**Introduction**

|  |  |
| --- | --- |
| Forecasting electricity price  A Time Series Analysis on Average Electricity Price in Boston Area | Abstract  **Data Title:** Electricity per KWH in Boston-Cambridge-Newton, MA-NH, average price, not seasonally adjusted  **Data Source:** https://data.bls.gov/timeseries/ APUS11A72610?amp%253bdata\_tool=XGtable&  output\_view=data&include\_graphs=true  **Objective:** Select a time series model to forecast the average electricity price from April 2018 to March 2019 in Boston Area  **Methods:** Fitting ARIMA, SARIMA, and Holt-Winter forecast model in the training set and compare the forecast accuracy based on the test set. Choosing the model with best forecast accuracy and predict  **Conclusion:** SARIMA model has the best forecast power. The forecast shows that the Average Electricity Price will decrease during the summer of 2018 and will increase afterwards. The price will reach the peak during the winter of 2018 – 2019  Zhe Zhao  MA 585 |

Electricity price has become a significant concern for students living off-campus in Boston Are since landlords usually do not cover its cost. Unlike the rent, which is stated in the contract and will remain unchanged for at least a year, the electricity price is more difficult to predict. Many factors would have impact on the Electricity price in general such as weather, location and transmission system. Students expect more expensive electricity bills during Boston’s winter storm season and heat waves summers.

The Average Electricity Price data in Boston Area has unique characteristics. The competitive retail electricity market in Massachusetts allows customers to choose between competitive retail suppliers. The privatized market of electricity supply part leads to abrupt, short-term and unanticipated price shocks. On the other hand, government-regulated electricity delivery part exhibits seasonality at daily, monthly and annual levels (Weron, 2014). Above features provides the possibility of time series analysis and challenges at the same time. Despite complexity of the dataset, time series analysis technics learned in the course are applied to examine the trend, periodical patterns and noises. Due to the limit of skillset and relevant knowledge, the modeling process failed to capture some patterns, such as shocks and price spikes, appeared in the data. More details will be discussed in the discussion section.

The goal of this research is to choose the best forecast model and to forecast the Average Electricity Price from April 2018 to March 2019 Boston-Cambridge-Newton, MA. By studying the historical data of the average electricity price, the present study identifies autoregressive integrated moving average (ARIMA), Seasonal ARIMA (SARIMA) and Holt-Winter model. The model selection process is based on each model’s forecast accuracy.

**Methods**

**Details of Data**

The time series data is acquired from the website of Bureau of Labor Statistics without the seasonal adjustment. The time phase is from January 2008 to March 2018 with 123 observations. This is a monthly dataset with the mean 0.1736. The unit of the data value is dollar per Kilowatt Hour.

The Average Electricity Price data contains curved pattern. From the start of 2009 to the end of 2014, the average electricity price data shows a concave-up curved pattern. Despite the curved pattern, the time series data in complete picture exists an increasing trend. This trend is well-captured by the linear estimation. We see that the least square estimate of the mean for each year’s Average Electricity Price from January 2008 to March 2018 appears to be increasing with a small but statistically significant slope of 0.0005. Figure 1 presents the time series plots of the data and the linear estimation of the trend.

With the existence of non-constant variance and price spikes, we choose variable stabilization based on the result of Box-Cox transformation. The standard of choosing transformation based on lambda is “square-root” for any lambda near 0.5 and “log” for any lambda near 0. Figure 2 indicates that lambda is near 0.5, so the “square-root” transformation is applied to the data. However, the transformation does not stabilize the variance very much due to the curved pattern. The reason that this research adopts the transformation is to minimize the effect of price spikes on the later modeling process.

