PHD

Socio-Economic Data Analytics and Applications in the Smart Grids

Ma, Qiuyang

Award date:
2020

Awarding institution:
University of Bath

Socio-Economic Data Analytics and Applications in the Smart Grids

By

Qiuyang Ma

BEng, SMIEEE

The thesis submitted for the degree of

Doctor of Philosophy

in

The Department of
Electronic and Electrical Engineering
University of Bath

September 2019

Contents

Contents ii
Abstract vi
Acknowledgements viii
List of Figures ix
List of Tables xi
List of Abbreviations xiii
Chapter 1 Introduction 1
1.1 Background and Motivation 2
1.1.1 The New Context: Climate Change and Growing Low Carbon Techniques 2
1.1.2 Motivations: Creating a Dynamic Energy Market 3
1.2 Research Questions: Analysis of the Socio-Economic Data 5
1.2.1 The Research Questions in Technical Field 6
1.2.2 The Research Questions in Application Field 7
1.3 Research Contributions 8
1.3.1 Contributions to the Techniques Field 8
1.3.2 Contributions to the Facilitating of Applications 9
1.4 The Layout of Thesis 9
Chapter 2 Review of Socio-Economic Data 12
2.1 Introduction 13
2.2 Effects of Socio-Economic Data 13
2.2.1 The Digitised Survey Socio-Economic Data 14
2.2.2 The Geo-Demographic Code 17
2.3 Applications and Methodologies in Socio-Economic Data Analytics 18
2.3.1 The Applications for the Socio-Economic Data Analysis 18
2.3.2 The Methodologies for the Socio-Economic Data Analysis 18
2.4 The Limitations of the Existing Research 20

Chapter 3 The Socio-Economic Criteria for the Impact on Wholesale Maret Cost Variation 22

3.1 Introduction 23

3.2 Impact Assessment of HHS on Customer with Different Socio-Economic Status 24

3.2.1 Background of the HHS 24

3.2.2 Problem Statement 26

3.2.3 The KLAM Beam Search Algorithm 28

3.2.4 The Implementation of the KLAM Searching Algorithm 35

3.2.5 The Results Analysis of the Case Study 42

3.2.6 Analysis and Discussion 45

3.2.7 Conclusions 50

3.3 Chapter Summary 50

Chapter 4 The Impact of Nework Cost Variation on Customers' Electricity Bills and Socio-Economic Status 52

4.1 Introduction 53

4.2 The Unit Home Equivalent Distribution Network Pricing Method 55

4.2.1 The Unit Home Equivalent 56

4.2.2 The Unit Home Equivalent in the Network 58

4.2.3 Theoretical Improvement of the UHE Pricing Method 61

4.3 The Case Study for the UHE Method 64

4.3.1 The LRIC Charging Model 67

4.3.2 The Proposed UHE Model 68

4.4 Results and Discussion of the Case Study 71

4.4.1 The Long-Term Impact on the Local Network Cost Change 71

4.4.2 The Long-Term Impact on the Local Network Consumption Change 72

4.5 Identification the Impact of UHE Energy Bill 76

4.5.1 The Network Cost Resulted by Using UHE Pricing Method 76

4.5.2 The Significant Socio-Economic Criteria for the Network Bill Variation 78

4.6 Chapter Summary 85

Chapter 5 The Impact of Socio-Economic Features on Cost-Reflective Customer Classification 87

5.1 Introduction 88

5.2 Background of the Research 89
5.3 The Cost-Reflective Customer Classification Framework 91
5.3.1 Supply Cost Quantification of Individual Customers Based on Smart Metering Data 93
5.3.2 Feature Generation from the Smart Metering and Socio-Economic Data 94
5.4 Implementation of the Proposed Framework.. 98
5.5 Results and Analysis 103
5.5.1 Comparison with Load-Profile Based Classification......................... 103
5.5.2 Results Analysis for the Selected Features 105
5.6 Chapter Summary 112
Chapter 6 The Impact of Socio-Economic Features on the Responsiveness to Different Tariff Plans 114
6.1 Introduction.. 115
6.2 Experimental Data Description.. 117
6.2.1 The TOU Tariff Types 117
6.2.2 The Definition of Customers' Responsiveness 119
6.2.3 The Load Features and Socio-economic Features 119
6.3 Proposed Responsiveness Analysis Framework....................................... 120
6.4 Results and Discussion 122
6.4.1 The Common Significant Features for Different TOU Tariff Plans.. 124
6.4.2 The Unique Significant Features for Different TOU Tariff Plans 132
6.5 Chapter Summary 138
Chapter 7 Conclusions 141
7.1 Conclusions.. 142
7.2 The Socio-Economic Criteria for the Wholesale Market Cost Variation .. 143
7.3 The Socio-Economic Criteria for the Network Cost Variation 144
7.4 The Impact of Socio-Economic Features on Cost-Reflective Customer Classification.. 145
7.5 The Impact of Socio-Economic Features on the Responsiveness to Different Tariff Plans... 146
7.6 The Limitations of the Research 148
Chapter 8 Future Works 150
8.1 Future Works 151
8.1.1 Development of Tariff Design based on Customers' Flexibility....... 151

8.1.2 Deeper Investigation between Socio-Economic Data and Customers' Usage Behaviour 151
Appendix A 153
Appendix B 180
Appendix C 186
Publications 190
References 191

Abstract

With the vigorous promotion of Advanced Metering Infrastructure and Half-Hourly Settlement (HHS) reform, the two-way communications between the residential customers and suppliers are built. The market signals are transmitted to the end-users from their accurate energy bill calculated in HHS process. The policymakers expect customers to mitigate the uncertainty in the energy market by modifying their usage behaviour following the market signals, and meanwhile reducing their energy bills. However, the policies also introduce uncertainty in customers' energy bills. Therefore, the impact of policies on customers from different socio-economic status needs to be assessed. Moreover, to timely launch appropriate interventions to assist the vulnerable customers, the socio-economic data needs to be analysed to obtain a more in-depth understanding of customers' usage behaviour. This thesis fills the research gaps by investigating the effect of socio-economic data from two aspects: 1) investigating the impact of interacted socio-economic data; 2) considering the effect of the collaboration of socio-economic data with other data sources, such as the smart metering data, the Time-Of-Use (TOU) tariff data and so on.

The investigation of the effect of interacted socio-economic data is triggered by the HHS reformation to the energy retail market. The HHS process provides more accurate energy bills to individual customers. Meanwhile, it also introduces uncertainty to customers' future energy bill. Hence, by analysing the effect of interacted socio-economic data on the variation of residential customers' energy bills, the impact of the HHS reform on customers with different socio-economic status can be assessed. A novel high-dimensional interaction-aware search method has been proposed, which is named the KLAM method. The KLAM method can detect the high-dimensional interacting significant factors, meanwhile minimising the information loss. The interacted significant socio-economic factors could describe the socio-economic characteristics of the new vulnerable customers under the HHS process. Additionally, a novel distribution network pricing method is proposed which removes the cross-

subsidies in network cost among customers. The impact of network cost variation on customers in different socio-economic status can be investigated.

Applying the socio-economic data with other data sources can explore the better performance of different demand-side appliances.

1) Socio-economic information can remedy the problem caused by the availability issues of other data sources. For example, the availability limitation of smart metering data for the new switch-in customers is a problem for the customer classification. Therefore, a cost-reflective classification framework has been proposed by collaborating socio-economic data with smart metering data. Three scenarios are established in the novel classification framework to estimate the energy cost level for the customers who 1) only have the smart metering data; 2) only have the socio-economic data;; 3) have both two datasets. The accuracy of energy cost prediction for those three scenarios is 74.88% and 53.31% and 75.00% respectively.

2) Furthermore, a responsiveness pre-evaluating framework has been proposed. This framework aims to identify the significant socio-economic criteria and load characteristics for customers' responsiveness to different TOU tariffs.

Acknowledgements

First of all, I would like to express my genuine appreciation to my supervisors, Dr Ran Li and Prof. Furong Li for their continuous support and invaluable suggestion and guidance on my research and throughout my PhD life in the University of Bath. Their guidance not only supports me to break through the challenges I met in my research but also help me to master the skills of analysis and solve problems. I am sincerely grateful for their efforts devoted to me.

I appreciate all the colleagues who have collaborated with me in researches and generously shared their priceless advice and experience with me. Dr Shuangyuan Wang, Dr Zhong Zhang, Dr Chenghong Gu, Dr Kang Ma, Dr Nathan Smith, Dr. Ignacio Hernando Gil, Dr Chen Zhao, Dr Zhipeng Zhang, Dr Zhimin Wang.

I would like to express my thanks to all my friends in the University of Bath for their selfless accompany with me, including Dr Minghao Xu, Dr Heng Shi, Dr Yuankai Bian, Dr Wei Wei, Dr Heather Wyman-Pain, Dr Hantao Wang, Miss Chi Zhang, Mr. Xiaohe Yan, Mr. Xinhe Yang, Miss Wangwei Kong, Miss Lanqing Shan, Mr. Jiahang Li, Miss Nini Yuan, Miss Yunting Liu. I treasure our friendship and every moment with you.

Last but not least, I would like to take this opportunity to express my gratitude to my family, Miss Ling Fang and Mr. Fuxiang Ma. I appreciate for everything my parents give to me.

List of Figures

Figure 1- 1: The electricity bill breakdown. 7

Figure 2- 1: The Acorn group names and categories 17

Figure 3- 1: The structure of the KLAM beam search algorithm 29
Figure 3- 2: The beam search structure in the KLAM beam search algorithm 30
Figure 3- 3: The example of information loss caused by the non-normal of treatments' distributions..32
Figure 3- 4: The example of information loss caused by the overlapping of treatments' distributions..34
Figure 3- 5:The flowchart of the KLAM beam search algorithm...............................41
Figure 3- 6: Interaction effect of one pair of factors found in Stage II.......................45

Figure 4- 1: Daily load profile for a PV net zero customer 54
Figure 4- 2: Examples for substation and home load profiles 57
Figure 4- 3: Investment projection: number of additional base homes can be connected to the network without triggering reinforcement 58
Figure 4- 4: The network structure of the test system 66
Figure 4- 5: Load profiles of seven substations 67
Figure 4- 6: The LRIC cost for all of new customers under every substation............68
Figure 4- 7: The UHE price for 10 customers under every substation 69
Figure 4- 8: The normalized UHE price of customers connected in Sub-V................70
Figure 4- 9: Home load profiles for the different normalized UHE price values........70
Figure 4- 10: Substations' load profiles for 10th year in UHE Scenario 1...................73
Figure 4- 11: Substations' load profiles for 10th year in UHE Scenario 2...................73
Figure 4- 12: Substations' load profiles for 10th year in LRIC model.........................74
Figure 4- 13: The original load profiles for Sub-III and Sub-VI 75
Figure 4- 14: The substation load profile for Sub-VI 76
Figure 4- 15: The typical load profile of ten classes..78

