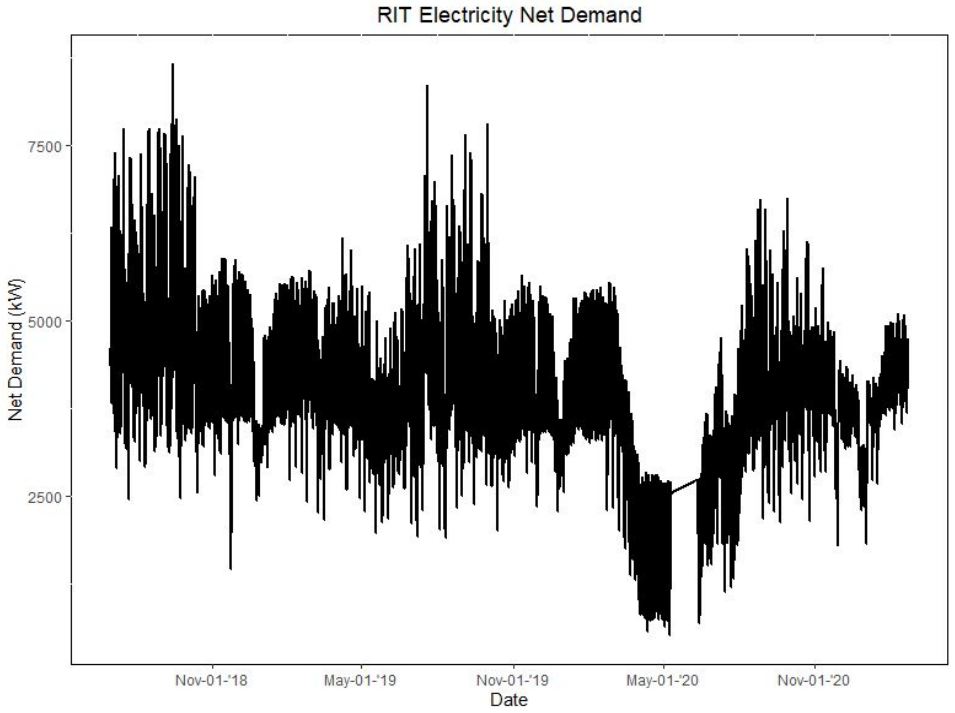
Evaluating Models to Forecast RIT's Energy Use

Aaron Kasinski & Trevor Martin

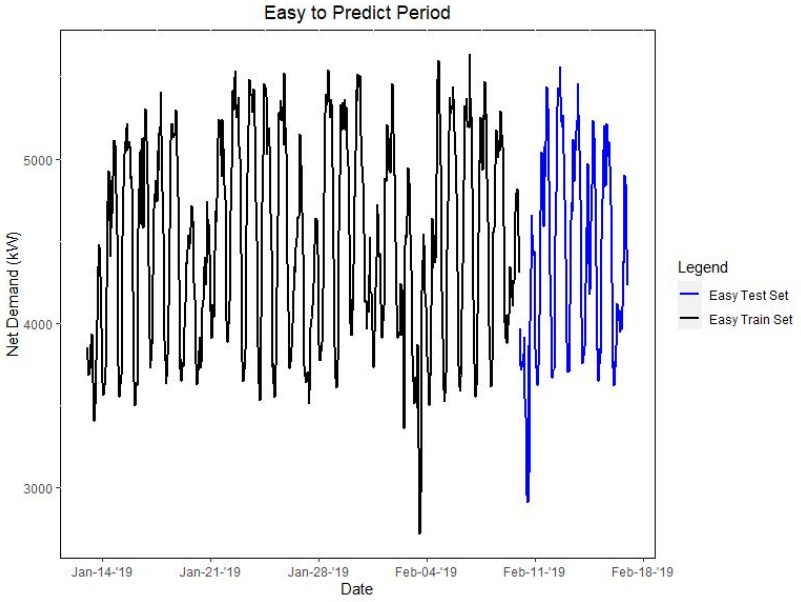
RIT Energy Data Overview

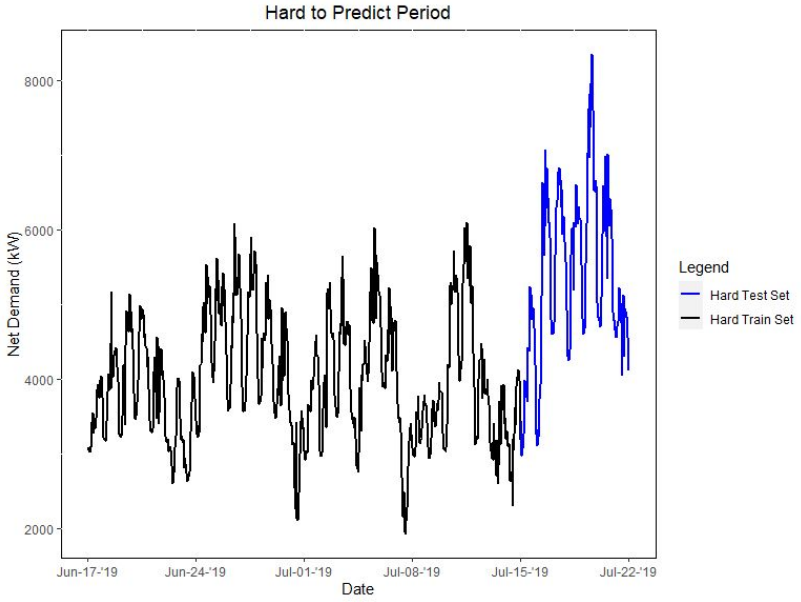
- Time series data
  - Energy demand
  - Solar production
  - Discrete and continuous predictor variables



Seasonal ARIMA Models

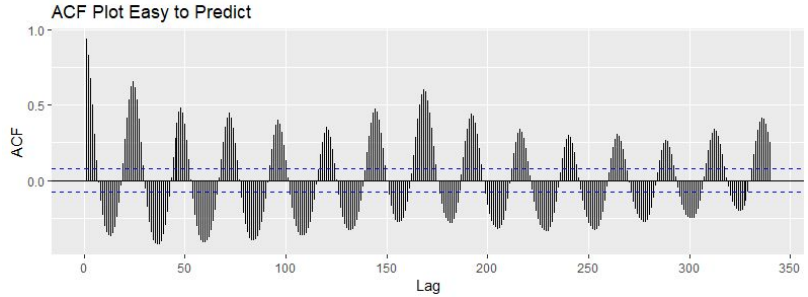
Training and Test Periods

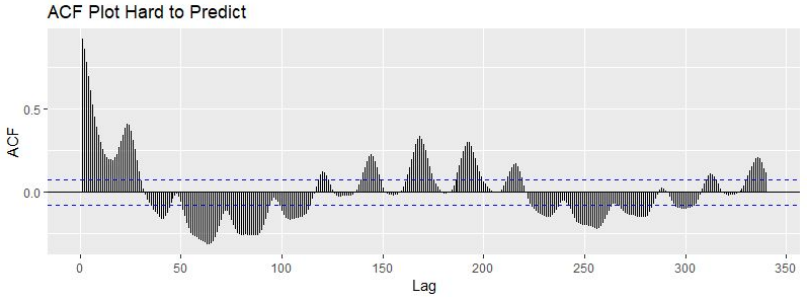


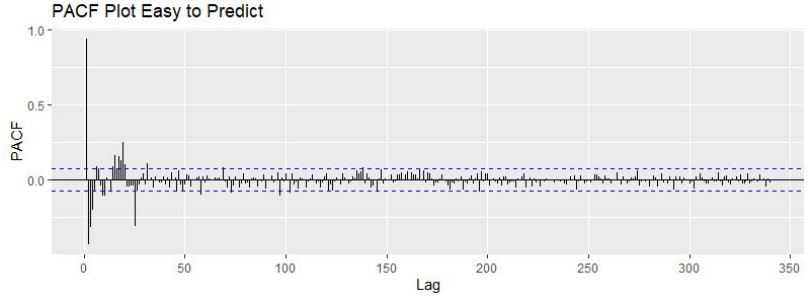


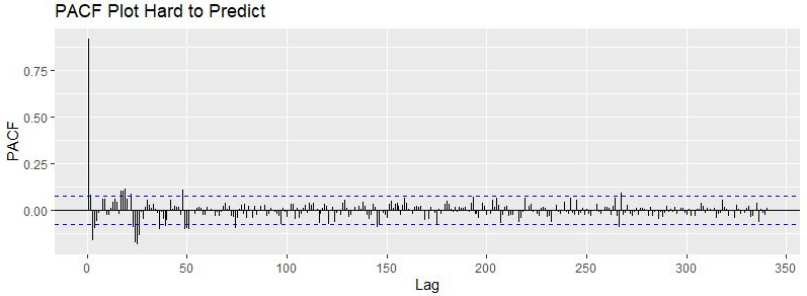
Date ranges taken from project part two.

ACF and PACF Plots of Raw Data



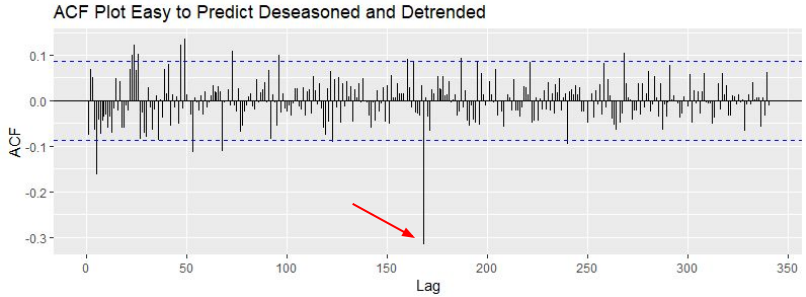


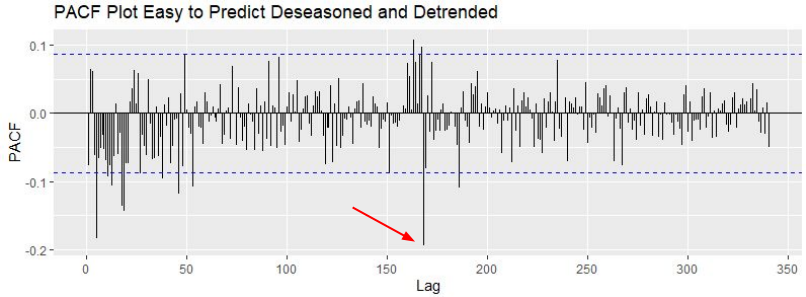


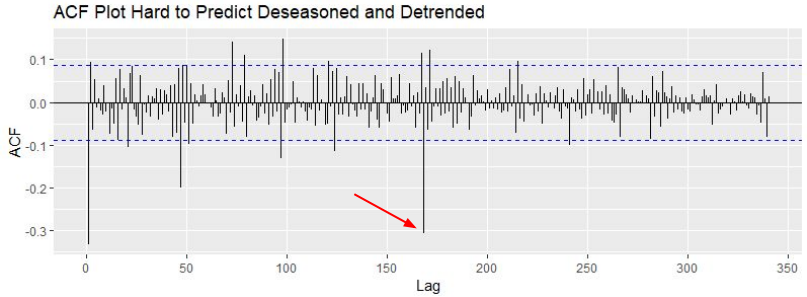


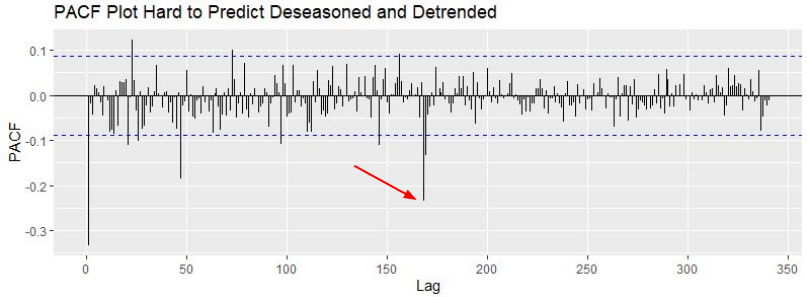
Data should be detrended and deseasoned.

Identifying Possible ARIMA Models









Weekly seasonality with correlation around recent hours.

