COVID-19 Outcome Prediction Based on Mobility Measures in Virginia

Anna Brower (aeb2de) and Claire Dozier (ced9mq)

Motivation

COVID-19 is a contagious respiratory disease that was first identified in Wuhan in 2019. This highly contagious disease spread quickly in China and soon to other countries. In January 2020, COVID-19 was declared a Public Health Emergency of International Concern, and by March, the outbreak was officially declared a pandemic. As of December 9, 2020, there have been more than 68.5 million confirmed cases and over 1.5 million deaths that have been attributed to COVID-19.[1] With the highly contagious nature of the disease and the many unknowns surrounding COVID-19, including what the long term health consequences will be, it is important to identify which policies work most effectively to reduce the spread of and deaths from COVID-19.

Countries around the world have implemented different policies in order to reduce deaths and to mitigate the spread of the disease in hope that hospitals will not be overwhelmed and fewer people will be infected. Some of these policies include social distancing, wearing of masks, increasing testing capacity, and isolation of those that have been infected. Other countries such as Sweden have opted for a different approach with fewer restrictions and no shutdowns with the aim of developing "herd immunity." Examining which of these different approaches and policies are most effective in reducing the spread of COVID-19 can aid governments in making their own policy decisions in hopes of keeping as many people healthy and alive as possible.

Background

The totality of efforts cannot be described here, but almost all agencies within the United States federal government are working in some way to assist Americans in surviving the financial collapse brought on by the pandemic, as well as conducting research on therapeutics and vaccines, and providing numerous other forms of support.[2] However, despite these efforts, there has been no nation-wide mask mandate or lockdown, and most implementations of COVID-19 prevention and mitigation measures are being led by individual states. As a result, each state has unique restrictions in place and there is little cohesion in policy at the national level. Additionally, as the pandemic has progressed, there have been marked differences in the ways in which different states have opted to handle the virus, which has produced drastically different outcomes. For example, states like New York and Virginia implemented tough lockdowns and restrictions early on, while others have taken little to no action and seen large jumps in cases in recent months. Due to this general lack of cohesion at the national level, we have chosen to narrow our analysis of the efficacy of different policies to the state of Virginia.

The choice has been made to limit the scope of this study to the state of Virginia for several reasons: (1) data for states in the United States is widely available and reputable (through sources such as the Centers for Disease Control and Prevention, Johns Hopkins University of

Medicine, and the National Institutes of Health), (2) choosing one state of interest enables simpler design of the study, and (3) Virginia implemented numerous social distancing restrictions in the early days of the pandemic but has since gone through a phased re-opening and seen cases rise steadily in the month of December.

Beginning on February 15, 2020, Google began to compile COVID-19 Community Mobility Reports for each state and the counties within states. These reports consist of aggregated and anonymized data on changes in movement trends over time by geography across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential.[3]

In addition to the Google COVID-19 Community Mobility data, we will be utilizing additional data from the COVID-19 Data Repository managed by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University on GitHub.[4] This data includes daily-updated time-series data about the number of confirmed COVID cases and number of confirmed COVID deaths across all counties and states in the United States since January 22, 2020. This data will be used to calculate the case fatality rate in Virginia over time since the start of the pandemic. Additionally, in order to calculate the infection rate over time we will use testing data from The COVID Tracking Project.[5]

The focus of this project is to utilize the Google COVID-19 mobility data, as it pertains to the state of Virginia, in conjunction with additional data sources for COVID-19 severity metrics, in order to forecast these metrics and gain insight on the efficacy of policies that limit public mobility.

Related Work

Chaudry et al. examined the impact that government actions, country preparedness and socioeconomic factors had on the mortality and number of cases of COVID-19. They examined COVID-19 deaths and cases from 20 different countries and attempted to predict these factors by using a multivariable negative binomial regression. They found that more cases were in countries with higher obesity, median population age, and longer time to border closures from the first reported case. As well, they found that there were significant increases in mortality with higher obesity and per capita gross domestic product. Overall, they found that low levels of national preparedness and lower testing were associated with increased national COVID-19 cases and mortality.[6]

Since the start of the pandemic, Johns Hopkins University of Medicine has been tracking the implementation of COVID-19 policies, cases, and deaths in different states. This data illustrates the time-series trends across the United States and is useful to gauge how different types of policies and when they are implemented has impacted the containment, or lack thereof, of the virus.[7]