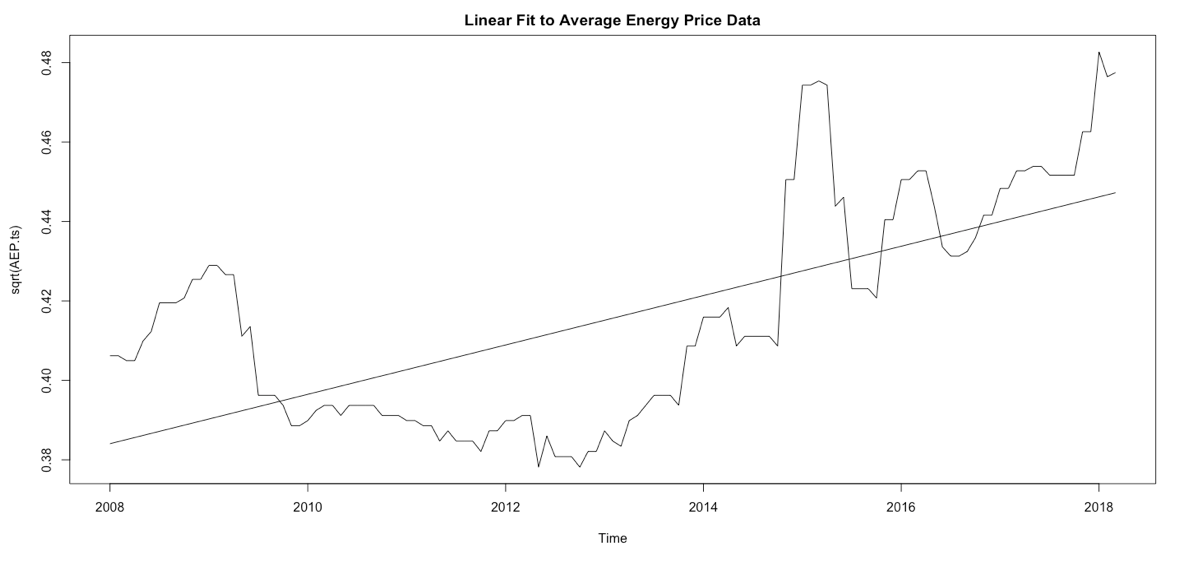


Figure 1 Time Series plots and the Linear Trend

The seasonal variations seem to be roughly constant despite the price shock around 2009 and 2015. To have a more complete picture of the trend component, seasonal component and random pattern of the data, the additive decomposition is adopted. In figure 3, we see clearly the trend component has obvious curved patterns but also an increasing linear trend. The seasonal component suggests a period of 1/10, which is near 1/12. In the later modeling and forecasting, we will adopt the period of 1/12 since it is closesr to the description of the monthly pattern, which was implied by the Average Electricity Price data.

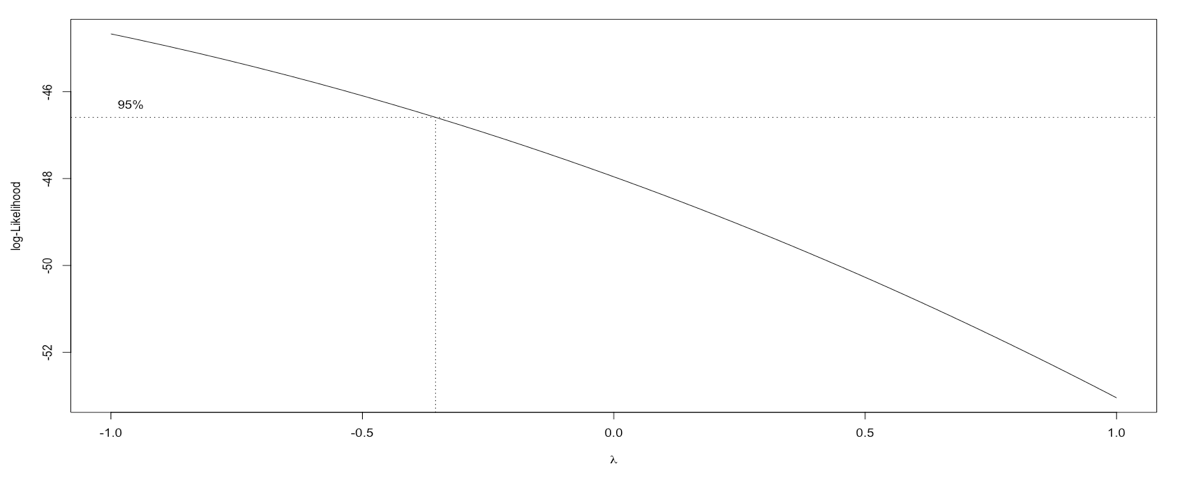


Figure 2 Box-Cox Transformation. The plot of log-likelihood against lambda. When lambda is near -0.5, the log-likelihood is maximized

**Statistic Techniques:**

The first step is to achieve the stationarity. As the trend component and seasonal component are overserved in the data decomposition, we eliminate two components by differencing techniques, which is defined by:



The elimination of the linear trend is conducted by differencing with one lag of the data. The Dicky-Fuller test is conducted to check the stationarity of the data. The test statistics is -5.7043 with the p-value, 0.01. The null-hypothesis of non-stationarity is rejected, so the data after the elimination of trend component is stationary. The second differencing is conducted on the basis of the first differencing data with the lag of 12 to eliminate the monthly seasonal component. Figure 4 shows a roughly stationary time series plot with constant mean zero and roughly stable variance despite some spikes.

