Analyzing COVID-19 Time-Series Data & Government Policy Responses for Forecasting Deaths

IACIS Annual International Conference - 2021

- Joshua Hammond, University of St. Thomas (MN)
  hamm9537@stthomas.edu
- Melinda Caouette, University of St. Thomas (MN)
  caou1088@stthomas.edu
- Dr. Manjeet Rege, University of St. Thomas (MN)
  rege@stthomas.edu

UNIVERSITY OF St. Thomas

https://doi.org/10.48009/2_iis_2021_122-135

Introduction & Background

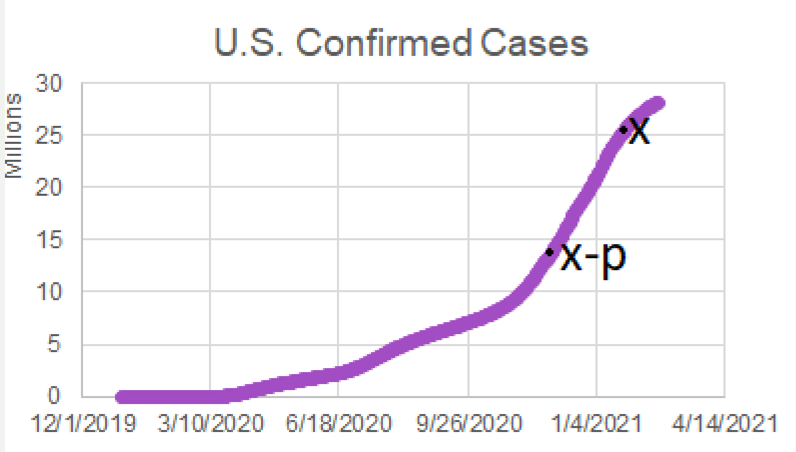
- Governments around the world implemented various **public health** and **policy countermeasures** intended to slow the continued outbreak of COVID-19 disease in 2020 and 2021
- How effective are these countermeasures, and is there a predictive relationship between **government policy response**, **confirmed case counts**, and **deaths** that would allow for better forecasting?

Research Hypothesis

H_1: More stringent government public health policy measures have some predictive power – *not necessarily causal* – which can improve forecast accuracy of COVID-19 deaths.

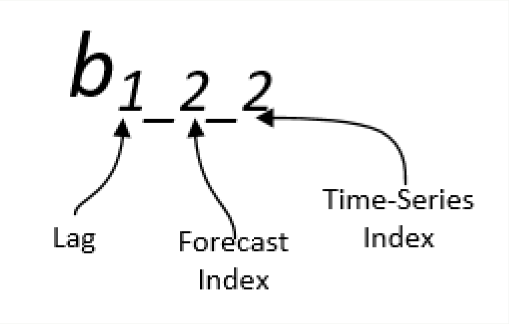
H_2: COVID-19 deaths can be forecast accurately using vector auto-regression and vector error correction models.

Vector Auto-Regression



- An Auto-Regressive (AR) model forecasts future data points of a time series using its own past steps

- Vector Auto-Regression (VAR) extends the concept of AR to include ***multiple time-series*** in the model. The relationships between each time series can be modeled as dependent (endogenous) or independent (exogenous) variables



$$x_t = c_1 + b_{1\_1\_1} * x_{t-1} + b_{1\_1\_2} * y_{t-1} + b_{2\_1\_1} * x_{t-2} + b_{2\_1\_2} * y_{t-2} + e_{1,t} \quad (1)$$

$$y_t = c_2 + b_{1\_2\_1} * x_{t-1} + b_{1\_2\_2} * y_{t-1} + b_{2\_2\_1} * x_{t-2} + b_{2\_2\_2} * y_{t-2} + e_{1,t} \quad (2)$$

$$\begin{aligned} Deaths_t = {} & c_1 + b_{1\_1\_1} * Deaths_{t-1} + b_{1\_1\_2} * Cases_{t-1} + b_{1\_1\_3} * StringencyIndex_{t-1} + \cdots \\ & + b_{p\_1\_1} * Deaths_{t-p} + b_{p\_1\_2} * Cases_{t-p} + b_{p\_1\_3} * StringencyIndex_{t-p} + e_{1,t} \end{aligned} \quad (3)$$

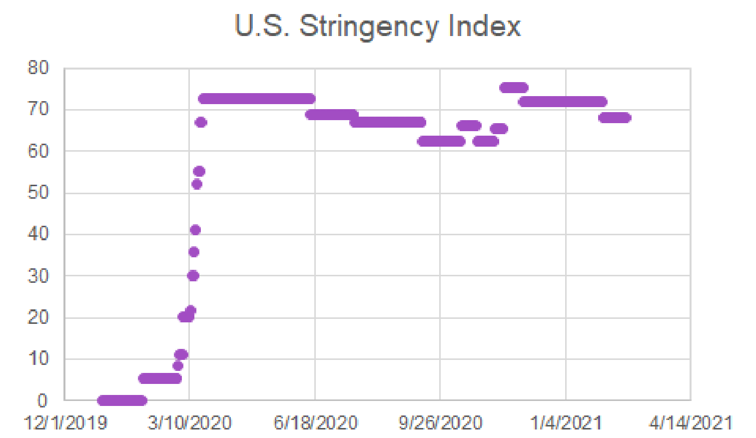
$$\begin{aligned} Cases_t = {} & c_2 + b_{1\_2\_1} * Deaths_{t-1} + b_{1\_2\_2} * Cases_{t-1} + b_{1\_2\_3} * StringencyIndex_{t-1} + \cdots \\ & + b_{p\_2\_1} * Deaths_{t-p} + b_{p\_2\_2} * Cases_{t-p} + b_{p\_2\_3} * StringencyIndex_{t-p} + e_{2,t} \end{aligned} \quad (4)$$

