

KADIR HAS UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING



DETERMINATION OF TIME-OF-USE PRICES IN ELECTRICITY MARKETS
USING CLUSTERING ANALYSES

GRADUATE THESIS

MOHSAN HUSSAIN

September, 2016

Mohsan Hussain

M.S. Thesis

2016

DETERMINATION OF TIME-OF-USE PRICES IN ELECTRICITY MARKETS
USING CLUSTERING ANALYSES

MOHSAN HUSSAIN

Submitted to the Graduate School of Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science
in
INDUSTRIAL ENGINEERING

KADIR HAS UNIVERSITY

September, 2016

KADIR HAS UNIVERSITY GRADUATE SCHOOL OF SCIENCE AND
ENGINEERING

DETERMINATION OF TIME-OF-USE PRICES IN ELECTRICITY MARKETS
USING CLUSTERING ANALYSES


MOHSAN HUSSAIN

APPROVED BY:

Assist. Prof. Emre ÇELEBI (Advisor)

Assoc. Prof. Ahmet Deniz YÜCEKAYA (Jury member)

Assist. Prof. Tamer DAĞ (Jury member)



APPROVAL DATE: 24 September 2016

“I, MOHSAN HUSSAIN, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.”

A handwritten signature in blue ink that reads "Mohsain Hussain". The signature is written in a cursive style with a large initial 'M' and a period at the end. A horizontal line is drawn underneath the signature.

MOHSAN HUSSAIN

DETERMINATION OF TIME-OF-USE PRICES IN ELECTRICITY MARKETS USING CLUSTERING ANALYSES

ABSTRACT

In this thesis, a clustering analysis to determine the blocks (clusters) of hours for time-of-use (TOU) pricing scheme is proposed and different clustering algorithms are compared using different measures, i.e., change in overall revenue, mean absolute percent error and adjusted coefficient of determination (R^2) from multiple linear regression analyses. Hourly electricity price and demand (load) data for two seasons (winter and summer) from Pennsylvania-New Jersey-Maryland (PJM) wholesale electricity market for 2014-2015 is used and based on detailed descriptive analyses and observations, three blocks of hours (off-peak, mid-peak and on-peak) are presented. In R software, two clustering algorithms (agglomerative hierarchical and k-means) are employed and several clusters for summer and winter weekday hours are formed. The average of the hourly electricity prices in the same cluster for off-peak, mid-peak and on-peak hours determines the TOU pricing scheme (hours in each cluster and prices for each clusters). These prices are compared to real-time pricing (RTP) rates in terms of change in overall revenue collected (price*load) and mean absolute percent error with respect to RTP rates.

Finally, in order to measure the significance of the TOU price and the demand relationship, multiple linear regression analyses are performed. In the regression models, dependent variable is the TOU price (or logarithm of it) and independent variables are the average load (or logarithm of it) of the TOU block of hours, lagged TOU price and lagged TOU average load as well as categorical variables for off-peak, mid-peak and on-peak hours for each TOU pricing scheme. Using Minitab software, different regression models are built and adjusted R^2 , significance of regression coefficients and the significance of the overall model are computed. The significant models (with 95% confidence) are reported and the TOU clusters with higher adjusted R^2 values are determined. Moreover, in order to measure the autocorrelation effect, Durbin-Watson statistics for each significant regression model are calculated and positive correlation among dependent and independent variables are reported. These analyses can be used by electricity market retailers, distribution

companies as well as regulatory bodies in determining TOU time blocks (clusters) and prices.

Keywords: TOU, Pricing Scheme, Clustering Algorithm, Multiple linear regression analysis, PJM wholesale market

ELEKTRİK PİYASALARINDA ÇOK ZAMANLI FİYATLARIN KÜMELEME ANALİZLERİ KULLANILARAK BELİRLENMESİ

ÖZET

Bu çalışmada, çok zamanlı (ÇZ) fiyatlandırma için saat öbeklerinin (kümelemelerinin) belirlenmesine yönelik bir kümeleme analizi önerilmiş ve farklı kümeleme algoritmaları farklı ölçülerle (çoklu doğrusal regresyon analizindeki düzeltilmiş belirlilik katsayısı $-R^2$, toplam gelirdeki değişim ve ortalama mutlak yüzde hata gibi) karşılaştırılmıştır. 2014-2015 yılları için Pennsylvania-New Jersey-Maryland (PJM) toptan elektrik piyasasından alınan mevsimlik (kış ve yaz) saatlik elektrik fiyat ve talep (yük) verileri kullanılmış ve detaylı açıklayıcı istatistiksel analiz ve gözlemlere dayanarak üç saat öbeği (gece, gündüz ve puant) sunulmuştur. Daha sonra, R yazılımında iki farklı kümeleme algoritması (hiyerarşik yığılmacı ve k-ortalamlar) kullanılmış ve yaz/kış işgünleri için kümelemeler oluşturulmuştur. Gece, gündüz ve puant saatler için aynı kümelemedeki saatlik elektrik fiyatları ortalamaları ile çok zamanlı fiyatlandırma planı (her kümelemedeki saatler ve her kümeleme için fiyatlar) belirlenmiştir. Bu fiyatlar, gerçek zamanlı fiyatlar (GZF) ile toplanan toplam gelir (fiyat*yük) ve GZF'ye göre ortalama mutlak yüzde hata bakımından karşılaştırılmıştır.

Son olarak, ÇZ fiyat ve talep ilişkisinin anlamlılığını ölçmek üzere çoklu doğrusal regresyon analizleri yapılmıştır. Regresyon modellerinde bağımlı değişken ÇZ fiyat (veya logaritması) ve bağımsız değişkenler ÇZ saat öbeğinin ortalama yükleri (veya logaritması), zamanı geciktirilmiş ÇZ fiyat, zamanı geciktirilmiş ÇZ ortalama yük ile her bir ÇZ fiyatlandırma planı için gece, gündüz ve puant saatlerini belirten gölge değişkenlerdir. Minitab yazılımı kullanılarak farklı regresyon modelleri oluşturulmuş ve düzeltilmiş belirlilik katsayısı (düzeltilmiş R^2), regresyon katsayıları anlamlılığı ve modelin anlamlılığı hesaplanmıştır. Anlamlı modeller (%95 güvenle) rapor edilmiş ve yüksek düzeltilmiş R^2 değeri olan ÇZ kümelemeleri belirlenmiştir. Bunun yanı sıra, otokorelasyonun etkisini ölçmek üzere her bir anlamlı model için Durbin-Watson istatistikleri hesaplanarak bağımlı ve bağımsız değişkenler arasındaki pozitif korelasyon rapor edilmiştir. Bu analizler elektrik piyasası perakende satıcıları,

dağıtım şirketleri ve düzenleyici kurumlar tarafından çok zamanlı saat öbeklerinin ve fiyatların belirlenmesi için kullanılabilir.

Anahtar sözcükler: Çok zamanlı fiyatlandırma, kümeleme algoritmaları, çoklu doğrusal regresyon analizi, PJM toptan elektrik piyasası

Acknowledgements

First of all, I would thank my dear ALLAH almighty for helping me come this far and paving the way ahead of me. I would like to express my sincere gratitude to Mr. and Mrs. Azam Hussain for their financial support during my studies.

I would like to express my utmost gratitude and love towards my family, especially my parents Mr. and Mrs. Muhammad Hussain, who have stood by me in difficult times and helped me for whatever I have right now. I also appreciate their patience for all the period I have been away from home.

I would like to thanks to Assist.Prof. Emre Çelebi, my thesis supervisor, for his guidance, help, comments and revisions in improving this thesis. Without his instruction and guidance, this thesis could not be accomplished.

I would also thank my great friend, Mr. Jasim Hummayun for his precious support and presence throughout this difficult time, may you be at the highest of highs. I am exceedingly grateful for all of your contributions.

Table of Contents

| | |
|--|------------|
| ABSTRACT | iii |
| ÖZET | v |
| Acknowledgements | vii |
| List of Figures | x |
| List of Tables..... | xii |
| Chapter 1: Introduction..... | 1 |
| 1.1 Overview of Electricity Markets Around the World | 1 |
| 1.2 Objectives and Contribution of this Study | 3 |
| 1.3 Scope and Outline of the Thesis | 4 |
| Chapter 2: Overview of Electricity Pricing and Demand Response in Electricity Markets | 6 |
| 2.1 Overview of Electricity Pricing | 6 |
| 2.1.1 TOU Pricing Scheme | 6 |
| 2.1.2 RTP Scheme..... | 8 |
| 2.1.3 Flat Rate Pricing Scheme | 10 |
| 2.1.4 Comparison of RTP and TOU Pricing Rates | 12 |
| 2.1.5 Scatter Diagrams for Hourly Electricity Prices | 13 |
| 2.1.6 Boxplot Figures for Seasonal (Winter/Summer) Hourly Electricity Prices | 14 |
| 2.2 Demand Response and its Effect on Electricity Prices..... | 16 |
| 2.2.1 Scatter Diagrams for Hourly Electricity Load for Each Season | 17 |
| 2.2.2 Boxplot Figures for Seasonal (Winter/Summer) Hourly Electricity Loads | 19 |
| 2.3 Price and Demand Relation in PJM Market | 20 |
| 2.4 How TOU Rates are Determined? | 21 |
| Chapter 3: Overview of Clustering Methods..... | 22 |
| 3.1 Hierarchical Clustering..... | 22 |

| | |
|---|----|
| 3.2 K-means Clustering | 25 |
| 3.3 How K is selected? | 28 |
| 3.4 Summary for Winter Clusters..... | 31 |
| 3.5 Summary for Summer Clusters | 33 |
| 3.6 Clustering of Hours Using Hourly Electricity Prices (RTP) | 35 |
| 3.6.1 Results for Winter Season..... | 35 |
| 3.6.2 Results for Summer Season..... | 39 |
| 3.6.3 MAPE and Change in Revenue | 44 |
| 3.7 Clustering of Hours Using Hourly Electricity Demands (Loads) | 45 |
| 3.7.1 Results for Winter Season..... | 46 |
| 3.7.2 Results for Summer Season..... | 50 |
| 3.7.3 MAPE and Change in Revenue | 55 |
| Chapter 4: Multiple Linear Regression Analyses of Clusters..... | 56 |
| 4.1 Overview of Multiple Linear Regression Analyses | 56 |
| 4.2 Results for Multiple Linear Regression Analysis of Clusters..... | 58 |
| 4.2.1 Regression Models for Winter Season | 58 |
| 4.2.2 Regression Models for Summer Season..... | 60 |
| 4.3 Discussion of the Results of Multiple Linear Regression Analyses of Clusters | 62 |
| Chapter 5: Summary and Future Research | 63 |
| 5.1 Summary | 63 |
| 5.2 Future Research..... | 64 |
| References | 65 |
| Appendices | 71 |
| Appendix A..... | 71 |
| Appendix B..... | 78 |
| Appendix C..... | 85 |

List of Figures

| | |
|---|----|
| Figure 1: Scatterplot Prices for Winter Season | 13 |
| Figure 2: Scatterplot Prices for Summer Season..... | 14 |
| Figure 3: Boxplot for Winter Season Prices..... | 15 |
| Figure 4: Boxplot for Summer Season Prices | 15 |
| Figure 5: Scatter Diagram of Loads for Winter Season..... | 18 |
| Figure 6: Scatter Diagram of Loads for Summer Season..... | 18 |
| Figure 7: Boxplot for Winter Season Demands | 19 |
| Figure 8: Boxplot for Summer Season Demands..... | 19 |
| Figure 9: Scree Plot for Winter Season..... | 29 |
| Figure 10: Scree Plot for Summer Season | 30 |
| Figure 11: TOU Prices of W1 | 35 |
| Figure 12: Single Linkage for Winter Prices | 36 |
| Figure 13: Complete Linkage for Winter Prices | 36 |
| Figure 14: Average Linkage for Winter Prices | 37 |
| Figure 15: Median Linkage for Winter Prices | 37 |
| Figure 16: Ward.d Linkage for Winter Prices | 38 |
| Figure 17: Ward.d2 Linkage for Winter Prices | 38 |
| Figure 18: K-means Clustering for Winter Prices..... | 39 |
| Figure 19: TOU Prices for Summer Cluster (S1)..... | 40 |
| Figure 20: Single Linkage for Summer Prices | 40 |
| Figure 21: Complete Linkage for Summer Prices..... | 41 |
| Figure 22: Average Linkage for Summer Prices..... | 41 |
| Figure 23: Median Linkage for Summer Prices | 42 |
| Figure 24: Ward.d Linkage for Summer Prices | 42 |
| Figure 25: Ward.d2 Linkage for Summer Prices | 43 |
| Figure 26: K-means Clustering of Summer Prices..... | 44 |
| Figure 27: Single Linkage for Winter Load | 46 |
| Figure 28: Complete Linkage for Winter Loads | 47 |
| Figure 29: Average Linkage for Winter Loads | 47 |
| Figure 30: Median Linkage for Winter Loads | 47 |
| Figure 31: Ward.d Linkage for Winter Loads | 48 |

Figure 32: Ward.d2 Linkage for Winter Loads 48

Figure 33: K-means Clustering for Winter Loads..... 49

Figure 34: Single Linkage for Summer Loads 51

Figure 35: Complete Linkage for Summer Loads 51

Figure 36: Average Linkage for Summer Loads 52

Figure 37: Median Linkage for Summer Loads 52

Figure 38: Ward.d Linkage for Summer Loads 53

Figure 39: Ward.d2 Linkage for Summer Loads 53

Figure 40: K-means Clustering for Summer Loads 54

List of Tables

| | |
|---|----|
| Table 1: Overall Winter Cluster (Prices and Demands)..... | 31 |
| Table 2: Overall Summer Cluster (Prices and Demands) | 33 |
| Table 3: Clustering by Observing the Prices (W1) | 36 |
| Table 4: Clustering by Observing the Prices Alternatively (W2) | 36 |
| Table 5: Clustering by Observing the Summer Prices (S1) | 40 |
| Table 6: Hierarchical Clustering of Summer Prices (S6-S7)..... | 43 |
| Table 7: MAPE and Change in Revenue measures (TOU vs RTP for winter and summer seasons) | 45 |
| Table 8: Clustering by Observing the Winter Demands (W10)..... | 46 |
| Table 9: Hierarchical Clustering of Winter Demands (W12,W16) | 49 |
| Table 10: K-means Clustering of Winter Demands (W17)..... | 50 |
| Table 11: Clustering by Observing Summer Demands (S9)..... | 50 |
| Table 12: K-means Clustering of Summer Demands (S16)..... | 54 |
| Table 13: MAPE and Change in Revenue measures (TOU vs RTP for winter and summer seasons) | 55 |
| Table 14: Correlation between Price and Demand (winter season)..... | 57 |
| Table 15: Correlation between Price and Demand (summer season) | 57 |
| Table 16: Regression Models for Winter Season..... | 58 |
| Table 17: Regression Models for Summer Season | 60 |

Chapter 1: Introduction

1.1 Overview of Electricity Markets around the World

Throughout the world, the electricity (or power) markets, which have long been influenced by upright integrated utilities, are undergoing vital changes in the composition of its regulation and operations. Owing to the current directives, these markets are evolving into a deregulated form in which the market coercion derives power prices. The market is now becoming competitive and the environment of the market is altering the customary centralised-operation procedures. This transformation is frequently known as “deregulation” (or else known as “restructuring”, “privatisation”, “liberalisation”, and “competitiveness”) in electricity markets. Many power corporations have broken apart into various agents that specialize in electricity supply, distribution, and production. These agents are essential to adopt latest management techniques to sustain in this competitive market. Also, it requires latest modelling methods that imitate how power market participants may respond to the changing financial, economic, and regulatory condition in which they operate.

Electricity is a type of a secondary energy source that is primary to our lives today, for many of us, life without electricity would be inconceivable. The productive life normally arrives at a pause when an event such as a blackout happens (all have had such an experience before), everything is stopped and people wait for the electricity to come back before being capable of doing nearly everything. In 1987, the electric competition was introduced earliest in Chile. Not long later, Wales and England and other advanced states followed. In United States (U.S.), the Energy Policy Act (EPA) of 1992 publically motivated a transformation of wholesale electricity competition. Competitive electricity wholesale markets slice up customary upright integrated (monopolistic and mostly state-owned) power entity, detaching electricity generation proprietor from the entities responsible for electricity distribution, transmission, and retail sale. Alternative to generating electricity to only cover the requirements of their electricity consumers, generation proprietary propose their power into a centralised wholesale market, and sold through the process of an auction (Chao and Huntington, 2013). In most cases, nevertheless, customers have very small impact on the design of these power markets. Juries composed of officials from producers, transmission

and distribution corporations, regulators and retailers take all resolutions. There is a clear and comprehensible reason for these circumstances, most customers with the feasible exception of the large ones, do not have the prowess and financial incentives needed to contribute efficiently in the market, due to time-consuming and composite tasks involved. Perhaps, as a result of this lack of representation, many electricity markets do not deal with most customers as an authentic load side that is able to take sensible decisions but simply as a demand that obeys the market outcome under all situations. Wholesale electricity markets have become utterly advanced and have been impartially successful especially over the last ten years. When we compare the retail electricity market with the wholesale market, the retail market has been very less successful, due to lack of customer participation, e.g., the chance to react spot electricity rates (Kirschen, 2003).

The proposed time-of-use (TOU) pricing method for electricity markets is mainly beneficial for regulatory bodies (who can set the retail prices), retailers and distribution companies. The liberalisation process for electricity markets in Europe has progressively focused on energy market amalgamation and associated cross border issue. The liberalisation signs of the federal energy market are now nearer to the long-term goals of a sole European electricity market. The interface point between the federal energy markets needs physical technical positioning and inter-connections. The announcement of the opening of European-Union power sectors the drivers have attained a degree of, institutions; systemize structures, and rules in the federal markets. The opening of markets has proceeded swiftly and in most instances, beyond the minimum requirements. Many customers and specifically substantially large consumers are seeing converging and lower electricity rates. The productivity of energy corporations has increased, while at the same time their profits in the highly competitive markets (e.g., Wales and England) appears to have decreased, reflecting continued redundancy in size (Jamash and Pollitt, 2005).

For numerous reasons, power utilities, and electric network corporations have been restricted to reorganize their functioning's from upright integrated mechanism to unfold market system (Bhattacharya et al., 2012). The philosophy of managing the system has also changed, with the deregulation and reorganising electricity transmission industry. The customary procedure was to supply all demands of

electricity whenever they happen, nevertheless, the current philosophy describes that the system will be more effective if variation in load is kept as low as possible.

Power deregulation has been utterly meagre in scope focusing on retailing and wholesale pricing. Many large generators were granted jurisdictional rights to sell electricity at deregulated rates (Bushnell et al., 2008). Transmission and distribution regions abide regulation but have been restructured to facilitate retail choice and wholesale markets. Many utilities have remained as the ownership of distribution lines but have renounced the day-to-day command of the network system to independent system operators (ISOs). ISO controls the power system and supplies the market contributors with peer access to the system network. It also supervises at the least one interchange organised through which companies can trade electricity (in some countries, e.g. Turkey, there is a separate entity called “Independent Market Operator”, IMO, which only deals with market operations, but not the system issues).

1.2 Objectives and Contribution of this Study

Much of the power demand in the U.S. is strenuous to a handful throughout the year, fundamentally throughout the summer season. This is usually the so-called “peak” load, which puts a considerable strain on the grid and hence, the risk of brownouts and blackouts is very high specifically in a summer season. It also notably raises the year round rates of electricity for customers. Better knowledge and decision making by consumers throughout the on-peak period of the load could substantially decrease this strain on the grid and the risk of electricity outages. Our main objective is to help the regulatory bodies, retailers or distribution companies in setting a TOU pricing scheme, which in turn can help minimize the power consumption during the peak hour periods.

Demand side management (DSM) is the modification of the customer demand for electricity through different techniques such as behavioural change through instructions and financial incentives. Normally, the objective of DSM is to motivate the customer to utilize less electricity throughout the peak hours demand period or to shift the time of electricity usage to off-peak periods such as weekends and night-time. On peak load, management does not certainly reduce total electricity

utilisation, but could be awaited to decrease the requirement for investment in the system network or power generation plants for meeting demand during the peak hour's period.

TOU pricing combined with smart meters indicates that consumer can even better control over their monthly energy consumption bill. By regulating their energy utilisation habits they can benefit when the prices of the electricity are lower. When the consumers know precisely how much energy they are utilising and what the prices are at a given specific hour or period, then the customers obtain the authority to make smarter decisions when utilising electricity. The electricity consumer will also benefit by minimizing their electricity utilisation patterns during the on-peak period through pricing with customer rewards for switching their utilisation pattern from on-peak hours to off-peak hours.

TOU pricing is a significant DSM technique, which could motivate and appreciate the customer's conducts and attain the purpose of decreasing the demand of the peak hour period and shift demand to off-peak periods. A simple TOU pricing scheme can be obtained by dividing the hours of the day into three patterns (namely, off-peak, mid-peak, and on-peak).

1.3 Scope and Outline of the Thesis

In this study, firstly we have analysed the overall hourly electricity prices and loads (demands) from Pennsylvania-New Jersey-Maryland interconnection (PJM), only the weekdays (excluding holidays and weekends) and winter and summer seasons separately. Then, we focused on the price and demand level of these weekdays and observe the behaviour of the prices with respect to their demand level at each hour of the day (hour 1 to hour 24). Based on the analyses in price level at different hours and comparing the price and demand at different hours, we have represented the three patterns of electricity pricing scheme as off-peak, mid-peak, and on-peak (i.e., TOU patterns) and then we converted these price as well as demand levels into three clusters.

This thesis consists of five chapters, listed below:

Chapter 1 provides the introduction, including history and importance of the electricity prices as well as loads and brief details of the previous studies done on the electricity prices in the U.S. competitive electricity markets. It also provides the objectives and contribution of this study and gives the basic structure of this thesis.

Chapter 2 includes the overview of electricity pricing schemes, (i.e., RTP, TOU pricing and fixed (flat) pricing schemes and comparison among these schemes). It also provides a background on demand response in electricity markets. It defines how the demand of electricity affects the price in the competitive market as well as the relationship between the demand and price of the electricity. It also provides information about how TOU blocks of hours are determined for pricing.

Chapter 3 gives the overview of clustering method (hierarchical and k-means). It also provides the detailed literature review of both clustering methods. It explains the single, average, median, complete, centroid, ward, and ward.d2 linkage methods in hierarchical clustering. On the other hand, for k-means clustering the selection of k (number of clusters) and its purpose as well as its drawbacks are presented. It also provides the cluster analyses for electricity prices and demands in two seasons (winter and summer). At the end of this chapter, there is a discussion of the results.

Chapter 4 presents the multiple linear regression analysis of clusters. It gives information about how the regression model are built and it analyses the significant models in details e.g., autocorrelation among the prices and demands (i.e., using the Durbin-Watson test to check positive or negative correlation). Finally, it provides the discussion of the regression models and best models that explains the relationship among dependent and independent variables.

Chapter 5 includes the conclusion of this research and further research areas that can be explored.

Chapter 2: Overview of Electricity Pricing and Demand Response in Electricity Markets

2.1 Overview of Electricity Pricing

Electricity (power) prices (rates) normally reflect the prices to construct, maintain, support and operate power generation plants and the grid (the composite system of electricity distribution and transmission lines). There are numerous principal component effects for electricity price such as fuel price, maintenance of power generation plants as well as distribution and transmission system, weather conditions, and regulations. Power prices are normally highest for commercial and residential customers, because it costs extra to distribute power to them. Industrial customers utilise more electrical energy and can obtain it at inflated voltage levels. Therefore, it is highly effective and less costly to distribute electricity to this type of consumers. The rates of power to industrial consumers are generally close to the wholesale price of electricity. In 2014, the mean retail rate of the electrical energy in the US was 10.45 cents/kWh. The price for the residential consumer was about 12.50 cents/kWh, whereas industrial consumers face an average price 7.01 cents/kWh. On the other hand, for the commercial user the average price was 10.75 cents/kWh, and for the transportation, it was 10.27 cents/kWh. There are several different pricing schemes for electricity pricing, which will be detailed in the next subsection:

- TOU Pricing Scheme
- RTP Scheme
- Flat Rate Pricing Scheme

2.1.1 TOU Pricing Scheme

In the TOU pricing scheme, both periods including the time and prices are fixed for some specific cycle. The widespread classification of TOU prices for electricity consumers are most frequently offered the rates established in advance, but can be changing over the day to apprehend the anticipated impact of varying electricity conditions (Faruqui et al., 2014). In TOU pricing scheme, the hours of the day are categorized into three blocks depending upon the demand and/or price.

- Off-peak
- Mid-peak
- On-Peak

Off-peak is the period of time at which both the prices and load of the electricity are usually low (i.e., due to the low load of electricity the production cost of electricity is also low, especially in base load conventional generators). The off-peak demand can be handled through cheaper power generation plants (e.g., hydro). TOU at peak periods means that the demand is the highest due to the commercial usage like production industries at this time, hence, the price of the electricity is the highest due to high generation rate. Mid-peak prices/loads are in between the off-peak and these on-peak price/load levels.

The distinctive feature of the many TOU rates is that they are established well in prior of the electricity supply/demand cycle and are not regulated to reflect real time conditions. To set up the rates that vary from fixed intervals of time is a technique to approximate the RTP rates. This, as a consequence, would miss the full variability of RTP if such TOU prices are established in advance (Hogan, 2014).

The principal forces at the back of the TOU prices are to consider the time difference in the prices formed at the wholesale market to generate electricity, motivate the consumers to minimize utilization during the peak demand hours or simply to move or change their electricity consumption to off-peak periods, permitting the advantage or other energy manufacturers to perform facilities more effectively.

