

Forecasting O'hare Blue Line Riders

November 28, 2017
Eloy Frias
Mohd Kamaludin



Introduction Problem statement

Mayor, Rahm Emanuel, recently proposed building an express train from O'hare to downtown Chicago. The city needs a baseline for potential riders in the near future. We will use the CTA's (Chicago Transit Authority) blue line O'hare station data to forecast the number of riders.

Data Properties

Assumption: Number of passengers on proposed express train from O'hare can be projected based on number of riders on CTA O'hare Blue Line

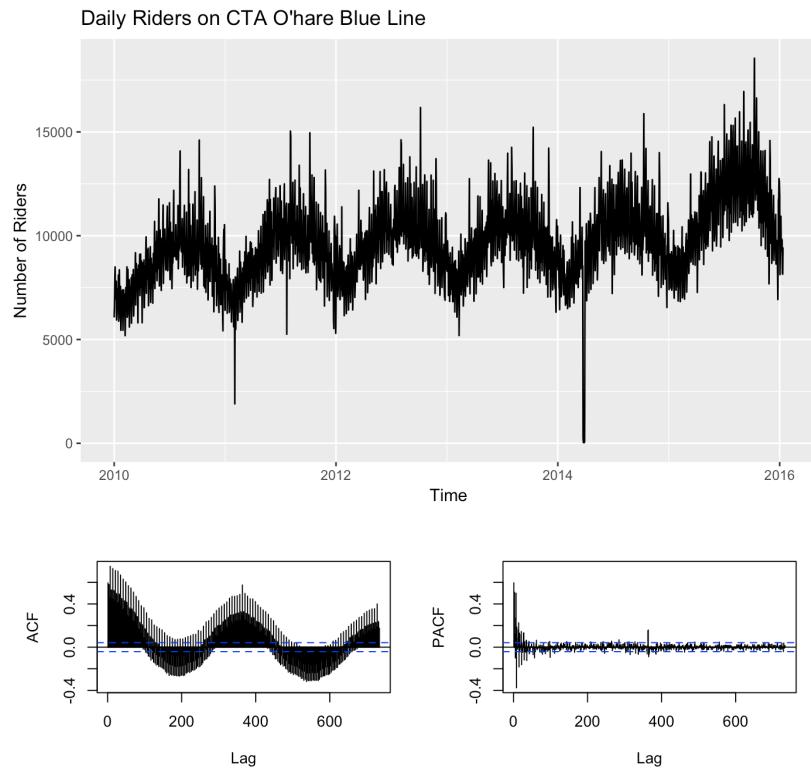
Univariate Time Series Data: Daily number of riders embark at O'hare blue line station (source: City of Chicago)

Data limitation:

- do not include riders exit from the station
- Not all Blue line riders are heading to downtown