$$\begin{aligned} & StringencyIndex_t \\ & = c_3 + b_{1\_3\_1} * Deaths_{t-1} + b_{1\_3\_2} * Cases_{t-1} + b_{1\_3\_3} * StringencyIndex_{t-1} + \cdots \\ & + b_{p\_3\_1} * Deaths_{t-p} + b_{p\_3\_2} * Cases_{t-p} + b_{p\_3\_3} * StringencyIndex_{t-p} + e_{3,t} \end{aligned} \quad (5)$$

UNIVERSITY OF St.Thomas

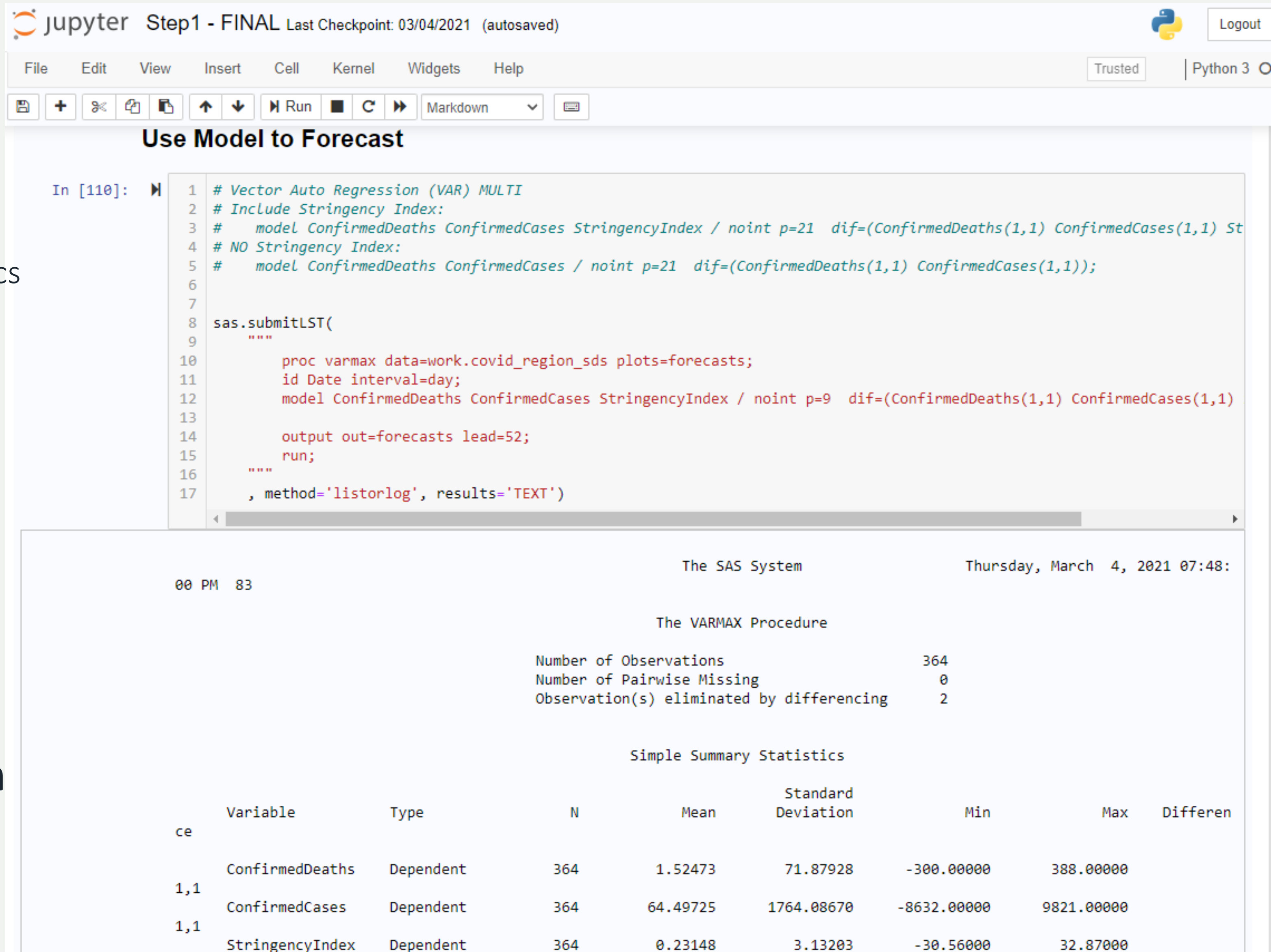
Data

- The source dataset was obtained from the Oxford University COVID-19 Government Response Tracker ([OxCGRT]) (Hale et al., 2020)
- The OxCGRT systematically collects 19 **indicators of government responses** for over 180 countries and many sub-national regions
  - Aggregates the indicators into the *government response index*, *containment and health index*, *economic support index*, and *stringency index*
- The source dataset originally had 49 columns and 120,268 rows from 1/1/2020 to 3/1/2021