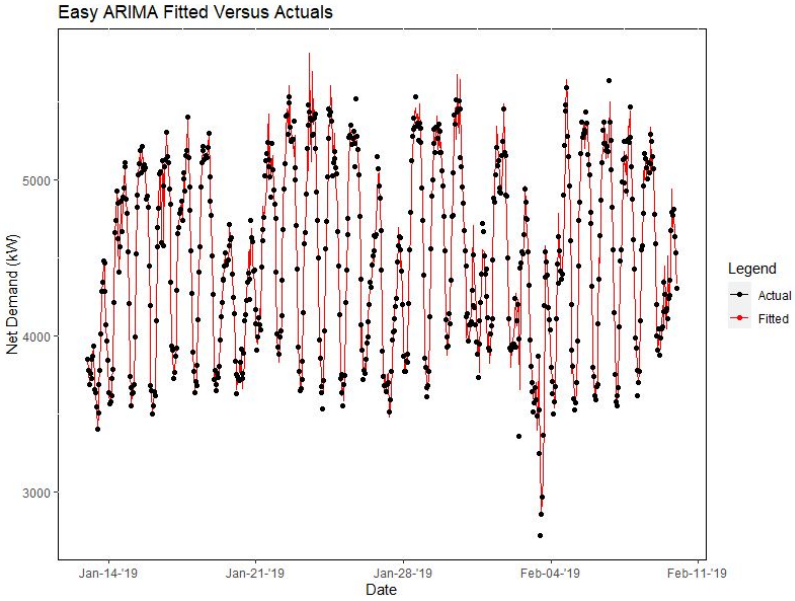
Testing Possible ARIMA Models

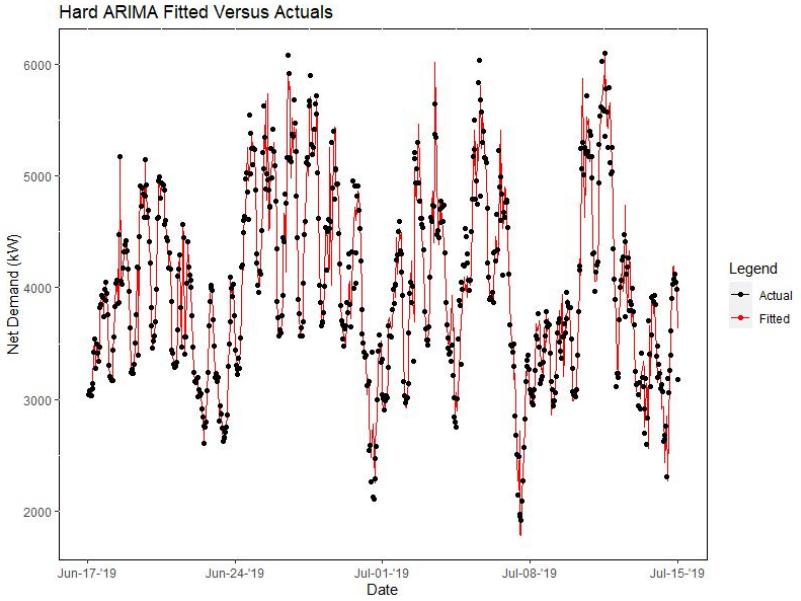
auto.arima function

Candidate model

Model	Easy AIC	Hard AIC
$(0,1,5),(0,1,0)_{168}$	6705.847	7437.471
$(1,1,0),(0,1,0)_{168}$	6729.023	7431.000
$(0,1,1),(1,1,1)_{168}$	6595.073	7306.360
$(0,1,0),(1,1,0)_{168}$	6641.087	7412.797
$(0,1,0),(0,1,1)_{168}$	6594.811	7380.169
$(1,1,0),(0,1,1)_{168}$	**6593.901**	**7306.021**
$(0,1,3),(1,1,0)_{168}$	6641.921	7343.545

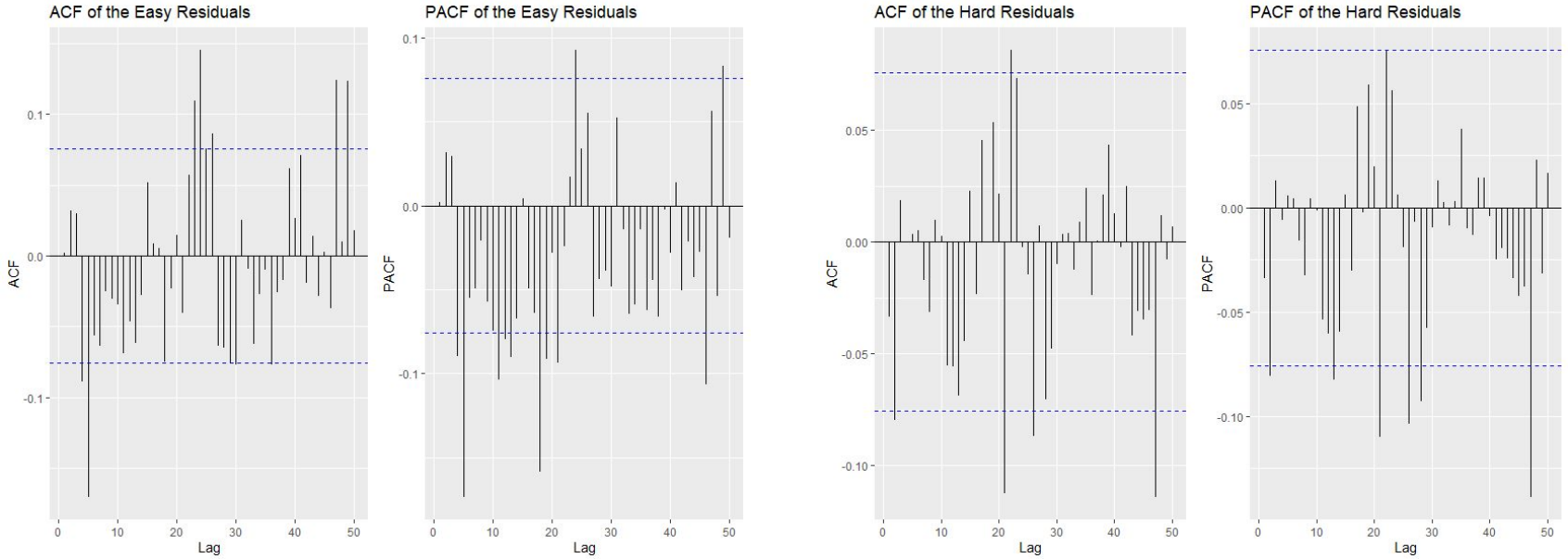
Fitting ARIMA Models to Training Data





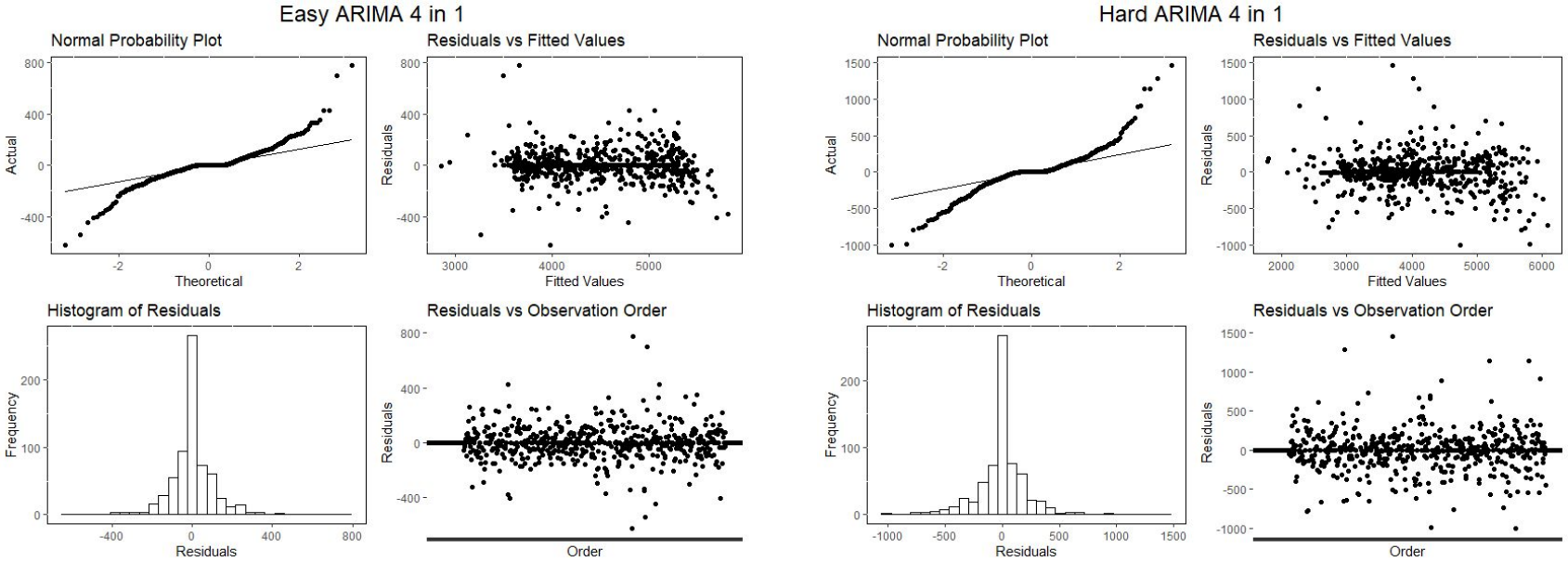
Models do not perfectly fit local max/min.

Residuals of the Best Fit Model



Residuals have some correlation.

Residuals of the Best Fit Model



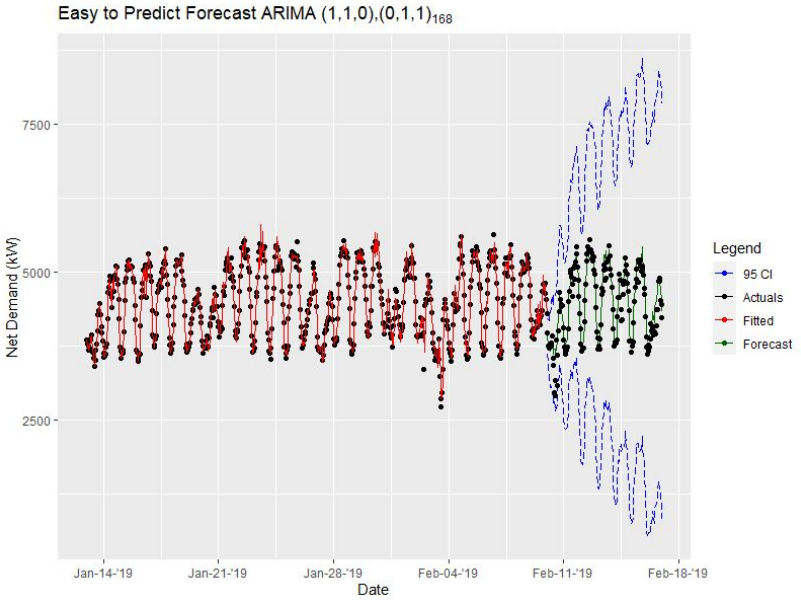
Normality, Equal Variance, Autocorrelation, and Linearity.

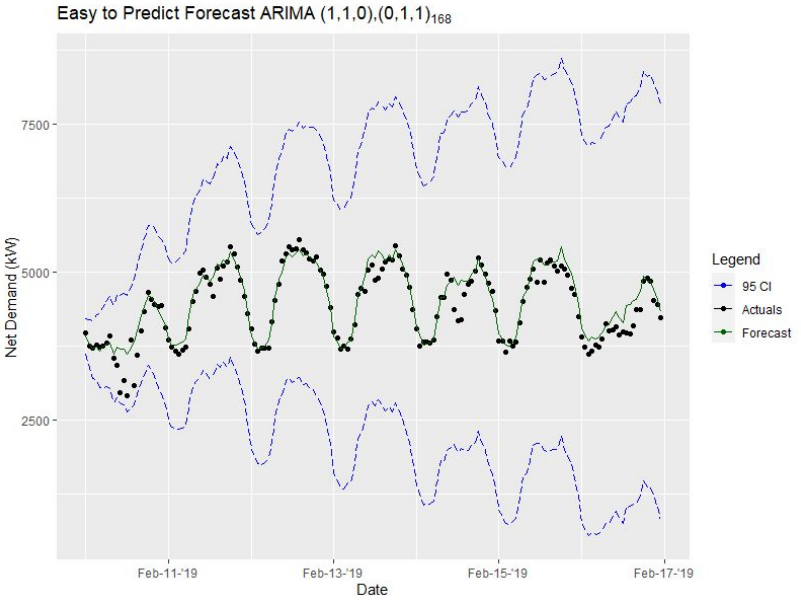
Picking the Best ARIMA Model

Model	Easy AIC	Easy MAPE	Hard AIC	Hard MAPE
$(0,1,5),(0,1,0)_{168}$	6705.847	4.080	7437.471	22.685
$(1,1,0),(0,1,0)_{168}$	6729.023	4.152	7431.000	22.876
$(0,1,1),(1,1,1)_{168}$	6595.073	3.333	7306.360	21.671
$(0,1,0),(1,1,0)_{168}$	6641.087	3.363	7412.797	23.472
$(0,1,0),(0,1,1)_{168}$	6594.811	3.253	7380.169	24.321
$(1,1,0),(0,1,1)_{168}$	6593.901	3.256	7306.021	22.374
$(0,1,3),(1,1,0)_{168}$	6641.921	3.324	7343.545	20.613

