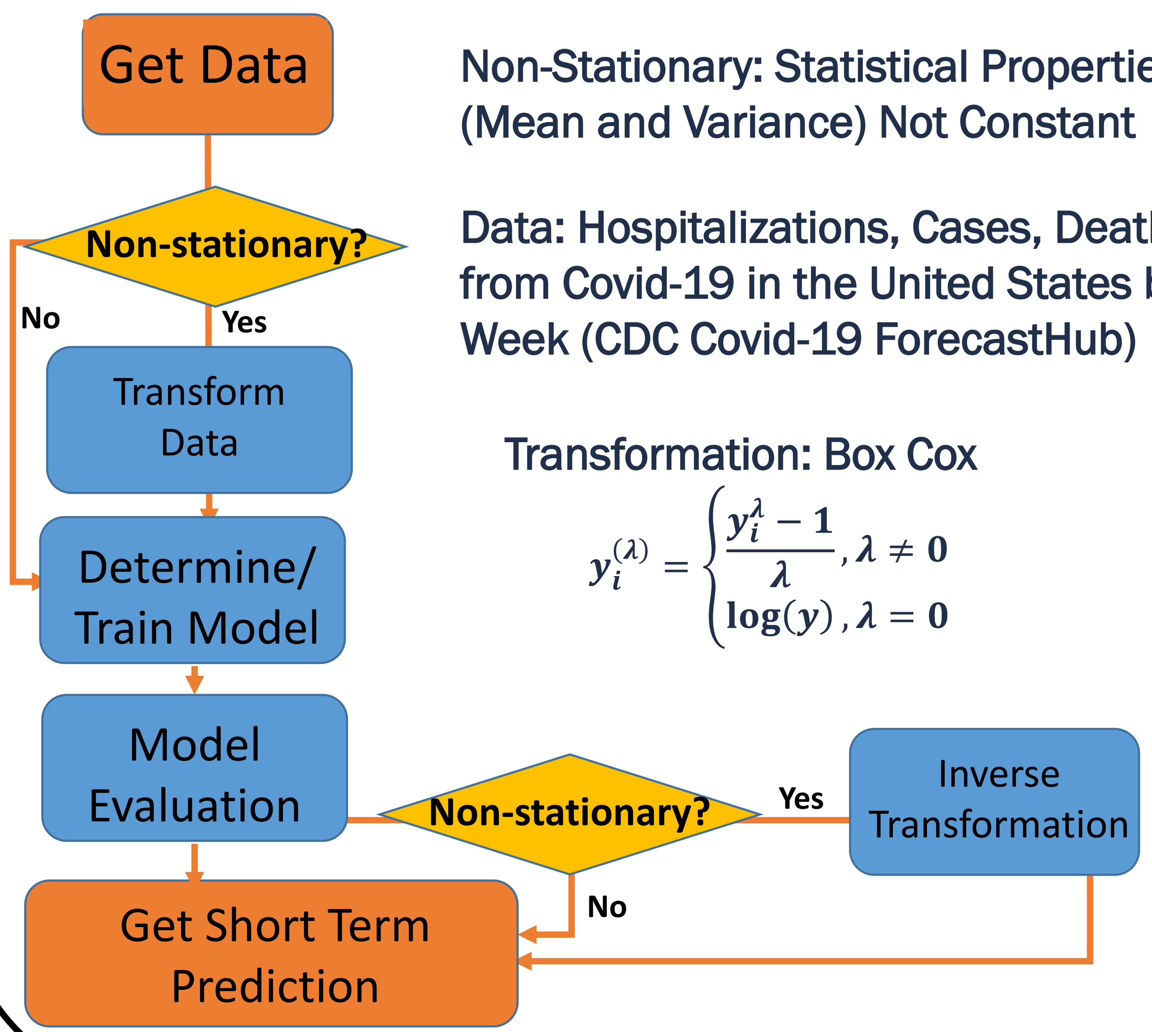
Multiple Covid-19 Time Series Forecasting

Finn Mokrzycki, Advisors Dr. Gursharn Kaur and Dr. Aniruddha Adiga

Background and Process



Non-Stationary: Statistical Properties (Mean and Variance) Not Constant

Data: Hospitalizations, Cases, Deaths from Covid-19 in the United States by Week (CDC Covid-19 ForecastHub)

Transformation: Box Cox

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases}$$

Current Project

Forecasting Multiple Time Series Simultaneously for Covid-19

Models Used:
ETS – Error, Trend, Seasonality
ARIMA – Autoregressive Integrated Moving Average
SARIMA – Seasonal Autoregressive Integrated Moving Average
VAR – Vector Autoregression
LSTM – Long Short-Term Memory (Recurrent Neural Network)

