UNIVERSITY OF WATERLOO

MSCI - 718

Statistical Data Analysis

Project Report

ZONAL ENERGY DEMAND OF ONTARIO

Team Members:

1. Vishwanath Sai Ravuri -20736411
2. Disha Bhadauria -20742667
3. Drishya Sarah Mathew -20740832

CONTENTS

EXECUTIVE SUMMARY

Electricity is an inseparable and indispensable part of modern society. Ontario's electricity system allows Ontarians to live, work and play safely and in comfort. Planning and forecasting is essential to a reliable, sustainable electricity future for Ontario.

This report summarizes statistical data analysis and prediction results of electricity demand of various zones of Ontario. The aim of the report is to document the different analysis carried out with electricity demand data of the year 2017 to predict best the model for forecasting the electricity demand for the coming year. Data has been analyzed and forecasted with Holts, ETS, ARIMA, SVM, Logistic Regression, residuals were verified to be normally distributed and best fit is found based on the least RMSE value. Analytic tools used in this project are R studio, SVM and Logistic regression.

INTRODUCTION

Ontario is Canada's second largest province in area (after Québec). The population is the largest in Canada at approximately 13,750,000 (2015), with about 94 per cent concentrated in Southern Ontario relatively close to the St. Lawrence River and near the U.S. border. Most people, about 86 per cent, live in urban areas, such as Toronto, Ottawa, London or Windsor.

Currently, most experts look for Ontario to be at, or near the top, in terms of provincial growth. The expectations are that Ontario will lead in both Gross Domestic Product (GDP) and employment. Ontario had been near, or at the top in growth, the past two years and that has not translated into increased growth for electricity. Both British Columbia and Ontario have led the nation in growth, primarily a result of their strong housing markets, though this has had little impact on electricity demand more homes will create demand as the housing stock grows, but the impact is relatively small.

DATA DESCRIPTION

The **Ontario** transmission system has been divided into ten **electrical zones** namely Northwest, Northeast Ottawa, East, Toronto, Essa, Bruce, Southwest, Niagara, and West. Since electricity consumption is highest in urban areas such as Toronto, Ottawa, Niagara and Bruce we have selected these zones for our analysis along with Total Ontario demand.

➢ Data Source

The IESO is responsible for forecasting electricity demand in Ontario and for assessing whether transmission and generation facilities are adequate to meet Ontario's needs. Electricity demand data set for training and testing has been taken from IESO website.

➢ Training Set

Our training set consists of electricity demand for the year 2017 for the zones Toronto, Ottawa, Niagara, Bruce and Total Ontario. From the data we can infer that a mild and wet summer led to lower peak days and reduced total demand for electricity in Ontario last year.

Demand

Ontario's total energy demand in 2017 was 132.1 terawatt-hours, a 4.9 TWh or 3.6 percent reduction from demand levels in 2016. The highest demand peak of the year was recorded at 21,786 megawatts (MW) on September 25 during a prolonged period of unseasonably high temperatures. The peak was low by historical standards and unusual in that it occurred in the fall. Estimates suggest that approximately 1,700 MW of demand response was observed that day. The mild and soggy weather also led to some unusual demand patterns. Generally, peaks were lower and demand patterns flatter than in the past, resulting in some occurrences of weekend peaks.

➢ Test Set

We have taken the electricity demand data from 1st January 2018 to 15th January 2018 in our test data. Looking ahead, Ontario's power system is well-positioned to maintain reliability in 2018 and into the future.

➢ Data Transformation

In data analysis **transformation** is the replacement of a variable by a function of that variable: for example, replacing a variable x by the square root of x or the logarithm of x. In a stronger sense, a transformation is a replacement that changes the shape of a distribution or relationship. In most cases we do transformations to bring a non-normal data to normality. Transformations are also done to make it meet the assumptions of a statistical test or procedure i.e. to reduce errors.

In our data we have transformed our variables into square root of itself. After data transformation errors of data models are reduced as well as it helped us to bring the data into normality.

INITIAL ANALYSIS

➢ Outlier Identification:

An outlier is an observation that appears to deviate markedly from other observations in the sample. Identification of potential outliers is important for the following reasons.