KPSS Score

KPSS Level = 7.2686, Truncation lag parameter = 11, p-value = 0.01



Data Transformations

Daily data converted to monthly due to Arima lag limitation

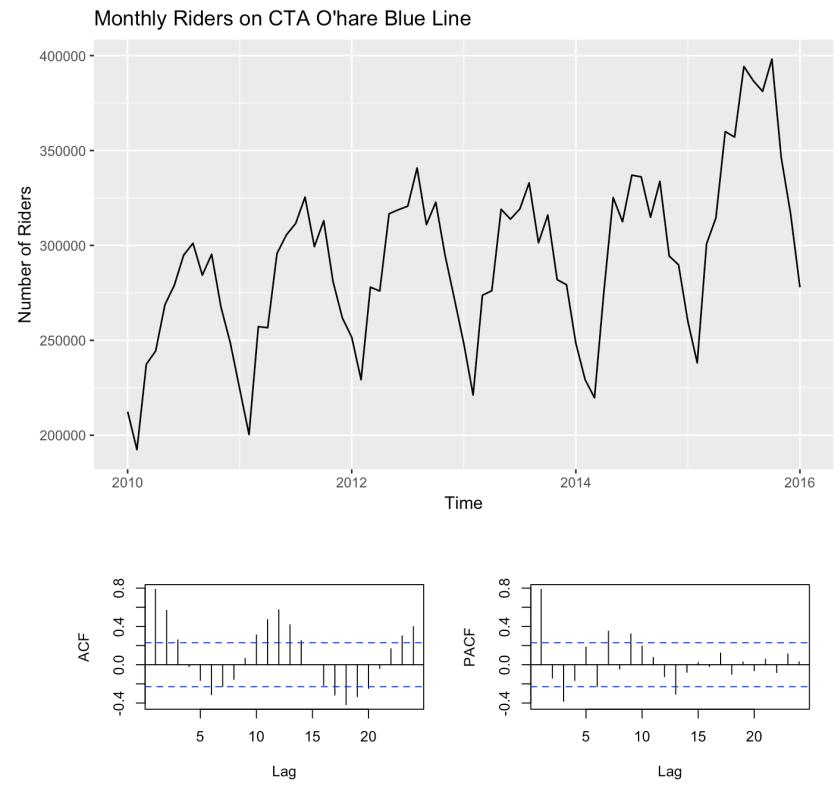
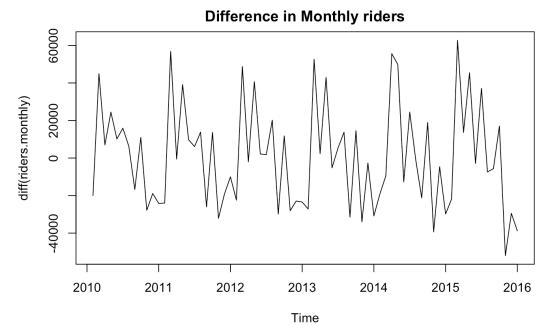
KPSS Score

KPSS Level = 1.147, Truncation lag parameter = 2, p-value = 0.01

Data is still non-stationary after transformation

KPSS on Differencing

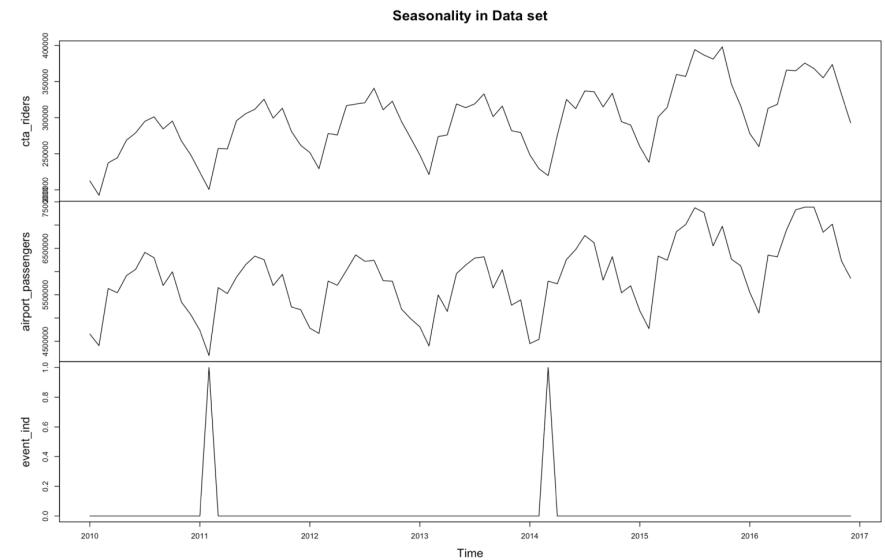
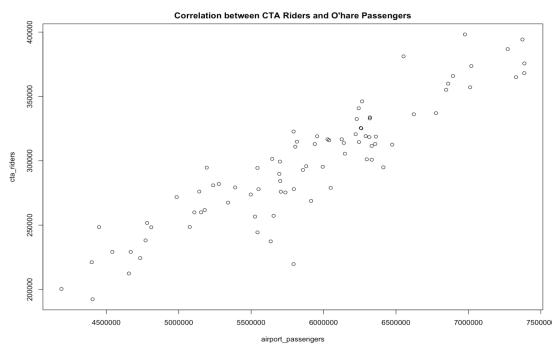
KPSS Level = 0.068291,
Truncation lag parameter = 1,
p-value = 0.1