Evaluating Using Probabilistic Forecast
Using Intervals for Prediction Instead of Points

Future Work

Using Models to Forecast National and Other States' Covid-19 Cases
Further Exploration of VAR Model for Modeling Three Time Series at Once (Deaths)
LSTM Models with Covid-19 Deaths (Both Separate from and Joined with Cases and Hospitalizations)
Determining Waves, Onsets, Peaks with Long Term Forecasts

Seasonal Models

ETS: Separates Model into Trend and Seasonal Components and Uses Residuals as Error

SARIMA: Same as ARIMA but Accounts for Seasonal Patterns

Little to No Seasonal Component in Covid-19 Time Series
Can Only Forecast Univariate Time Series

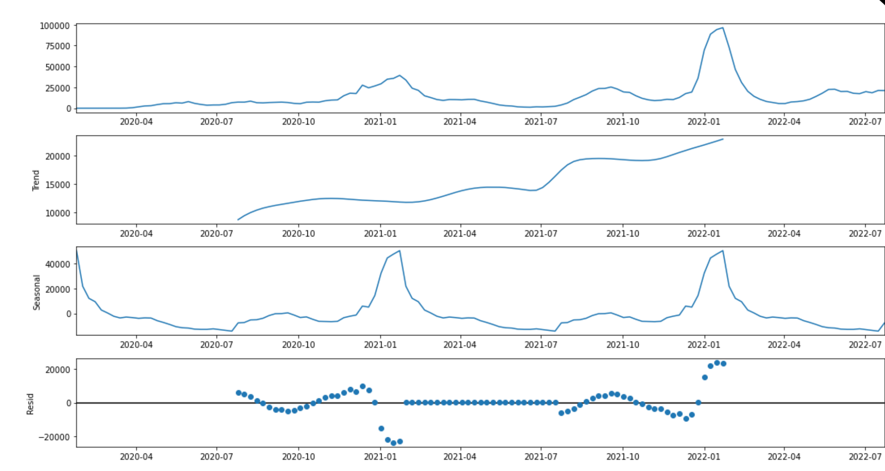
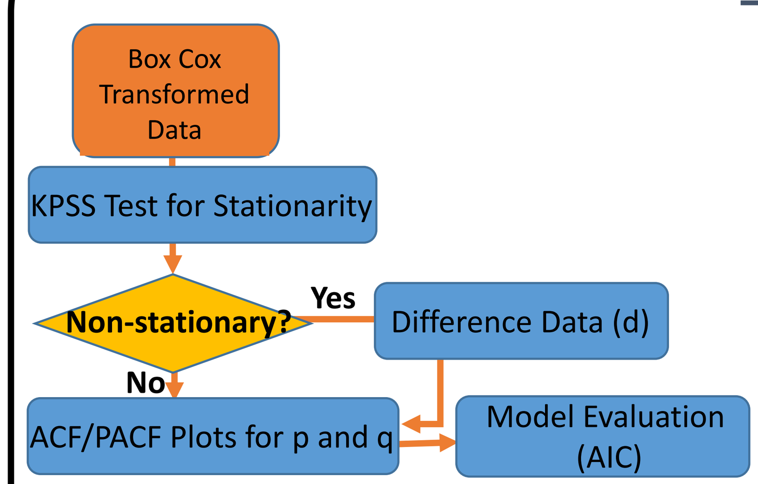
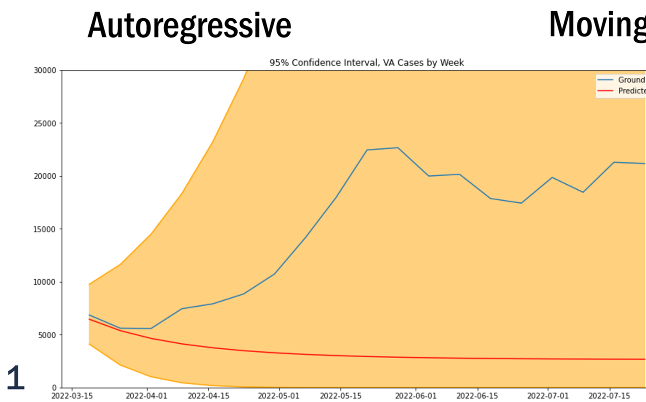