TOU rates for domestic electricity consumers still have to obtain extensive acceptance. This is because of their capability to save customers, such as above \$1.2 billion every year in California only as claimed by the Energy Power Research Institute (EPRI) (King, 2001). Put in an application to all domestic consumers in the U.S., about 10% minimization of electricity during the peak demand hours would translate to around 20,000 MW (this is around the similar peak load hours for Electric and Pacific Gas, the nation's biggest amalgamate of gas and electric utility).

Time differentiated rates should target at giving an economic inducement to consumers to change their utilization patterns by decreasing peak hour's demand and

moving energy consumption from peak demand hours to off-demand hours. The details on cross-price and rates flexibility of domestic electricity or power demand by TOU are extraordinarily important to evaluate the effectiveness of time differentiated rates policy. From the hypothetical viewpoint, the implementation of the TOU rates should provide increase in public welfare. Time differentiated rates policy can also allow, prices that are adjacent to marginal rates and forefront to welfare. Nevertheless, cost responsiveness in a TOU price infrastructure can be much higher in the long period of time when consumers have the chance to respond to rates increase by buying more efficient apparatus and appliances. In the limited period of time or in short run domestic customers can decrease utilization only by preceding usage or by moving consumption to off-peak time period (Filippini, 2011).

2.1.2 RTP Scheme

In RTP scheme, prices of electricity change on an hourly basis. Normally, the prices of the electricity are fixed and announced only on an hour ahead or day ahead basis. These pricing strategies can be utilized efficiently to influence the consumer usage during the peak demand hours. It considers the wholesale rates (the marginal price of electricity), weather conditions, generator breakdown, and shortage in production due to some uncertain conditions and another incident in an electricity wholesale market. Utilities can exercise different retail charges for different days and for the different hours of the day.

From the producer's point of view, RTP rates decreases the entire payments to producers in the wholesale market, owing to the fact that the reduction in demand during the peak hours when rates of electricity are very high. Moreover, RTP rates can decrease the capability of producers to market power practice. When producers tend to withhold their capacity, retail rates also increase. After that, the profitability of rate increases is balanced by the demand of response (i.e. the rate increase can be counteracted by the reduction in demand and due to this reduction in demand it can also reduce the profitability) and practices of the power market is also discouraged. Ultimately, RTP can decrease the demand for excess capacity by either moving consumption from peak hours to off-peak hours or by minimizing consumption at peak hours of the specific period of time (Borenstein et al., 2002).

In spite of the fact that RTP is a vital conceptual progress over TOU pricing strategy, it normally has evasive benefits. The unreliability and uncertainty of rates switch the price risk to consumers and accordingly, it has failed to entice more customers. Furthermore, the billing costs and incremental metering connected with the application of the RTP rates can demoralize utilities and customers. The examination for Electric Company and Pacific Gas calculated these costs around \$1 billion (Faruqui and Sergici, 2010).

Final customers may not respond to RTP scheme for two principal reasons. Firstly, the price of examining and estimation of hourly rates and persistently optimising the utilization may be immense for small consumers. Secondly, adjusting utilization freely may not be possible for customers due to physical characteristics and configuration of the transmission network, in particular, frequently directed intervention (because of the scarcity and congestion in transmission grid) that can be controlled by the transmission network operator normally takes place at the zones level. This means that discrete customers cannot have their control for being served by the transmission network operator (Joskow and Tirole, 2004).

A compulsory requirement for application of a RTP scheme is the metering technologies, which can calculate the utilization of consumers on a specific period of time or interval basis. These specific time period or interval meters can preserve an individual utilization measurement for every single hour in a billing interval. Consequently, utilities can achieve their billing and metering process and the end user can collect the electricity pricing information. The price of a domestic period or interval meter is particularly six-time the price of a commercial period meter and a traditional domestic meter is about two times the price of the domestic period meter (Waters, 2004).

A further possibility to the usual electricity retail market is the price model such as conservation with the inclined block rates (IBR). In IBR, the marginal cost rises by the entire amount consumed. That cost is further away from the definite threshold in the entire hourly/daily/monthly domestic load, the energy cost will rise to a value which is very high as compared to the normal price of electricity. This produces

incentives for customers to preserve their load of electricity at daytime at different hours in such a way to avoid compensate electricity at a higher price. On the other hand, IBR helps in the balancing of load and decreasing the peak to average ratio (PAR, load factor) as well. Increasing demand needs the impending activation of high priced/unreliable production set, or load reduction request are commonly sent out by the utilities when electricity demand is high enough to put the grid reliability at risk. Usually in a smart grid, the proceed notice for a reduction in load commonly sent from one side to the other through the communication network to every meter asking electricity plan to take suitable action. This will reschedule the section of the forthcoming electricity utilization automatically to some later hours of the day which is leading to the great reduction in the entire load of the electricity distribution (Mohsenian and Leon, 2010).

In 2001, the electricity market in the California demonstrated very high rates for electricity and scarcity threats. The difficulty that comes into view in California and another retail market are innate to the design of market and design response is incisive in a most promising solution. A comprehensive study of RTP efficiency revealed that efficiency gains from RTP are notable or meaningful when the demand flexibility is extremely low. Excluding this, it is divulged that the TOU toll, which is, off-peak and peak toll price of the electricity simply, represents very little efficiency gains when compared to the RTP.

The most often debated subject about RTP is the compulsory or voluntary application of the specified class of end users. Normally, RTP is related with huge numbers of the end-user. In execution of the entire programme RTP is voluntary. Compulsory RTP does not mean inevitably that end users are uncovered to the direct risk of the electricity market. A day-ahead or forward contract is a fine opportunity to decrease the risk and it can also decrease the volatility of price compensate (Faria and Vale, 2011).

2.1.3 Flat Rate Pricing Scheme

The most frequently used retail pricing strategy all around the world is the fixed pricing per kWh of electricity consumed before deregulation and even after

deregulation. Regulated rates for small commercial and residential customers in the U.S. are most commonly fixed for a year (Deweese, 2001).

The fixed electricity rates is determined on the basis of allocated price and an allocated output determined where this allocated price schedule and the price schedule of the marginal supply intersects. The vertical sum output of the allocated price schedule and the revealed load price schedules equal to the marginal price of supply (Clarke, 1971). Fixed pricing strategy mainly focuses on the usage based rather than the time based. Due to this feature of the fixed pricing, it faces a lot of criticism and that is why end-users are billed on their collective consumption over a period. The price of the electricity that the customer pays is time invariant in the fixed pricing scheme. Consequently, customers are retained from the RTP variations that occur in the wholesale market and therefore their monthly bills remains the same by following the effective fixed rate scheme. Due to this stability in price, the retailer is able to encounter their revenue.

The problem today is that customers are unconcerned to electricity prices and have no curiosity of cutting power during the price spikes, because regional or state regulatory bodies seclude them from the volatility of prices. Consequently, the end-users become insensitive to changes in electricity price. Most of the inefficiency and incompleteness of the market are due to this insensitive demand in the wholesale market.

The major problem when dealing with fixed pricing scheme is its unevenness. The electricity production cost is high at peak hours but the distribution price to the end-user remains fixed in all the hours in the day. Plants that generate imperative electricity to fulfil the spike or peak demand is high priced to run as compared to the hydro or nuclear plants that fulfill the mid-peak or off-peak demand.

On the other hand, one may have to wait up until these contract finishes to get lower rates if the market rates fall. The certitude of acquiring fixed prices could charge a customer more money. Utmost weather conditions will not change consumer rates, so it is the great loss for producers to produce electricity during peak hours.

2.1.4 Comparison of RTP and TOU Pricing Rates

In economic hypothesis, efficient pricing is attained when the electricity is priced at the marginal cost of providing the electricity demand of the last increment and this can be achieved only by a perfectly competitive market. Nevertheless, the time varying pricing concept commenced before the studies on peak demand pricing (Çelebi and Fuller, 2007). Peak demand and their cost have been a concern because of the capacity needs for these demands. In peak demand pricing, the marginal cost of electricity is high throughout the periods of the peak demand and is reflected in customer prices, for example, the prices in the TOU periods. In TOU pricing scheme, both specific time periods and the pricing in that period are known as forecasted and are fixed for some time period (e.g., season). While in the case of real-time pricing, normally prices are fixed and vary on an hourly basis and known only on an hourly ahead or day ahead basis. On the other hand, RTP rates reflects the weather condition, wholesale prices, generation shortage, generator failures or other contingency that may happen in an electricity wholesale market.

The theoretical part of the peak demand literature on pricing was not capable of giving empirical answers to the problem and numerous large-scale pilot projects has been organized with TOU pricing scheme above the past three decades. Studies for these pilot projects can be found in (King and Chatterjee, 2003; Aigner, 1984). These studies gathered data that enable econometricians to approximate the factors of electricity demand functions such as self and cross price elasticities, lag elasticities and elasticities of substitution. Few countries even executed TOU pricing scheme on a national level (Chick, 2002). Another study for California (State-wide Pilot Pricing) has demonstrated that industrial and residential and small to medium commercial consumers cut energy utilization in peak demand periods in response to TOU prices (Faruqui and George, 2005).

Hogan (2014) chooses RTP rather than the TOU rates because of larger error in capturing RTP rates due to TOU rates. There is a considerable efficiency difference between even the best TOU plan and RTP. In this study, mean absolute percent error (MAPE) between TOU and RTP rates is employed to measure this efficiency.

2.1.5 Scatter Diagrams for Hourly Electricity Prices

We are using PJM data in our study, because it is easily available and accessible from PJM website (PJM, 2016). PJM started in 1927, established the world's first ongoing power pool. Further utilities joined in 1956, 1965, and 1981. During this time PJM was operated by a section of single member utility.

In this study, only 2014-2015 data for winter season and 2015 data for summer season are used. In winter case, three months (December, January, and February) are selected and data for 12 weeks is obtained. Only weekdays (Monday to Friday) data are employed because the weekends including any holidays are assumed to be off-peak, however, a similar analyses can be performed on weekends and holidays. In the weekday's data for winter season, there are 56 observations (weekdays) of 12 weeks. Similarly summer season composed of months June, July, and August with 14 weeks and 65 observations of weekdays.

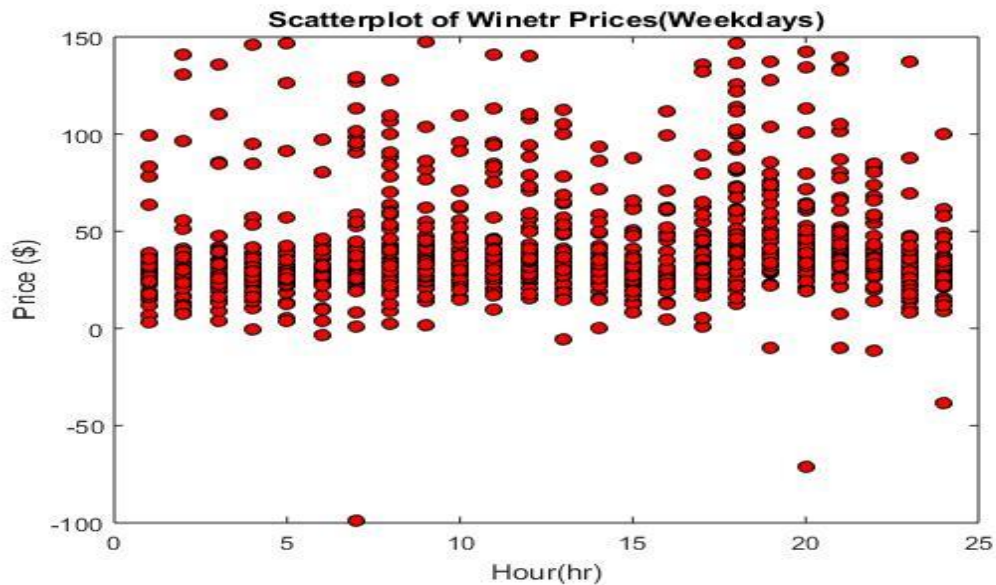


Figure 1: Scatterplot Prices for Winter Season

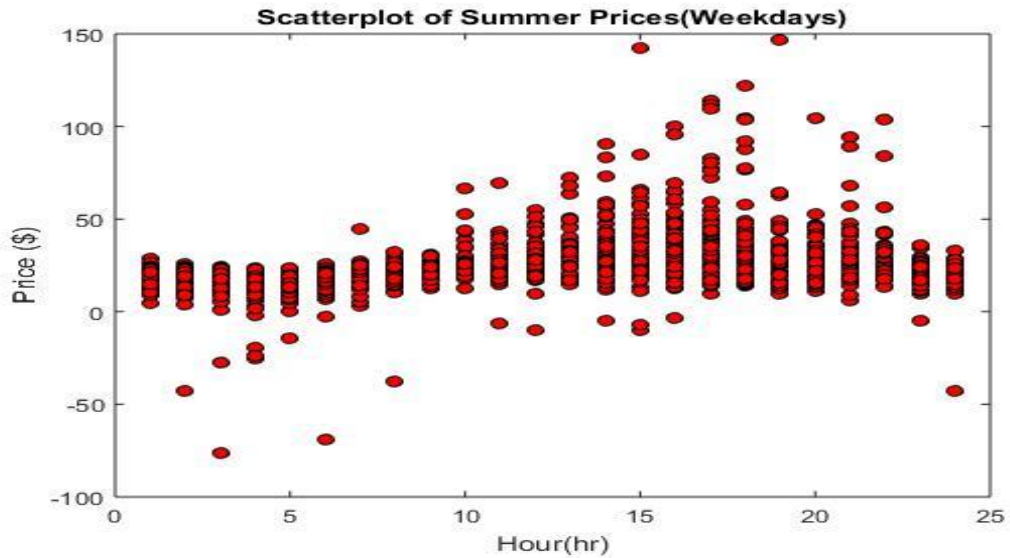


Figure 2: Scatterplot Prices for Summer Season

Figure 1 and Figure 2 are scatterplot figures for all hourly price observations in winter (56 observations) and in summer (65 observations) seasons, respectively. It can be observed that the winter season data is highly diversified. There is more amount of noise in the winter season as compared to the summer season. Also it should be noted that there are more price outliers for the winter season than the summer season.

2.1.6 Boxplot Figures for Seasonal (Winter/Summer) Hourly Electricity Prices

Boxplot figures normally depict the distribution of data based on the five number summary: minimum, first quartile, median, third quartile, and maximum. Also it can identify outliers using a simple criterion (i.e., any data point more than 1.5 times the interquartile range (IQR) below the first quartile or above the third quartile values). We have plotted the boxplot figures of electricity prices for 24-hour (day) to check the sparsity in the dataset of both seasons (winter and summer).

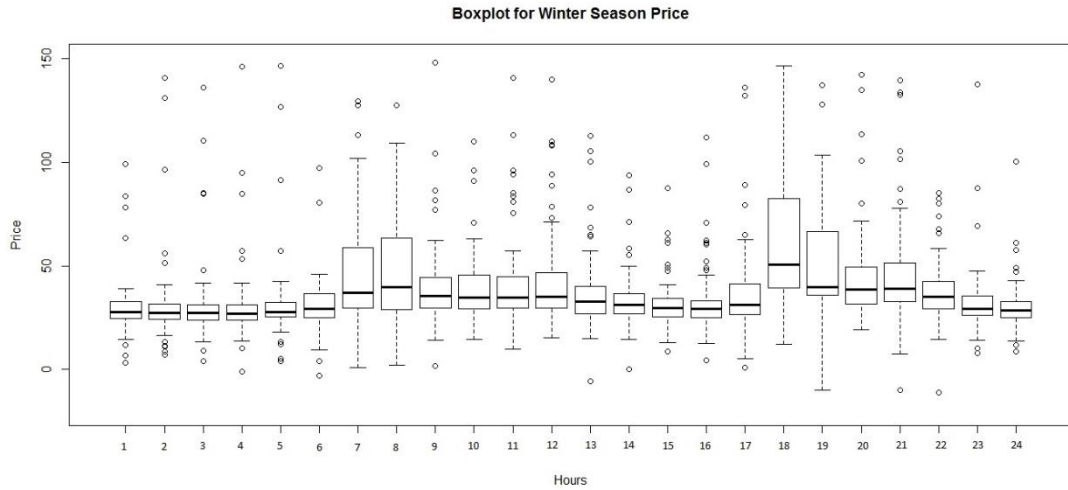


Figure 3: Boxplot for Winter Season Prices

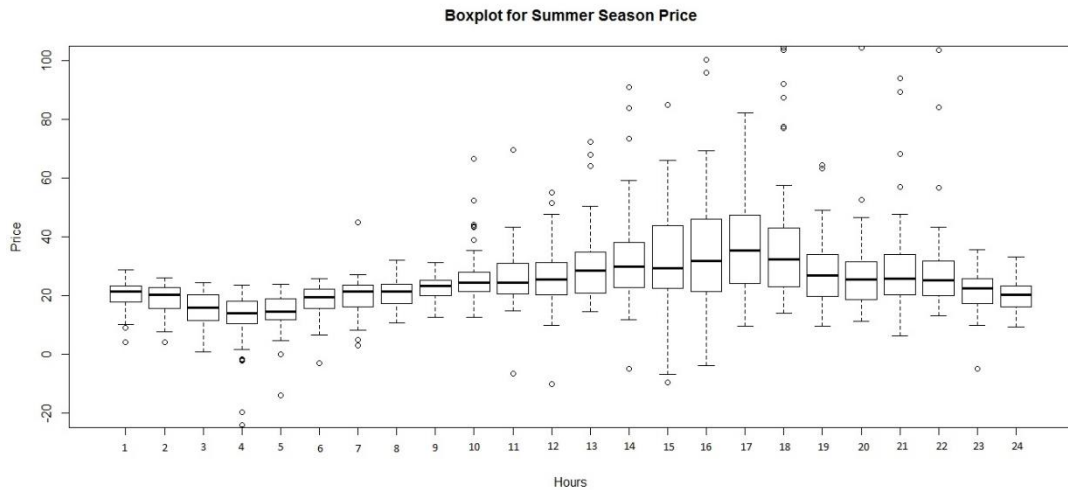


Figure 4: Boxplot for Summer Season Prices

In the above boxplots, the horizontal line (dark black line) in the box is the median. If the horizontal line (dark black line) in the box is situated on the upper side then it is a positively (left) skewed distribution and if it occurs on the down-side of the box then it is negatively (right) skewed distribution. Note that Figure 3 and Figure 4 are zoomed to better reflect the price fluctuations during the day.

The vertical dotted lines (called “whiskers”) on both sides of the box above and below represent the spreading of data or the distribution of the data. As we can see from the Figures 3 and Figure 4 (e.g. by observation), the prices can be clustered into three groups such as off-peak, mid-peak and on-peak. Finally, the (o) represents the

outliers which occurs either above or below of the whisker limits. Outliers can be defined as the abnormal levels of observations that are situated far away from the normal level.

2.2 Demand Response and its Effect on Electricity Prices

Demand response (DR) is about shifts in utilisation period of electricity by end-user from their normal consumption patterns in reaction to variations in the electricity price over time” (Albadi and Saadany, 2008). DR incorporates all the utilization patterns of electricity modifications by end-user that has planned to change consumption timing, the level of immediate demand, or the consumption of total electricity. Eventually, the ultimate aim of the DR is to minimize the peak demand. In order to reduce the peak load or the peak demand, the dynamic pricing performance is calculated by using the elasticity of the demand price which depicts the customer’s sensitivity to the electricity prices.

DR programs have numerous advantages in North America and all around the world (Charles River Associate, 2005). As an example, New York Independent System Operator (NYISO) incentive based programs paid out more than 14,000 programme participants about US\$ 7.2 million in incentive for freeing 700 MW peak of off-peak load in the summer season of 2003. Presently, one-third of its customers are on the TOU pricing scheme.

Shifting electricity load and exerting less expensive electricity to the system, the customers can take the benefit from DR. Since they incorporate bottleneck relief, enhance reliability, lower volume requirement and system benefits from economic load response which should be greater than end-use customer benefits per unit (Spees and Lave, 2007). Perhaps to concede the end-user customer time to aim and respond without having to invest in automated technology, day-ahead prices have been utilized in closely all associated programmes. Although the price in the day-ahead market is a powerful forecaster of the RTP rate, it cannot interact with unpredicted system circumstances such as power failures or other crisis. Benefits of the system from instantaneous load reduction and load shedding or power failure in emergency conditions can only be accumulated from the active management of load or instant

prices. The prompt response needs automated technology that takes actions in the favour of the end-use customer in response to the broadcasted prices. Providing end-users with knowledge on both day-ahead and RTP rates would enable real time response. Alongside this advantage, DR is adequate to give system reliability, cost minimization, efficiency in the market, risk handling, market power reduction and at the last environmental benefits.

DR markets in the U.S. are usually run by the regional transmission operators (RTOs). These programs normally include subsidies of one class to another (Walawalker et al., 2008). Two types of DR programs are offered by PJM.

- Economic DR Program
- Emergency DR program

In the economic demand response, if the LMP in a specified region is over a trigger point (set at \$75/MWh by PJM) then PJM pays the Location Marginal Price (LMP) to customers. PJM remunerates the end-user the difference between the LMP and the generation and transmission (G&T) unit of the end-user bill, when the LMP is below or equal to \$75/MWh. PJM suggests this economic DR program in both real-time and day-ahead markets. The major difference between non-compliance there is no restriction in the real-time market while effective bidding represents a responsibility to reduce load into the day-ahead DR market program.

The emergency DR programme, is a discretionary programme for legitimacy that offers electricity payments to end-user that curtail load during the emergency in the system. The payments that are offered to the customer are above of \$500/MWh or regional LMP for that hour. There is no punishment in this programme for non-compliance because this programme is hardly used by PJM (less than twice a year).

2.2.1 Scatter Diagrams for Hourly Electricity Load for Each Season

In this section, the hourly load data for summer and winter seasons can be observed from the following scatter diagrams.

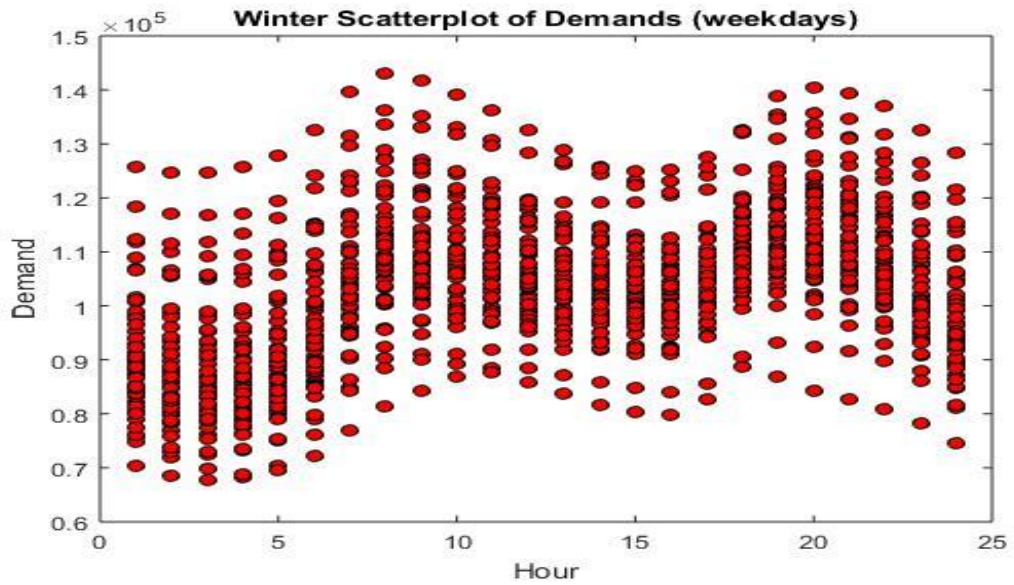


Figure 5: Scatter Diagram of Loads for Winter Season

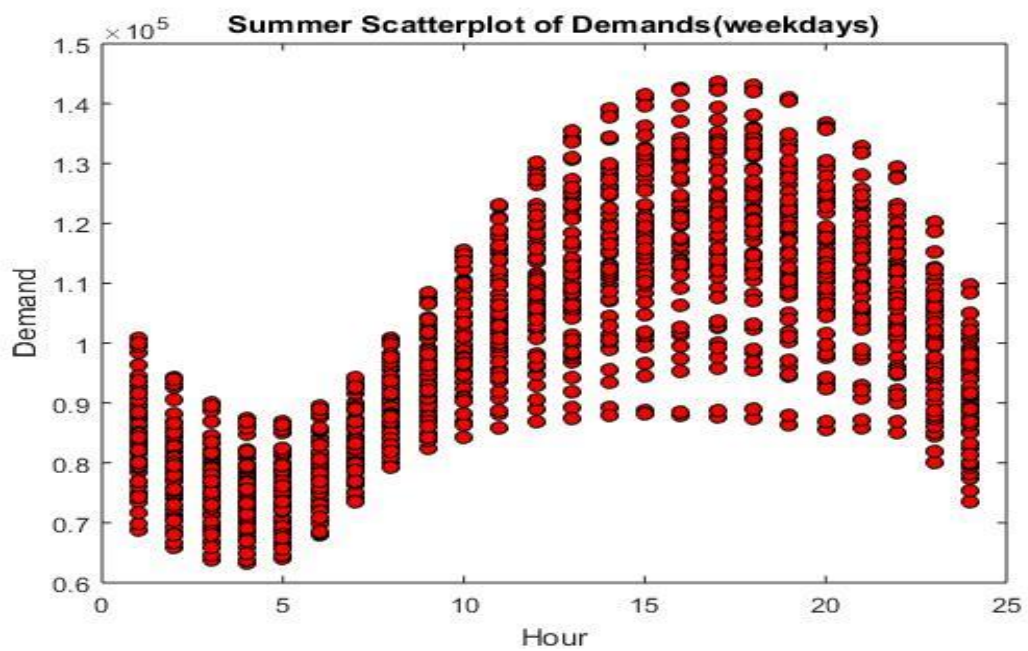


Figure 6: Scatter Diagram of Loads for Summer Season

Figure 5 and Figure 6 are scatter diagrams for all the observations in winter (56 days) and summer (65 days) seasons, respectively. It can be observed that the winter season load is highly diversified. There is more amount of variation in the winter season as compared to the summer season. Also it should be noted that there are more outliers for the winter season than the summer season.