Selecting Predictors

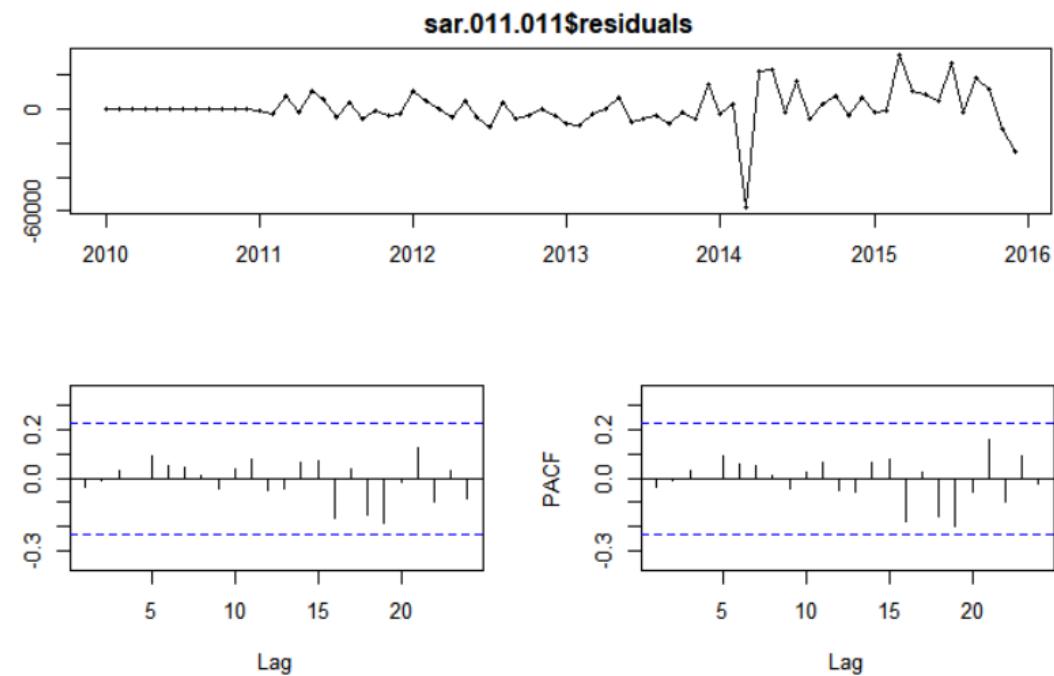
We also use predictors

- Number of Monthly Passenger Volume at O'hare airport (Source: Chicago Department of Aviation)
- Event Indicator for when CTA Blue Line is Closed/Partially Open (Source: Feature Engineering)