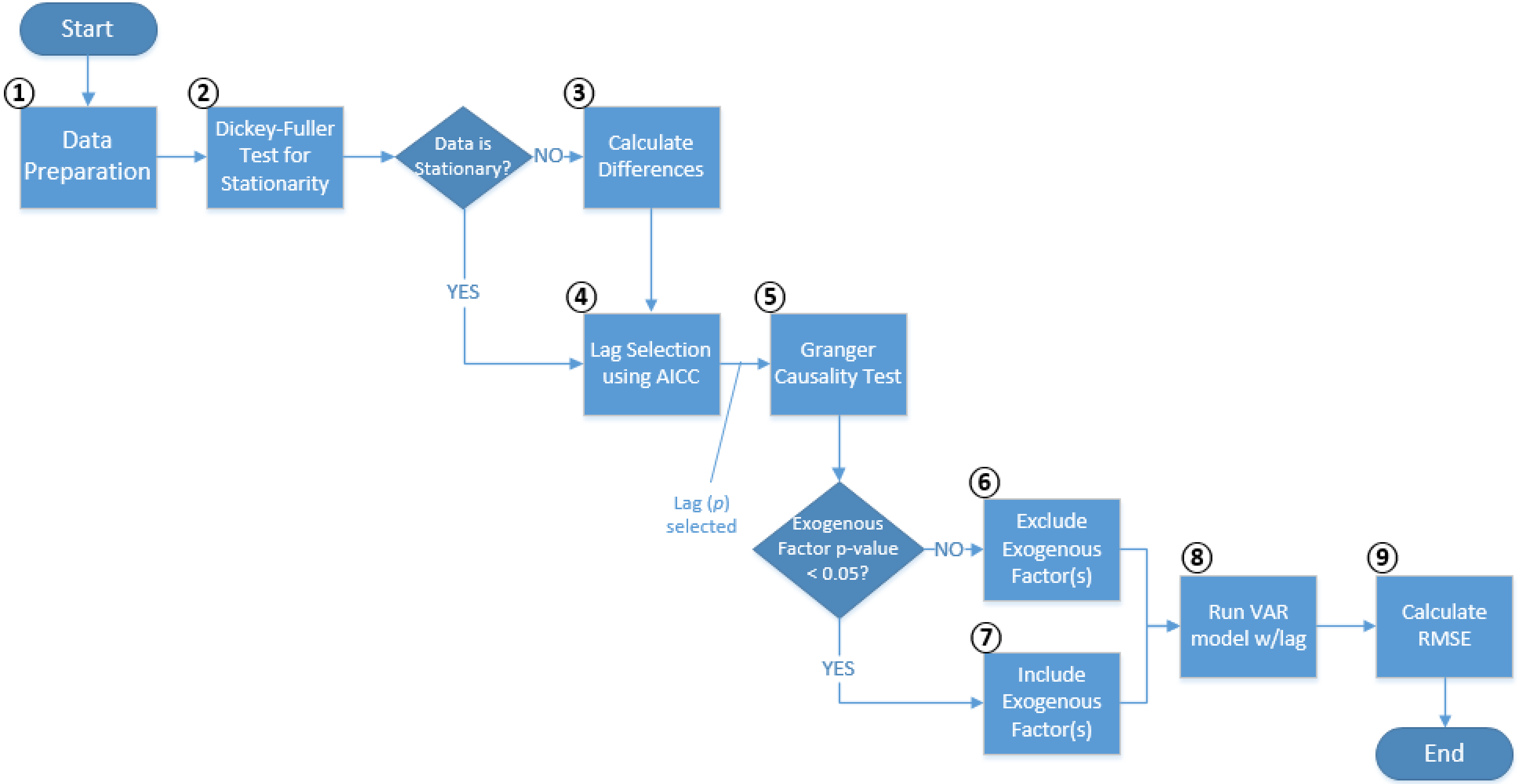
Software Tools

- Analysis was performed using **SAS/ETS® software** version 15.1
  - Running on SAS On-Demand for Academics (ODA) system using Base SAS software version 9.04.01
  - The ODA session runs inside a Jupyter notebook via the SASPy Python interface
- The SAS/ETS® software **VARMAX** procedure was used to perform vector auto-regression
- Data plots generated with **Seaborn**



