How Do Economic Indicators Impact the General Public's Interest in US Jobless Claims?

March 12, 2019

Time Series and Forecasting

Summary

In this project, our main goal is to forecast future interest in US jobless claims using several major economic indicators. We selected the indicators that reflect the overall strength of the US economy, such as Jobless Claims (officially called Initial Claims (ICSA)), St. Louis Fed Financial Stress Index (STLFSI), S&P 500 Index and Russell 2000 Index.

For this project, we will be using an autoregressive–moving-average model with exogenous inputs (ARIMAX). This model contains the AR(*p*) and MA(*q*) models and a linear combination of the last *b* terms of known and external time series, denoted *d*.

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \sum_{i=1}^{b} \eta_i d_{t-i}.$$

We constructed different models and compare the RMSE with baseline models. The baseline models, persistence and moving average, have RMSE of 10.63 and 11.98. we also tried Holt exponential smoothing which yields RMSE of 12.29, as well as simple exponential smoothing which yields RMSE of 8.91.

We felt like we could improve the forecasting by ARIMA and ARIMAX models so we fitted ARIMA(1,0,1); however the model has the RMSE of 10.33, which is no any better than a simple exponential smoothing model.

Finally, after performing log differencing to all the explanatory variable indexes, we fitted ARIMAX(2,0,1), and that gives us the best RMSE of 8.19.

Summary Statistics

Explanatory variables:

1) Percentage change in Initial Claims (ICSA)

> Jobless claims are a statistic reported weekly by the U.S. Department of Labor that counts people filing to receive unemployment insurance benefits. The purpose of reviewing jobless claims is to determine the direction of the unemployment rate and the overall health of the economy.

We decided to use percentage change rather than the raw ICSA values. By using the percent change transformation, it will make the time series stationary.

2) Logged-differenced St Louis Fed Financial Stress Index

> This index measures the degree of financial stress in U.S. markets on a weekly basis. The average of the index is zero. A value that is above zero indicates that there is an increase in the financial stress while a value below zero signals market conditions that are less stressful than average.

3 & 4) Logged-differenced S&P 500 and Logged-differenced Russell 2000

> Both S&P 500 and Russell 2000 are indexes that represent the stock market performances by widely quoted measure of the overall performance of capital stocks. Assumption: if the unemployment rate is higher, the general income will be limited; people have less cash in hand and therefore spend less money (demand for products decreases). This can mean that stock prices may go down in many areas because there isn't as much of a demand for certain goods.

For St Louis Fed Financial Stress Index, Standard & Poor's 500, and Russell 2000, we decided to apply log difference transformations on each of the variables. By performing log difference transformations, the data will make highly skewed distributions less skewed. This allows the use of making patterns in the data more interpretable and to meet the assumptions of inferential statistics. Also, the log difference transformation allows it to make non-stationary data into stationary data.

Dependent variable

<u>Google Search Volume Index (SVI)</u>

We use the Google Search Volume Index (SVI) as an indicator for interest in US jobless claims. This index has a range of 0 to 100, and provides daily and weekly reports on the volume of queries for search terms used all around the globe.

We chose this index as our dependent variable because there is evidence that simple seasonal autoregressive models that incorporates relevant Google search trends as variables tend to outperform models that do not (Varian, 2009).

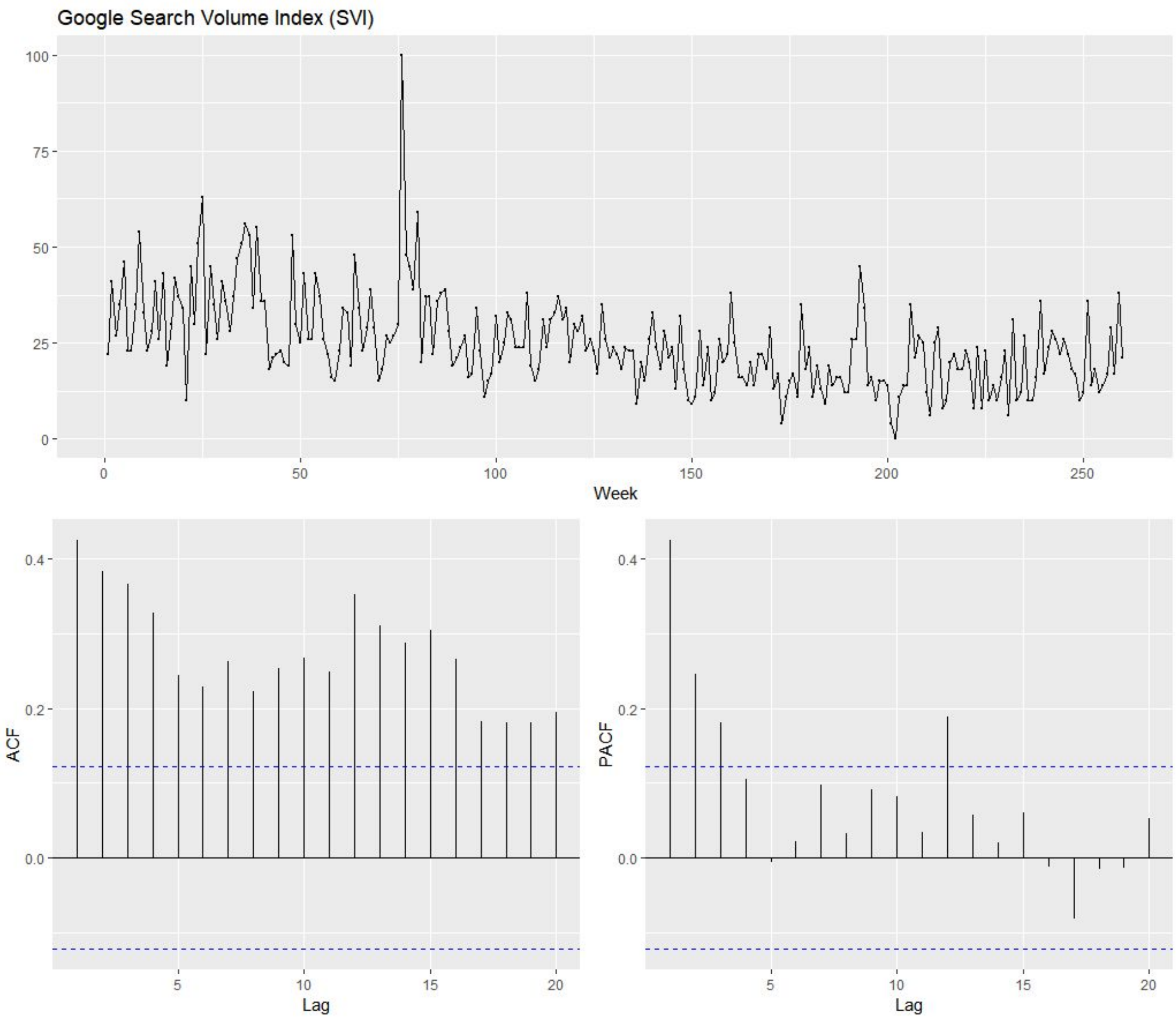
In our project, we focus on search queries that originated from the United States. For our analysis, we focus on data from the past 5 years.

