HEC MONTRĒAL

MATH60638A- Forecasting Methods

Report Part III:

Presented to:

Prof. Debbie Dupuis

Presented by:

Dhaval Patel– 11296446

Hamid Aliakbarlou– 11291818

Negar Aminpour – 11305212

March 2023

1. Introduction

In the last phase of the project, we applied SARIMA and ARX models to the data. We then did overall comparisons and selected a winner to predict on our test set. In this final part, we finally conclude our analysis on the NYISO electricity demand.

2. ARIMA & SARIMA

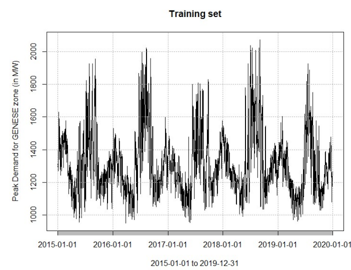


Figure 1.1: daily peak of hourly demand (2015-2022)

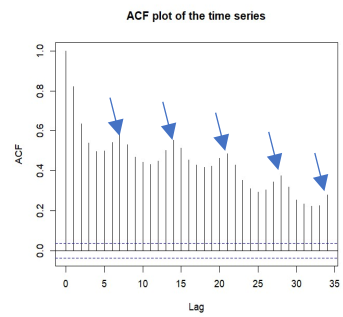


Figure 1.2: ACF of the plot series

In Part II of the report, the auto.arima function was used to fit an ARIMA model, which returned an AR(5) structure on the errors for our regression model. However, in Figures 1.1 and 1.2, we observe weekly seasonality in the time series and particularly ACF plots of daily peak electricity demand, so we excluded ARIMA and directly used SARIMA models instead. ACF of the time series evidently shows that it is non-stationary and we need differencing. The Yt in the time series was transformed using a Box-Cox method to achieve a normal distribution. The effectiveness of this

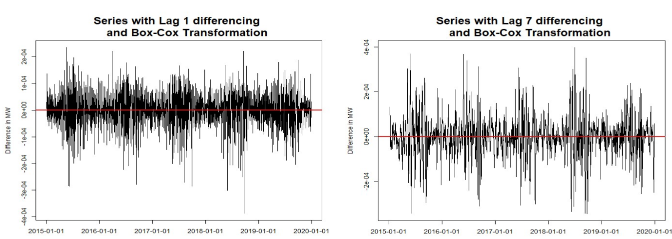


Figure 1.4: Differentiation of Transformed series at lag7 and lag 1

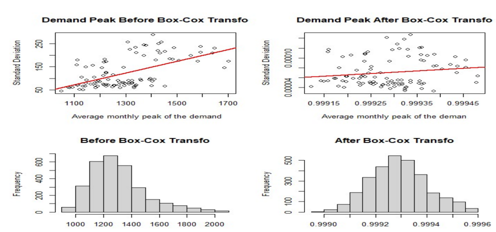


Figure 1.3: Box-Cox transformation effect

transformation is shown in Figure 1.3. We also tried differencing at lag 1 and lag 7 which could be indicative of a non-seasonal and seasonal differencing. It turned out that differencing at lag 1 makes the series closer to being stationary compared to differencing at lag 7 (Figure 1.4). However, to some extent the variance is non-constant through time and seasonality is still evident in the differenced series. To have an initial estimation of an appropriate SARIMA model for the data, the ACF and PACF plots of transformed series (Figure 1.5) were analyzed. Both plots indicates either sinusoidal or exponential decay with noise at lags 7, indicating P, and Q could be greater than or equal to 1. It means we need Seasonal AR and MA error terms. We thus go through an iterative process of creating SARIMA models until we find an adequate one. Hence, we did an incremental process by considering residuals behaviour and Ljung_Box tests and we achieved some adequate models. We selected one of them which seemed to be more consistent with the normality assumption according to residuals analysis.