Chosen model

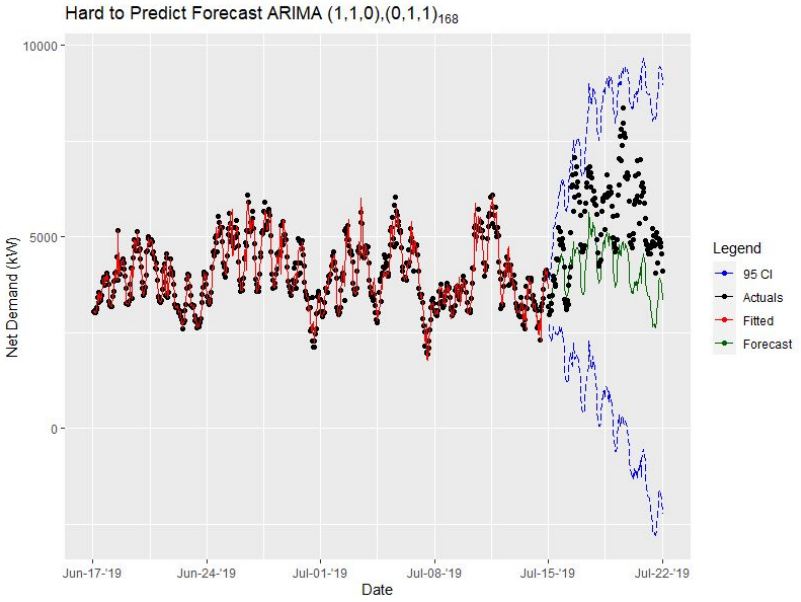
Plots of the Best Easy Model Forecast

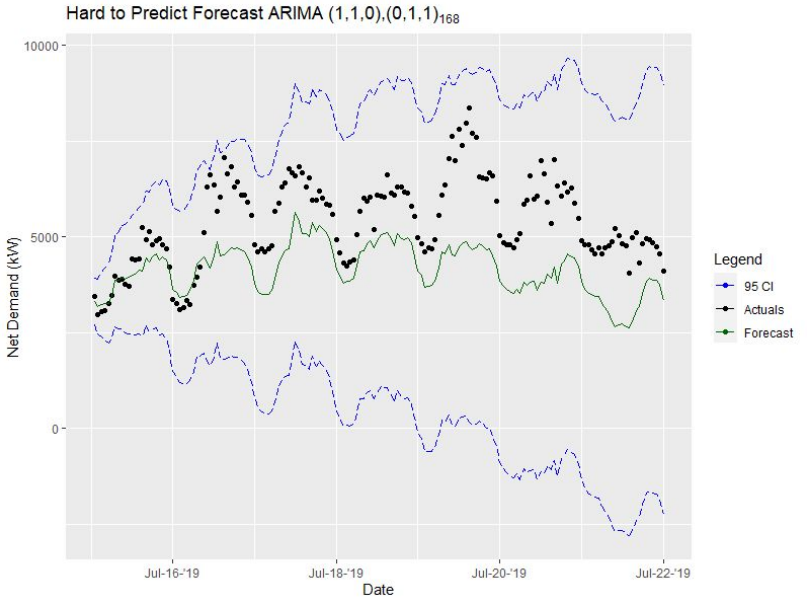




MAPE = 3.256

Plots of the Best Hard Model Forecast



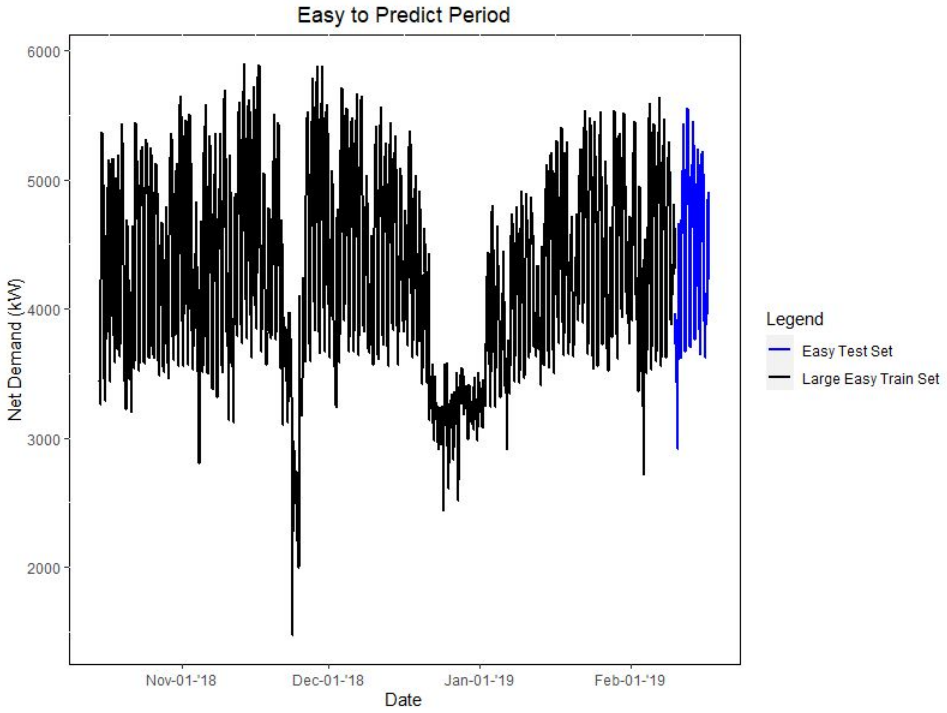


MAPE = 22.374

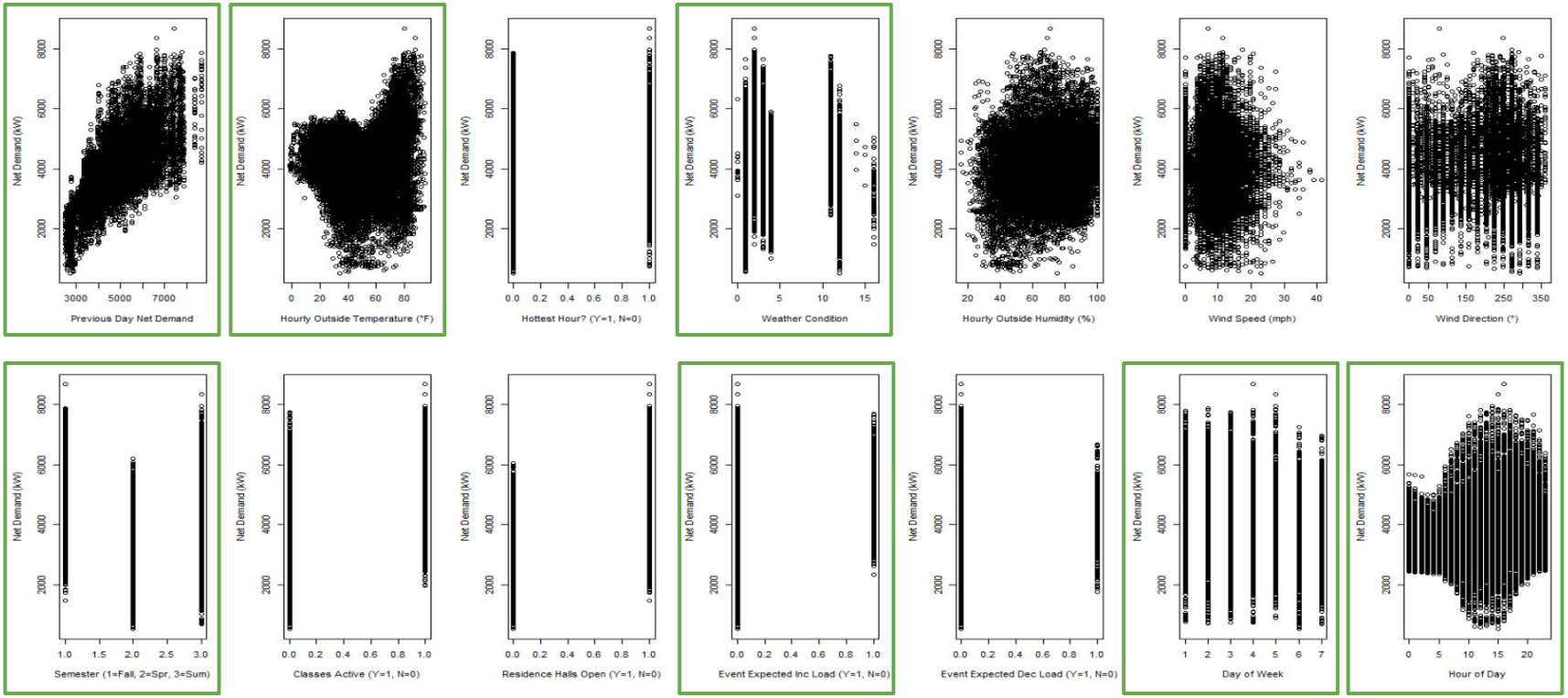
Dynamic Regression Models

Lengthening the Easy Training Set

- Period was lengthened to capture a wider range of regressor values
- New easy training period of 3 months
  - 6x larger than the original training set

