TEAM PROJECT

Part 3

Team 3
Yang Yu, Chaoyang Zheng, Meng Peng & Katherine Gimon-Valencia
April 19/2019

Introduction

This report is the last in a three-part series which systematically explores increasingly complex forecasting methods in order to accurately predict daily electricity demand for the following day for the Rockland Electric Company (RECO). This report begins by addressing some changes from the previous report including the calculation of smoothing methods on the entire validation set, validation of the beta coefficients of the explanatory variables and a look at why the percentage bias for multiple regression was so large. Next the report will explore calculations of ARIMA and SARIMA models, compare the forecasting methods from all 3 reports, followed by our final forecasting recommendation and conclusion for RECO.

Previously Double-Seasonal Holt Winters and TBATS had only been run on portions of the validation set. Also, an attempt was made to run state space models with frequency 356, causing R to crash. To correct both issues, the frequency was changed to 7 for the state space model (remained an MNM) and the other two methods were run on the entire validation set; the results of which can be seen in Table 1. TBATS is the best of the three methods with an MAPE of 5.83% outperforming the performance benchmark (the naïve no-change method) of 6.84%.

Table 1: Performance of Smoothing Methods in the Validation Dataset

Method	Bias	% Bias	MAPE (%)
State Space (MNM)	18.2	0.887	6.93
DSHW	6.13	0.643	6.18
TBATS	1.12	0.588	5.83

Secondly, in multiple regression, humidity and wind chill had the wrong signs; this has been corrected in Figure 1. Furthermore, it was noted that the beta coefficients for HDD & CDD were small. However, after validation, this appears to make sense considering the average MW/day between 2011 and 2018 was 4138, which is not a particularly large figure (compared to other PJM zones) and that consumers in this region supplement their electricity use with gas[1]. Lastly, the model (with arma errors) was re-run following the identification of a large percentage bias, suspected to be the result of two large residuals. This was confirmed to be the case. One was linked to Memorial Day which, despite its identification as a Holiday, did not show a decrease in demand (in numerous years). Therefore, it was reclassified as a non-Holiday. The second was linked to a heat wave which took place in July 2013 when "the 14-20th marked the most uncomfortable July interval [thanks to] scorching daytime maximums, high humidity levels and high night-time minimums"[2] which correspond to the days with high residuals. Because it was a real weather event which the model simply had a hard time capturing, the data was left as is. Table 2 shows the improved regression model with Memorial Day marked as a regular day and the correct humidity and wind chill signs.

```
> print(summary(mreg))

Call:
lm(formula = Yt ~ HDDt + CDDt + lag(HDDt, 1) + lag(HDDt, 2) +
    lag(CDDt, 1) + lag(CDDt, 2) + Rhumt + WINDCHILLt + factor(WEEKENDt) +
    factor(HOLIDAYt), x = T, y = T)

Residuals:
    Min      1Q  Median      3Q     Max
-1060.5  -120.6    -3.4   116.5  2333.6

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        3723.5996    10.6888  348.37  < 2e-16 ***
HDDt                 16.7498     2.5606    6.54  8.4e-11 ***
CDDt                 89.4250     3.2240   27.74  < 2e-16 ***
lag(HDDt, 1)          5.0797     1.9672    2.58  0.00992 **
lag(HDDt, 2)          5.7382     1.4812    3.87  0.00011 ***
lag(CDDt, 1)         30.4648     3.1598    9.64  < 2e-16 ***
lag(CDDt, 2)          5.8107     2.3097    2.52  0.01199 *
Rhumt                 0.1969     0.0368    5.35  1.0e-07 ***
WINDCHILLt            0.1509     0.6831    0.22  0.82518
factor(WEEKENDt)1  -548.6288    13.0819  -41.94  < 2e-16 ***
factor(HOLIDAYt)1  -230.1317    29.5316   -7.79  1.2e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 224 on 1446 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.922,     Adjusted R-squared:  0.921
F-statistic: 1.7e+03 on 10 and 1446 DF,  p-value: <2e-16
```

Figure 1

Table 2: Performance of Linear Regression in the Validation Dataset			
Method	**Bias**	**% Bias**	**MAPE (%)**
Linear Regression with ARMA Errors	94.6	2.92	4.84

[1] *New Jersey State Profile and Energy Estimates.* [US Energy Information Administration]. Retrieved April 18, 2019 from https://www.eia.gov/state/?sid=NJ

[2] Robinson, Dave. (4 August 2013). *Yet Another Hot Summer Month: July 2013 Summary*. [NJ Weather & Climate Network]. Retrieved April 4 2019 from https://www.njweather.org/content/yet-another-hot-summer-month-july-2013-summary