2.2.2 Boxplot Figures for Seasonal (Winter/Summer) Hourly Electricity Loads

The following boxplot figures show the hourly distribution of load data, median values and the outliers of each of the winter and the summer seasons.

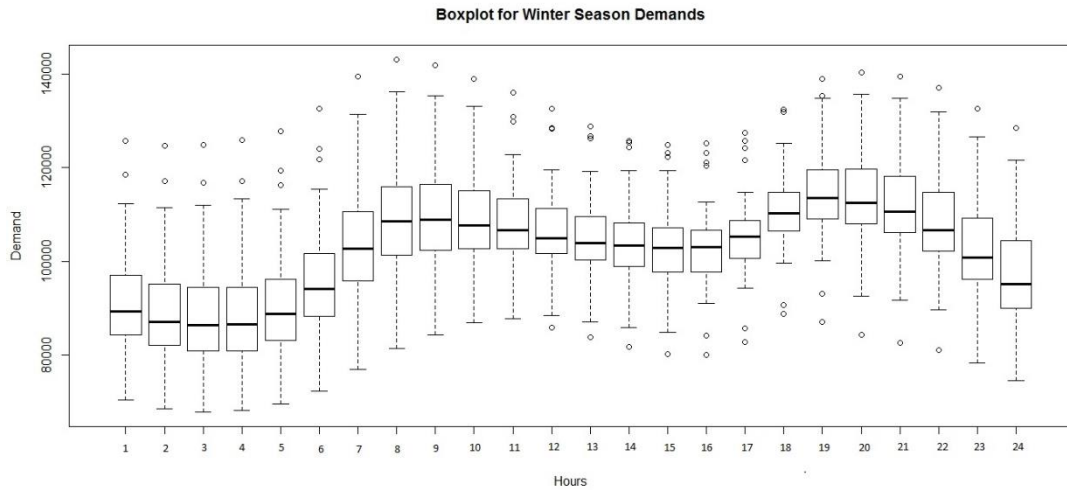


Figure 7: Boxplot for Winter Season Demands

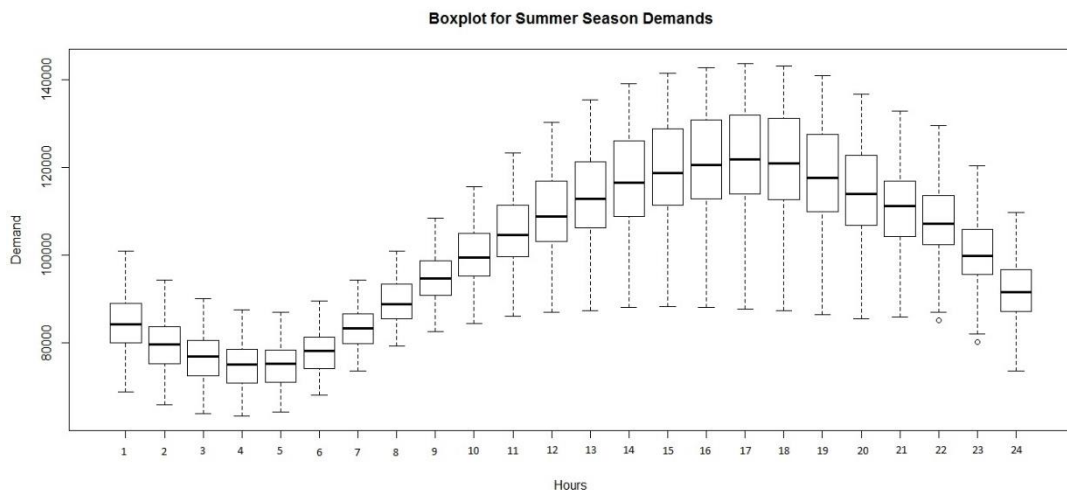


Figure 8: Boxplot for Summer Season Demands

In the above boxplots, the horizontal line (dark black line) in the box is the median. If the horizontal line (dark black line) in the box is situated on the upper side then it is a positively (left) skewed distribution and if it occurs on the down- side of the box then it is negatively (right) skewed distribution. Note that Figure 3 and Figure 4 are zoomed to better reflect the price fluctuations during the day.

The vertical dotted lines (called “whiskers”) on both sides of the box above and below represent the spreading of data or the distribution of the data. As we can see

from the Figure 3 and Figure 4, the loads can be also clustered into three groups such as off-peak, mid-peak and on-peak by observing the figures. Finally, the (o) represents the outliers which occurs either above or below of the whisker limits. Note that there are more outliers in winter season than the summer season.

2.3 Price and Demand Relation in PJM Market

We can examine the relationship between prices and demands by comparing the winter and summer seasons. By looking at the price and load boxplots of both seasons, it can be observed that if the demand of electricity is increased then price of electricity is increased as well in both cases. Next, we can inspect the prices and demands of electricity on an hourly basis for each season.

By observing the Figure 3 and Figure 7, the initial hours from hour 1 to hour 6 in both price and load figures are similar for winter season. Because the price levels from hour 1 to hour 6 is less as compared to the remaining hours in the boxplots we assumed these prices are off-peak. For hour 7 and hour 8, the prices of the electricity and the demand of the electricity are also at its peak. On the other hand, from hour 18 to hour 21 the prices of the electricity is very high and when we observe the load figures the demand of the electricity is also at its peak. Therefore, it can be observed that these hours (h6, h7, and h18 to h21) are on-peak hours. The rest of the hours of the electricity prices are in between the off-peak and on-peak period and it is similar for the demand of the electricity in these periods, which can be the mid-peak period for winter season.

Secondly, we compare the price and load boxplots for the summer season. By looking at the Figure 4 and Figure 8, the prices of the electricity are decreased first from hour 2 to hour 5, then it continuously increased from hour 6 to hour 18 and after this hour, it starts decreasing again from hour 19 till the end of hour 24. The load Figure 8 shows a similar behaviour. The next step is to group the electricity prices by looking at the hours in these boxplots. By focusing on this pattern, we divide the prices in the boxplot into three groups such as off-peak, mid-peak and on-peak. The off-peak prices in summer season are from hour 1 to hour 9 and hour 23

and hour 24. The mid-peak prices are from hour 10 to hour 12 and from hour 19 to hour 22. The on-peak period in summer season is from hour 13 to hour 18.

2.4 How TOU Rates are Determined?

Generally, TOU blocks are determined in a market by the observation of load only. Then, TOU rates are chosen accordingly, but not based on the price levels, mostly to ensure cost recovery and load reduction. Initially, we made a boxplot to check the relationship between the price and demand levels in the PJM market. After finding the results, we classified the hours of the day into three groups (off-peak, mid-peak, and on-peak) where the whole data is clustered using the three groups by using several clustering methods (namely hierarchical and k-means methods) for hours of the day (i.e., TOU) and compare them using multiple linear regression analyses (i.e., the relation of price demand (load) and blocks of hours is better reflected) and also measure the revenue change (compared to RTP) as well as percent error of TOU prices (i.e., compared to the RTP rate using mean absolute percent error or MAPE). Then we select the best overall TOU prices for different seasons.

Chapter 3: Overview of Clustering Methods

3.1 Hierarchical Clustering

Hierarchical clustering is a method that seeks to develop a hierarchy of clusters. Some of the literature on hierarchical clustering is reviewed below.

The general idea of the clustering of data is the method of recognizing the cluster or natural grouping within the multidimensional data depending upon the similarity measure between data, such as Euclidean distance. In hierarchical clustering, the algorithm is used to create a cluster tree (dendrogram) by using merge techniques or splitting heuristics. Cluster tree can be defined as “a cluster tree depicting a sequence with each cluster being a split up of the data set” (Leung et al., 2000). Hierarchical algorithms that are used for the partition to create the cluster tree are called divisive algorithm. Similarly, the most frequent algorithms that are used for merging to create the dendrogram are called agglomerative. According to Leung et al. (2000), this clustering algorithm is not sensitive to initiation, vigorous in the presence of the noise and creates clustering that is alike to that can be recognized by human eyes (e.g., by observation).

The technique of establishing hierarchical clustering of a mutually exclusive subsidiary, every one of which has components that are maximally identical with respect to stated attributes, is recommended for use in large scale studies, especially when a detailed optimal solution for an identified number of groupings is not practical. By repeating this procedure until only a single group remains, the complete hierarchical shape and a quantitative approximation of the loss with every phase in the grouping of the data can be obtained (Ward, 1963).

A dendrogram tree illustrates the hierarchical associations of the clusters explained by the analysis. There are variety of criteria that have been practiced to explain the heights of nodes joining the dendrogram clusters for incremental cluster analysis of sum of squares, including (1) rise in dispersion at every phase (Gordon, 1981), (2) entire dispersion at every phase (Murtagh, 1983; Anderberg, 1971), (3) dispersion within cluster of discrete clusters (Pielou, 1984), (4) average dispersion with-in cluster of discrete clusters (Birks et al., 1975). All of these scales somewhat gives

different details about the analysis. The dendrogram or tree diagrams are subject to reversals with scales (1) and (4). The tree diagram of (3) and (4) give details about the separate clusters. The scale (3) dispersion with-in cluster is extremely dependent on the size of the cluster. Entire dispersion with-in the cluster (2) does not illustrate the incremental formation of the cluster and subject to reversals. The connecting node of every merger raises the connecting nodes of all former mergers. Normally, the tree diagrams of (2) and (4) give the most practical information.

Hierarchical clustering procedures are based on the utilisation of the proximity matrix specifying the similarity between the data points of each pair to be clustered. The end result of this technique is a dendrogram which represents the nested grouping of patterns and similarity levels at which grouping change (Jain et al., 1999).

On the other hand, divisive hierarchical clustering algorithms initiate with the entire pattern allocated to a single cluster. Then, abpartitioning technique is applied to that single cluster at every level until every cluster comprises one pattern or uniform pattern. After that, the most two alike clusters are merged with each other. Therefore, this step is replicated until the entire patterns are allocated to a single cluster (Turi, 2001). There are numerous agglomerative hierarchical clustering algorithms that were suggested in many studies which contrast with each other in a way that most two alike clusters are estimated. But there are two most famous agglomerative hierarchical clustering algorithms, complete linkage (Anderberg, 1973) and single linkage (Sneath and Sokal, 1973). In our study, we also used the median, average, ward.d and ward.d2 algorithms for clustering the prices and loads.

In the beginning of the agglomerative hierarchical clustering procedure, each and every component is in a cluster of its own. These specified clusters are then consecutively merged into big clusters until the entire component end-up being in the similar cluster. The two clusters are divided by the smallest distance are combined at every stage. The meaning of the smallest distance is what we differentiate between the contrasting agglomerative hierarchical clustering. Single linkage clustering in statistics is one of the various techniques which we are using for agglomerative hierarchical clustering. The distance between the two clusters is directed by a single

element pair, specifically, that component pair (one in every cluster) that are nearest to each other. The smallest distance of these links that continues to exist at each stage leads to the fusion of the pair of clusters whose components are involved. This procedure is also called nearest neighbour clustering. The outcome of the clustering can be seen as a dendrogram, which depicts the succession of cluster fusion and the time interval (distance) at which every fusion was happening (Sneath and Sokal, 1973).

In dendrogram, the horizontal axis appears for the variance and distance between the clusters. On the other hand, the vertical axis illustrates the clusters and objects. The dendrogram is impartially very easy to explain. However, our main attention is to find the similarity between the data and clustering. Each and every connecting fusion (point) of three clusters is depicted on the diagram by cleaving of a horizontal line into a pair of horizontal lines. The split position of the horizontal lines, which are represented by the small erect bar, gives the variance (distance) between the clusters.

On the other hand, complete linkage clustering algorithm combines the clusters whose distance between their most faraway patterns is smallest. Generally, complete linkage clustering algorithm creates dense clusters as compared to single linkage clustering algorithm which creates elongated clusters. Consequently, complete linkage clustering algorithms are normally better than single linkage clustering algorithms (Jain et al., 1999).

In the average linkage agglomerative hierarchical clustering, we define the distance between the pair of clusters to be the mean distance between the dataset in the first cluster and dataset in the second cluster. By following this method, the distance between clusters, at every stage of the procedure, we amalgamate the two clusters that have the shortest mean linkage distance (Huth et al., 1993).

In the centroid linkage, agglomerative hierarchical clustering combines the clusters in such a way that the distance between their centroids are the minimum or we define the distances between the pair of clusters as the distance in-between the two average-vectors of the clusters. At every stage of the procedure, we amalgamate the two

clusters that have the shortest mean linkage distance. One drawback of this method is that when combined with very large clusters the characteristics of the small cluster is lost (Turi, 2001).

The median linkage agglomerative hierarchical clustering is defined as the median of all distances between pairs of objects in order to decide how far they are apart (Ultsch et al., 1995). The median method (Evratt and Brian, 1974; Gordon, 1967) assumes the clusters are of equal size, hence the new group will always be between the two groups being merged.

In the ward linkage agglomerative hierarchical clustering, the technique does not directly explain a measure of the distances between a couple of clusters or points. Generally, it is based on the analysis of variance (ANOVA) technique. At each step, pair of clusters are merged, which gives the smallest increase in the merged sum-of-squared error from the single univariate ANOVAs way that can be done for every variable with classes (groups) as explained by the clusters at that phase of the procedure. The main difference between ward.d and ward.d2 is the distances measurement. The ward.d2 criterion values are “on a scale of distances” whereas the Ward criterion values are “on a scale of distances squared” (Murtagh and Legendre, 2011).

3.2 K-means Clustering

The main objective of k-means clustering is to partition the dataset into k clusters in which every observation is assigned to the cluster with the nearest average. The literature on this clustering algorithm is reviewed below.

Martinez et al. (2007) explain k-means clustering as the process used to create a grouping of the larger data set which represents the intention of behaviour of a system as precise as possible. The k-means clustering technique is applied to larger data set to extract the useful information for electricity time series data. This k-means technique split the whole year data into alike groups of hours depending upon their behavior at different days.

Xu and Wunsch (2005) applied the k-means clustering which plays an essential role in the clustering of data. The examination of the unlabelled data of cluster analysis is either constructed by hierarchical structure or established by the different set of groups according to their similarities and dissimilarities. The main aim of the clustering is to remove the inherent uncertainty within the cluster and among the clusters as well. In each iteration of k-means clustering, the algorithm removes the noise and outliers that are present in the larger data set.

Meshram et al. (2013) discussed k-means algorithms in clustering the identical types of load patterns. Typically, if the number of clusters is less due to the distribution of data, than k-means algorithm enhances the rate of classification. Accurate load forecasting plays a critical role to minimize the generation cost. It is also necessary for the consumer's reliability and also demonstrates that the electricity consumption can be clustered on the basis of the load value and also the load pattern.

Li et al. (2014) provide a study of domestic customers where active DSR can create benefit in terms of minimizing costs of electricity for customers and avoiding the electricity use especially in peak demand periods for the distribution network operators (DNOs). K-means clustering is also used to convert the RTP into a TOU pricing scheme. It is done by taking the average of the prices in each cluster and this average is the price for the corresponding TOU hours.

Hernandez et al. (2012) conducted a literature survey to understand the electricity utilisation patterns that is highly important for the implementation of green trends and the optimization of resources. A real industrial park has been examined by clustering the separate days (working and non-working days) according to their load curves and the consumption behaviour of the electricity on a daily basis. Significant consumption behaviours have been recognized properly by the system with an absolute unsupervised fashion (k-means clustering).

Alsabti et al. (1997) arrange the entire samples of big data or patterns in a k-dimensional tree formation such that one can discover the entire patterns which are nearest to the given specimen efficiently. All the specimens are strong candidates for the nearest prototype at the foundation level. However, the entire prototype is

arranged to their closest cluster by using clustering constraints. He applied the approach that will recalculate the distance at each (or a few) iterations to reduce the distance from the point to the centroid of the cluster.

Azevedo and Vale (2006) proposed a new procedure for price prediction, based on artificial neural network (ANN) and k-means clustering techniques. The principal concept of this method is that it has obtained statistical information from the historical data. To find the patterns (similarities and dissimilarities) through ANN is very difficult, hence, in order to find the patterns they used k-means clustering techniques. K-means clustering technique enables to probe the structure of a group and utilizes every group to train one particular neural network. These enable every neural network to acquire knowledge in a more effective way by the information from the previous data.

Martinez et al. (2011) presented a new method which is based on the similarity of pattern sequence to forecast the behaviour of the time series. In order to find the grouping and labelling of the samples from the data set, k-means clustering techniques are used. Consequently, the projection of the set of data points is given as follows. Firstly, former pattern sequence of the day to be forecasted is extracted. Then, this succession is obtained from the previous data and the prophecy is measured by taking the mean of all the samples instantly after the matched order or sequence. This proposed algorithm provides very competitive results in the New York market and it has been successfully applied to the electricity demand time series and prices.

Räsänen and Kolehmainen (2009) suggested a valuable estimation technique for clustering of time series and its application on concentrating or generation of more authentic electricity load curves, especially in the case of small customers. This approach is very less responsive to faulty value and this technique can also manage time series of distant length. The accomplishment of this approach was calculated using the data for 1035 actual customers. After the application of this approach the three principal advantages are reported as: 1) managing the time series of different length, 2) less responsive to missing value, and 3) ability to minimize the dimensionality of actual time series.

Mahmoudi et al. (2009) provided a clustering method that divides a set of the specific characteristic vector into the cluster. This implementation of clustering the load curves in deregulated and regulated markets was for better rate design, load and price forecasting, system analysis, and distributed resource selection. In this study, they have developed a new characteristic which they calculated the performance of the algorithm of clustering on the basis of faraway dataset implementation on the power system.

Bobric et al. (2009) provided the study of distribution network uncertainty and how to cluster loads of buses for specific time at specific regions. The degree of uncertainty is about the load level of the network, power losses, bus loads, and bus voltage levels. First of all, it is very important to measure the load profile of buses by using the available historical data. Then, using this data, they classified the electricity demand profiles of large scale sets of data. They classified this information into different clusters by using k-means clustering algorithms, which represents all the information or points in clusters that are very much similar to each other and very different when it is compared to the information or points in the other clusters.

3.3 How K is selected?

Determination of k (number of clusters) for a dataset is a problem that occurred repeatedly and it is a difficult issue to resolve in any clustering analyses. The performance of the clustering algorithm may be affected by the choice of k. Consequently, instead of using a one pre-defined value for k, a number of values might be chosen. It is important to have the number of values that are supposed to be reasonably large to reflect the specific characteristics of the data sets (Pham et al., 2005). The k-means clustering algorithm executed in many data analysis and data mining software packages require the analyst to input the k value (number of clusters) to be determined by the user. When k-means clustering is used as a pre-processing tool, the number of clusters is determined by the specific requirements of the main processing algorithm (Hansen and Larsen, 1996).

There are various statistical measurements available for choosing k. These measurements are most frequently applied in the form of merging with probabilistic

clustering techniques. The Bayesian information criterion and Akeikes information criterion (Pelleg and Moore, 2000) is computed on datasets which are built by a group of Gaussian distribution. The measurement applied by Hardy (1996) are based on the supposition that the dataset apt the Poisson-distribution. Monte-Carlo methods, which are related with the null-hypothesis, are utilized for evaluating the results of clustering and also for determining the value of k (Halkidi et al., 2002). Illustration verification is used most extensively due to its clarity and explanation possibilities. Visual examples are often used to illustrate the drawbacks of an algorithm or to present the expected clustering results (Bilmes et al., 1997; Lindeberg, 1994).

In this study, we have clustered our data by the visualization of the scree plot. Scree plots are commonly used in factor analysis and principal component analysis to visually evaluate which factors or components describe most of the variability in the data. We also used scree plots to visually evaluate the sparsity in the data. The best possible pattern in a scree plot is a vertical curve, accompanied by the bend and then a horizontal or flat line.

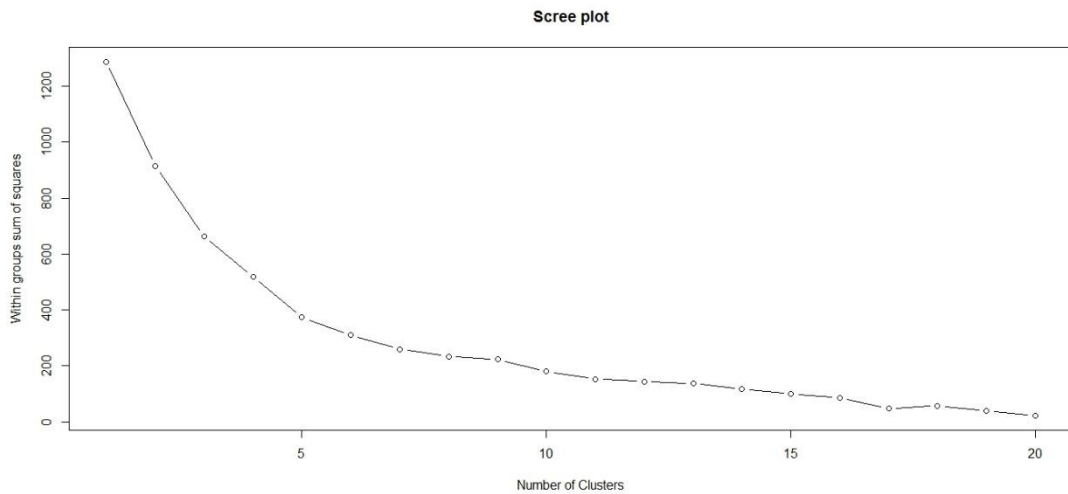


Figure 9: Scree Plot for Winter Season

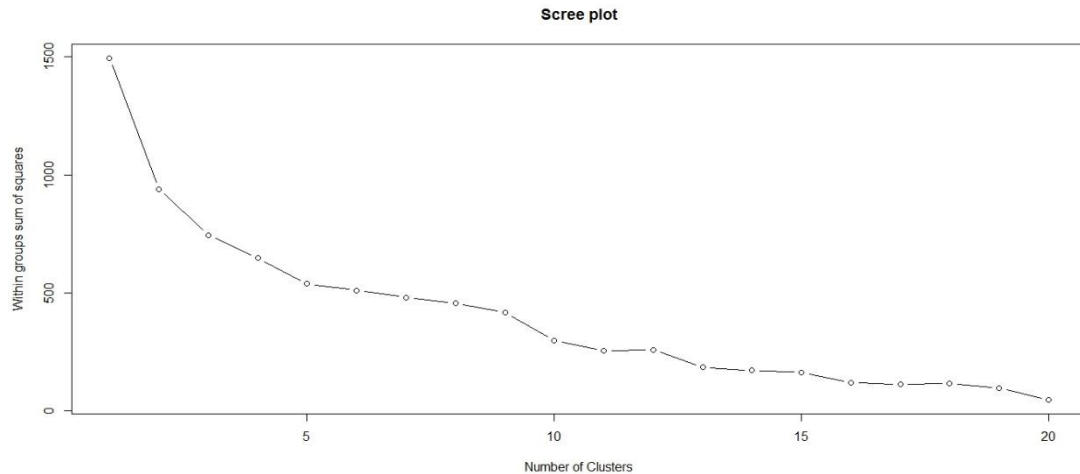


Figure 10: Scree Plot for Summer Season

The above plots show the k clusters (from 1 to 24 to reflect each hour) of the dataset. When we see the diagrams there is a sharp decrease in between the initial three dataset points but after this point, the data set shows very similar behavior and the decrease in the value of within group sum of squares is not significant.

K-means algorithm also has some several drawbacks. To predict the value of k (cluster) is very difficult. It did not work well with the universal cluster. Different beginning partitions can give different ending clusters. It does not work properly with k (in the real dataset) of different density and size. It has to run several times to come over random starting centroids. However, its ease of practice and application in many studies encourage us to use it in this study.

3.4 Summary for Winter Clusters

Table 1 summarizes the overall clusters for the winter season (prices and demands) in this study. It represent the blocks of hours for off-peak, mid-peak, and on-peak periods.