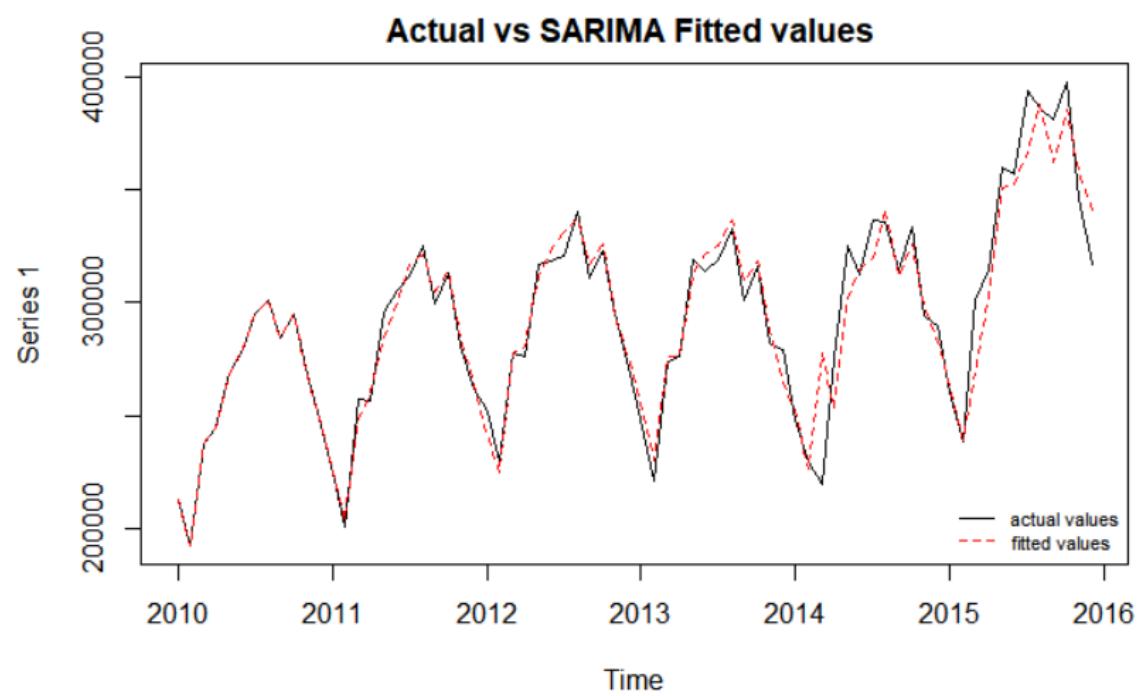
Model 1: Seasonal Arima

- SARIMA(1,0,1)(0,1,1) had the lowest AICc and BIC
- AICc of 1295 and BIC of 1301
- The ACF plot shows there is no autocorrelation of the residuals
- Ljung-Box test also shows that the residuals have no remaining autocorrelation



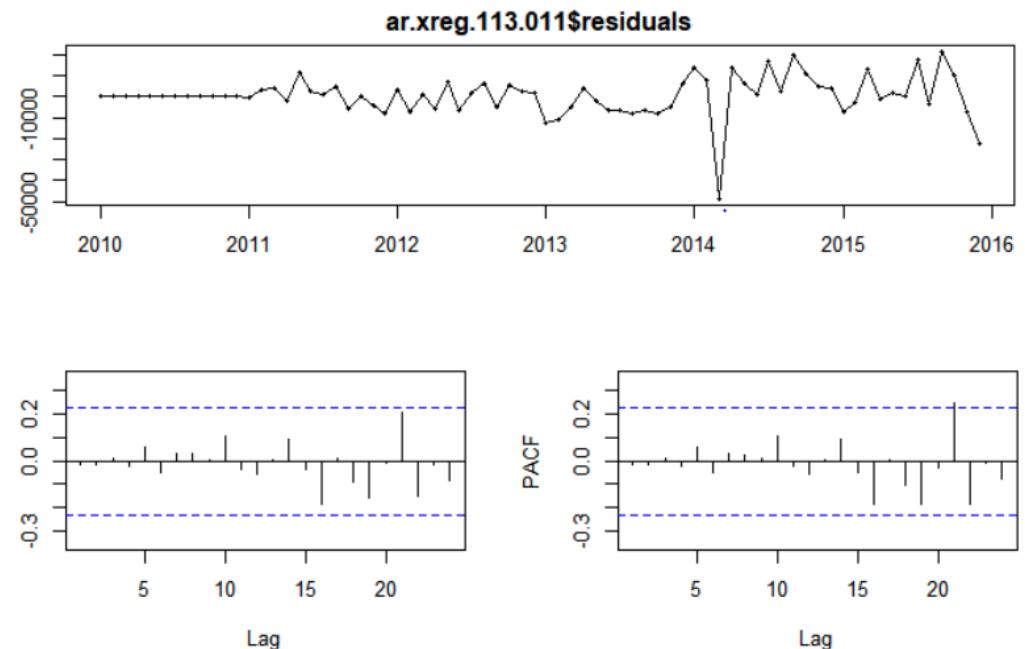
Model 1: Seasonal Arima

- The fitted values follow the trend of actual values.
- After 2011, the fitted values get progressively worse.