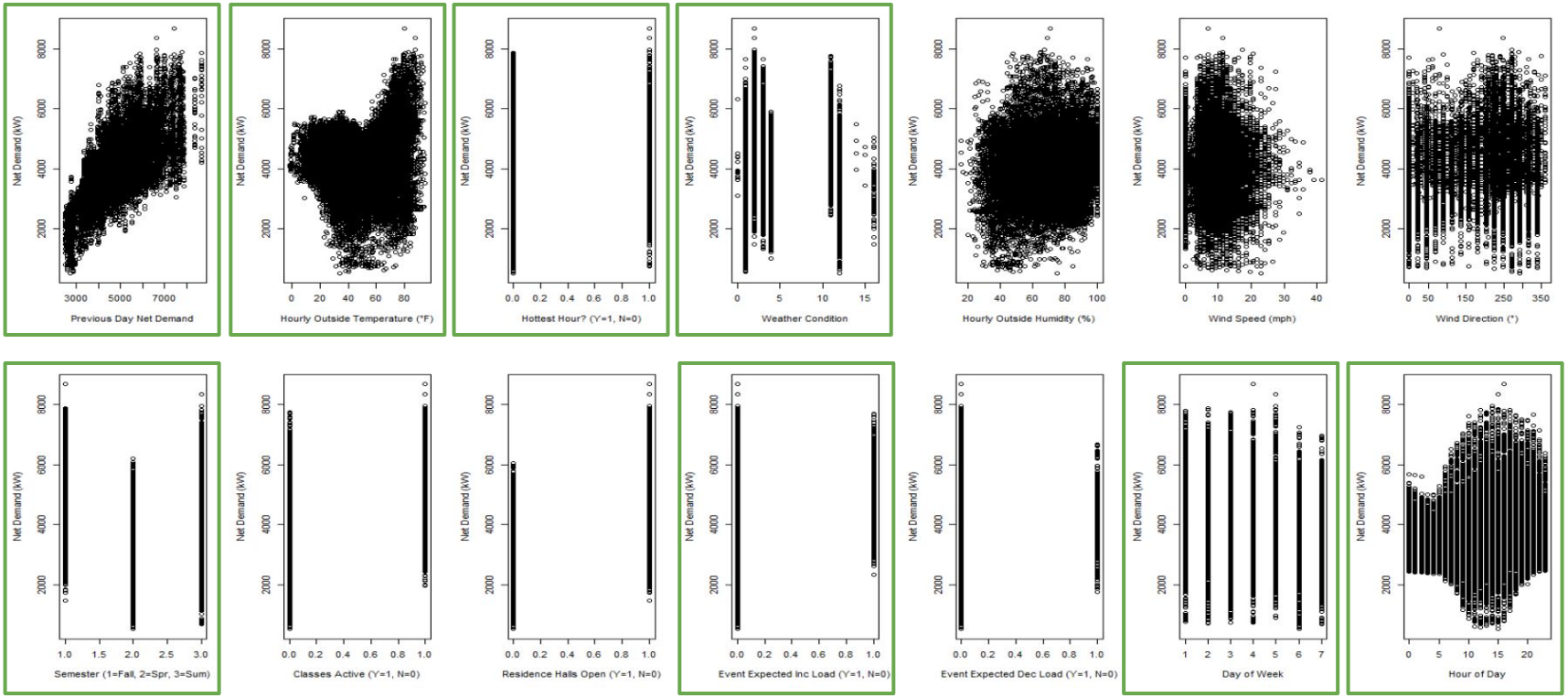


Identifying Potential Regressors



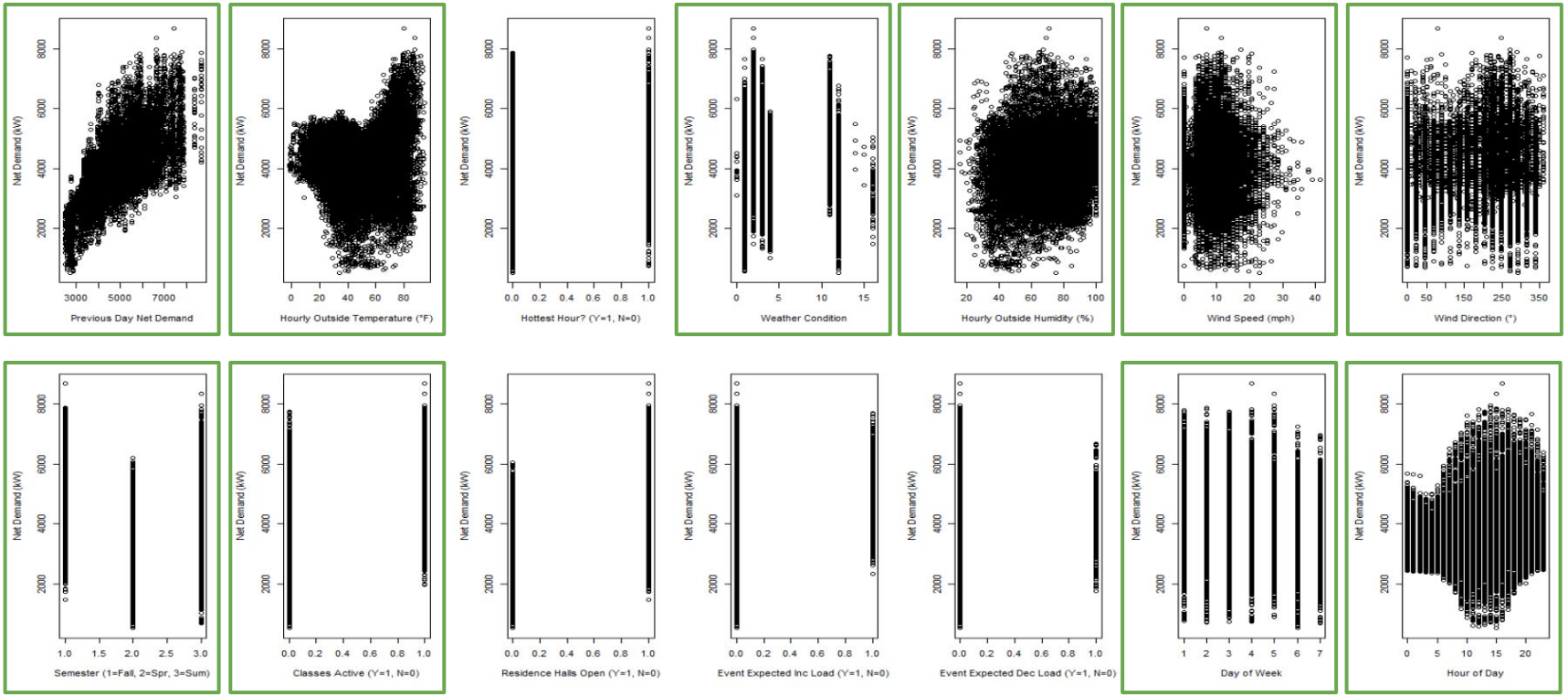
AIC = 37349.204

Identifying Potential Regressors



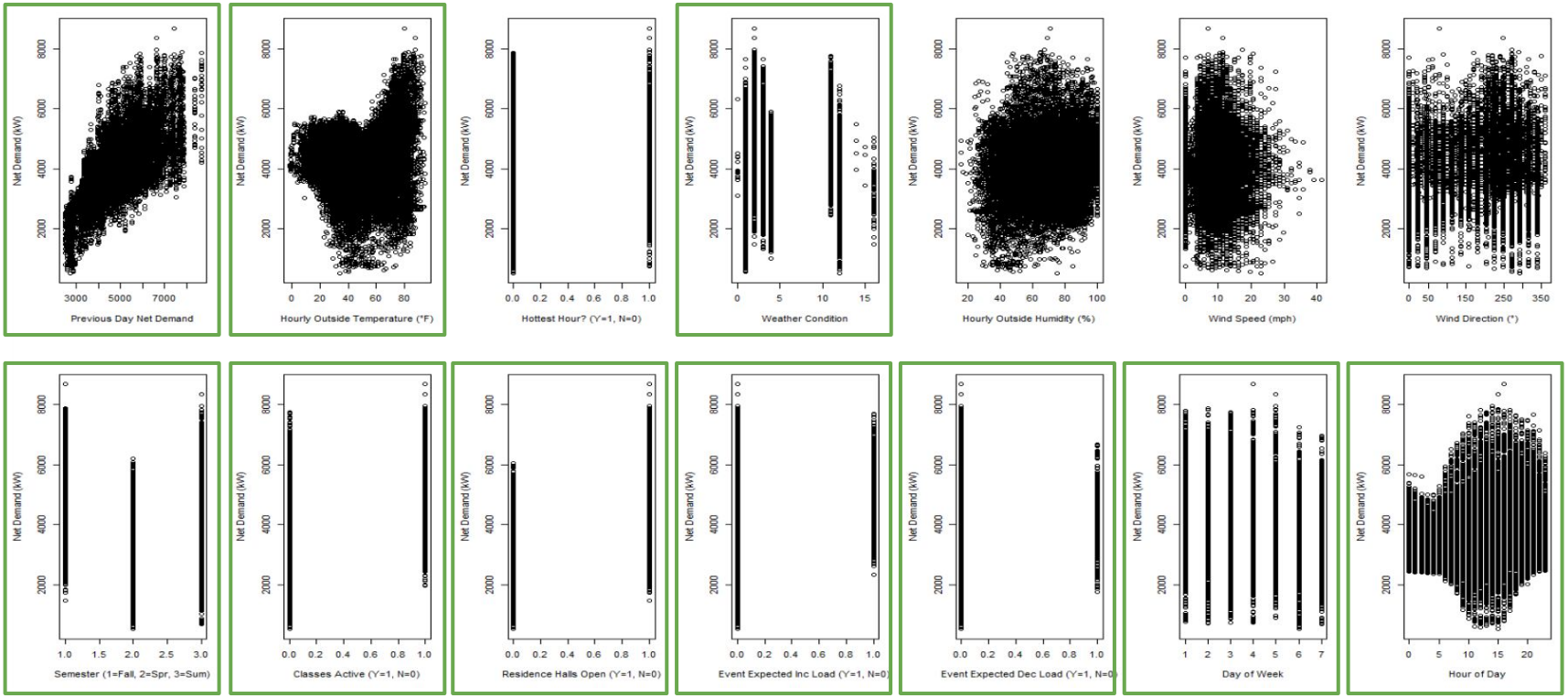
AIC = 37348.770

Identifying Potential Regressors



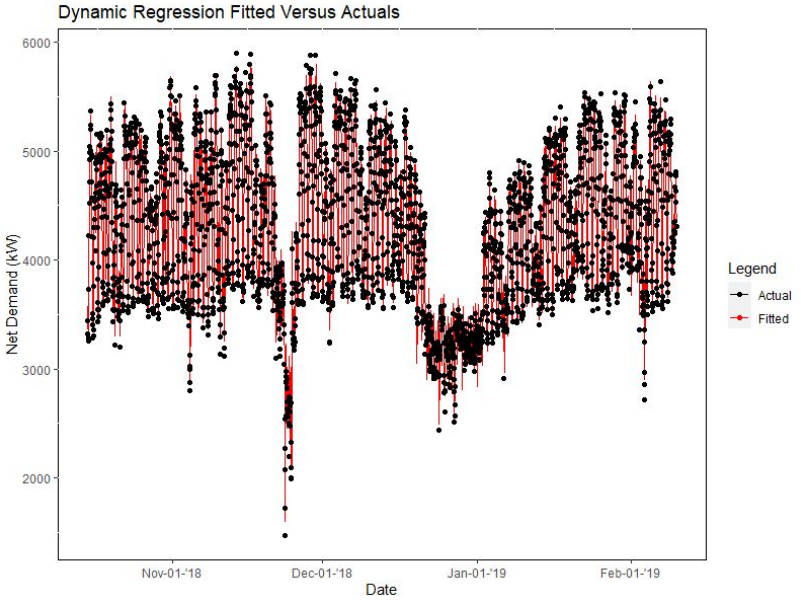
AIC = 37343.316

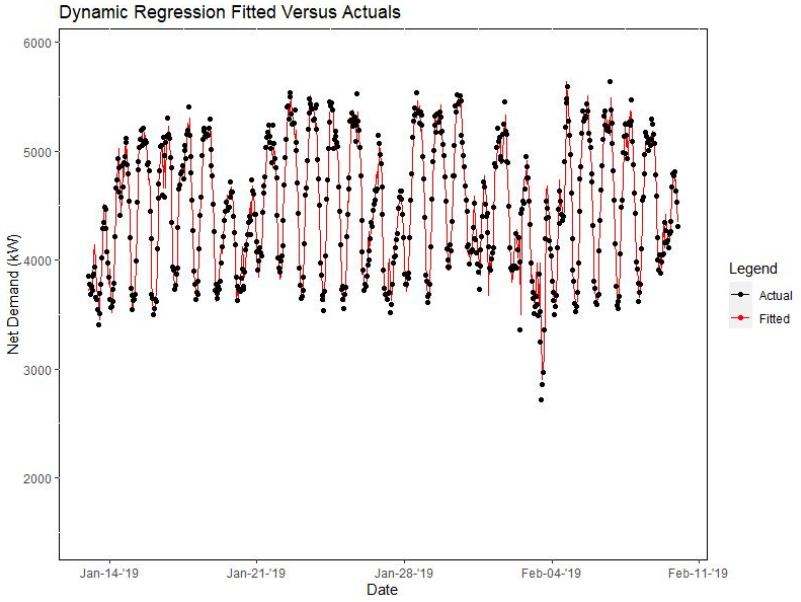
Identifying Potential Regressors



AIC = 37312.308

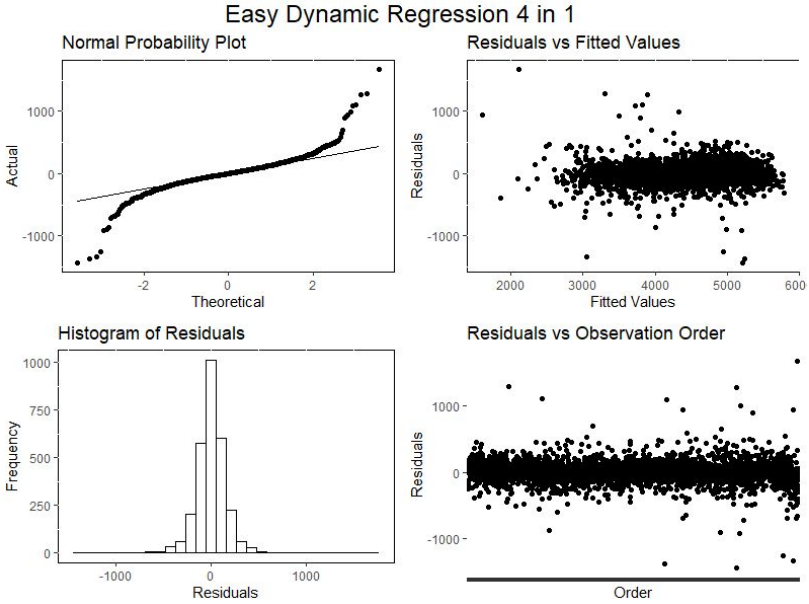
Fit of Dynamic Regression Model

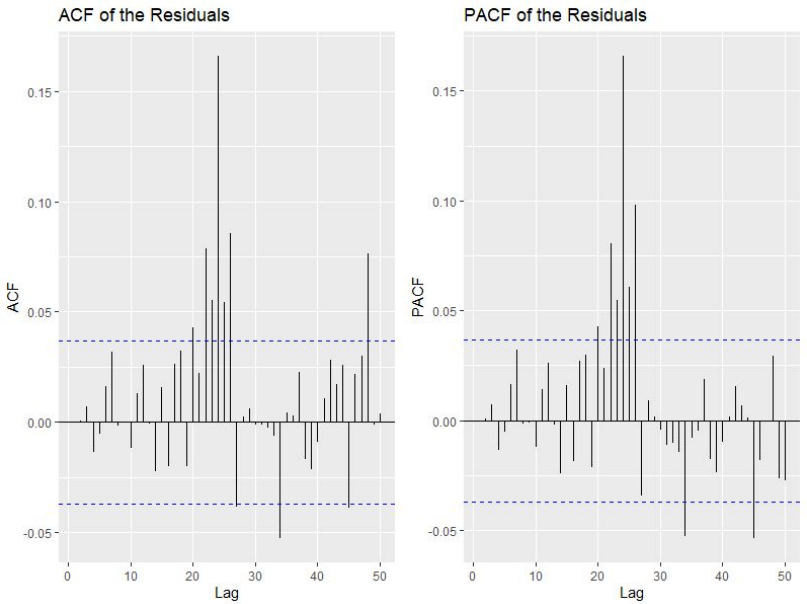




Model appears to fit weekend days well.

Dynamic Regression Residuals

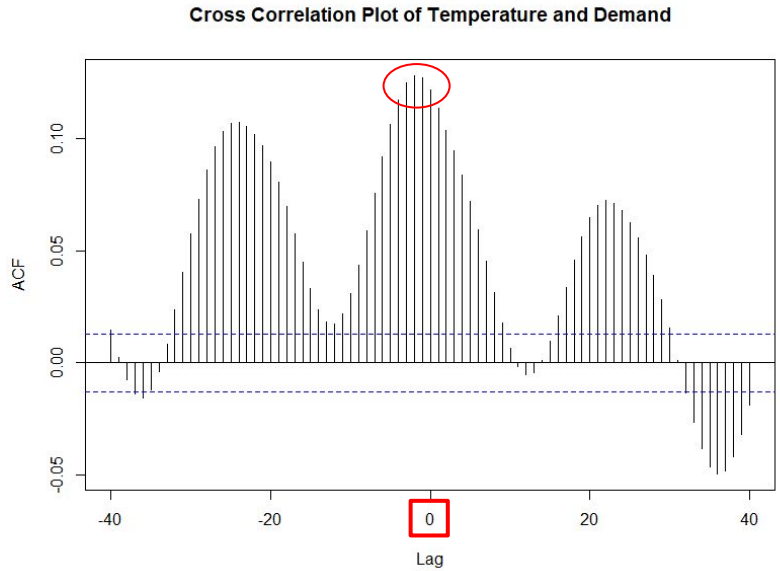


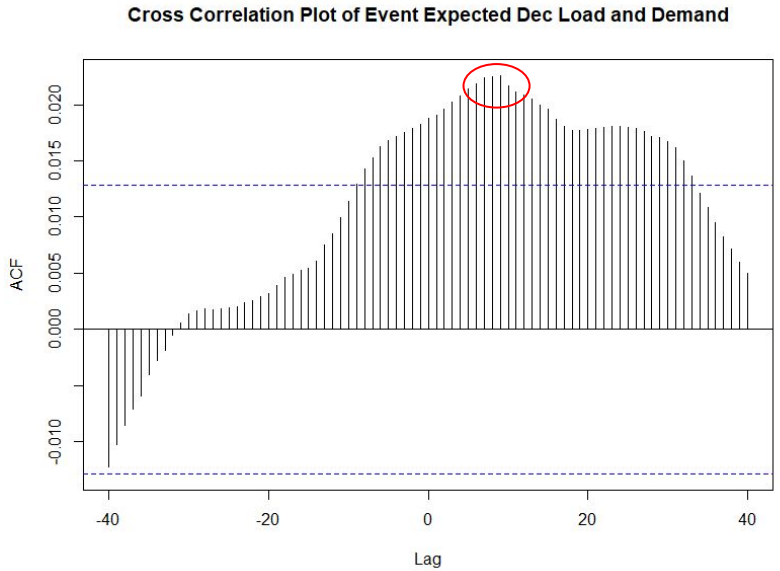


Residuals pass assumptions.