Table 1: Overall Winter Cluster (Prices and Demands)

| Cluster Type | Off-peak | Mid-peak | On-peak | Cluster Name | Method |
|--------------|--------------------------------|------------------|----------------------|----------------------------------|---|
| W1 | (h1-h6) | (h9-h17) | (h7-h8 and h18-h21) | Winter Cluster 1 | By Observation of Winter Prices |
| W2 | (h1-h6 and h23-h24) | (h9-h17) | (h7-h8 and h18-h22) | Winter Cluster 2 | By Observation of Winter Prices (Alternative) |
| W3,W4,W5,W6 | (h1-h6 and h9-h17 and h19-h24) | (h7-h8) | (h18) | Winter Cluster 3,4,5 and 6 | Single, Complete, Average and Median Linkages Hierarchical Clustering of Prices |
| W7,W8 | (h1-h6 and h9-h17 and h23-h24) | (h7-h8) | (h18-h22) | Winter Cluster 7 and Cluster 8 | Ward.d and Ward.d2 Linkages of Hierarchical Clustering of Prices |
| W9 | (h1) | (h2-h11) | (h12-h24) | Winter cluster 9 | K-means Clustering of Winter Prices |
| W10 | (h1-h6 and h24) | (h11-h17 and 23) | (h7-h10 and h18-h22) | Winter Cluster 10 | By Observation of Winter Loads |
| W11,W13 | (h24) | (h1-h6) | (h7-h23) | Winter Cluster 11 and Cluster 13 | Single and Average Linkage Hierarchical Clustering of Loads |

| | | | | | |
|---------|-----------------|------------------------------|----------------------|---------------------------------|--|
| W12,W16 | (h1-h6 and h24) | (h13-h17 and h23) | (h7-h12 and h18-h22) | Winter Cluster 12 and Cluster16 | Complete and Ward.d2 Linkage of Hierarchical Clustering of Loads |
| W14 | (h1-h6) | (h23-h24) | (h7-h22) | Winter Cluster 14 | Median Linkage of Hierarchical Clustering of Loads |
| W15 | (h1-h5) | (h6 and h13-h17 and h23-h24) | (h7-h12 and h18-h22) | Winter Cluster 15 | Ward.d Linkage of Hierarchical Clustering of Loads |
| W17 | (h1-h9) | (h18-h24) | (h10-h18) | Winter Cluster 17 | K-means Clustering of Loads |

There are several clusters for winter season according to clustering of prices and demands, but we have used only those clusters which are significant and practical (i.e., some clustering results have only one or two hours for some clusters). We consider and analyse only those clusters that have at least 5 consecutive hours. The details about each of these significant clusters are given in the subsequent sections.

3.5 Summary for Summer Clusters

Table 2 summarizes the overall clusters for summer season (prices and demands) in this study.

Table 2: Overall Summer Cluster (Prices and Demands)

| Cluster Type | Off-peak | Mid-peak | On-peak | Cluster Name | Method |
|--------------|----------------------|--------------------------------|-----------|--------------------------------|---|
| S1 | (h1-h9 and h23-h24) | (h10-h12 and h19-h22) | (h13-h18) | Summer Cluster 1 | By Observation of Summer Prices |
| S2 | (h1-h15 and h18-h24) | (h16) | (h17) | Summer Cluster 2 | Single Linkage Hierarchical Clustering of Prices |
| S3 | (h3-h5) | (h1-h2 and h6-h12 and h18-h24) | (h13-h17) | Summer Cluster 3 | Complete Linkage of Hierarchical Clustering of Prices |
| S4 | (h1-h12 and h19-h24) | (h13-h16) | (h17-h18) | Summer Cluster 4 | Average Linkage of Hierarchical Clustering of Prices |
| S5 | (h3-h5) | (h1-h2 and h6-h17 and h19-h24) | (h18) | Summer Cluster 5 | Median Linkage of Hierarchical Clustering of Prices |
| S6,S7 | (h1-h7 and h23-h24) | (h8-h12 and h19-h22) | (h13-h18) | Summer Cluster 6 and Cluster 7 | Ward.d and Ward.d2 Linkage of Hierarchical Clustering of Prices |
| S8 | (h4-h5 and h24) | (h1-h3 and h6-16) | (h17-h23) | Summer Cluster 8 | K-means Clustering of Prices |
| S9 | (h1-h8 and h24) | (h9-h12 and h21-h23) | (h13-h20) | Summer Cluster 9 | Clustering by Observation of Demands |

| | | | | | |
|---------|-----------------|-----------------------------|-----------|---------------------------------|--|
| S10 | (h8 and h24) | (h1-h7) | (h9-h23) | Summer Cluster 10 | Single Linkage Hierarchical Clustering of Demands |
| S11 | (h2-h6) | (h1 and h7-h10 and h23-h24) | (h11-h22) | Summer Cluster 11 | Complete Linkage of Hierarchical Clustering of Demands |
| S12 | (h1-h8 and h24) | (h9-h10 and h23) | (h11-h22) | Summer Cluster 12 | Average Linkage of Hierarchical Clustering of Demands |
| S13,S15 | (h1-h8 and h24) | (h9-h11 and h23) | (h12-h22) | Summer Cluster 13 and Cluster15 | Median and Ward.d2 Linkage of Hierarchical Clustering of Demands |
| S14 | (h2-h6) | (h1 and h7-h10 and h23-h24) | (h11-h22) | Summer Cluster 14 | Ward.d Linkage of Hierarchical Clustering of Demands |
| S16 | (h1-h11) | (h19-h24) | (h12-h18) | Summer Cluster 16 | K-means Clustering of Demands |

There are several clusters for summer season according to clustering of prices and demands, but we have used only those clusters which are significant and practical (i.e., some clustering results have only one or two hours for some clusters). We consider and analyse only those clusters that have at least 5 consecutive hours. The details about each of these significant clusters are given in the subsequent sections.

3.6 Clustering of Hours Using Hourly Electricity Prices (RTP)

We have used only weekdays (Monday to Friday) data, because the weekends including national holidays are supposed to be off-peak. However, a similar analysis can be performed for weekends and national holidays. The weekday's data for winter have 56 observations (weekdays), whereas for summer it has 65 observations. In this study, only one year (2014-2015) data is used, but historical time-series data for many years can be incorporated.

3.6.1 Results for Winter Season

In this section, the winter price data is clustered by observation and by hierarchical and k-means clustering techniques.

3.6.1.1 Clustering by Observation

First we clustered the prices of winter season and found the TOU blocks (off-peak, mid-peak, and on-peak) and we have used the median of the dataset to determine the TOU prices.

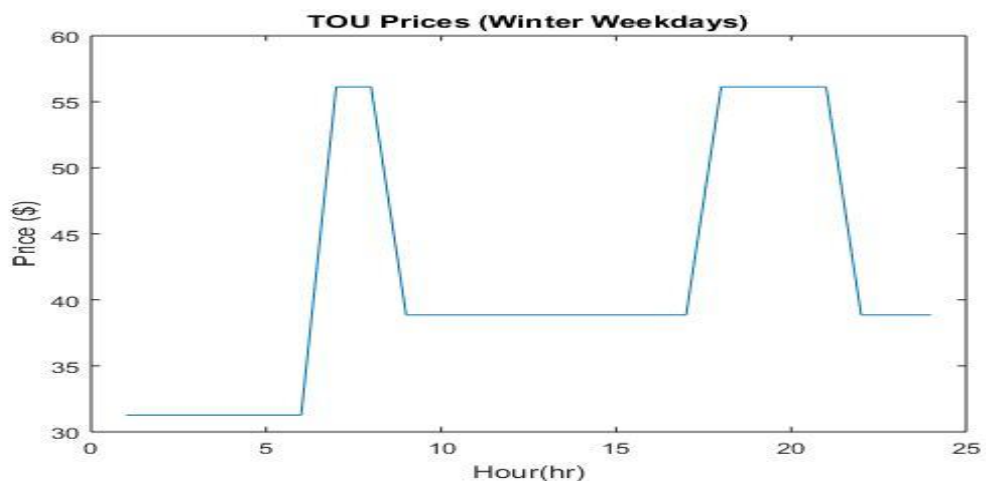


Figure 11: TOU Prices of W1

Table 3 and Table 4 represent the different pricing schemes by observation. It also gives the prices at each TOU blocks and the hours of the day. We are using the median because of the skewed distribution of prices and the presence of outliers.

Table 3: Clustering by Observing the Prices (W1)

| W1 | | Off-peak (h1 to h6) | Mid-peak (h9 to h17 and h22-h24) | On-peak (h7 to h8 and h18 to h21) |
|----|-------------|------------------------|--|---|
| | Prices | 31.267 | 38.859 | 56.129 |
| | Total Hours | 6 | 12 | 6 |

Table 4: Clustering by Observing the Prices Alternatively (W2)

| W2 | | Off-peak (h1-h6 and h23-h24) | Mid-peak (h9-h17) | On-peak (h7-h8 and h18-h22) |
|----|-------------|------------------------------------|----------------------|-----------------------------------|
| | Prices | 31.96 | 41.269 | 50.764 |
| | Total Hours | 8 | 9 | 7 |

3.6.1.2 Clustering by Hierarchical Algorithm

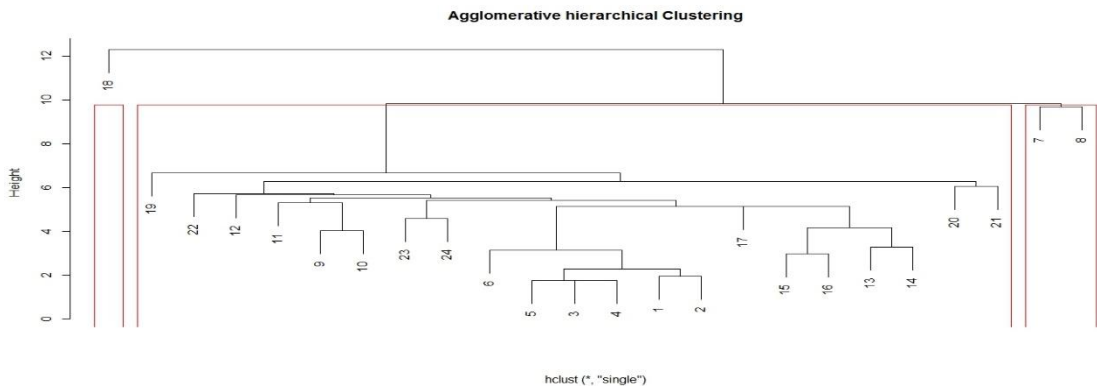


Figure 12: Single Linkage for Winter Prices

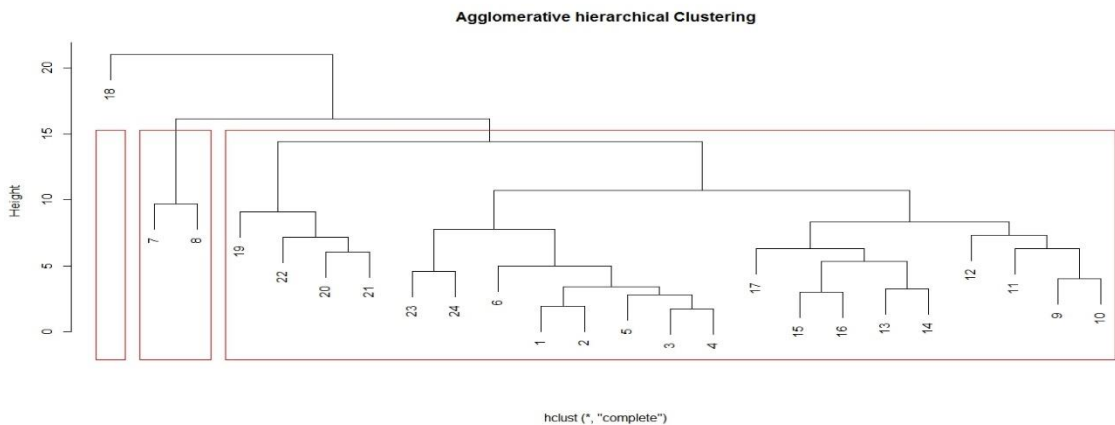


Figure 13: Complete Linkage for Winter Prices

Figure 12 and Figure 13 presents the hierarchical clustering method with single and complete linkage methods, respectively. For both methods, it can be observed that

one cluster is very large as compared to the other clusters which is not a good practice for TOU pricing (i.e., almost all hours are grouped as off-peak or mid-peak). The main difference in these diagrams is only the sequence of clusters. First cluster contains only one hour in both single and complete linkage methods. Third cluster in single linkage contains only two hours, hours 7 and 8, while in complete linkage these two hours are in the second cluster. The rest of the hours are clustered in a single cluster, the second cluster in single linkage method and the third cluster in complete linkage method.

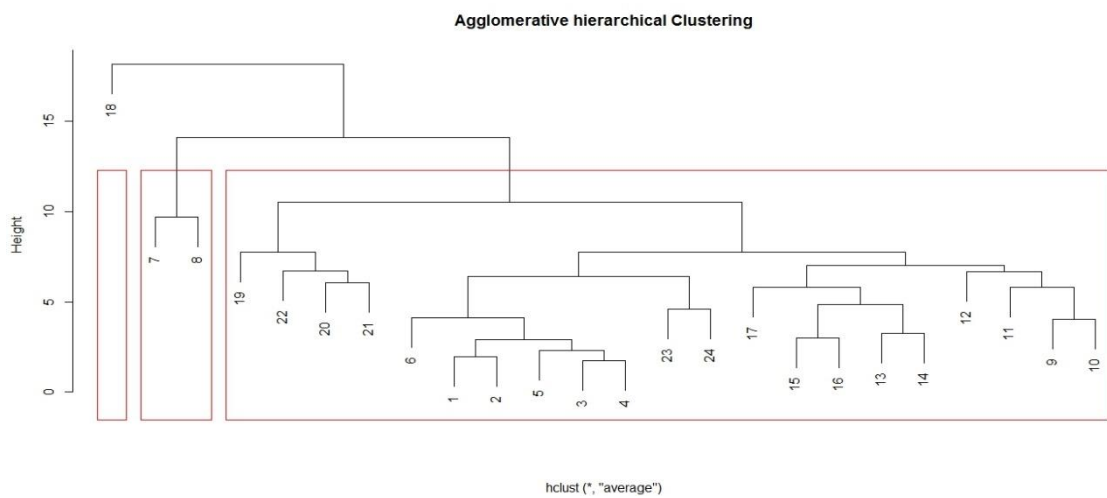


Figure 14: Average Linkage for Winter Prices

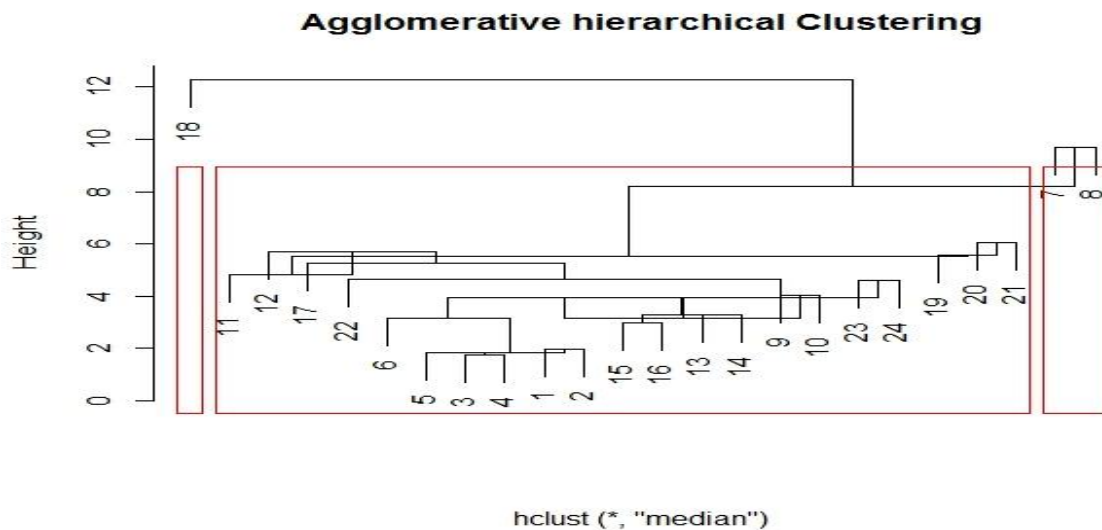


Figure 15: Median Linkage for Winter Prices

In Figure 14 and Figure 15, clustering results for average and median linkage methods are presented. They are similar to the single linkage and complete linkage

methods. One cluster is very large and has many hours. The remainder of the clusters have one or two hours in the cluster.

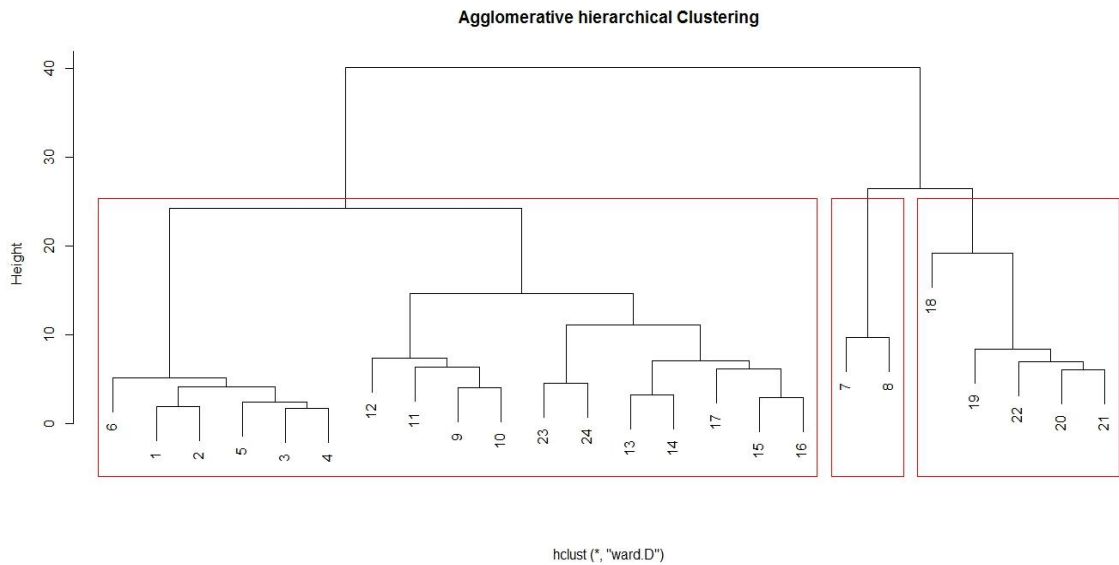


Figure 16: Ward.d Linkage for Winter Prices

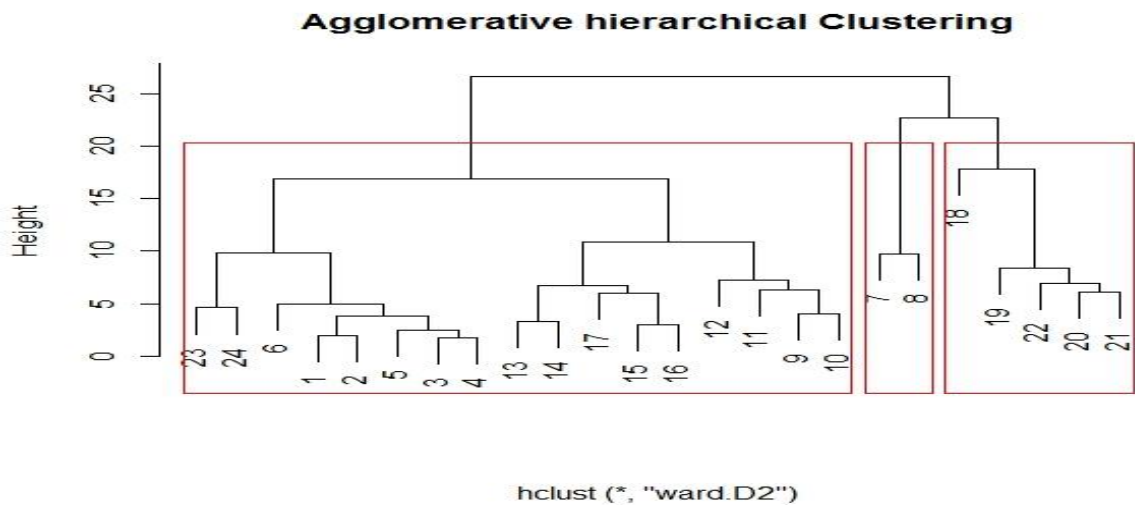


Figure 17: Ward.d2 Linkage for Winter Prices

In Figure 16 and Figure 17 of ward.d and ward.d2 linkage methods, clustering results seem better than the other linkage methods (single, complete, average, and median) but not an ideal case for the winter season. The ward.d and ward.d2 linkage methods are similar to each other such that the initial (first) cluster contains 17 hours (i.e., hour 1 to hour 6, hour 9 to hour 17 and hour 23 to hour 24). The second cluster contains two hours (i.e., hour 7 and hour 8). The last cluster includes five hours (i.e., hour 18 to hour 22).

3.6.1.3 Clustering by K-means Algorithm

We clustered the prices for winter season into three clusters as the scree plot suggested.

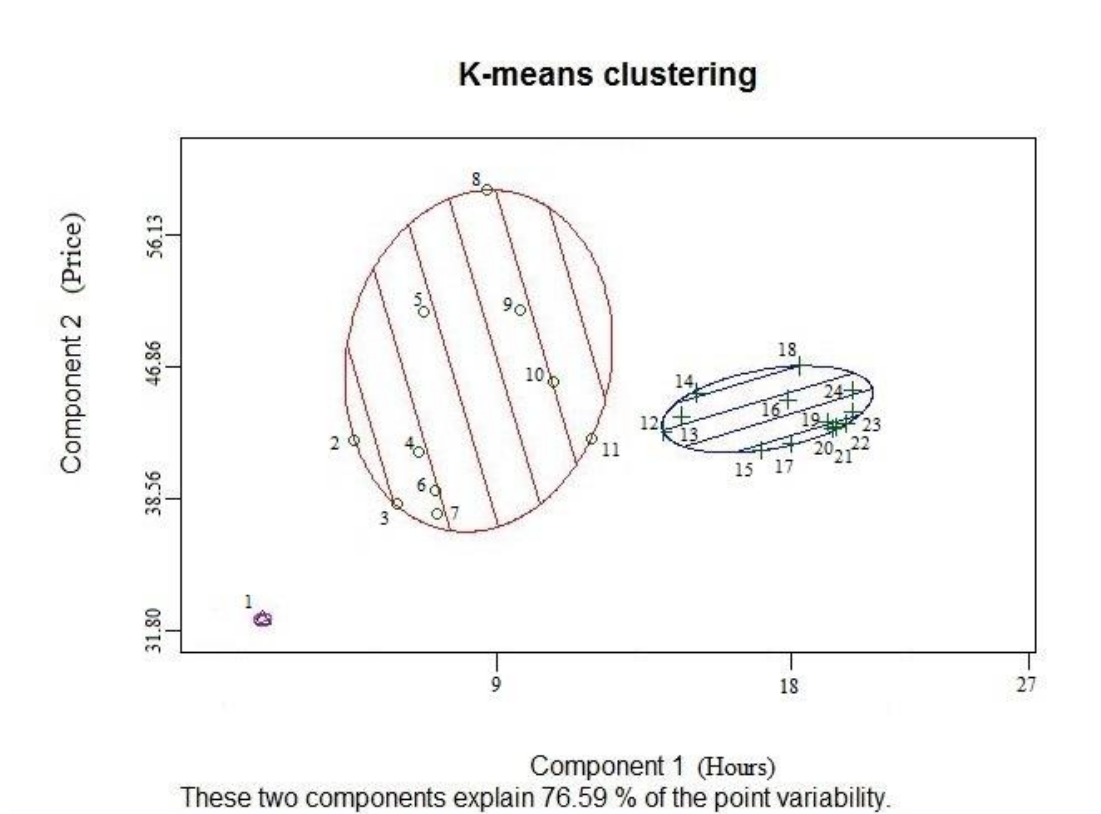


Figure 18: K-means Clustering for Winter Prices

We have repeated the k-means algorithm to get an accurate k-means clustering result because the algorithm starts from random centroids. The numeric numbers on above Figure 18 depicts the number of hours. First cluster has only one hour which is the off-peak block with a TOU price of \$31.8. Second cluster is very large as compared to cluster one and it includes 10 hours (i.e., hour 2 to hour 11). This is the mid-peak period with a TOU price of \$38.42. Third cluster is the largest cluster of the winter prices which contains 13 hours (i.e., hour 12 to hour 24). This corresponds to the on-peak period with a TOU price of \$45.362.

3.6.2 Results for Summer Season

In this section, similar clustering analyses are performed for the prices in the summer season.

3.6.2.1 Clustering by Observation

We clustered the price data for summer season into three clusters by observation.

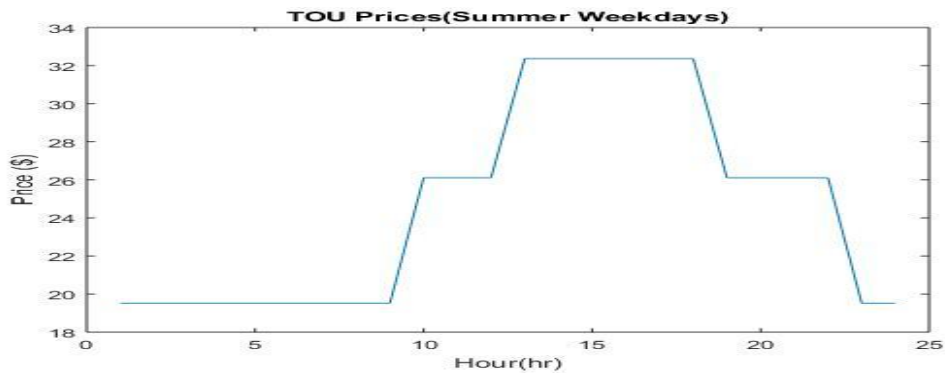


Figure 19: TOU Prices for Summer Cluster (S1)

Table 5 represents the pricing scheme of the TOU prices for summer season. Figure 19 provides a depiction of the TOU pricing scheme for this cluster (off-peak, mid-peak, and on-peak prices).