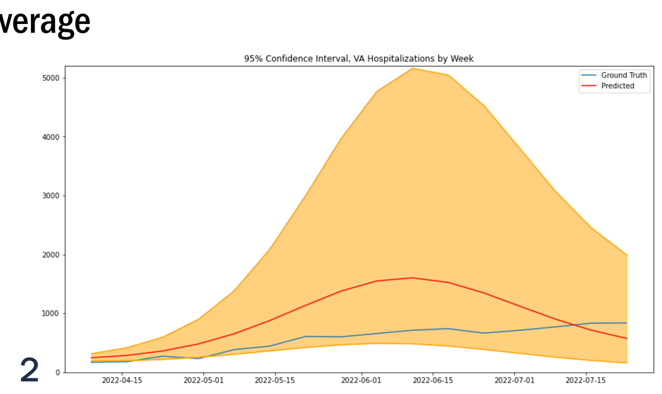
Figure: Covid-19 Cases (Virginia) Broken into Trend, Seasonal, and Error Components

ARIMA(p,d,q)

$$y'_t = c + \underbrace{\phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p}}_{\text{Autoregressive}} + \underbrace{\theta_{t-1}\varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}}_{\text{Moving Average}} + \varepsilon_t$$



1


2


3

	Variable	Order	Train MAE	Test MAE
0	Cases	(2,1,0)	2443.33	11663.5
1	Hospitalizations	(3,0,2)	93.9341	418.744

Figures:
1: Predictions for Cases (95% CI)
2: Predictions for Hospitalizations (95% CI)
3: MAE Comparison

VAR(p)

Takes Multivariate Series and Captures Their Relationship Over Time

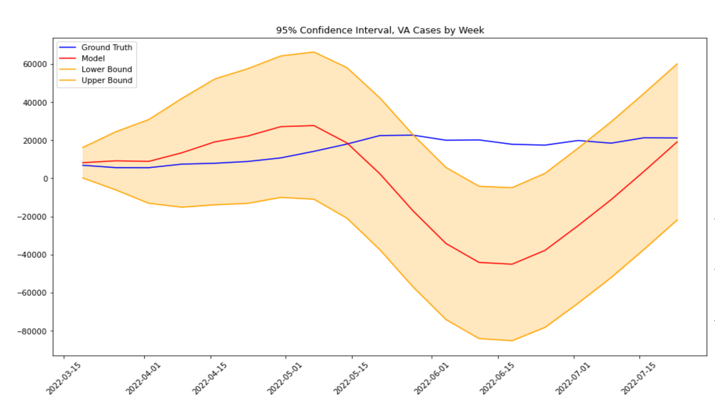


Figure 1

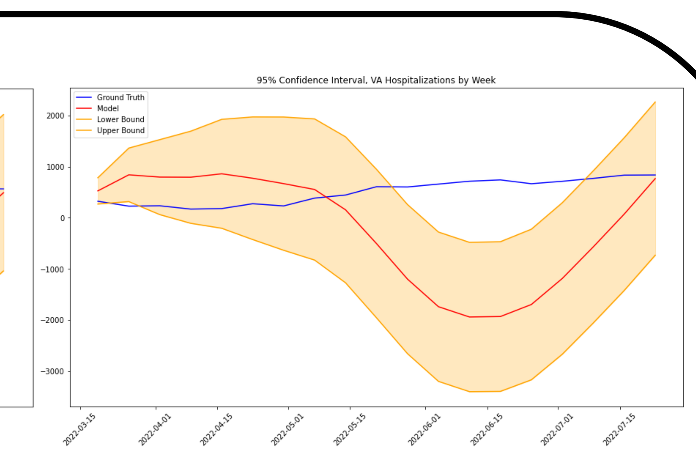
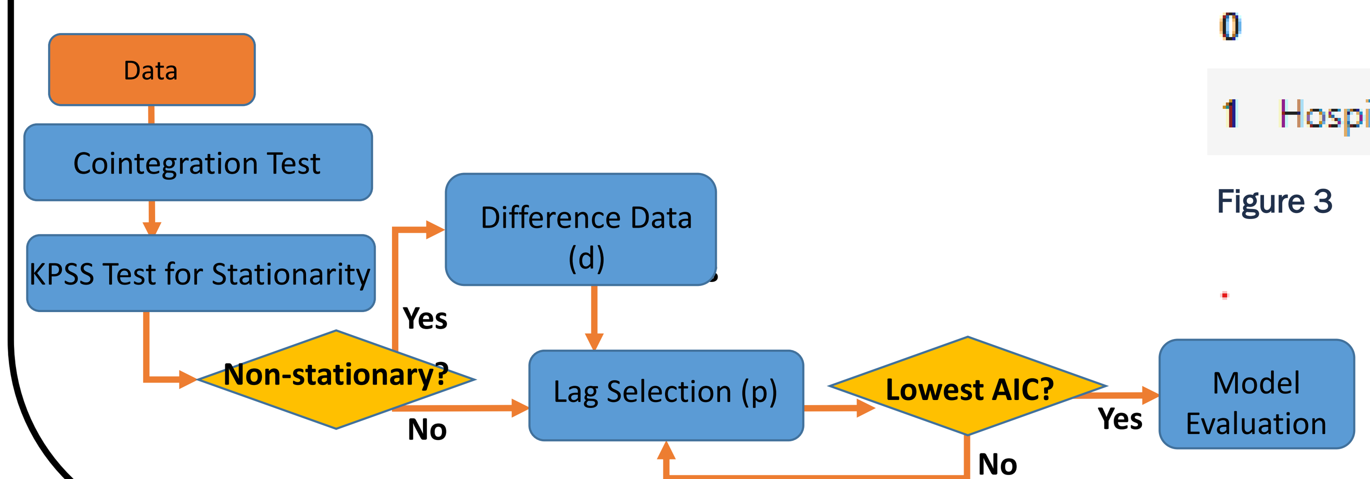