Dynamic Regression Models with Lagged Predictors

Identifying Potential Lagged Regressors

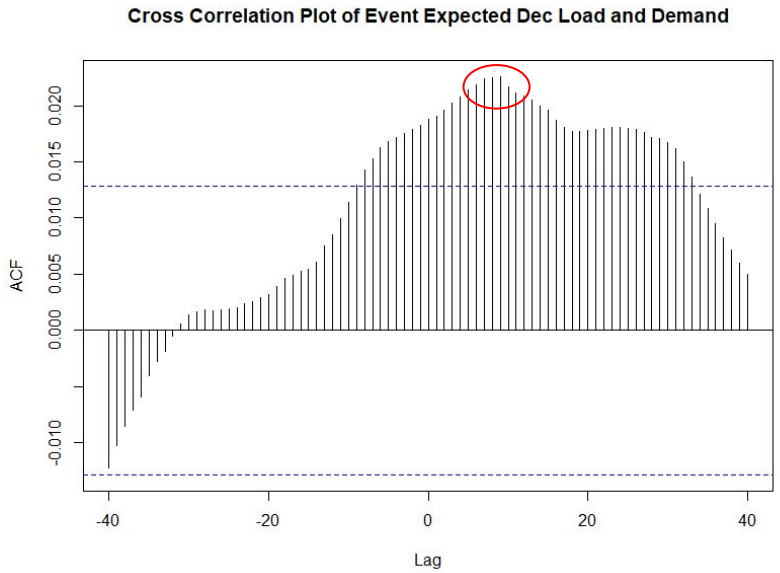




Multiple regressors were analyzed

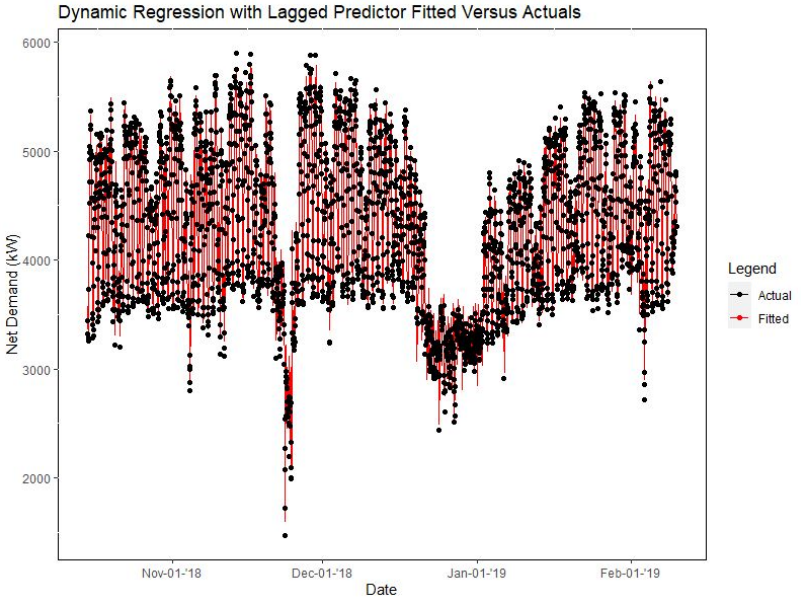
Different Lags Tested

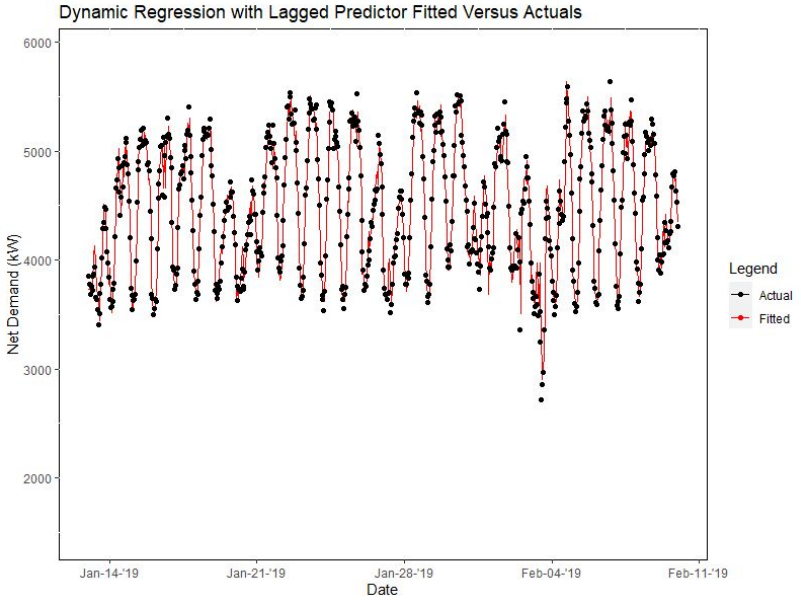
Lag	AIC
7	37310.182
8	37312.313
9	37304.738
10	37304.441
11	37311.593



Lag 10 results in the lowest AIC.