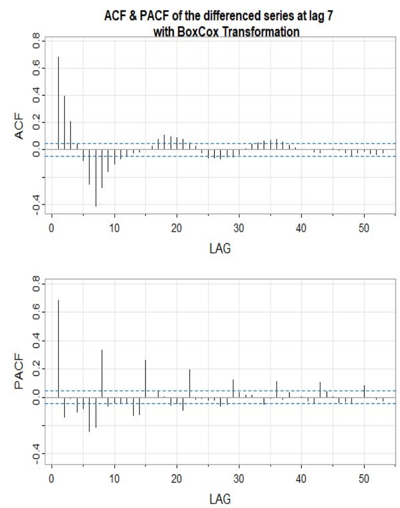


Figure 1.5: ACF & PACF of the differenced series at lag 7 with BoxCox transformation

From the differenced plot and ACF/PACF plots, we found that seasonality could be set to 7. Finally, the selected adequate model is (p,d,q) = (3,1,2) and (P,D,Q) = (2,0,1) with S=7. Although multiple SARIMA models were found to be adequate, only the one described above is discussed in this report. The SARIMA model was validated using residual plots (Figure 1.6). For this purpose, the ACF plot of the residuals, QQ plot of the standardized residuals, and Ljung-Box test were used.

Based on Ljung-Box test, the p-values were all above the threshold, and satisfied the 5% significance level. However, the QQ plot indicated a heavy tail effect, which suggests that the errors do not perfectly follow a normal distribution. We are aware of this breach in normality assumptions, however we still move forward as it is our most adequate model.

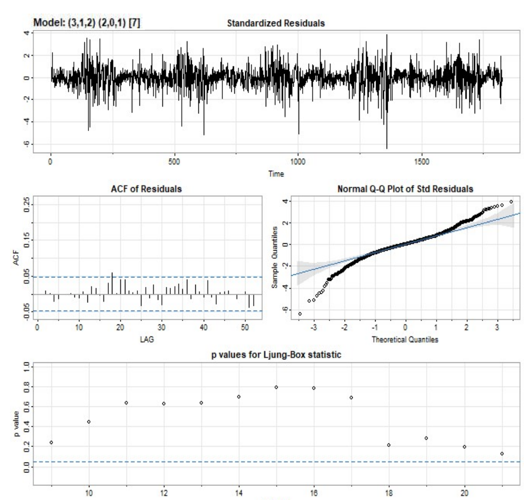
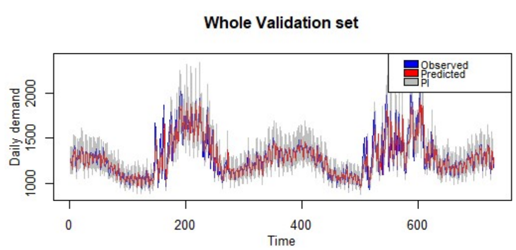


Figure 1.6: Residual plots of the model (p,d,q) = (3, 0,2) and (P,D,Q) = (2,0,1) with

The SARIMA model produced similar MAPE values for both expanding and rolling windows (Table 1). We did a Diebold-Mariano test and could not reject H0, thereby confirming that they are not significantly different. The rolling window had an MAPE of 4.95%, which was better than the Naïve benchmark but not as good as TBATS and regression with ARMA errors. The prediction interval for the model covered 93% of the values within the 95% PI which sounds good enough.

Table 1: Coverage & MAPE of SARIMA model with Expanding &Rolling windows

Model	MAPE (total)	Coverage of PI	MAPE (Winter)	MAPE (Spring)	MAPE (Summer)	MAPE (Fall)	Coverag (Winter)	Coverag (Spring)	Coverag (Sumer)	Coverag (Fall)
SARIMA (Expand)	**4.948%**	**93.1%**	**2.90%**	**6.37%**	**7.70%**	**2.83%**	**100%**	**88.46%**	**85.32%**	**98.91%**
SARIMA (Roll)	**4.95%**	**93%**	**2.89%**	**6.33%**	**7.69%**	**2.85%**	**100%**	**88.46%**	**84.78%**	**98.91%**
Regression ARMA Errors	**2.92%**	**94%**	**2.21%**	**3.57%**	**3.42%**	**2.49%**	**99%**	**92%**	**88%**	**97%**