Figure 5- 1: The structure of the proposed customer classification framework 91
Figure 5- 2: The breakdown for gas and electricity bills..93
Figure 5- 3: Two-class classification problems example for SVM 102
Figure 5- 4: The impact of top 7 smart metering features on cost groups................107
Figure 5- 5: The impact of four smart metering features on cost groups 108
Figure 5- 6: The distributions of three cost groups for 10 socio-economic features .111

Figure 6- 1: The four types of weekday TOU Tariffs...118
Figure 6- 2: The four types of weekends TOU Tariffs 118
Figure 6- 3: The divided time period for consumption related load feature.............120
Figure 6- 4: The flowchart of the propose framework..122
Figure 6- 5: Population breakdown for responsiveness groups................................124

List of Tables

Table 3- 1: The XOR digital logic gate example for interaction effect 27
Table 3- 2:Notations of the algorithm...35
Table 3- 3: Original factors and their P-values detected in Stage I 43
Table 3- 4: Original factors and their P-values detected in Stage II by KL-divergence 44
Table 3- 5: Socio-economic factors show different impact on the cross-subsides......48

Table 4- 1: Parameters of the Test Network and DC-PTDF..65
Table 4- 2: The network investment cost in 10 years 71
Table 4- 3: The number of customers injected into each substation in the 10-years period 71
Table 4- 4: The difference between peak load and valley load 75
Table 4- 5:The UHE price for different load pattern customers..................................77
Table 4- 6:The P-value of the original significant socio-economic factors.................79
Table 4- 7: The P-value of the socio-economic factors detected in *Stage II* 79
Table 4- 8: The significant socio-economic factor combinations and their impact on cross-subsidies 81

Table 5- 1: The explanation of 56 load features 96
Table 5- 2: Comparison results between ReliefF-SVM and other methods 99
Table 5- 3: Classification accuracy results based on hierarchical clustering algorithm 105
Table 5- 4: The smart meter features selected by ReliefF algorithm.........................106
Table 5- 5: The socio-economic features selected by ReliefF algorithm 109

Table 6- 1: The Number of Participants for Different TOU Tariffs..........................123
Table 6- 2: The Common Significant Features for Different TOU Tariffs 125
Table 6- 3: The significant interacting-feature combinations for common significant features...127

Table 6- 4: The unique significant features for Tariff A 133
Table 6- 5: The unique significant features for Tariff B...134
Table 6- 6: The unique significant features for Tariff C...135
Table 6- 7: The unique significant features for Tariff D 136

List of Abbreviations

the Office of Gas and Electricity Markets	Ofgem
Carbon Dioxide	CO_2
Department of Energy & Climate Change	DECC
Low Carbon Technologies	LCTs
Electric Vehicles	EVs
Demand-side Response	DR
Typical Load Profiles	TLPs
Energy Management System	EMS
Half- Hourly Settlement	HHS
Analysis of Variance	ANOVA
Kullback-Leibler divergence	KL-divergence
Gaussian Mixture Model	GMM
Unit Home Equivalent	UHE
Chief Income Earner	CIE
Grid Supply Point	GSP
Question	Qu
Locational Marginal Pricing	LMP
Long-run Incremental Cost Pricing	LRIC
Typical Headroom Profile	THP
Power Transfer Distribution Factor	PTDF
Substation	Sub
United Kingdom	UK
United States of America	USA

Chapter 1

Introduction

THIS chapter introduces the background, motivations, challenges and contributions of this work. The structure of this thesis also be given in this chapter.

1.1 Background and Motivation

1.1.1 The New Context: Climate Change and Growing Low Carbon Techniques

Since the industrial revolution, the amount of Carbon Dioxide (CO_2) emission increased by nearly 40% [1, 2]. To retard the rapid growth of greenhouse gas emissions, the UK has signed the Kyoto Protocol promising to cut off 34% of the greenhouse gas emission by 2020 with respect to the emission level in 1900 [3, 4, 5]. Since the power sector is the biggest source of the CO_2 emission in the UK, it brings a huge opportunity and motivation for achieving the reduction goal [6]. To step forward to a greener future and fulfil the environmental commitment, the Low Carbon Technologies (LCTs) have been significantly developed in the recent decade. Generally, the LCTs are deployed to reduce the carbon emission from two perspectives:

- ***Energy generation aspect:***

  Traditional fossil energy fuel is the primary source of the carbon emission which threatens the environment. Therefore, the clean and sustainable renewable generations have been promoted vigorously. The Ofgem introduced a portfolio of projects and funds to encourage providing cleaning energy in the UK [7-10]. Consequently, in 2018, the electricity generation from renewable sources, such as the wind, solar and bioenergy, was up to 110 TWh, which was 33% of the total generation. This has grown from 2000 when the renewable generation only occupied 2.6% [8].

- ***Customer consumption aspect:***

  The growing peak demand also consequents on the tremendous need for energy use. Thereupon, the innovations which can improve the system utilisation by providing the flexibility to the demand-side users are also exploited. The flexible demand, such as the electric vehicles and battery storages, could support the end-users to modify their usage behaviour. Through the appropriate incentive (the Demand-side Response) and optimal planning (the Energy Management System), the flexibility at demand-side could bring great advantage to the system. For instance, it can reduce the system peak demand, deferring network reinforcement and increasing the system utilisation.

1.1.2 Motivations: Creating a Dynamic Energy Market

To achieve a more environmentally friendly future, a dynamic energy market needs to be created. The two changes in the new context of the energy market introduce the two motivations of socio-economic data analytics.

- The first motivation is the requirement of a more in-depth understanding of customers' usage behaviour.

In the new context of the power system, the penetration of renewable energy is high. Due to the inherently volatile and intermittent of renewable generation output, the suppliers face the rapidly increasing uncertainty in the demand prediction, which is caused by the additional flexible load at the demand-side, such as the Electric Vehicles (EVs) and storages. Under this circumstance, an in-depth understanding of customers' usage behaviour is required in the new dynamic energy market.

Traditionally, the demand of customers is highly accumulated, especially for domestic customers. The suppliers estimated the usage for each domestic customer based on the Typical Load Profiles (TLPs) [11]. However, there are only two TLPs to represent the 53 million volatile residential customers in the UK. The inaccurate estimation impedes the suppliers to understand the demand of residential customers.

When the suppliers cannot access to the accurate customers' usage data, the effect of socio-economic data on customers' usage behaviour is worth to be investigated. Moreover, collaborating the socio-economic data with other data, such as the Time Of Use (TOU) tariff data, could help the suppliers to obtain a more in-depth understanding of customers' usage behaviour.

- The second motivation is the unblocking of the communication between the end-users and suppliers.

Traditionally, the suppliers had been blocked from the residential customers' accurate half-hourly data due to the absence of smart meters. The residential customers cannot receive the energy market signals because they were charged with a fixed tariff.

However, due to the high uncertainty in the new dynamic smart market, the end-users are expected to be participated more actively to reduce the uncertainty bare by the suppliers. Therefore, two critical strategic policies had been launched by the Department of Energy & Climate Change (DECC) and the Ofgem to unblock the two-way communication between the end-users and the suppliers, which are:

i) Rolling out the smart meters to small and residential customers

In 2013, the DECC arranged to roll out 53 million smart meters to all residential and small business customers in the UK by the end of 2020 [12]. The smart meters realize the real-time remotely communication between the customers and suppliers.

ii) Electricity settlement reform- Half-Hourly Settlement (HHS)

The electricity settlement is a top-down allocation process which aims to apportion the regional consumption to each supplier. The original electricity settlement for domestic customers estimated their demand based on TLPs [11]. The inaccurate TLP estimation causes large cross-subsidies and unfair issues among residential customers. Therefore, in April of 2014, the Ofgem decided to lead an electricity settlement reform project to realise the opportunity brought by the widely deploying of smart meters [13]. By utilising the accurate half-hourly usage data in the settlement process, the cost-reflective energy bill can transmit the energy market signals to the end-users.

Those strategic policies aim to send market signals to guide the customers changing their usage behaviour. The policymakers expect customers can mitigate the uncertainty in the dynamic smart market, and meanwhile, reducing their energy bill. However, it also introduces uncertainty in customers' energy bill. Additionally, if the customers fail to respond to the market signals, the timely and appropriate interventions need to be launched by suppliers to support the end-users.

Socio-economic information can describe the social status of customers. By analysing the effect of interacted socio-economic factors on customers’ bill variation, the impact of the policy on customers from different social status could be assessed. Moreover, with the socio-economic characteristics for the vulnerable customer groups under the new dynamic smart market, more tailored interventions (such as DSR) could be designed by the suppliers.

1.2 Research Questions: Analysis of the Socio-Economic Data

With the smart meter rolling-out and the HHS reformation, the energy market becomes more dynamic, transparent, and fairer for customers. Meanwhile, the new environment of the energy market also presents a huge challenge to suppliers. In the dynamic smarter energy market, a vast number of the DR schemes or Energy Management System (EMS) are developed to better facilitate the LCTs. It requires the suppliers to have a better understanding of the energy behaviours of their consumers. The smart metering data for residential customers has been widely investigated in many literatures to obtain deeper insights into residential customers behaviour [14, 15], since it intuitively and accurately demonstrates the customers' usage pattern. However, the effect of socio-economic data is rarely investigated.

Apart from the smart metering data, the socio-economic data provides necessary information related to customers' usage behaviour from a different perspective. The social-economic data can provide a variety of information, such as the social class, income level and the appliances owned by the customer and so on, about the individual customer through the digitization devices. Therefore, the socio-economic data is equally important as the smart metering data, especially for the customer analysis in the DSR and EMS programmes.

Nowadays, there are two unsolved research gaps of the new dynamic energy market, which calls for the support from socio-economic data analytic.

i) Impact assessment of HHS on customers with different socio-economic status is necessary for the policymaker and suppliers to ensure the electricity supply for customers is affordable, secure and low-carbon [16].

ii) Lack of research of interconnecting socio-economic data with other dataset domains to improve the demand-side applications' performance.

1.2.1 The Research Questions in Technical Field

To handle the research gaps by utilizing the socio-economic data, there are two technical problems:

- **The interaction effect among the socio-economic data**

In order to describe the socio-economic status of customers' life comprehensively, massive information needs to be digitalized into plenty of socio-economic factors. The effect of interaction among factors could be significant on customers' usage behaviour, consumption and bill change caused by HHS. However, there is no efficient methodology which can search the interaction among factors, meanwhile, avoid the curse of dimensionality and minimize the information loss.

- **The energy cost change for individual customer**

Figure 1-1 demonstrates the electricity bill breakdown, which shows that the wholesale costs and the network costs are the two largest parts for the total bill. Therefore, the bill change caused by transforming to the HHS process is constituted by the wholesale and network cost changes.