1. An outlier may indicate bad data.

2. In some cases, it may not be possible to determine if an outlying point is bad data. Outliers may be due to random variation or may indicate something scientifically interesting. In any event, we typically do not want to simply delete the outlying observation. However, if the data contains significant outliers, we may need to consider the use of robust statistical techniques.

The box plot and the histogram can be useful graphical tools in checking the normality assumption and in identifying potential outliers. Outliers that are identified in the Ontario electricity demand are as follows.

Total Ontario Demand : No Outliers

Toronto : 3 Outliers [7141, 7315, 7352]

Ottawa : 5 Outliers [1207, 1284, 1280, 1263, 1236]

Niagara : 3 Outliers [605, 629, 640]

Bruce : 11 Outliers [264, 255, 283, 281, 309, 269, 222,163, 135, 171, 153]

➢ Stationarity Test:

A stationary time series is one whose properties do not depend on the time at which the series is observed**.** So time series with trends, or with seasonality, are not stationary the trend and seasonality will affect the value of the time series at different times.

adf.test:

The null-hypothesis for an ADF test is that the data are non-stationary. So large p-values are indicative of non-stationarity, and small p-values suggest stationarity. Using the usual 5% threshold, differencing is required if the p-value is greater than 0.05.

Another popular unit root test is the *Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test*. This reverses the hypotheses, so the null-hypothesis is that the data are stationary. In this case, small p-values (e.g., less than 0.05) suggest that differencing is required.

By performing above two stationarity test for the training data it is shown that data is almost stationary with no trend and seasonality with very few exceptions (Niagara).

➢ Implemented Models:

- Exponential Smoothing (Holts)
- Error Trend Seasonality (ETS)
- ARIMA
- Support Vector Machine (SVM)
- Logistic Regression

ONTARIO DEMAND:

Data consists of electricity demand for entire Ontario for the year 2017.
The Forecast RMSE values of the 5 fitted models are-:

- ➢ HOLTS : 3.12
- ➢ ETS : 3.14
- ➢ ARIMA : 3.75
- ➢ SVM : 1.76
- ➢ Logistic Regression : 1.80

Also plotted qqplot of the fitted models (see appendix) and all the errors are found to be normally distributed. Out the above models SVM has the least RMSE value hence SVM is found to be the best fit.

TORONTO:

Data consists of electricity demand for Toronto for the year 2017.
The Forecast RMSE values of the 5 fitted models are-:

- ➢ HOLTS : 3.198
- ➢ ETS : 3.22
- ➢ ARIMA : 2.19
- ➢ SVM : 1.59
- ➢ Logistic Regression : 1.64

Also plotted qqplot of the fitted models (see appendix) and all the errors are found to be normally distributed. Out the above models SVM has the least RMSE value hence SVM is found to be the best fit.

OTTAWA:

Data consists of electricity demand for Ottawa for the year 2017.
The Forecast RMSE values of the 5 fitted models are-:

- ➢ HOLTS : 2.48
- ➢ ETS : 2.45

- ARIMA : 1.07
- SVM : 1.93
- Logistic Regression : 1.70

Also plotted qqplot of the fitted models (see appendix) and all the errors are found to be normally distributed. Out the above models ARIMA has the least RMSE value hence ARIMA is found to be the best fit.

NIAGARA:

Data consists of electricity demand for Niagara for the year 2017.
The Forecast RMSE values of the 5 fitted models are-:

- HOLTS : 0.935
- ETS : 0.77
- ARIMA : 0.719
- SVM : 0.579
- Logistic Regression : 0.601

Also plotted qqplot of the fitted models (see appendix) and all the errors are found to be normally distributed. Out the above models SVM has the least RMSE value hence SVM is found to be the best fit.

BRUCE:

Data consists of electricity demand for Bruce for the year 2017.
The Forecast RMSE values of the 5 fitted models are-:

- HOLTS :1.062
- ETS : 1.053
- ARIMA : 0.9746
- SVM : 1.002
- Logistic Regression : 0.8908

Also plotted qqplot of the fitted models (see appendix) and all the errors are found to be normally distributed. Out the above models Logistic Regression has the least RMSE value hence Logistic Regression is found to be the best fit.

CONCLUSION

This report mainly concentrate on forecast results of Ontario Daily Electricity Demand for 15days from Jan 1st 2018 to Jan 15th 2018. From the analysis of the residuals Total Ontario demand forecasts can be completely reliable whereas Bruce forecast cannot be completely reliable even though forecast errors are very less, as residuals doesn’t satisfy basic assumption i.e. not normally distributed. As a final note forecasts on Ontario demand, Toronto, Niagara and Ottawa are reliable and models developed on them can be used for forecasting.