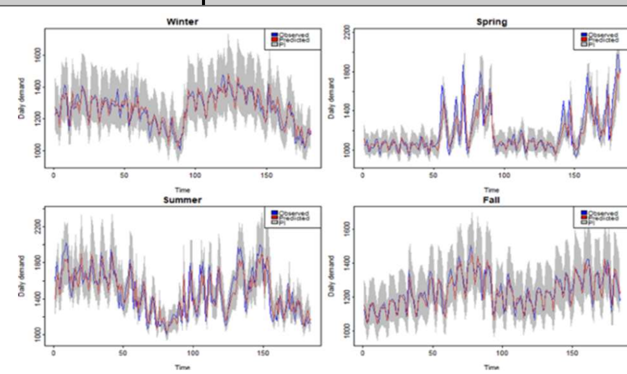


Figure 1.7: PI for SARIMA for a) whole validation and for b) different seasons

Our model performs better in Winter and Fall with more accurate predictions and better coverage but performs worse in Spring and Summer with less accurate predictions and lower coverage (Figure 1.7). However, the prediction uncertainty is greater in Winter and Fall compared to Spring and Summer due to the prediction interval clouds being thicker in cold seasons.

3. ARX

Initially we fit ARX(1) and ARX(2) models in the in-sample with CDD and HDD variables, however the results were not great. The ACF plots on the in sample showed peaks surpassing the 5% dotted lines on every 7th lag up to 0.6 dependency (Figure 1.8).

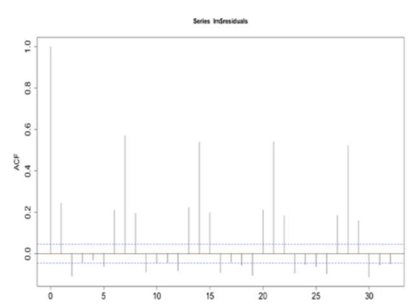


Figure 1.8: ACF ARX(1) with CDD + HDD

We thus scrapped these models and decided to use days of the week variables instead in an ARX(1) and ARX(2) model structure.

After looking at their respective MAPE and coverage by season in the validation set (Figure 1.9) we notice that their results are very close. After running a Diebold-Mariano test, we get a p-value greater than 0.05 thus we cannot reject H0 and we conclude that there is not enough evidence that the forecast accuracy of the two models is significantly different. We thus keep the simpler ARX(1) model. After viewing the ACF on the in-sample of the ARX(1) model with days of the week variables, we notice that there are peaks at each lag of 7. However, after forecasting on the validation data and viewing the ACF of the forecast residuals, we notice that no longer have the same pattern as seen in Figure 1.10. We have thus selected the ARX(1) with days of the week as our best ARX model in this section. We notice that in Winter and Fall, our model has almost 100% coverage with low MAPE and in Spring and Summer we have more than double the MAPE than in the other two seasons with lower coverage, 89% and 82% respectively (Figure 1.9). It should also be noted that the prediction interval cloud is thicker in Winter and Fall, and thinner in Spring and Summer. This implies that although our coverage is greater in Winter and Fall, the forecasts are more uncertain and variable compared to Summer and Spring even though their coverage is smaller (Figure 1.11).

	Winter	Spring	Summer	Fall
MAPE ARX(1)	3.30	7.08	7.86	3.18
MAPE ARX(2)	3.31	7.07	7.86	3.19
Coverage ARX(1)	100%	89%	82%	99%
Coverage ARX(2)	100%	92%	81%	99%