<u>Summary Statistics</u>:

	Baseline models					
	Persistence (random walk)	Average of all previous observations	Holt Linear exponential smoothing	Simple smoothing	Arima(1,01)	Arimax(2,0,1)
RMSE	11.98249	10.62995	12.28643	8.905084	10.3345	8.186519

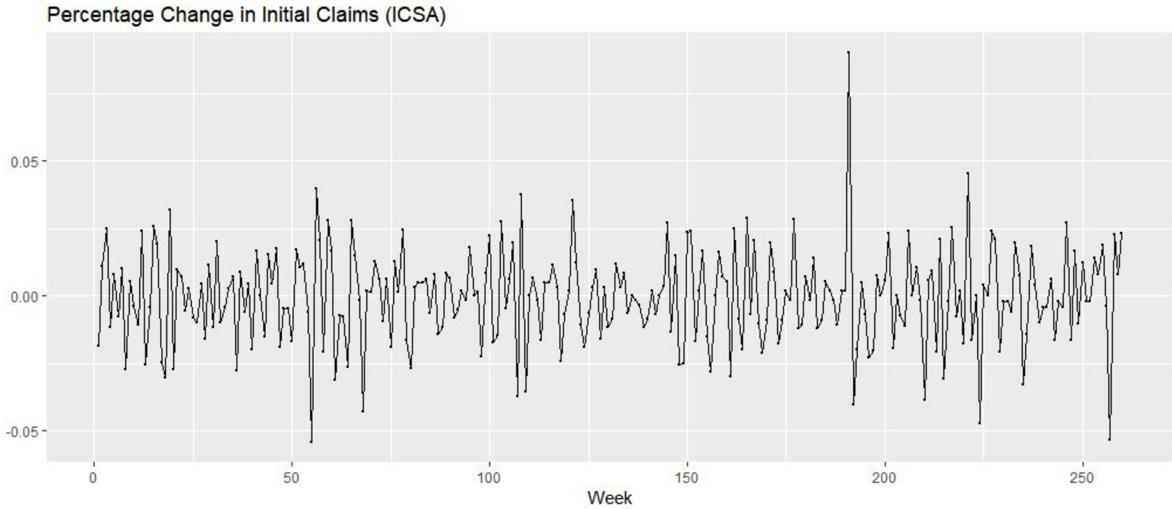
The table above shows the RMSE for each of the models tested. Arimax(2,0,1) performed the best with the lowest RMSE value of 8.186519.

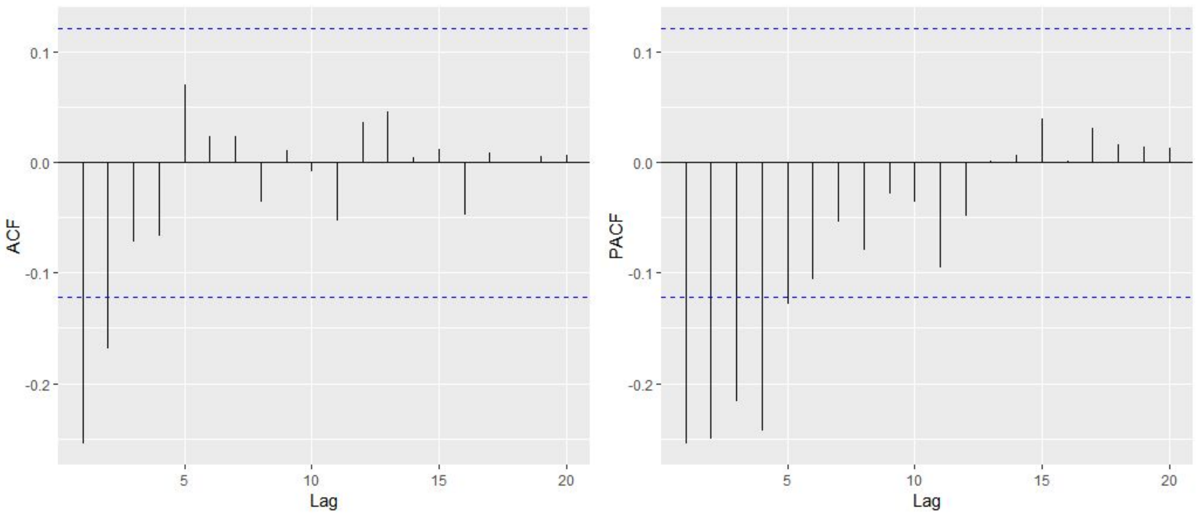
Time series and acf plots for each variable:



