US Domestic Flight Passenger Data

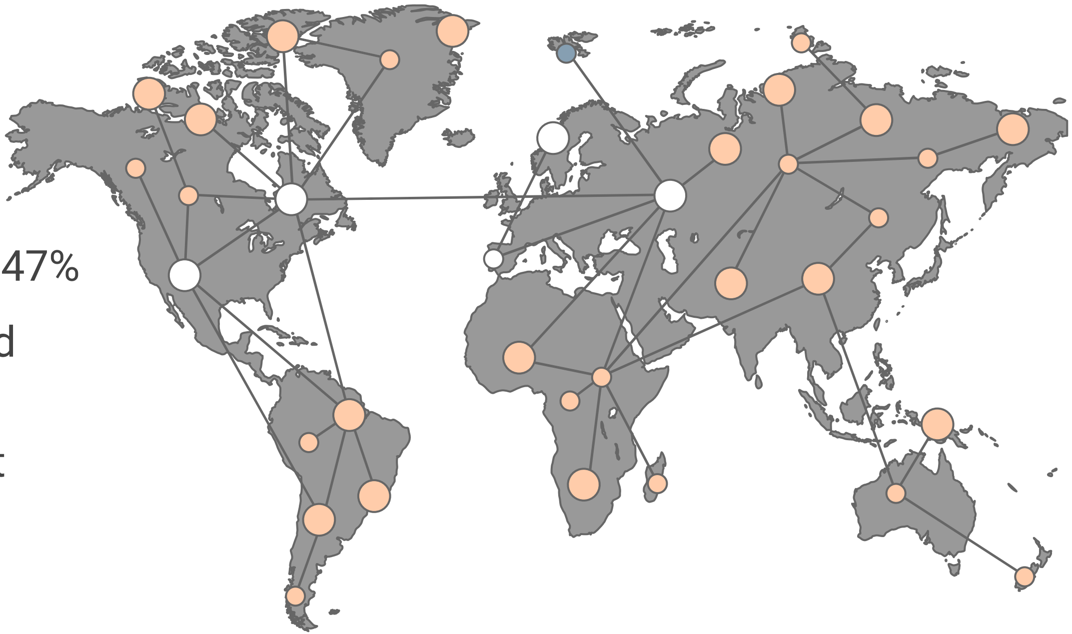
Presented by:
Elio Aybar, Lev Tyomkin, Matt Fligiel & Matt Norgren

Introduction

Over the course of 17 years:

- Domestic Air Travel has increased 47%
- International Air Travel has doubled

Related to the United States and aircraft leaving or returning to it.