ARIMA/SARIMA Model

First, it was determined that because the data shows a dual seasonality, an ARIMA model would be inappropriate to provide reliable forecasts and instead SARIMA was examined. The seasonality was set to 7 to reflect the more immediate weekly cycle. It was also important to systematically test the parameters while simultaneously asking ourselves what they meant in terms of the data. We had a difficult time finding an appropriate model with normally-distributed residuals and a Ljung-Box test with large p-values. As such, it was suggested to cut the training set from 4 to 2 years (covering 2013-2014) and to begin by looking at perhaps D=1; the diagnostics of this model can be found in Figure 2. The ACF provides evidence of activity at larger seasonal lags, as well as local ones as there is decay at seasonal lags and pollution around these points. We decided to try to address this first before looking at stationarity as the model did not appear to be weakly stationary either. As such, we began testing

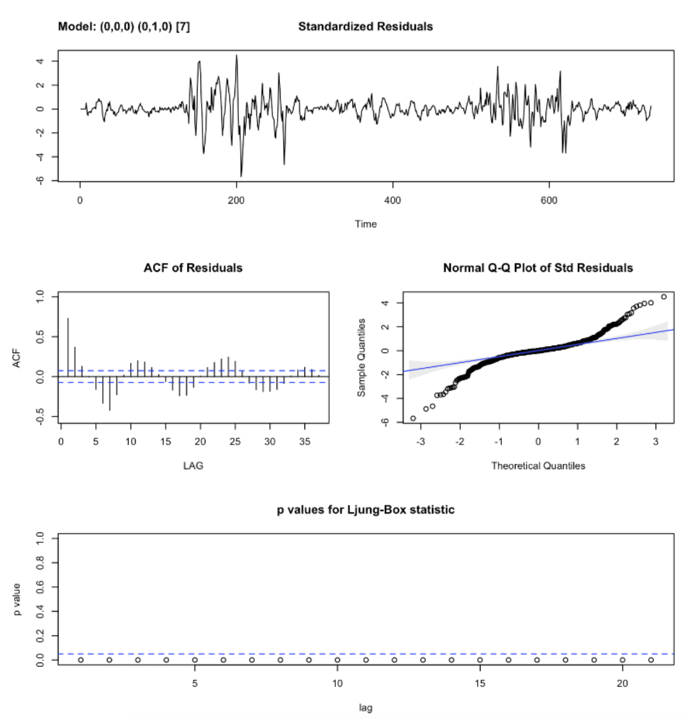

Figure 2

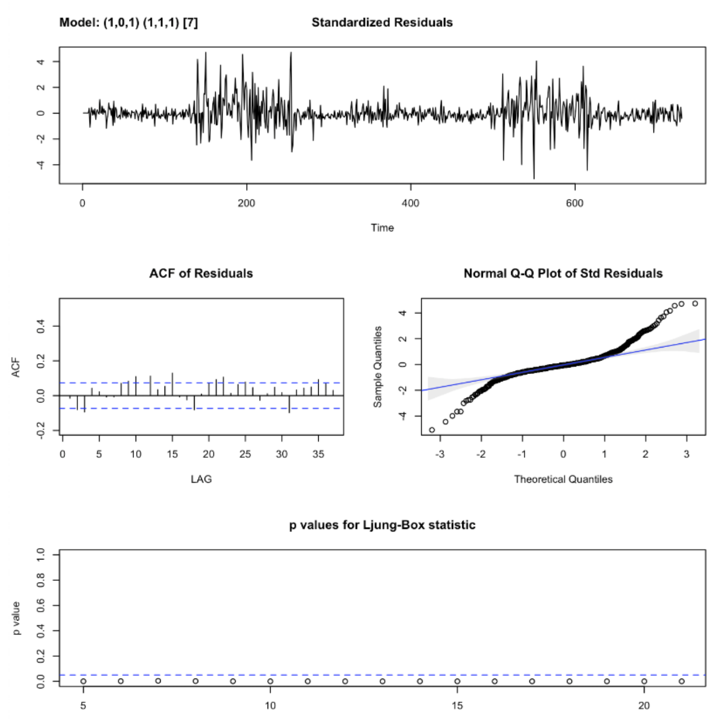

Figure 3

models with p=1, q=1, P=1 and Q=1. These results can be seen in Figure 3. Although there is an improvement in the ACF, the model still does not appear adequate. We began testing larger values of p to explore whether electricity demand is linked to demand at previous days, or local time-steps. Finally, after having tested many combinations and even testing with P=0 because we were not seeing any improvement, 2 models of p=7,d=0,q=1,P=0,D=1,Q=1,S=7 and p=6,d=1,q=1,P=0,D=1,Q=1,S=7 emerged. They can be seen in Figure 4; both passed the Ljung-Box test. In the end, we selected the second model with p=6 to favour a smaller model because the ACF's were almost identical. In addition, the p-values of the Ljung-Box are much higher for the first few lags of this second model. As such, we can conclude that the errors are not correlated however in all models the errors remain not