REFERENCES

https://www.researchgate.net.

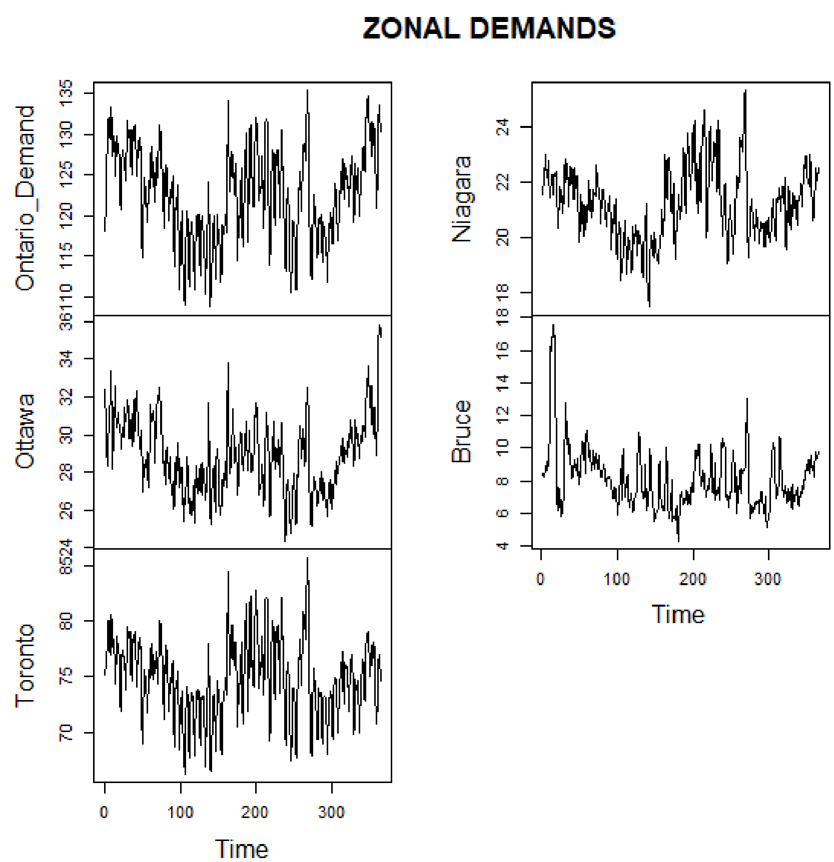
IESO. (n.d.). *http://www.ieso.ca/corporate-ieso/media/year-end-data*.

http://www.ieso.ca/en/power-data/data-directory.

https://www.r-bloggers.com/machine-learning-using-support-vector-machines/.

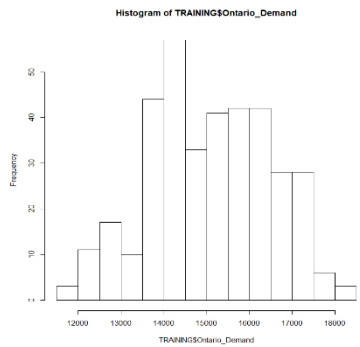
APPENDIX

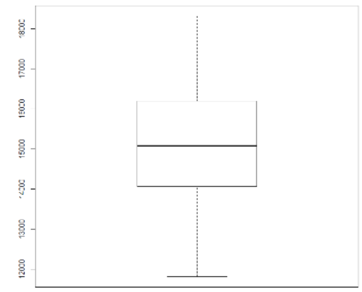
1) Time series plot of zones: Ontario demand, Ottawa, Toronto, Niagara, Bruce



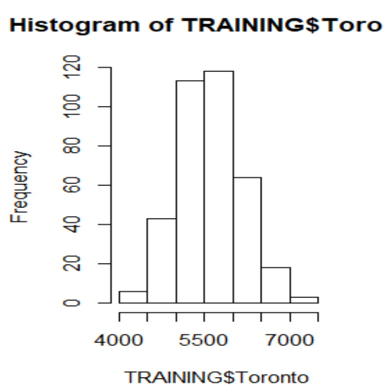
2) Outliers (Histogram and box plots)