With the accurate smart metering usage data, the costs changes in the energy wholesale market can be easily obtained. However, the network costs variation is difficult to be ascertained for the individual customer. The widely utilised Distribution Use of System (DUoS) charging methodology, which depends on the annual peak value, is designed for the large customers and the retailers. For the small individual customers who have volatile load patterns, the existing DUoS charging method is not cost-reflective. Therefore, before assessing the impact of HHS on customers' bills, a cost-reflective distribution network pricing needs to be proposed for the individual residential customer.

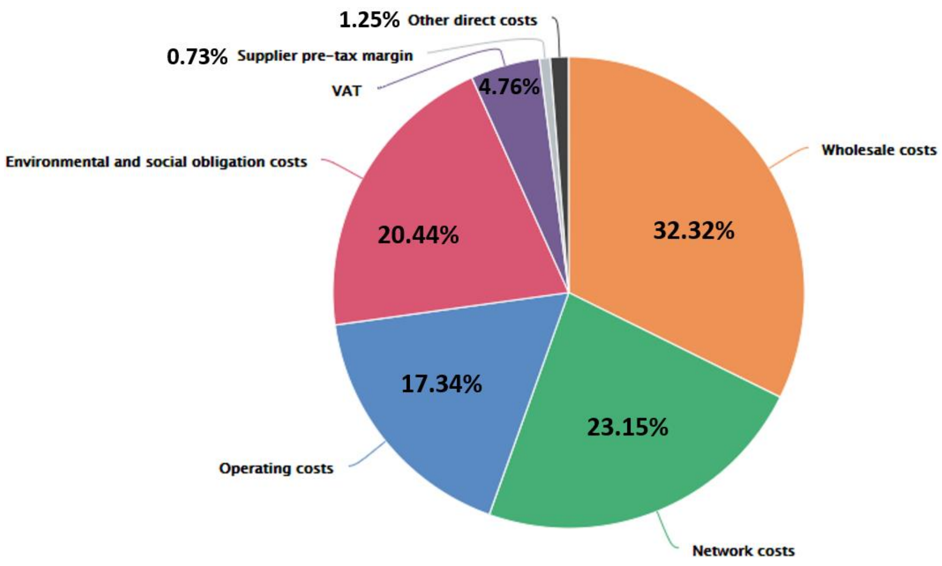


Figure 1- 1: The electricity bill breakdown.

1.2.2 The Research Questions in Application Field

The main advantage of adding socio-economic data into the analysis:

- **Breakthrough the limitation caused by the availability issues of smart metering data**

In the new dynamic smarter energy market, the customers' accurate smart metering data is crucial for designing and customer targeting of the demand-side application like DSR schemes or EMS. However, due to the full competition of the UK's energy retail market, the customers' historical smart metering data is not open to customers' new suppliers. , although plenty of existing research analysis significant load characteristics for customer targeting or segmentation, the different degree of data-availability for different customers would be the main limitation in practical demand-side applications.

The socio-economic data could remedy the missing of usage data. In this way, accurate customer targeting, or segmentation can be achieved by analysing socio-economic data combined with partial, even no smart metering data. However, there are few studies analysis the interconnecting effect of socio-economic data with other data sources, such as the TOU tariff data, usage data.

1.3 Research Contributions

This thesis aims to utilise the socio-economic data to support customers and the network participants better fitted in the new dynamic smarter power market. There are four original contributions for this research.

1) Two contributions are to the techniques field

   Two novel algorithms proposed to explore the effect of socio-economic data on customers' energy bill variation. A high-dimensional interaction-aware beam search algorithm, KLAM algorithm, is proposed to assess the impact of wholesale market cost variation (caused by HHS reform) on customers from different socio-economic status. Furthermore, a novel distribution UHE network pricing method has been proposed. The UHE network pricing removes the cross-subsidies among network cost by allocating the network cost to individual customers based on their usage behaviour. Therefore, the significant socio-economic criteria for the network cost level for different residential customers can be figured out.

2) Two contributions are to the facilitating of demand-side applications

   The socio-economic data have facilitated two demand-side applications. The first application is building a cost-reflective customer classification framework. By collaborating socio-economic data with load features, the accuracy of the classification has been improved. Additionally, even when the load data is inaccessible, the cost level of customers can still be estimated based on socio-economic data in the proposed framework. The second application is building a customers' responsiveness pre-evaluation framework to figure out the socio-economic criteria of the benefited customers under different TOU tariffs.

1.3.1 Contributions to the Techniques Field

- This thesis proposed a high-dimensional interaction-aware beam searching method, which is called the KLAM algorithm, in the sensitivity analysis between the socio-economic makeups and the amount of energy bill changing in the HHS process. The proposed KLAM algorithm can take the interaction effect among socio-economic factors into consideration with the minimized information loss in dimensionality reduction. The proposed KLAM algorithm has figured out several significant socio-economic characteristics of households who would experience an energy bill increasing under the HHS process.

- For the purpose of obtaining the value of network cost cross-subsidies for each residential customer, a novel Unit Home Equivalent (UHE) distribution network pricing is proposed in this thesis to calculate individuals' network cost. The UHE pricing method considers the likelihood of future peaks created by the load profiles for customers at different time point, instead of only calculate the network cost depends on the contribution at the historical annual peak point.

1.3.2 Contributions to the Facilitating of Applications

By collaborating the socio-economic data with load features and TOU tariff data, two applications can be facilitated.

- A cost-reflective customer classification framework has been proposed in this research. The proposed framework aims to rapidly classify customers based on their energy cost from the perspective of suppliers. By utilizing the socio-economic data and/or the smart metering data, three scenarios are built based on the available data for customers. Through the proposed framework, even the smart metering data is inaccessible, the energy-cost level for a new customer still could be identified by only using the influential socio-economic features.

- A responsiveness pre-evaluating framework has been proposed in this thesis. The main objective for the proposed framework is to identify the significant criteria for customers’ responsiveness to different TOU tariffs. The framework analyses the interaction effect among households’ socio-economic features and the load characteristics to achieve the appropriate customer-targeting for different TOU tariff plans.

1.4 The Layout of Thesis

The rest of the thesis is organised as follows:

Chapter two presents a comprehensive literature review of the research of socio-economic data of residential customers. The literature review introduces the widely investigated socio-economic factors and their effect in different applications. The difference in research methods and their limitations are also demonstrated in this chapter.

Chapter three proposes a novel high-dimensional interaction-aware KLAM beam searching method. It adopts the Analysis Of Variance (ANOVA) to searching the significant socio-economic factors for customers' bill change after removing the wholesale market cost cross-subsidies. Meanwhile, to minimize the information loss, the Kullback-Leibler divergence (KL-divergence) and Gaussian Mixture Model (GMM) are adopted in the novel search method to recycle the significant information hidden in the factors, which are abandoned to reduce the computation burden in ANOVA.

Chapter four follows the previous chapter and adopts the novel KLAM beam search method to detect the impact of removing the cross-subsidies in the energy network cost on customers in different socio-economic status. This chapter proposes a new distribution network pricing for electricity retail market, called the Unit Home Equivalent (UHE) pricing method, to remove the cross-subsidies in network cost among customers. The UHE value for individual customer measures the additional number of the same customer can be connected to the network without triggering the reinforcement relative to a unit constant base home. Then, the network investment cost is allocated based on the UHE value to the individual household. Finally, the significant interacted socio-economic factors which positively impacts on customers' accurate network cost depict the socio-economic status of adversely affected customers.

Chapter five focus on the effect of socio-economic data collaborating with other data sources. The first application is the customers' energy cost classification for customers with different availability of input data. In this chapter, a cost-reflective customer classification framework has been proposed by collaborating customers' smart metering data with socio-economic factors. Three scenarios are built based on the available data for customers to estimate the energy cost level for new switched-in customers with partial or even without smart metering data.

Chapter six focus on the effect of socio-economic factors on customers' responsiveness to different DR programs. A framework is established to pre-evaluate customers' responsiveness for different tariff plans by taking advantages from the interaction effect among customers' socio-economic and load factors.

Chapter seven demonstrates the conclusions of the research and the key contributions of the work.

Chapter eight outlines the potential research topics in future work.

Chapter 2

Review of Socio-Economic Data

THIS chapter summarises the socio-economic data which has been studied in previous works for different purposes. It also introduces the popular methods adopted in literature and discusses the limitations of existing research in socio-economic data analytics.

2.1 Introduction

Following the transformation to the smarter power grid, many researchers already found that they need to obtain more in-depth insights into residential customers' consumption and usage behaviour from the household characteristics information due to the inherently volatile usage pattern for the domestic customers.

Therefore, in this chapter, the literature review from three aspects to introduce the existing research related to socio-economic information of residential customers.

1) The effects of the widely investigated socio-economic data

   In this sub-section, the conclusions of the widely investigated socio-economic factors are presents. The conclusions including their effects on consumption, energy-saving, and customers usage behaviour fields.

2) The applications and methodologies in socio-economic data studies

   In this sub-section, the applications and the methodologies which were used to investigate the socio-economic data in the previous studies are introduced.

3) The limitations of the existing investigation of socio-economic data

   Finally, the limitations of methodologies and datasets of the literature review related to socio-economic data analytics are presented.

2.2 Effects of Socio-Economic Data

In recent years, a broad range of data analytics studies focuses on the influence of the data which depicts customers' household characteristics. In this thesis, the socio-economic data is defined as the data which is digitalised from customers' society's characteristics except for the smart metering data.

The socio-economic data investigated in literature can be separated into two groups:

1) The digitised data collected from a survey of questionnaire

2) The geodemographic code which is labelled by government bodies

2.2.1 The Digitised Survey Socio-Economic Data

This kind of research always utilises plenty of socio-economic questions to describe the socio-economic status for one customer. Generally, the comprehensive state can be depicted by four categories of factors: 1) The dwelling factors; 2) The electric appliances factors; 3) The personal information factors, and 4) The psychological factors.

- **Dwelling factors**

The socio-economic data which describe the dwelling of households include: 1) dwelling type; 2) age of the house; 3) the floor area of the house (i.e. Number of bedrooms, Number of floor and so on). Some factors about the cooling and heating system and the energy-saving devices, such as the energy-saving light bulbs, are also included.

Generally, the dwelling type of customers have a positive effect on customers' consumption level [17-22]. The customers who live in detached houses have significantly higher energy consumption than other dwelling types, such as the semi-detached, bungalows, and end-of-terrace houses.

However, the conclusions on the influence of floor area on the consumption level are not as consistence as the dwelling type. Some research reported that the floor area value has a significant impact on customers' usage [20, 23, 24]. The main reason is due to the larger dwelling size often related to the more considerable need for electrical heating and cooling [25, 26]. In contrary, in [17, 27], the authors found the difference in floor area did not influence the average consumption. The author in [17] believes the insignificant effect of the floor area is related to the similarity architecture of the dwelling samples.

The age of the house also achieved different conclusions for the impact on the consumption level. The positive effect summarized in [21, 22, 28] and the contrary conclusion obtained in [25, 29-31]. Additionally, the house-age also has a positive impact on the energy-saving features, as found in [32]. The contrary conclusion in [33] has been found that the age of the house does not have a significant effect on energy savings.

