Forecasting Price Volatility in ERCOT

Jade Forest, Vicky Janita, and Tatiana Sokolova

4/20/2022

Introduction, Motivation, Relevance, and Objectives

ERCOT, or the Texas Interconnection, is only weakly connected to other interconnections within the continental United States. It has the sole responsibility of providing its own power to its consumers with access to limited imports. From 2011 to 2021, Solar and Wind Generation increased from being 8.5% to 28% of ERCOT's total fuel mix.[1] In recent years, concerns have been brought forth surrounding the costs of electricity increasing with increased renewable energy utilization. [2]

Additionally, in 2021, from February 13th through 22nd, the state of Texas experienced the costliest winter storm on record, with 164 hours of freezing temperatures, over $195 billion in resulting damages, and 250 deaths due to devastating power outages.[3] While the blackouts, and resulting spikes in electricity costs, were initially blamed on the grid's reliance on solar and wind, the leading factor that contributed to the shortage of electricity was the grid's heavy reliance on gas and the breakdowns that ensued due to much of the gas system not being winterized.[4]

This project seeks to explore and predict the price volatility of electricity before and after the large increase of renewable energy utilization in ERCOT and compare it to the price volatility of electricity after extreme weather events due to climate change. To do so, ERCOT's hourly settlement point price (SPP) will be forecasted and compared against the exogenous variables of renewable energy generation (from solar and wind) and demand (aka load) data.

Dataset Information

The three datasets used in this analysis were collected from ERCOT's historic data archive. The datasets include:

1. Settlement Point Price for each load zone in ERCOT[5],

[1] *https://www.ercot.com/gridinfo/generation*
[2] *https://www.forbes.com/sites/michaelshellenberger/2018/04/23/if-solar-and-wind-are-so-cheap-why-are-they-making-electricity-more-expensive/?sh=20f433181dc6*
[3] *https://www.austintexas.gov/sites/default/files/files/HSEM/2021-Winter-Storm-Uri-AAR-Findings-Report.pdf*
[4] *https://insideclimatenews.org/news/05022022/texas-storms-extreme-weather-renewable-energy/*
[5] *https://www.ercot.com/mktinfo/prices*

2. Generation by fuel type for entire ERCOT territory[6],
3. Load (demand) for each load zone in ERCOT[7],

The ERCOT data structures proved challenging to import. The most challenging aspect was the varying data structures throughout the same datasets. For each data set yearly .xlsx files were downloaded. These files contained separate sheets for each month in addition to superfluous metadata and summary sheets. Over the years ERCOT had updated the variable naming conventions, sheet names, and group names within the data. Additionally, some sheets had text and numeric information in the same column. For these reasons, tailored user-defined functions (UDFs) and for loops had to be built for each data set. The process of importing the data was incredibly time consuming, iterative, and provided a solid foundation for data wrangling in R.

In general, the following steps were applied to each data set.

1. Data was downloaded from ERCOT and stored in separate directories.
2. File names were extracted from each directory.
3. A user-defined function was applied via a loop to each directory to accomplish the following steps.
4. Excel file was opened.
5. Sheet names were extracted.
6. Each sheet was imported into R.
7. Variable names were standardized.
8. Numeric + Character columns were separated.
9. Variable classes were specified.
10. NAs were removed with na.approx().
11. Each monthly sheet was bound to a data frame.
12. A load zone was selected.
13. Variables were filtered into the desired date range.
14. Variables were converted into a time series function using mstl() with weekly, monthly, and yearly seasonality.

In addition to dplyr and other common tools, functions used for data wrangling included:

- filter_time(): filter by time values
- list.files(): get names from directory
- gc(): to clear items from memory
- lapply: to loop over lists

Below, the summary statistics for the final dataset are listed below, in addition to the first 10 rows of data. The summary statistics table was created using the st() function from the vtable package. The observations are hourly. Patterns of note from the summary table are that the Price

[6] https://www.ercot.com/gridinfo/generation
[7] https://www.ercot.com/gridinfo/load

minimum is negative and there is a large range for the Price, Wind Generation, and Solar Generation datasets. The correlation was also calculated between exogenous variables and the settlement point price, shown in Figure 1. There is relatively low correlation between the three exogenous variables and electricity price on an hourly basis.

Table 1: Summary statistics for January 1, 2012 – December 31, 2021

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Price ($/MWh)	87672	45.1	292.05	-10.94	18.85	36.31	9026.99
Wind Generation (MWh)	87672	6658.38	4610.69	0	3007.29	9368.55	24391.38
Solar Generation (MWh)	87672	412.78	1033.11	0	0	190.28	7210.28
Load (MWh)	87672	1164.37	243.16	631.8	989.09	1296.2	2117.4

Table 2: Top ten rows from the cleaned and wrangled data frame. Shown as a representation of the full data set.

Representation of Processed Data in Load Zone West

Date	Price ($/MWh)	Solar Generation (MWh)	Wind Generation (MWh)	Load (MWh)
2018-02-17 00:00:00	22.18	0.000000	4860.566	1057.6558
2018-02-17 01:00:00	19.45	0.000000	6308.652	1026.3656
2018-02-17 02:00:00	18.68	0.000000	7071.044	1002.2675
2018-02-17 03:00:00	18.08	0.000000	6064.856	1003.9397
2018-02-17 04:00:00	17.89	0.000000	6832.127	998.9583
2018-02-17 05:00:00	17.44	0.000000	8413.915	1004.8233
2018-02-17 06:00:00	18.21	0.000000	9159.690	1037.9753
2018-02-17 07:00:00	20.59	0.000000	9487.905	1077.7160
2018-02-17 08:00:00	21.17	9.082502	8944.382	1128.6305
2018-02-17 09:00:00	23.71	290.638823	8769.760	1165.4867

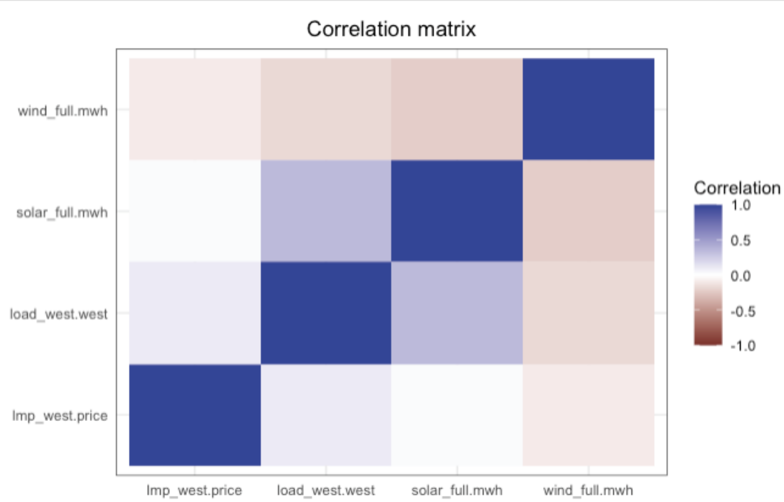


Figure 1: Correlation matrix between price and exogenous variables.

Analysis (Methods and Models)

After conducting the descriptive statistics shown in the summary table above, each time series was plotted (Figure 2) and decomposed into the component parts. The series was seasonally decomposed using the mstl() function with weekly, monthly, and yearly seasonality specified. Additionally, the PACF and ACF plots were analyzed.

