Richard Y Liu

Stat 410 Final Project

Due April 24, 2020

**The Economic Impact of Pandemics and Implications for Covid-19**

**Introduction**

In this paper I aim to analyze the economic impacts of different pandemics and epidemics in our recent history, namely, H1N1 or Swine Flu and West African Ebola. With this analysis, I will regress different aspects of these pandemics against their economic impact and extrapolate with the Covid-19 data and generate a best case and worst case scenarios in terms of economic impact. I decided to investigate this topic due to the unprecedented effects of Covid-19 on our lives, specifically, how it has led to a severe economic downturn unlike anything that we’ve experienced in the past.

**Data Description**

*H1N1 or Swine Flu***.** This data came from Kaggle, and was collected by WHO through their various collaborations with governments around the world. These governments provide data in a variety of ways, through non-profit organizations or federal institutions. This is the most recent pandemic that we have had, while the other two viral outbreaks are classified as epidemic, so this disease would be the most relevant to our comparison to Covid-19. The final statistics of this disease were 60.8 million cases with 12,469 deaths, resulting in a mortality rate of 0.02%. This mortality rate is significantly lower than that of our current Covid-19, however the scale at which this impacted the world is the significant fact that we can use to compare economic impacts, as so many cases resulted in a higher economic impact.

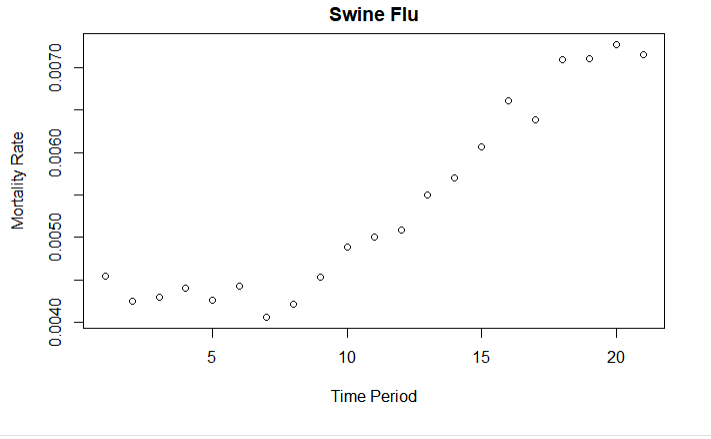
*West African Ebola***.** This is the most widespread outbreak of Ebola virus disease in history, and caused major loss of life and heavily disrupted social and economic patterns in Guinea, Liberia, and Sierra Leone. The final statistics for the Ebola outbreak were 28,616 confirmed cases and 11,310 deaths, resulting in a mortality rate of 40%. While this disease was not nearly as widespread as Covid-19 currently is, we are already at 2.8 million cases worldwide and 196,000 deaths, the much higher death rate is an important outlier that will allow us to examine the local economic impact of this disease, and possibly extrapolate to the rest of the world. I do recognize that health care within these three nations is severely different from that of the rest of the world, but I believe this to be an effective comparison in a worst case scenario, due to the myriad of issues that we have had thus far in combatting Covid-19.

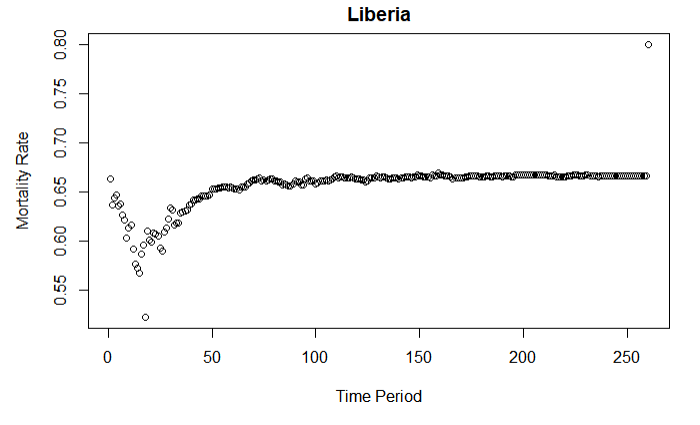
*Covid-19.* Our current pandemic that has resulted in most of the world governments shutting down cross country travel as well as heavily restricting inter-state travel. This data is the most recent of our data as well as the most detailed. This data will be the data that we extrapolate from in order to determine the magnitudes of the possible economic costs of Covid-19. As of today, the current number of cases is 2.8 million with 196,000 deaths, resulting in a mortality rate of 7%. However, this number can be deceiving due to the fact that a large number of these cases are unresolved. If we were to look at only the resolved cases, which number 974,000, the mortality rate jumps to 20%, a statistic that look dangerously close to our Ebola numbers rather than our Swine Flu numbers.