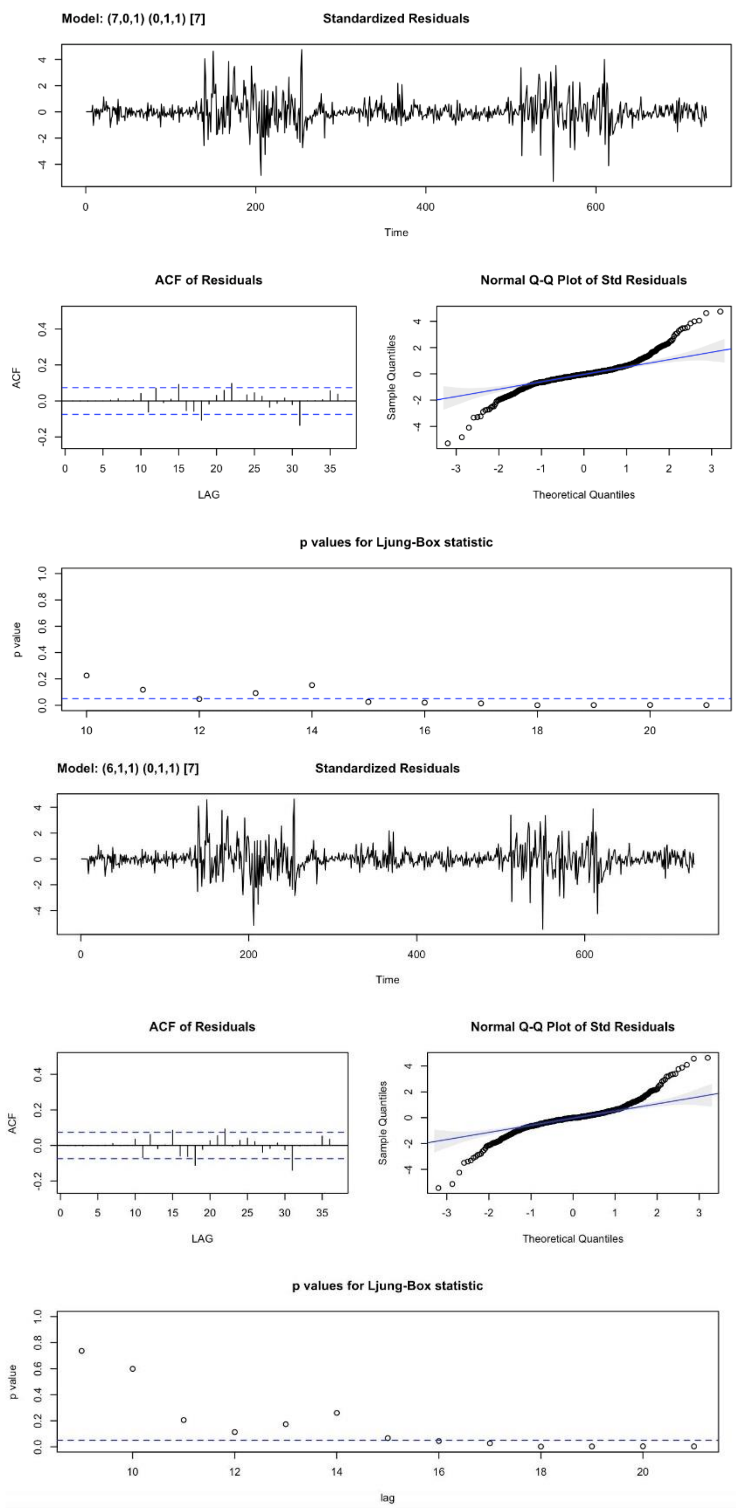

Figure 4

normally distributed and the residuals are showing patterns of larger and smaller sections, which is not ideal.

The next model which was tested was an ARX which is able to test exogenous variables and lagged demand values. The variables which were tested included HDD, CDD, humidity, wind chill, day of week, holidays and lag-1 of electricity demand. The results of both SARIMA and ARX can be found in Table 3.

Table 3: Performance of ARIMA Family Models in the Validation Dataset

Method	Bias	% Bias	MAPE (%)
SARIMA	2.91	0.322	5.18
ARX	70.4	2.07	3.85

The results show that ARX is the winner and outperforms the performance measure.