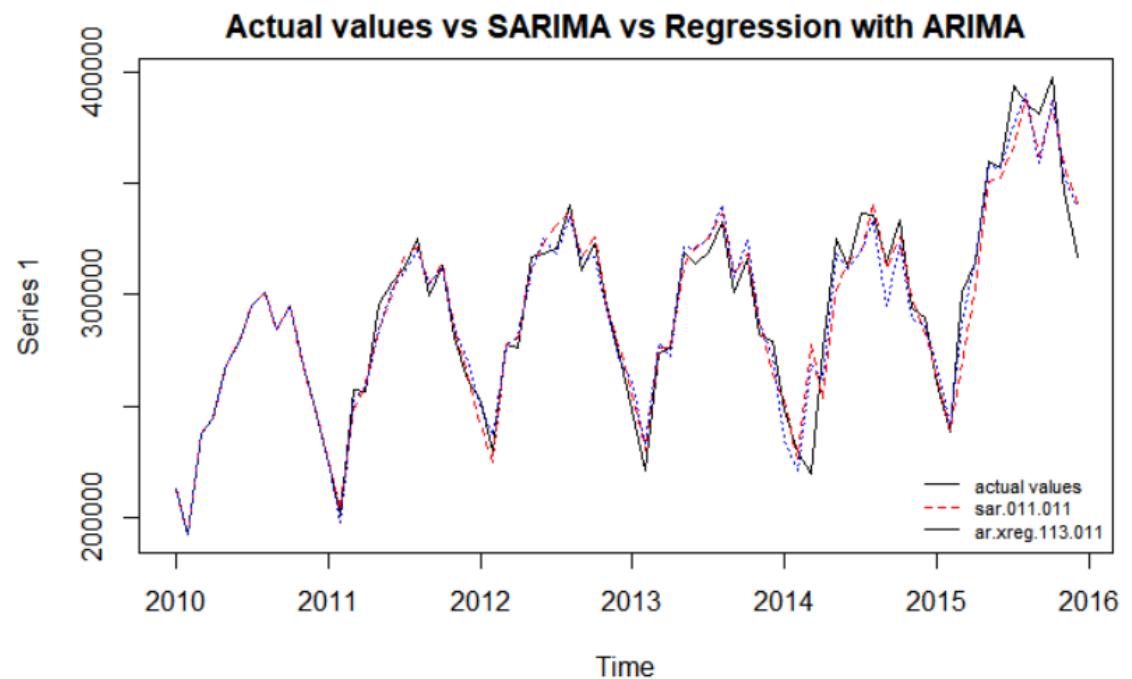
Model 2: Regression with Arima errors

- Regression with ARIMA errors has AICc of 1287 and BIC of 1301
- The ACF plot shows there is no autocorrelation of the residuals
- This model has slightly lower AICc and BIC values compared to the SARIMA model



Model 2: Regression with Arima errors

- The fitted values follow the trend of actual values
- After 2011, the fitted values get progressively worse