- **Appliance Factors**

The electric appliances factors attract the most attention of the research related to customers' consumption investigation. The appliance related factors include the questions of 1) the ownership of appliances; 2) the frequency of using appliances; 3) the consumption of appliances.

The conclusions of the existing research on the influence of the appliance-related factors show that the number of the owned electric appliance has a significant positive effect on consumption level [17, 24, 34, 35]. The frequency of use of washing machine and tumble dryers also been reported that have a positive effect on consumption level in [17, 20].

Moreover, appliance factors have been investigated in the analysis of customers' responsiveness in DR programmes. This kind of research mainly focuses on controllable (flexible) appliances. For example, in [36], the ownership of the wet appliances (i.e. dishwashers, washing machines and tumble dryers) has no significant effect associated with the amount of load shifting in DR programs. However, the opposite results had been found in [37-39], which demonstrate the ownership of the wet appliances makes a higher willingness to accept the DR compared to other appliance (i.e. the cooking appliance).

However, based on the review in [40], there are several appliances had been studied infrequently, such as the laptop computer, electric heater, and water-pump and so on. Therefore, the effect of those appliances on customers' consumption is inconclusive.

- **Personal Information**

The widely used socio-economic factors which describe the personal information of customers can be categorized into: 1) family composition, including the number of occupants, presence of children, teenager, adults and elder people; 2) the income level of the household; 3) employment status of the chief Income Earner (CIE); 4) the education level of the CIE; 5) the socio-economic class for the household; 6) the age of the occupants and 7) the tenure type.

From [27, 35], the employment status of CIE had been found consistently that it is insignificant on household consumption. The income of CIE demonstrates its positive effect on customer consumption level in [19, 41].

Other personal information related factors, such as the presence of children in a household, socio-economic status class, have either a mix of effect which is difficult to be summarized, or been infrequently investigated [40]. The factors like education level or tenure type have nearly no effect on the household consumption level [36].

For example, the positive effect of family composition on residential consumption level has been proofed in many literature[18, 20, 21, 24, 42, 43]. McLoughlin et al. in [18] found that the consumption of household living with children is significantly higher than the building where adults are living together. However, the research in [35] revealed that the number of children under three would have a significant impact on usage level contrary to previous studies. Therefore, the conclusion of the effect of the family composition is difficult to give.

- **Psychological Factors**

In [44, 45], several Norm Activation Model (NAM) variables have been summarized to analysis the pro-environmental behaviour (which is formed as an altruistic behaviour in the research) of people. It found that people who have the awareness of consequences of their behaviour for the environment (i.e. people who believe the energy usage will damage the environment) and feel responsible for the consequence will achieve larger energy usage reduction. In [46, 47], the authors reveal that the psychological variables, such as the attitudes to energy saving, have significant effect on energy conservation.

However, the attitudes on energy conservation, climate change and renewable resources investigated in [33] demonstrate that there is no effect for the environmental attitude on the customers' demand shifting. The similar results also have been discovered in [48, 49] that the attitudes of customers to climate change have small influence on electricity demand.

2.2.2 The Geo-Demographic Code

Some research adopts a geo-demographic code to represent the customers' socio-economic status. This geo-demographic code is assigned by government bodies after classification people based on comprehensive information. For example, the Acorn Group Label has been applied to describe the demographic composition for customers in [50] to support the investigation of residential consumer responsiveness to time-varying pricing. The Acorn Label, which is displayed in Figure 2-1 [50], is a geo-demographic classification of the UK's households, which is licensed by the Consolidated Analysis Centres Incorporation (CACI) [51].

Label	Acorn group	Acorn category
A	Wealthy executives	Wealthy achievers
B	Affluent greys	Wealthy achievers
C	Flourishing families	Wealthy achievers
D	Prosperous professionals	Urban prosperity
E	Educated urbanites	Urban prosperity
F	Aspiring singles	Urban prosperity
G	Starting out	Comfortably off
H	Secure families	Comfortably off
I	Settled suburbia	Comfortably off
J	Prudent pensioners	Comfortably off
K	Asian communities	Moderate means
L	Post industrial families	Moderate means
M	Blue collar roots	Moderate means
N	Struggling families	Hard pressed
O	Burdened singles	Hard pressed
P	High rise hardship	Hard pressed
Q	Inner city adversity	Hard pressed

Figure 2- 1: The Acorn group names and categories

2.3 Applications and Methodologies in Socio-Economic Data Analytics

2.3.1 The Applications for the Socio-Economic Data Analysis

The socio-economic data can give assistance in many kinds of researches. The main two applications for the existing socio-economic data analysis are:

- **Electricity Consumption**

  The most studied application is the socio-economic data impact on domestic electricity consumption [18, 19, 22, 52]. All the dwelling, personal information, appliances and psychologic factors have been investigated to find the determinants of residential electricity consumption [53, 54]. In [25, 41], the socio-economic data are utilised to improve the prediction of domestic energy demand. On the contrary, there are some studies [55-57] utilizing the smart metering usage data to identify the household socio-economic characteristics.

- **Responsiveness of DR and Energy Conservation**

  The appliance stock information is important in the studies of customers' responsiveness analysis due to some electrical appliances provide flexibility to the customers participating in the DSR programmes [58, 59]. Moreover, personal information such as how many people would stay at home during the daytime and environment attitudes also have a significant influence on the amount of demand shifting [36]. Therefore, the analysis of the responsiveness of residential customers in DSR is a primary application for socio-economic data.

  Furthermore, the relationship between socio-economic data and energy conservation also has been investigated in [21, 32, 60, 61].

2.3.2 The Methodologies for the Socio-Economic Data Analysis

There are various methods used in previous socio-economic data analysis. The literature has been categorised into two groups:

- **Modelling Method**

In the modelling method group, the series of regression models are widely adopted. Especially for the application of finding the relationship between the socio-economic factors and the residential consumption level [17, 18, 22, 25, 33, 53, 62].

For example, simple and multiple regression were used in [25] to determine the strength of the effect of the dwelling types and appliance factors on the difference in gas and electricity consumption. In [33] the effect of five psychological factors related to the attitudes of energy-saving, the house size and customer age factors on households' energy saving has been assessed based on multiple regression. The author in [18], examined the influence of dwelling, customer personal information factors and cooking type factors on total electricity consumption, maximum demand, load factor and time of use of maximum demand through a multiple linear regression model.

The ordinary linear least squares regression is adopted in [22], which is used to detect the relationship between the energy use and the ownership of domestic appliances.

Furthermore, in [63], there are three modelling techniques for the energy consumption prediction based on appliance ownership and power rating data and the dwelling related data, which are regression model, decision tree and neural network.

- **Statistical Analysis**

The statistical analysis also widely utilised to assess the effect of socio-economic data on different appliances [28, 29, 42].

The correlation value is an important statistical value to test the relationship between the energy consumption and the dwelling age in [28, 41, 42]. The effect of floor area and average annual electricity consumption was found in [27] by a clear positive correlation value.

Additionally, the mean value and variance value also could be used to analyse the socio-economic characteristics for different customer groups [28, 60, 64, 65].

For example, in [28], the customers were firstly segmented based on their socio-economic characteristics (i.e. the type of house, the age of house), then, the mean and standard deviation of the energy consumption for each customer groups are calculated to demonstrate the relationship between the socio-economic data and usage level.

In [64], the usage patterns for customers are firstly classified, then the percentage values of the ownership of appliances have been summarized to obtain the socio-economic characteristics for customers in different load profile classes.

2.4 The Limitations of the Existing Research

By review the previous research of the effect of socio-economic data, there are mainly two limitations for applying the widely used methodology to our research:

- **The Interaction Effect**

The interaction effect among socio-economic factors does not be taken into consideration in most of the literature papers.

Although the modelling method, like the regression methods, contains the effect among several socio-economic factors, however, the number factor is very limited for the modelling techniques. To comprehensively describe the socio-economic status for different households, the dimensionality of input data would be too high to be handled by the modelling methods used in previous literature.

The mixed inconclusive effect of many socio-economic factors resulted by different studies may also due to the lack of the context of customers' socio-economic status.

- **The Requirement of the Interpretability**

In the literature review, the neural network techniques are utilised to detect the relationship between socio-economic status and consumption. Although the neural network can take the interaction effect among socio-economic factors into consideration, the interpretability of the important socio-economic characteristics also is required for the research in this thesis.

Chapter 3

The Socio-Economic Criteria for the Impact on Wholesale Market Cost Variation

T**HIS** chapter conducts assessments of the impact of HHS. A high-dimensional interaction-aware searching method has been proposed to investigate the socio-economic status for the customers be adversely impacted.

3.1 Introduction

As mentioned above, the cross-subsidies of the wholesale market cost among customers have been removed by introducing in the precise smart metering data in HHS. The original electricity settlement process allocates customers' accumulative consumption into each half-hour period based on the TLPs due to the absence of smart metering data. With the rolling out of smart meters, the HHS reform had been first set in 2013, which aims to use the correct half-hourly smart metering data to help create retail energy markets that work better for consumers.

However, the impact of HHS could adverse for specific customers. For example, some vulnerable customers who had been subsidised in traditional settlement process may experience a considerable bill increase after transforming to the HHS. Therefore, to ensure the energy is affordable for end-users, the power system regulator and suppliers must assess the impact of HHS reform on customers in different socio-economic status.

However, the socio-economic data for customers is relatively unheeded compared to usage data in the earlier studies which are introduced in Chapter 2. The existing literatures universally investigate the effect of single socio-economic factor, which is insufficient to describe the state of customers' life. But the interaction effect among factors could influence customers' usage significantly, which should be investigated. In this chapter, we will propose a novel algorithm, the KLAM beam search, which takes the high-dimensional interaction effect of socio-economic factors into consideration meanwhile minimized the information loss. By using the proposed KLAM searching algorithm, the socio-economic factor combinations which have a significant effect on customers' bill change in HHS process can be found.

Therefore, the contribution for this chapter are: 1) This research, for the first time, assesses the impact of HHS on residential customers. The results provide a more in-depth insight into the new vulnerable customers under the HHS, and meanwhile, could support the policymakers and suppliers to adopt measures timely; 2) A novel searching algorithm has been proposed, which can consider the high-dimensional interaction effect among factors meanwhile minimising the latent information loss caused during the dimensionality reduction.

The content of this chapter is under review by the IEEE Transactions of Power System, which titled as " Impact Assessment of Smart Meters on Electricity Cross-subsidies: A High-dimensional Interaction- Aware KLAM Algorithm". This chapter is structured in an alternative-based format. The rest of this chapter is constructed as follows: Section 3.2 demonstrates the submitted paper, which including the introduction of the sensitivity analysis between bill changing and the socio-economic status of customers, the details of the proposed KLAM methodology, the case study and the results analysis. The conclusions are presented in Section 3.3.