Table 5: Clustering by Observing the Summer Prices (S1)

| S1 | | Off-peak (h1-h9 and h23 and h24) | Mid-peak (h10-h12 and h19-h22) | On-peak (h13-h18) |
|-------------|--------|--|--------------------------------------|----------------------|
| | Prices | | 19.521 | 26.113 |
| Total Hours | | 11 | 7 | 6 |

3.6.2.2 Clustering by Hierarchical Algorithm

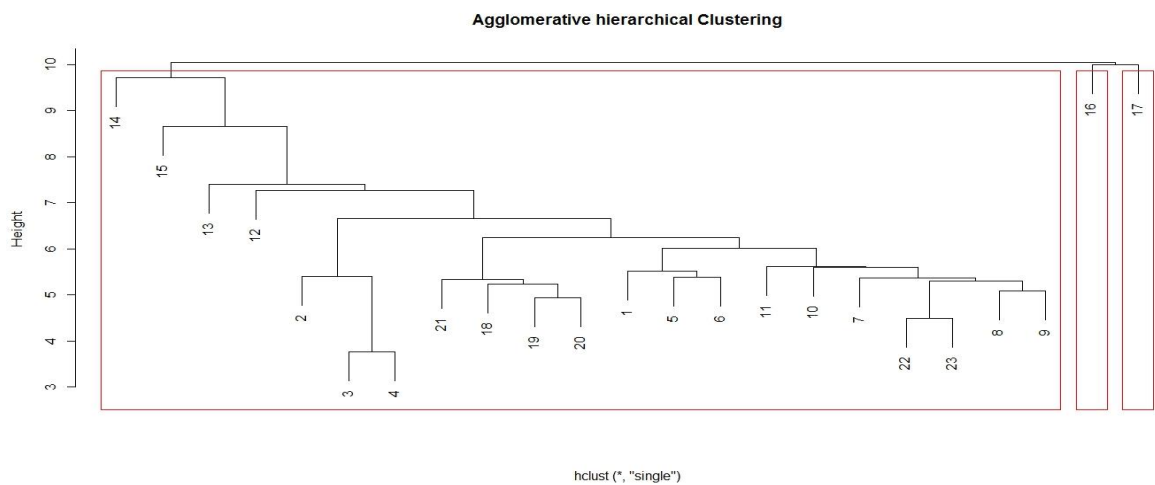


Figure 20: Single Linkage for Summer Prices

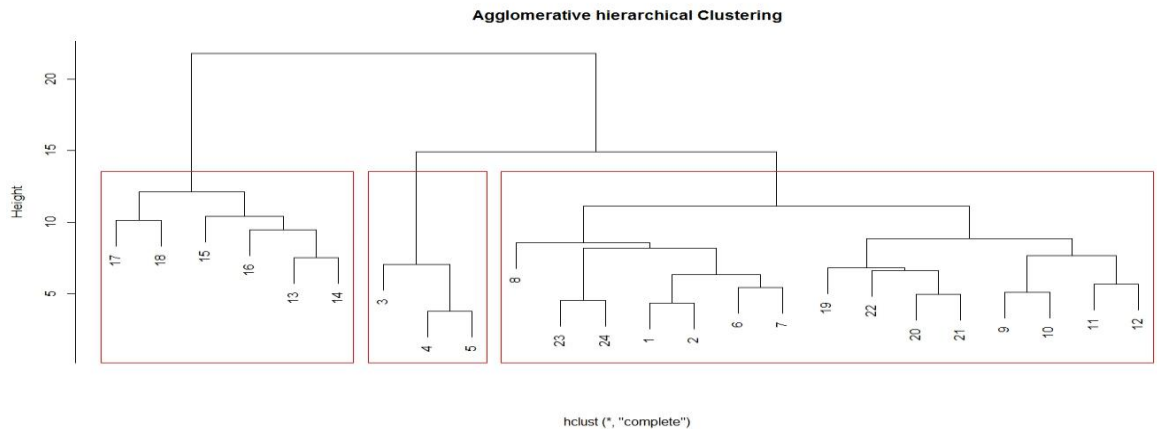


Figure 21: Complete Linkage for Summer Prices

Figure 20 shows the results for the single linkage method. One of the clusters is very large as compared to other clusters. It contains 22 hours (i.e., hour 1 to hour 15 and hour 18 to hour 24). On the other hand, the other two clusters contain only a single hour. The complete linkage method is presented in Figure 21. It can be observed that the complete linkage method seems better as compared to the single linkage method. Cluster one contains total 6 hours (i.e., hour 13 to hour 18). Second cluster contains 3 hours (i.e., hour 3 to hour 5). Third cluster is the largest cluster, which contains 15 hours (i.e., hour 1 to hour 2, hour 6 to hour 12, and hour 19 to hour 24).

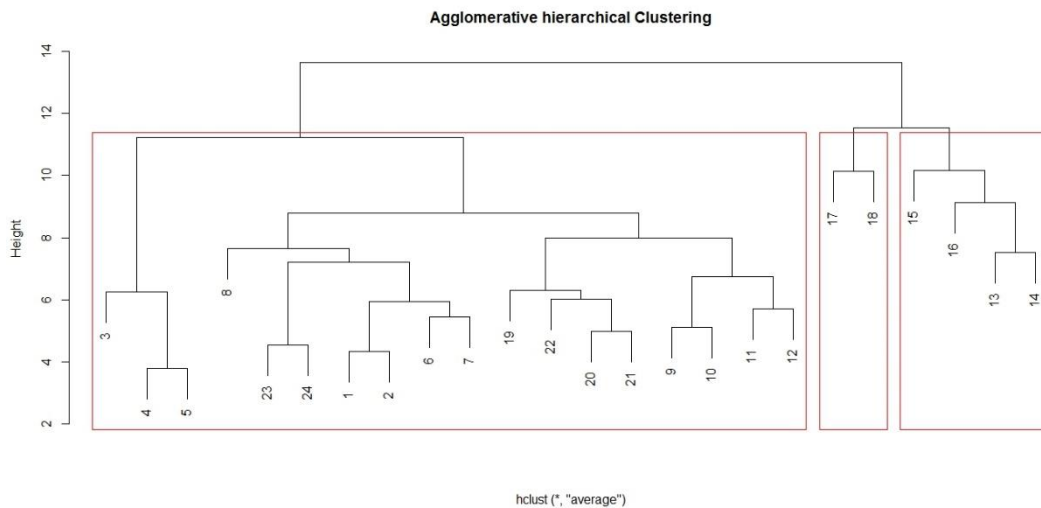
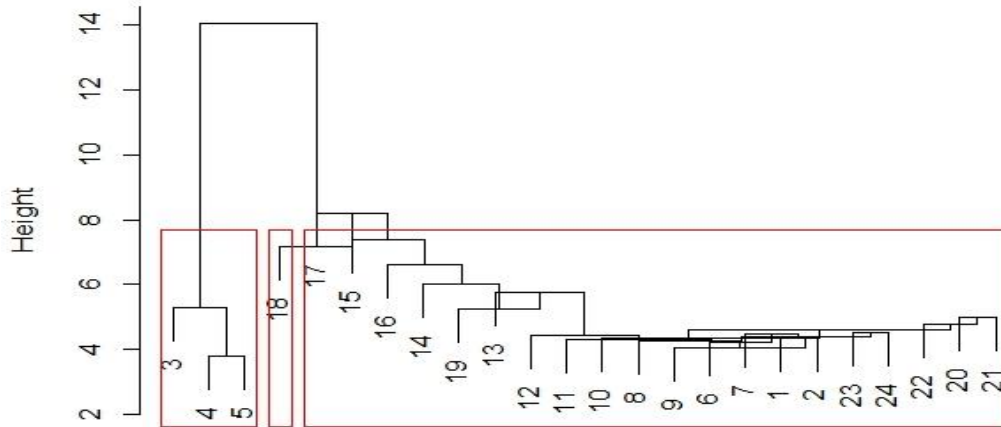


Figure 22: Average Linkage for Summer Prices

Agglomerative hierarchical Clustering

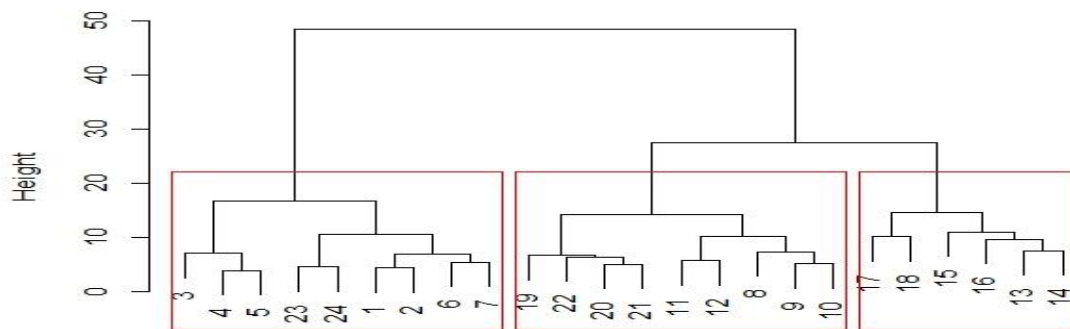


hclust (*, "median")

Figure 23: Median Linkage for Summer Prices

Figure 22 and Figure 23 display the results for the average linkage and median linkage methods, respectively. In both methods, the problem of having a very large cluster persists. The main difference between average and median linkage methods is that, in average linkage method first cluster is the largest cluster, whereas in median linkage method third cluster is the largest one.

Agglomerative hierarchical Clustering



hclust (*, "ward.D")

Figure 24: Ward.d Linkage for Summer Prices

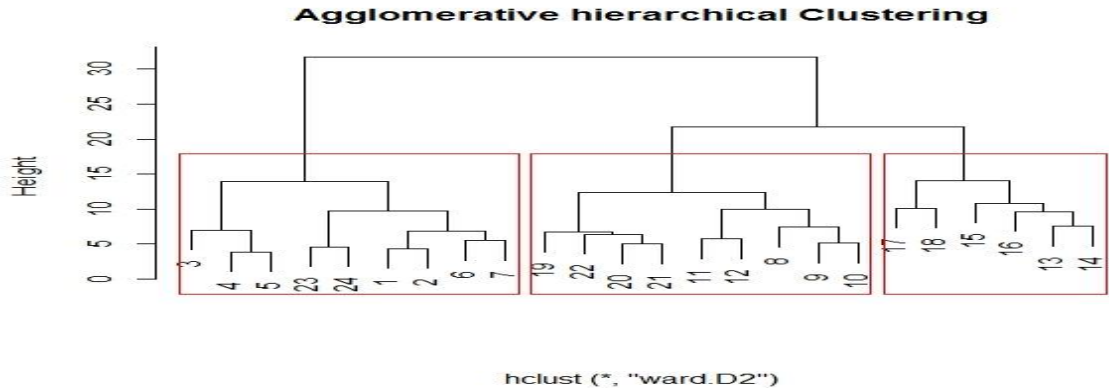


Figure 25: Ward.d2 Linkage for Summer Prices

Results for the ward.d linkage and ward.d2 linkage methods are displayed in Figure 24 and Figure 25, respectively. These two methods provide a balance among clusters in terms of number of hours included in each cluster. Ward.d linkage method and ward.d2 linkage method are similar to each other. Cluster one in both linkages hold the 9 hours (i.e., hour 1 to hour 7 and hour 23 to hour 24). The second cluster accommodates 9 hours (i.e., hour 8 to hour 12 and hour 19 to hour 22). On the other hand, the third cluster contains 6 hours (i.e., hour 13 to hour 18). Table 6 gives the details about the ward.d linkage and ward.d2 linkage methods.

Table 6: Hierarchical Clustering of Summer Prices (S6-S7)

| | | Off-peak (h1-h7 and h23-24) | Mid-peak (h8-h12 and h19-22) | On-peak (h13-h17) |
|-------|-------------|-----------------------------------|------------------------------------|----------------------|
| S6,S7 | Prices | 18.3979 | 25.20357 | 32.36881 |
| | Total Hours | 9 | 9 | 6 |

3.6.2.3 Clustering by K-means Algorithm

We have clustered the prices for winter season into three clusters as the scree plot suggested.

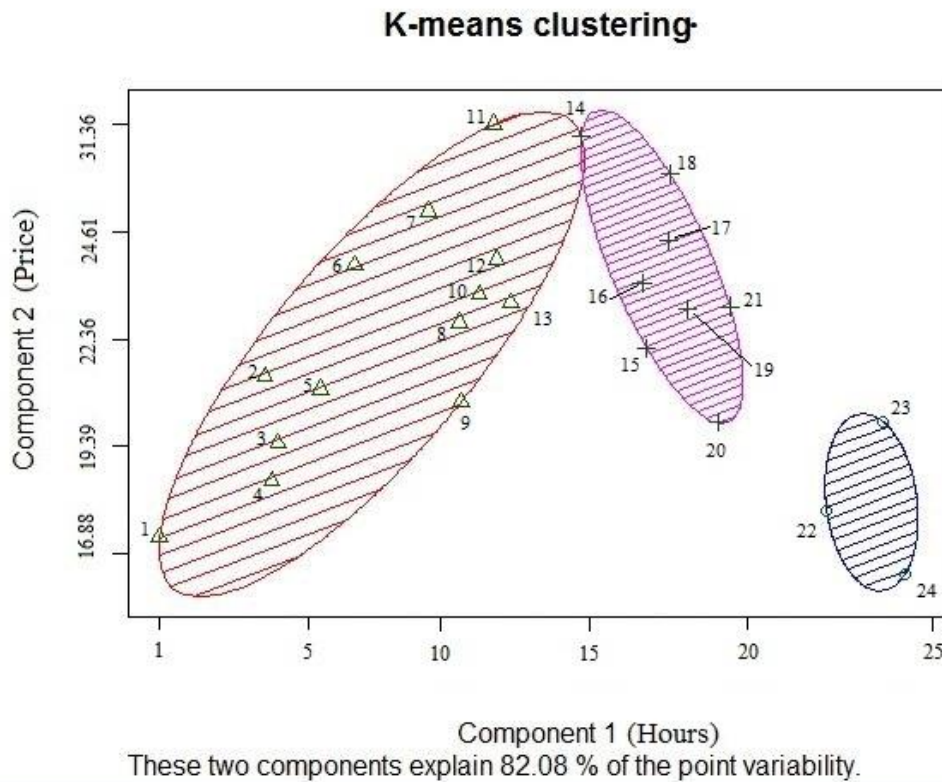


Figure 26: K-means Clustering of Summer Prices

As observed from Figure 26, the first cluster is the largest and holds 13 hours (i.e., hour 1 to hour 13). It is the mid-peak period with a TOU price of \$21.899. Second cluster accommodates 8 hours (i.e., hour 14 to hour 21) and it is the on-peak period with a TOU price of \$23.66. Finally, last cluster has only three hours (i.e., hour 22 to hour 24) and it is the off-peak period with a TOU price of \$17.89.

3.6.3 MAPE and Change in Revenue

Table 7 gives the details about the mean absolute percent error (MAPE) and the change in revenue measures for all the clusters in winter and summer seasons when compared to RTP in the market.

Table 7: MAPE and Change in Revenue measures (TOU vs RTP for winter and summer seasons)

| Cluster Type | Off-peak Price | Mid-peak Prices | On-peak Prices | Change in Revenue (%) | MAPE (%) |
|--------------|----------------|-----------------|----------------|-----------------------|----------|
| W1 | 31.267 | 38.859 | 56.129 | 9.513 | 67.47923 |
| W2 | 31.96 | 41.269 | 50.764 | 10.373 | 67.70904 |
| S1 | 19.521 | 26.113 | 32.369 | 8.292 | 90.50 |
| S6,S7 | 18.3979 | 25.20357 | 32.36881 | 9.019 | 86.20 |

We found the revenue difference between the TOU price and RTP by multiplying the hourly demands with corresponding RTP and TOU prices in both seasons and computed the percent change in revenue when TOU prices are used instead of RTP. The values in “Change in Revenue (%)” column in Table 7 reflect the reduction in revenues collected when TOU prices are used. For the winter season, MAPEs are low as compared to the summer season. Note that any pricing scheme other than RTP will have a degree of error compared to RTP. The revenue requirements are less than 11% for all clustering results based on price data.

We have calculated the MAPE by using the following formula:

$$MAPE = 100 * \frac{1}{n} \sum | Revenue_{TOU} - Revenue_{RTP} | / Revenue_{RTP}$$

where $n = 24 * 56$ for winter season and $n = 24 * 65$ for summer season, and $| . |$ denotes the absolute value.

Finally, Table 7 also shows the MAPE results for the winter and summer season clusters. The MAPE values for W1 and W2 clusters are lower than these values for S1, S6, and S7 clusters.

3.7 Clustering of Hours Using Hourly Electricity Demands (Loads)

In this section, we have analysed the demand (load) data similar to analyses for price data in section 3.6.

3.7.1 Results for Winter Season

By using the load data in both seasons (winter and summer) we have performed similar clustering analyses as in section 3.6.

3.7.1.1 Clustering By Observation

By simply observing the hourly load dataset we have categorized the hours of the day into three blocks (off-peak, mid-peak, on-peak)

Table 8: Clustering by Observing the Winter Demands (W10)

| W10 | | Off-peak (h1-h6 and h24) | Mid-peak (h11-h17 and h23) | On-peak (h7-h10 and h18-h22) |
|-----|-------------|--------------------------------|----------------------------------|------------------------------------|
| | Prices | 31.1766 | 39.385 | 53.131 |
| | Total Hours | 7 | 8 | 9 |

3.7.1.2 Clustering by Hierarchical Algorithm

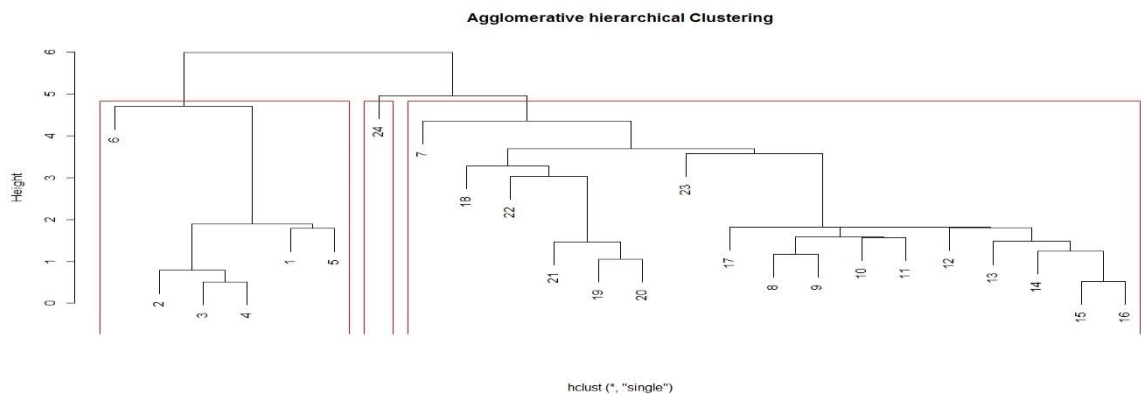


Figure 27: Single Linkage for Winter Load

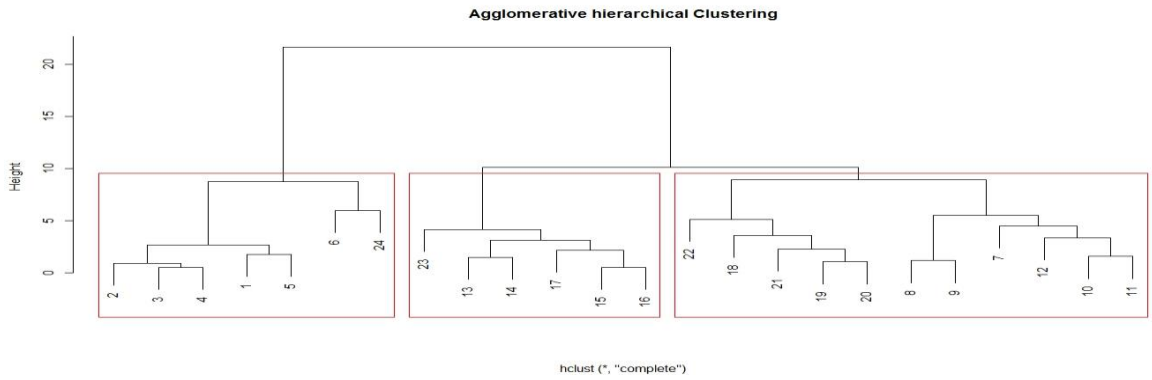


Figure 28: Complete Linkage for Winter Loads

Figure 27 and Figure 28 shows the results for the single linkage and complete linkage methods of the winter loads, respectively. First cluster holds 6 hours (i.e., hour 1 to hour 6). Second cluster contains only a single hour (i.e., hour is 24), while the rest of the hours accommodate in the third cluster. Complete linkage method is far better than the single linkage method. First cluster in complete linkage method contains 7 hour (i.e., hour 1 to hour 6 and hour 24). Second cluster carries total of 6 hours (i.e., hour 13 to hour 17 and hour 23). Third cluster holds 11 hours (i.e., hour 7 to hour 12 and hour 18 to hour 22).

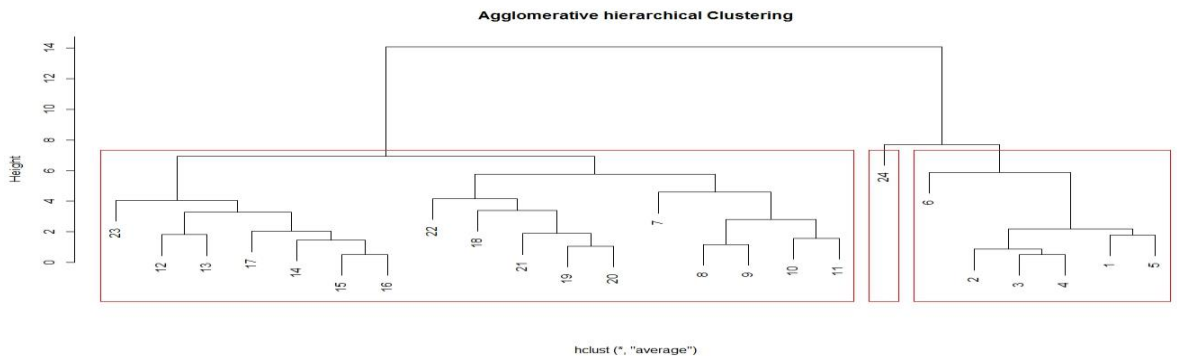


Figure 29: Average Linkage for Winter Loads

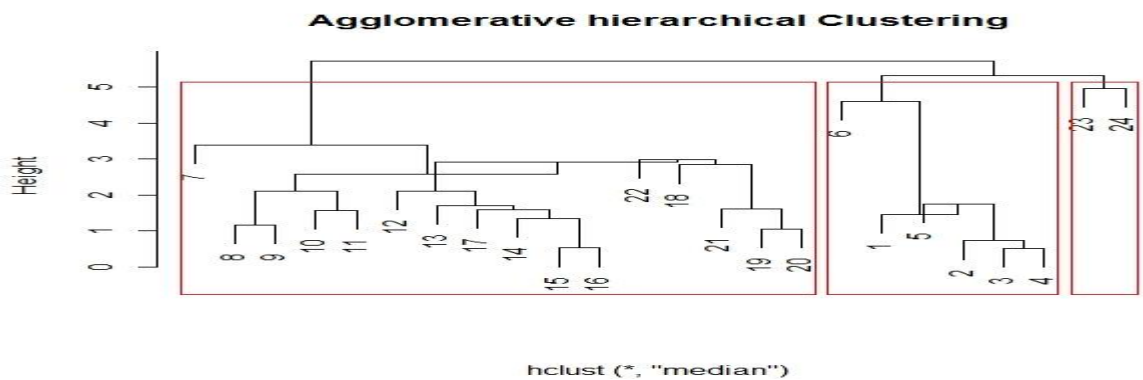


Figure 30: Median Linkage for Winter Loads

In average linkage method (Figure 29) and median linkage method (Figure 30), one cluster is very large and the other clusters are very small. The third cluster in average linkage method is quite similar to the second cluster of the median linkage method. But the main difference in both methods is the sequence of clusters.