UNIVERSITY OF St.Thomas

Methodology



UNIVERSITY OF St.Thomas

1 Data Preparation

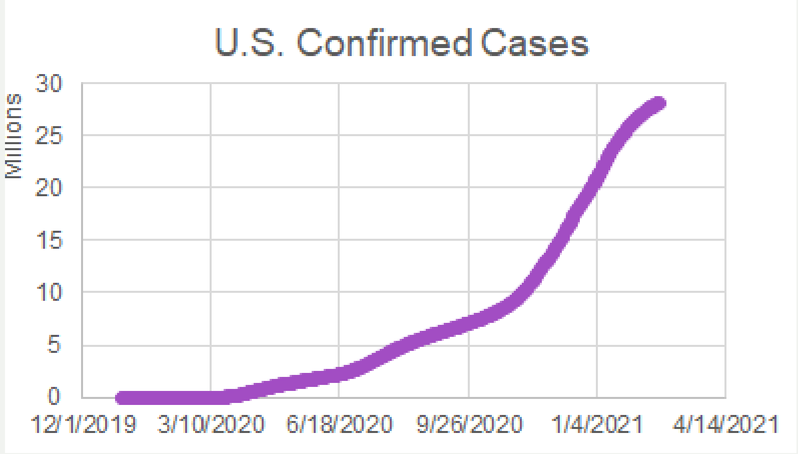
Name	Description	Data Type	Example
CountryName	Country Name	String	United States
RegionName	Sub-National Regions (State, etc.)	String	
Jurisdiction	NAT_TOTAL or STATE_TOTAL	String	NAT_TOTAL
Date	Reported date	Date	3/21/2020
C1_SchoolClosing	Record closings of schools and universities; ordinal scale of 0 to 3	Integer	3
C2_WorkspaceClosing	Record closings of workspace; ordinal scale of 0 to 3	Integer	3
C3_CancelPublicEvents	Record cancelling public events; ordinal scale of 0 to 2	Integer	2
C4_RestrictionsGatherings	Record limits on gatherings; ordinal scale of 0 to 4	Integer	4
C5_ClosePublicTransport	Record closings of public transport; ordinal scale of 0 to 2	Integer	1
C6_StayAtHome	Record orders to 'shelter-in-place'; ordinal scale of 0 to 3	Integer	2
C7_RestrictionsIntMvt	Record restrictions on internal movement between cities/regions; ordinal scale of 0 to 2	Integer	2
C8_InternationalTravel	Record restrictions on international travel; ordinal scale of 0 to 4	Integer	3
E1_Income support	Record if cash payments to people who lose jobs; ordinal scale of 0 to 2	Integer	0
E2_Debt/contract relief	Record if freezing financial obligations for households; ordinal scale of 0 to 2	Integer	0
E3_Fiscal measures	Announced economic stimulus spending	Float	0
E4_International support	Announced offers of aid spending to other countries	Float	0
H1_PublicInfo	Record of public info campaigns; ordinal scale of 0 to 2	Integer	2
H2_TestingPolicy	Record who has access to testing; ordinal scale of 0 to 3	Integer	3

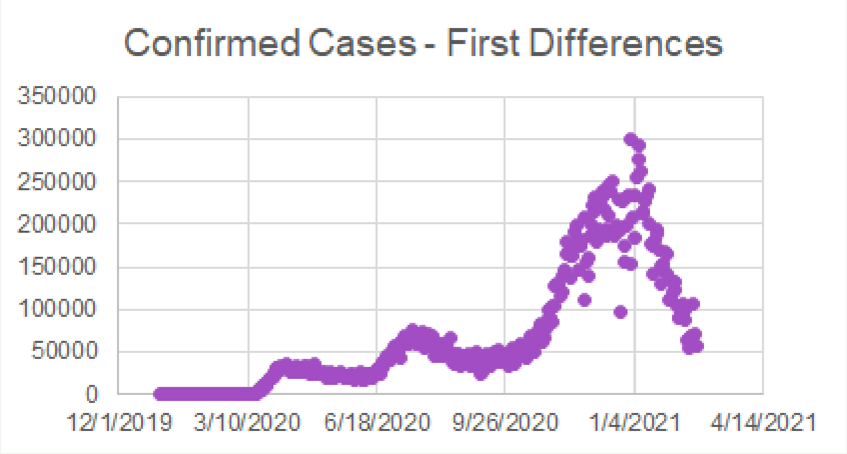
- Prepared using Jupyter notebooks and *pandas*
  - Transformed into a SAS data store for processing
- All rows that correspond to the period after 2/21/2021 were removed
  - Some countries had a lag in reporting COVID-19 cases
- Missing values for "*ConfirmedCases*" and "*ConfirmedDeaths*" were populated with zero
  - These columns were originally blank until the first case was reported
- Other missing data were imputed w/ mean values
- Unused columns were removed
- Long column names were truncated, and white space was removed to meet the SAS naming conventions