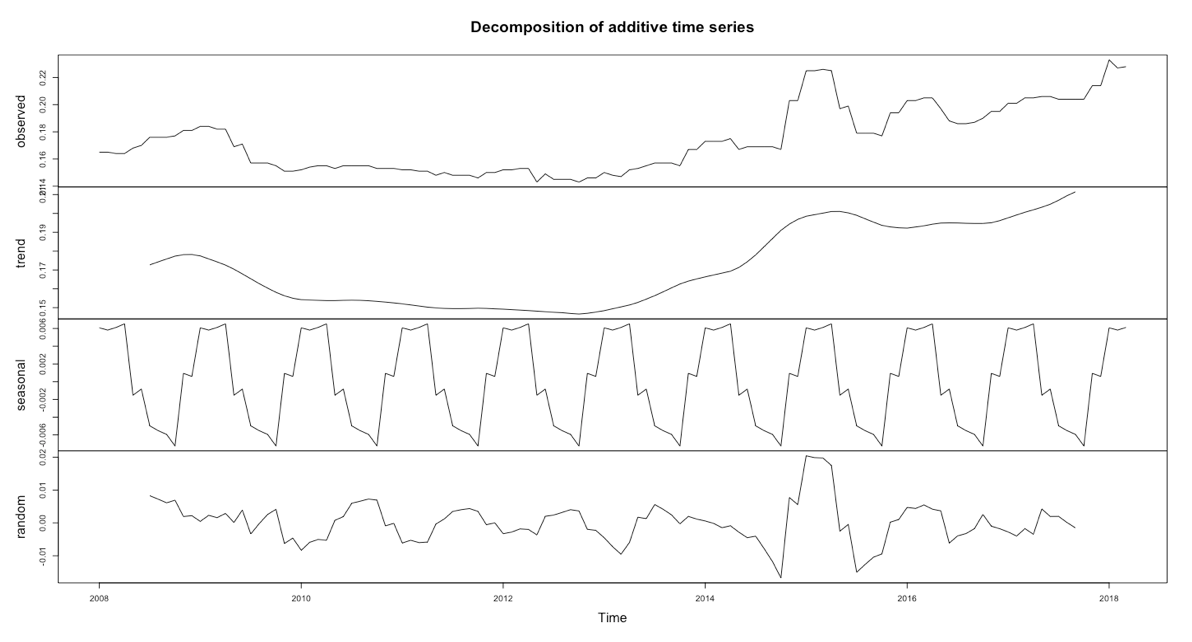


Figure 3 Additive decomposition

The dataset is divided into two part, the training set to develop the model, and the test set to examine the forecasting accuracy of models developed in the training set. The training set contains previous 111 observations (Jan 2008 to February 2017) from the original dataset and the last 12 (Mar 2017 to Mar 2018) data points from as the test set. The test group is set to contain 12 data points in accordance to the purpose of the present research, which is to forecast the average electricity price of the April 2018 to March 2019 (12 months).

ARIMA (p, d, q) model is fitted based on looking for orders by fitting all subsets of ARIMA (20, 1, 13) model to the differenced dataset and choosing the best-fit orders on the basis of the smallest criterion BIC (See Appendix). SARIMA (p, d, q) × (P, D, Q) model is fitted based on Auto-Correlation Function (ACF) and Partial Auto-correlation Function (PACF) plots of the differenced dataset. Two plots are developed in figure 5 to help determine the order of ARIMA part and seasonal part.

For determining the model, forecast accuracy is the most important criteria. After model selection process, different models will be fitted on the training set to forecast the average electricity price of March 2017 to March 2018. The forecast values will be compared with observed values from test set and accuracy is measured by, first, calculating the mean square root of errors, which is defined by the following, N is the number of observations:

sqrt (sum (observed value – forecast value) / N)