The time series plots in Figure 2 show positive, seasonal trends in both wind and solar generation. Load has a positive, but less severe, trend with apparent seasonality as well. The LMP Price time series does not feature a significant positive trend, however, extreme volatility is present. Note that the large spike above $7500/MWh occurs during the Texas Winter Storm of February 2021 that was referenced in the Introduction.

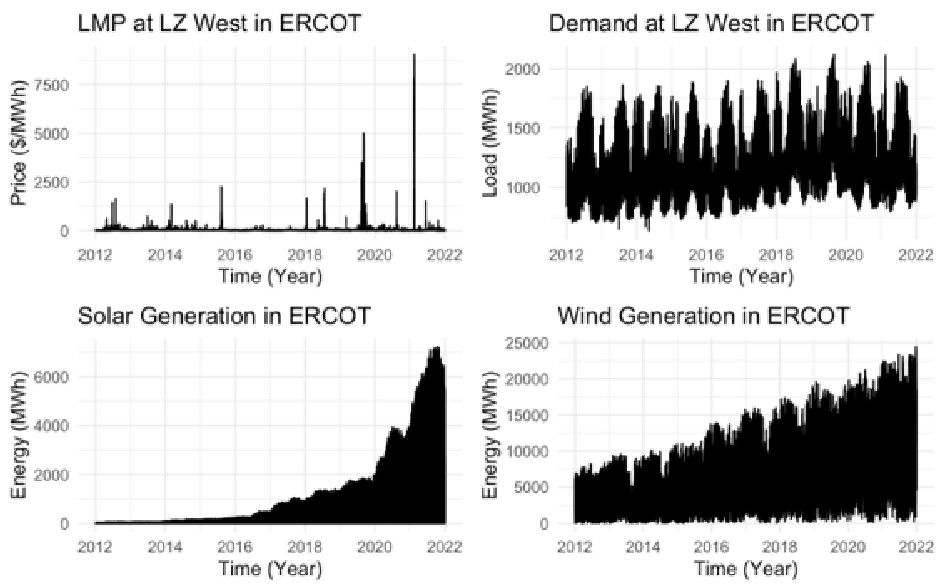


Figure 2: Time series plots of the four variables.

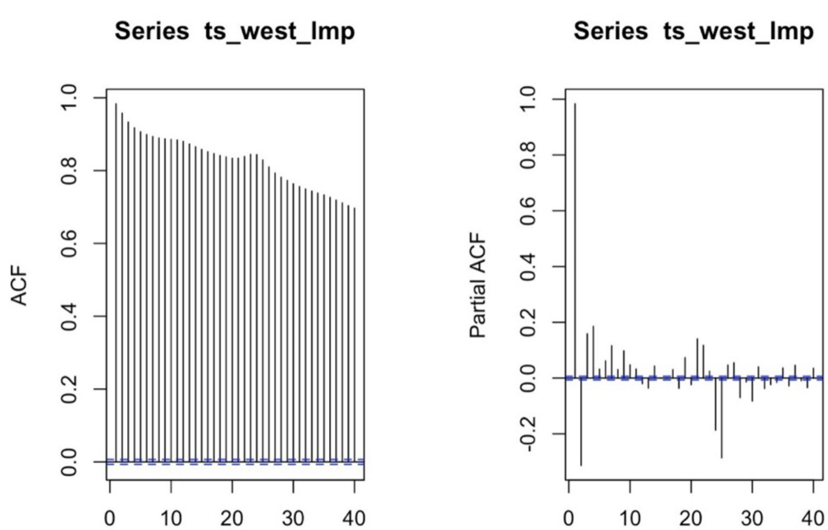


Figure 3: ACF and PACF plot of full LZ West Price data set.

We computed the ACF and PACF plot (Figure 3) of the full dataset using the Acf() and Pacf() functions to examine the correlation of the residuals and determine the order of the ARIMA model. Visual inspection of the plots show a slow decrease in the ACF, without noticeable seasonal scalloping. The PACF shows a cut off at lag 1. Based on this information, these plots suggest that an AR(1) model could be effective. However, this data has far more complexity and variation for an AR(1) model to be effective.

Eight models were developed, trained, and tested in this analysis:

- Seasonal Naïve (SNAIVE) was used as a preliminary baseline. The results are not discussed.
- ARIMA: Applied solely to price data using the auto.arima() function.
- ARIMA + Fourier: An arima model with fourier term.
- ARIMA + Fourier + XREG: An arima model incorporating hourly solar generation, wind generation, and demand as exogenous variables.
- ETS+STL: Implement seasonal trend decomposition and exponential smoothing method using stlf() function from the forecast package.
- TBATS: An exponential smoothing method of forecasting for complex seasonal data using tbats() function to the train dataset and forecast() to forecast 1-week hourly price (168 hours).
- Neural Network (NN): Single hidden layer neural network using nnetar() function to the train dataset and forecast() to forecast 1-week hourly price (168 hours).
- Neural Network (NN) + XREG: Neural network model incorporating hourly solar generation, wind generation, and demand as exogenous variables.

The test and train data sets were filtered into the groups described below using the filter_time() function.

After examining the time series as a whole and developing the initial models, the models were applied to three distinct datasets. Group 1 uses a training dataset from 02-13-2015 through 02-12-2018 and a test dataset from 02-13-2018 to 02-20-2018. Group 2 uses a training dataset from 02-13-2018 through 02-12-2020 and a test dataset from 02-13-2020 to 02-20-2020. Group 3 uses the same date range as group 2, however, outliers outside of the 99^{th} percentile are removed. The idea behind these three model groups is to examine how the models perform in a period of lower volatility (Group 1), a period of increased volatility (Group 2), and a period of increased volatility that has been artificially dampened (Group 3). It is hypothesized that the model Groups 1 and 3 will perform better, with lower overall RMSE and MAPE scores, when compared to Group 2. Technical difficulty caused some models not to function of all machines involved in this study. Given more time and computing resources, these models would be rerun to ensure parity across all groups.

Group 1: 02-13-2015 to-13-2018

Group 1 Time Series Decomposition

Using the mstl() function, the data was decomposed the training set to see the trend as well as the hourly, weekly, and yearly seasonality presented in Figure 4. The data shows a decreasing trend with several significantly high SPPs mainly at the beginning period of the training set. There are no clear patterns shown in the hourly and weekly seasonality plot. However, we observed significant spikes on the yearly seasonal plot at an approximately 6-month interval. Similarly, the remainder plot also shows significant spikes at 6-month intervals, indicating the time series may not have been fully decomposed.