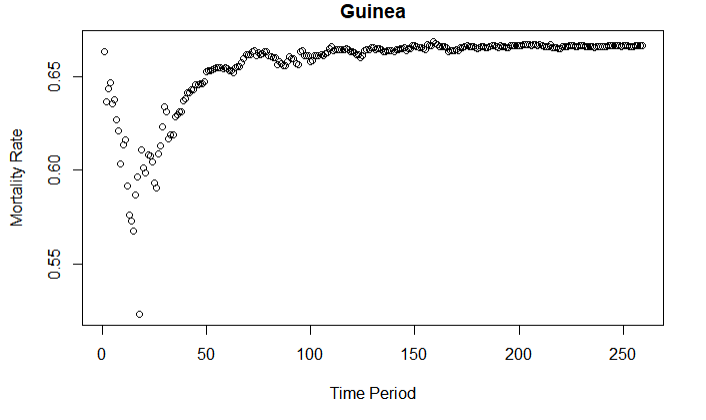
*GDP.* We will be using the GDP’s of various countries that have been impacted by these diseases. So for the Swine Flu, we will be using the US GDP data before, during and after the resolution of the disease. This disease occurred during the period 2009-2010. For the West African Ebola Virus, we will be using GDP data for Guinea, Liberia and Sierra Leone, the three nations that were most affected by this virus. This disease occurred during the period 2014-2016.

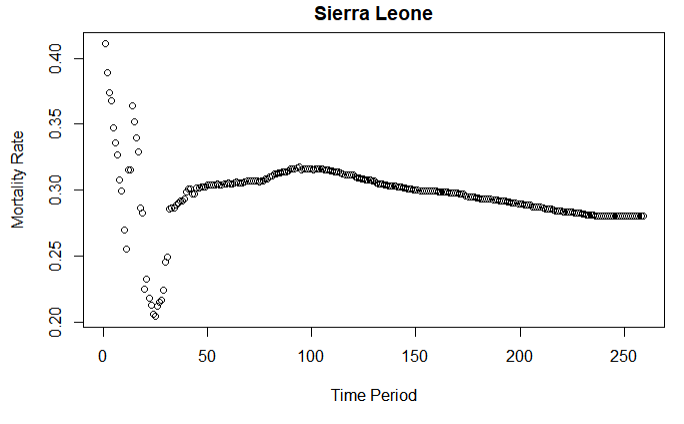
**Exploratory Data Analysis**

*H1N1 or Swine Flu (2009 – 2010).* The data as it was provided details the country of interest, the number of cases, the number of deaths, and the date at which these statistics were recorded. The first most important aspect of this data is that it only covers a 1.5 month period from May 23rd 2009 to July 6th 2009, arguably the virus’s most significant period. The virus itself took place for a whole year. The variable are the number of cases, the number of deaths, country of interest, and the date of recording. The number of cases and number of deaths can be combined in order to determine the mortality rate of the virus at a certain time within a certain country. Another important observation of the data is that the dates are backward, i.e. most recent dates are first in the file. We notice a clear upward almost linear trend during these 1.5 months that lead us to believe that this disease will get much worse much more quickly.



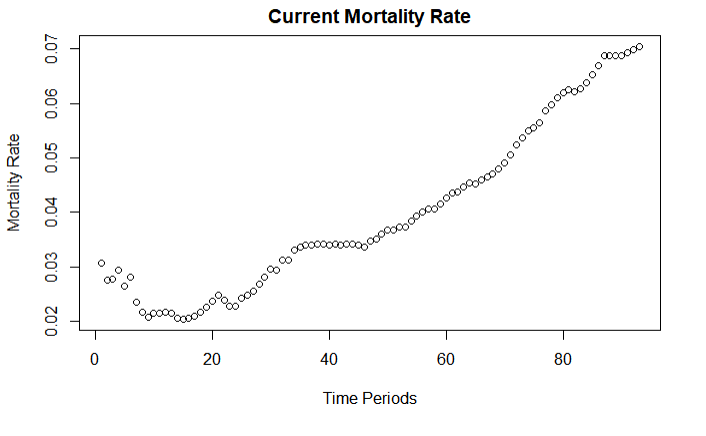
*West African Ebola (2014-2016).* In the 5 nations that are provided in the data, there is a noticeable trend in 4 of the nations, where there is scattered data in the beginning of the disease’s spread, but then the mortality rates settle down into a comfortable average as time goes on. This is due to the nature of the data being cumulative, as more and more cases are being added the overall average is being solidified. In the case of Guinea and Liberia, that mortality rate is 65%. However for Sierra Leone, the mortality rate seemed to be on a downward trend towards the end of the spread, settling in at around 27%. 