Overall Comparisons

Before selecting a method, we noted that up until this point, the residuals in the summer were higher vs the rest of the year for all methods; Figure 5 shows, as an example, the residuals of ARX & SARIMA in the last 2 years of the training set. We considered that not all models might perform equally throughout the year and that the possibility might exist to create a hybrid model. As such,

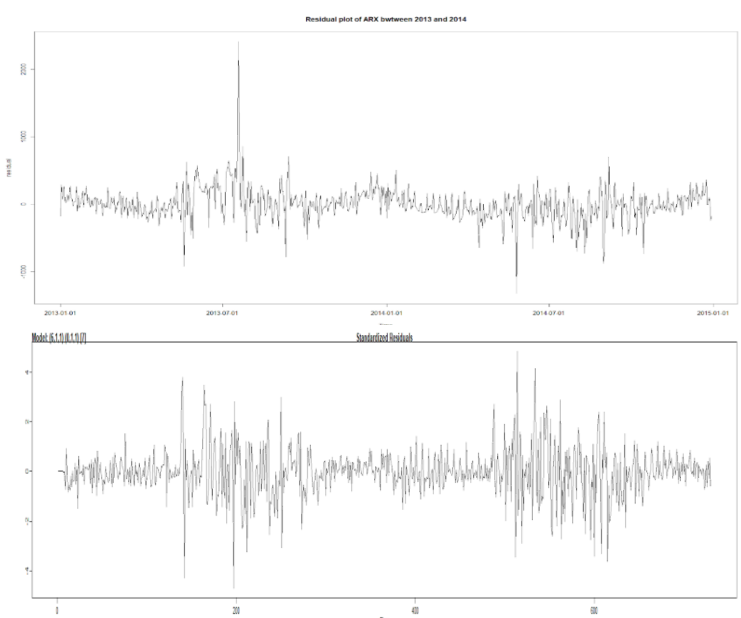

Figure 5

the data was analyzed to determine how the “summer” and “not summer” periods should be split. Based on a typical year, electricity demand begins to peak during May 1-Sept. 30 and thus, this was considered “summer”. Table 4 shows the subsequent calculations of all the methods during the “summer” and “not summer” periods of the validation dataset.

Table 4: Comparison of Methods in Summer & Not-Summer in the Validation Dataset

Method	Summer Period			Not-Summer Period		
	Bias	% Bias	MAPE (%)	Bias	% Bias	MAPE (%)
TBATS	-39.6	0.0414	7.99	30.7	0.962	3.96
Regression	-20	0.00278	3.68	156	4.46	5.39
SARIMA	12.2	0.845	7.49	6.54	0.248	3.6
ARX	-10.2	-0.00258	3.23	117	3.31	4.23

Based on the performance of the methods, it was determined that a hybrid model of ARX running from May1- Sept. 30 and SARIMA running for the rest of year would be tested in the validation set, alongside the winning method from each of the larger families. These results can be found in Table 5.

Table 5: Performance of Smoothing Methods in the Validation Dataset

Method	Bias	% Bias	MAPE (%)
Naïve: No Change (Performance Measure)	1.14	0.459	6.84
TBATS	1.12	0.588	5.83
Regression	94.6	2.92	4.84
SARIMA	2.91	0.322	5.18
ARX	70.4	2.07	3.85
Hybrid	8.78	0.37	3.31

As a result, the winning method in validation is the hybrid model with an MAPE of 3.31%.

Recommendation & Conclusion

Simply relying on the MAPE as the primary KPI, we would recommend a hybrid model of ARX & SARIMA. Based on its performance in the test set, RECO can expect that when the model is used it will yield an MAPE of 4.33%. We do note that the MAPE is higher out-of-sample which could be an indication of slight overfitting in the in-sample. It would be important to continuously monitor the model's performance out-of-sample to ensure the model can *actually* produce accurate forecasts. Also, a user would need to be aware that by using this model, they would need to obtain forecasts for HDD, CDD, humidity and windchill (the other variables of weekday/weekend and Holidays are easier to determine for future periods). The accuracy of these forecasts will have a direct impact on the error. In this case where we are forecasting for the next day only, it should not be too difficult to obtain these forecasts from organizations such as the National Oceanic and Atmospheric Administration, and they should be fairly reliable. However, for forecasting periods further into the future, obtaining reliable forecasts for these variables could be problematic. A solution to this could be to simply use SARIMA which is the best performing method of those not relying on explanatory variables. For the sake of comparison, it was run in the test set and yielded an MAPE of 5.82%. In conclusion, provided RECO has access to reliable forecasts, we would recommend using a hybrid model; if this is not the case, a SARIMA (p=6,d=1,q=1,P=0,D=1,Q=1,S=7) could also be an option.