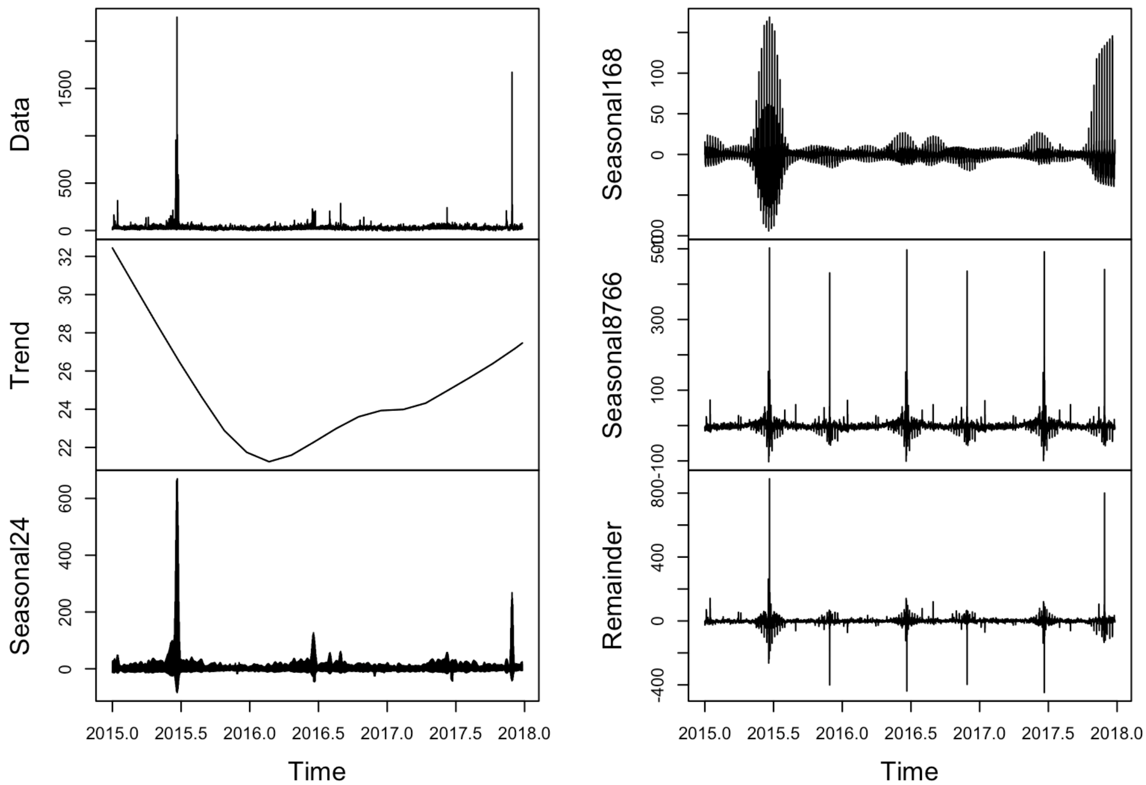


Figure 4. Decomposition of Group 1 Training Set

Forecast Accuracy

Table 3 summarizes the forecast accuracy score of models ran for Group 1. In general, the scores are relatively low with NN + XREG model showing the best model with the lowest RMSE and MAPE scores compared to the other models. The STL+ETS model appears to be the worst performing model with the highest RMSE score of 21.655. Due to a technical difficulty, the ARIMA + Fourier and ARIMA + Fourier + XREG models were not tested on this dataset.

Table 3. Forecast Accuracy Score on Group 1

Forecast Accuracy for Hourly SPP at Load Zone West (2015-2018)

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
ARIMA	-7.0904704	9.963362	8.264926	-90.89313	94.13112	0.8051147	6.770860
ARIMA + XREG	11.2846271	12.341847	11.285052	72.45485	72.45719	0.5606942	3.378733
STL+ETS	-4.8414284	21.654732	10.277607	-53.51476	78.47954	0.7380773	5.570049
TBATS	-6.2138251	10.298791	7.884719	-63.11396	69.72490	0.8088097	4.093063
NN	-3.3523772	7.552977	5.902631	-52.67041	62.13597	0.8001974	4.409079
NN + XREG	-0.7122782	5.347643	4.098679	-26.61614	40.43742	0.6624576	2.936040

Top forecasts graph

Figure 5 compares 4 out of 6 models run for Group 1. The black line represents the actual SPP and the colored lines are the forecast results. The ARIMA model, which is the third best according to its RMSE and MAPE score, shows a flat forecast for the entire 168 hours. Interestingly, the STL+ETS model that has the highest RMSE and MAPE scores were able to capture the test set's peaks and dips pattern accurately. The significantly high spike by STL+ETS model is most likely due to the high actual SPP at the end of the training set period. The NN and NN+XREG models were able to follow the test set's general pattern and are forecasting close to the actual SPP.

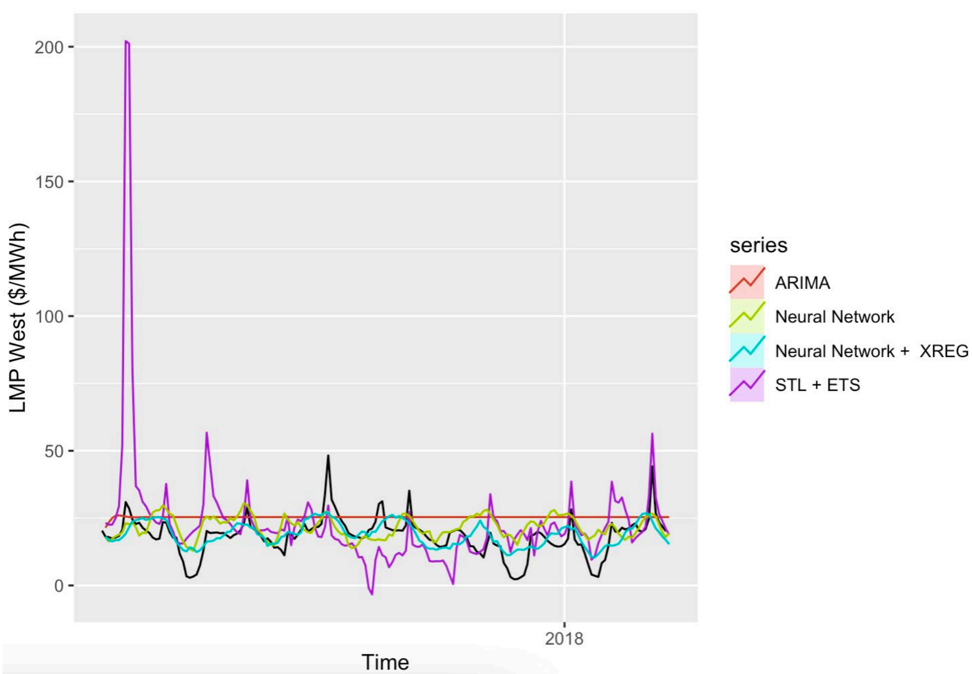


Figure 5. Forecast Comparison of 4 Models (ARIMA, NN, NN+XREG, STL+ETS)

Residuals

Figure 6, Figure 7, and Figure 8 plot the residuals, ACF, and histogram of the ARIMA, NN+XREG, and NN models, respectively. The residual plots for all 3 models show several residuals that significantly deviate from the mean. Due to extreme outliers in the data, it is difficult to evaluate the potential presence of a trend on the residual series. The ACF plots show significant self-correlation for most of the residuals. Finally, the residuals distribution of all 3 models violate normality due to the extreme outliers and have long positive tails.