Data Overview

What: Air Travel Passenger counts with All Carriers & Airports

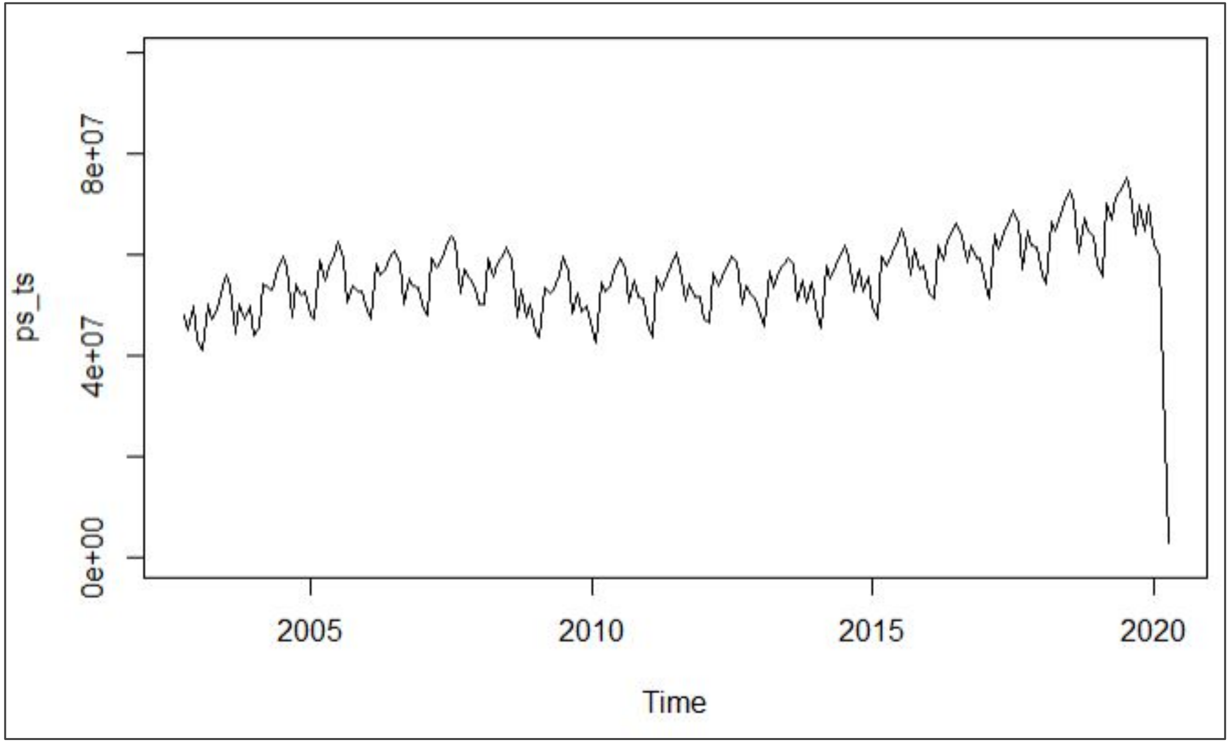
Source: U.S. Dept of Transportation

Date Range: October 2002 - December 2019 (2020 removed)

Key Features:

- Domestic Passenger Count
- International Passenger Count
- Total Passenger Count

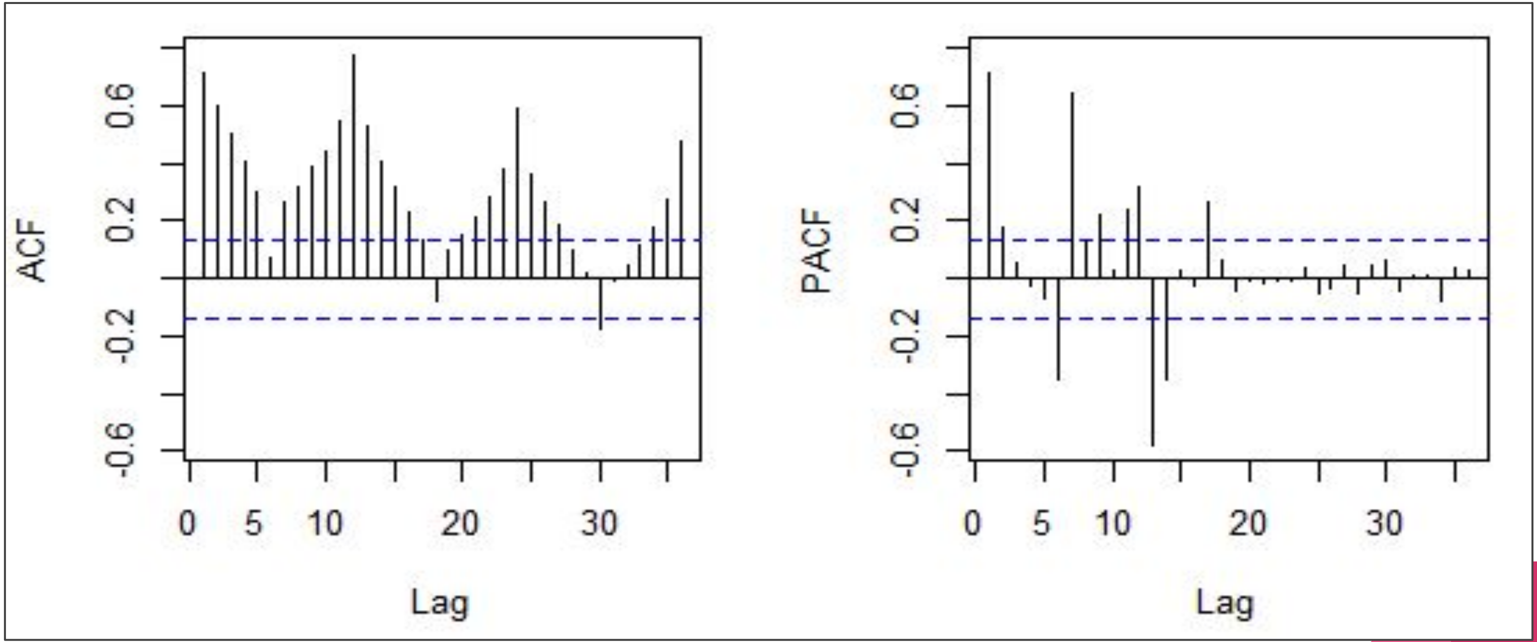
Data Overview (Cont.)