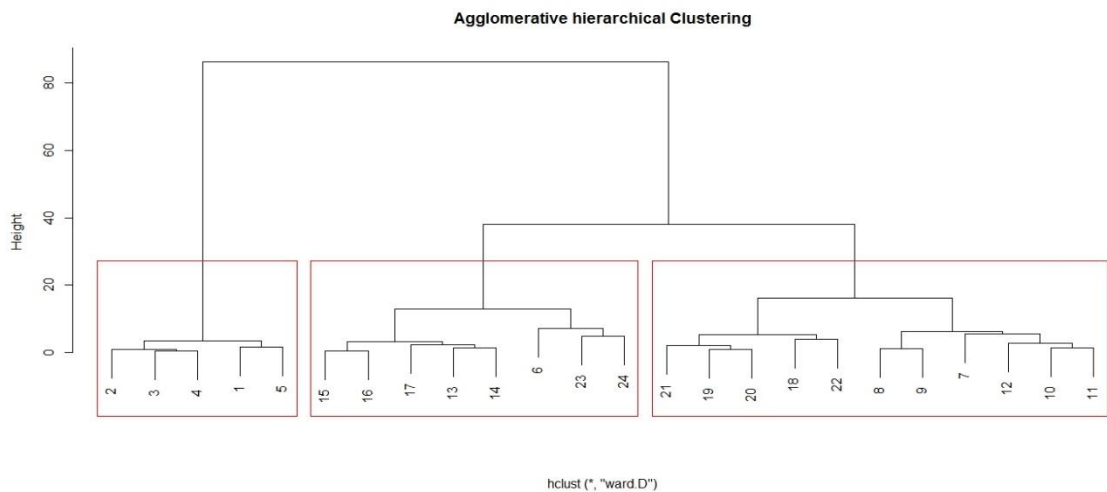


Figure 31: Ward.d Linkage for Winter Loads

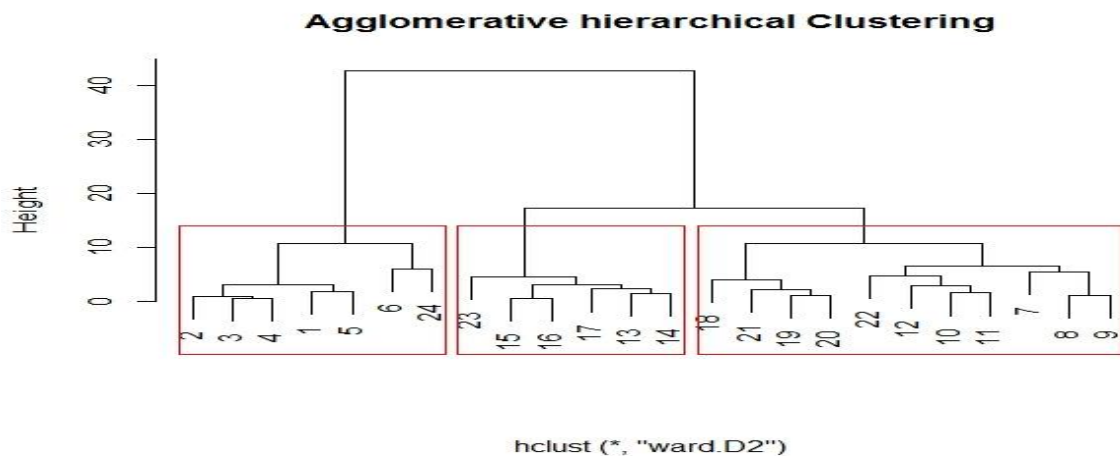


Figure 32: Ward.d2 Linkage for Winter Loads

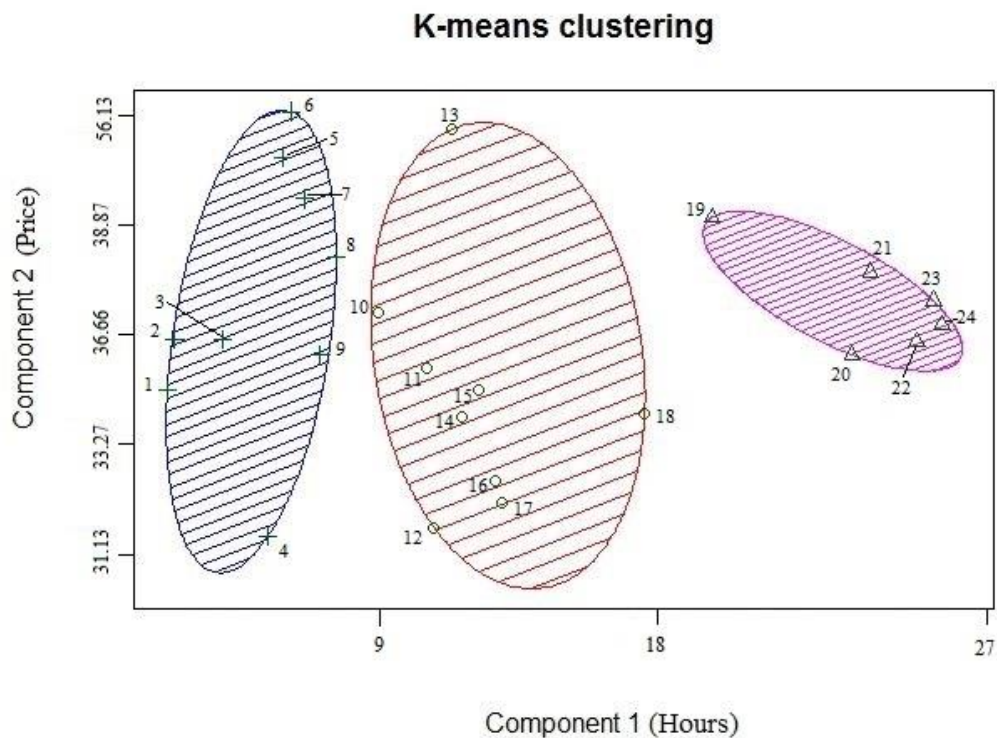
Figure 31 and Figure 32 displays the results for ward.d linkage and ward.d2 linkage methods, respectively. In these methods, it can be easily observed that ward.d2 linkage method is better. In ward.d linkage method first cluster consists of 5 hours (i.e., hour 1 to hour 5). Second cluster contains 8 hours (i.e., hour 13 to hour 17, hour 23 to hour 24 and hour 6). Third cluster accommodates a total of 11 hours (i.e., hour 7 hour 12 and hour 18 to hour 22). In ward.d2 linkage method, first cluster contains 7 hours, second cluster contains 6 hours and the rest of the hours are in the third cluster. The details for the ward.d2 method are given in the Table 9.

Table 9: Hierarchical Clustering of Winter Demands (W12,W16)

| W12,W16 | | Off-peak (h1-h6 and h24) | Mid-peak (h13-h17 and h23) | On-peak (h7-h12 and h18-h22) |
|---------|-------------|--------------------------------|----------------------------------|------------------------------------|
| | Prices | 31.1766 | 36.6834 | 52.1051 |
| | Total Hours | 7 | 6 | 11 |

3.7.1.3 Clustering by K-means Algorithm

Winter load data are clustered into three clusters by using the k-means clustering algorithm.



These two components explain 98.24 % of the point variability.

Figure 33: K-means Clustering for Winter Loads

In Figure 33, the results depict the three clusters. These clusters are more balanced in terms of number of hours in each cluster, compared to hierarchical clustering algorithm results. First cluster contains a total of 9 hours (i.e., hour 1 to hour 9) and it is the on-peak period with a TOU price of \$46.367. Second cluster has 9 hours (i.e., hour 10 to hour 18) and it is the off-peak period with a TOU price of \$35.017. Third cluster holds 6 hours (i.e., hour 18 to hour 24) and it is the mid-peak period with a

TOU price of \$39.129. The point variability between two components shows how much the dataset in the statistical distribution diverges from the mean. Table 10 summarizes the TOU pricing scheme for the k-means clustering algorithm.

Table 10: K-means Clustering of Winter Demands (W17)

| | | Off-peak (h1-h9) | Mid-peak (h10-h18) | On-peak (h19-h24) |
|-----|-------------|---------------------|-----------------------|----------------------|
| W17 | Prices | 37.56026 | 43.36107 | 47.20006 |
| | Total Hours | 9 | 9 | 6 |

3.7.2 Results for Summer Season

In this section, similar clustering analyses are performed for the loads in the summer season.

3.7.2.1 Clustering by Observation

By observation of the load data in summer, we have clustered the hours of the day as presented in Table 11.

Table 11: Clustering by Observing Summer Demands (S9)

| | | Off-peak (h1-h8 and h24) | Mid-peak (h9-h12 and h21-h23) | On-peak (h13-h20) |
|----|-------------|--------------------------------|-------------------------------------|----------------------|
| S9 | Prices | 18.339 | 24.859 | 30.945 |
| | Total Hours | 9 | 7 | 8 |

3.7.2.2 Clustering by Hierarchical Algorithm

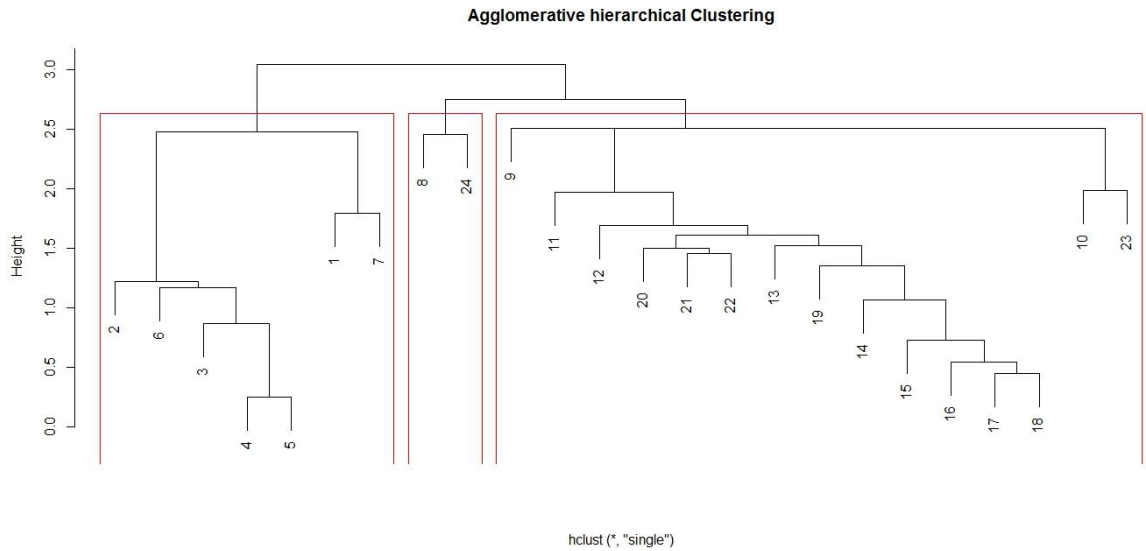


Figure 34: Single Linkage for Summer Loads

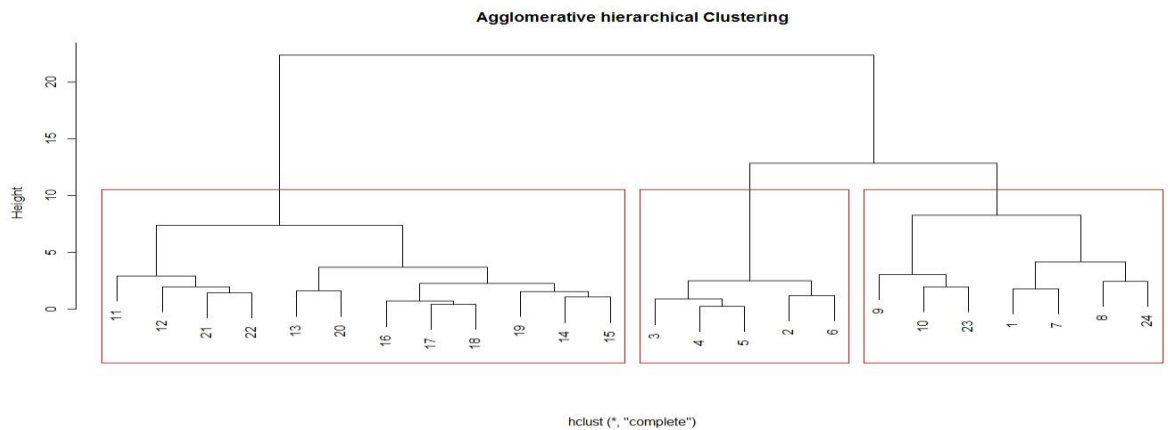


Figure 35: Complete Linkage for Summer Loads

In Figure 34 and Figure 35, the results for the single and complete linkage methods are depicted, respectively. The main drawback of the single linkage method is that it gives one large cluster and two small clusters, as we have examined in the clustering of price in winter and summer seasons as well as the clustering of load data in winter season. In single linkage method, the third cluster is the largest cluster and it contains 15 hours (i.e., hour 9 to hour 23). The second cluster holds only two hours (i.e., hour 8 and hour 24). The rest of the hours are in the first cluster. Compared to single linkage method, the complete linkage method has balance among clusters. First cluster holds a total of 12 hours, whereas second and third clusters contain 5 and 7 hours, respectively.

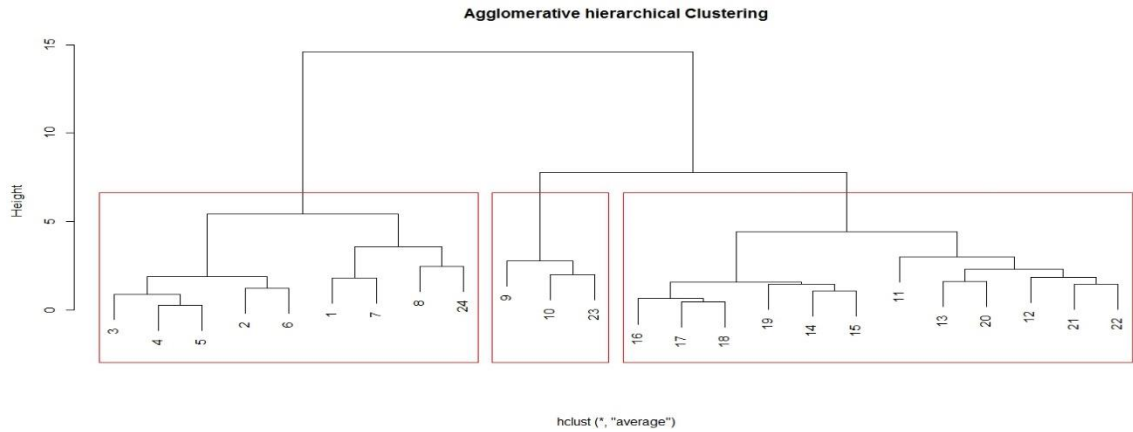


Figure 36: Average Linkage for Summer Loads

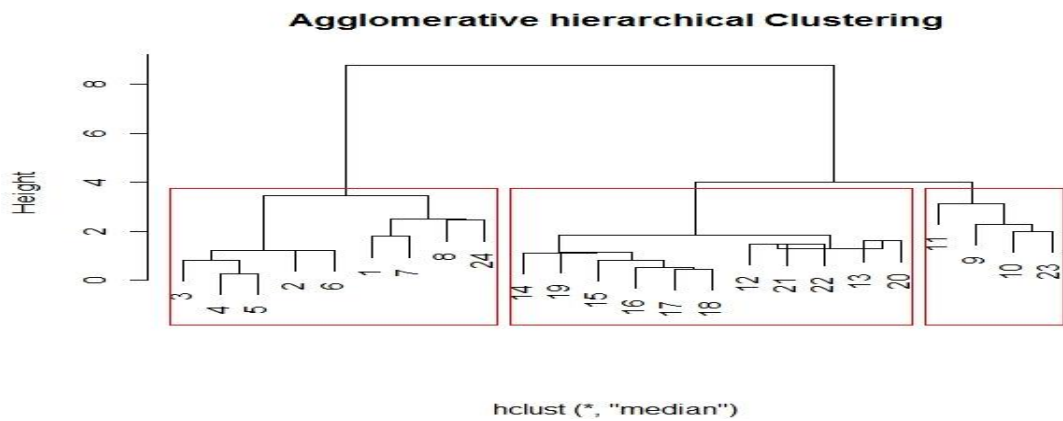


Figure 37: Median Linkage for Summer Loads

In Figure 36, the first cluster contains 9 hours (i.e., hour 1 to hour 8 and hour 24). Second cluster is the smallest cluster that carries only three hours (i.e., hour 9 to hour 10 and hour 23). The rest of the hours are in the third cluster. In median linkage method (Figure 37), the first cluster contains 9 hours (i.e., hour 1 to hour 8 and hour 24), the second one has 11 hours (i.e., hour 12 to hour 22), and the third cluster holds only four hours (i.e., hour 9 to hour 11 and hour 23).

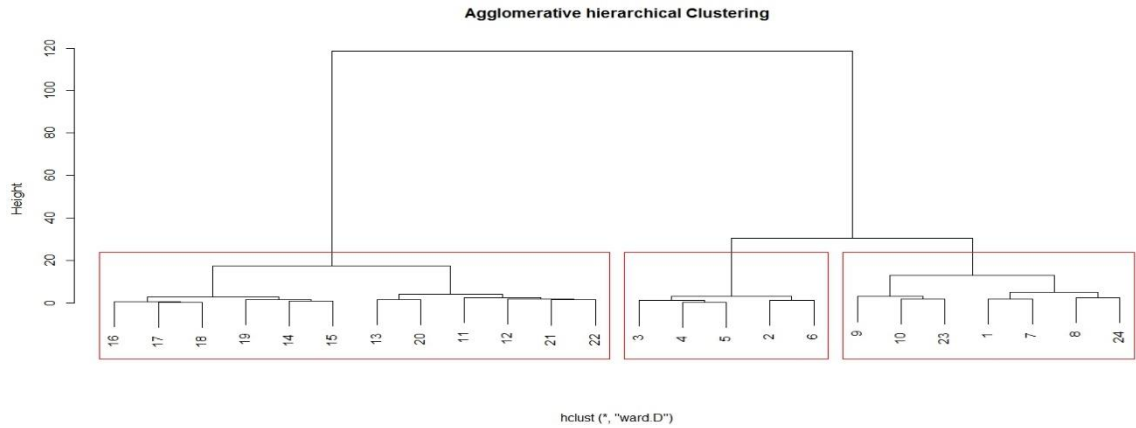


Figure 38: Ward.d Linkage for Summer Loads

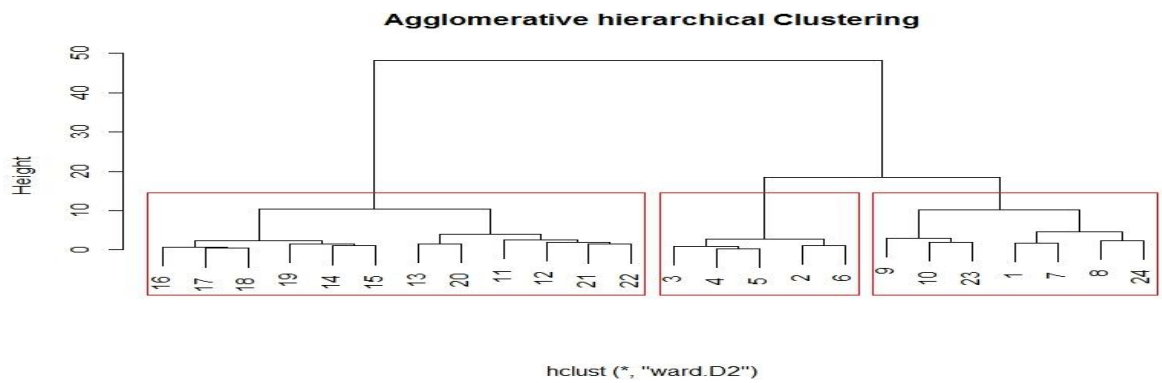


Figure 39: Ward.d2 Linkage for Summer Loads

Figure 38 and Figure 39 are exactly same for ward.d and ward.d2 linkage methods, respectively. In both methods, the first cluster holds a total of 12 hours (i.e., hour 11 hour 22). The second cluster in both methods contains 5 hours (i.e., hour 2 to hour 6). Finally, the third cluster accommodates 7 hours (i.e., hour 7 to hour 10, hour 23 to hour 24 and hour 1) in both methods.

3.7.2.3 Clustering by K-means Algorithm

We have clustered the loads for summer season into three clusters as the scree plot suggested.

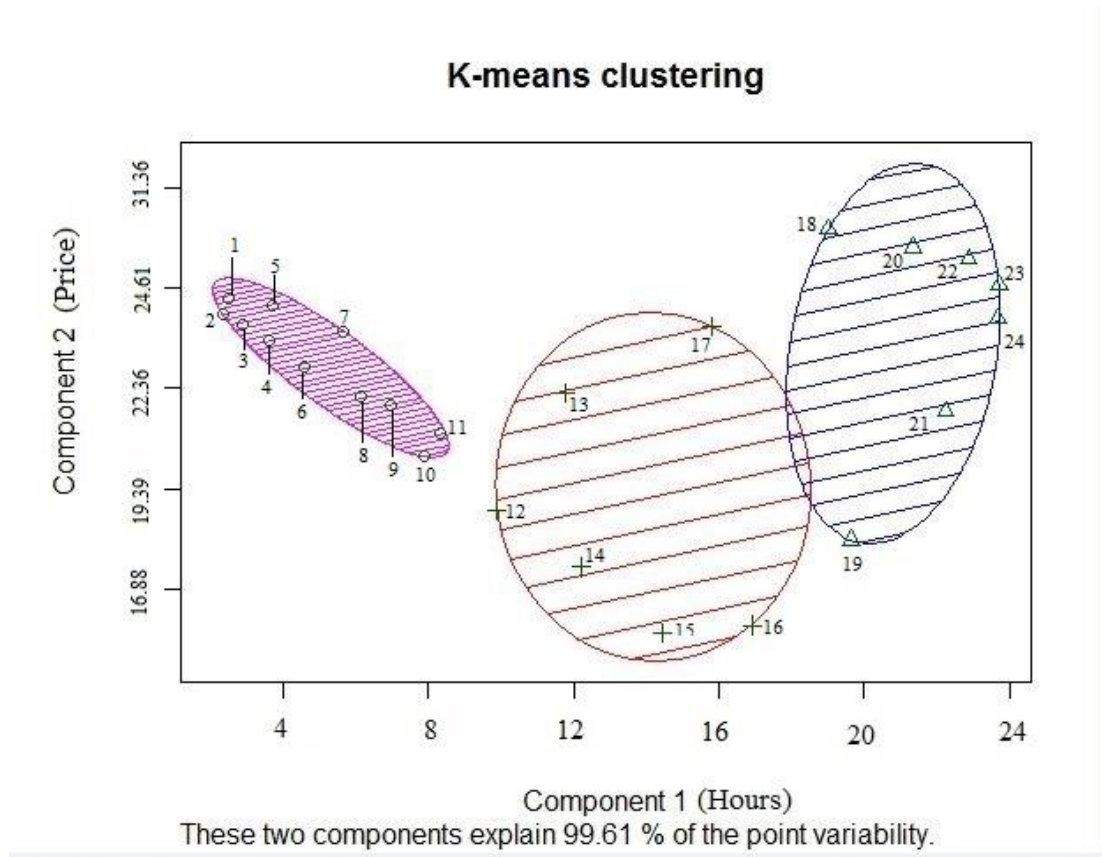


Figure 40: K-means Clustering for Summer Loads

Table 12 shows the number of hours in each cluster and the TOU prices at corresponding TOU clusters. The first cluster is the largest cluster in Figure 40.

Table 12: K-means Clustering of Summer Demands (S16)

| S16 | | Off-peak (h1-h11) | Mid-peak (h19-h24) | On-peak (h12-h18) |
|-------------|--------|----------------------|-----------------------|----------------------|
| | Prices | 19.88266 | 24.60857 | 31.36208 |
| Total Hours | 11 | 6 | 7 | |

3.7.3 MAPE and Change in Revenue

Table 13 gives the change in revenue and mean percent error of the clusters by load data in winter and summer seasons.

Table 13: MAPE and Change in Revenue measures (TOU vs RTP for winter and summer seasons)

| Cluster Type | Off-peak Prices | Mid-peak Prices | On-peak Prices | Change in Revenue (%) | MAPE (%) |
|--------------|-----------------|-----------------|----------------|-----------------------|----------|
| W10 | 31.1766 | 39.385 | 39.385 | 7.478 | 69.96397 |
| W12,W16 | 31.1766 | 36.6834 | 36.6834 | 7.449 | 68.09051 |
| W17 | 37.56026 | 43.36107 | 43.36107 | 8.431 | 73.27241 |
| S9 | 18.339 | 24.859 | 24.859 | 8.875 | 86.37 |
| S16 | 19.88266 | 24.60857 | 24.60857 | 9.253 | 92.50 |

For the winter season, MAPEs are low as compared to the summer season. MAPE basically represents a measure of the accuracy of the procedure in statistics. The error terms compared to RTP are large in this study. However, any pricing scheme other than RTP will have a degree of error compared to RTP. Note that the revenue requirements less than 10% for all clustering results based on load data.

Chapter 4: Multiple Linear Regression Analyses of Clusters

4.1 Overview of Multiple Linear Regression Analyses

A lot of forecasting models have been developed by using demographic, social, economic and geographic factors for electricity market outcomes. In this study, we are using regression models: a) to decide on the TOU pricing scheme that improve the relation between price and load levels and b) to choose among different clustering algorithms we have presented in chapter 3 by using the regression analyses results. In data modelling, regression analysis is a statistical procedure for approximating the connection among different variables. It contains many procedures for analysing and modelling a number of variables when the centre of attention is the relationship between one or more independent variables (or predictors) and a dependent variable. Normally, the examiner seeks to ascertain the normal impact of one variable on another variable, such as the price is directly affected by demand (Sykes, 1993).

In our models for both winter and summer seasons, we first find the median value of TOU prices/loads for each week (12 weeks for winter and 14 weeks for summer), and then compute the average TOU values for summer and winter season (i.e., a total of 24 price and demand values). In our case, these average TOU prices are the dependent variable, while the average TOU demands and categorical (off-peak, mid-peak, on-peak) variables are the independent variables.

In multiple linear regression analysis, dependent variable is normally influenced not only by ratio scale factor (price) but also by the quantitative elements. Such elements usually show the absence or presence of attribute. There is only one way to quantify such quality: by creating categorical (dummy) variables. These variables has the value 1 or 0, 1 representing the presence (or possession) of that quality and 0 representing the absence of that quality. When one variable is present, the rest of the other categorical variables are 0 (Gujarati and Porter, 2009). After creating the categorical variables for three clusters (i.e., off-peak, mid-peak, on-peak), we have performed the multiple linear regression analysis. We have also included logarithm

(log) and lag of price and load data in our analyses. Lag of price (P_{t-1}) and lag of load (D_{t-1}) as independent variables do not result in significant models (e.g., models without significant coefficients), hence they are excluded from the main body of the thesis, but the results for those models can be found in Appendix A and B.