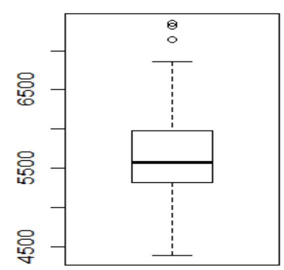
a) Ontario



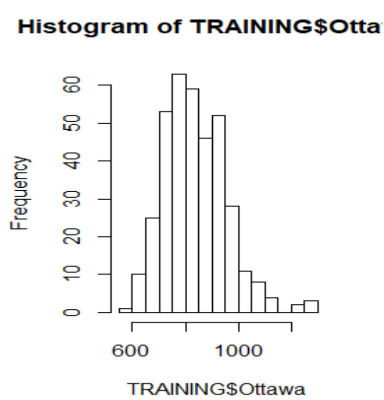


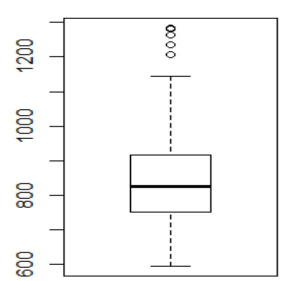
b) Toronto



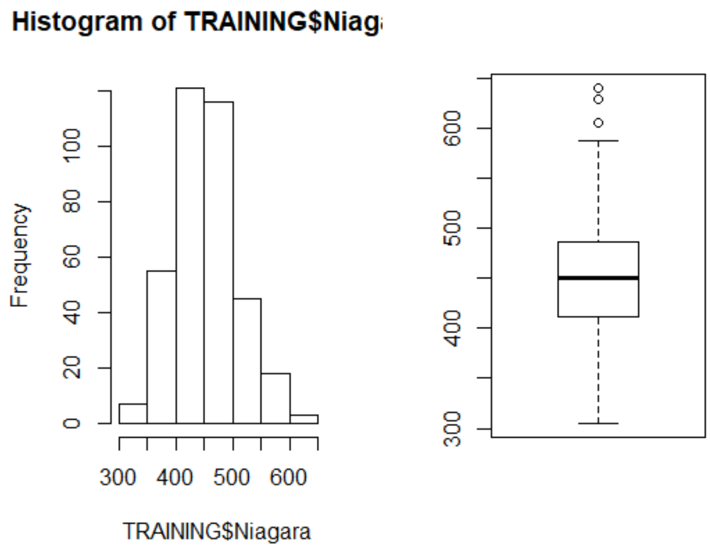


c) Ottawa

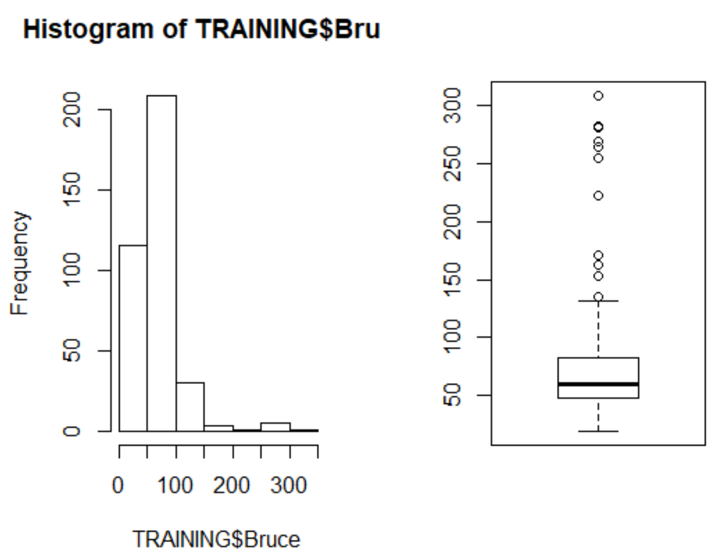




d) Niagara

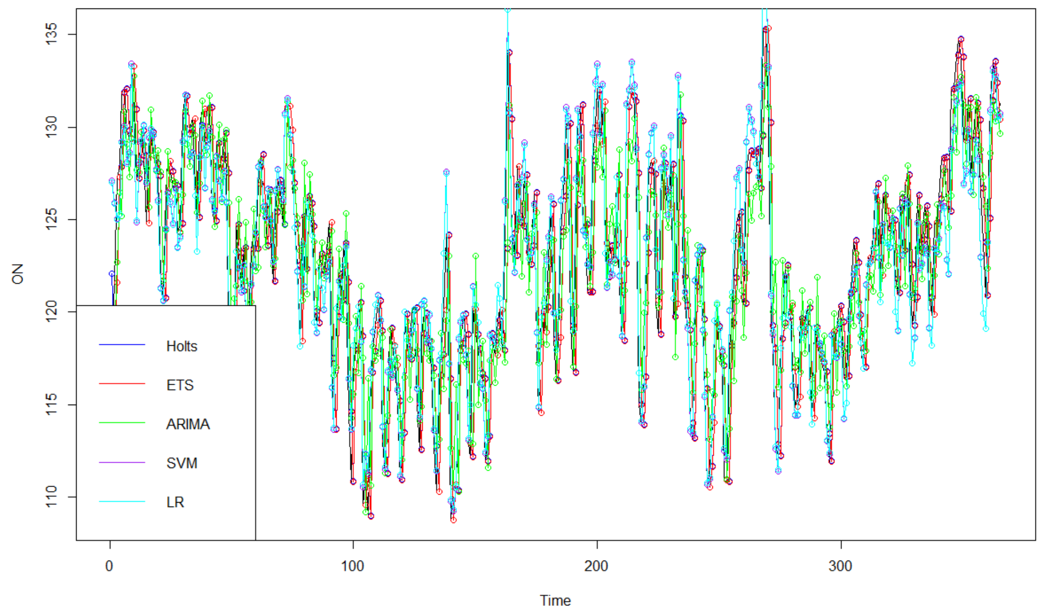


e) Bruce