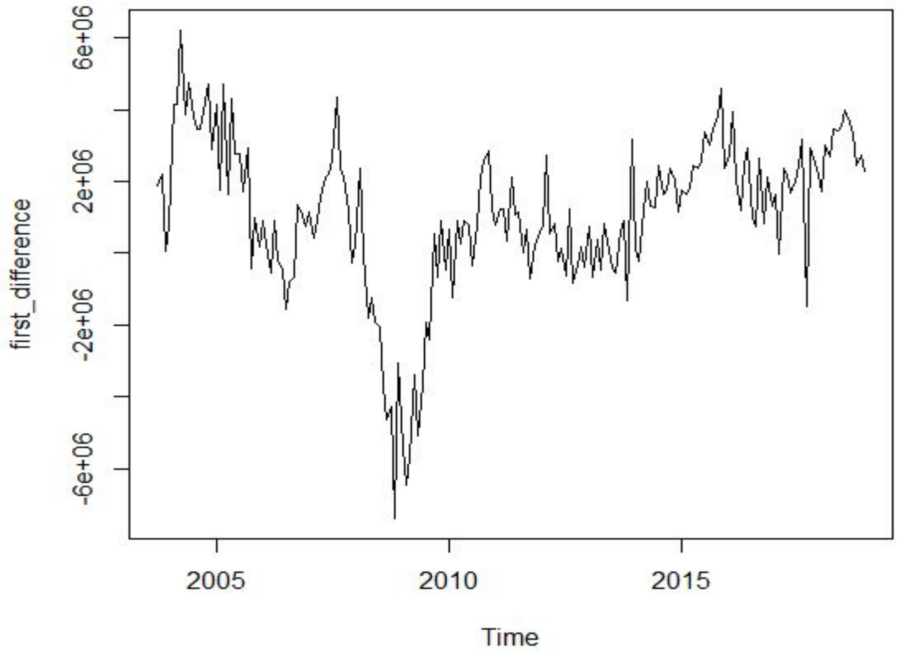
Key Features:

- Seasonality
- Trend

Key Charts (ACF/PACF) & Interpretation



First difference



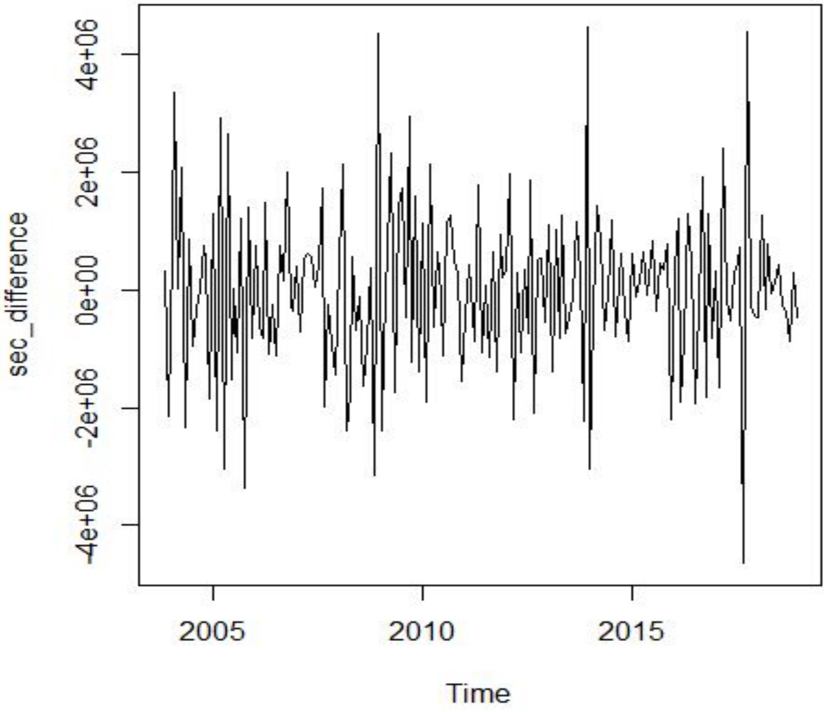
KPSS test
H0: series is stationary
HA: series is not stationary

```
        KPSS Test for Level Stationarity

data:  first_difference
KPSS Level = 0.52706, Truncation lag parameter = 4,
 p-value = 0.03557
```