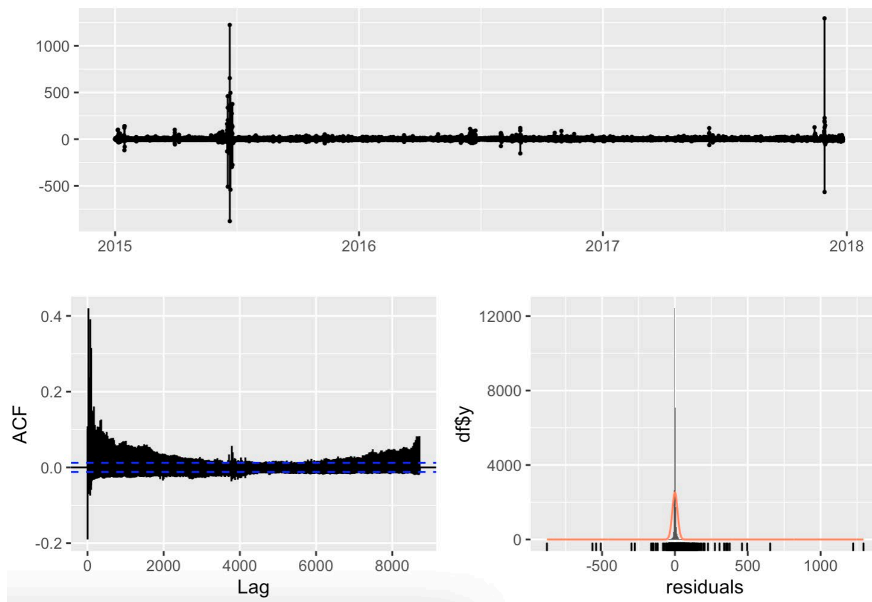


Figure 6. Checkresiduals() plots for the ARIMA(2,1,1) model applied to Group 1

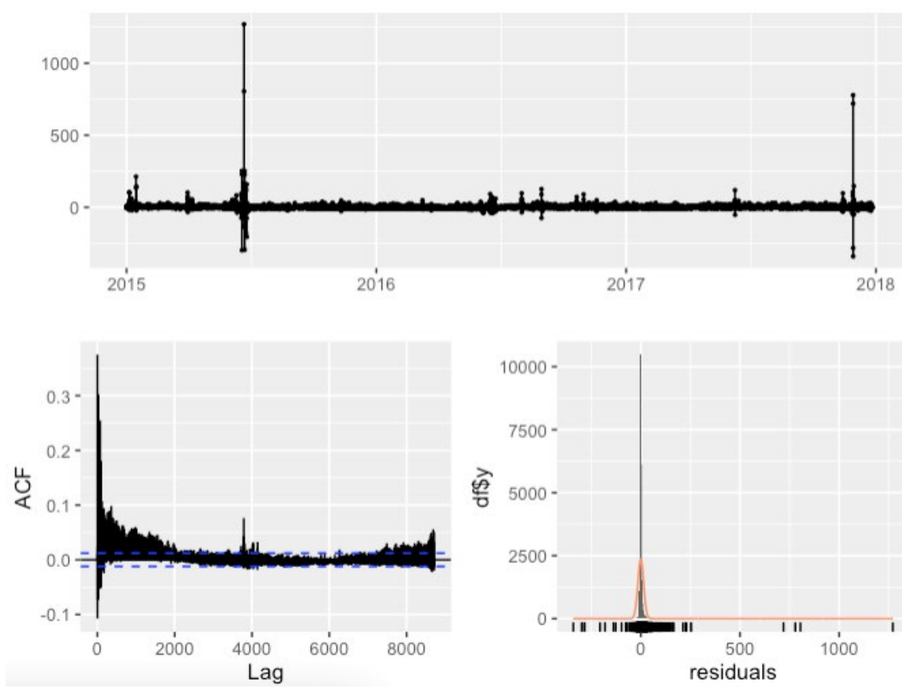


Figure 7. Checkresiduals() plots for NN+XREG model applied to Group 1

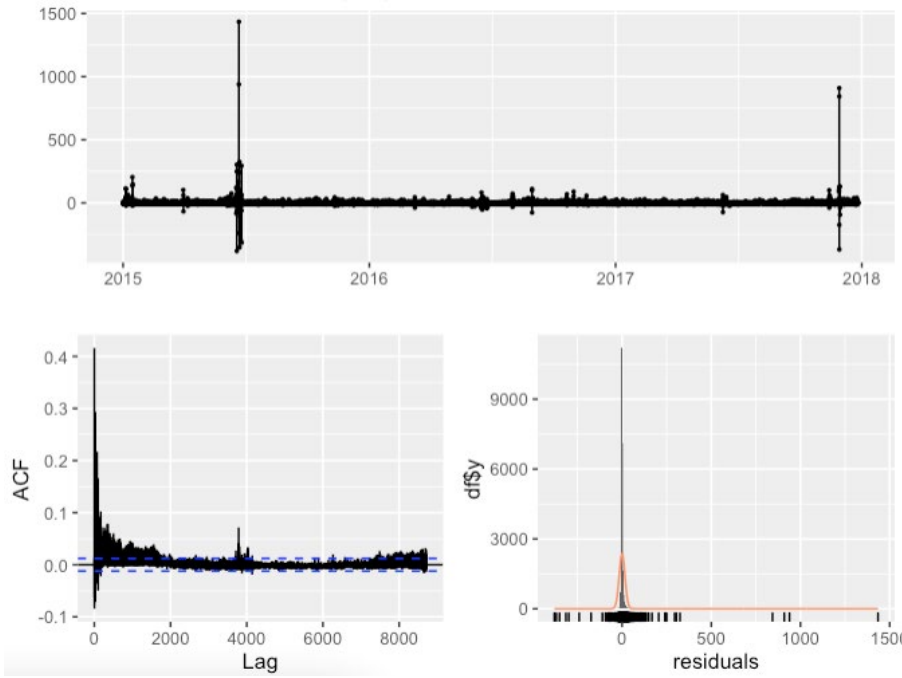


Figure 8. Checkresiduals() plots for the NN model applied to Group 1

Group 2: 02-13-2018 to 02-13-2021

Group 2 Time Series Decomposition

To decompose Group 2's training set, we used the mstl() function again. The decomposition (Figure 9) shows a slightly increasing trend through 2020 with an extremely increasing trend starting in 2020 (driven by the last few data points that led into the winter storm). There appears to be some yearly seasonality (most likely in the winter and summer months) as with the Group 1 dataset, but no explicit hourly or weekly seasonality. This is likely due to external factors impacting the SPP that we could not account for (e.g. politics and weather). Additionally, the remainder plot also shows significant spikes at 6-month intervals, indicating the time series may not have been fully decomposed by the mstl() function alone.

