Project Proposal: Jiner Zheng

Predicting Excess Mortality During COVID-19

Oct 8, 2021

Excess mortality is defined as the increase in deaths from all causes during a crisis relative to 'normally' expected deaths. Researchers have proposed that excess mortality is a more reliable measurement compared to the number of deaths purely attributed to a pandemic, because the number of deaths caused by diseases may be undercounted from misdiagnosis and variance in governments' testing capacity and reporting policy. In this project, I am interested in examining the social burden resulted from COVID-19 using excess deaths as the indicator and also detecting cross-country differences in excess mortality under consideration of different countries' policy responses on the pandemic. The following four countries were chosen specifically according to different levels of stringency in governmental interventions and different extents of how much COVID-19 has spread:

- **United States:** *stringency index = 61.57, total number of confirmed cases = 45,030,695, total number of COVID-19 deaths = 730,413*
- **United Kingdom:** *stringency index = 41.2, total number of confirmed cases = 8,081,300, total number of COVID-19 deaths = 137,541*
- **Brazil:** *stringency index = 51.39, total number of confirmed cases = 21,532,558, total number of COVID-19 deaths = 599,865*
- **Russia:** *stringency index = 52.31, total number of confirmed cases = 7,717,356, total number of COVID-19 deaths = 214,485*

The stringency index is a composite measure of governmental responses to COVID-19, with a scale from 0 to 100 (100=strictest) based on nine policy responses indicators including school and workplace closures, travel bans, restrictions on public gatherings, and other interventions (Ritchie et al., 2020). The number of confirmed cases and death cases is based on data updated until October 8th, 2021, 17:22 GMT.

The COVID-19 started in Wuhan, China and has become a global threat since it has infected billions of people across 213 countries, declared by the World Health Organization (WHO) as a global pandemic on March 11, 2020. Many researchers have constructed various forecasting models to estimate the COVID-19 spread and death cases aiming to help allocate

hospitalizations and governmental interventions based on predicted number of cases. Liu, Magal, Seydi, and Web (2020) formulated a mathematical model to estimate cumulative COVID-19 cases in China with consideration of some important features of this epidemic, including the time and extent of government interventions, reported and unreported cases, asymptomatic and symptomatic cases. Their model captured major factors that might affect the outbreak spread and indicated the importance of governmental restrictions against the infection spread. Besides, Salaheldin and Abotaleb (2020) used the exponential growth and ARIMA models to predict daily confirmed, recovered, and death cases in China, Italy, and the US. Their findings suggested that the number of confirmed and death cases in China would be increasing slightly while the growth rate of recovered cases in China would be much higher. Besides, the exponential growth model seemed to outperform ARIMA in predicting COVID cases, especially for data in Italy and the US in that this model captured the large number of increases in expected values for these Western countries.

Existing literature focusing on excess deaths during COVID-19 is relatively limited, but some researchers have proposed useful models to estimate excess mortality. For instance, Giattino et al. (2021) used a regression model to get the expected number of deaths in 2020 under normal conditions using historical deaths data from 2015-2019 for each country and then estimate excess deaths as *Excess deaths = Reported Deaths - Expected Deaths.* They also presented a measure of excess deaths for cross-country comparisons, the P-score of excess deaths:

$$P\text{-}score = \frac{Reported\ Deaths\ -\ Projected\ Deaths}{Projected\ Deaths} * 100$$

My project will build upon this method of estimating excess deaths as well as its comparable metric of *p-score.* Instead of fitting regression model for predicting expected deaths in all countries, I would like to use different forecasting methods including ARIMA, Holt's Winter method, Regression, and Long Short-Term Memory respectively for the four countries based on their different time series features.

The data used in this project comes from the World Mortality Dataset, which contains country-level data on deaths from all causes in 2015-2021 collected from various sources

including the Short-Term Mortality Fluctuations (SMTF) dataset from the Human Mortality Database (HMD), EuroStat and many other national open data sources (Karlinsky & Kobak, 2021). The preliminary data exploration showed some seasonality and trend in all four countries' deaths time series (as indicated in the decomposition plot from **figure 2**). More specifically, US deaths time series has a relatively clear trend, and Brazil deaths time series has relatively clear seasonality, while Russia deaths time series indicated short-term autocorrelation (as presented in the correlograms from **figure 1**). Moreover, we can tell from the decomposed plots in **figure 2** that the number of deaths from all causes increased sharply cross all four countries since the year of 2019 when the COVID-19 pandemic started.