We first examine the correlation among the prices and the demands for each hour of a day in both winter and summer seasons.

Table 14: Correlation between Price and Demand (winter season)

| | | | | | | | | |
|-------------|-------|-------|-------|-------|-------|-------|--------|--------|
| Hours | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H8 |
| Correlation | 0.59 | 0.589 | 0.621 | 0.621 | 0.595 | 0.598 | 0.5915 | 0.5334 |
| Hours | H9 | H10 | H11 | H12 | H13 | H14 | H15 | H16 |
| Correlation | 0.627 | 0.689 | 0.682 | 0.692 | 0.668 | 0.584 | 0.4672 | 0.6096 |
| Hours | H17 | H18 | H19 | H20 | H21 | H22 | H23 | H24 |
| Correlation | 0.627 | 0.621 | 0.621 | 0.594 | 0.546 | 0.505 | 0.5239 | 0.5883 |

Table 15: Correlation between Price and Demand (summer season)

| | | | | | | | | |
|-------------|--------|--------|--------|--------|-------|--------|--------|--------|
| Hours | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H8 |
| Correlation | 0.7015 | 0.5311 | 0.4893 | 0.5836 | 0.573 | 0.2882 | 0.5956 | 0.3693 |
| Hours | H9 | H10 | H11 | H12 | H13 | H14 | H15 | H16 |
| Correlation | 0.5354 | 0.5955 | 0.6432 | 0.6764 | 0.495 | 0.5991 | 0.4436 | 0.6542 |
| Hours | H17 | H18 | H19 | H20 | H21 | H22 | H23 | H24 |
| Correlation | 0.4649 | 0.4003 | 0.5929 | 0.6168 | 0.575 | 0.4842 | 0.6157 | 0.5491 |

Table 14 and Table 15 show the correlation among the prices and the demands for each hour of a day in both winter and summer. It is obvious that the price and demand are positively correlated for every hour and season. We have performed a Durbin-Watson test to confirm this in our regression analyses.

In econometrics, Durbin-Watson test is normally used to detect the existence of autocorrelation (a connection between values split from each other by a specified lag of time) in the residuals (errors of prediction) from an analysis of regression (Gujarati and Porter, 2009). To observe the autocorrelation at significance level (α), the test statistics is compared to the upper and lower critical values.

- $H_0: \rho=0$ Against $H_1: \rho>0$. Reject H_0 at α level if $d < d_u$. Then there is remarkable positive autocorrelation.

- $H_0: \rho=0$ Against $H_1: \rho<0$. Reject H_0 at α level if the approximated $(4-d) < d_u$. Then there is remarkable proof of negative autocorrelation.
- $H_0: \rho=0$ Against $H_1: \rho\neq 0$. Reject H_0 at 2α level if $d < d_u$ or $(4-d) < d_u$, than the test is inconclusive either the autocorrelation is positive or negative.

The above conditions are used to observe the correlation between dependent and independent variables. Minitab is used to find the value of the Durbin-Watson statistics. After finding the “d” value on Minitab and by following the above conditions, it is compared to the upper and lower values in the Durbin-Watson table at 95% significance level.

4.2 Results for Multiple Linear Regression Analysis of Clusters

4.2.1 Regression Models for Winter Season

Table 16 shows the significant regression models for winter season for each significant cluster found in section 3.6 and 3.7 for winter season. The regression models are not only significant as a whole but also each independent variable’s regression coefficient is significant.

Table 16: Regression Models for Winter Season

| Clu. Type | Reg. Mod | Dep. Var. | Ind. Var. | Adj. R Squ. | Sig. | B Coe. | P Value | Durbin Watson Test |
|-----------|----------|-----------|-------------------------------------|-------------|----------|--|--|--------------------------------------|
| W1 | 1 | Log Pt | Log Dt, Off peak, Mid peak, On peak | 0.919 | 4.78E-10 | 2.3013621 -14.285719 -25.959484 -12.651630 | 0.00027361 0.02603775 3.8259E-06 0.01371609 | * d=1.23782 du=1.79 dl=0.99 |
| | 2 | Log Pt | Dt, Off peak, Mid peak, On peak | 0.901 | 4.26E-10 | 0.0012432 -14.935227 -26.044454 -13.291040 | 0.00024272 0.01794515 3.3841E-06 0.01053947 | * d=1.32748 du=1.79 dl=0.99 |
| W1 | 3 | Pt | Dt Off peak Mid peak On peak | 0.9188 | 4.26E-10 | 0.00124323 -14.935227 -26.044454 -13.291040 | 0.0002 0.01794515 3.3841E-06 0.01053947 | * d=1.3796 du=1.79 dl=0.99 |

| | | | | | | | | |
|-------------|---|-----------|--|------------|--------------|--|--|--------------------------------------|
| W1 | 4 | Pt | Log Dt Off peak Mid peak On peak | 0.917 8 | 4.78E- 10 | 291.301361 -14.285719 -25.959484 -12.651630 | 0.00027361 0.02603775 3.8260E-06 0.01371609 | * d=1.3129 du=1.79 dl=0.99 |
| W2 | 5 | Log Pt | Log Dt, Off peak, Mid peak, On peak | 0.899 | 3.24E- 09 | 2.0043206 -0.1614348 -0.2108863 -0.0948629 | 0.00091441 0.00309029 3.2756E-05 0.03627891 | * d=1.19637 du=1.79 dl=0.99 |
| | 6 | Log Pt | Dt, Off peak, Mid peak, On peak | 0.904 | 2.16E- 09 | 9.0102E-06 -0.1592220 -0.2121049 -0.1017739 | 0.00059411 0.00287620 2.3731E-05 0.02449651 | * d=1.23782 du=1.79 dl=0.99 |
| W10 | 7 | Log Pt | Log Dt, Off peak, Mid peak, On peak | 0.907 | 2.53E- 10 | 1.833776 -0.052795 -0.061588 -0.010865 | 2.7953E-05 0.04836555 0.02534380 0.00835644 | * d=1.19637 du=1.79 dl=0.99 |
| W12, W16 | 8 | Log Pt | Log Dt, Off peak, Mid peak, On peak | 0.837 | 4.86E- 08 | 1.70482952 -0.1609860 -0.1858552 -0.0823610 | 0.04912460 0.02447336 0.00091076 0.01664629 | * d=1.3338 du=1.79 dl=0.99 |
| W17 | 9 | Log Pt | Log Dt, Off peak, Mid peak, On peak | 0.709 | 1.08E- 05 | 3.14270506 0.06742078 -0.0038290 -0.0018451 | 1.2431E-06 0.00325183 0.00517815 0.00977066 | * d=1.50811 du=1.79 dl=0.99 |

4.2.2 Regression Models for Summer Season

Table 17 depicts the significant regression models for summer season for each significant cluster found in section 3.6 and 3.7 for summer season.

Table 17: Regression Models for Summer Season

| Clu. Type | Reg. Mod | Dep. Var. | Ind. Var. | Adj. R Squ. | Sig. | B Coe. | P Value | Durbin Watson Test |
|-----------|----------|-----------|-------------------------------------|-------------|----------|--|--|--------------------------------------|
| S1 | 1 | Log Pt | Log Dt, Off peak, Mid peak, On peak | 0.834 | 3.44E-07 | 2.96052953 -0.1004265 -0.2080328 -0.1469958 | 0.00500161 0.00657479 0.00056359 0.01542864 | * d=1.67509 du=1.79 dl=0.99 |
| S6,S7 | 2 | Log Pt | Log Dt, Off peak, Mid peak, On peak | 0.9014 | 1.6E-10 | 2.05360275 0.04670751 -0.0528751 -0.0648154 | 2.8021E-06 0.00279769 0.00175899 0.01771372 | * d=1.61127 du=1.79 dl=0.99 |
| S9 | 3 | Log Pt | Log Dt, Off peak, Mid peak, On peak | 0.903 | 1.6E-10 | 2.05360275 0.04670751 -0.0528751 -0.0648154 | 2.8021E-06 0.00279769 0.00175899 0.01771372 | * d=1.12769 du=1.79 dl=0.99 |
| S16 | 4 | Log Pt | Log Dt, Off peak, Mid peak, On peak | 0.902 | 1.67E-10 | 1.05880037 -0.0180840 0.01108163 0.06440308 | 5.5381E-06 0.04549675 0.00586719 0.01034916 | * d=1.26035 du=1.79 dl=0.99 |

- Clu. Type: cluster type
- Reg. Model: regression model
- Dep. Variable: dependent variable
- Ind. Variable: independent variable
- Sig.: significance level
- B Coe.: B Coefficient
- Adj. R Squ.: adjusted R square

In regression models, the adjusted R-square value represents a statistical measure that reflects how much the independent variable in the defined regression model has explained the variation in the dependent variable. It is also known as the coefficient of determination for multiple regression. Usually, for multiple regression models the F-test is used to check the overall significance level of the model instead of a t-test that can evaluate only a single regression coefficient of an independent variable at a time. Hence, for multiple regression, the most suitable test is the F-test to check the significance of the model between the variables. On the other hand, the t-test is used to evaluate the significance of the each independent variables' regression coefficient in the model. The p-value for every coefficient is the result of the null hypothesis that shows whether the null hypothesis (i.e., the coefficient is not significantly different than 0) can be rejected or not. If the p-value (<5%) then we can reject the null hypothesis with 95% confidence level, because variation in the independent variable is interconnected with the variation of the dependent variable. On the other hand, if the p-value is higher than 5%, then this test recommends that there is no statistical evidence that links the variation in independent variable(s) to variation of the dependent variable.

In Table 15 and Table 16, we have evaluated and presented different regression models where all coefficients of independent variables are significant (non-significant models or models with non-significant coefficients are not presented but included in Appendix A and Appendix B).

In some of the models presented, the coefficient of the regression for “log Dt” independent variable represents the “price elasticity of demand”, denoted by β , when the dependent variable is “log Pt” and can be computed as:

$$B = -\frac{1}{b}$$

In this equation, “b” represents the value of the coefficient of the regression for “log Dt” independent variable.

In regression analysis, the Durbin-Watson test statistic is used to detect whether autocorrelation exist or not. In other words, whether residuals from multiple linear regression model are correlated or not. In Table 15 and Table 16, the column, (*)

sign represents a positive correlation between the hourly price and load levels of electricity in the regression model. In our case, price is directly proportional to demand. If the demand for the electricity is increased then the price of the electricity is increased as well. In all the cases, the d-value which depicts the Durbin-Watson test statistic is between upper and lower limit values of the test which means that the price and load are positively correlated.

4.3 Discussion of the Results of Multiple Linear Regression Analyses of Clusters

Table 16 and Table 17 represent the significance level, adjusted R-square and Durbin-Watson test values. Note that all the regression coefficients in the models are significant. It can be observed easily that the winter season cluster W1 has the highest adjusted R-square value. The logarithmic model does not affect too much in W1 case because we can see that there is only a slight R-square difference by using the log. MAPE and change in revenue can be observed in section 3.4.5 and Table 7. W1 cluster gives the lowest change in revenue and the lowest MAPE as compared to all other winter season clusters.

Next, we compare the results for the summer season. S6, S7, S9, and S16 clusters have R-square value very close to each other. Change in revenue and MAPE can be observed from Table 7 and Table 13 in section 3.4.5 and section 3.4.8. Hence, S6, S7 and S9 clusters give the lowest MAPE as compared to the S16, but S9 cluster also gives the lowest change in revenue as compared to all other summer season clusters.

Chapter 5: Summary and Future Research

5.1 Summary

This thesis presents a novel methodology to determine the block of hours and prices for TOU pricing scheme using different clustering methods, namely agglomerative hierarchical and k-means algorithms. We have illustrated our approach using 2014-2015 winter and summer seasons' hourly electricity prices and demands from the PJM market. Based on our detailed descriptive analyses and observations, we propose three clusters (off-peak, mid-peak and on-peak) for TOU pricing scheme. The median of RTPs in the same cluster for off-peak, mid-peak and on-peak hours determines the TOU pricing scheme (hours in each cluster and price for it). We have also compared the performance of TOU pricing schemes from the clustering algorithms by several measures, MAPE and change in overall revenue in comparison to RTP scheme.

Furthermore, we have analysed the strength of the relationship among price and demand variables for different TOU pricing clusters. For this purpose, we have performed multiple regression analyses, where dependent variable is the mean of RTPs within the TOU block of hours and independent variables are mean load of the TOU block of hours, as well as categorical variables for off-peak, mid-peak and on-peak hours for each TOU scheme. Using Minitab software, different regression models (using logarithms and lag of prices and demands) are built. Adjusted coefficient of determination (adjusted R^2), significance of coefficients and the overall model are reported. The significant models (with 95% confidence) are reported and the ones with higher adjusted R^2 values are offered for TOU pricing schemes. Moreover, for auto-correlation, Durbin-Watson statistics are calculated and positive correlation among dependent and independent variables are reported.

Different clustering methods (by observation, hierarchical and k-means) are applied to cluster hours for winter and summer seasons based on prices and loads. By observing the results of the winter and summer seasons, it is verified that clustering by observation is better than the other clustering techniques. In winter season, W1

cluster gives the best results because it has the highest R^2 value, lowest percent error if the pricing scheme is shifted from RTP to TOU pricing scheme. For summer season, it is the S6, S7 and S9 clusters that are more suitable.

It is necessary to ensure that the customers are showing much more knowledge about and response to their electricity bills before considering shifting to a TOU pricing scheme. Smart meters and connected power utilities have the potential to provide contribution to this serious problem, but on the other hand education will be critical to ensure customers can switch to the beneficial pricing schemes for them. TOU pricing is progressively seen as a viable DSM alternative by policy-makers and particularly, in those countries where the implementation of smart metering technologies has arrived (or will reach in the forthcoming years) double-digit penetration (Torriti et al., 2010).

5.2 Future Research

There are several ways in which this research can be extended to conduct a more thorough investigation of the hourly electricity price and demand levels. An outlier analysis can be performed to eliminate the adverse effects of the outliers in the data and to avoid the sparsity within the statistics or in the boxplots (Ratcliff, 1993). We can also include additional independent variables (e.g., other factors as dummy variables such as temperature, humidity and precipitation forecast) in the multiple regression analyses. Moreover, an autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) models can be used to avoid the adverse effects of autocorrelation among the different variables (Chujai et al., 2013). Furthermore, other clustering method such as expectation-maximization algorithms or fuzzy c-means clustering (Hathaway and Bezdek, 2001) can be employed to investigate other possible clusters for TOU pricing schemes.

References

1. Azevedo, F. and Vale, Z.A., 2006. Forecasting electricity prices with historical statistical information using neural networks and clustering techniques. In IEEE PES Power Systems Conference and Exposition. (PSC06), pp. 44-50.
2. Alsabti, K., Ranka, S., and Singh, V., 1997. An efficient k-means clustering algorithm. Electrical Engineering and Computer Science. Paper 43.
3. Albadi, M.H. and El-Saadany, E.F., 2008. A summary of demand response in electricity markets. Electric power systems research, 78(11), pp.1989-1996.
4. Aigner, D.J., 1984. The welfare econometrics of peak-load pricing for electricity: Editor's Introduction. Journal of Econometrics, 26(1), pp.1-15.
5. Anderberg, M. R., 1973. Cluster analysis for applications. Academic Press New York
6. Anderson, A.J.B., 1971. Numeric examination of multivariate soil samples. Journal of the International Association for Mathematical Geology, 3(1), pp.1-14.
7. Bhattacharya, K., Bollen, M. and Daalder, J.E., 2012. Operation of restructured power systems. Springer Science & Business Media.
8. Bobric, E.C., Cartina, G. and Grigoras, G., 2009. Clustering techniques in load profile analysis for distribution stations. Advances in Electrical and Computer Engineering, 9(1), pp.63-66.
9. Bushnell, J.B., Mansur, E.T. and Saravia, C., 2008. Vertical arrangements, market structure, and competition: An analysis of restructured US electricity markets. The American Economic Review, 98(1), pp.237-266.
10. Borenstein, S., Jaske, M. and Rosenfeld, A., 2002. Dynamic pricing, advanced metering, and demand response in electricity markets. Center for the Study of Energy Markets.
11. Bilmes, J., Vahdat, A., Hsu, W. and Im, E.J., 1997. Empirical observations of probabilistic heuristics for the clustering problem. Technical Report TR-97-018, International Computer Science Institute, University of California, Berkeley, CA.

12. Birks, H.J.B., Webb, T. and Berti, A.A., 1975. Numerical analysis of pollen samples from central Canada: a comparison of methods. *Review of Palaeobotany and Palynology*, 20(3), pp.133-169.
13. Chao, H.P. and Huntington, H.G. eds., 2013. *Designing competitive electricity markets* (Vol. 13). Springer Science & Business Media.
14. Chujai, P., Kerdprasop, N. and Kerdprasop, K., 2013, March. Time series analysis of household electric consumption with ARIMA and ARMA models. In *Proc. IMECS Conf.*, Hong Kong.
15. Çelebi, E. and Fuller, J.D., 2007. A model for efficient consumer pricing schemes in electricity markets. *IEEE Transactions on Power Systems*, 22(1), pp.60-67.
16. Chick, S.E. and Ng, S.H., 2002, December. Simulation input analysis: joint criterion for factor identification and parameter estimation. In *Proceedings of the 34th Conference on Winter Simulation: exploring new frontiers*, pp. 400-406.
17. Clarke, E.H., 1971. Multipart pricing of public goods. *Public choice*, 11(1), pp.17-33.
18. Charles River Association, 2005. *Primer on demand-side management with an emphasis on price-responsive programs*. Prepared for The World Bank by Charles River Associates. Tech. Rep.
19. Dewees, D.N., 2001. *Price and Environment in Electricity Restructuring*. Law and Economics Programme, Faculty of Law, University of Toronto.
20. Everitt, B., 1974. *Cluster analysis*. London: Heinemann Educational [for] the Social Science Research Council.
21. Faruqui, A., Hledik, R. and Lessem, N., 2014. Smart By Default. Time varying rates from the get to go *Fortnightly Magazine*. Public Utilities Fortnightly.
22. Filippini, M., 2011. Short-and long-run time-of-use price elasticities in Swiss residential electricity demand. *Energy Policy*, 39(10), pp.5811-5817.
23. Faria, P. and Vale, Z., 2011. Demand response in electrical energy supply: An optimal real time pricing approach. *Energy*, 36(8), pp.5374-5384.
24. Faruqui, A. and George, S., 2005. Quantifying customer response to dynamic pricing. *The Electricity Journal*, 18(4), pp.53-63.

25. Faruqui, A. and Sergici, S., 2010. Household response to dynamic pricing of electricity: a survey of 15 experiments. *Journal of regulatory Economics*, 38(2), pp.193-225.
26. Faruqui, A. and George, S.S., 2002. The value of dynamic pricing in mass markets. *The Electricity Journal*, 15(6), pp.45-55.
27. Gower, J.C., 1967. A comparison of some methods of cluster analysis. *Biometrics*, pp.623-637.
28. Gordon, A. D., 1981. *Classification: methods for the exploratory analysis of multivariate data*. London; New York: Chapman and Hall
29. Gujarati, D. N., and Porter, D. C., 2009. *Basic econometrics*. Boston, Mass, McGraw-Hill Irwin.
30. Hogan, W.W., 2014. Time-of-Use Rates and Real-Time Prices. Working Paper, August.
31. Hernández, L., Baladrón, C., Aguiar, J.M., Carro, B. and Sánchez-Esguevillas, A., 2012. Classification and clustering of electricity demand patterns in industrial parks. *Energies*, 5(12), pp.5215-5228.
32. Halkidi, M., Batistakis, Y. and Vazirgiannis, M., 2002. Cluster validity methods: part I. *ACM Sigmod Record*, 31(2), pp.40-45.
33. Hathaway, R.J. and Bezdek, J.C., 2001. Fuzzy c-means clustering of incomplete data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 31(5), pp.735-744.
34. Hardy, A., 1996. On the number of clusters. *Computational Statistics & Data Analysis*, 23(1), pp.83-96.
35. Hansen, L.K. and Larsen, J., 1996, June. Unsupervised learning and generalization. In *IEEE International Conference on Neural Networks* (pp. 25-30).
36. Huth, R., Nemesova, I. and Klimperová, N., 1993. Weather categorization based on the average linkage clustering technique: An application to European mid-latitudes. *International Journal of Climatology*, 13(8), pp.817-835.
37. Jamasb, T. and Pollitt, M., 2005. Electricity market reform in the European Union: review of progress toward liberalization & integration. *The Energy Journal*, pp.11-41.

38. Joskow, P.L. and Tirole, J., 2004. Retail electricity competition (No. w10473). National Bureau of Economic Research.
39. Jain, A.K., Murty, M.N. and Flynn, P.J., 1999. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), pp.264-323.
40. Kirschen, D.S., 2003. Demand-side view of electricity markets. *IEEE Transactions on Power Systems*, 18(2), pp.520-527.
41. King, C.S., 2001. The economics of real-time and time-of-use pricing for residential consumers. American Energy Institute, Tech. Rep.
42. Li, R., Wang, Z., Le Blond, S. and Li, F., 2014, July. Development of time-of-use price by clustering techniques. In *PES General Meeting| Conference & Exposition, 2014 IEEE* (pp. 1-5). IEEE.
43. Leung, Y., Zhang, J.S. and Xu, Z.B., 2000. Clustering by scale-space filtering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12), pp.1396-1410.
44. Lindeberg, T., 1994. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2), pp.225-270.
45. Murtagh, F. and Legendre, P., 2011. Ward's hierarchical clustering method: clustering criterion and agglomerative algorithm. *ArXiv preprint arXiv: 1111.6285*.
46. Mohsenian-Rad, A.H. and Leon-Garcia, A., 2010. Optimal residential load control with price prediction in real-time electricity pricing environments. *Smart Grid, IEEE Transactions on*, 1(2), pp.120-133.
47. Mahmoudi-Kohan, N., Moghaddam, M.P. and Bidaki, S.M., 2009. Evaluating performance of WFA K-means and modified follow the leader methods for clustering load curves. In *IEEE PES Power Systems Conference and Exposition (PSCE'09)*, pp. 1-5.
48. Martínez-Álvarez, F., Troncoso, A., Riquelme, J.C. and Riquelme, J.M., 2007. Partitioning-clustering techniques applied to the electricity price time series. In *Intelligent Data Engineering and Automated Learning-IDEAL 2007*, pp. 990-999. Springer Berlin Heidelberg.
49. Meshram, R.A., Deorankar, A. and Chatur, D., 2013. Predicting Electricity Load Pattern of Customers Using Clustering Technique. *Journal of Information, Knowledge and Research in Computer Engineering*, 2(2),458-462

50. Murtagh, F., 1983. A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26(4), pp.354-359.
51. PJM, 2016. Metered Load Data [Online] Available at: <http://www.pjm.com/markets-and-operations/ops-analysis/historical-load-data.aspx> [Accessed November. 2015].
52. PJM, 2016. Locational Marginal Pricing [Online] Available at: <https://dataminer.pjm.com/dataminerui/pages/public/lmp.jsf> [Accessed November. 2015].
53. Pham, D.T., Dimov, S.S. and Nguyen, C.D., 2005. Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1), pp.103-119.
54. Pelleg, D. and Moore, A.W., 2000, June. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In *ICML (Vol. 1)*.
55. Pielou, E.C., 1984. *The interpretation of ecological data: a primer on classification and ordination*. John Wiley & Sons.
56. Räsänen, T. and Kolehmainen, M., 2009. Feature-based clustering for electricity use time series data. In *Adaptive and Natural Computing Algorithms*, pp. 401-412. Springer Berlin Heidelberg.
57. Ratcliff, R., 1993. Methods for dealing with reaction time outliers. *Psychological bulletin*, 114(3), p.510.
58. Spees, K. and Lave, L.B., 2007. Demand response and electricity market efficiency. *The Electricity Journal*, 20(3), pp.69-85.
59. Sykes, A. O., 1993. *An Introduction to Regression Analysis*, Coase-Sandor Institute for Law & Economics Working Paper No. 20.
60. Sneath, P.H. and Sokal, R.R., 1973. *Numerical taxonomy. The principles and practices of numerical classification*. WF Freeman and Co., San Francisco, 573.
61. Torriti, J., Hassan, M.G. and Leach, M., 2010. Demand response experience in Europe: Policies, programmes and implementation. *Energy*, 35(4), pp.1575-1583.
62. Turi, R.H., 2001. *Clustering-based colour image segmentation*. PhD thesis: Monash University.
63. Ultsch, A., Vetter, C. and Vetter, C., 1995. Self-Organizing-Feature-Maps versus statistical clustering methods: a benchmark. *Fachbereich Mathematik*.

64. Walawalkar, R., Blumsack, S., Apt, J. and Fernands, S., 2008. An economic welfare analysis of demand response in the PJM electricity market. *Energy Policy*, 36(10), pp.3692-3702.
65. Waters, G., 2004. Payable on Demand. *Public Utilities Fortnightly*, 142(2), pp.55-57.
66. Ward Jr, J.H., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301), pp.236-244.
67. Xu, R. and Wunsch, D., 2005. Survey of clustering algorithms. *Neural Networks, IEEE Transactions on*, 16(3), pp.645-678.

Appendices

All the regression model results for winter season used for the analysis in this thesis are provided in the appendices in tabular form. Moreover, the R scripts used to analyse the data are also provided.