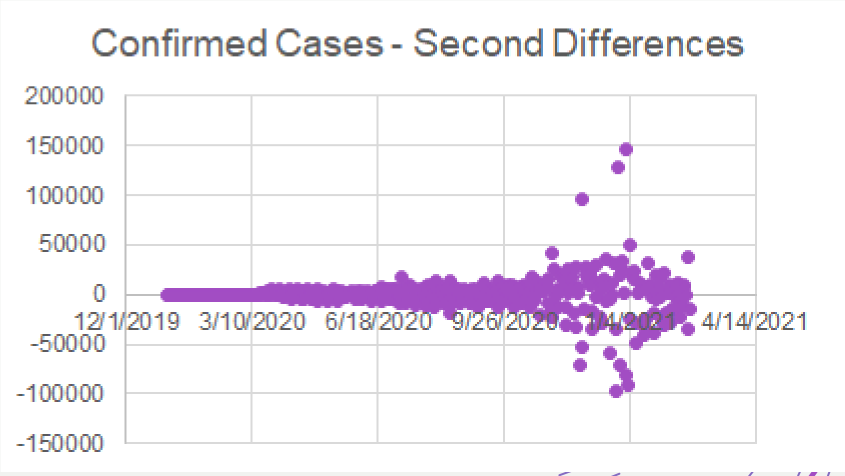
UNIVERSITY OF St.Thomas

❷ Test for Stationarity

- Time-Series analysis techniques generally require the data to be “stationary” – mean and variance are constant over time
- Dickey-Fuller test for Stationarity tells us if we need to transform the data







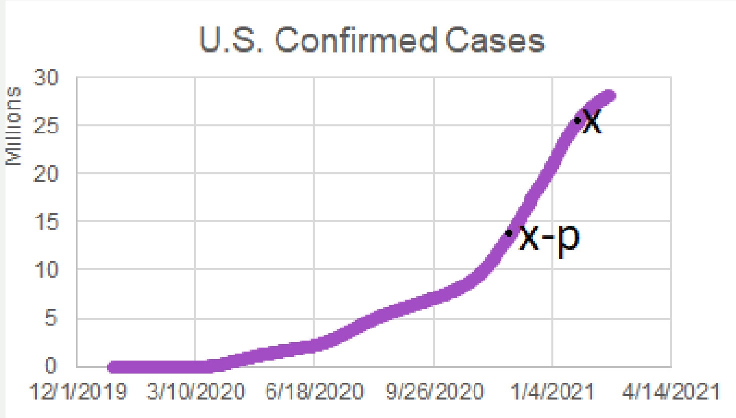
UNIVERSITY OF St.Thomas

3 Calculate Differences

- We can make the data stationary by performing first or second differences on the data if necessary ($t_i = t_i - t_{i-1}$)

4 Lag Selection for Auto-Regression

- The lag (p) is the number of steps in the past to look
  - The optimal lag can vary by data set
- We used Akaike's information corrected criteria (AICC) looking out to a maximum of 21 days (lags)
  - The model with the smallest AICC value was selected and this number of lags was used in the forecast model



5 Granger Causality

- Given two time-series, a **policy stringency index** and **confirmed deaths**, you can state that the stringency index *granger-causes* confirmed deaths if including the past values of the stringency index improves the predictive accuracy of confirmed deaths in a model *more* than considering confirmed deaths alone
- This fits well with VAR models as there can be multiple time-series influencing each other
- The test states a *null-hypothesis* that stringency index does not granger-cause confirmed deaths, with a significantly small p-value indicating that the null-hypothesis can be rejected
  - Our threshold for p-value cut off was 0.05
- Time-series that were above the threshold were excluded from the forecast models
- The SAS code is shown in the appendix

8 Run VAR Model(s)

- Models were trained on one year of data running from 1/1/2020 to 12/31/2020
- Four models were used for each country:
  1. Multivariate Vector Auto Regression (VAR)
  2. Univariate Auto Regression (AR)
  3. Multivariate Vector Error Correction (VEC)
  4. Univariate Vector Error Correction (VEC)

UNIVERSITY OF St.Thomas

9 Calculate Root Mean Squared Error (RMSE)

- Models were evaluated by calculating the root mean square error (RMSE) for 52 days of forecasted deaths running from 1/1/2021 to 2/21/2021

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}$$

Results

- The Granger causality p-values for most of the OxCGRT composite indexes (government response index, containment and health index, economic support index, and stringency index) were very high (>> 0.05) indicating that **we cannot reject the null hypothesis that these indexes do not granger-cause confirmed deaths**
  - In a handful of countries, we did find that the stringency index granger-caused confirmed deaths and we included it in the model
- The OxCGRT Stringency Index was found to granger-cause confirmed deaths in the models for Brazil, Italy and Japan
  - The RMSE was improved for Italy and Brazil while the error rate was improved for all three countries
- The forecasts for 52 days out have less than 5% error for 11 out of 14 countries