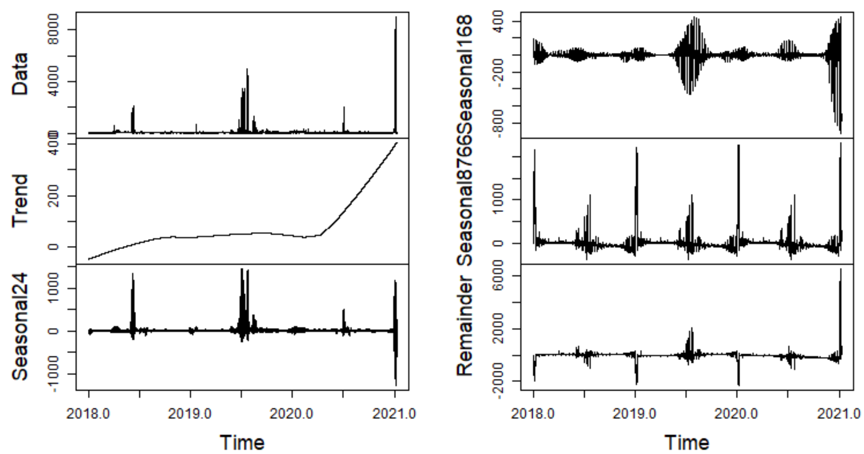


Figure 9. Decomposition of Group 2 Training Set

Forecast Accuracy

The accuracy of the Group 2 models is shown in Table 4 below. As expected, the events of this week could have not been predicted with anything close to accuracy, for all the RMSE values from the test set show an error in the multiple thousands range from that of the train set (which is particularly extreme given the highest RMSE in Group 1 was in the 20s range). The MAPE values performed even worse than the RMSE values with the values reaching 5 digits. Interestingly, the MAPE and RMSE were not correlated like in Group 1 with the worst RMSE value model being the best MAPE value model. Since the outcome range is so wide, it was determined that relative error measurements (RMSE) would be better to use than the absolute error measurements (MAPE), thus only RMSE values were used to determine the best models for Group 2. Additionally, due to lack of computing power, the ARIMA and ARIMA+XREG model would not run a 168 hour forecast (with it not being able to perform past 72 hours) for this date range. Therefore, these two models were excluded from this analysis.

Table 4: Forecast Accuracy for Group 2 models.

Forecast Accuracy for Hourly SPP at Load Zone West

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
STL+ETS	4370.642	5400.429	4773.591	-3583.45740	3711.4204	0.9591155	173.276544
ARIMA+Fourier	1079.779	3741.688	3278.277	-17260.67519	17315.6035	0.9636656	874.233286
ARIMA + Fourier + XREG	4177.333	5398.400	4571.769	-1437.02689	1553.9900	0.9634656	64.453006
TBATS	4773.944	5743.386	5034.066	-2432.48961	2570.9746	0.9535028	112.727798
NN	5220.933	6114.865	5269.165	-1278.09640	1433.8966	0.9544349	60.545600
NN + XREG	5443.672	6357.120	5446.165	12.61222	155.4673	0.9569870	3.251691

Top forecasts graph

With the above-mentioned limitations in mind, based on the RMSE values, the Neural Network + XREG model was the least likely to predict this extreme event (where previously it performed the best). Instead, the top 3 forecasting models were the ARIMA+Fourier model, ARIMA+Fourier+XREG model, and the STL+ETS model. To display the variety of how each model predicted the LMPs during Texas' freeze, all of the working models were plotted on a single graph in Figure 10 below.

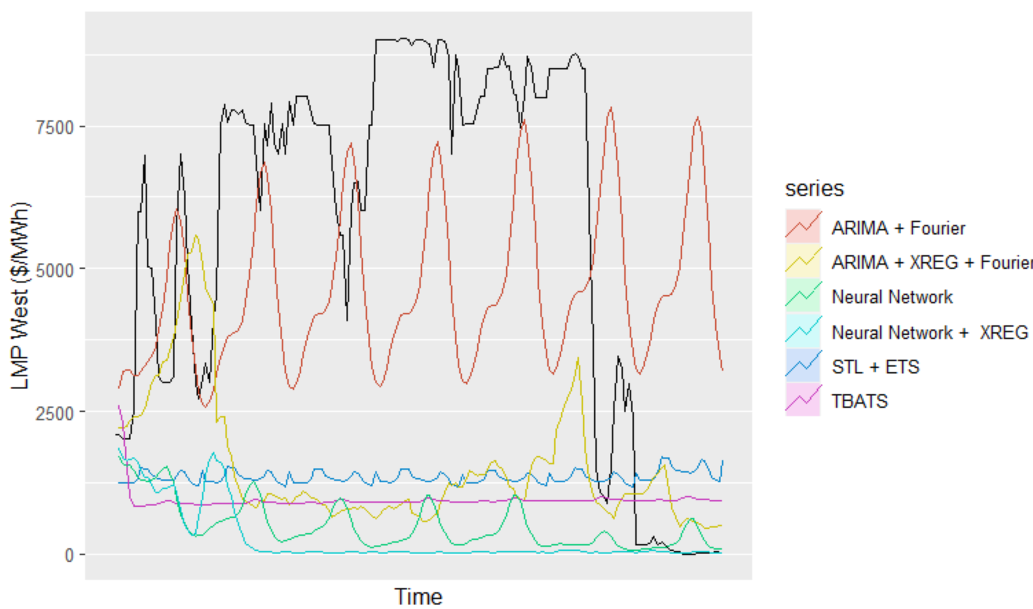


Figure 10: Timeseries plot comparing Group 2 models with the test dataset (black).

Residuals

Figure 11 - Figure 13 depict Group 2's residuals, ACF, and histogram for the best three models: the ARIMA+Fourier model, ARIMA+Fourier+XREG model, and the STL+ETS model. The

time series residual plots show a great deal of variance around the mean, particularly in the last year, highlighting the volatility of the SPP in general. The residual ACF plots show a high degree of autocorrelation and a decreasing trend for the two ARIMA+Fourier models. The residual ACF plot for the STL+ETS model has a steeper decreasing trend which cuts off significantly around lag 500. Surprisingly, the distribution graphs for the ARIMA+Fourier models do not have the extreme tails of the other two Groups and the STL+ETS model, but all show a relatively normal distribution.