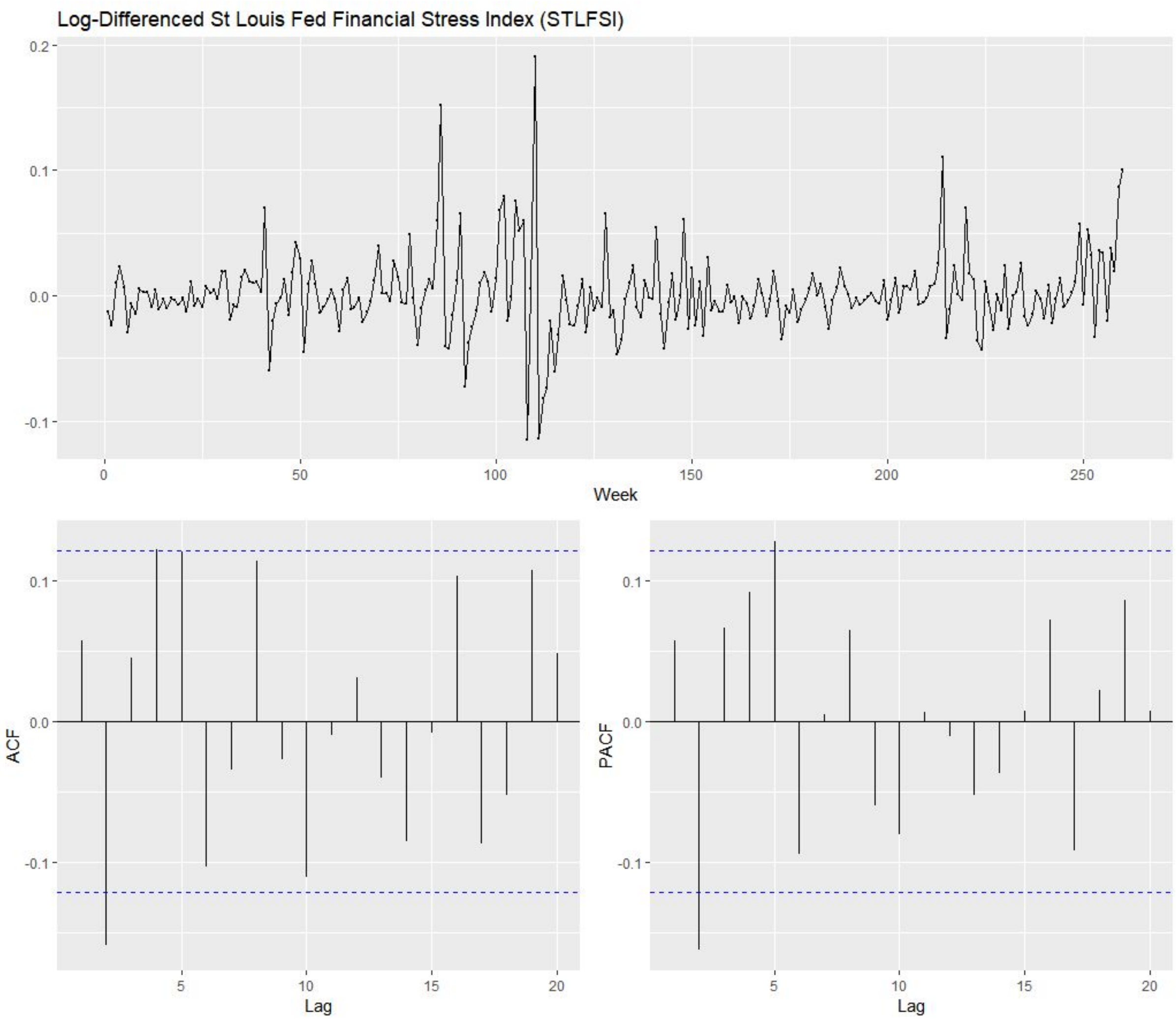
Based on the time series plot of Google Search Volume Index, it shows that it is stationary because there is no trend or seasonality.

We tried a total of six models: two baseline models (persistence and moving average), holt exponential smoothing, simple smoothing, ARIMAX(1,0,1) and ARIMAX(2,0,1). Of the six models tested, we concluded that the results in an ARIMAX(2,0,1) model gives the smallest out-of-sample RMSE, when compared to the plain ARIMA model and the 2 baseline models.