3.2 Impact Assessment of HHS on Customer with Different Socio-Economic Status

3.2.1 Background of the HHS

Since the deregulation of electricity retail market in the UK in the early 1900s, energy suppliers have arisen who purchase electricity based on the estimation of their customers' consumption in each half-hour period. Therefore, the settlement is an essential process to allocate regional consumption to each supplier in a top-down fashion. However, the UK's wholesale electricity market prices change every half hourly while customers (end-users) are usually recorded biannually due to the absence of smart meters. This means a large mismatch between the top-down wholesale cost and bottom-up customer bills. In the UK, this problem has traditionally been addressed by typical load profiles (TLPs) [11]. 27 million residential customers are grouped into two classes that differentiate customers with and without economic electric heaters, each represented by a TLP to allocate customers' accumulated consumption into each half-hourly settlement period. Due to the inevitable errors in the estimated TLPs, there clearly exists a cross-subsidy in the market: some customers are overcharged while others are undercharged.

To remove the unfair cross-subsidy and hence increase the market liquidity, the UK government arranges to roll out 53 million smart meters to all households and small business by the end of 2020 [66]. With the accurate and timely smart metering data, the UK's electricity retail market is expected to increase transparency efficiency and substantially reduce cross-subsidies [67]. Although the effect of reducing overall cross-subsidy is visible, a significant uncertainty of individual customers' future electricity bills has been brought by introducing smart meter into the settlement. Therefore, it is crucial to identify customer groups that might be adversely impacted by the smart metering based half-hourly settlement, such that early intervention could be introduced to support the new vulnerable customers by policymakers. This paper fills this gap by identifying those socio-economic factors which are significant in determining the variation in the electricity bills. It is essentially a sensitivity analysis, which can evaluate the contribution of each socio-economic factor of households on the energy bill uncertainty caused by the degree of cross-subsidies.

In the existing literature, the impact of different socio-economic status on load profile groups has been investigated in many fields. The pioneering research [17, 18, 40] analyse the correlation between socio-economic factors and energy consumption. Later, the mapping between usage and socio-economic data also has been conducted for the customer behaviour segmentation [68], energy conservation [21, 32] and load profile group assignment for new customers without smart meter installations [69]. In previous works, the model-based methods are widely employed to assume the functional form between energy consumption and several socio-economic factors, for example, the regression method [21, 32]. When more socio-economic factors are taken into account, to avoid the curse of dimensionality, dimension reduction methods, such as feature selection [68] or feature extraction [53, 70], are always adopted with the model-based algorithms to consider only a partial set of socio-economic factors.

However, there are two challenges to directly adopt the previous algorithms in this paper's analysis: 1) the literature works only focus on the energy consumption rather than the degree of cross-subsidies which lead to a future bill changing. The degree of the cross-subsidies more related to the difference between the real load pattern and the estimated one which is more complicated than the amount of energy consumption; 2) the interaction effect among different factors would be ignored during the rough discard of factors which finally lead to the latent information loss. Those "abandoned" factors, which are not salient on its own, may carry significant effects when they combined with other factors.

To overcome the hurdles, this paper proposes a novel KLAM beam search algorithm to consider the interaction effect between socio-economic factors meanwhile minimising the latent information loss. The frame of the algorithm is beam search, which aims to avoid the curse of dimensionality by leaving the factor which satisfies the pruning rule. This paper adopts the results of the Analysis of Variance (ANOVA) as the pruning rule.

The Kullback-Leibler (KL) divergence and Gaussian Mixture Model (GMM) are implemented to recycle the pruned socio-economic factors to minimize the potential loss. The KLAM algorithm will be conducted on the dataset collected from the Smart Metering Electricity Customer Behaviour Trails (CBTs). The socio-economic status of the new vulnerable customer groups under the half-hourly settlement will be first revealed.

The contributions of this research are from two aspects.

- For the policy impact assessment, it is the first time to investigate the impact of transforming to the half-hourly settlement on the bill change of customers in different social classes. The socio-economic status of the new vulnerable customer groups has been comprehensively depicted. This could guide the policy maker and suppliers to deploy opportune interventions.
- For the technique aspect, this paper proposed a novel beam search algorithm, the KLAM, which can consider the high-dimensional interaction effect between factors meanwhile minimizing the information loss. It can be applied to solve the problem when the high-dimensional interaction among factors need to be considered.

The rest of the paper is constructed as follow. Section II states the problems to address this sensitivity analysis between the socio-economic status and the degree of cross-subsidies. The proposed KLAM beam search algorithm will be described in Section III. Section IV demonstrates the case study on CBTs dataset and the analysis and discussion will be displayed in Section V. The conclusions are finally drawn in Section VI.

3.2.2 Problem Statement

The object of this research is to assess the impact of transferring to the half-hourly settlement on customers with different socio-economic status. The sensitivity analysis can be applied to evaluate the contribution of each socio-economic factor on the future bill uncertainty caused by the elimination of cross-subsidies. However, as the cross-subsidies are indirectly influenced by a combination of several socio-economic factors, the interaction effect among the factors could be significant and should be taken into consideration. Therefore, two challenges have emerged and need to be tackled:

1) Considering the high-dimensional interaction effects;
2) The trade-off between the computational expense and latent information loss.

A. The high-dimensional interaction-aware sensitivity analysis

The interaction among factors indicates a phenomenon that some factors influence the output by grouping rather than individual. A typical illustration of the interaction effect is the XOR digital logic gate problem, which is demonstrated in Table 3-1. Neither FACTOR A nor FACTOR B has a significant effect on the final output. However, once only we analysed the two factors together, the variation of the output could be explained perfectly. This example illustrates that the significant effect on the final output is from the interaction between FACTOR A and FACTOR B. The XOR digital logic gate is a two-dimensional interaction example. For this paper, the higher-dimensional interaction (interaction among three or more factors) could have great potential for supporting the sensitivity analysis between the degree of cross-subsidies and the socio-economic factors.

Table 3- 1: The XOR digital logic gate example for interaction effect

FACTOR A	FACTOR B	FINAL OUTPUT OF XOR GATE
0	0	1
0	1	1
1	0	0
1	1	0

Sensitivity analysis methods can be classified into three categories: 1) mathematical; 2) statistical; or 3) graphical. For the statistical methods, input factors are assigned probability distribution and assessing the effect of variance in factors on the output distribution [71]. By allowing one or more factors to vary simultaneously, the statistical methods show superiority over the others owing to the aware of interaction effect among factors. The widely applied statistical methods for the sensitivity analysis include the regression analysis, Mutual Information (MI) index, Response Surface Methods (RSM), Fourier Amplitude Sensitivity Test (FAST), and ANOVA.

However, the socio-economic factors, such as personal information factors (age, education level and job), environmental awareness and household appliance factors etc., are collected from the questionnaire to depict the status of the customers. Thus, the socio-economic data is discrete, and the number of factors is considerable. The methods like RSM and regression analysis are not appropriate to deal with the sensitivity analysis which has large inputs. Moreover, the reliability to the FAST method is poor for the discrete inputs [72]. Finally, the ANOVA has been employed to evaluate the contribution of interaction among factors on the future bill uncertainty. Because ANOVA but not MI is often significant if one factor with a small frequency had a remarkable difference in the output [73]. Therefore, ANOVA is more suitable for the dataset with a relatively small sample size.

B. Computational expense and latent information loss

By considering the high-dimensional interaction effect instead of the single socio-economic factor, the computational complexity of this problem is exponential complexity with the increase of the input size.

To address the computation issue, the traditional dimensionality reduction techniques such as feature selection or feature extraction will abandon some feature to decrease the computational burden. However, the abandoned factor which does not demonstrate significant effect by its own could have a significant effect when it interacts with other factors. This kind of latent information hiding in the interaction of non-promising factors would be lost through roughly dimensionality reduction process.

Consequently, this paper adopts the beam search structure which is an optimization of best-first search that decrease the computation burden. Therefore, the beam search allows every factor to interact with all the other factors but keeps several most promising factors' combinations to do further exploration, which called the beam width. The heuristic pruning rule keeps the size of the beam small to reduce the computational burden. In the paper, the ANOVA test result will be used as the pruning rule. Meanwhile, all the pruned socio-economic factors would be recycled to minimise the occurrence of information loss caused by undetected by ANOVA.

3.2.3 The KLAM Beam Search Algorithm

To overcome the two challenges mentioned above, this paper proposes a novel KLAM beam search algorithm. The fundamental object of the proposed algorithm is to detect the significant socio-economic factors which influence the variation of customers' cross-subsidies.

In the proposed KLAM algorithm, the ANOVA is applied as the pruned rule of the beam search to reduce the computational burden. Then, the KL-divergency and GMM are employed to recycle all pruned off factors to minimize the latent information loss. Therefore, as Figure 3-1 shown, the novel KLAM algorithm consists of two stages which are: 1) Significant factors detection; 2) Pruned factors recycling. The detailed introduction of the algorithms applied at each stage would be further discussed in the following sub-section.

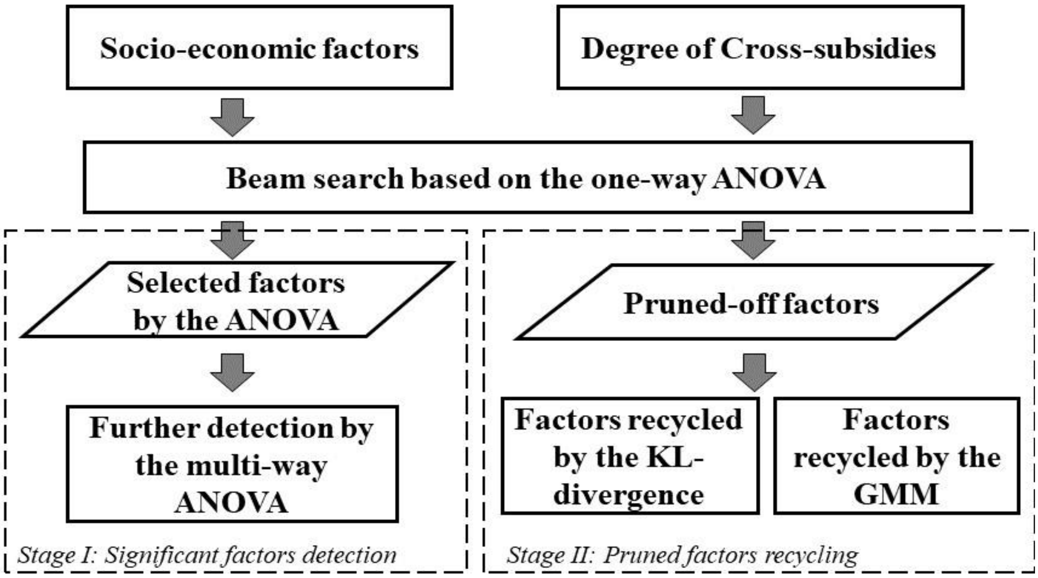


Figure 3- 1: The structure of the KLAM beam search algorithm

- ***Algorithm used at Stage I***

In this stage of the KLAM method, the primary objective is to reveal the significant socio-economic factors or their combinations, which influence the variation of customers' future energy bills after the cross-subsidies have been removed. This is implemented by the high-dimensional interaction-aware sensitivity analysis based on ANOVA. Since considering the high-dimensional interaction among factors is an exponential complexity problem, therefore, to reduce the computational burden, the beam search strategy has been applied to detect the significant socio-economic factors' combination.