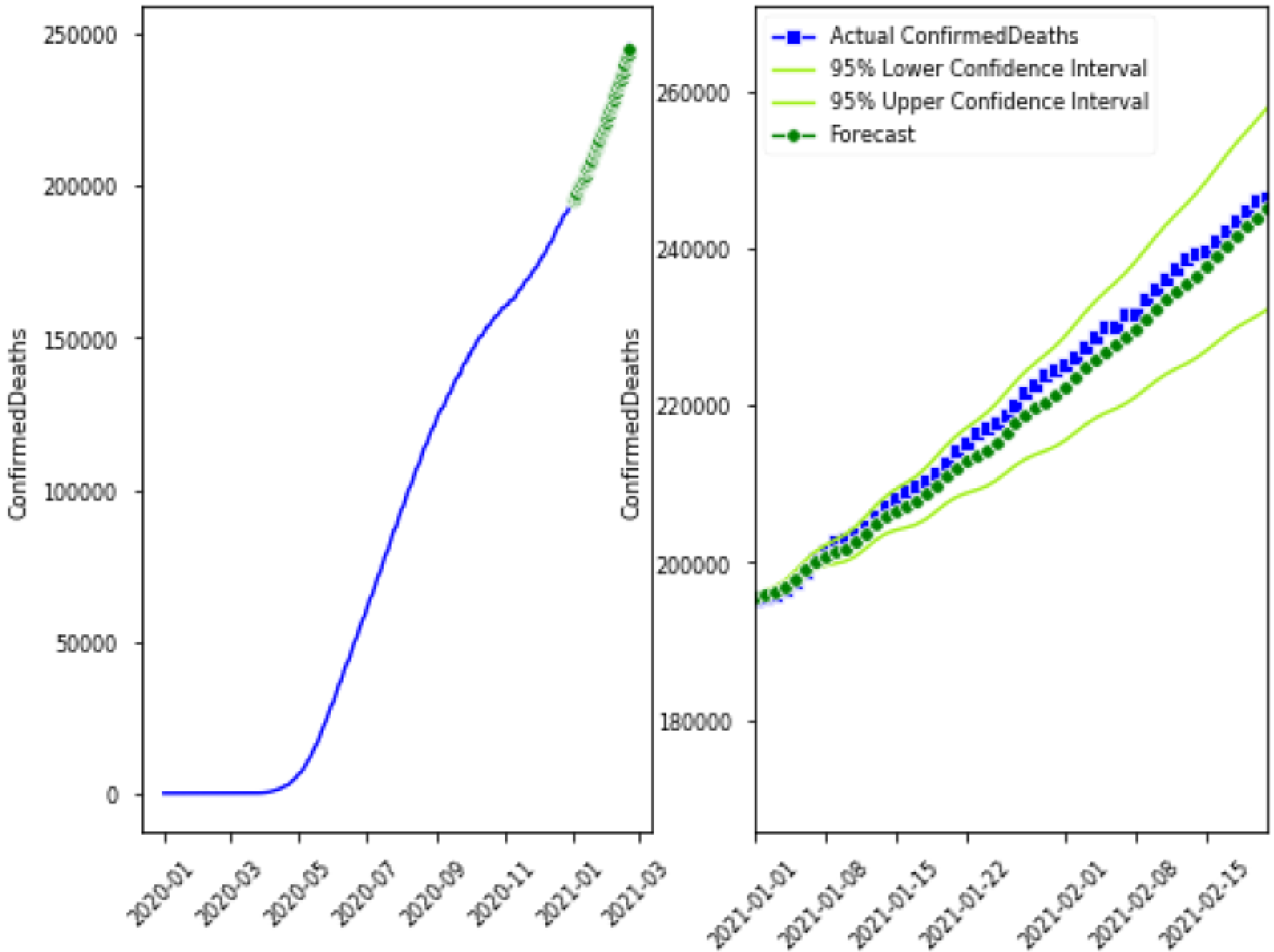
Results: Summary of 52-Day Forecast

Country	Multivariate VAR's RMSE	Univariate AR's RMSE	Multivariate VEC's RMSE	Univariate VEC's RMSE	Forecast on 02/21/2021	Cumulative Deaths on 02/21/2021	Error Rate
China	1.7	1.5	81.6	146.4	4,634	4,636	-0.04%
United States	6494.7	5985.6	9833.9	11733.3	500028	499009	0.20%
Australia	10.0	2.4	19.2	21.5	913	909	0.44%
Brazil	2149.5	7368.5	3793.4	7957.4	244955	246504	-0.63%
Japan	1036.7	988.9	266.7	718.4	7,421	7,485	-0.86%
Israel	117.6	605.2	191.3	634.6	5643	5577	1.18%
Sweden	670.3	840.8	606.6	952.1	12415	12649	-1.85%
India	6358.7	1300.8	2304.7	7262.0	159,386	156,385	1.92%
Italy	4245.2	3304.5	1183.4	3270.0	98,521	95,718	2.93%
New Zealand	2.3	0.4	2.6	0.5	25	26	-3.85%
South Africa	1768.3	2284.1	7695.7	2178.3	51,013	49,053	4.00%
United Kingdom	5977.5	5644.1	7092.9	8593.3	128803	120580	6.82%
Spain	2930.8	3788.8	3750.5	2956.9	61,260	67,101	-8.70%
South Korea	206.6	259.2	169.6	680.8	1,901	1,562	21.70%