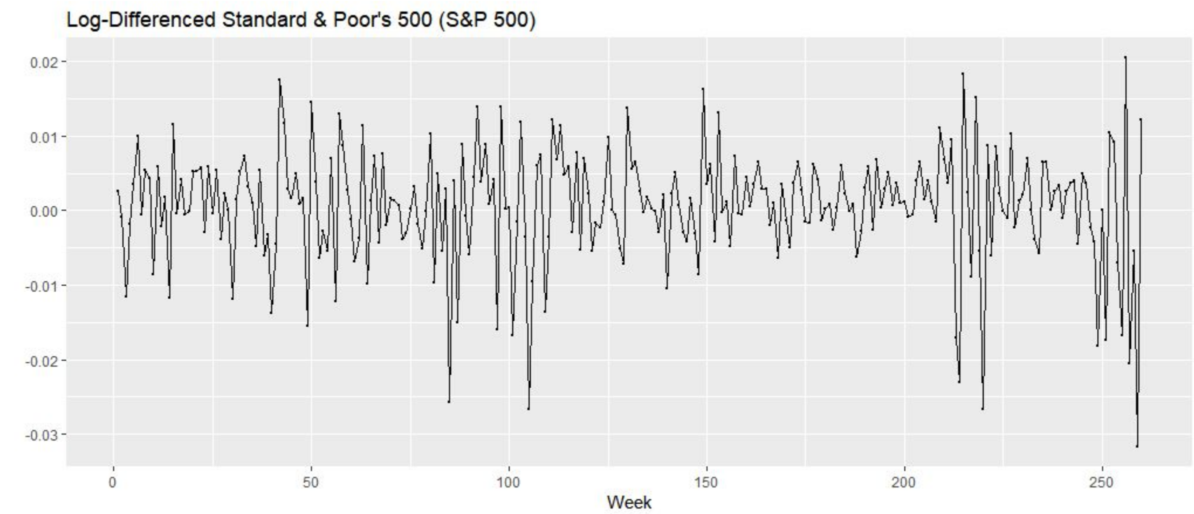


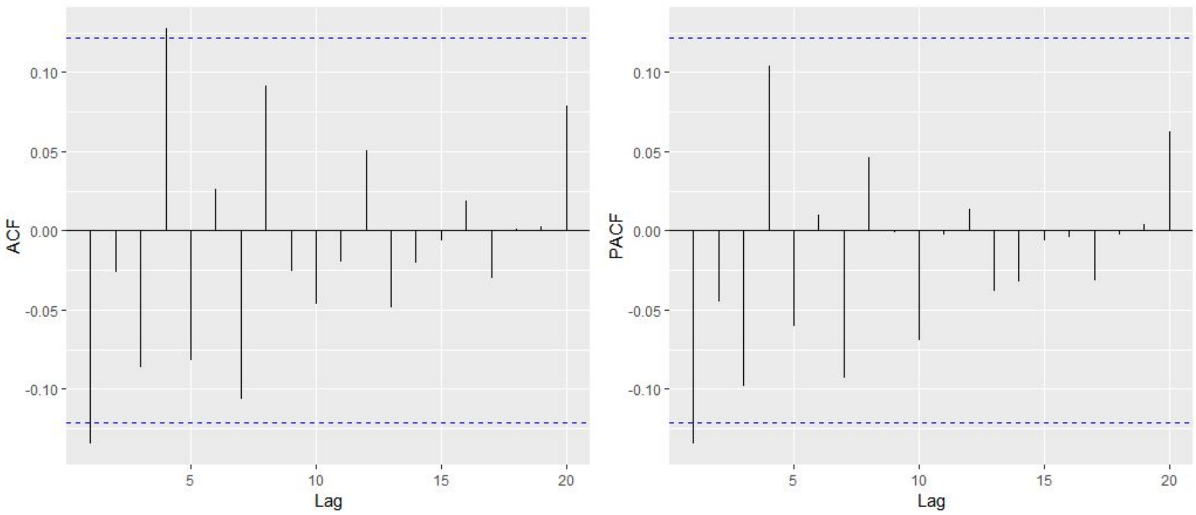


By applying percentage change in Initial Claims, it turns the time series into stationary which is illustrated in the time series plot above.

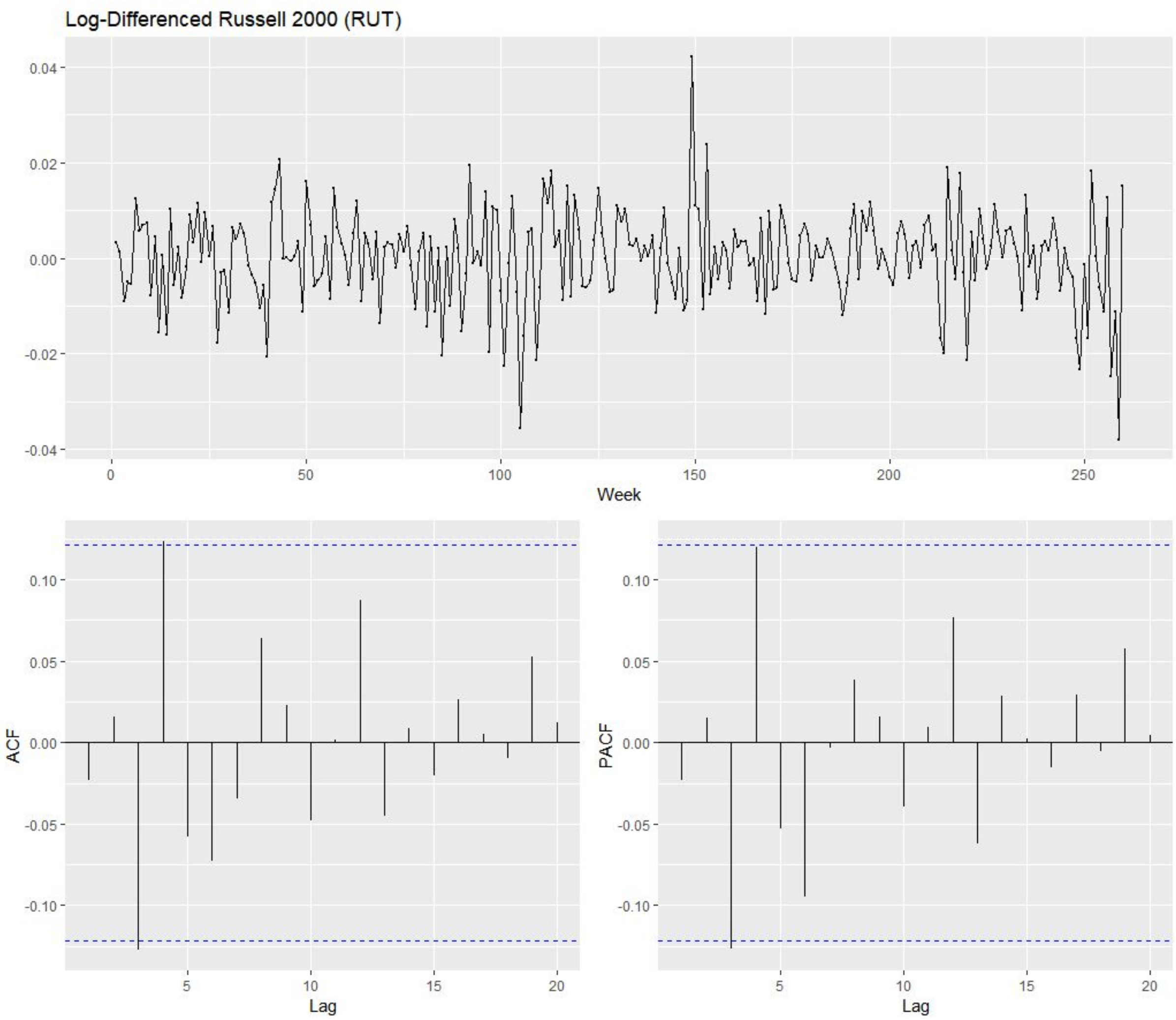


By applying logged difference in St Louis Fed Financial Stress Index (STLFSI), it turns the time series into stationary which is illustrated in the time series plot above.





By applying logged difference in Standard & Poor’s 500 (S&P 500), it turns the time series into stationary which is illustrated in the time series plot above.



By applying logged difference in Russell 2000 (RUT), it turns the time series into stationary which is illustrated in the time series plot above.

Outline analysis

We decided to to go with ARIMAX for our final model; we chose ARIMAX as it is best model out of the six models we have built, it also allows us to use other explanatory variables for forecasting.

We did not do any transformation to our dependent variable -- SVI values, as it is stationary and the it also has a fixed range of zero to 100, we are worried that applying any transformation may distort its value.