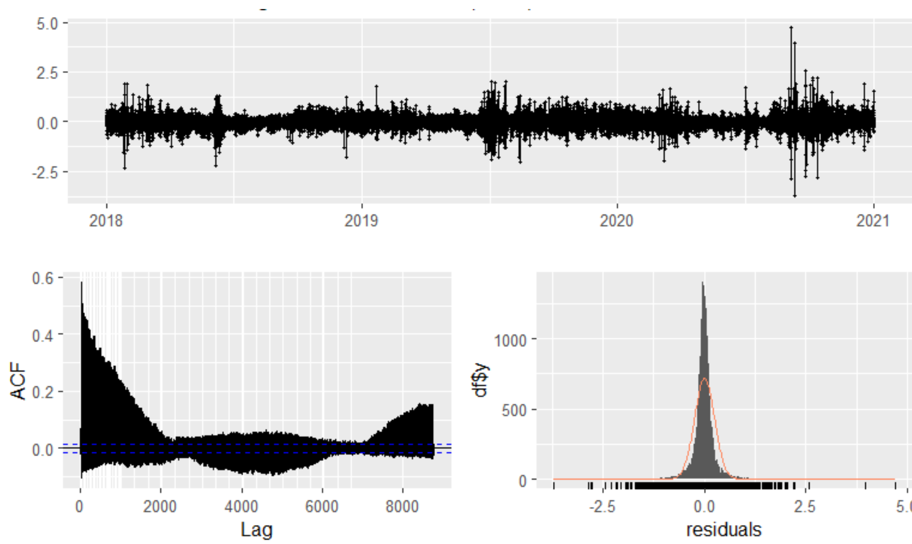


Figure 11. Checkresiduals() plots for the ARIMA+Fourier model applied to Group 2

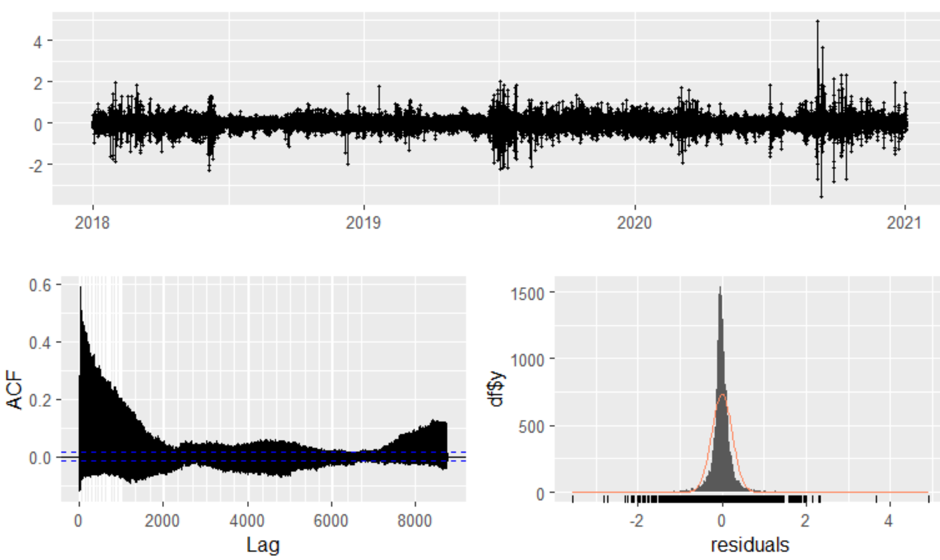


Figure 12. Checkresiduals() plots for the ARIMA+Fourier+XREG model applied to Group 2

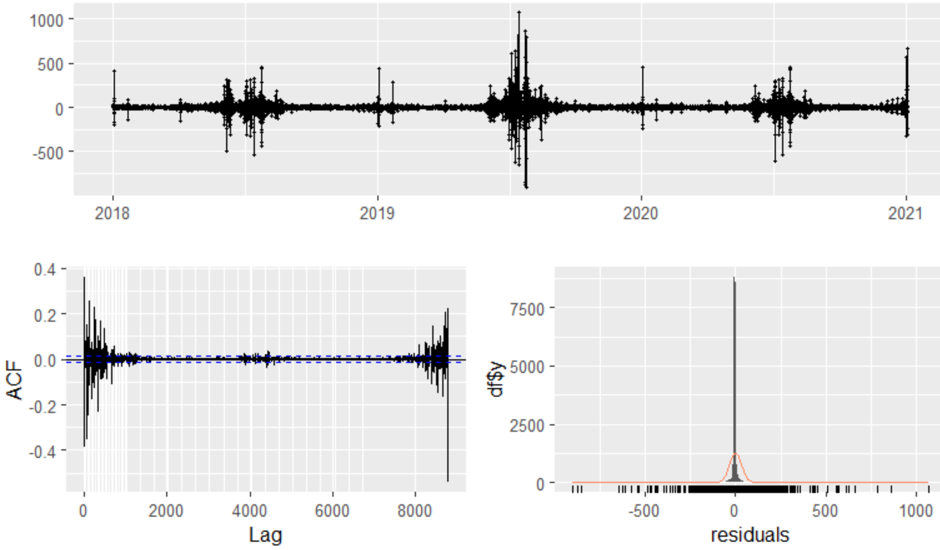


Figure 13. Checkresiduals() plots for the STL + ETS model applied to Group 2

Group 3: 2-13-2018 to 02-13-2021, Outliers Removed

The figure below compares original data histograms and box plots (on top) to those of the data with outliers outside of the 99th percentile removed (on bottom). The original data has a high positive skew. Removing the outliers decreased the intensity of the volatility, while retailing some positive skew.

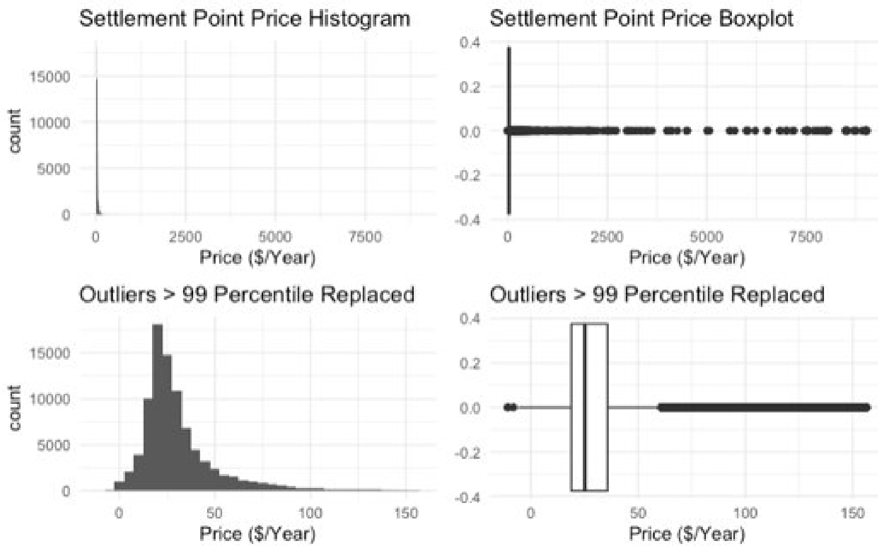


Figure 14: Comparison between original data histogram and boxplots with outlier removed histogram and boxplots.

Group 3 Time Series Decomposition

The time series decomposition of the outlier removed data shows a nonmonotonic decreasing trend. The trend is heavily influenced by the volatility of the data and may not represent true conditions. The seasonal components of the trend do not show neat seasonality present in some timeseries data. Again, the chaotic seasonality is probably due to external factors impacting price.

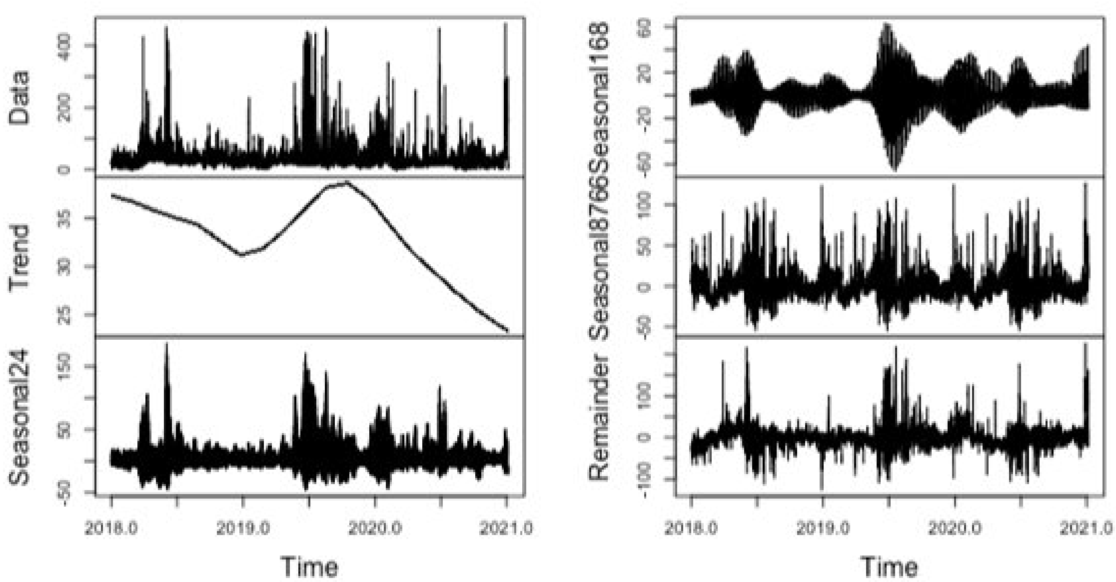


Figure 15: Time series decomposition of Group 3 dataset.