Augmented Dickey-Fuller test
H0: series is not stationary
HA: series is stationary

```
        Augmented Dickey-Fuller Test

data:  first_difference
Dickey-Fuller = -2.7333, Lag order = 5, p-value = 0.2696
alternative hypothesis: stationary
```

Second difference



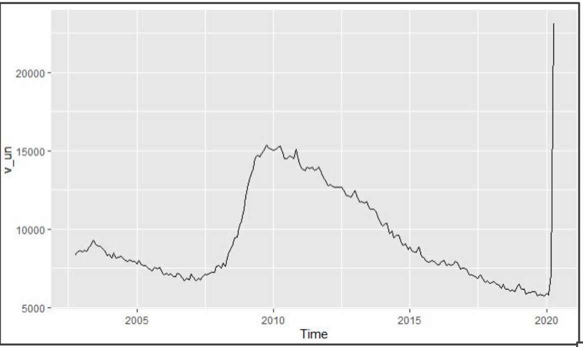
KPSS test
H0: series is stationary
HA: series is not stationary

```
        KPSS Test for Level Stationarity

data:  sec_difference
KPSS Level = 0.043021, Truncation lag parameter = 4,
 p-value = 0.1
```

Augmented Dickey-Fuller test
H0: series is not stationary
HA: series is stationary

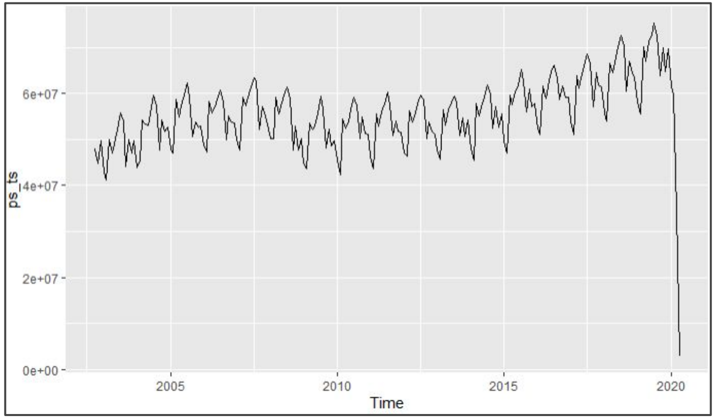
```
        Augmented Dickey-Fuller Test

data:  sec_difference
Dickey-Fuller = -6.6216, Lag order = 5,
p-value = 0.01
alternative hypothesis: stationary
```

Ordinary Least Squares (OLS)