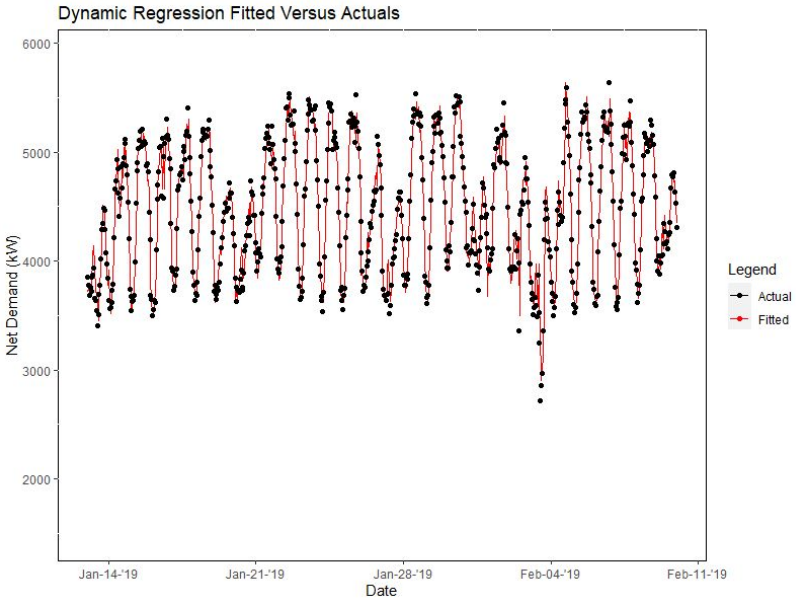
Fit Lagged Dynamic Regression Model



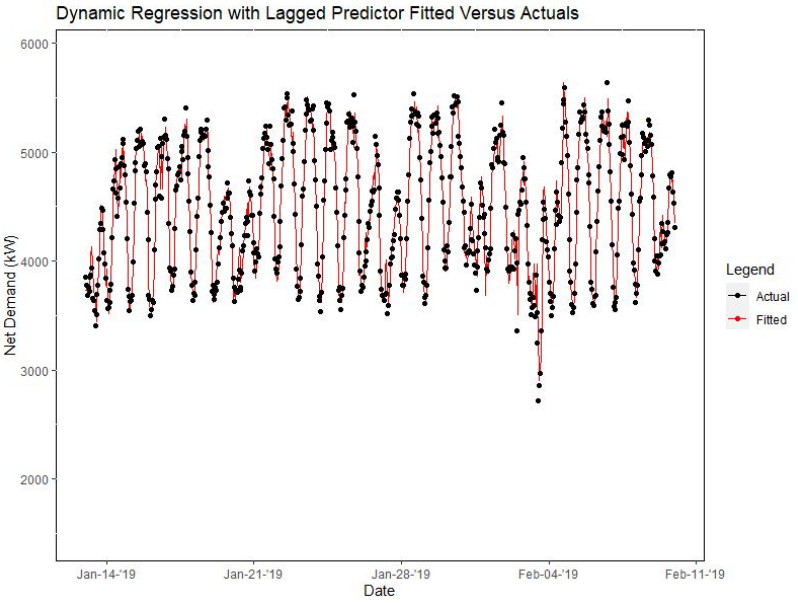


This model includes the lagged regressor

Original vs. Lagged Model



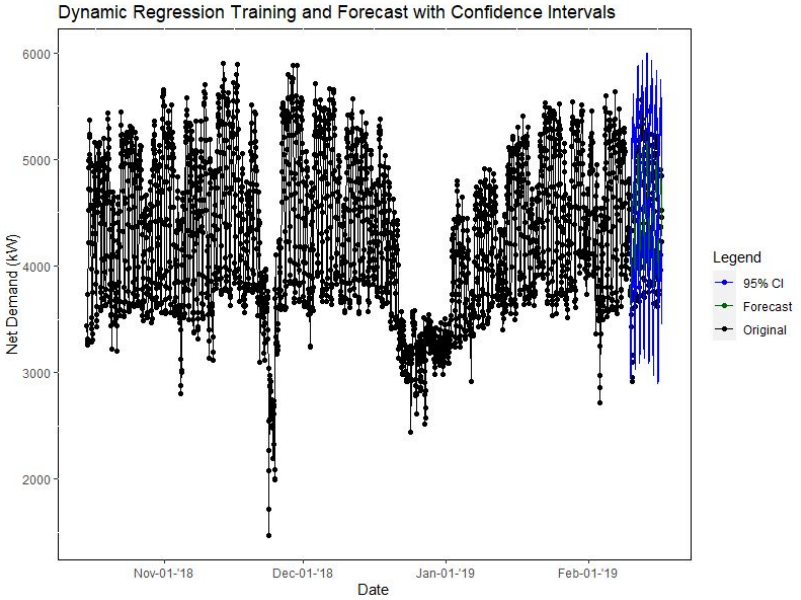
AIC = 37312.308

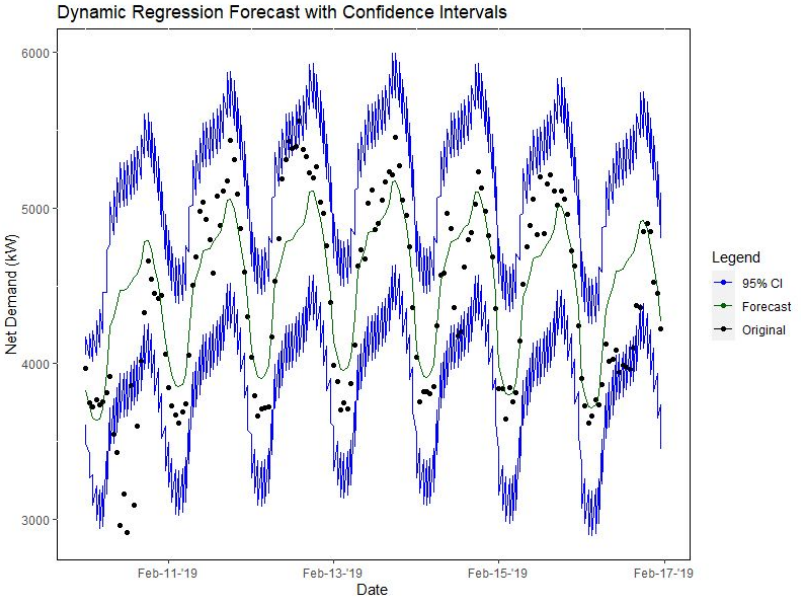


AIC = 37304.441

Forecasting the Dynamic Regression Model with Lagged Predictors

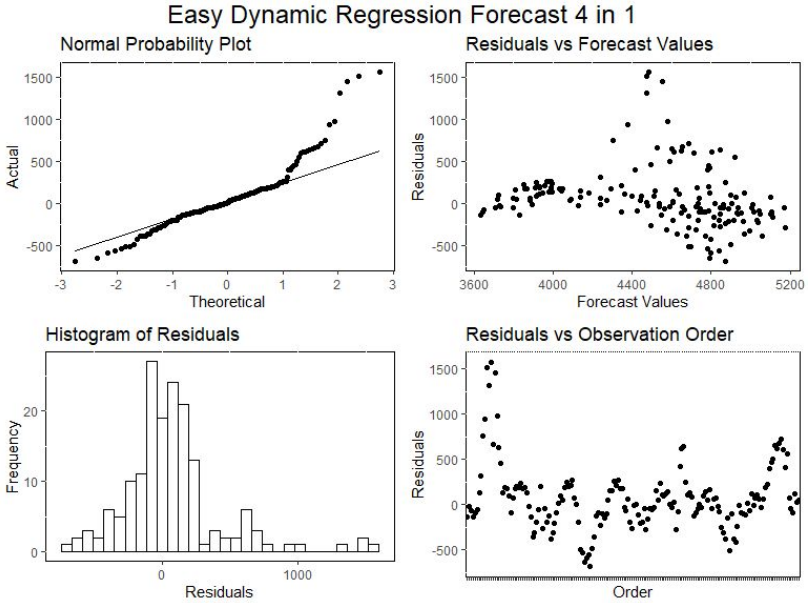
Easy Week Forecast

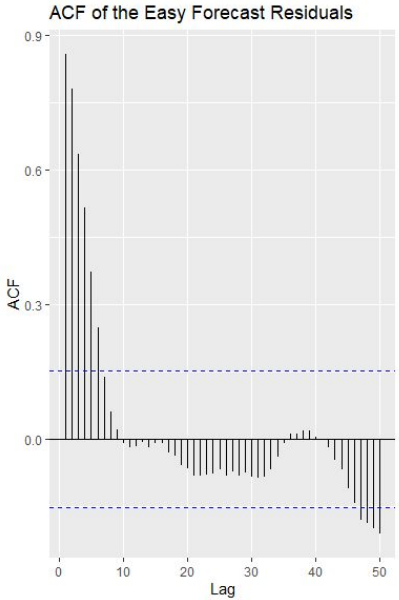


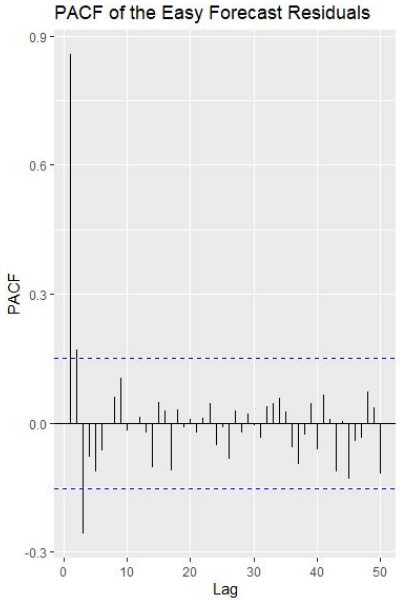


MAPE = 5.917

Easy Week Forecast Residuals

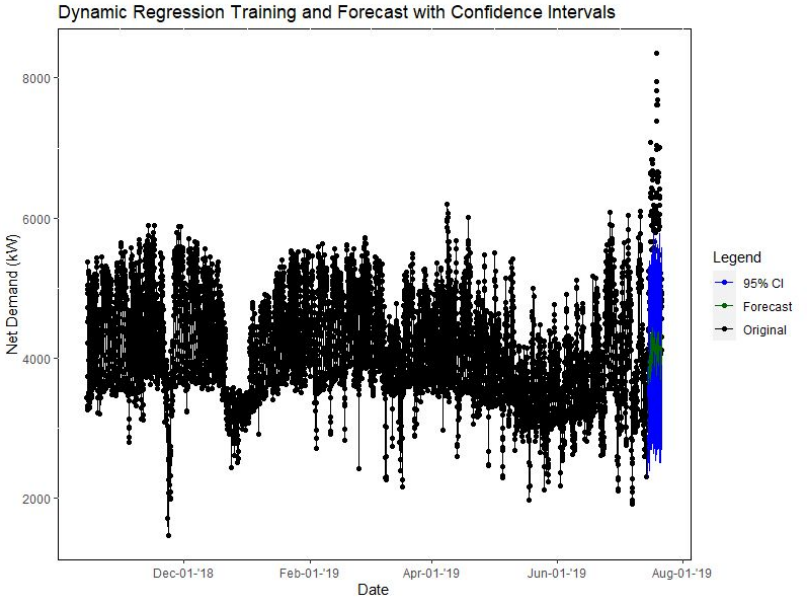


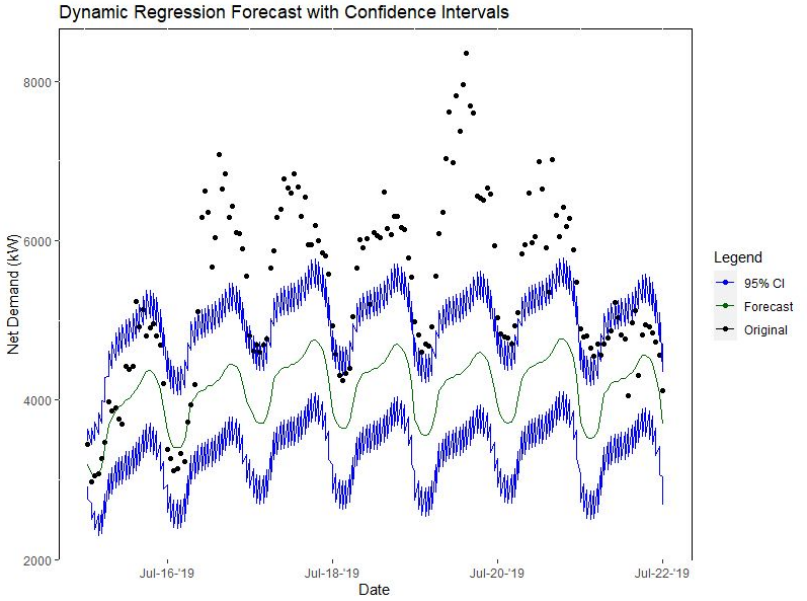




Residuals do not pass assumptions.

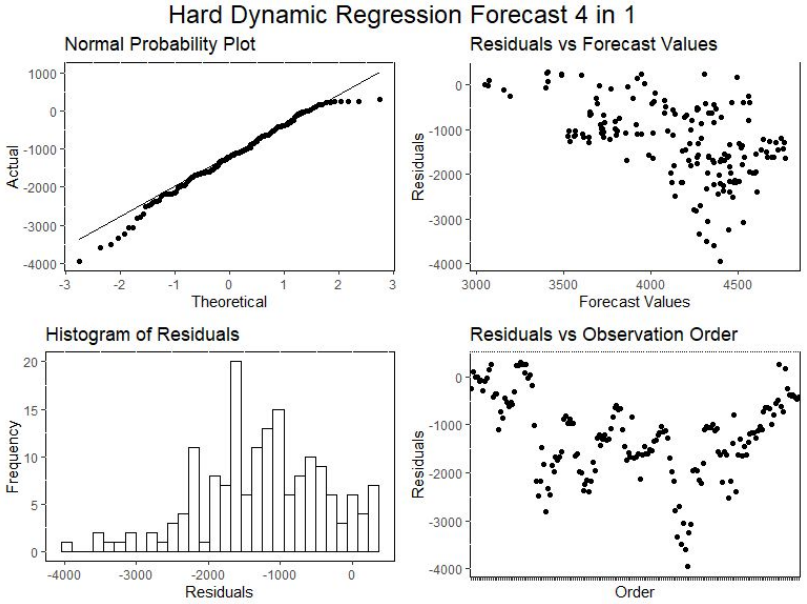
Forecasting Hard Week

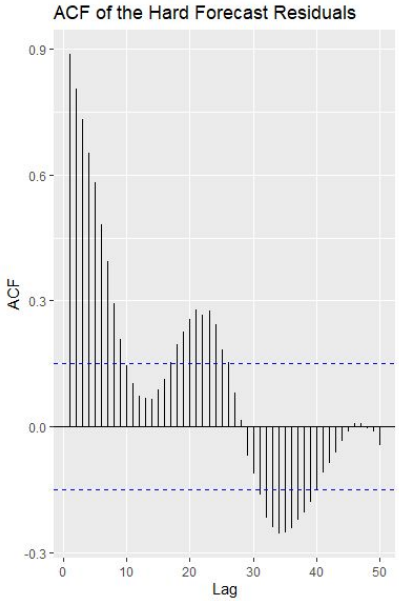


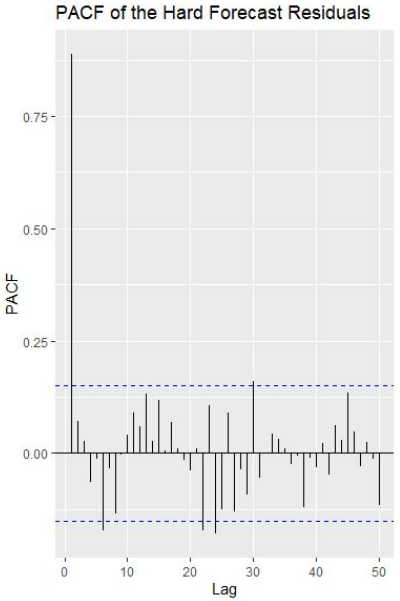


MAPE = 21.643

Forecasting Hard Week Residuals

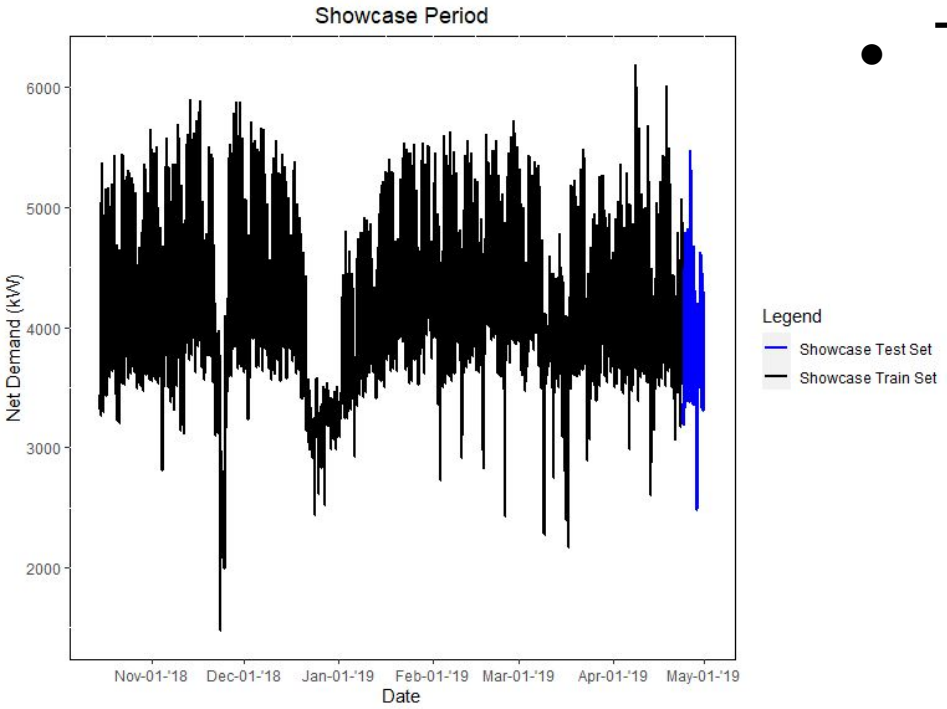






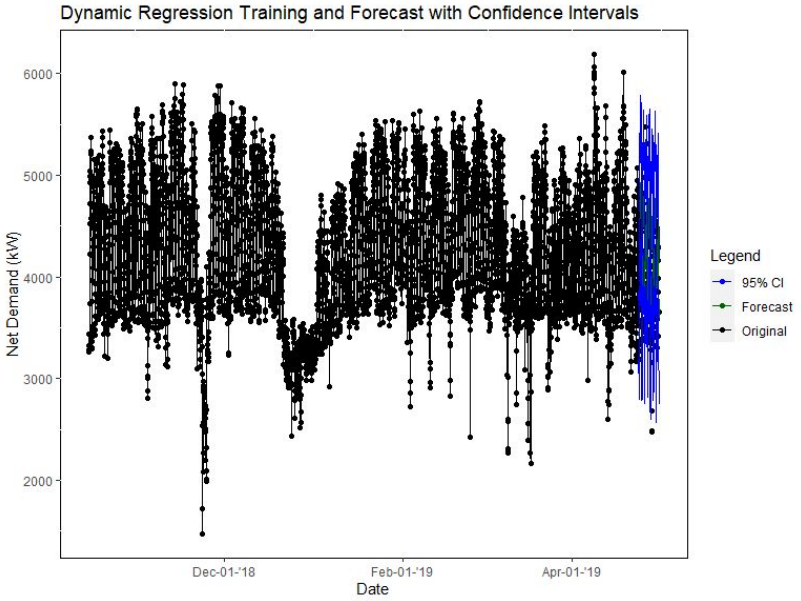
Residuals do not pass assumptions.