1) Beam search

The structure of the beam search is demonstrated as Figure 3-2. The black node is the expanded node, which represents the social-economic factor at this node has a considerable effect on the cross-subsidies which make it has the chance to interact with other socio-economic factors in the further multi-way ANOVA. The number of the expanded nodes is the beam width of the beam search. On the opposite side, the white node indicates that the social-economic factor does not perform a significant influence on the cross-subsidies, which should be pruned off to reduce the computational burden. In other words, the ANOVA test result performs as the pruning rule in the proposed KLAM algorithm.

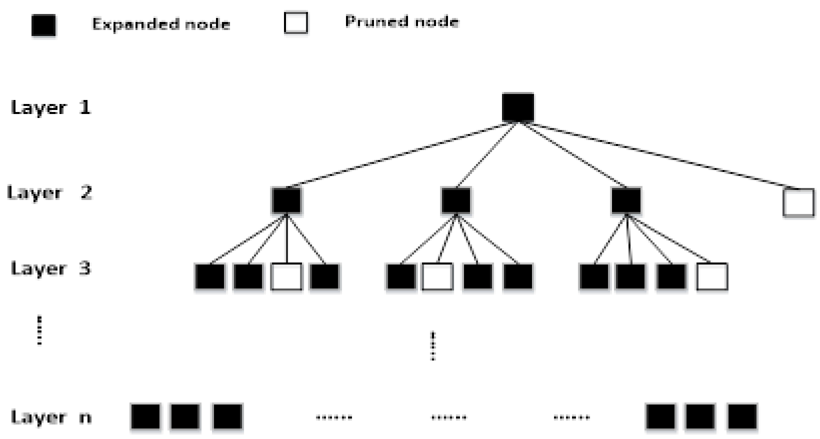


Figure 3- 2: The beam search structure in the KLAM beam search algorithm

2) ANOVA

ANOVA is the most efficient parametric method, which is devised to test the differences between several different groups of treatments, thus circumventing the problem of making multiple comparisons between the group means using t-tests. In this paper, each socio-economic factor can take on a certain number of choices, which are defined as levels of the socio-economic factor. The levels' combinations of multi socio-economic factors are called treatment.

In the typical application of ANOVA, the null hypothesis (h_0) assumes that the means of the treatments are identical. Then, the test result of ANOVA would be utilized to decide whether the null hypothesis should be accepted or not. The test result of ANOVA can be calculated by the F-test. The F-value is a ratio of two variances, the between treatments' variances and the within treatments' variances, which formulated as (3-1) shown:

$$F = \frac{\sum_{t=1}^{T} J \times (\overline{X_t} - \bar{X})^2 / (T-1)}{\sum_{t=1}^{T}\sum_{j=1}^{J} (x_{tj} - \bar{X})^2 / T \times (J-1)} \tag{3-1}$$

where the subscript j denotes the observations of every treatment t, j taking value from 1 to J. x_{tj} denotes the j^{th} observation in the t^{th} treatment group, $\overline{X_t}$ is the mean in the t^{th} treatment group and the $\bar{X}$ represents the overall mean value of the data.

The F-value indicates the number of times that the variance between groups exceeds the variance within the group. With the magnitude of F-value, the probability of obtaining this F-value by chance can be tabulated, which is called "p-value". If there is no significant difference between treatments just as the null hypothesis said, then the variance calculated from the between-groups sums of squares and the within-group sums of squares should be the same value. Therefore, a threshold value has been set to determine whether the null hypothesis should be accepted or not. If the magnitude of F-value is equal or greater than the value tabulated at the 5% level of probability, the null hypothesis that the effect of treatments is identical is rejected. The p-value is gained by tabulated F-test results calculated through ANOVA. With the magnitude of F-value, the p-value represents the probability of obtaining this F-value by chance. The higher the F-value is, the larger the relative variance among the group-means is. Thereby, the p-value will be smaller for the factor which has a significant effect on the variation in ANOVA significant test.

For this high-dimensional interaction-aware sensitivity analysis, there are more than one socio-economic factor (independent). Hence, the multi-factor ANOVA model is used. Using the two-way ANOVA as an example, the null hypotheses are: 1) there is no difference in means of factor A; 2) there is no difference in means of factor B; and 3) there in no interaction between factors A and B. Therefore, the F-value for each of those three cases the null hypothesis is needed to be tested in the two-way ANOVA model. Specifically, the variation of the response variable can be modeled as a linear combination of the effects of two factors and their interaction effect, which is given in (3-2)

$$V_{mdk} = \mu + e_{mdk} + A_m + B_d + (AB)_{md} \tag{3-2}$$

In this model, V_{mdk} is the value of the k^{th} observation of the m^{th} level of factor A and the d^{th} level of factor B. μ is the overall mean of the observations and e_{mdk} is a random element reflects the natural variation and errors of measurement. A_m and B_d represent the main effects of the two factors, and the interaction effect of those two factors are expressed by $(AB)_{md}$.

Therefore, due to the F-test of the interaction term, $(AB)_{md}$, the interacting factors could be detected even the effect of each factor own is insignificant.

- ***Algorithm used at Stage II***

Due to the high computation expense of the high-dimensional interaction-aware sensitivity analysis, the beam search has to prune the non-promising factors according to the ANOVA results to decrease the computational burden. However, the beam search cannot guarantee that it will find the optimal solution. Therefore, *Stage II* aims to ameliorate the latent information loss by recycling the pruned socio-economic factors.

For the proposed KLAM beam search method, the p-value from the ANOVA is applied as the pruning rule to decide which factors would be expanded next. Therefore, the main information loss at *Stage I* of the proposed KLAM methods always happen in two circumstances where the F-value fails to reflect the significance:1) Small F-value due to the non-normal distribution; 2) Small F-value due to the overlapping of distributions. The detailed solutions for the information loss occurred in those two circumstances will be demonstrated respectively in the following sub-sections.

1) *Due to the non-normal distribution*

The ANOVA, as a sensitivity analysis method, has a disadvantage which is if the distribution of the response variable significantly departures from normality, the test result of ANOVA may not be robust [76]. Therefore, if the distribution of cross-subsidies β_c for a treatment group departures from the assumption of normality, the F-value and its corresponding tabulated p-value will fail to reveal the significant impact of this treatment. An extreme instance is demonstrated in Figure 3-3 where the blue distribution and the red distribution (the non-normal distribution) have the same mean values but different shapes. However, the unusual distributions cannot be detected through the ANOVA at *Stage I* which causing the information loss.

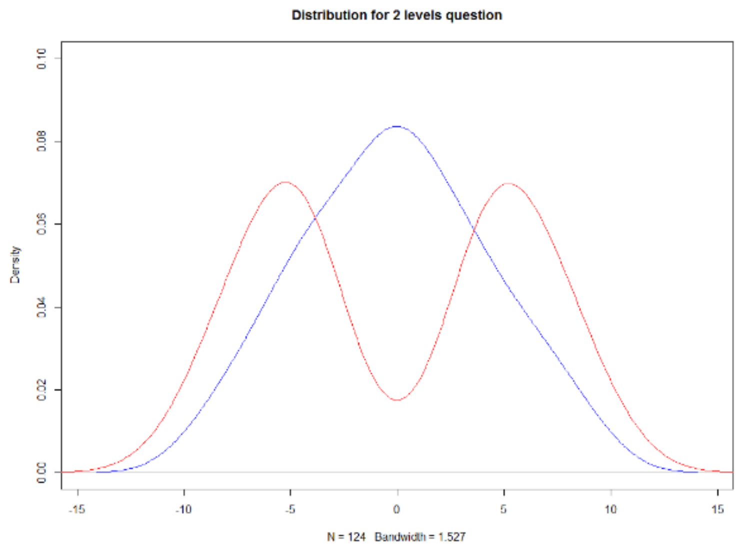


Figure 3- 3: The example of information loss caused by the non-normal of treatments' distributions

Hence, the proposed KLAM method adopts the KL-divergence at the Stage II to avoid this kind of latent information loss. The KL-divergence measures how one probability distribution diverges from a second expected one [77]. Suppose that

p_0 and p_1 are two probability densities, the KL-divergence is defined as (3-3) displayed.

$$D_{KL}(p_0 \parallel p_1) = \int_{-\infty}^{\infty} p_0(x) \log \frac{p_0(x)}{p_1(x)} dx \tag{3-3}$$

However, the KL-divergence is not symmetric. $D_{KL}(p_0 \parallel p_1)$, which represents the distance from p_0 to p_1, is usually different from the distance from p_1 to p_0. The alternative averaged KL-divergence, $J(p_0, p_1)$ [78] is employed here just as (3-4) formulated below:

$$J(p_0, p_1) = \frac{D_{KL}(p_0 \parallel p_1) + D_{KL}(p_1 \parallel p_0)}{2} \tag{3-4}$$

All the pruned-off socio-economic factors will be recycled through the KL-divergence to avoid the information loss due to the non-normal distributions.

2) Due to the overlapping of the distributions

When several factors are interacting, the significance of one distinguish treatment could be reduced due to the overlapping of other treatments' distributions. The similar mean values for the overlapped treatments will lead to a relatively small growth of the F-value at Stage I. Consequently, there is a high probability that the corresponding interacted socio-economic factors would be pruned off at the second step of the *Stage I*. For example, there are two factors, and every factor has two levels. The distributions of the four treatments are shown as Figure 3-4. The interaction effect between those two factors is not such significant due to there are three distributions do not distinguish from each other. However, the distribution labelled with $d_{A,1}$, which represents the level of first factor is "A" and the level for the second question is "1", is outstanding and demonstrates a different effect from other three treatments. Therefore, this kind of information is worth to do further analysis, however, cannot be detected in ANOVA.