In August 2020, the Pew Research Center published an analysis of public reception in different countries to government COVID responses. With respect to this report, it is notable that of those polled in the United States, only 47% said that the country had done a good job of handling the virus (these responses are very politically polarized). Additionally, in the U.S. 77% of respondents claimed that the country is more divided now than prior to the start of the pandemic. This study is useful for gaining a more holistic understanding of the impact that implementing different policies can have on members of different countries. Our modeling may show that policies affecting mobility are effective and useful for forecasting, but public resistance to and acceptance of such policies must be considered if they are to be successfully implemented and enforced.[8]

In their paper, “Vector Autoregressive Models For Multivariate Time Series Analysis On Covid-19 Pandemic in Nigeria,” Ajo, Awogbemi, and Ilugbusi used Vector Autoregression (VAR) models to model and forecast confirmed cases, new cases, and total deaths variables in Nigeria. The paper demonstrates that VAR models might be a particularly useful tool for public health officials trying to forecast future trends in infections and deaths during the pandemic.[9]

Voko and Pitter in their paper, “The effect of social distance measures on COVID-19 epidemics in Europe: an interrupted time series analysis,” conducted an interrupted time series analysis to characterize the changepoint in the COVID-19 pandemic in different European countries. This research utilized the aforementioned Google COVID-19 mobility data to calculate a social distancing index and interrupted time series analyses using STATA software.[10]

Claim / Target Task

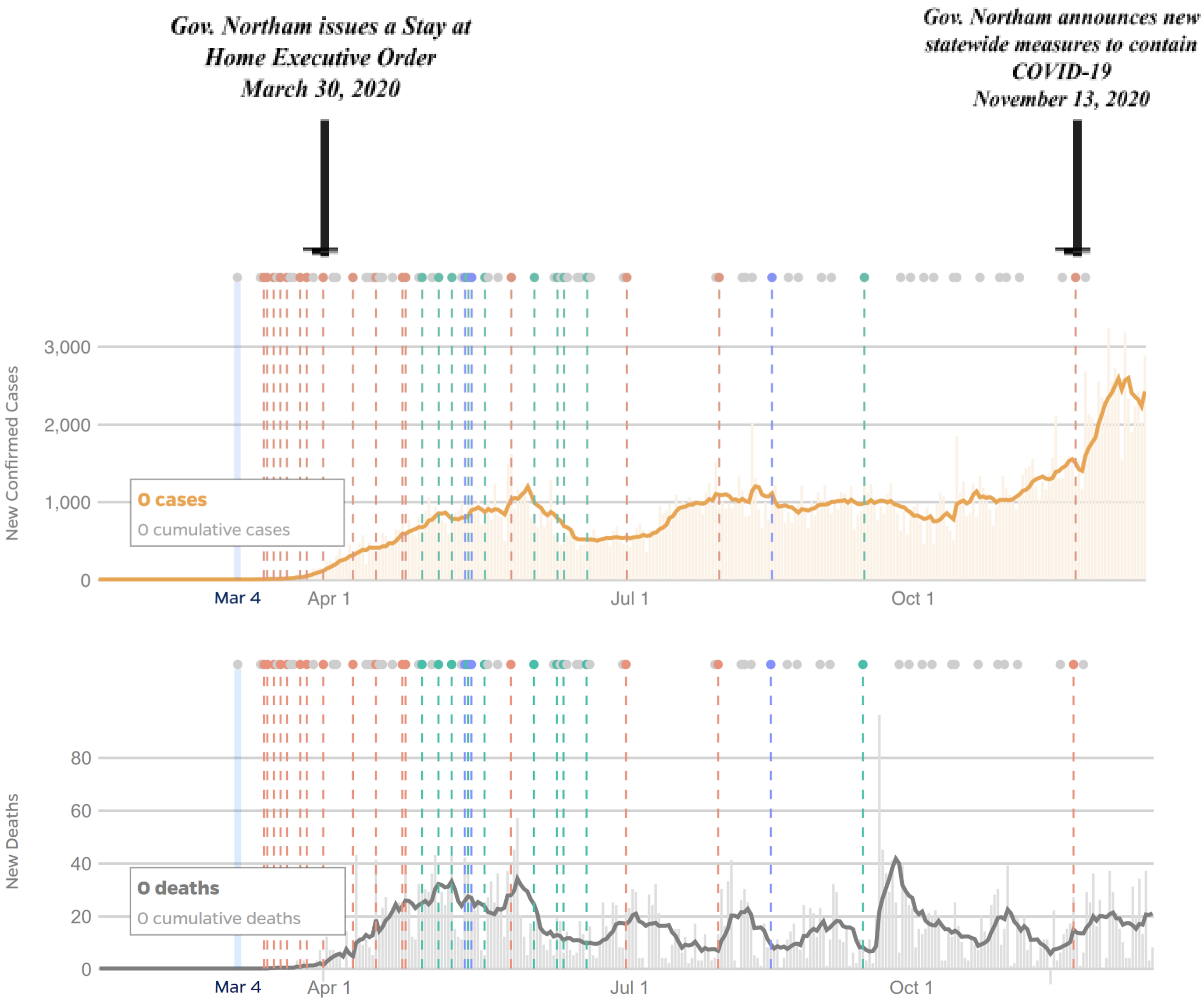
For this project, the forecast of outcomes (as measured by COVID-19 severity metrics) over time is reasonable and extremely necessary because, as the winter months are fast-approaching and the number of cases continues to rise precipitously, being able to forecast these metrics will be essential in convincing policy-makers to take action and implement the most effective prevention measures. As it currently stands, there has been great resistance to implementing nation-wide policies and restrictions in order to contain the spread of COVID-19.