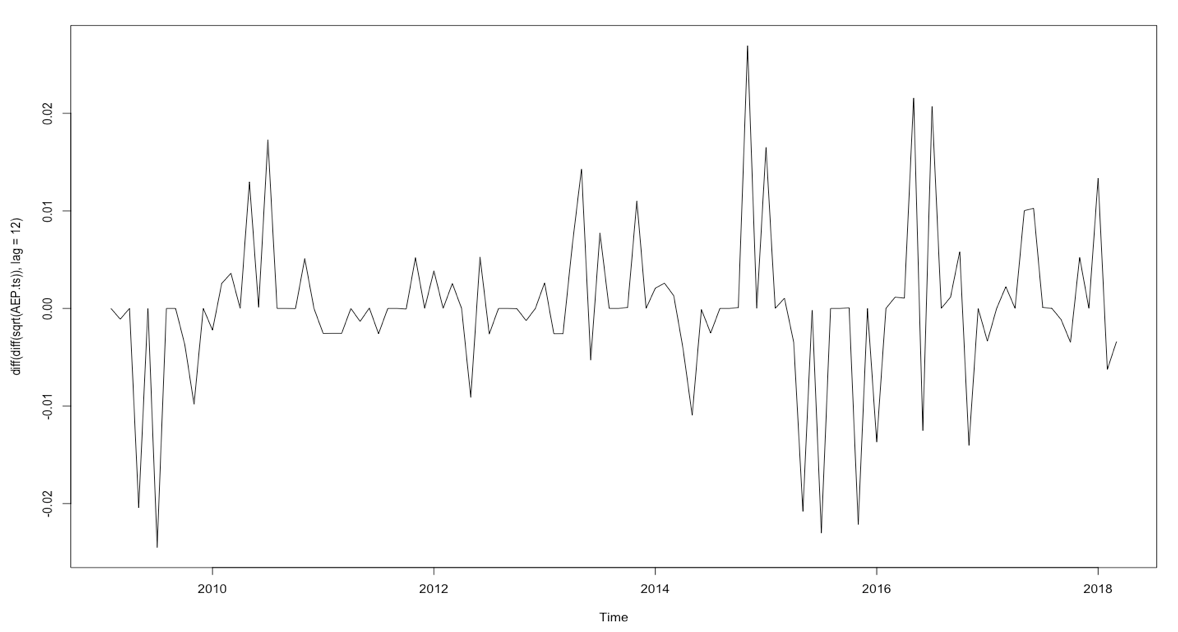


Figure 4 Time series plot of the data after the elimination of trend component and seasonal component by differencing

Second measurement is called Mean Absolute Percent Errors, which measures the mean proportion of true values that accounts to forecast errors. The following formula is defined to this measurement:

(1 / N) \* sum (abs (observed value – forecast value) / observed value)

Besides the ARIMA and SARIMA forecasts, Holt-Winter Forecasts, which utilizes the double-exponential smoother to calculate forecasts, is also generated to be compared with ARMA forecasting.

For model validation process, the study develops, firstly, standardized residuals plot and check the randomness of its distribution; secondly, ACF of residuals to check if any Auto-Correlation is not captured by the selected model; finally, the plot of p-value from Ljung-Box Test to see if the residuals has white noise distribution.

**Results**

The result of subsets fitting comparison is the ARIMA model with the least BIC, -22, ARIMA (18, 1, 12). In the fitted model, coefficient of X­­t - 1,Xt - 3, Xt - 5 to Xt - 17, et - 1 to et - 6, et – 8 to et – 11 fixed to be zero. The summary table of the chosen ARIMA (18, 1, 12) model is presented by table 1.

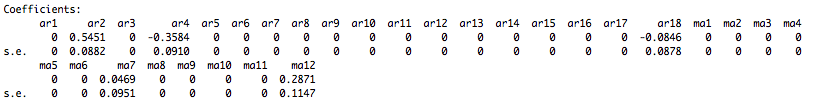
****

Table ARIMA (18, 1, 12)

The result of SARIMA is summarized in table 2. The order is (0, 1, 2) × (2, 1, 0). The ACF and PACF plots of the differenced dataset are presented in figure 5 to help determine the order of SARIMA model. For the seasonal part, notice that ACF of lag 1, lag 13, lag 25 and lag 37 indicates a pattern of sine wave decaying to zero while the PACF is significant till lag 13. The seasonal part is similar to AR (2) process, so we determine the order to be (2, 1, 0). If we only look at ACF and PACF at a few lags after lag 1, notice that ACF presents a sine wave decaying to zero while PACF has significance at lag 2. This pattern is similar with the process of MA (2). Therefore, for the non-seasonal part, we determine the order to be (0, 1, 2).

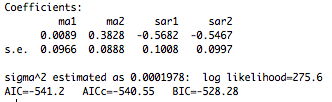
****

Table Result of SARIMA (2, 1, 0) X (0, 1, 2)

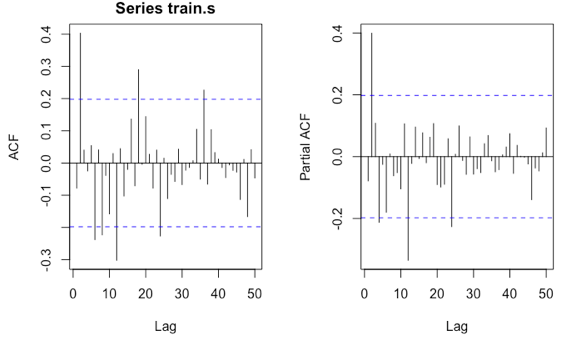


Figure 5 ACF and PACF plots for differenced dataset

The Average Electricity Price from March 2017 to March 2018 is forecast by the chosen ARIMA (18, 1, 12) and SARIMA (0, 1, 2) × (2, 1, 0) models, along with Holt-Winter forecasting. The result of forecast is compared with the test set of the original data, and the accuracy of forecasts is measured by root mean squared errors (RMSE) and mean absolute percentage error (MAPE). Table 3 presents the calculation results of the accuracy measurements of three forecasts. Figure 6 presents the forecast result of ARIMA (18, 1, 12); figure 7 is the result of SARIMA (0, 1, 2) × (2, 1, 0) and the result of Holt-Winter forecast is figure 8.

