United States COVID-19 Hospital Resource Forecasting

Description

I will produce forecasts which show hospital bed use, need for intensive care beds, and ventilator use due to COVID-19 based on projected deaths for all 50 U.S. states. These projections are produced by models and data from IHME (Institute for Health Metrics and Evaluation) based on observed death rates from COVID-19, and include uncertainty intervals. These forecasts were developed in order to provide hospitals, policy makers, and the public with crucial information about how expected need of resources align with existing resources, so that cities and states can best prepare.

Hypothesis

NULL Hypothesis (H_0): Input time series is stationary
Alternate Hypothesis (H_1): Input time series is non-stationary
The confidence level will be at 95%.

After this is verified by different statistical tests, I will check the forecasts by comparing them with the data given by IHME and see how closely they correlate.

Proposed Approach

IHME's database has three columns for each dimension. One representing the mean of the dimension in different locations distributed by date, second having lower uncertainty bound, and third having upper uncertainty bound corresponding to each dimension.
In this database, we are targeting three dimensions:

- All beds - number of beds needed by day
    - Mean
    - Lower uncertainty bound
    - Upper uncertainty bound
- ICU beds - number of intensive care unit beds needed by day
    - Mean
    - Lower uncertainty bound
    - Upper uncertainty bound
- Invasive Ventilators - number of intensive ventilation needed by day
    - Mean
    - Lower uncertainty bound
    - Upper uncertainty bound

All these columns are distributed by location and date. Clearly, each of these columns form a time series with time being represented by date column and value being represented by each of the above mentioned columns.

The proposed idea is to use time series forecasting using ARIMA and SARIMAX in order to see how many COVID beds, ICU beds, and invasive ventilators will be needed in the coming year and how far the predicted upper and lower forecast values go from the given upper and lower uncertainty bounds in the dataset.

Also, I'll be predicting the mean values of each of these dimensions for the upcoming year and these predictions can then help us and authorities prepare for additional amenities that should be available in times of need.

Evaluation (Testing for Accuracy)

For accuracy testing, I will implement the following:

- Comparing the observed data points and forecasts, as well as calculating the Mean Squared Error and Root Mean Squared Error. The less they are, the better the model.
- Running model diagnosis and observing the Q-Q plots and observing how much closer the points are to the theoretical quantiles, histogram of forecasts and residuals and seeing if they're having normal distribution.
- Checking the Akaike Information Criterion (AIC) value. After running grid search, the parameters which have lowest AIC value corresponding to them will be used for model building.
- After executing the above, I can then obtain quarterly, biannual, or annual forecasts and compare the already acquired data from the source for upcoming months and compare the same with the forecasts derived from the model.

Acknowledgements

1. IHME: COVID-19 Projections. (2020). Retrieved May 3, 2020, from https://covid19.healthdata.org/united-states-of-america
2. Hyndman, R. J., & Athanasopoulos, G. (2018, May 6). Forecasting: Principles and Practice. Retrieved May 10, 2020, from https://otexts.com/fpp2/arima.html
3. Perktold, J., Skipper, S., & Taylor, J. (2019). SARIMAX: Introduction¶. Retrieved May 10, 2020, from https://www.statsmodels.org/dev/examples/notebooks/generated/statespace_sarimax_stata.html