Appendix A

Table A.1 with log model of energy price of winter season (2014-2015)

| Reg. Mod | Dep. Var. | Ind. Var. | Adj. R Squ. | Sig. | B Coe. | P Value | Durbin Watson Test |
|----------|-----------|---|-------------|----------|--|--|--------------------|
| 1 | Log Pt | Pt-1 Off peak Mid peak On peak | 0.8234 | 6.18E-07 | 0.00191457 -0.2586918 -0.2078586 -0.0632317 | | |
| 2 | Log Pt | Dt-1 Off peak Mid peak On peak | 0.8368 | 2.97E-07 | 6.01816E-06 -0.243886896 -0.266416202 -0.111676763 | | |
| 3 | Log Pt | Dt Off peak Mid peak On peak | 0.901 | 4.73E-10 | 1.23287E-05 -0.098301896 -0.205086078 -0.117800746 | 5.36584E-05 0.042862741 1.10882E-05 0.008901584 | * |
| 4 | Log Pt | Log Pt-1 Off peak Mid peak On peak | 0.828 | 4.85E-07 | 0.25043769 -0.2516725 -0.2078034 -0.0631536 | | |
| 5 | Log Pt | Log Dt-1 Off peak Mid peak On peak | 0.8346 | 3.36E-07 | 1.40347549 -0.2411887 -0.26553973 -0.10936173 | | |
| 6 | Log Pt | Log Dt Off peak Mid peak On peak | 0.919 | 4.99E-10 | 2.899442482 -0.09117969 -0.204255225 -0.111766962 | 5.67486E-05 0.039197151 1.20811E-05 0.011694975 | * |
| 7 | Log Pt | Pt-1 Dt-1 Off peak Mid peak On peak | 0.8385 | 1.46E-06 | 0.000678798 4.99885E-06 -0.243238914 -0.257755324 -0.109886093 | | |
| 8 | Log Pt | Pt-1 Dt Off peak Mid peak On peak | 0.9179 | 3.71E-09 | 9.46517E-05 1.2221E-05 -0.099010688 -0.205386618 -0.118733388 | | |

| | | | | | | | |
|----|-----------|---|--------|----------|---|--|--|
| 9 | Log Pt | Pt-1 Log Pt-1 Off peak Mid peak On peak | 0.8487 | 8.21E-07 | -0.019371801 2.540304185 -0.204101265 -0.200681561 -0.028675317 | | |
| 10 | Log Pt | Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.837 | 1.58E-06 | 0.000798584 1.123502867 -0.240971761 -0.255510779 -0.107698367 | | |
| 11 | Log Pt | Pt-1 Log Dt Off peak Mid peak On peak | 0.9176 | 3.85E-09 | 0.000183069 2.852245969 -0.092554345 -0.204852573 -0.113725432 | | |
| 12 | Log Pt | Dt-1 Dt Off peak Mid peak On peak | 0.9183 | 3.58E-09 | -7.69796E-07 1.2841E-05 -0.095002696 -0.196894294 -0.110879867 | | |
| 13 | Log Pt | Dt-1 Log Pt-1 Off peak Mid peak On peak | 0.8399 | 1.35E-06 | 4.55519E-06 0.1109524 -0.240968815 -0.253514374 -0.106791852 | | |
| 14 | Log Pt | Dt-1 Log Dt-1 Off peak Mid peak On peak | 0.8442 | 1.06E-06 | 4.81682E-05 -10.18227016 -0.271462911 -0.25635736 -0.107597979 | | |
| 15 | Log Pt | Dt-1 Log Dt Off peak Mid peak On peak | 0.9176 | 3.87E-09 | -4.59582E-07 2.969532734 -0.089150307 -0.199341956 -0.107432671 | | |
| 16 | Log Pt | Dt Log Pt-1 Off peak Mid peak On peak | 0.9182 | 3.59E-09 | 1.20132E-05 0.0320082 -0.099767214 -0.205844002 -0.119935379 | | |
| 17 | Log Pt | Dt Log Dt-1 Off peak Mid peak On peak | 0.9184 | 3.54E-09 | 1.2906E-05 -0.212748567 -0.095230706 -0.195501238 -0.109733813 | | |
| 18 | Log Pt | Dt Log Dt Off peak Mid peak On peak | 0.9179 | 3.73E-09 | 1.3564E-05 -0.291581763 -0.099084541 -0.205168129 -0.118374145 | | |

| | | | | | | | |
|----|--------|---|--------|----------|--|--|--|
| 19 | Log Pt | Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.8388 | 1.45E-06 | 0.124900507 1.018172237 -0.238672689 -0.251198909 -0.104424919 | | |
| 20 | Log Pt | Log Pt-1 Log Dt Off peak Mid peak On peak | 0.918 | 3.68E-09 | 0.04175268 2.805959988 -0.093118631 -0.205275252 -0.114847612 | | |
| 21 | Log Pt | Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9176 | 3.87E-09 | -4.59582E-07 2.969532734 -0.089150307 -0.199341956 -0.107432671 | | |
| 22 | Log Pt | Pt-1 Dt-1 Dt Off peak Mid peak On peak | 0.9186 | 2.36E-08 | 0.000308254 -1.19762E-06 1.27747E-05 -0.095477438 -0.193320333 -0.11007081 | | |
| 23 | Log Pt | Pt-1 Dt-1 Log Pt-1 Off peak Mid peak On peak | 0.8534 | 3.08E-06 | -0.016422956 2.98615E-06 2.100293097 -0.204325945 -0.231731443 -0.062530734 | | |
| 24 | Log Pt | Pt-1 Dt-1 Log Dt-1 Off peak Mid peak On peak | 0.8444 | 5.02E-06 | 0.00021792 4.56307E-05 -9.648330134 -0.269808847 -0.254104357 -0.10723699 | | |
| 25 | Log Pt | Pt-1 Dt-1 Log Dt Off peak Mid peak On peak | 0.918 | 2.5E-08 | 0.00036028 -9.66687E-07 2.953987754 -0.089616402 -0.195096213 -0.106504468 | | |
| 26 | Log Pt | Pt-1 Dt Log Pt-1 Off peak Mid peak On peak | 0.9309 | 6.02E-09 | -0.015195554 1.15261E-05 1.837073977 -0.068611445 -0.20033697 -0.090587526 | | |

| | | | | | | | |
|----|-----------|---|--------|----------|--|--|--|
| 27 | Log Pt | Pt-1 Dt Log Dt-1 Off peak Mid peak On peak | 0.9188 | 2.32E-08 | 0.00031632 1.28161E-05 -0.312392937 -0.096161001 -0.192016405 -0.109072362 | | |
| 28 | Log Pt | Pt-1 Dt Log Dt Off peak Mid peak On peak | 0.9179 | 2.52E-08 | 9.43748E-05 1.22622E-05 -0.009665195 -0.099034557 -0.205388459 -0.118749666 | | |
| 29 | Log Pt | Pt-1 Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.8525 | 3.23E-06 | -0.01668881 2.144329613 0.639339979 -0.202526917 -0.228917237 -0.059366024 | | |
| 30 | Log Pt | Pt-1 Log Pt-1 Log Dt Off peak Mid peak On peak | 0.93 | 6.69E-09 | -0.014811878 1.801663136 2.6842069 -0.063624975 -0.199939519 -0.086242126 | | |
| 31 | Log Pt | Pt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9182 | 2.46E-08 | 0.000377353 -0.266506704 2.968277528 -0.089999121 -0.19342668 -0.105231601 | | |
| 32 | Log Pt | Dt-1 Dt Log Pt-1 Off peak Mid peak On peak | 0.9194 | 2.17E-08 | -1.61081E-06 1.2727E-05 0.068356085 -0.094527596 -0.189563305 -0.107877427 | | |
| 33 | Log Pt | Dt-1 Dt Log Dt-1 Off peak Mid peak On peak | 0.9197 | 2.1E-08 | 1.85452E-05 1.25447E-05 -4.628126802 -0.110972582 -0.193926632 -0.109044335 | | |
| 34 | Log Pt | Dt-1 Dt Log Dt Off peak Mid peak On peak | 0.9183 | 2.42E-08 | -8.91348E-07 1.86807E-05 -1.35940573 -0.098130577 -0.195983337 -0.112460341 | | |

| | | | | | | | |
|----|--------|---|--------|----------|---|--|--|
| 35 | Log Pt | Dt-1 Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.845 | 4.86E-06 | 4.25345E-05 0.058659446 -9.008158925 -0.266740377 -0.250696168 -0.105485696 | | |
| 36 | Log Pt | Dt-1 Log Pt-1 Log Dt Off peak Mid peak On peak | 0.9189 | 2.29E-08 | -1.37449E-06 0.07354509 2.944399951 -0.08852567 -0.191357631 -0.104230615 | | |
| 37 | Log Pt | Dt Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.9196 | 2.13E-08 | 1.27518E-05 0.068474409 -0.404678376 -0.095594772 -0.188475744 -0.107022852 | | |
| 38 | Log Pt | Dt Log Pt-1 Log Dt Off peak Mid peak On peak | 0.9196 | 2.13E-08 | 1.27518E-05 0.068474409 -0.404678376 -0.095594772 -0.188475744 -0.107022852 | | |
| 39 | Log Pt | Dt Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9185 | 2.39E-08 | 1.86874E-05 -0.238220741 -1.348427363 -0.098482357 -0.194733103 -0.111419669 | | |
| 40 | Log Pt | Log Pt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9182 | 2.44E-08 | 9.86727E-06 0.033633774 0.502789307 -0.098492077 -0.205741009 -0.119055046 | | |
| 41 | Log Pt | Pt-1 Dt-1 Dt Log Pt-1 Off peak Mid peak On peak | 0.9191 | 2.24E-08 | 0.074826086 -0.358286436 2.955237792 -0.089258048 -0.189870232 -0.103074704 | | |
| 42 | Log Pt | Pt-1 Dt-1 Dt Log Dt-1 Off peak Mid peak On peak | 0.9358 | 2.22E-08 | -0.018084827 -3.45323E-06 1.29635E-05 2.258210306 -0.051454862 -0.16438746 -0.059157618 | | |

| | | | | | | | |
|----|-----------|---|--------|----------|--|--|--|
| 43 | Log Pt | Pt-1 Dt-1 Dt Log Dt Off peak Mid peak On peak | 0.9198 | 1.27E-07 | 0.000106347 1.7322E-05 1.25383E-05 -4.370382909 -0.110246992 -0.19285889 -0.108867433 | | |
| 44 | Log Pt | Dt-1 Dt Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.9186 | 1.41E-07 | 0.000292631 -1.25868E-06 1.6753E-05 -0.925315492 -0.097582453 -0.192881393 -0.111187604 | | |
| 45 | Log Pt | Dt-1 Dt Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9202 | 1.21E-07 | 1.40136E-05 1.25257E-05 0.047650938 -3.682768611 -0.107379323 -0.189422395 -0.107326271 | | |
| 46 | Log Pt | Dt Log Pt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9194 | 1.31E-07 | -1.6583E-06 1.56982E-05 0.067193048 -0.691224629 -0.096126131 -0.189224838 -0.108732144 | | |
| 47 | Log Pt | Pt-1 Dt-1 Dt Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.9196 | 1.29E-07 | 1.50085E-05 0.067397541 -0.411591738 -0.525760379 -0.096856886 -0.18828673 -0.107722812 | | |
| 48 | Log Pt | Pt-1 Dt Log Pt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9368 | 1.2E-07 | -0.018206643 1.4012E-05 1.274E-05 2.249799695 -4.119580133 -0.065540798 -0.164060257 -0.058212924 | | |

| | | | | | | | |
|----|-----------|---|--------|----------|---|--|--|
| 49 | Log Pt | Dt-1 Dt Log Pt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9368 | 1.2E-07 | -0.018206643 1.4012E-05 1.274E-05 2.249799695 -4.119580133 -0.065540798 -0.164060257 -0.058212924 | | |
| 50 | Log Pt | Pt-1 Dt-1 Dt Log Pt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9375 | 6.22E-07 | -0.019270067 5.75813E-06 3.36281E-05 2.380790528 -2.277905696 -4.832629461 -0.067757156 -0.160283972 -0.061599394 | | |

Note: “*” represents the significant models where price and demand variables are positively correlated.

Appendix B

Table B.1 without log model of energy price of winter season (2014-2015)

| Reg. Mod | Dep. Var. | Ind. Var. | Adj. R Squ. | Sig. | B Coe. | P Value | Durbin Watson Test |
|----------|-----------|---|-------------|----------|---|--|--------------------|
| 1 | Pt | Pt-1 Off peak Mid peak On peak | 0.8448 | 1.86E-07 | 0.172096356 -31.29563235 -26.24966911 -7.408873602 | | |
| 2 | Pt | Dt-1 Off peak Mid peak On peak | 0.857 | 8.66E-08 | 0.000593879 -29.68482142 -32.08790733 -12.49409922 | | |
| 3 | Pt | Dt Off peak Mid peak On peak | 0.9188 | 4.26E-10 | 0.001243234 -14.9352274 -26.04445367 -13.29103952 | 0.0002 0.01794515 3.38413E-06 0.010539473 | * |
| 4 | Pt | Log Pt-1 Off peak Mid peak On peak | 0.848 | 1.53E-07 | 22.87317935 -30.62991781 -26.25444189 -7.451791113 | | |
| 5 | Pt | Log Dt-1 Off peak Mid peak On peak | 0.8556 | 9.48E-08 | 139.0801207 -29.4042029 -32.02822309 -12.299233 | | |
| 6 | Pt | Log Dt Off peak Mid peak On peak | 0.9178 | 4.78E-10 | 291.3013609 -14.28571858 -25.95948414 -12.65162997 | 0.000273607 0.026037751 3.82595E-06 0.013716094 | * |
| 7 | Pt | Pt-1 Dt-1 Off peak Mid peak On peak | 0.8574 | 4.89E-07 | 0.817117073 0.222608375 9.82798E-05 0.0002454 0.111728776 | | |
| 8 | Pt | Pt-1 Dt Off peak Mid peak On peak | 0.9189 | 3.35E-09 | -0.015706737 0.001261121 -14.81760874 -25.99458126 -13.13627458 | | |
| 9 | Pt | Pt-1 Log Pt-1 Off peak Mid peak On peak | 0.8653 | 2.97E-07 | -2.060588535 266.4474642 -25.56974172 -25.496888 -3.784320138 | | |

| | | | | | | | |
|----|----|---|--------|----------|---|--|--|
| 10 | Pt | Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.8563 | 5.23E-07 | 0.052085379 120.8197035 -29.39005003 -31.37411262 -12.19074479 | | |
| 11 | Pt | Pt-1 Log Dt Off peak Mid peak On peak | 0.9178 | 3.76E-09 | -0.005622349 292.7508443 -14.24350065 -25.94113857 -12.59148209 | | |
| 12 | Pt | Dt-1 Dt Off peak Mid peak On peak | 0.9192 | 3.23E-09 | -9.76725E-05 0.001308234 -14.51662194 -25.0050726 -12.4129121 | | |
| 13 | Pt | Dt-1 Log Pt-1 Off peak Mid peak On peak | 0.8581 | 4.69E-07 | 0.000490209 7.862416766 -29.47803751 -31.17364542 -12.1479399 | | |
| 14 | Pt | Dt-1 Log Dt-1 Off peak Mid peak On peak | 0.861 | 3.92E-07 | 0.004194976 -869.9242654 -32.04078375 -31.22852811 -12.1456275 | | |
| 15 | Pt | Dt-1 Log Dt Off peak Mid peak On peak | 0.918 | 3.69E-09 | -6.22885E-05 300.8008955 -14.01067044 -25.29357459 -12.06419094 | | |
| 16 | Pt | Dt Log Pt-1 Off peak Mid peak On peak | 0.9188 | 3.38E-09 | 0.001240013 0.326790292 -14.95018768 -26.05219175 -13.31283323 | | |
| 17 | Pt | Dt Log Dt-1 Off peak Mid peak On peak | 0.9193 | 3.2E-09 | 0.0013115 -25.15965535 -14.57202833 -24.91095009 -12.33704374 | | |
| 18 | Pt | Dt Log Dt Off peak Mid peak On peak | 0.9192 | 3.23E-09 | 0.002719477 -348.4759151 -15.87058395 -26.14251492 -13.97632244 | | |
| 19 | Pt | Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.8572 | 4.96E-07 | 9.239283615 110.5780258 -29.2180815 -30.96738667 -11.93404141 | | |

| | | | | | | | |
|----|----|---|--------|----------|---|--|--|
| 20 | Pt | Log Pt-1 Log Dt Off peak Mid peak On peak | 0.9179 | 3.75E-09 | 1.449935672 288.0550159 -14.35305173 -25.99490641 -12.75861098 | | |
| 21 | Pt | Log Dt-1 Log Dt Off peak Mid peak On peak | 0.918 | 3.69E-09 | -6.22885E-05 300.8008955 -14.01067044 -25.29357459 -12.06419094 | | |
| 22 | Pt | Pt-1 Dt-1 Dt Off peak Mid peak On peak | 0.9179 | 3.75E-09 | 1.449935672 288.0550159 -14.35305173 -25.99490641 -12.75861098 | | |
| 23 | Pt | Pt-1 Dt-1 Log Pt-1 Off peak Mid peak On peak | 0.9192 | 2.21E-08 | 0.002277322 -0.000100833 0.001307744 -14.52012924 -24.97866881 -12.40693493 | | |
| 24 | Pt | Pt-1 Dt-1 Log Dt-1 Off peak Mid peak On peak | 0.8693 | 1.2E-06 | -1.740735281 0.0003239 218.7206495 -25.5941122 -28.86478485 -7.456525794 | | |
| 25 | Pt | Pt-1 Dt-1 Log Dt Off peak Mid peak On peak | 0.861 | 1.99E-06 | -0.001522557 0.004212705 -873.6547806 -32.05234031 -31.24426932 -12.14814964 | | |
| 26 | Pt | Pt-1 Dt Log Pt-1 Off peak Mid peak On peak | 0.9295 | 7.1E-09 | -1.630231607 0.00118775 193.9804904 -11.60768899 -25.46137832 -10.1642941 | | |
| 27 | Pt | Pt-1 Dt Log Dt-1 Off peak Mid peak On peak | 0.9193 | 2.2E-08 | 0.002763909 0.001310715 -26.03031767 -14.58015697 -24.88050067 -32.40177918 | | |

| | | | | | | | |
|----|----|---|--------|----------|--|--|--|
| 28 | Pt | Pt-1 Dt Log Dt Off peak Mid peak On peak | 0.9195 | 2.16E-08 | -0.028095765 0.003107014 -432.4036646 -15.8854645 -26.07692188 -13.86452844 | | |
| 29 | Pt | Pt-1 Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.8687 | 1.25E-06 | -1.764647368 222.7704004 70.52092502 -25.39608692 -28.61135923 -7.169587906 | | |
| 30 | Pt | Pt-1 Log Pt-1 Log Dt Off peak Mid peak On peak | 0.9281 | 8.41E-09 | -1.5934931 190.7848157 274.9565618 -11.18006195 -25.42087689 -9.681172489 | | |
| 31 | Pt | Pt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9181 | 2.49E-08 | 0.00927797 -20.43931683 301.6497041 -14.04753172 -25.06484743 -11.94006089 | | |
| 32 | Pt | Dt-1 Dt Log Pt-1 Off peak Mid peak On peak | 0.918 | 2.51E-08 | 0.007811957 -7.3284E-05 300.4638337 -14.02077677 -25.20151416 -12.04406472 | | |
| 33 | Pt | Dt-1 Dt Log Dt-1 Off peak Mid peak On peak | 0.9194 | 2.16E-08 | -0.000140776 0.001302387 3.503409123 -14.49227195 -24.62934231 -12.25903004 | | |
| 34 | Pt | Dt-1 Dt Log Dt Off peak Mid peak On peak | 0.9197 | 2.11E-08 | 0.001150957 0.001289077 -299.1876175 -15.54900334 -24.81322658 -12.29425323 | | |
| 35 | Pt | Dt-1 Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.8611 | 1.97E-06 | 0.003889353 3.182197054 -806.2302904 -31.78459254 -30.921416 -12.03103897 | | |

| | | | | | | | |
|----|----|---|--------|----------|---|--|--|
| 36 | Pt | Dt-1 Log Pt-1 Log Dt Off peak Mid peak On peak | 0.9183 | 2.44E-08 | -0.000112777 4.058493537 299.4139744 -13.97620062 -24.85296964 -11.88748948 | | |
| 37 | Pt | Dt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9183 | 2.44E-08 | -0.000112777 4.058493537 299.4139744 -13.97620062 -24.85296964 -11.88748948 | | |
| 38 | Pt | Dt Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.9201 | 2.03E-08 | -0.000144287 0.003547759 -521.3304969 -15.71616055 -24.65572165 -13.01902218 | | |
| 39 | Pt | Dt Log Pt-1 Log Dt Off peak Mid peak On peak | 0.9195 | 2.14E-08 | 0.001303686 3.470549624 -34.88740524 -14.59048062 -24.55487067 -12.19964165 | | |
| 40 | Pt | Dt Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9201 | 2.02E-08 | 0.003464615 -34.64597287 -502.1797609 -15.78300451 -24.62488201 -12.96488832 | | |
| 41 | Pt | Log Pt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9192 | 2.21E-08 | 0.002814659 -0.865999581 -368.9293144 -15.88583866 -26.12776446 -13.95879066 | | |
| 42 | Pt | Pt-1 Dt-1 Dt Log Pt-1 Off peak Mid peak On peak | 0.9184 | 2.41E-08 | 4.151762549 -29.26907271 300.2497921 -14.03767365 -24.73644219 -11.79686066 | | |
| 43 | Pt | Pt-1 Dt-1 Dt Log Dt-1 Off peak Mid peak On peak | 0.9329 | 3.13E-08 | -1.91090047 -0.000335453 0.001327382 234.8903839 -9.941069655 -21.96918223 -7.111139823 | | |

| | | | | | | | |
|----|----|---|--------|----------|---|--|--|
| 44 | Pt | Pt-1 Dt-1 Dt Log Dt Off peak Mid peak On peak | 0.9197 | 1.27E-07 | -0.013000407 0.001300495 0.001289857 -330.6955051 -15.63770302 -24.94375264 -12.31587857 | | |
| 45 | Pt | Dt-1 Dt Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.9201 | 1.23E-07 | -0.006691962 -0.000135887 0.003591843 -531.2573745 -15.72869519 -24.72665763 -13.04812743 | | |
| 46 | Pt | Dt-1 Dt Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9197 | 1.27E-07 | 0.000956003 0.00128826 2.049979611 -258.5175887 -15.3944186 -24.61945088 -12.22034079 | | |
| 47 | Pt | Dt Log Pt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9202 | 1.22E-07 | 0.0033279 2.504596494 -41.08870634 -471.6081879 -15.72259949 -24.38532485 -12.8275074 | | |
| 48 | Pt | Pt-1 Dt-1 Dt Log Pt-1 Log Dt-1 Off peak Mid peak On peak | 0.9202 | 6.61E-07 | 0.000197867 0.003205755 2.267340805 -87.00967684 -443.8943438 -15.82246441 -24.40865449 -12.79489531 | | |
| 49 | Pt | Pt-1 Dt Log Pt-1 Log Dt-1 Log Dt Off peak Mid peak On peak | 0.9364 | 1.53E-05 | -2.167778958 -0.000437929 0.006104294 264.1267837 -1110.49091 -11.89739567 -21.06781595 -7.792270412 | | |

| | | | | | | | |
|----|----|----------|--------|----------|--------------|--|--|
| 50 | Pt | Pt-1 | 0.9364 | 6.97E-07 | -2.185320586 | | |
| | | Dt-1 | | | -0.001104196 | | |
| | | Dt | | | 0.006524184 | | |
| | | Log Pt-1 | | | 266.9617596 | | |
| | | Log Dt-1 | | | 155.07265 | | |
| | | Log Dt | | | -1206.146819 | | |
| | | Off peak | | | -11.53567599 | | |
| | | Mid peak | | | -21.00249053 | | |
| | | On peak | | | -7.886502872 | | |

- Clu. Type: cluster type
- Reg. Model: regression model
- Dep. Variable: dependent variable
- Ind. Variable: independent variable
- Sig.: significance level
- B Coe.: B Coefficient
- Adj. R Squ.: adjusted R square

Note: “*” represents the significant models where price and demand variables are positively correlated.

Appendix C

```
Price<-read.table
("C:\\Users\\Hussain\\Desktop\\Price.csv",header=T,sep=",
")

boxplot (Price$h1,Price$h2,Price$h3,Price$h4,Price$h5,Price
$h6,Price$h7,Price$h8,Price$h9,Price$h10,Price$h11,Price
$h12,Price$h13,Price$h14,Price$h15,Price$h16,Price$h17,Pr
ice$h18,Price$h19,Price$h20,Price$h21,Price$h22,Price$h23
,Price$h24, ylab='Price', main = 'Boxplot of Winter
Season (Prices)', xlab='Hours')

hclust.fit1 <- hclust(d, method="Single")

Plot(hclust.fit1, main="Agglomerative hierarchical
Clustering",xlab="") # display dendrogram

groups1 <- cutree(hclust.fit1, k=3) # cut tree into 3
clusters

# draw dendrogram with red borders around the 3 clusters
rect.hclust(hclust.fit1, k=3, border="red")

#perform K-means clustering with K=3

fit1 <- kmeans(data,3) # 3 cluster solution

# get cluster means

Aggregate (Price,by=list(fit1$cluster),FUN=mean)

#for visualization

par(mfrow=c(1,2))

clusplot(data,groups1,color=TRUE, shade=TRUE,

labels=0.1, lines=0, main="Agglomerative
clustering by Ward's method")

clusplot(data, fit1$cluster, color=TRUE, shade=TRUE,

labels=0.1, lines=0, main="K-means clustering")
```