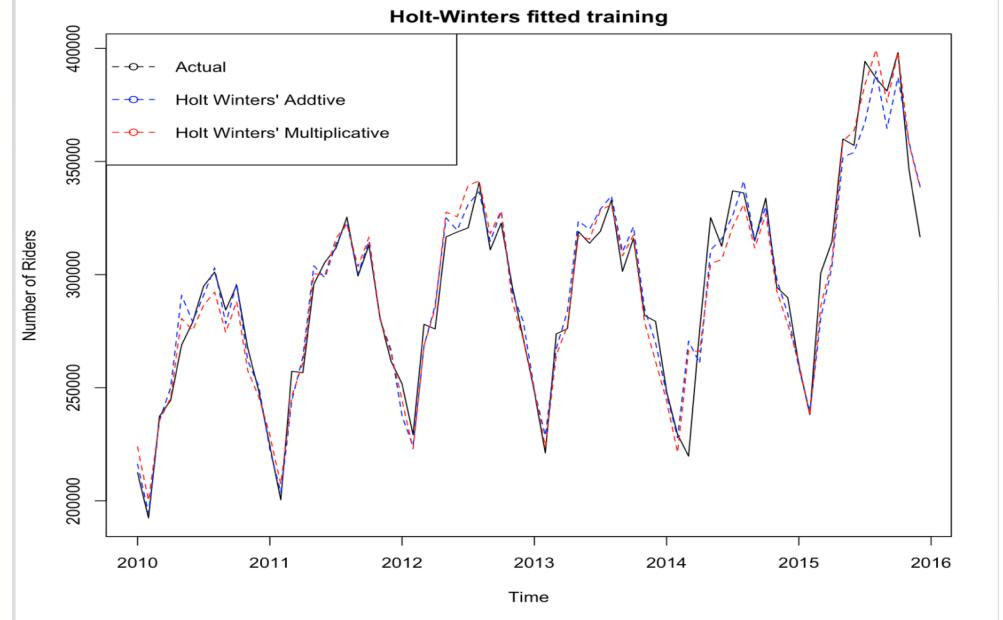
Model 3: Holt Winters

- Holt Winters model is great for Seasonality data
- Additive Model

	ME	RMSE
Training set	-363.9611	10458.70
Test set	-22917.4552	25236.62

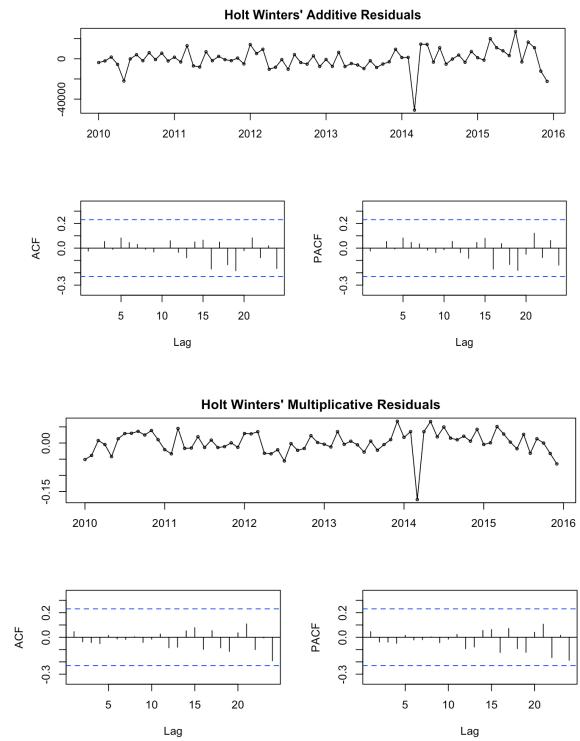
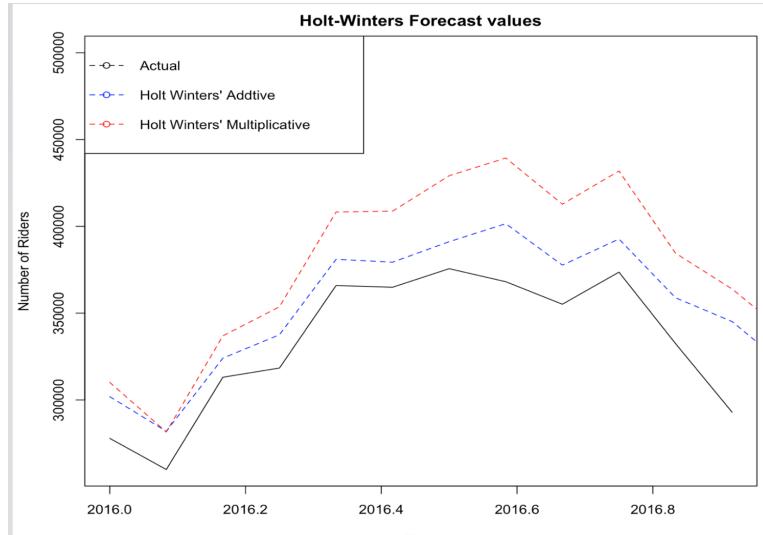
- Multiplicative Model