We decided to use percentage change in ICSA instead of raw ICSA values, so as to make it stationary. Initially we took the log-differences of the ICSA value, the approximations become a lot less accurate as they further apart from zero. For this reason, we opted for percentage changes in ICSA instead of log-differences, because ICSA values are pretty big (in the 100000s). For the remaining three explanatory variables, we applied log-differencing. We also lagged our explanatory variables by 2 so it captures the stickiness in economy.

Contributions Section

Idea Generation:

We originally wanted to do stock index forecasting, but as we learned that there are also multiple teams doing the similar project, we pivoted to forecast future US jobless claims.

Team Formation:

We formed our team on the first week of class, Chris and Alan have known each other for two years so they knew they were going to do the project together, and Carson approached Alan on the first day of class to form a team of three.

Ordering of names of authors:

We ordered the name by first letter of our first name by descending alphabetical order.

Teamwork:

After deciding on the final topic, we had a general meeting to look up some potential predictors of jobless claims in US. After doing some researches, Carson suggested to include St. Louis Fed Financial Stress Index in to the model as it reflects the stress in economy, and it is related to how businesses are doing in short run. Chris suggested to include stock index such as S&P 500 and Russell 2000 as the indexes reflect how well the top companies are doing; he also suggested to do log difference to the indexes before proceeding to perform forecasting. Alan and Chris wrote the majority of the final written submission while Carson did majority of the coding. Everyone on the team contributed to both parts of the project.

Appendix

R Code

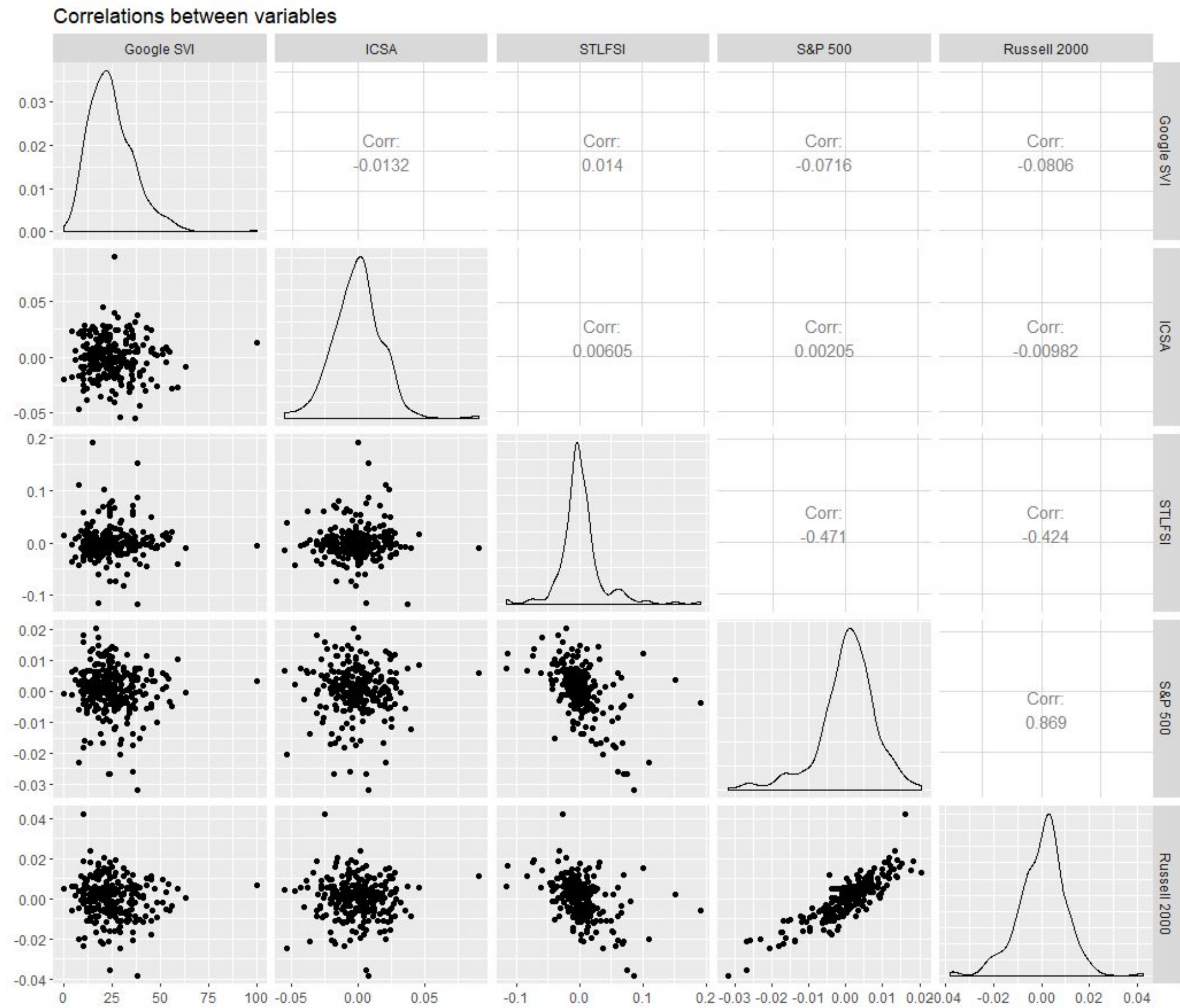
```
### ======================= STAT 443 Term Project ======================= ###

# Project Title: How Do Economic Indicators Impact the General Public's Interest in US Jobless
Claims?
# Dependent variable: Google Search Volume Index (SVI)
# Explanatory variables: Initial Claims, St Louis Fed Financial Stress Index, Standard & Poor's
500, Russell 2000

# This results in an ARIMAX(2,0,1) model that gives the smallest out-of-sample RMSE, when
compared to the plain ARIMA model and the 2 baseline models.

# =========================== Beginning of Script =========================== #

# Clean workspace and set working directory
rm(list=ls())
setwd("E:/2018-19 W2/STAT 443/Term Project"); getwd()

# Load required packages
pacman::p_load(readxl, aTSA, xts, forecast, ggplot2, GGally, xtable)

# Import data
Data<-read_excel("Data.xlsx", sheet="Data", col_types=c("date", "numeric", "numeric",
"numeric", "numeric", "numeric"))

# Define time series data and set up matrix of explanatory variables
SVI<-ts(Data$`Google SVI`); ICSA<-ts(Data$ICSA); STLFSI<-ts(Data$STLFSI);
SP500<-ts(Data$`S&P 500`); RUT<-ts(Data$`Russell 2000`)
Data2<-Data; Data2$Week<-NULL; Data2$`Google SVI`<-NULL; Matrix<-data.matrix(Data2);
rm(Data2)

# Define training and holdout sets (Training set: 1st to 210th observations | Holdout set: 211th to
260th observations)
SVI_train<-SVI[1:210]; SVI_holdout<-SVI[211:260]
```