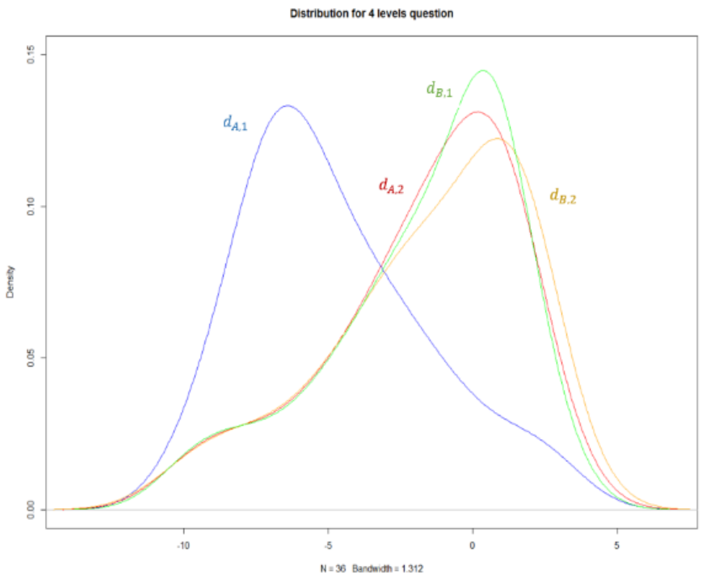


Figure 3- 4: The example of information loss caused by the overlapping of treatments' distributions

To ameliorate this kind of information loss, the Gaussian Mixture Model (GMM) has been adopted in the Stage II. The GMM can make statistical inference about the properties of the sub-population when the probability density function (pdf) of the "hidden" subpopulation is the Gaussian distribution, which is also known as Gaussian mixture components. By this way, the outstanding distribution will be separated from other overlapped treatments, as Gaussian mixture component.

When GMM is applied at the *Stage II*, there are two essential purity values still need to be tested.

1) A ratio ρ_t is employed to illustrate the purity of a specific treatment t in g^{th} Gaussian mixture component.
2) A ratio ρ_g to represents the purity of the g^{th} Gaussian mixture component for all observations with treatment t.

Only both of ρ_t and $\rho_g$are high and over the purity threshold value Γ_ρ, it can guarantee that in the g^{th} Gaussian mixture component, all observations are mainly from one treatment t and most of the observations with treatment t are only contained in Gaussian mixture component g. In this way, the specific treatment for the interacting factors could be taken into consideration again

3.2.4 The Implementation of the KLAM Searching Algorithm

To validate the efficacy of the KLAM beam search algorithm, this section demonstrates the proposed algorithm on a real residential smart metering database of Irish households which is collected from the Smart Metering Electricity CBTs and publicized by Commission for Energy Regulation of Ireland. The smart data is recorded for over 1000 customers at a half-hourly basis, from July 2009 to December 2010. Finally, once the incomplete data were removed, the yearly smart metering data for 993 residential households are involved in presenting the case study. Additionally, the Irish project also surveys the views of the socio-economic status for those households through a questionnaire which is demonstrated in Appendix A. With the real smart metering data and the socio-economic information of the real Irish households, the implementation of the high-dimensional interaction-aware sensitivity analysis can be achieved through four steps. The four steps are 1) Socio-economic factors cleaning; 2) The cross-subsidies quantification; 3) *Stage I:* significant socio-economic factors detection; and 4) *Stage II:* pruned socio-economic factors recycling. The detailed introduction of the four steps would be given in the following sub-section and the notations used in this section are listed in Table 3-2.

Table 3- 2:Notations of the algorithm

Notation	***Description***
q	Socio-economic question index
f_q	Socio-economic factor index for question q
c	Customer indicator
l_c^q	The level chosen by customer c for question q
Q	The set for all socio-economic questions
L_{f_q}	Total number of levels for question q
δ_{sc}	Energy bill calculated by half-hourly settlement for customer c
d	Day indicator
h	Half-hour indicator
t_{start}	The start date of the smart meter record
t_{end}	The end date of the smart meter record
S_{cdh}	Smart metering data of a customer at one half-hourly period
p_{dh}	The wholesale price of the h^{th}half-hour on d^{th}day
δ_{pc}	Energy bill calculated by TLP-based settlement for customer c
LP_{dh}	The load profile value for the h^{th}half-hour on d^{th}day
α_c	The annual advanced consumption for customer c
E_{cdh}	The TLP estimated consumtion for customer c the in the h^{th}half-hour on d^{th}day
GSP_{dh}	The GSP correction factor of the h^{th}half-hour on d^{th}day

β_c	The degree of cross-subsidy for a customer c
W	Maximum beam width for the search
w	Counter for beam width
n	Index for the interaction factors'number
$f_q^{selected}$	Factors which pass the one-way ANOVA significant test
$I_{nw}^{f_q^{selected}}$	Factor subset which is used to store factors which interacted with $f_q^{selected}$
f_q^{rest}	Factors in the complement of the set $I_{nw}^{f_q^{selected}}$
$f_{q_{rest}}^{nw}$	The w^{th} new factors added in the $I_{nw}^{f_q^{selected}}$ which interacted with $f_q^{selected}$ at n-way ANOVA
f_q^{Pruned}	All socio-economic factors which do not be marked as $f_q^{selected}$
t	The index of the treatment, take values from 1 to T
T	The total number of the treatments
g	The index of Gaussian misture component
G	Total number of Gaussian Misture components
ρ_t	The purity of a specific treatment t in g^{th}Gaussian mixture component.
ρ_g	The purity of the g^{th} Gaussian mixture component for all observations with treatment t.
Γ_ρ	The threshold value of purity $\rho_g$and ρ_t

A The socio-economic factor definition

The socio-economic information for every household is collected through a questionnaire. Two types of variables are applied to translate the survey answers to the factors: 1) dummy variable; and 2) ordinal variable.

The dummy type factors only have two value, "1" and "0", which are used to sort data into mutually exclusive categories, i.e. smoker/ non-smoker [74]. The variables in the ordinal factor type have natural, ordered categories [75]. These factors translate the questions whose options exist on an ordinal scale, for example, "Is your general health poor, reasonable, good, or excellent?" may have those answers coded as 1, 2, 3, and 4 respectively.

Therefore, every question can be treated as a socio-economic factor and the options for a question are defined as levels of the factor. Among multi-factors, the combination of levels defined as treatment. Each socio-economic factor contains the level chosen by every customer which is defined in (3-5) and the notations are listed in the following table.

$$f_q = \{l_1^q, l_2^q, \cdots, l_c^q\} \tag{3-5}$$

where the question index $q \in [1, Q]$and the total number of the level for the question

q is represented by L_{f_q}. Finally, the customers' answers of the questionnaire are translated into 142 dummy or ordinal socio-economic factors

B. The socio-economic factor definition

The uncertainty The uncertainty of customers' future bill is caused by the transferring from traditional TLP settlement to the smart metering based half-hourly settlement which can remove the cross-subsidies. Therefore, the cross-subsidies will be quantified by comparing the difference of the electricity bills accessed under two settlement process.

Due to nearly 60% of the energy cost is coming from the wholesale market in the UK, the calculation will be scaled up based on the wholesale market cost. The accurate bills without cross-subsidies can be quantified by multiplying the usage with the corresponding wholesale market price and accumulating in (3-6):

$$\delta_{sc} = \sum_{d=t_{start}}^{d=t_{end}} \sum_{h=1}^{h=48} S_{cdh} \times p_{dh} \tag{3-6}$$

The energy bills estimated through the traditional TLP settlement process is commonly used when the smart meters are absence. Customers' accumulative consumption has been read seasonally or half-annually. The TLP represents the load shape for all residential customers. The real half-annual accumulative consumption would be allocated into each half-hour following the shape of TLP.

Therefore, the first step of the traditional TLP settlement is to gain the annual advanced consumption, α_c, by scaling up the real accumulated metered consumption to the annual level, based on the TLP fraction between a year and the corresponding period, just as (3-7) shown below.

$$\alpha_c = \sum_{h=1}^{h=48} \left(\sum_{d=t_{start}}^{d=t_{end}} S_{cdh} \times \frac{\sum_{d=1}^{d=366} LP_{dh}}{\sum_{d=t_{start}}^{d=t_{end}} LP_{dh}} \right) \tag{3-7}$$

Then, the next step is allocating the annual advanced consumption α_c into each half-hour period as (3-8) demonstrating.

$$E_{cdh} = \frac{LP_{dh}}{\sum_{d=1}^{d=366} \sum_{h=1}^{h=48} (LP_{dh})} \times \alpha_c \tag{3-8}$$

In the practical traditional settlement process, the third step aims to correct the profiled error by creating a Grid Supply Point Group Correction Factors (GGCFs). The GGCFs are used to ensure that the total energy allocated to suppliers in each settlement period (30 minutes) equals to the energy entering the Grid Supply Point (GSP) groups from the transmission system. However, due to the limit customer numbers, this research assumes that all customers are connected to the same supply point and the GSP Group Take equals to the total customers' real consumption sum during each half-hour period. Therefore, for every half hour, the GGCFs, GSP_{dh}, are calculated as (3-9) shown to fix the over or under energy accounting issue.

$$GSP_{dh} = \sum_{c=1}^{c=C} [\frac{S_{cdh}}{\frac{\alpha_c}{\sum_{d=1}^{d=365} \sum_{h=1}^{h=48} LP_{dh}} \times LP_{dh}}] \tag{3-9}$$

Finally, the energy bill estimated through the traditional TLP settlement process can be calculated in (3-10).

$$\delta_{pc} = \sum_{d=t_{start}}^{d=t_{end}} \sum_{h=1}^{h=48} E_{cdh} \times GSP_{dh} \times p_{dh} \tag{3-10}$$

To better express the future bills' uncertainty for every customer instead of simply minus the two bills, a parameter β_c is developed in (3-11) to describe the degree of impact on customers' bill after removing the cross-subsidies.

$$\beta_c = \frac{(\delta_{sc} - \delta_{pc}) \times 100\%}{\delta_{pc}} \tag{3-11}$$

C. Stage I: significant socio-economic factors detection

After the generation of the socio-economic factors and quantification the cross-subsidies, the proposed KLAM beam search method would be employed to reveal the significant socio-economic characteristics which influence the variation of customers' future bill. The flowchart of *Stage I* has been displayed in Figure 3-5. The whole procedure could be introduced in 3 steps：

1) Step 1: Using one-way ANOVA to test the significance of every individual socio-economic factor on cross subsidies. The significant factors who pass the one-ANOVA (p-value < 5%) are marked as $f_q^{selected}$.
2) Step 2: For every $f_q^{selected}$, it has the chance to interact with all other socio-economic factors (two-way ANOVA). However, to reduce the computation burden, the beam search sets the beam width W as three. Therefore, only the top three factors which obtain the smallest p-value (meanwhile, p-value requires to smaller

than 5%) when interacting with the $f_q^{selected}$. Those three socio-economic factors which have significant interaction effect with the $f_q^{selected}$ will be marked as as $f_{q_{rest}}^{nw}$ and respectively store in corresponding set $I_{nw}^{f_q^{selected}}$ (where n $= 2$ now and w changes from 1 to 3).

3) Step 3: Then, the interaction effect test for each $f_q^{selected}$ would further move to next layer and repeat step 2. Socio-economic factors which belong to set $\complement_Q I_{nw}^{f_q^{selected}}$ would have chance to do the ANOVA test with factors saved in set $I_{nw}^{f_q^{selected}}$.

4) Step 4: Finally, there are n socio-economic factors stored in set $I_{nw}^{f_q^{selected}}$ which passed the n-way ANOVA and have a significant interaction effect with $f_q^{selected}$. Until the complement set, $\complement_Q I_{nw}^{f_q^{selected}}$, becomes an empty set, the factors combination stored in every $I_{nw}^{f_q^{selected}}$ would be output.