RMSE: Lower is better

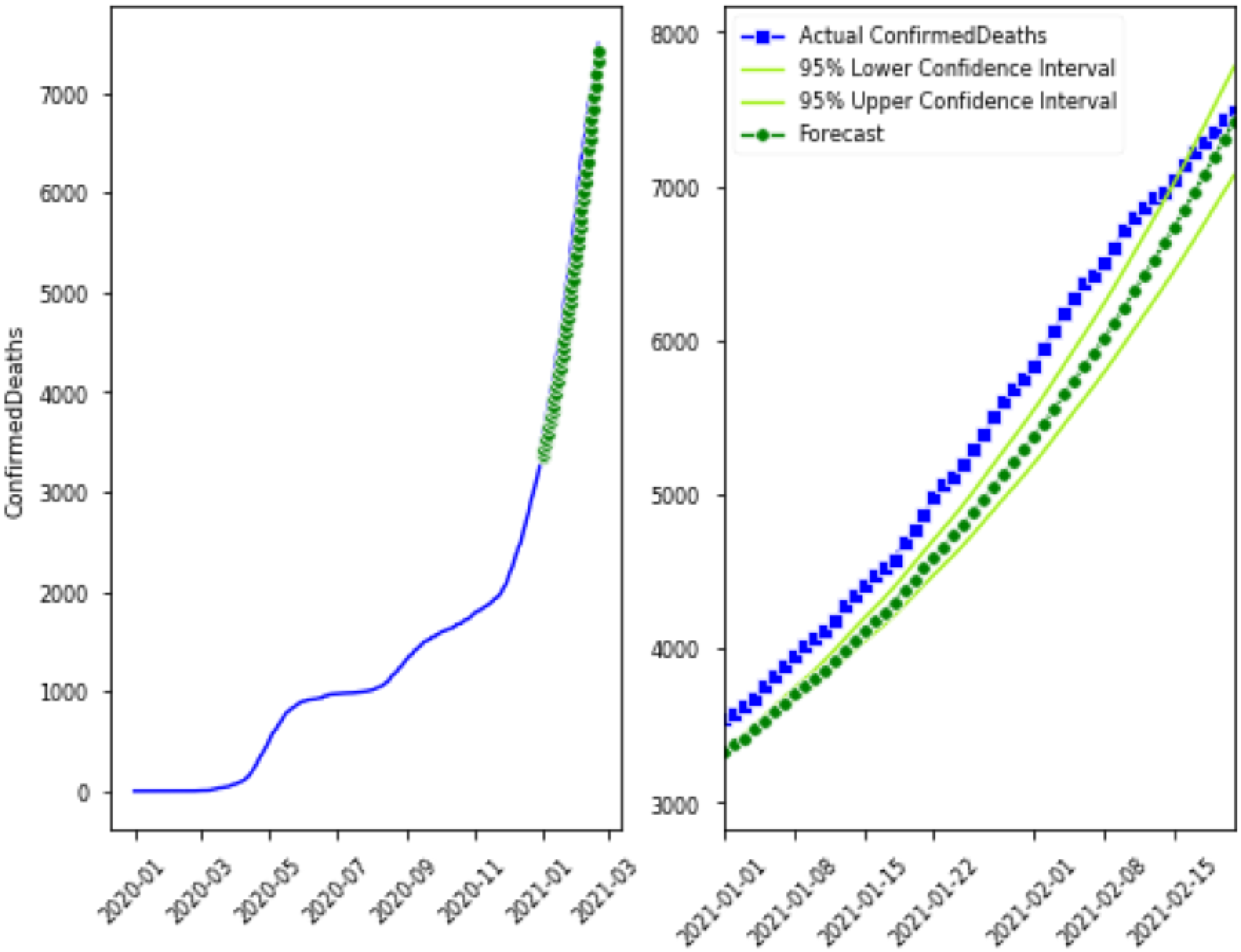
UNIVERSITY OF St.Thomas

Figure 1: Brazil (Multivariate VAR w/ Stringency Index)