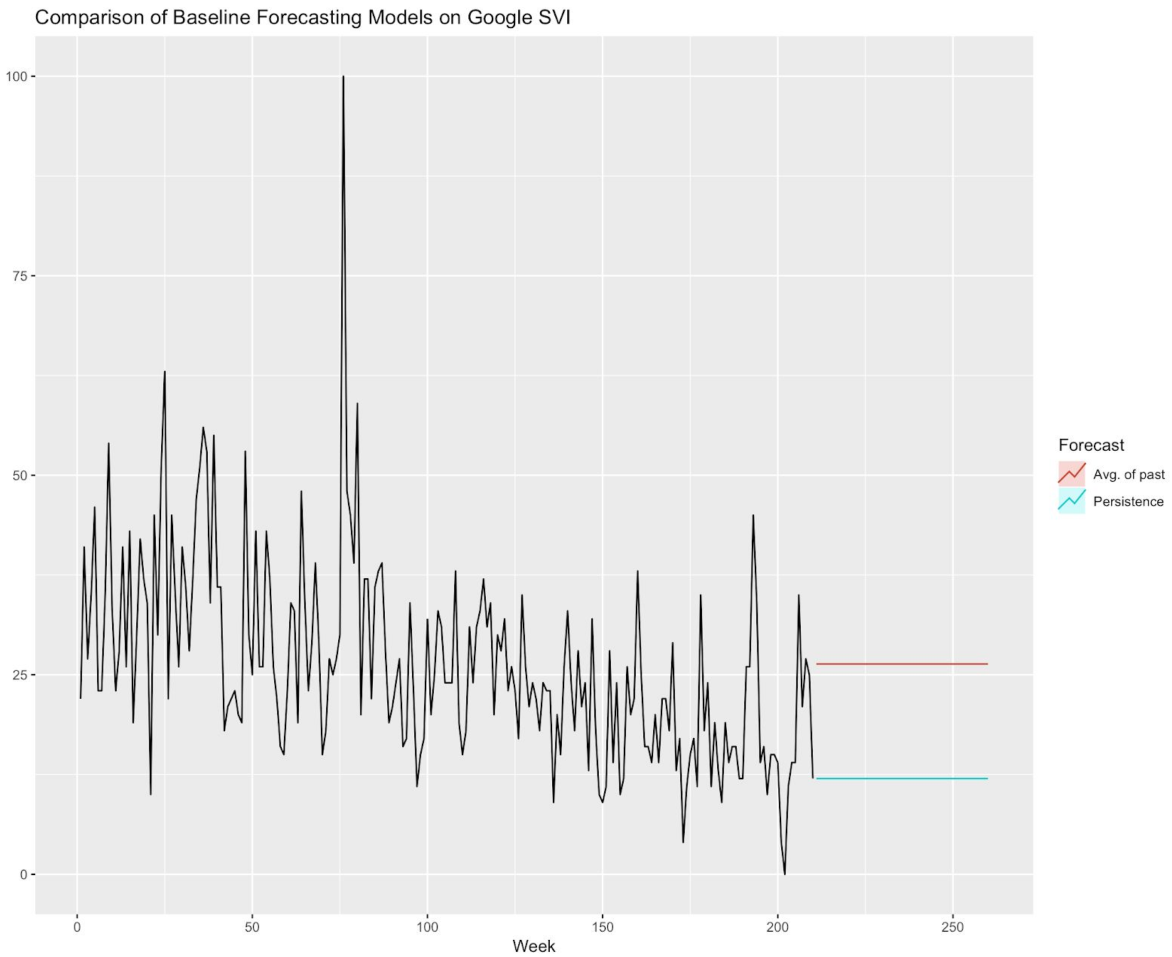
```
Matrix_train<-Matrix[1:210, 1:4]; Matrix_holdout<-Matrix[211:260, 1:4]

# ===================== Testing for stationarity in variables ===================== #

# Google Search Volume Index (SVI)
stationary.test(SVI, method="adf")# Stationary except if we don't allow for drift and trend at lags 2, 3 and 4

# Percentage Change in Initial Claims (ICSA)
stationary.test(ICSA, method="adf")# Stationary

# Log-differenced St Louis Fed Financial Stress Index (STLFSI)
stationary.test(STLFSI, method="adf")# Stationary

# Log-differenced Standard & Poor's 500 (S&P 500)
stationary.test(SP500, method="adf")# Stationary

# Log-differenced Russell 2000 (RUT)
stationary.test(RUT, method="adf")# Stationary

# ============================ Correlation plots =========================== #

# Simple plot of correlations between variables
Data2<-Data; Data2$Week<-NULL; Figure_1<-ggpairs(Data2, title="Correlations between variables"); rm(Data2)

# Autocorrelation (ACF) & partial autocorrelation (PACF) plots for all variables
Figure_2<-ggtsdisplay(SVI, lag.max=20, main="Google Search Volume Index (SVI)", xlab="Week")
Figure_3<-ggtsdisplay(ICSA, lag.max=20, main="Percentage Change in Initial Claims (ICSA)", xlab="Week")
Figure_4<-ggtsdisplay(STLFSI, lag.max=20, main="Log-Differenced St Louis Fed Financial Stress Index (STLFSI)", xlab="Week")
Figure_5<-ggtsdisplay(SP500, lag.max=20, main="Log-Differenced Standard & Poor's 500 (S&P 500)", xlab="Week")
```

```
Figure_6<-ggtsdisplay(RUT, lag.max=20, main="Log-Differenced Russell 2000 (RUT)",
xlab="Week")

# =================Out-of-sample RMSEs of baseline models =================== #

# Persistence (random walk) model
fc<-NULL; zt<-SVI_holdout; fcerror<-NULL; MSE<-NULL; fc[1]<-SVI_train[210]
for(i in 2:50){
  fc[i]<-SVI_holdout[i-1]
}
fcerror<-zt-fc; MSE<-sum(fcerror^2); RMSE_rwf<-sqrt(MSE/50); rm(fc); rm(zt); rm(fcerror);
rm(MSE); rm(i)

# Average of all previous observations
fc<-NULL; zt<-SVI_holdout; fcerror<-NULL; MSE<-NULL; cumsum<-sum(SVI_train);
fc[1]<-cumsum/210
for(i in 2:50){
  cumsum<-cumsum+SVI_holdout[i-1]
  fc[i]<-cumsum/(210+i-1)
}
fcerror<-zt-fc; MSE<-sum(fcerror^2); RMSE_avgpo<-sqrt(MSE/50); rm(fc); rm(zt); rm(fcerror);
rm(MSE); rm(cumsum); rm(i)

# Plotting forecasts with the "forecast" package
Figure_7<-autoplot(naive(ts(SVI[1:210]), h=50), main="Forecasts from Persistence (Random
Walk) Model on Google SVI", xlab="Week", ylab="")
Figure_8<-autoplot(meanf(ts(SVI[1:210]), h=50), main="Forecasts from Average of Past
Observations Model on Google SVI", xlab="Week", ylab="")
Figure_9<-autoplot(ts(SVI[1:210])) + autolayer(meanf(ts(SVI[1:210]), h=50), series="Avg. of
past", PI=FALSE) + autolayer(naive(ts(SVI[1:210]), h=50), series="Persistence", PI=FALSE) +
  ggtitle("Comparison of Baseline Forecasting Models on Google SVI") + xlab("Week") + ylab("")
+ guides(colour=guide_legend(title="Forecast"))

# ==================== Fitting ARIMA/ARIMAX models ==================== #
# ARIMA(1,0,1) model (without explanatory variables) # out-of-sample RMSE lower than
baseline models
```