Forecast Accuracy

The accuracy of the Group 3 models is show below. Note that the Group 3 accuracy measures are greatly reduced from the Group 2 accuracy measures previously displayed, which suggests that removing the outliers increased the models forecasting ability. That said, the high MAPE values indicate that the models are not performing well. Overall, the Neural Network (NN) model performed the best based considering both the RMSE and MAPE values. The NN model fails to perform significantly better than the ARIMA model, this when considered with the high MAPE scores suggests that none of the models are adequately forecasting the data.

Table 5: Forecast Accuracy for Group 3 models

Forecast Accuracy for Hourly SPP at Load Zone West (Outliers Removed)

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
ARIMA	-10.83605	36.97711	24.15054	-191.45853	199.14488	0.84303	7.17767
ARIMA + XREG	3.33422	40.84142	24.44637	-52.82229	82.36112	0.88102	1.57365
STL+ETS	17.95983	44.81581	29.99052	-81.46890	149.80594	0.77934	4.60503
ARIMA+Fourier	-9.51390	44.54317	28.84090	-240.59701	255.91628	0.88454	10.60048
ARIMA + Fourier + XREG	31.29897	52.13200	38.43296	23.12619	79.79601	0.88558	0.83820
TBATS	-9.43452	39.13705	24.65583	-214.89507	224.70425	0.85961	8.47396
NN	13.92020	36.87887	22.38343	-45.98108	92.87312	0.82914	3.35736
NN + XREG	-23.12302	58.17318	44.09231	-126.47737	148.82374	0.92063	3.20522

Top forecasts graph

The graph below compares the top three Group 3 models with the test data (in black). The large spike towards the end of the test data is the February 2020 Texas Freeze, who's magnitude was reduced by removing outliers. The top performing ARIMA(2,1,5) and NN(1,9) models both fail to capture the spike or the normal variation in the test data, again suggesting that these models are not effective for this forecasting application. Additionally, the high order of MA terms in the Arima model likely suggests overfitting.

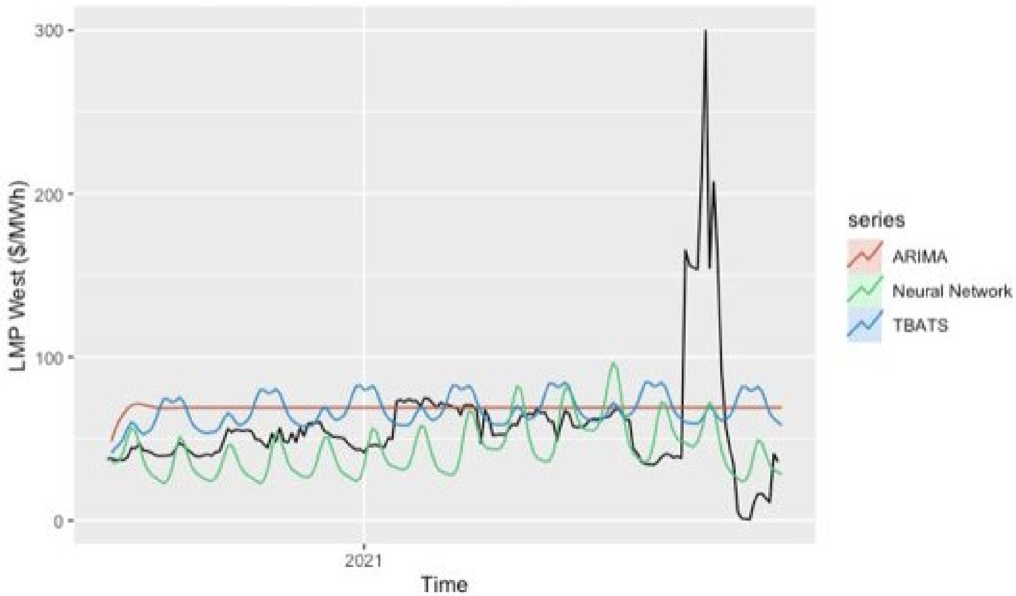


Figure 16: Timeseries plot comparing top three Group 3 models with the test dataset (black).

Residuals

The time series residual plots still show a great deal of variance around the mean, indicating that the models do not follow the original data well. Looking at the residuals in the three figures below there is evidently a non-normal distribution even after the 99th percentile outlines were removed. This likely occurred because the outliers were replaced with a mean value that was right skewed because of the extreme outliers. The residual ACF plots show a high degree of autocorrelation and a decreasing trend, likely because the models could not be able to predict the extreme price volatility of SPPs in ERCOT.