	ME	RMSE
Training set	223.1454	9898.676
Test set	-46915.5898	49540.340



Model 3: Holt Winters

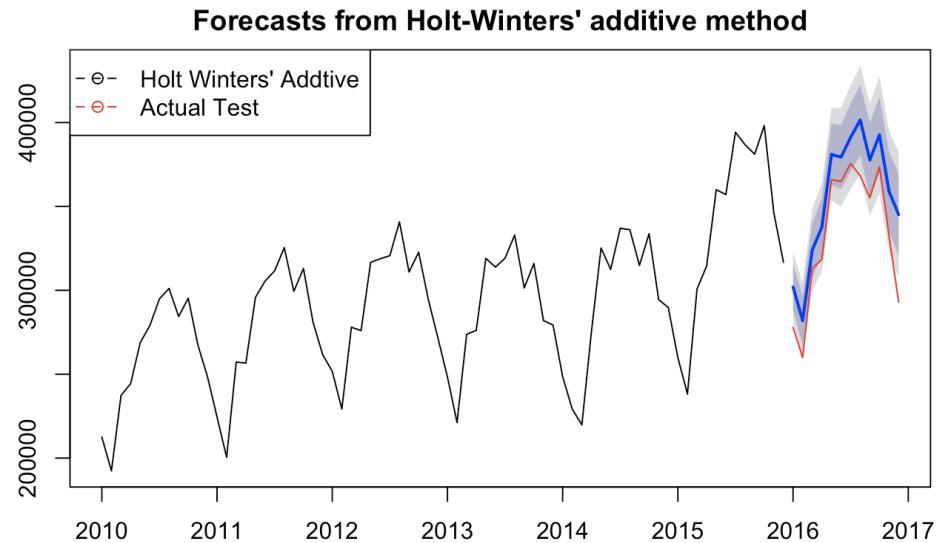
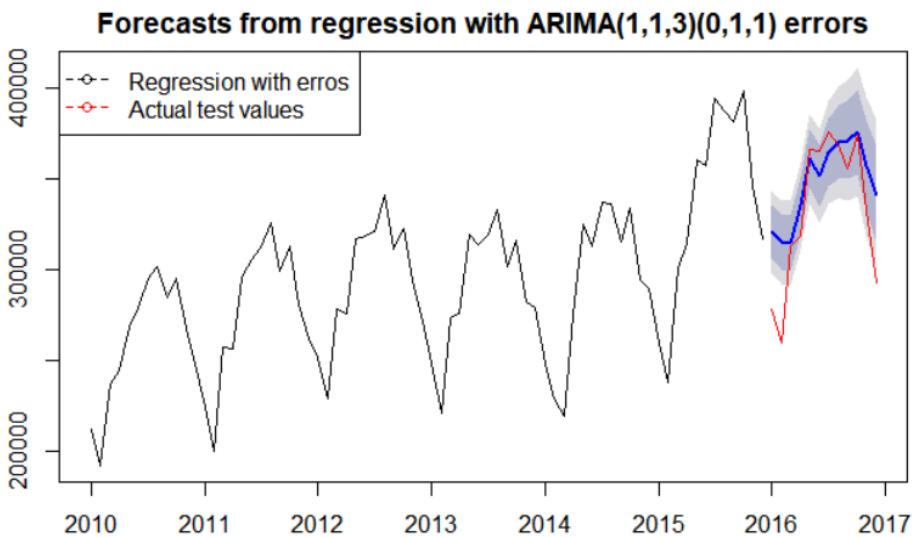
- Additive model forecast values are closer to actual values than Multiplicative model
- Using predictors in the model did not significantly change the forecast values



Result Model Comparison

Models	AIC	AICh	BIC	Train RMSE	Test RMSE	Sum of Squared Forecast Error
Arima with Xreg (1,1,3)(0,1,1)[12] errors	1287.1	1289.3	1301.64	9619.7	26869.49	8663637486
Arima with Seasonality (0,1,1) (0,1,1)	1295.47	1295.91	1301.7	11468.6		
Holt Winters Additive				10458.70	25236.62	7642642607
Holt Winters Multiplicative				9898.676	49540.340	29450943278
Holt Winters Additive with Xreg				10458.70	25236.62	7642642607

Forecast plot



Summary

- CTA Blue Line riders data clearly have seasonality pattern
- We tried multiple Time Series models with Holt Winters Additive model having the most accurate forecast
- Improve models by using more predictors:
 - Number of Taxis to/from the Airport
 - Number of Uber/Lyft to/from the Airport
 - Weather
 - Intervention model

Appendix

CTA Rider

<https://data.cityofchicago.org/Transportation/CTA-Ridership-L-Station-Entries-Daily-Totals/5neh-572f>

O'hare Passenger Volume

<https://www.ohareairport.net/passenger-statistics.shtml>

<http://www.flychicago.com/business/CDA/factsfigures/Pages/airtraffic.aspx>

O'hare express train proposal

<http://www.chicagotribune.com/news/columnists/wisniewski/ct-met-ohare-express-rfq-20171128-story.html>