If this study can demonstrate that publicly-accessible mobility measures and testing/death data can help to forecast the chosen COVID-19 metrics accurately using a VAR model, this research could provide beneficial information to public health officials and decision makers and aid them in enacting more successful virus mitigation efforts.

Intuitive figure showing why the Target Task is necessary and important

The two figures below are from JHU and show the new confirmed cases and new deaths per day from COVID-19, as well as the timing of policy choices to combat the spread of COVID-19 in Virginia.[7] On March 30, 2020, Virginia Governor Ralph Northam announced a Stay at Home Executive Order for non-essential workers that was in effect until June 10, 2020. Violating this order was a class 1 misdemeanor. It appears that this enforcement of stay-at-home order was largely effective, with the number of new COVID-19 cases tending to remain under 1,000 per day into October. The number of deaths per day has also remained fairly low. However, beginning in early November the number of new confirmed cases per day has steadily been rising (the number of new deaths has remained fairly steady). On November 13, 2020, Gov.

Northam announced new restrictions that include limiting gatherings to 25 people or less, an expanded mask mandate, on-site alcohol curfew, and increased enforcement in an effort to combat the rising cases. Our target task is to use vector autoregressive time series analysis to formulate a model that can forecast future infection and mortality rates and, in doing so, motivate public officials and citizens alike to take action to limit the spread of COVID-19.



Proposed Solution

Our team proposes that a VAR model is appropriate for modeling and forecasting COVID-19 infection and mortality rates using mobility data as additional endogenous predictors. VAR models are capable of handling multivariate time series data and have been shown to be one of the most flexible and accurate types of models for this purpose. They are especially effective at describing dynamic system behavior, as is the case with COVID-19 data.[9] The basic structure of a VAR model is that a variable is a linear combination of past lags of itself and past lags of other endogenous variables:

$$Y_t = v + A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + u_t \text{ where } u_t \text{ is appx. Normal,}$$
$$A_i \text{ is a KxK coefficient matrix, and p is the lag order (model hyperparameter)}$$[11]

All variables are treated as endogenous variables and are used to predict one another. In this study, eight different equations were produced by the model, and the two of primary interest were the equations for the infection_rate variable and the equation for the death_rate variable.

Team Contributions

I. Recommended a suitable modelling approach for handling multivariate, quantitative time series data relating to COVID-19.
II. Applied this versatile modeling approach to publicly accessible mobility, infection, and death data using the Python *statsmodels* package.
III. Demonstrated forecasting capabilities/accuracy of a VAR model for COVID-19 outcomes using selected endogenous variables, thus providing clear benefit to public health officials and decision-makers who must be able to predict the course of the pandemic and implement or reimplement effective policies to contain the spread.

Data Summary

To begin, we had to determine the COVID-19 mortality rate using the Johns Hopkins COVID-19 confirmed case counts and mortality counts per day. First, we filtered the Johns Hopkins data so that it only contained the data from the state of Virginia. Then, for both the case and mortality counts we summed all of the counties in Virginia in order to find the total number of cases and fatalities per day. These values were used to determine the overall mortality rate.