Figure 2

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

	Variable	Train MAE	Test MAE
0	Cases	1956.02	2052.46
1	Hospitalizations	67.1032	70.4928

Figure 3



Figures:
1: Predictions for Cases with 95% CI
2: Predictions for Hospitalizations with 95% CI
3: MAE Comparisons

LSTM

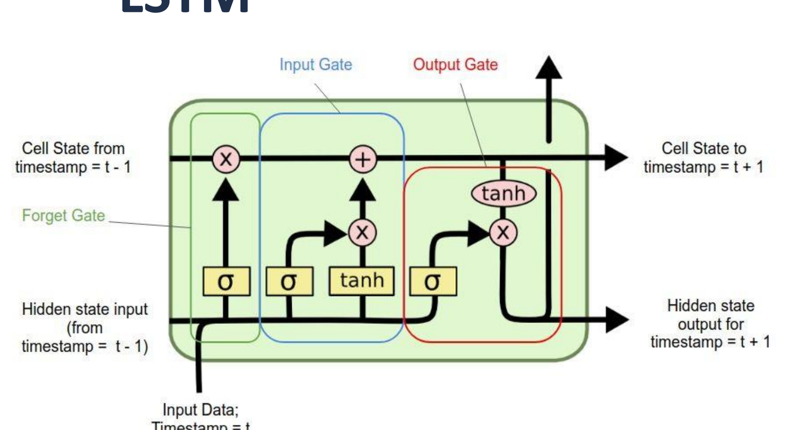


Figure 1: LSTM Model

Sequential Model

Past Information Flows Through Memory Cell ("Cell State" t-1) and "Remembers" Past Information

Current Data Input and Run Through Sigmoid (How Much Information to "Forget")

Data Outputs Through Hidden State And Cell State

Model	Variable	Train MAE	Test MAE
1 (Multivariate)	Cases	2449.86	1859.87
1	Hospitalizations	165.736	187.444
2 (Case Only)	Cases	2451.01	1916.83
3 (Hospitalization Only)	Hospitalizations	80.1825	65.0098

Figure 2: LSTM with Input $[y_{t-1}, y_t]$

Model	Variable	Train MAE	Test MAE
1 (Multivariate)	Cases	2380.38	1802.63
1	Hospitalizations	132.89	165.176
2 (Case Only)	Cases	1944.93	2126.45
3 (Hospitalization Only)	Hospitalizations	49.3901	70.2071

Figure 3: LSTM with Input $[y_{t-3}, y_{t-2}, y_{t-1}, y_t]$

Model	Variable	Train MAE	Test MAE
1 (Multivariate)	Cases	2294.07	1832.14
1	Hospitalizations	198.021	352.975
2 (Case Only)	Cases	1056.82	2773.04
3 (Hospitalization Only)	Hospitalizations	49.15	60.2068