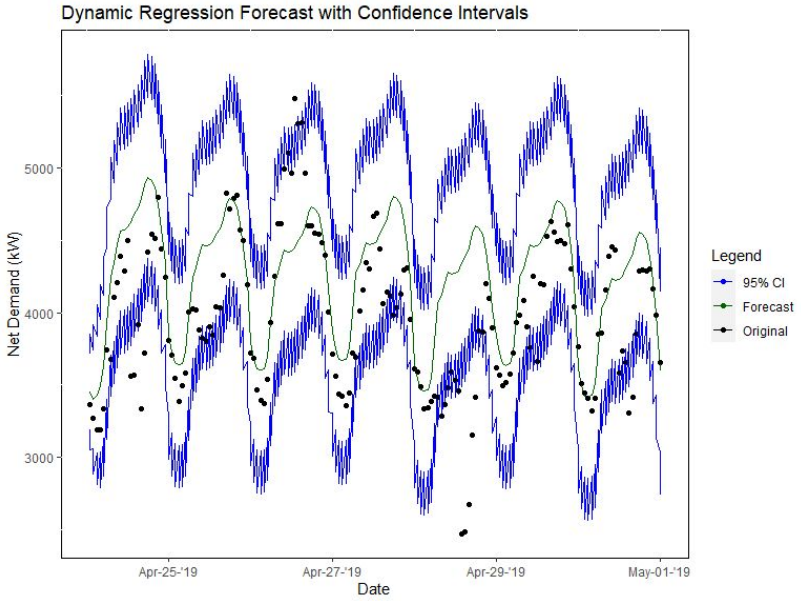
Identify a Showcase Period



- Test set 4/24/19-4/30/19
  - Imagine RIT April 27th 2019
  - Last Day of Classes April 29th

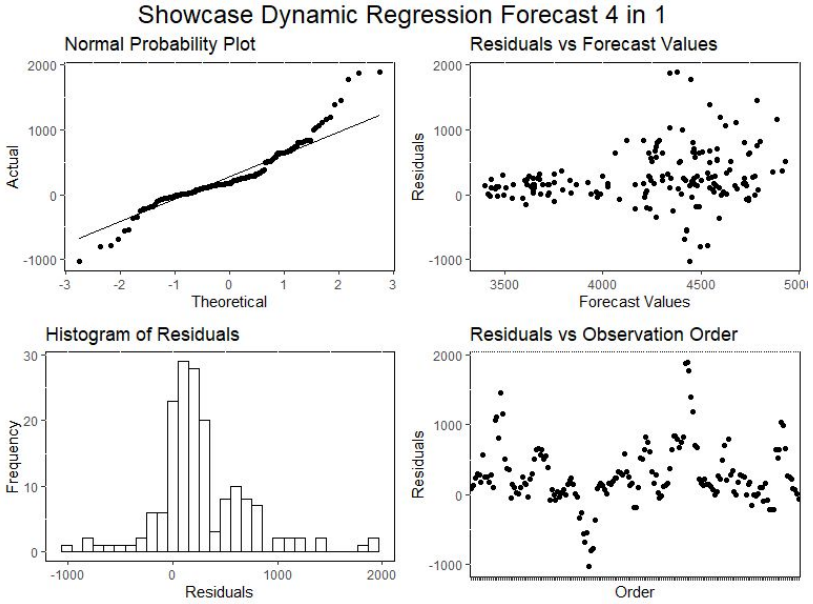
Showcase Week Forecast

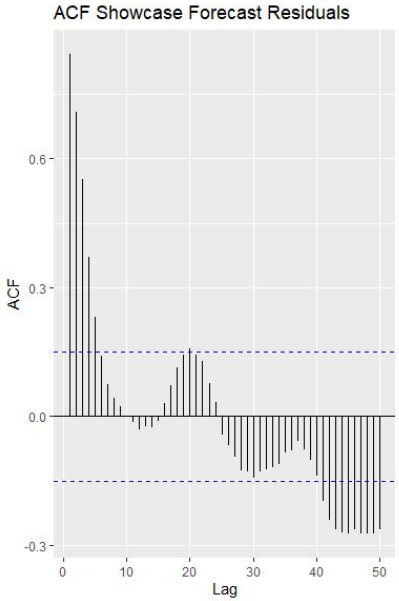


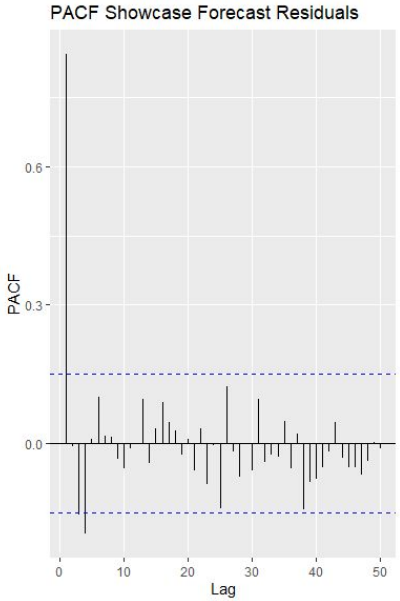


MAPE = 9.691

Showcase Week Forecast Residuals



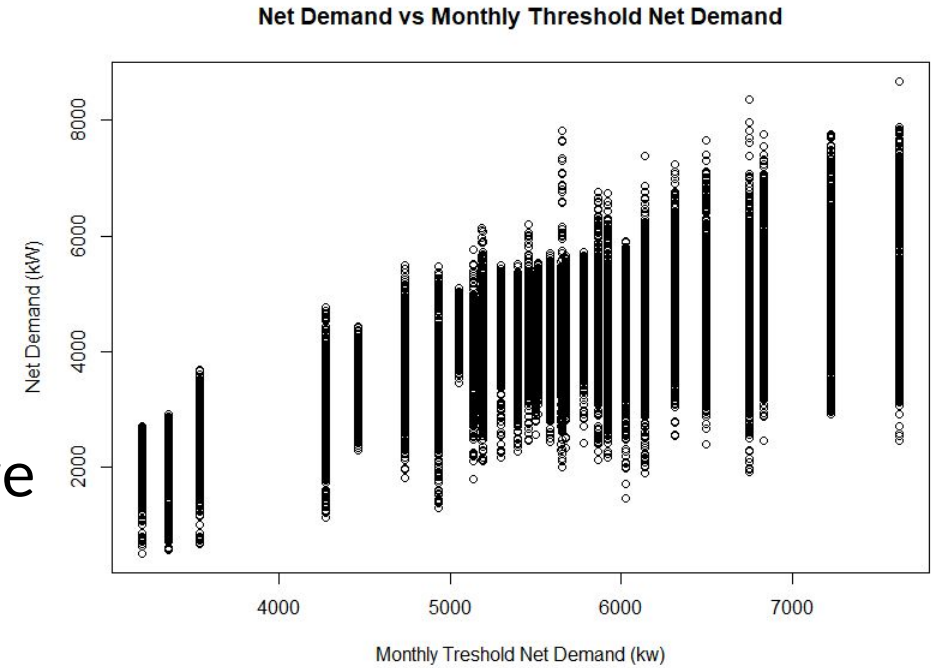




Residuals do not pass assumptions.

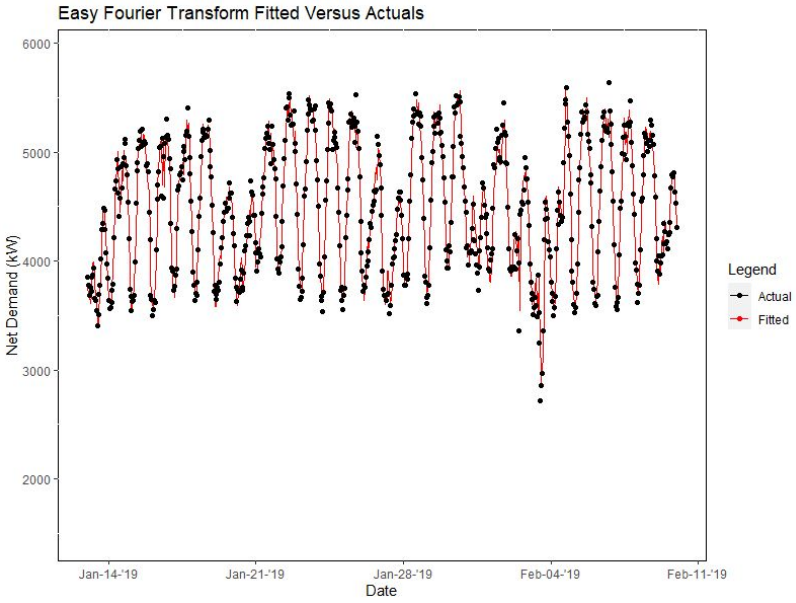
Other Possible Regressors

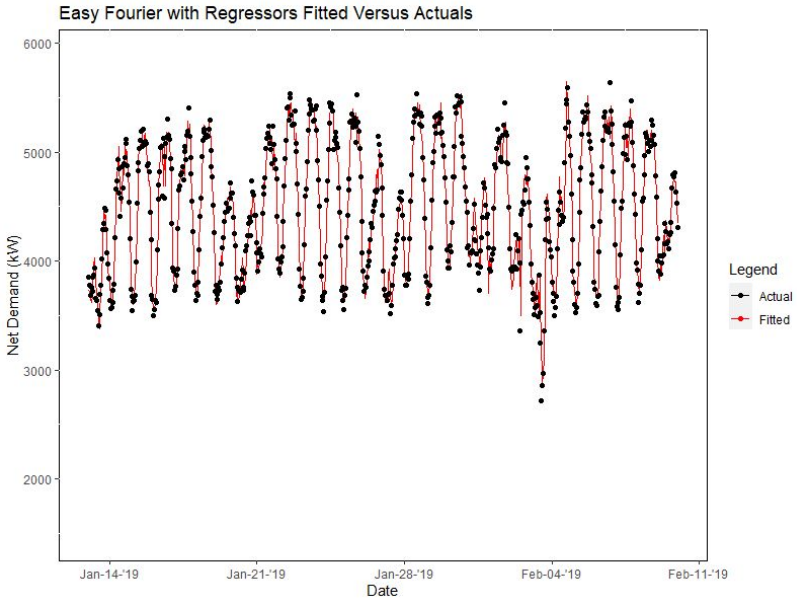
- Monthly threshold net demand should be investigated
- Not enough repeat special events
  - Two graduation days
- Thermostat set points were not helpful