The same procedure was conducted for the infection rate. The COVID Tracking Project provides daily data on the total number of tests conducted and the total number of positive tests by state. To find the infection rate, we divided the number of positive tests by the total number of tests conducted per day.

The final data set used was the Google COVID-19 Community Mobility Reports. This data contains information about the percent change in mobility trends. These values are aggregated and anonymized and contain information on categories of places including retail, recreation, groceries and pharmacies, parks, transit stations, workplaces, residential. This data is given by state for each day, so once it was filtered to contain only Virginia values, the COVID-19 outcome metrics were merged with the Google mobility data by date.

The final data frame contained the Google mobility data for each of the six locations (retail, recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential) and the two different COVID-19 outcome metrics, infection rate and fatality rate. The training and testing data were split by date so that the VAR model could forecast the infection rate and mortality rate for future data. The training data is from March 15, 2020 to November 20, 2020, and the testing data goes from November 21, 2020, to December 4, 2020. The March 15, 2020, cutoff for the training data ensured that the data did not contain disproportionately high infection and mortality rates during the beginning stages of the pandemic due to the lack of widespread testing, and the November 20th cutoff allows enough testing data to determine if our model is able to accurately predict the infection rate and mortality rate. The use of this easily accessible, real world time series data makes it a good choice to use for a VAR model that can aid in predicting the course of the pandemic and provide another consideration for deciding future policy when it comes to controlling the spread of COVID-19.

Implementation

To begin, we cleaned and prepared the data. This step in the process involved calculating the infection rate and the mortality rate, as well as merging the three datasets, the Google COVID-19 Community Mobility Reports, the COVID Tracking Project, and the Johns Hopkins COVID-19 time series data, by date. Once the data had been organized into the appropriate manner, we began the process of building a VAR model in order to predict mortality and infection rate using the mobility data as an indicator of policy efficacy.

The VAR model has several assumptions, the first of which requires that the time series are stationary, meaning the properties of the series do not depend on the time at which the series is observed. To check for stationarity we implemented and analyzed the results of the Augmented Dickey-Fuller (ADF) Test. This is one of the most widely used unit root tests and can determine how strongly a time series is defined by a trend. The hypothesis for the test is as follows:

H_0: The time series has a unit root, meaning it is non-stationary
H_1: The time series does not have a unit root, meaning it is stationary

After running the ADF test for the first time, we observed that none of the eight different time series were stationary at a significance level of 0.05. The VAR class in the statsmodels package assumes that the time series are stationary for modeling. As a result, we chose to transform the data to make it stationary by taking first-differences. After differencing once, we found that all of the time series became stationary.

The next requirement for a VAR model is that “each of the time series in the system influences each other. That is, you can predict the series with past values of itself along with other series in the system.”[12] To test for this causation, we did a Granger causality test on the stationary data. The hypothesis for this test is that:

H_0: The past values of the time series (X) do not Granger-cause the other series (Y).
H_1: The past values of the time series (X) do Granger-cause the other series (Y).

From this test, we concluded that almost all of the variables in the system were interchangeably Granger-causing one other, which meant the system of multiple time series was appropriate to use a VAR model to forecast.

The next step involved determining the hyperparameter for the model, which for a VAR model is the lag order (p). In order to select the right order (p) of the VAR model, we iteratively fit increasing orders of VAR models and used the Akaike Information Criterion (AIC) metric to choose the best order model, which is the model order with the lowest AIC. From this step, we determined that a lag of 8 was most appropriate. (See the figure below for the AIC values for each lag order.)

```
Lag Order = 1
AIC :  -52.008546881947424
BIC :  -50.99145206851185
FPE :  2.588716767945276e-23
HQIC:  -51.59914949091318

Lag Order = 2
AIC :  -53.08333571882643
BIC :  -51.156616728994024
FPE :  8.850104290588633e-24
HQIC:  -52.307711588994216

Lag Order = 3
AIC :  -53.83261876571636
BIC :  -50.991008776543914
FPE :  4.199593534684066e-24
HQIC:  -52.68856434091248

Lag Order = 4
AIC :  -54.13694279440751
BIC :  -50.375123585114245
FPE :  3.120879655442439e-24
HQIC:  -52.622233151169844

Lag Order = 5
AIC :  -55.548649592155854
BIC :  -50.86125084497761
FPE :  7.701393204115712e-25
HQIC:  -53.66103815726308

Lag Order = 6
AIC :  -56.71969422315171
BIC :  -51.10129281202493
FPE :  2.4331890208502293e-25
HQIC:  -54.456912486465114

Lag Order = 7
AIC :  -56.8118387897965
BIC :  -50.25695805661424
FPE :  2.279573861952717e-25
HQIC:  -54.17159601232189

Lag Order = 8
AIC :  -56.86669347620239
BIC :  -49.369802494378106
FPE :  2.238928285938323e-25
HQIC:  -53.84667639327737

Lag Order = 9
AIC :  -56.85081777271644
BIC :  -48.406330597784724
FPE :  2.388746079005469e-25
HQIC:  -53.44869029222617

Lag Order = 10
AIC :  -56.82698104417961
BIC :  -47.429255951156236
FPE :  2.6063262114734817e-25
HQIC:  -53.04038393970211
```