The best forecast model is SARIMA (0, 1, 2) × (2, 1, 0) because it has the least RMSE and MAPE among three models, and it well captures the pattern of the time series data plots of Average Electricity Price. The ARIMA forecast relatively well captured the variability of the data, but its accuracy is the least among three methods. Even though Holt-Winter forecast has relatively more accuracy, but it does not smooth the pattern of the seasonal pattern of the data. As we can see from the forecast plot of Holt-Winter Method, the forecast is simply a straight line indicating the weighted mean of previous few data points.

****

Table RMSE and MAPE

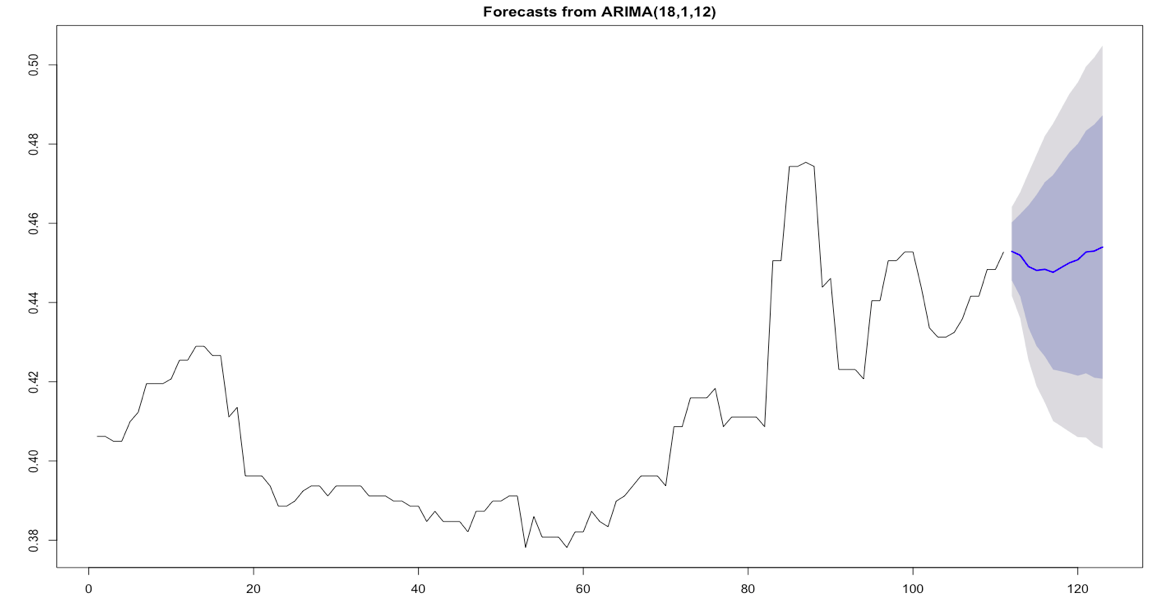