Figure 1.9: Coverage & MAPE of ARX

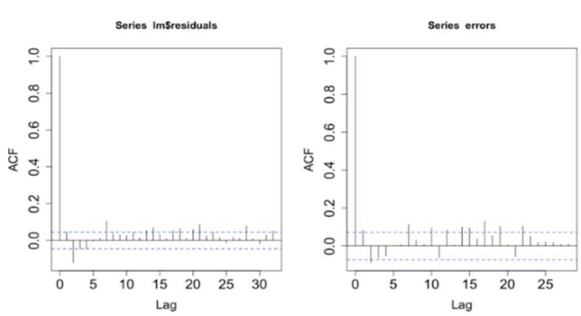


Figure 1.10: ACF of ARX(1) with Days of the

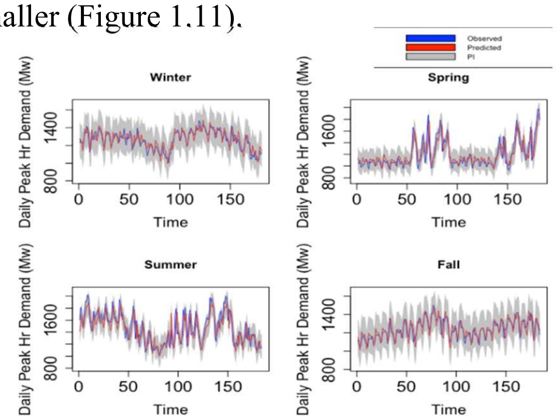


Figure 1.11: PI for ARX(1) with Days of the Week

4. Overall Comparison

We selected the best models/methods from each category (naïve, smoothing, regression, AR/ARX, and ARIMA/SARIMA) based on in-sample criteria such as residual analysis, ACF plots, and MAPE comparison. The Ljung-Box test was also used to validate the selected SARIMA models. The details of the selected models can be found in Table 2. We use MAPE, coverage and prediction interval cloud size to select the best model by section.

Table 2: Comparing winners of various model types

Model	MAPE (total)	Cov. of PI	MAPE (Winter)	MAPE (Spring)	MAPE (Summer)	MAPE (Fall)	Cove (Winter)	Cov (Spring)	Coverag (Sumer)	Cov. (Fall)	Time
SARIMA (Roll)	4.95%	93%	2.89%	6.33%	7.69%	2.85%	100%	88.46%	84.78%	99%	Few sec
ARX(1) w/ DOW	5.36%	92%	3.30%	7.08%	7.87%	3.18%	100%	89%	82%	99%	Few seco
Regression ARMA(5,0,0) Errors	2.92%	94%	2.21%	3.57%	3.42%	2.49%	99%	92%	88%	97%	29 min
TBATS (weekly + annual season) Expanding (Fit 7 days)	4.90%	92%	2.9%	6.14%	7.66%	2.86%	100%	88%	81%	99%	1 hour
Random Walk	6.03%	N/A	3.73%	7.45%	8.85%	4.07%	N/A	N/A	N/A	N/A	Few sec

It's important to note that a small difference in MAPE can have a significant impact on the financial cost of forecasting. For instance, each 1% difference in MAPE results in an average difference of 1261 MW, 1193 MW, 1466 MW, and 1237 MW in peak hourly loads per day for Winter, Spring, Summer, and Fall, respectively (Table 3). As such, even a 1% change in MAPE can have a significant financial impact, with an estimated cost difference of approximately \$1765, \$1670, \$2052 and \$1731 per season on average for each peak hourly daily load, based on the average cost of electricity in the GENESE region in 2023 (0.14 cents / KW)[1].

Table 3: Average Observation of Peak Hourly Load for the Day by Season

Average Observation (Winter)	Average Observation (Spring)	Average Observation (Summer)	Average Observation (Fall)	Avg cost 1% change in MAPE (Winter)	Avg cost 1% change in MAPE (Spring)	Avg cost 1% change in MAPE (Summer)	Avg cost 1% change in MAPE (Fall)
1260.89 (MW)	1192.95 (MW)	1465.49 (MW)	1236.75 (MW)	1765.40$	1670.13$	2051.68$	1731.45$