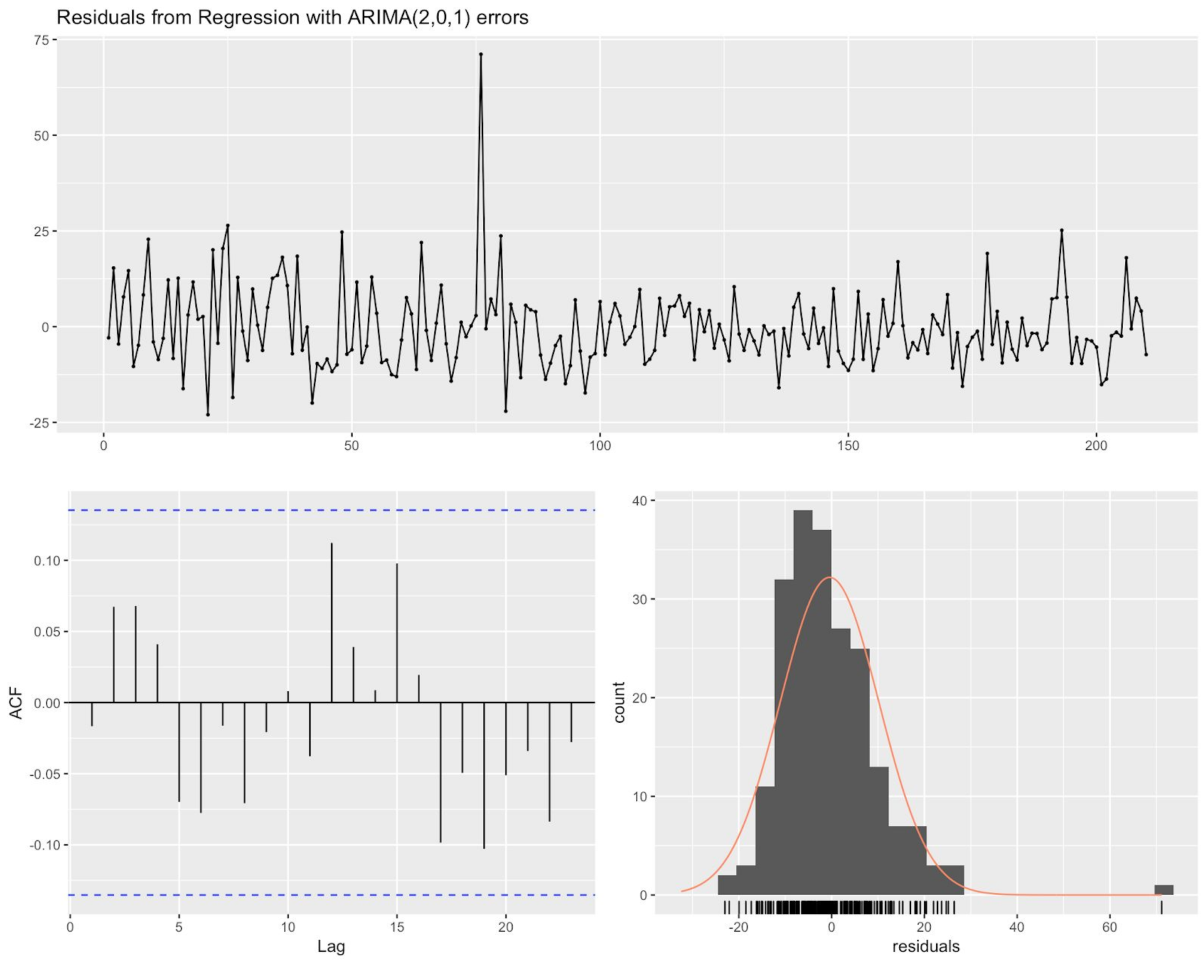
```
Model1<-auto.arima(SVI_train, approximation=FALSE, test="adf"); print(Model1);
checkresiduals(Model1); Figure_10<-recordPlot()
Model1_fc<-forecast(Model1, h=50); fcerror<-SVI_holdout-Model1_fc$mean;
MSE<-sum(fcerror^2); RMSE_Model1<-sqrt(MSE/50); rm(fcerror); rm(MSE)
Figure_11<-autoplot(Model1_fc, main="Forecasts from ARIMA(1,0,1) on Google SVI",
xlab="Week", ylab="")

# ARIMAX(2,0,1) model (with explanatory variables) # lowest out-of-sample RMSE
Model2<-auto.arima(SVI_train, approximation=FALSE, xreg=Matrix_train, test="adf");
print(Model2); checkresiduals(Model2); Figure_12<-recordPlot()
Model2_fc<-forecast(Model2, h=50, xreg=Matrix_holdout);
fcerror<-SVI_holdout-Model2_fc$mean; MSE<-sum(fcerror^2); RMSE_Model2<-sqrt(MSE/50);
rm(fcerror); rm(MSE)
Figure_13<-autoplot(Model2_fc, main="Forecasts from ARIMAX(2,0,1) on Google SVI",
xlab="Week", ylab="")

# ===================== End of Script ===================== #
print(Figure_1)# Figure 1: Correlations bewteen variables
print(Figure_2)# Figure 2: Google Search Volume Index (SVI)
print(Figure_3)# Figure 3: Percentage Change in Initial Claims (ICSA)
print(Figure_4)# Figure 4: Log-Differenced St Louis Fed Financial Stress Index (STLFSI)
print(Figure_5)# Figure 5: Log-Differenced Standard & Poor's 500 (S&P 500)
print(Figure_6)# Figure 6: Log-Differenced Russell 2000 (RUT)
print(Figure_7)# Figure 7: Forecasts from Persistence (Random Walk) Model on Google SVI
print(Figure_8)# Figure 8: Forecasts from Average of Past Observations Model on Google SVI
print(Figure_9)# Figure 9: Comparison of Baseline Forecasting Models on Google SVI
print(Figure_10)# Figure 10: Residuals from ARIMA(1,0,1) model
print(Figure_11)# Figure 11: Forecasts from ARIMA(1,0,1) on Google SVI
print(Figure_12)# Figure 12: Residuals from ARIMA(2,0,1) model
print(Figure_13)# Figure 13: Forecasts from ARIMAX(2,0,1) on Google SVI

### ===================== END OF DOCUMENT ===================== ###
```