****

Figure 6 Forecast of ARIMA

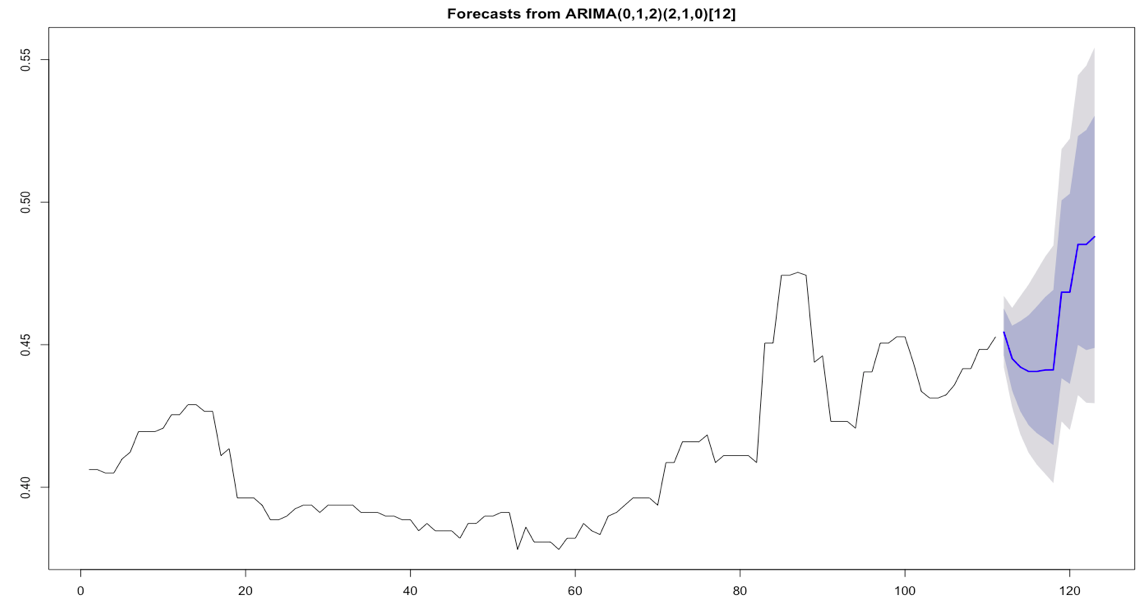


Figure 7 Forecast of SARIMA

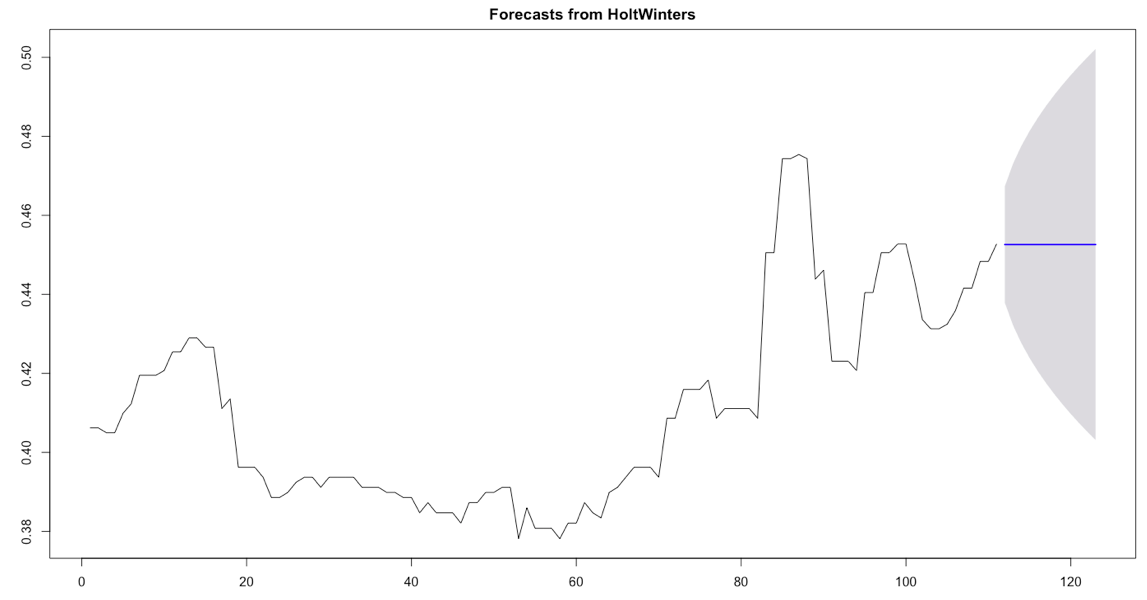


Figure 8 Forecast of Holt-Winter Method

**Forecast**

We choose SARIMA (0, 1, 2) × (2, 1, 0) and fit it into the original dataset to forecast the Average Electricity Price in Boston-Cambridge-Newton from April 2018 to March 2019. The result of forecast is presented by table 4 and figure 9 visualizes the forecast data. Table 4 also gives the 95% and 80% confidence intervals for the forecasting data. From figure 9, we observe that the average electricity price will decrease at start till the end of 2018 and will increase afterwards.