The final steps in our process were building the model and making predictions for our testing data for mortality rate and infection rate. To fit the model, we used the lag 8 value that we found previously. Once the model was fitted, we checked for serial correlation in the model residuals using the Durbin Watson test. Serial correlation of residuals refers to any patterns that might remain in the residuals of the model following model training and fitting. If there is correlation remaining in the residuals, this might suggest that there patterns that were not accounted for. Thus, we used the Durbin Watson statistic as a measure of any remaining serial correlation in residuals. A value of 2.0 means that there is no autocorrelation detected in the sample. Since the output of the results of the test are very close to the desired 2.0 value, the serial correlation was acceptable and there are likely not significant patterns remaining in the residuals of the fitted model.

In order to generate a forecast, the VAR model requires up to the chosen lag order number of observations from the past data (in our case, 8 lags). Once the forecast input array was generated, this could be passed to the forecast function and the next 14 steps could be forecast. Following the forecast generation, the forecasts were inverted by rolling back the first-differences of the training data, which provided the actual forecast values. Following the predictions, we performed an analysis of our results.

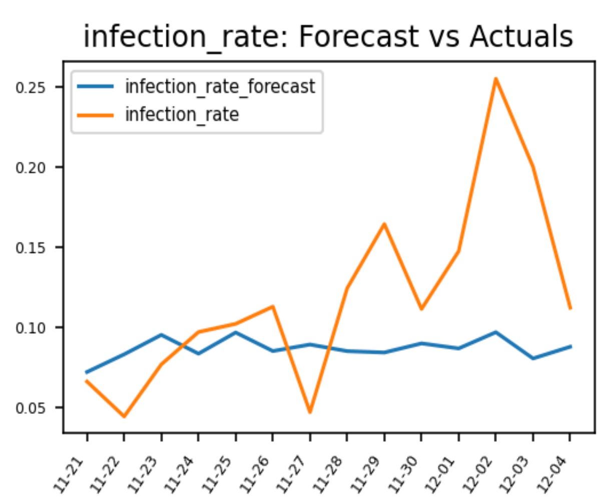
Technical challenges

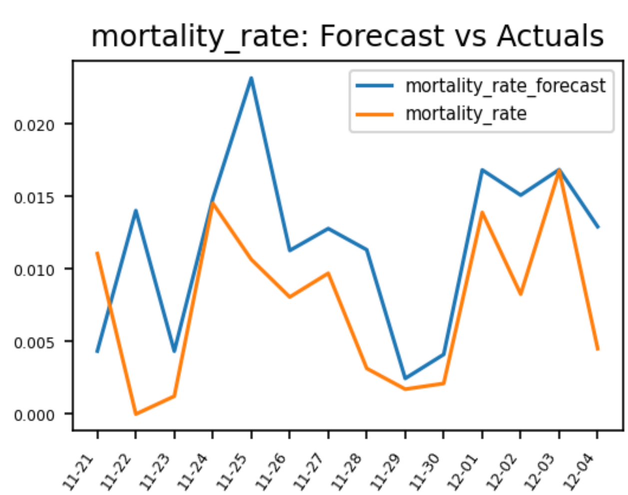
Our original plan involved modeling the time at which different policies (that were one-hot encoded) were implemented to determine how the type of policy and when it went into effect impacted COVID-19 outcome metrics. However, we ran into a number of challenges with this plan. The first problem that we were concerned with was how only having one policy change would impact our model, especially since a number of the policies were implemented early on in the pandemic before there was widespread testing, and it would be harder to represent how that affected mortality rate and infection rate. We considered doing an interrupted time series analysis based on the implementation of several policies, but this idea became technically infeasible as we looked into the packages that were available to do time series analysis and found that Python was lacking in the statistical methods that would be needed to carry out this type of work.