Among the models in Table 2, we could start with our benchmark and iteratively compare the models with each other. However, we observe that our regression model outperforms all other models and methods in terms of MAPE and it also has the closest coverage to 95% for all seasons in both cases (Table 2). Due to a lack of space in the report, we could not include predictions intervals for all five section winners, but the regression also had the smallest prediction interval clouds. We thus decided to select the regression model with ARMA errors as our best model. We still needed to verify if its prediction accuracy per season was significantly different than the other models. Hence we ran DM tests comparing the regression model with the random walk, the TBATS, the ARX and the SARIMA model "winners" per section for Winter, Spring, Summer and Fall. The results of these are contained in Table 4. We observe that for all four seasons, the prediction accuracy for our regression model is significantly different than the other models and methods. We, thus, go forward with the regression model with ARMA(5,0) errors as our best model.

Table 4: Diebold-Mariano Test by Season Comparing Regression with Random Walk, TBATS, ARX and SARIMA

Comparisons	Season	Sest_Statistic	p_value
Regression with ARMA errors - Random Walk	winter	7.623685	1.37E-12
	spring	6.516458	6.92E-10
	summer	8.620641	3.10E-15
	fall	8.019615	1.22E-13
Regression with ARMA errors - TBATS	winter	4.381468	2.00E-05
	spring	4.839559	2.78E-06
	summer	8.223156	3.57E-14
	fall	1.68773	9.32E-02
Regression with ARMA errors - ARX(1)	winter	-5.971972	1.22E-08
	spring	-5.214681	5.00E-07
	summer	-8.481275	7.34E-15
	fall	-2.509767	1.29E-02
Regression with ARMA errors - SARIMA	winter	-4.128015	5.59E-05
	spring	-5.379643	2.29E-07
	summer	-8.484339	7.20E-15
	fall	-1.657481	9.91E-02

5. Results on the Test Set

Based on the comparison we did in section 4, Table 5 presents results for prediction on test & validation set in cold seasons and warm seasons using our "winning" model.

Table 5: Reg. ARMA errors model performance on TEST & VALIDATION set.

Model	MAPE (total)	Coverage of PI	MAPE (Winter)	MAPE (Spring)	MAPE (Summer)	MAPE (Fall)	Coverag (Winter)	Coverag (Spring)	Coverag (Summer)	Coverag (Fall)
Reg. ARMA (5,0,0) (Test)	**3%**	**95.3%**	**2.57%**	**4.33%**	**3%**	**2.14%**	**97.8%**	**90%**	**94%**	**98.9%**
Reg. ARMA (5,0,0) (Valid)	**2.92%**	**94%**	**2.21%**	**3.57%**	**3.42%**	**2.49%**	**99%**	**92%**	**88%**	**97%**

We observe that the MAPE on the test set actually improved in Summer and Fall and

Winter and Spring compared to the validation set. It should be noted that the differences in MAPE between the test and validation sets are all less than 1% as seen in Table 5. Furthermore, the coverages got closer to 95% in Winter and Summer and strayed further from 95% in Spring and Fall. Globally, the MAPE only increased by 0.08% and the coverage by 1.3% confirming good performance of our winning model on the test set.

6. Insights & Conclusion

In conclusion, the linear regression with ARMA(5,0) errors proved to be the most effective model for predicting electricity demand in both warm and cold seasons. This model demonstrated superior performance in terms of MAPE and coverage, justifying its selection despite longer computation times. When predicting on the unobserved test set with our regression model, the results were quite similar to our validation set as mentioned in the results section with each season having a MAPE change less than 1%. It is crucial to recognize that even a small difference in MAPE can lead to significant cost savings due to the impact on peak hourly loads. This highlights the importance of selecting an accurate model for demand forecasting, as it can have considerable financial implications as seen in the comparisons section. While the model performed well in general, the higher MAPE values during warm seasons could indicate that some important factors or patterns were not captured by the model. Future research could explore different cut-offs for the CDD variable or incorporate additional features, such as humidity, to further enhance the model's accuracy and utility for electricity demand forecasting.

References

[1] Cost of electricity in Genesee County, NY (2023). EnergySage
https://www.energysage.com/local-data/electricity-cost/ny/genesee-county/