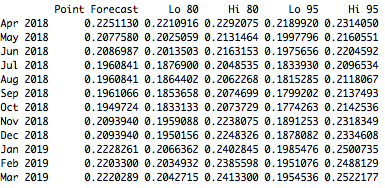


Table Forecast of April 2018 to March 2019

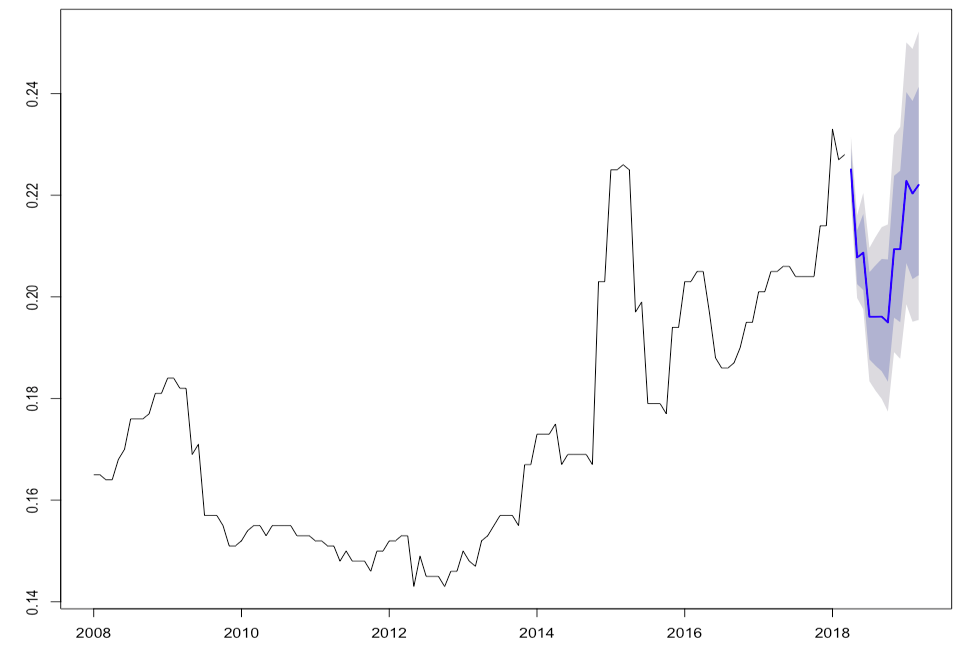
****

Figure 9 Forecast Plot

**Model Validation**

The diagnostic plots of SARIMA (0, 1, 2) × (2, 1, 0) model for the original dataset is presented in figure 10. The standardized residual plot shows randomness of residuals’ distribution. There is no significant ACF detected and the p-value of Ljung-Box Test are larger than 0.025, by which we fail to reject the null hypothesis of white noise process. The residuals of SARIMA (0, 1, 2) × (2, 1, 0) model present roughly white noise process.

To further check the randomness of the residuals’ distribution and independence, we look at Normal QQ plot and fit the ARIMA model to residuals. Figure 11 of normal QQ plot is developed, we find that despite two tails of data at both ends indicating non-normality, majority of data points indicates normal distribution. The auto.arima function is fitted in R studio and the result is ARIMA (0, 0, 0), which shows that there is not significant ARIMA process detected and ACF and PACF has no significance among residuals.

**Discussion**

Based on experiences, the electricity demand will decrease as the summer comes because a significant portion of population, students, will leave Boston Area. Previous years’ data also indicates decrease of electricity price during June, July and August. The electricity price will hike up when students come back for schooling. Another significant factor is the upcoming winter. People will increase their time at home and electricity demand will increase by more use of heater or other electronic devises in order keep buildings warm. Previous years’ data presents such a pattern. Electricity price from December to March are always higher than other months.

This Study only uses SARIMA model to forecast the Average Electricity Price. There is curved pattern in the dataset that is not well captured by the model. Many factors could be considered to estimate the Average Electricity Price, but the study fails to combine the parametric predictions due to the limit of knowledge. More advanced techniques, no matter for non-parametric estimations or parametric modeling, could be applied in the future research.