Figure 1: Correlograms

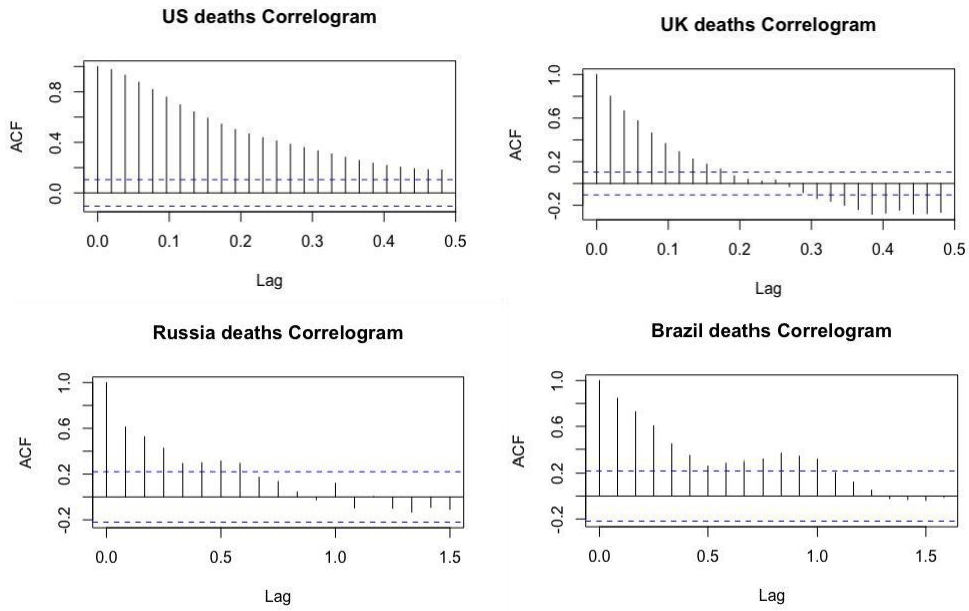
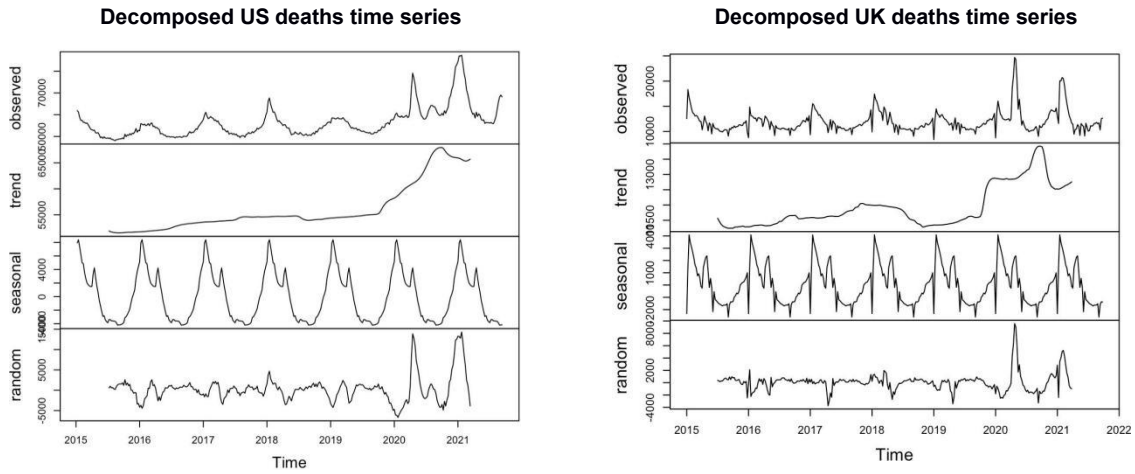
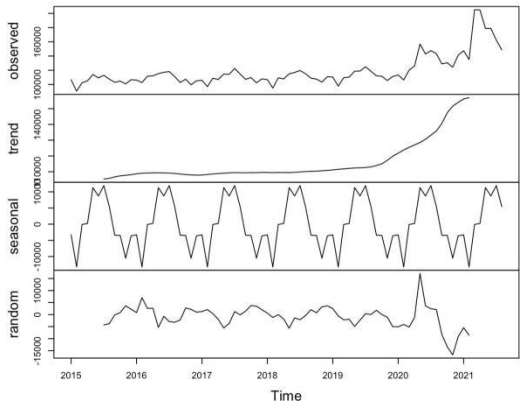


Figure 2: Decomposed Time Series Plots



Decomposed Brazil deaths time series



Decomposed Russia deaths time series