Technical solution

Due to these technical challenges, we decided to shift our focus from using categorical policies to using the Google COVID-19 Community Mobility data. This data provided quantitative, easily-modeled indicators of social distancing policy efficacy. The data also allowed us to use a VAR model, which can be easily implemented through the Python *statsmodels* package. This allowed us to use a model that had more documentation and was more accessible than other models/ techniques that were previously considered.

Experimental Results





Based on our visualizations, it appears the forecast for the infection rate is quite steady and does not fluctuate to the degree the actual infection rate fluctuates. There might be a few reasons for this, including:

- Potentially weaker Granger-causal relationship between infection rate and the other endogenous time series predictors
- Additional predictors may be needed to more accurately forecast the infection rate (mobility data and mortality rate may not be sufficient)
- Infection rates are highly dependent on the number of tests conducted per day, and this can vary widely, making infection rate a more difficult metric to forecast

However, the model appears to perform much better at forecasting the mortality rate, with the forecast closely following the patterns of the actual mortality rate. The model does appear to slightly over-estimate the mortality rate on each day (making it a more conservative estimate).

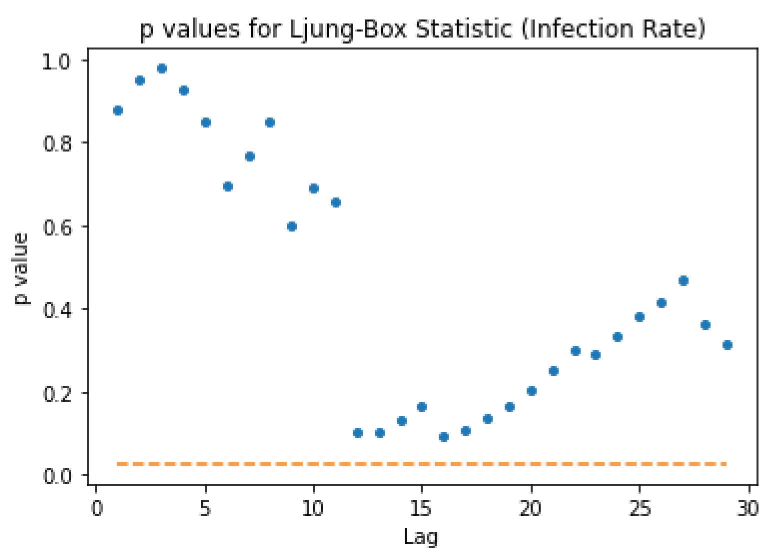
Experimental Analysis

The first step of our analysis involved conducting the Ljung-Box Test, which determines whether the first H sample autocorrelations of the residuals, considered together are significant. We do not want significance in this test as that would imply autocorrelation. This test offers an effective means of testing the adequacy of the model up to a certain number of lags. The Ljung-Box hypothesis can be defined in the following way:

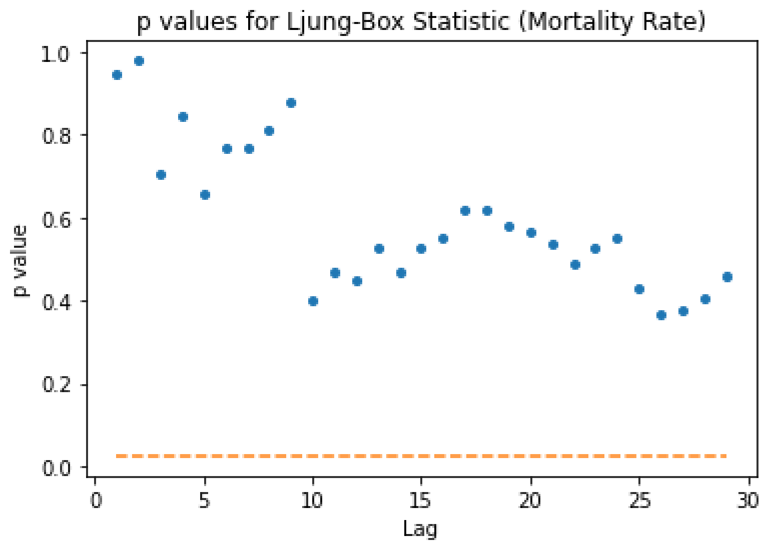
H_0: Autocorrelation is not significant (i.e. the model is adequate up to lag H)

H_1: Autocorrelation is significant (i.e. the model is not adequate up to lag H)

The below graphs display the p-values for the different lags of the residuals from the fitted model. The orange-dashed line indicates the significance level of p=0.05. If points fall below that line, then the model is not adequate for that lag. The higher the points are, the higher the p-value and the stronger the evidence to accept the null hypothesis.