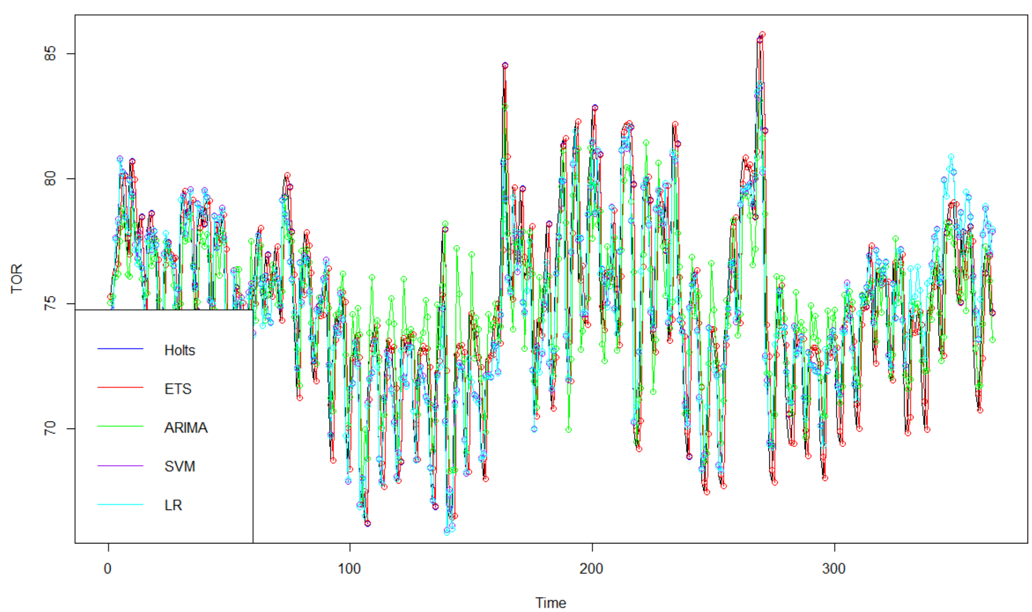


3) Models

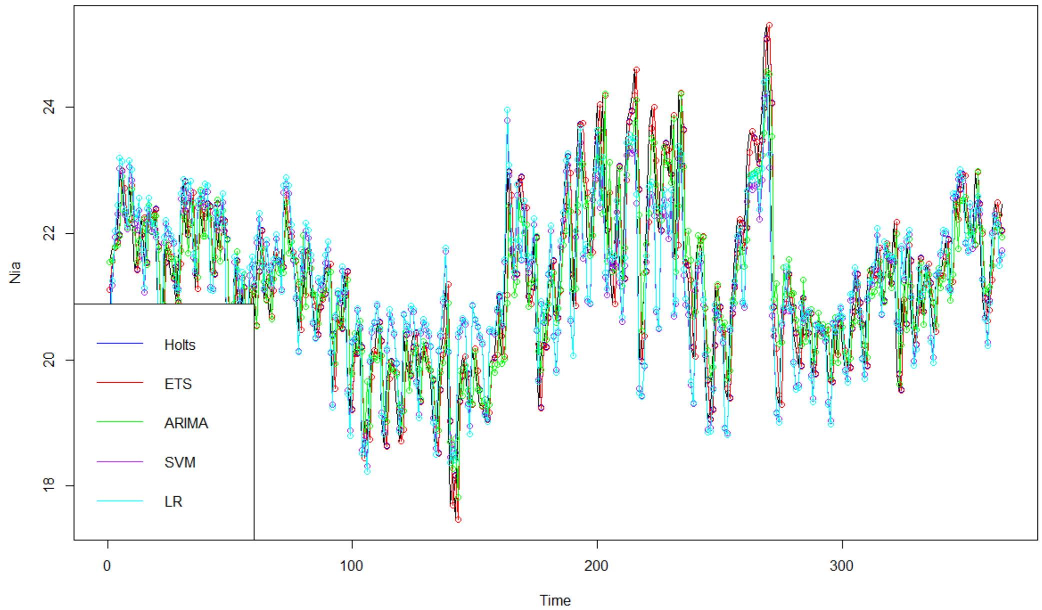
a) Ontario Deamnd



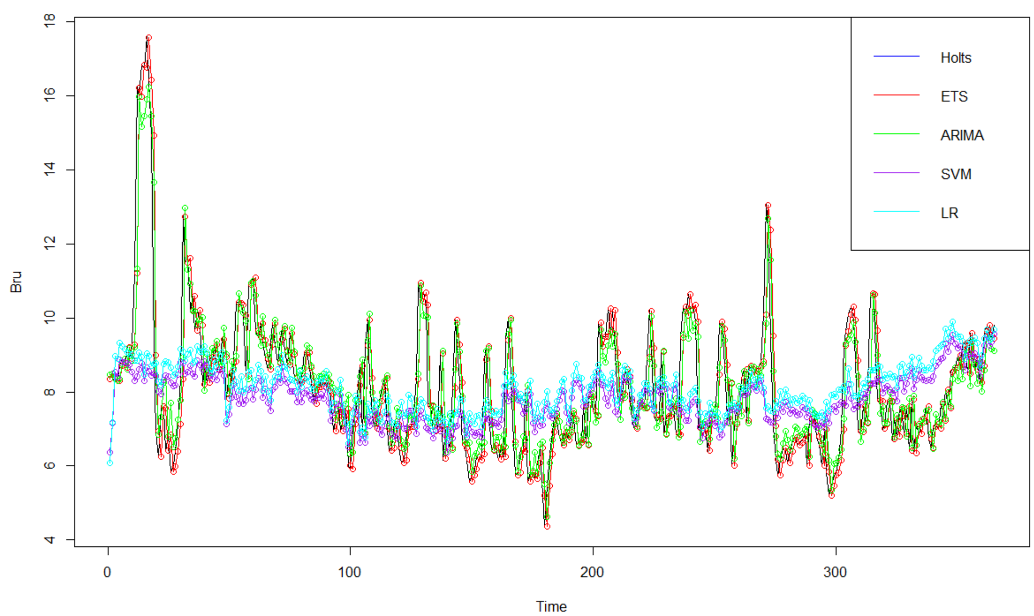
b) Toronto