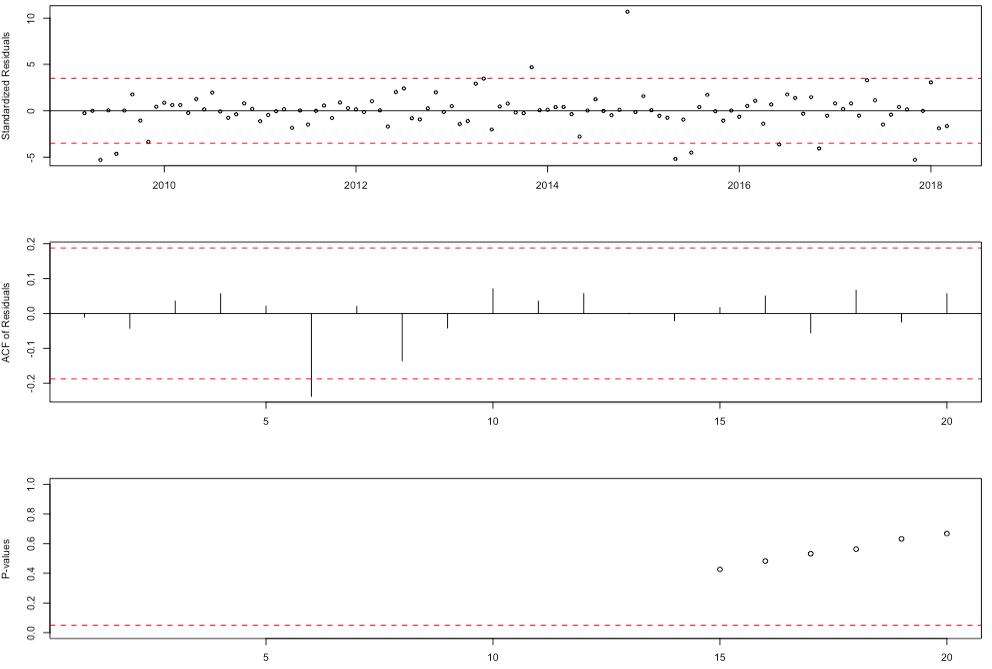


Figure 9 Diagnostic Plots

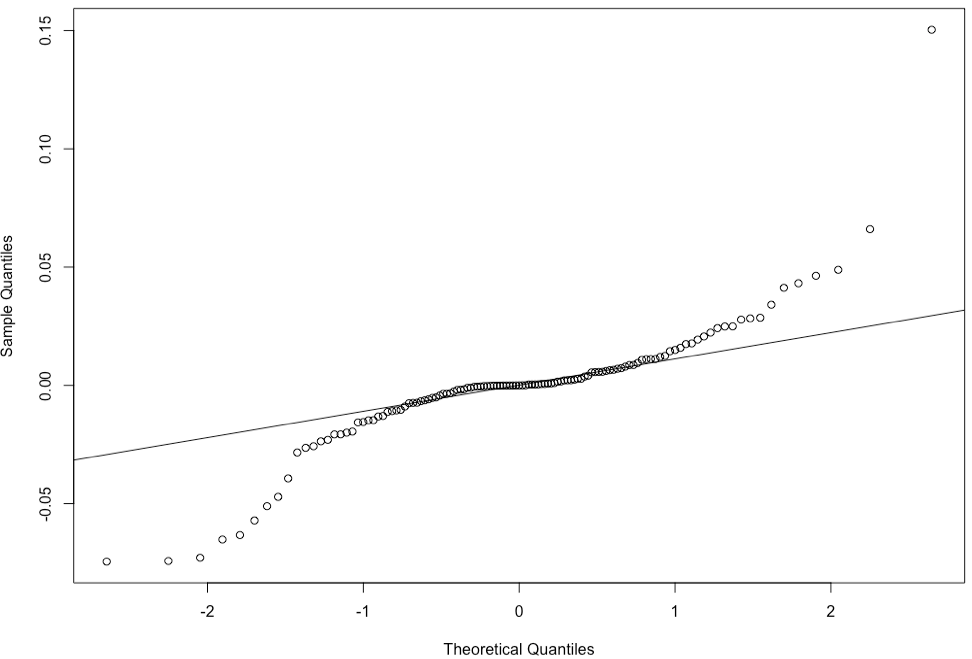


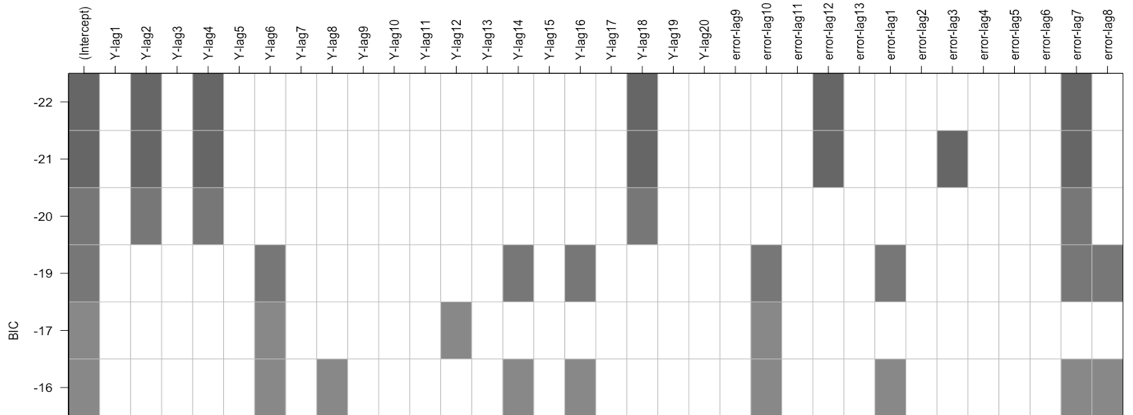
Figure 10 Normal QQ Plot

**Conclusion**

From April 2018 to August 2018, the average Electricity Price in Boston-Cambridge-Newton, MA will drop from around $0.26 per KWH to $0.196 per KWH. As the summer ends, the average price will increase afterwards from $0.1961 per KWH, and reaches the peak in January 2019, 0.223 per KWH. During the winter of 2018 – 2019, from November 2018 to February 2019, the average electricity price in the whole area will be stabilized around $0.22 per KWH level and the trend will continue to March 2019.

**Appendix**

Subset-ARMA fiiting:



The best fit model is ARMA (18, 12) with coefficients of Xt – 1, Xt – 2, Xt – 5 to Xt – 17, et – 1 to et – 6, and et – 8­ to et – 11 setting to be zero.

**Reference**

Rafal Weron, 2014. "[Electricity price forecasting: A review of the state-of-the-art with a look into the future](https://ideas.repec.org/p/wuu/wpaper/hsc1407.html)," [HSC Research Reports](https://ideas.repec.org/s/wuu/wpaper.html) HSC/14/07, Hugo Steinhaus Center, Wroclaw University of Technology.

Data Source: https://data.bls.gov/timeseries/APUS11A72610?amp%253bdata\_tool=XGtable&output\_view=data&include\_graphs=true