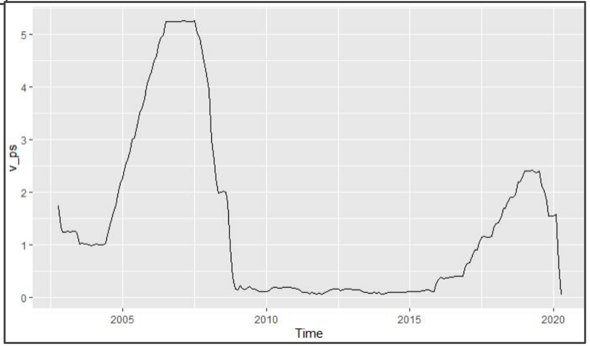


Unemployment Level

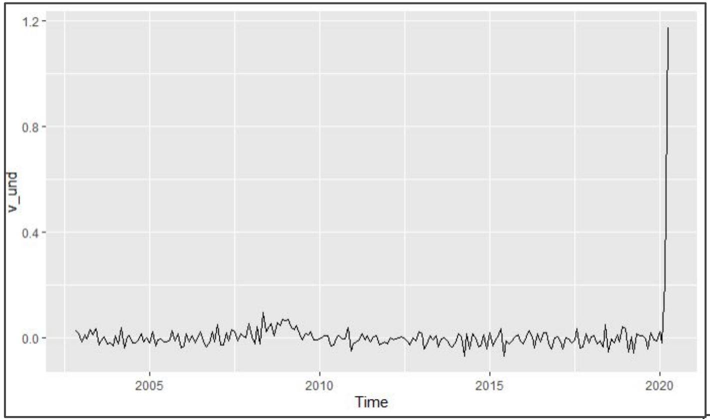
Domestic Passenger Data



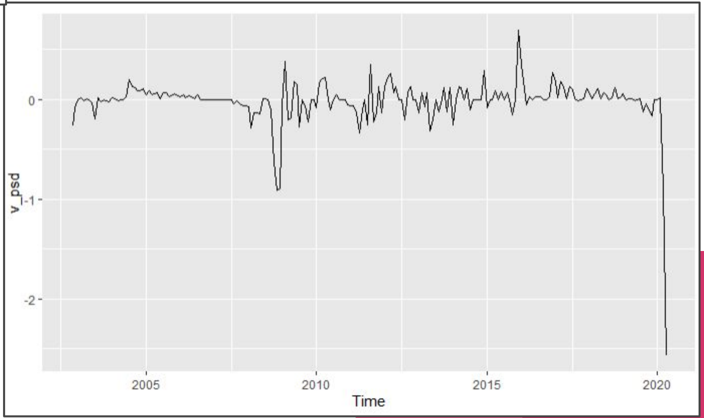
Personal Savings (US)



Ordinary Least Squares (OLS)



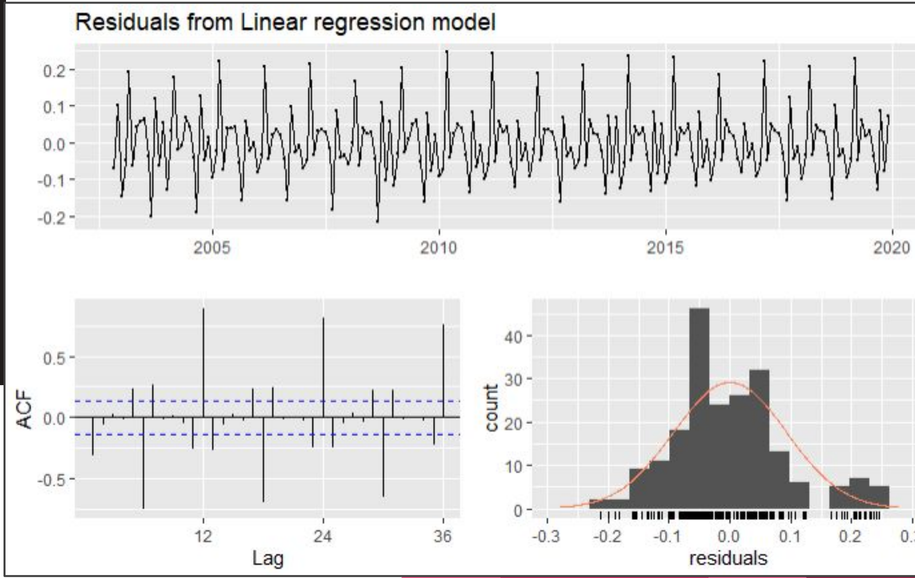
Log Differenced
Unemployment Level



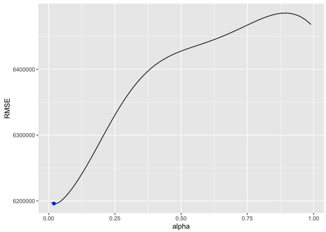
Log Differenced
Personal Savings (US)