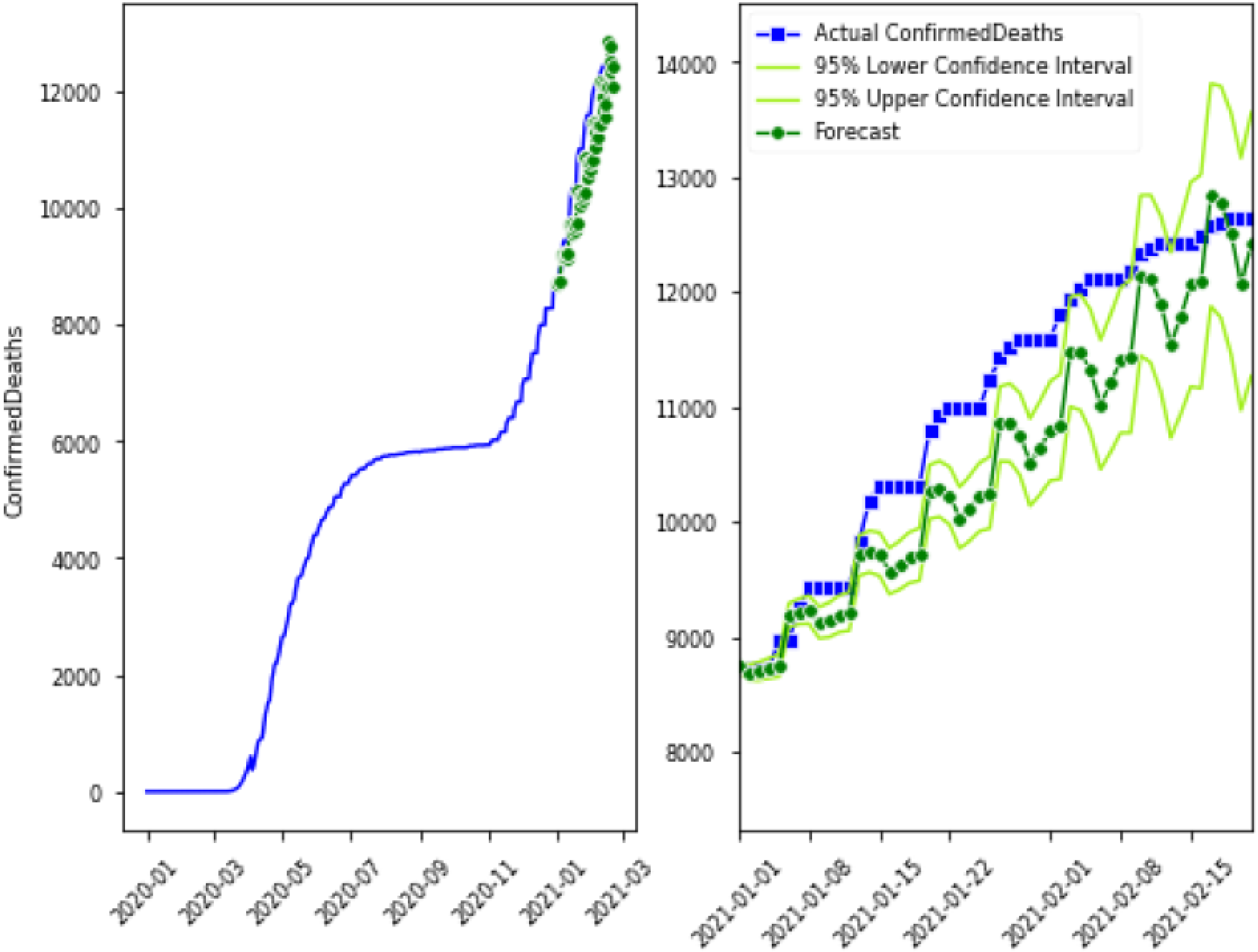
UNIVERSITY OF St.Thomas

Figure 3: Japan (Multivariate VEC w/ Stringency Index)



UNIVERSITY OF St.Thomas

Figure 4: Sweden (Multivariate VEC)



Conclusions

- In this study we attempted to identify relationships between various indexes of government policy, confirmed case count, and confirmed deaths and quantify their predictive power using Vector Auto Regression (VAR) methods
- *H_1: More stringent government public health policy measures have some predictive power – not necessarily causal – which can improve forecast accuracy of COVID-19 deaths*
  - In a handful of countries (3) including the stringency index also improved the forecast, however most of the other policy indexes did not improve the forecast accuracy.
- *H_2: COVID-19 deaths can be forecast accurately using vector auto-regression and vector error correction models*
  - We found that the multivariate VAC and VEC models outperformed the univariate models in eight out of fourteen countries studied
  - We interpret this to mean that for most countries the multivariate models outperform simple univariate AR/VEC models.

UNIVERSITY OF St.Thomas

Questions?

Cleansed data and related materials are located here: https://github.com/joshhammond/IIS_COVID19_Data

UNIVERSITY OF St.Thomas

Appendix

VARMAX Code Example

– Use dif to convert data to stationary
– Use minic to find the best lag

```
proc varmax data=work.covid_region_sds plots;
id Date interval=day;
model ConfirmedDeaths ConfirmedCases StringencyIndex /
minic=(type=aicc p=21 q=0)  dif=(ConfirmedDeaths(1,1)
ConfirmedCases(1,1) StringencyIndex(1));
```

UNIVERSITY OF St.Thomas

VARMAX Code Example

– Test granger causality

```
proc varmax data=work.covid_region_sds;
id Date interval=day;
model ConfirmedDeaths ConfirmedCases StringencyIndex /
noint p=11  dif=(ConfirmedDeaths(1,1) StringencyIndex(1)
GovernmentResponseIndex(1) ContainmentHealthIndex(1)
ConfirmedCases(1,1));
causal group1=(ConfirmedDeaths) group2=(StringencyIndex);
causal group1=(ConfirmedDeaths) group2=(ConfirmedCases);
```

UNIVERSITY OF St.Thomas

VARMAX Code Example

– Generate forecasts

– Use cointeg for VEC, omit for VAR

```
proc varmax data=work.covid_region_sds plots=forecasts;
id Date interval=day;
model ConfirmedDeaths ConfirmedCases StringencyIndex /
noint p=11  dif=(ConfirmedDeaths(1,1) StringencyIndex(1)
GovernmentResponseIndex(1) ContainmentHealthIndex(1)
ConfirmedCases(1,1));
cointeg rank=1;
output out=forecasts lead=52;
```

UNIVERSITY OF St.Thomas