A strange aspect of the data was that Nigeria’s data seemed wrong. After a few periods of time, the mortality rate settled in at exactly 40% with no variation for the 200+ periods, indicating some anomalies in the data that I can’t explain. This is why Nigeria isn’t not being included in the analysis, due to the abnormality of the data provided.

*Covid-19.* This is our most current data, however it is also the one with the most unusual patterns. The first graph is a graph of mortality rates over time, but the second graph is one of mortality rates over resolved cases only over time. In the first graph, there’s a clear linear trend that indicates that this pandemic has no sign of slowing down while in the second graph it seems that the mortality rate amongst resolved cases has settled at around 23%, which would imply that when this pandemic is resolved, the mortality rates of all cases would be 23% or, as of our current numbers, result in a minimum of 644,000 deaths, if the disease were to stop spread at this instant.



*GDP.* The variables for the GDP data are real GDP and year. Amongst the 5 graphs, world, US, Guinea, Liberia and Sierra Leone, 3 of them share the common feature that their economy take a downturn at around the 2008 mark and starts to climb again after 2009/2010. While it would be nice to attribute this solely to Swine Flu, there was the much more likely factor of the depression of ’08 that would cause these dips. An interesting note however is that neither Sierra Leone nor Liberia have dips during the Swine Flu or ’08 Depression period, but both Guinea and Sierra Leone both have dips during 2014, when we expect them to due to the Ebola outbreak. Strangely enough, Liberia is the country that seems to be least impacted, with no noticeable dip during the Swine Flu and only a slowdown in growth during the Ebola disease. These are not results I would’ve expected given the mortality rates of Ebola during this time.

**Data Analysis**

Models:

GDP ~ cases + deaths + mortality rate

GDP growth rate ~ cases + deaths + mortality rate

I am using this multiple linear regression model in order to determine which of these parameters has the greatest effect on GDP and the growth rate of GDP. I’m using both GDP and growth rate because GDP in an absolute sense may not be able to paint a clear picture while impact on growth rate may be more pronounced, due to a temporary halt or slowdown in the workings of an economy. As for why countries was not included in these regressions, there were too many countries that would need to be sampled and none of the data came organized by region of the world, and it would’ve been too tedious to try and sort them all properly. I am also doing this model only for Ebola, due to the existence of the ’08 Depression, which I believe will cancel out or be greater in magnitude than the effect of the Swine Flu on the economy.

The model diagnostics should inform us how much each parameter affects GDP and GDP growth rate, which do not contradict our assumptions that there is an effect. The results we got from the analysis are as follows.

GDP Guinea

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.792e+09 4.398e+05 19990.867 <2e-16 \*\*\*

cases -2.506e+02 2.294e+02 -1.092 0.2748

deaths 1.016e+03 6.337e+02 1.603 0.1091

mort\_rate -2.783e+06 1.210e+06 -2.299 0.0216 \*

GDP Growth Guinea

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.844e+00 1.539e-02 249.777 <2e-16 \*\*\*

cases 1.199e-06 8.026e-06 0.149 0.881

deaths -2.827e-06 2.218e-05 -0.127 0.899

mort\_rate 5.304e-03 4.235e-02 0.125 0.900

GDP Sierra Leone

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4286051064 7821758 547.965 < 2e-16 \*\*\*

cases 8809 4079 2.159 0.030908 \*

deaths -38127 11270 -3.383 0.000728 \*\*\*

mort\_rate 107366587 21522802 4.989 6.51e-07 \*\*\*

GDP Growth Sierra Leone

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.829e+01 2.486e-01 -73.547 < 2e-16 \*\*\*

cases 2.949e-04 1.297e-04 2.274 0.023024 \*

deaths -1.262e-03 3.582e-04 -3.521 0.000437 \*\*\*

mort\_rate 3.536e+00 6.842e-01 5.168 2.56e-07 \*\*\*

GDP Liberia

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.175e+09 3.926e+05 8086.593 < 2e-16 \*\*\*

cases -3.352e+02 2.047e+02 -1.637 0.101720

deaths 1.478e+03 5.656e+02 2.612 0.009048 \*\*

mort\_rate -4.191e+06 1.080e+06 -3.880 0.000107 \*\*\*

GDP Growth Liberia

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.452e-02 7.709e-03 7.073 1.97e-12 \*\*\*

cases 7.328e-06 4.020e-06 1.823 0.06846 .

deaths -3.210e-05 1.111e-05 -2.890 0.00388 \*\*

mort\_rate 9.083e-02 2.121e-02 4.282 1.92e-05 \*\*\*

For Guinea, the model seems to work the least well, with the only predictor with some accuracy being the mortality rate in predicting the GDP. For Sierra Leone, and Liberia, both deaths and mortality rate are statistically significant when predicting the GDP and the GDP growth rate which would indicate a significant level of effect due to deaths and mortality rate in those two countries.