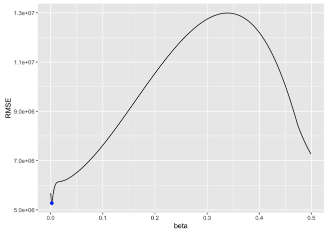
Ordinary Least Squares (OLS)

```
Call:
tslm(formula = ps_tsd ~ cbind(v_und, v_psd))

Residuals:
      Min        1Q    Median        3Q       Max 
-0.216167 -0.069341 -0.006777  0.060631  0.291563 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)              -0.003604   0.007360  -0.490    0.625    
cbind(v_und, v_psd)v_und -1.802367   0.133308 -13.520   <2e-16 ***
cbind(v_und, v_psd)v_psd  0.060850   0.047328   1.286    0.200    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1064 on 207 degrees of freedom
Multiple R-squared:  0.7141,	Adjusted R-squared:  0.7113 
F-statistic: 258.5 on 2 and 207 DF,  p-value: < 2.2e-16
```



Exponential Smoothing



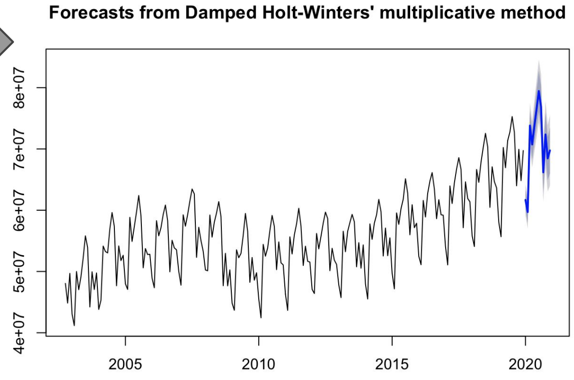


Using the fpp2() library, we applied a couple methods for Exponential Smoothing

- Simple Exponential Smoothing and Holt’s Method
    - Used an algorithm to identify the optimal alpha and beta values respectively
    - Looked at lowest RMSE in relation to x-plot of alpha/beta values from 0 - 100
    - Then forecasted 12 months on differenced data
        - Used differenced approach to try to address clear trend
- Concluded that both models overfitted the data and despite forecasting on differenced data, saw our performance metrics remain higher for test/train
    - Lesson here reaffirms that it is difficult to apply SES and Holt on data with trend and seasonality like our plane data

Holt-Winters

```
##                    ME    RMSE      MAE        MPE     MAPE      MASE       ACF1
## Training set 68685.88 1063227 821880.7 0.07729496 1.510296 0.4088466 0.01007767
```



Using hw() we forecasted the next 12 months of our T.S.

- We experimented with the following four methods
  - Linear trend with additive seasonality
  - Linear trend with multiplicative seasonality
  - Exponential smoothing
    - Linear trend with additive seasonality and damping
    - Linear trend with multiplicative seasonality and damping
- Comparing the four approaches, we identified our best Holt Winters model to be multiplicative with damped trend
  - The multiplicative approach makes sense because we, from plotting our time series object, we observe data with a positive trend but with seasonality that appears to increase over time
  - By dampening the trend to be turn flat over time, we achieve better accuracy metrics (lowest RMSE) as early as 12 months forecasted out

ARIMA

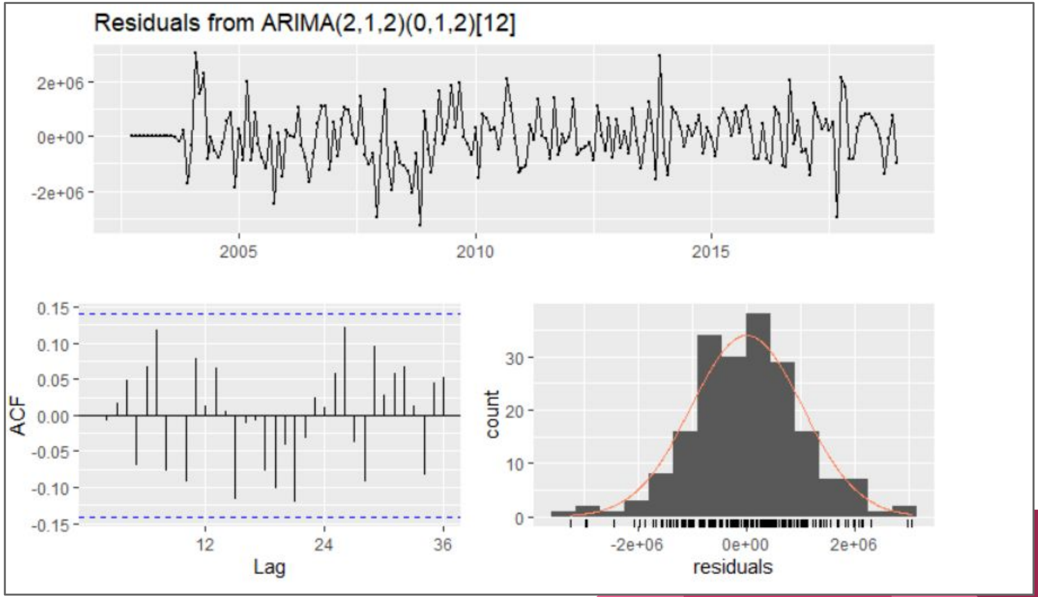
An initial Auto-ARIMA identified our best model as an ARIMA(2,1,2)(1,1,2)[12].