For the infection rate p-values, it appears that they are quite high (above p=0.5) until lag 11. Following lag 11, there is a significant dip in p-values, but they rise fairly steadily up until lag 30. This visual analysis suggests that the model is fairly adequate at least until lag 11, and likely slightly beyond. This analysis is supported by the visualization of the forecast. For the first 11 days in the forecast (November 21, 2020, to December 1, 2020), the forecast is fairly close to the actual infection rate. The large deviation in the forecast begins after day 11.



For the mortality rate p-values, the p-values are well above p=0.6 up until lag 9 and they remain fairly high for all subsequent lags, never dipping far below p=0.4 up until at least lag 30. This suggests that the model is very adequate at forecasting the mortality rate. This is confirmed through the visual evaluation of the mortality rate forecast above, in which the forecast closely mirrors the actual mortality rate for the entire 14-day testing period.

The three statistical metrics used to evaluate how well our forecasts predicted the actual data were mean squared error (MSE), root mean squared error (RMSE), and correlation. MSE and RMSE provide simple measures of the imperfection of the fit of the estimator of the data, and the correlation refers to the existence of some statistical relationship between the forecasted and actual data values. The following values are the results found:

Forecast accuracy of infection rate:

- MSE: 0.0040523103
- RMSE: 0.0636577592
- Correlation: 0.2264818128

Forecast accuracy of mortality rate:

- MSE: 4.47247e-05
- RMSE: 0.0066876498
- Correlation: 0.5593556241

For the infection rate forecast, the MSE and RMSE were both quite small (under 0.1), which suggests that the model did an adequate job of estimating the infection rate. However, the correlation coefficient is only about 0.23, which might indicate that there are other variables that are predictive and are missing in the model. It should be noted that because real data is used, there is likely noise and the data is not perfect.

For the mortality rate forecast, the MSE and RMSE were both incredibly low (less than 0.05), which suggests that the VAR model provided a fairly accurate forecast of the mortality rate. Additionally, the correlation coefficient is much higher for the mortality rate than the infection rate, indicating that there seems to be evidence of moderately significant correlation between the selected variables and the mortality rate.

Conclusion

This research demonstrated the application of an accessible and adaptable VAR model from the Python *statsmodels* package to COVID-19 outcome forecasting. This type of model has the potential to be extremely useful to multiple stakeholders including healthcare providers, government decision makers, and the general public. Containing the spread of COVID-19 requires being able to forecast its course and implement effective policies accordingly, such as increased social distancing, mask mandates, and stay-at-home orders.

Future Work

There are several ways in which this work can be continued and expanded upon:

1. Reproduce these results with data from other states to determine how well this approach generalizes. Additionally, it may be helpful to take a lower-level approach and apply this model to county-level (instead of state-level) data. This might be particularly beneficial for large counties (e.g., Fairfax Country, Virginia).
2. Utilize an alternative programming language (specifically tailored to statistics), such as STATA, to conduct an interrupted time series analysis of different policy implementations (such as mask mandates, stay-at-home orders, and school closings) to provide a more holistic understanding of policy impact on state-level COVID-19 outcomes.
3. In addition to the endogenous predictors utilized in this research, it may be necessary to include other predictors in the model, particularly to predict infection rate.

References

1. https://coronavirus.jhu.edu/map.html
2. https://www.usa.gov/coronavirus
3. https://www.google.com/covid19/mobility/
4. https://github.com/CSSEGISandData/COVID-19
5. https://covidtracking.com/data/download
6. https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(20)30208-X/fultext
7. https://coronavirus.jhu.edu/data/state-timeline
8. https://www.pewresearch.org/global/2020/08/27/most-approve-of-national-response-to-covid-19-in-14-advanced-economies/
9. https://abjournals.org/african-journal-of-biology-and-medical-research-ajbmr/wp-content/uploads/sites/17/journal/published_paper/volume-3/issue-2/AJBMR_O2OW66XY.pdf
10. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7288252/
11. https://www.statsmodels.org/stable/vector_ar.html#
12. https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/