Fourier Transforms with and without Regressors

Easy Fourier Models

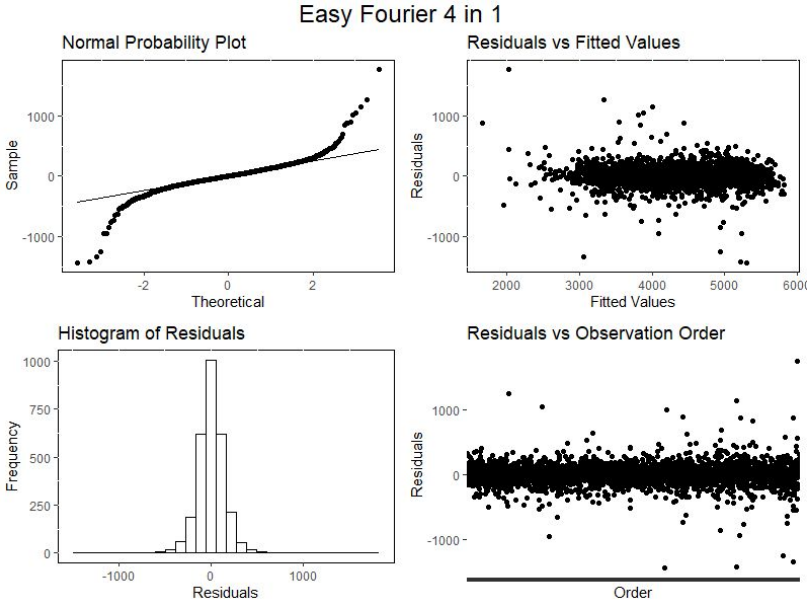


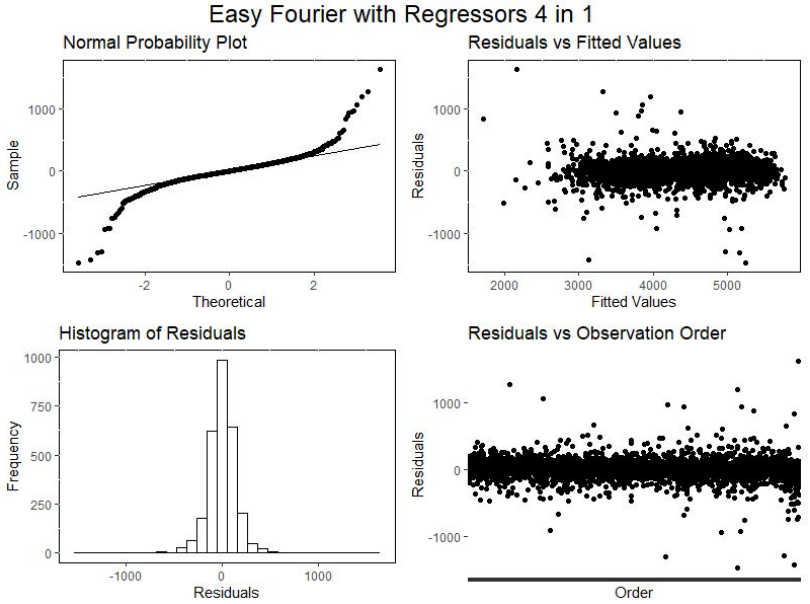


AIC = 37133.127

AIC = 37165.543

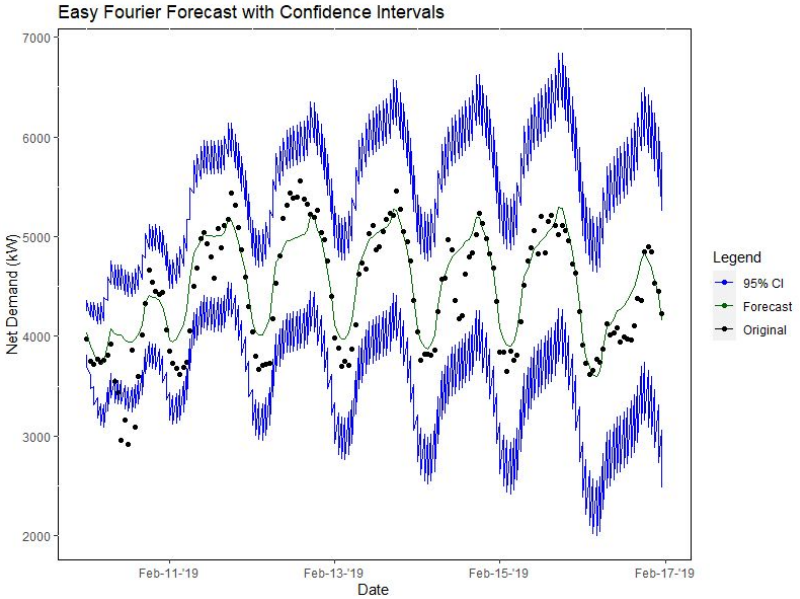
Easy Fourier Model Residuals

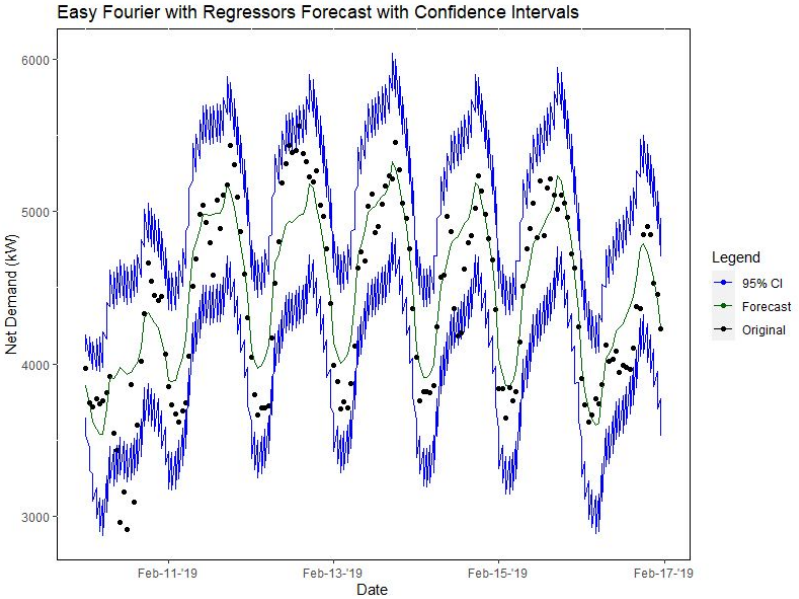




Both residuals pass assumptions.

Easy Fourier Model Forecasts

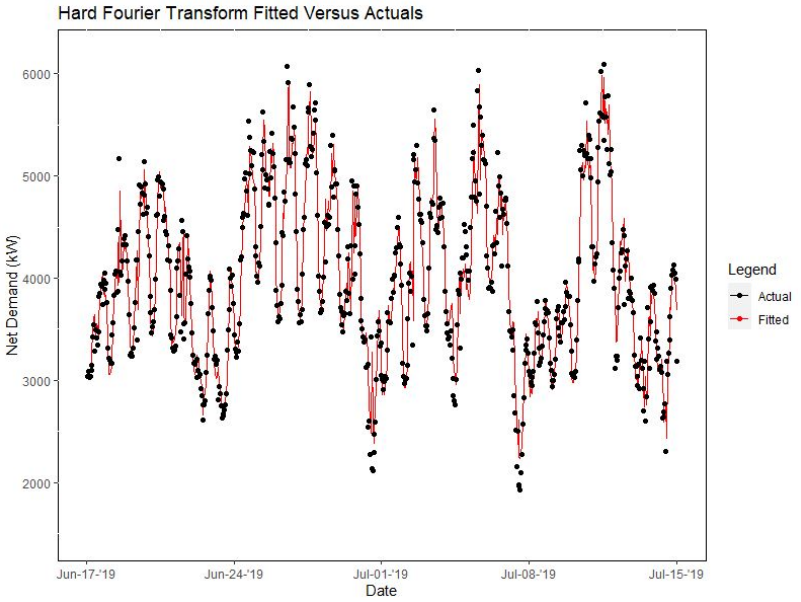




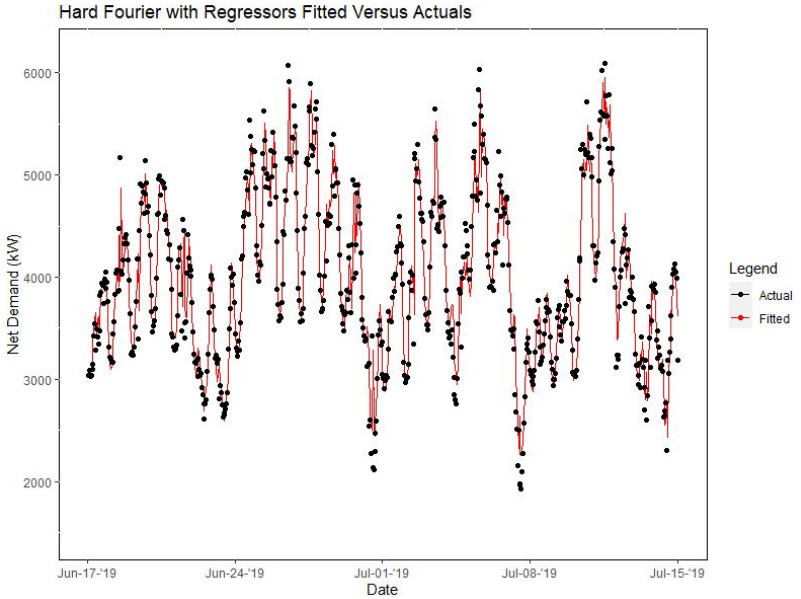
MAPE = 4.713

MAPE = 4.616

Hard Fourier Models

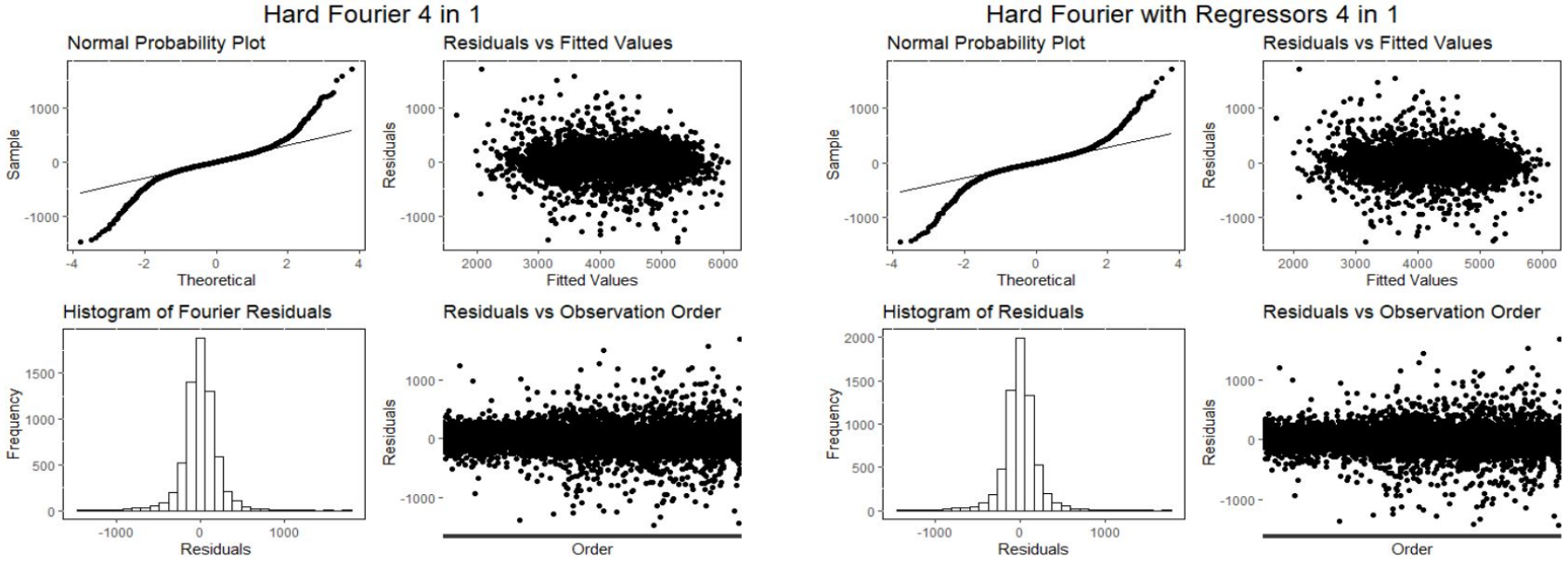


AIC = 89364.208



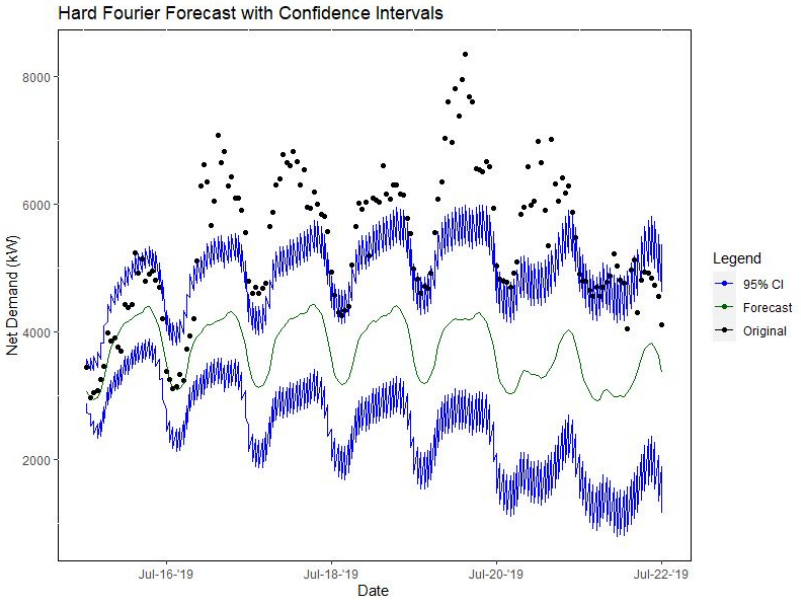
AIC = 89094.111

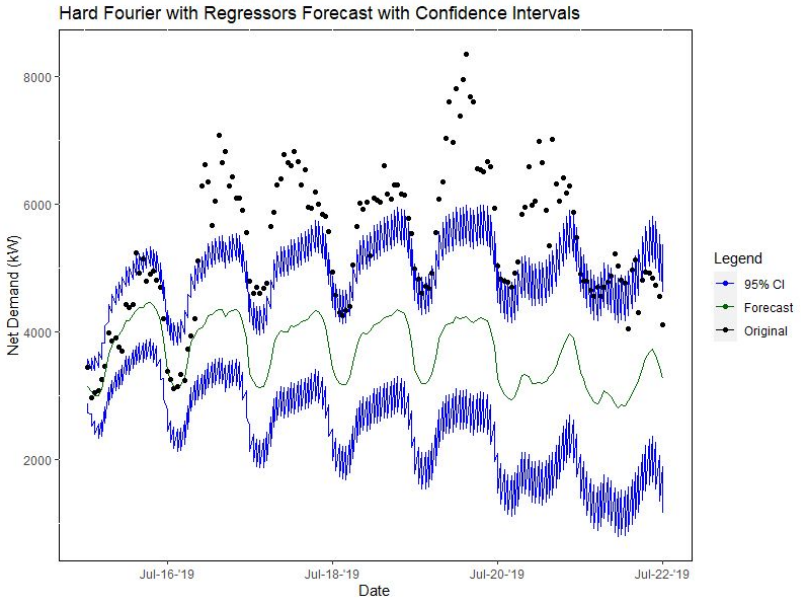
Hard Fourier Model Residuals



Both residuals pass assumptions.

Hard Fourier Model Forecasts





MAPE = 29.097

MAPE = 29.419

Final Table Comparing All Models

Model	Easy AIC	Easy MAPE	Hard AIC	Hard MAPE
Holt Winters	N/A	4.701	N/A	16.616
Seasonal ARIMA $(1,1,0),(0,1,1)_{168}$	6593.901	3.256	7306.021	22.374
Dynamic Regression	37312.308	5.906	N/A	N/A
Dynamic Regression with Lagged Predictors	37304.441	5.917	89637.764	21.643
Fourier Model Easy → i=9, j=6 Hard → i=10, j=10	37133.127	4.713	89364.208	29.097
Fourier Model with Regressors Easy → i=9, j=6 Hard → i=10, j=10	37165.543	4.616	89094.111	29.419


auto.arima (seasonal = TRUE)
Wait, the best model is seasonal ARIMA?
Always has been.

Thank You