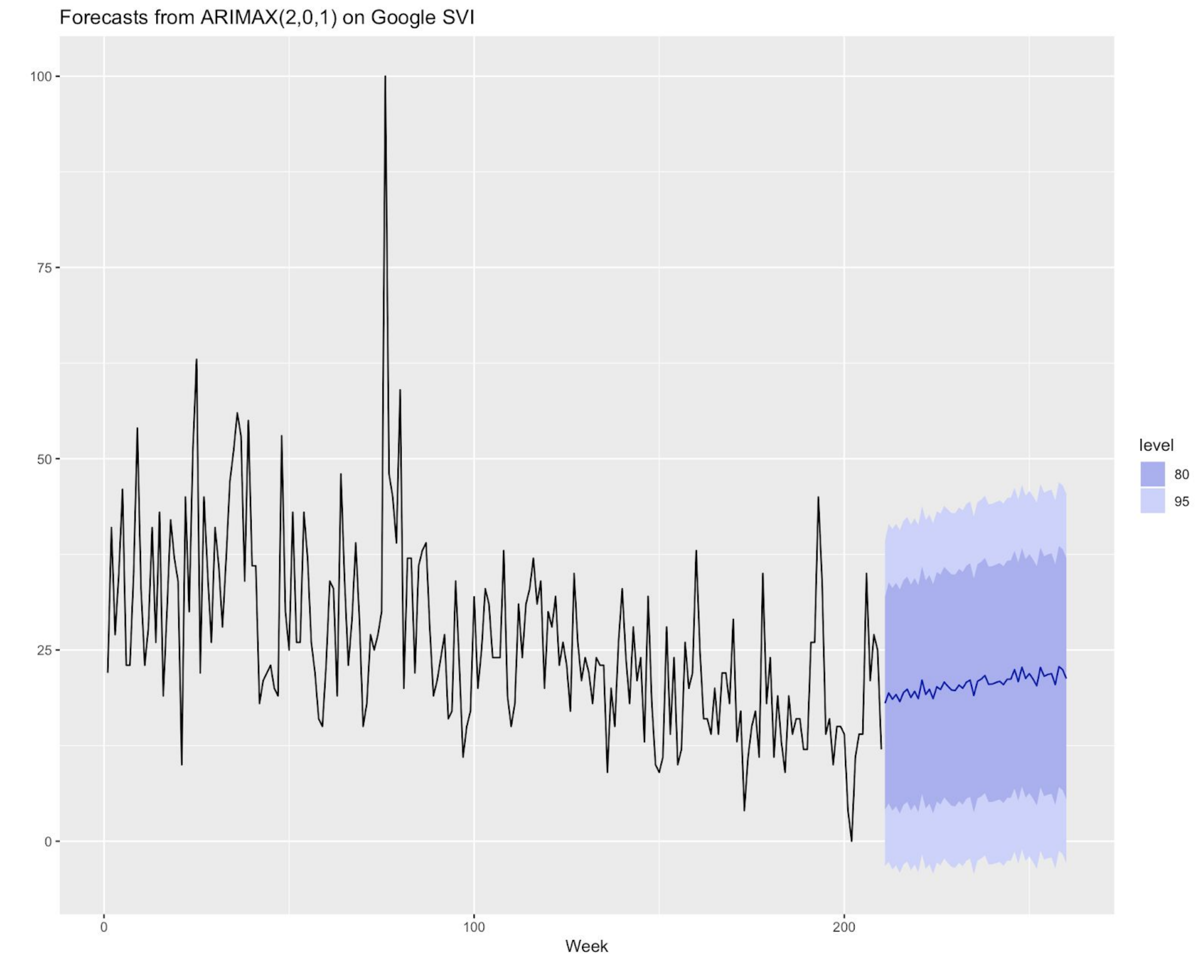
Secondary plots





Comparison of Baseline Forecasting Models on Google SVI
100
75
50
25
0
0
50
100
150
200
250
Week
Forecast
Avg. of past
Persistence


Residuals from Regression with ARIMA(2,0,1) errors
75
50
25
0
-25
0
50
100
150
200
ACF
0.10
0.05
0.00
-0.05
-0.10
0
5
10
15
20
Lag
count
40
30
20
10
0
-20
0
20
40
60
residuals


Forecasts from ARIMAX(2,0,1) on Google SVI
100
75
50
25
0
0
100
200
Week
level
80
95