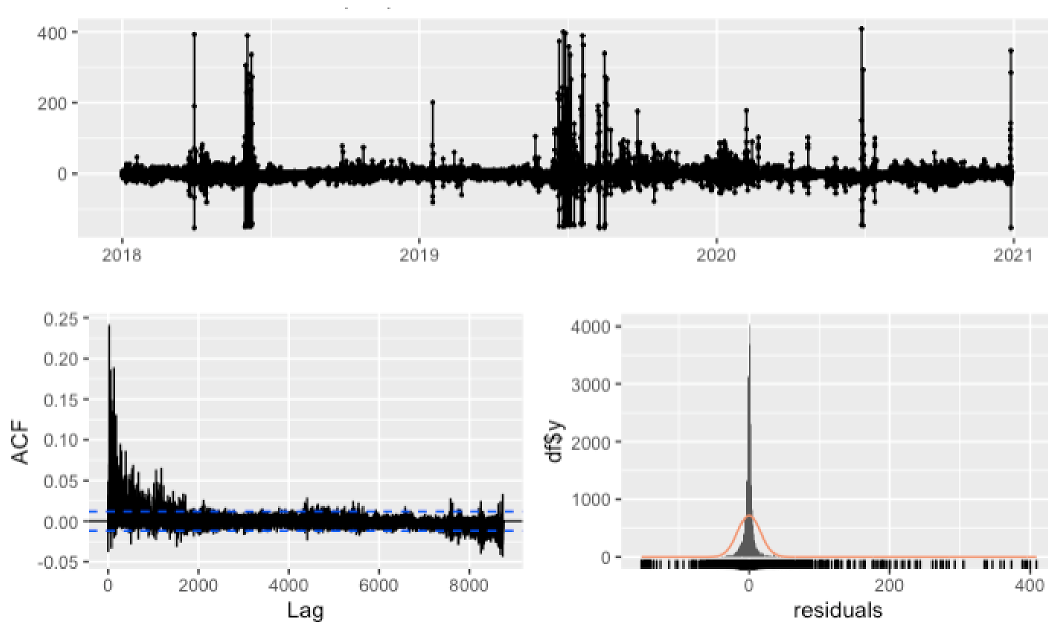
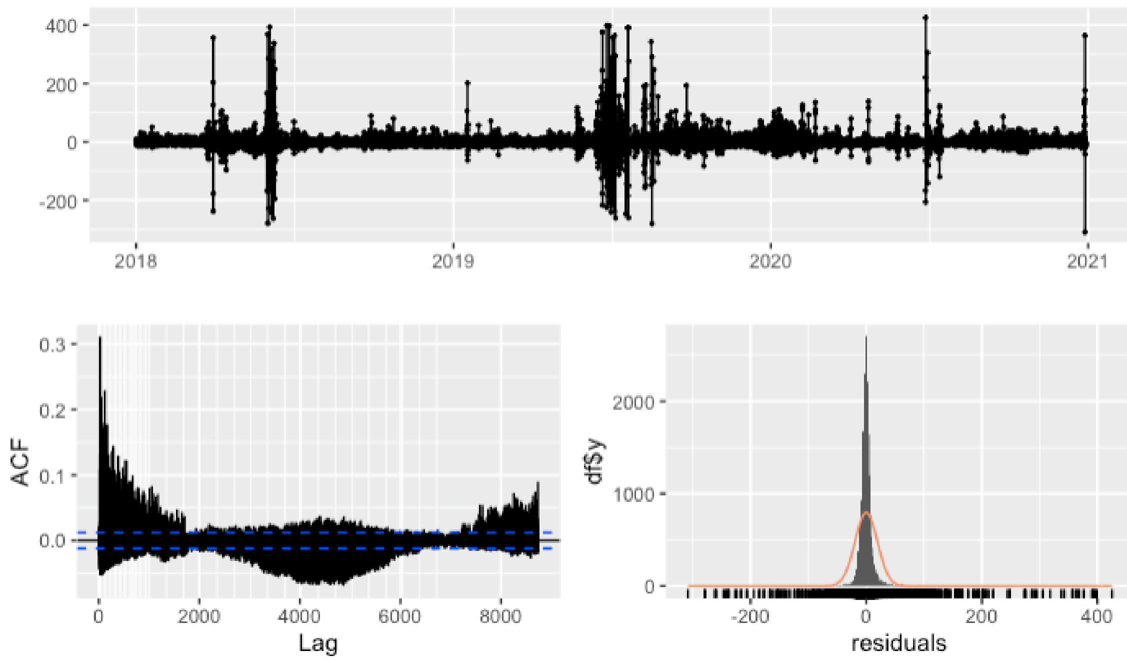
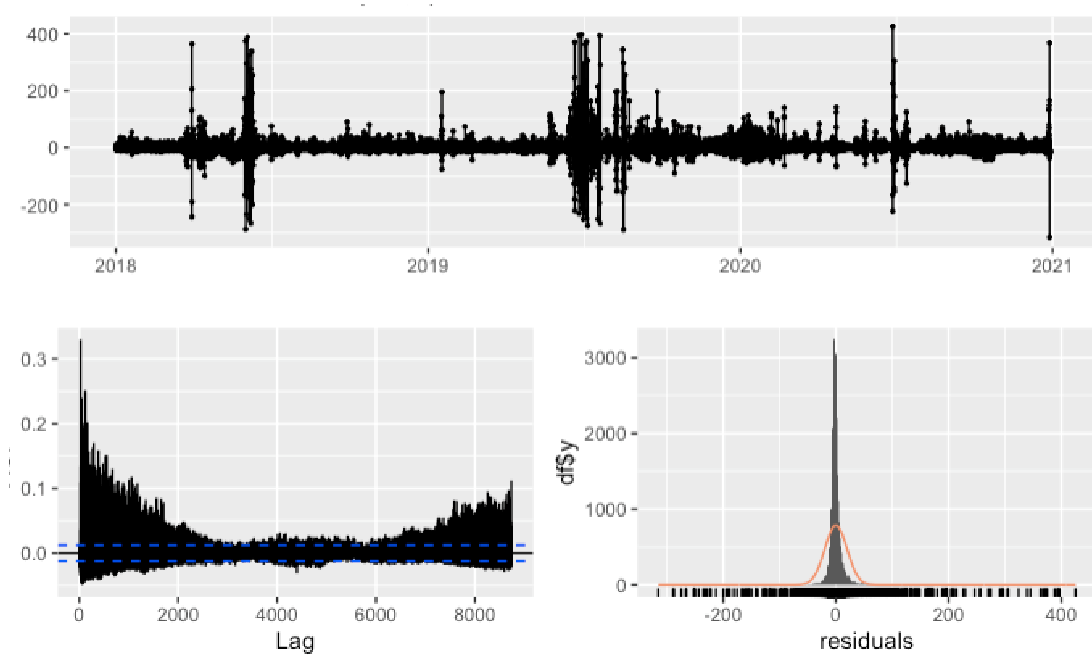


Figure 17: Checkresiduals() plots for the Neural Network model applied to group 3 data.



-

Figure 18: Checkresiduals() plots for the TBATS model applied to group 3.



-

Figure 19: Checkresiduals() plots for the ARIMA(2,1,5) model applied to group 3.

Conclusions

The models deployed across the three groups failed to generate accurate forecasts when compared with the test data. Our models were able to capture and produce a forecast that follows the trend of the test sets but failed to come close in predicting prices that existed in the extreme ranges (such as the price during a winter storm). The forecasts for Group 1 (the dataset that contained the earliest timeframe) performed better than that of Group 2 which contained the unrefined February 2021 Winter Storm data. Removing outliers (e.g. the storm) from Group 2's data to produce the Group 3 time series resulted in an improved model, but still Group 1's forecast outperformed. Thus, we were able to see that even without unexpected winter storms, price volatility in ERCOT has increased over recent years. Adding in the exogenous variables of wind and solar generation as well as demand did not improve our forecasts as we would have suspected, but rather made the models worse. Therefore, this showed that Texas' increased reliance on renewables is not driving up electricity prices in any significant way.

As mentioned throughout the analysis section, several of our models hung up our computers when forecasting the next 168 hours but worked with 72 hours. Therefore, if this model were to be used again, we would suggest that it not be used to predict beyond 72 hours into the future. Additionally, it would be interesting to improve upon our neural network forecasts by adding multiple hidden layers and nodes per layer to see if predictions could come closer to that of our test data. Since we saw worse forecasts when bringing in solar and wind generation data, it would be interesting to explore if natural gas production influences SPPs more and could be another potential exogenous variable to add to the model.

All in all, this exercise showed how increasingly difficult it has become to forecast SPPs due to increased variability caused by factors we could not possibly predict, such a climate change-induced extreme weather events. While ERCOT has been criticized in the past for not having a capacity market and letting economic competition alone keep electricity prices stable and the grid reliable, even the largest power market in the United States, PJM, has been struggling with forecasting SPPs. In fact, this year's Base Residual Auction (for setting future prices) was changed from auctioning electricity prices three years ahead of the delivery period to just one year ahead,[8] despite efforts to improve their models[9]. Without being able to time-travel, the increasing wave of random climate-change induced weather events, along with increasing global political tensions, could make even forecasting prices for the next 72 hours virtually unreliable. Therefore, we will not be betting on electricity prices based on these models any time soon.

---

[8] *https://www.icf.com/insights/energy/pjm-2022-2023-bra-auction-analysis*

[9] *https://pjm.com/-/media/library/reports-notices/reliability-pricing-model/20180425-pjm-2018-variable-resource-requirement-curve-study.ashx%20and%20write%20700-1200*