Figure 4: LSTM with Input $[y_{t-5}, y_{t-4}, y_{t-3}, y_{t-2}, y_{t-1}, y_t]$

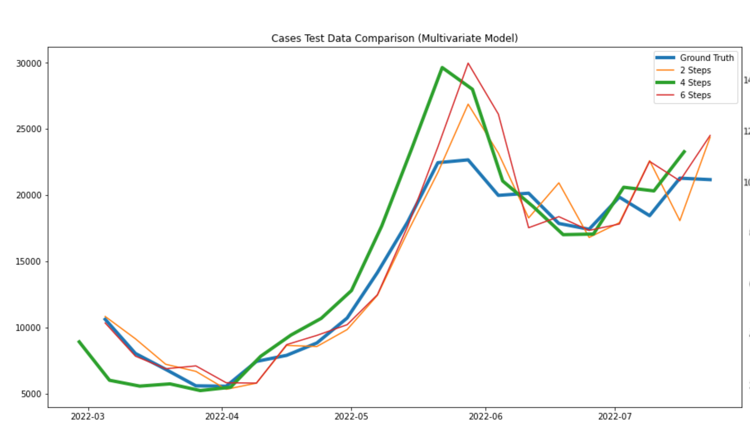


Figure 5

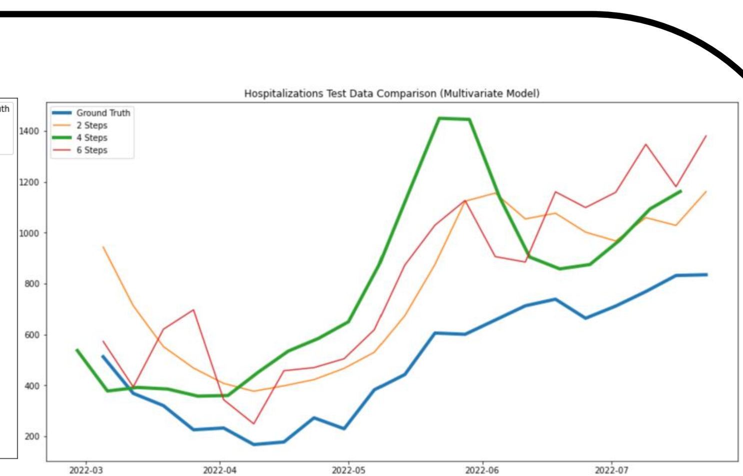


Figure 6

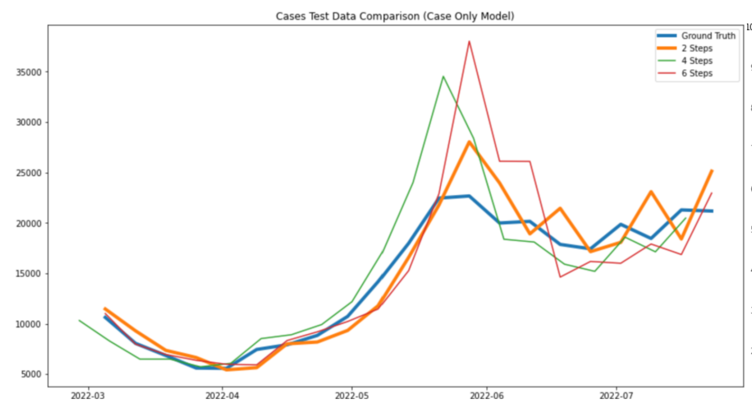


Figure 7

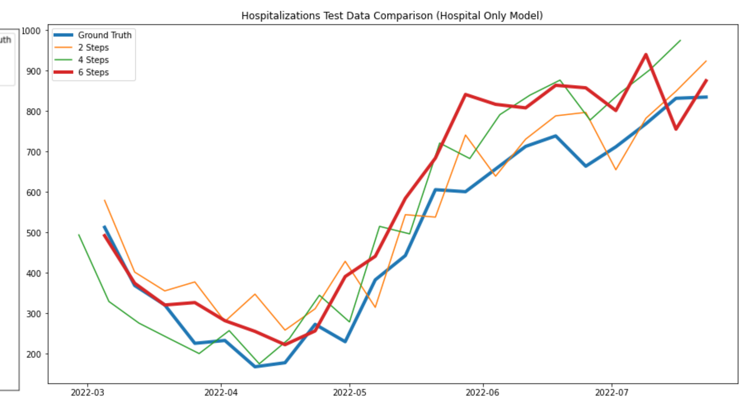


Figure 8

Figures:
5: Multivariate Models for Cases (Lowest Test MAE Bolded)
6: Multivariate Models for Hospitalizations (Lowest Test MAE Bolded)
7: Case Only Model (Lowest Test MAE Bolded)
8: Hospitalization Only Model (Lowest Test MAE Bolded)