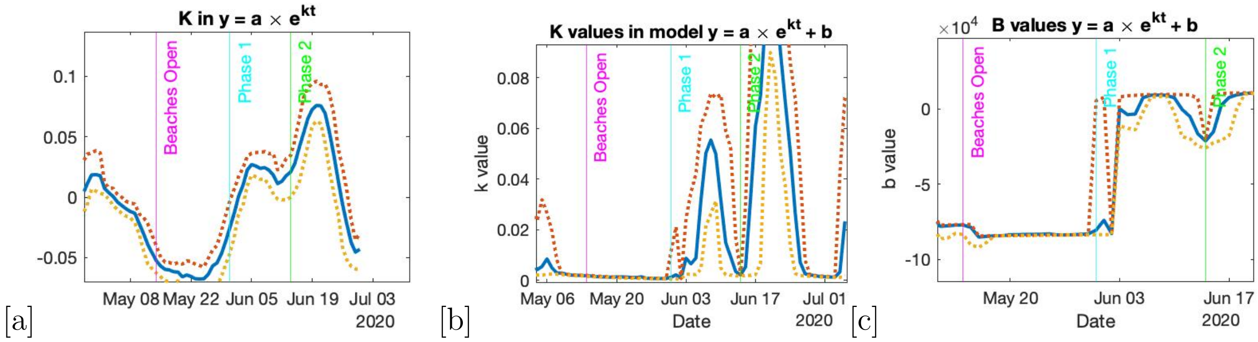


Figure 15: Figure 15a is the same graphic from Figure 11. This shows the k values calculated from the daily data modeled in 14-day windows for the equation $y = ae^{kt}$. The red and yellow dotted lines are the upper and lower limites of the 95% confidence interval. Figures 15b and 15c show a similar time interval with for the cumulative data modeled in 14-day windows for the equation $y = ae^{kt} + b$. The red and yellow dotted lines are the upper and lower limits of the 95% confidence interval.

With the information available during this study, the fall after Phase 2 might be attributed to Governey Carney's response to bars and taprooms, as well as restaurants.

4.3 Conclusion

The overall theme in the association of the data and graphics with the policies lies in the fact of control. Delaware had timely responses to each of the outbreaks that occurred in the State during the pandemic. Each of these responses were successful in their own way. Some responses came within days, as with the overcrowded beaches example: the response to close beaches came just 2 days after the incident. This timely behavior by the policy makers of Governer Carney's office and the State of Delaware truly did help control the spread of the virus. Most of the outbreaks were from isolated events like the nursing homes and chicken factories. The state itself was put under control with shelter-in-place and state of emergency orders that included and was not limited to closing businesses and requiring masks. These had profound effects that were displayed in two models looking at Cumulative and Daily COVID-19 cases in Delaware.

5 Acknowledgments

I would like to thank Professor Piovoso from the Engineering Department at Swarthmore College for his support and mentorship during this project. I would also like to thank the Swarthmore College donors[10] for making this project possible. Additional recognition is given to the New York Times for uploading COVID-19 cases data to their GITHUB Repository.

6 References

1. Record of Delaware Actions:
   https://governor.delaware.gov/health-soe/
2. New York Times COVID-19 Cases Data:
   https://github.com/nytimes/covid-19-data/blob/master/us-states.csv
3. MATLAB fmincon Function:
   https://www.mathworks.com/help/optim/ug/fmincon.html
4. My GITHUB Repository:
   https://github.com/tatabas/delaware-covid-analysis
5. MATLAB Bootstrap Function:
   https://www.mathworks.com/help/stats/bootstrp.html
6. News Article about March 20:
   https://www.delmarvanow.com/story/news/local/delaware/2020/03/20/rehoboth-beach-dra
   2887892001/
7. Study outlining the incubation of the virus:
   https://pubmed.ncbi.nlm.nih.gov/32150748/
8. Chicken Factory news:
   https://www.foxnews.com/food-drink/delaware-chicken-company-attendance-kill-chicken
9. MassGen study:
   https://jamanetwork.com/journals/jama/fullarticle/2768533
10. Swarthmore College Summer Research Program:
    https://www.swarthmore.edu/summer-research-opportunities