These are the prediction intervals of gdp and gdp growth using these models and inputting cases = 3 million, deaths = 210000, and mort\_rate = 0.07, the current statistics for Covid-19

fit lwr upr

1 8253453324 7151628517 9355278131

fit lwr upr

1 6.848099 -31.70996 45.40616

fit lwr upr

1 22713462630 3117804690 42309120571

fit lwr upr

1 601.8127 -21.0919 1224.717

fit lwr upr

1 2478992726 1495500425 3462485027

fit lwr upr

1 15.30394 -4.009358 34.61723

And these are the prediction intervals of GDP and GDP Growth using these models and inputting cases = 1050000, deaths = 210000, mort\_rate = .2, which are the current statistics for resolved cases for Covid-19.

fit lwr upr

1 8741670351 8506047579 8977293123

fit lwr upr

1 4.510684 -3.734871 12.75624

fit lwr upr

1 5550116474 1359628828 9740604119

fit lwr upr

1 27.17249 -106.0343 160.3793

fit lwr upr

1 3132048185 2921730555 3342365815

fit lwr upr

1 1.025879 -3.104226 5.155984

The GDP numbers seem a little harder to understand, but the GDP growth numbers are a little easier to interpret. According to these models, when put into the scenarios of Covid-19, growth rates are predicted to be anywhere between -32% and 45% growth rate for Guinea, -21% and 1224% growth for Sierra Leone, and between -4% and 35% growth rate for Liberia. Now this analysis and conclusion seem counter intuitive and lead me to question whether this model is truly and accurate method for predicting the economic impacts of a coronavirus situation.

**Summary and Discussion**

As of writing this conclusion, I have some positive and negative conclusions about my analysis. On the positive side, deaths and mortality rate do seem to have some level of reliability in predicting the economic effects of their own diseases and within their own countries. However, on the negative side, it would be foolish to attempt to use models that are designed for detecting the economic impact of Ebola in Guinea and try to use them to predict the effects of an entirely different disease such as Covid – 19 and in a different country, such as the United States. Which brings me to the main limitations of this analysis. There simply wasn’t enough time or data where there could be. The main areas that can be improved are more detailed GDP and GDP growth numbers, not on a yearly basis, but on a quarterly or even monthly basis. This nuance in data would allow a more accurate regression and determination of whether an effect exists or not. Another area that can be improved is the organization of the data itself. For example, in the Covid data set, all the countries were arranged alphabetically, rather than by region, or by simple cumulative statistics. So instead, you would be forced to organize the data yourself into a format that you can work with, and there simply wasn’t enough time to do that. A more organized dataset would’ve yielded more results in this analysis as well.

**References**

* <https://www.worldometers.info/coronavirus/>
* [https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.MKTP.CD&country=GIN#](https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.MKTP.CD&country=GIN)
* <https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=GN>
* <https://www.worldbank.org/en/search?q=summary+on+ebola+recovery+plan%3A+Sierra+Leone&currentTab=1>
* <https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/cost-of-ebola.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fvhf%2Febola%2Foutbreaks%2F2014-west-africa%2Fcost-of-ebola.html>
* [https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.MKTP.CD&country=BRA#](https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.MKTP.CD&country=BRA)
* <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=US>
* <https://www.kaggle.com/cdc/zika-virus-epidemic>
* <https://www.kaggle.com/imdevskp/ebola-outbreak-20142016-complete-dataset>
* <https://www.kaggle.com/de5d5fe61fcaa6ad7a66/pandemic-2009-h1n1-swine-flu-influenza-a-dataset>
* <https://www.kaggle.com/imdevskp/corona-virus-report/data>
* <https://www.livescience.com/worst-epidemics-and-pandemics-in-history.html>
* <https://fred.stlouisfed.org/series/GDP>
* <https://wwwn.cdc.gov/nndss/data-and-statistics.html>
* <https://www.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html>
* <https://www.ncbi.nlm.nih.gov/books/NBK525302/>