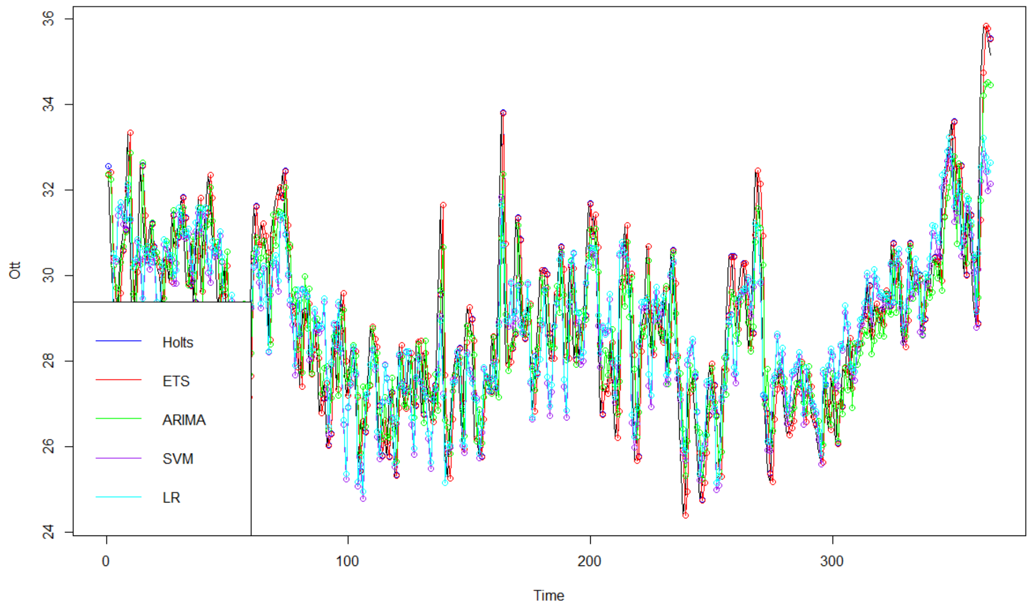
c) Ottawa



d) Niagara

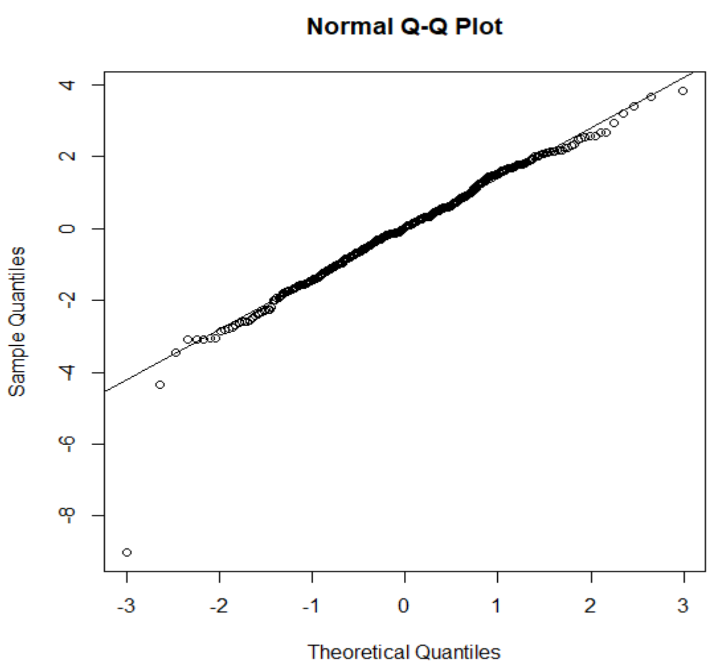


e) Bruce

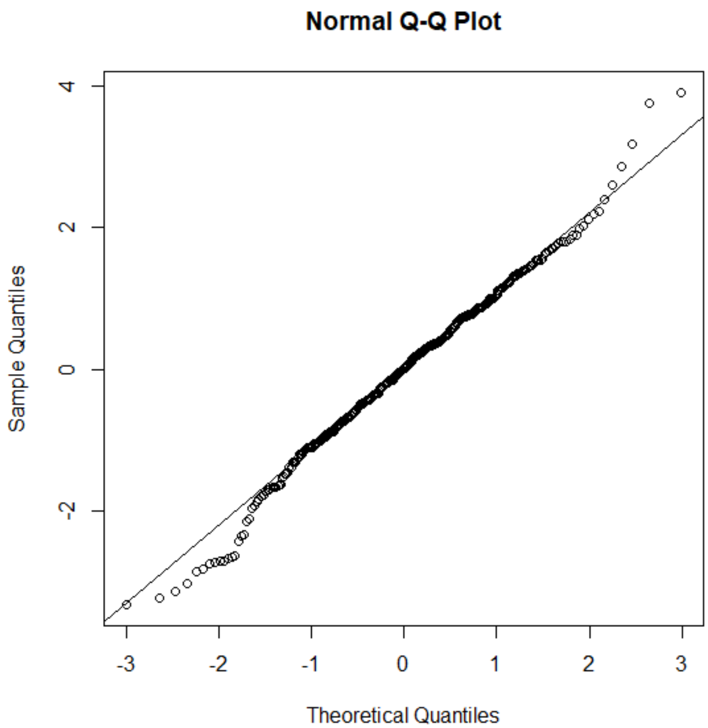