- This makes sense, as there is a strong seasonal pattern, as well as a trend, implying the need for differencing
- Overall, this model performed well, with AICc 5590
- Could it be trimmed?

ARIMA (Cont)

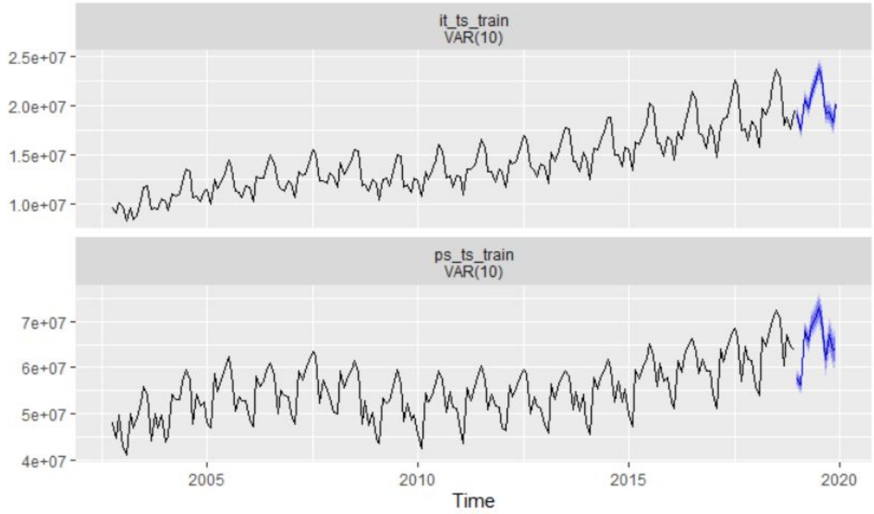
Tried multiple models

- Decreased Parameters
- Final ARIMA
  ARIMA(2,1,2)(0,1,2)[12]
  - Performed better on AICc; Error Metrics
- Residuals non concerning:



VAR Modeling

- With a combination of international and domestic flight data, we were able to create a strong VAR model
    - We tried multiple, but inevitably a high order model was required to account for seasonality



Future Work

Given the unique circumstances surrounding 2020:

- Additional data could be added
- An intervention would need to be calculated
    - Pulse Function
    - Intervention effects that die out gradually:

$$m_t = \delta m_{t-1} + \omega P_t^{(T)}$$

- A shorter window training window

Conclusion - Performance

- Ultimately, the ARIMA model did better than any others
  - While the VAR would be expected to do well, variations between patterns of international and domestic travel prevented it from performing quite as strongly as the ARIMA
  - OLS was not able to properly use the time aspect of the dataset
  - This model outperformed holt-winters as well on RMSE and other statistics

```
[1] "ARIMA(2,1,2)(0,1,2)[12]"
               ME    RMSE     MAE      MPE    MAPE       ACF1 Theil's U
Test set 733490.6 1534955 1287674 0.964031 1.89908 -0.2215569 0.2396795
[1] "Training Stats - AIC: 5588.6048156741 , AICc: 5589.24849383502 , BIC: 5611.03286248363"
```

Conclusion

Individual Contributions:

- Matt N: Intro, OLS, Future Work
- Elio: Key Charts, Data Overview, OLS
- Lev: Holt-Winters, Exponential Smoothing
- Matt F: Arima, VAR