4) QQ Norm plots

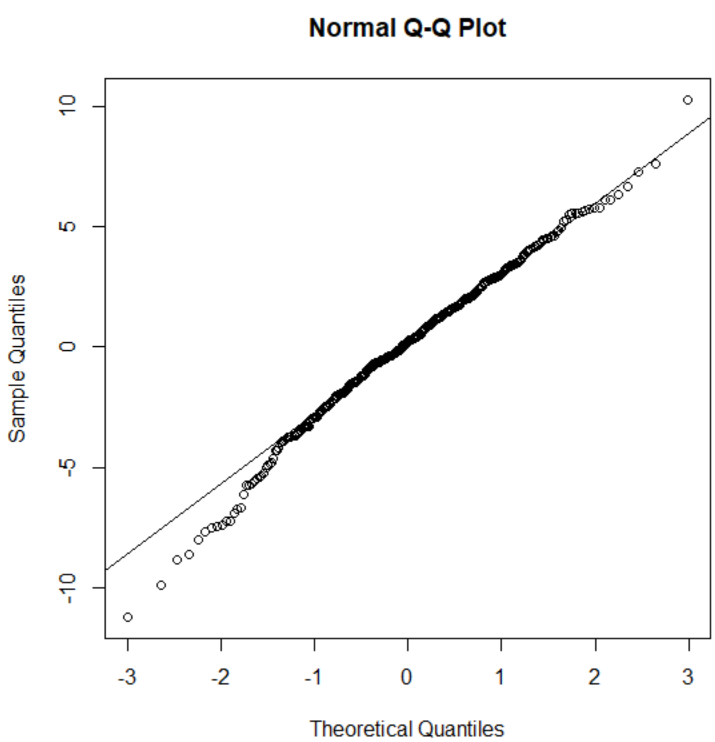
a) Ontario Demand



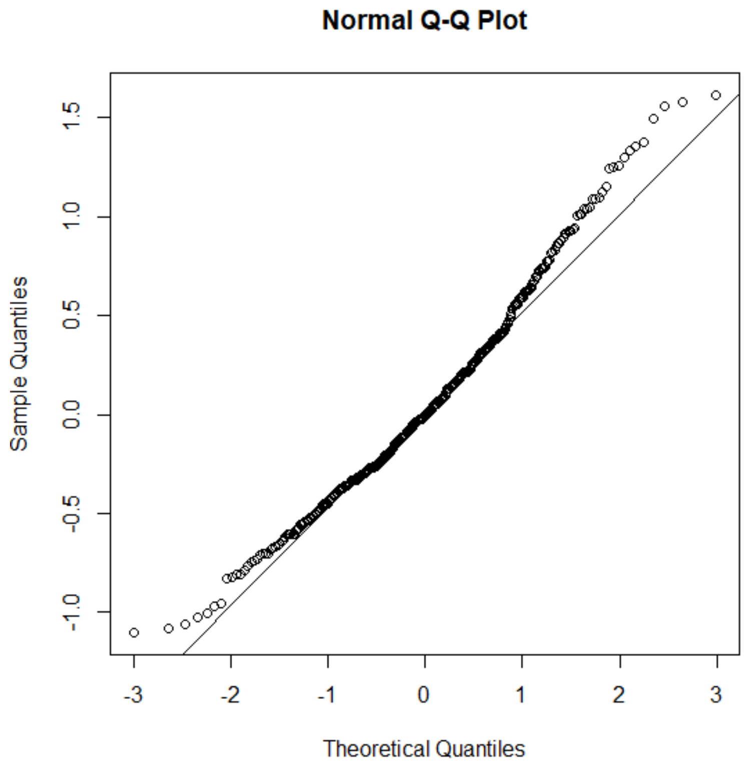
b) Toronto



c) Ottawa



d) Niagara



e) Bruce