D. Stage II: pruned socio-economic factors recycling

Due to the high computation expense, the beam search must prune the non-promising factors according to the ANOVA results, which could cause the latent information loss. Therefore, to minimise the information loss, *Stage II* is built to recycle all pruned socio-economic factors by KL-divergence and GMM algorithms.

The Program 2 displays the information recycling algorithm with pseudo code of KL-divergence and GMM below.

Program 2: Information recycling based on KL-divergence and GMM

1: Load all the pruned off social-economic vectors f_q^{Pruned} and the degree of cross-subsidies β_c

2: **For** every f_q^{Pruned}**:**

3: Calculate the distribution of each level, p_0 and p_1

4: Calculate the averaged KL-divergence value of the $D_{KL}(p_0 \parallel p_1)$ and $D_{KL}(p_1 \parallel p_0)$ as equation (3-6) shown

5: Sort each pair of the two levels by their KL-divergence value and do the significant test of the two levels' effect on the cross-subsidies β_c.

6: Sent the factor f_q^{Pruned} back to Stage I if any of its two levels have relatively high KL-divergence and passed the significant test.

End

For the GMM algorithm

For each two of the pruned-off factors: f_{qA}^{Pruned} and f_{qB}^{Pruned}**:**

Calculate the maximum treatment number of the two socio-economic factors: $T = L_{f_{qA}^{Pruned}} \times L_{f_{qA}^{Pruned}}$

For G from 2 to T**:**

Assign the cross-subsidies β_c into G Gaussian mixture components and label the customers with the mixture component number g they belong to.

Calculate the purity for the specific t^{th} treatment in customers labeled by g^{th} Gaussian mixture component: ρ_t

Calculate the purity for the g^{th} Gaussian mixture component label in all customers have the t^{th} treatment: ρ_g

If $(\rho_g \geq \Gamma_\rho) \wedge (\rho_t \geq \Gamma_\rho)$ then:

Do the significant test of the t^{th} treatment:

If P-value < 5% then:

Output the specific t^{th} treatment of those two socio-economic f_{qA}^{Pruned} and f_{qB}^{Pruned}

End if

End if

G=G+1

End For

End

Terminate

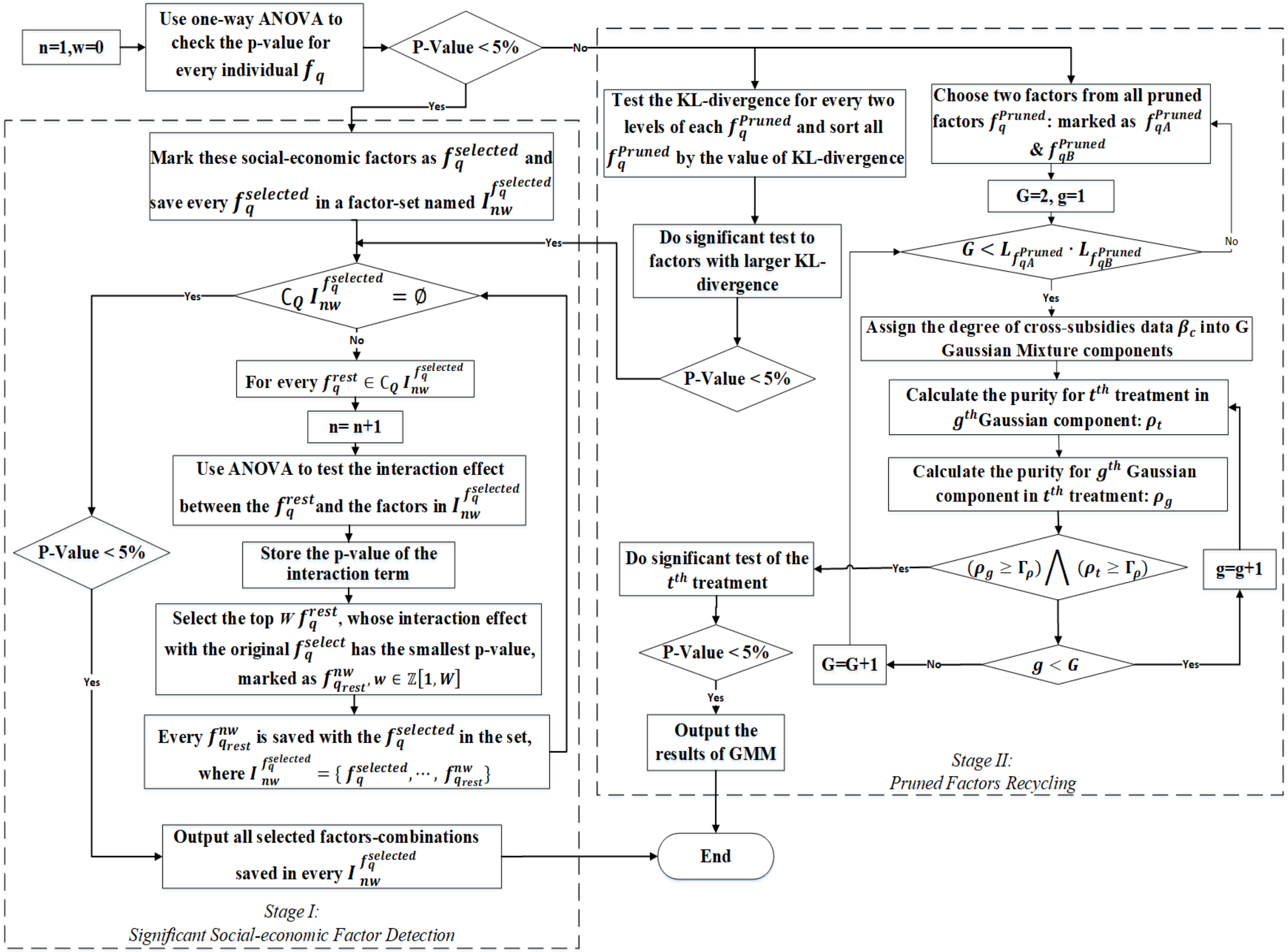


Figure 3- 5:The flowchart of the KLAM beam search algorithm

3.2.5 The Results Analysis of the Case Study

Then, the answers of customers are translated into 142 dummy or ordinal socio-economic factors.

After quantifying the degree of the cross-subsidiesβ_c for each household, the sensitivity analysis between the β_c and the interacting socio-economic factors will be explored through the KLAM beam search algorithm. The beam width W is set as 3. The inner implications of social status and the uncertainty of energy bills under the new half-hourly settlement will be discussed.

In general, 69 socio-economic factors have been highlighted by the proposed KLAM algorithm that the interaction-effect among them illustrates a significant influence on the change of customers' future energy bill. Among them, 51 factors are detected at *Stage I* of the KLAM algorithm and the rest 18 factors are discovered through the recycling at *Stage II*. In the following sub-section, a detailed introduction of those significant interacted factors would be given.

A. Significant socio-economic factors detected at Stage I

Seventeen of the 51 socio-economic factors demonstrate their impaction individually, with a p-value which less than 0.05, through the one-way ANOVA at the first step of *Stage I*. Those 17 original factors are marked as the $f_q^{selected}$ and start to interact with other factors. Due to the difference in $L_{f_q}$for different factors and the limitation of the involved household number, three original factors show a significant effect on the β_c (p-value <0.05) in a four-way interaction and other 14 original factors have more outstanding significance in a five-way interaction. The detailed interacting socio-economic factor groups are illustrated in the Table C-1 in the Appendix C. Table 3-3 only lists the content of each original factor with the p-value before and after the interaction. It illustrates the amelioration of the p-value for each original factor after interacting with other factors. The bold text represents the original factors in the four-way interaction. After the interaction, the p-value for every factor combination is decreased. It is evident that the interaction effect among the 4 or 5 socio-economic factors could promote their influence on the amount of cross-subsidies β_c.

Table 3- 3: Original factors and their P-values detected in Stage I

Number	*Content of Question*	*Individually P-value*	*P-value for the interaction term*
Qu 84	The number of the electric cookers you own	2.83×10^{-6}	$\mathbf{3.48\times 10^{-7}}$
Qu 58	Describes how you cook	1.91×10^{-5}	$\mathbf{1.37\times 10^{-6}}$
Qu 97	The number of the Lap-tops you own	2.63×10^{-4}	$\mathbf{7.63\times 10^{-5}}$
Qu 92	Do you have Lap-top	2.99×10^{-4}	$\mathbf{1.33\times 10^{-4}}$
Qu 104	**How often would you use electricity cooker**	7.46×10^{-4}	$\mathbf{5.10\times 10^{-4}}$
Qu 41	**Do you have plug in electricity heaters**	1.33×10^{-3}	$\mathbf{1.00\times 10^{-3}}$
Qu 3	The employment status of the CIE*	2.36×10^{-3}	$\mathbf{1.17\times 10^{-3}}$
Qu 90	**The number of TVs greater than 21 inch you won**	2.41×10^{-3}	$\mathbf{1.51\times 10^{-3}}$
Qu 9	The description the people you live with	4.23×10^{-3}	$\mathbf{2.69\times 10^{-3}}$
Qu 105	How often would you use the plug-in electricity heater	6.97×10^{-3}	$\mathbf{5.20\times 10^{-3}}$
Qu 2	Age of the CIE*	1.13×10^{-2}	$\mathbf{5.57\times 10^{-3}}$
Qu 5	Do you have internet access in your home	2.46×10^{-2}	$\mathbf{9.07\times 10^{-3}}$
Qu 11	How many people typically in the house during the day	1.25×10^{-2}	$\mathbf{9.82\times 10^{-3}}$
Qu 53	Do you use solid fuel boiler to heat water	2.46×10^{-2}	$\mathbf{1.48\times 10^{-2}}$
Qu 4	the Occupation of CIE*	3.40×10^{-2}	$\mathbf{2.44\times 10^{-2}}$
Qu 91	Do you have the Desk-top computer	4.21×10^{-2}	$\mathbf{3.03\times 10^{-2}}$
Qu 99	The number of the Wash Machines you own	3.72×10^{-2}	$\mathbf{3.28\times 10^{-2}}$

B. Significant socio-economic factors recycled in Stage II

To ameliorate the latent information loss, the KL-divergence and GMM are adopted at *Stage II* to recycle the abandoned factors which do not pass the significant test in the one-way ANOVA. The results of both will be detailed illustrated in the following sub-section.

1) *Significant results detected by KL-divergence*

Firstly, for every pruned-off factor at this stage, the KL-divergence between every two levels of each factor has been evaluated. Then, there are 5 socio-economic factors whose two specific levels have high KL-divergence value and pass the significant test. The passing of significant test proofs that the different choice of those two levels would impact the final bill changing. Finally, the 5 factors are sent back to the *Stage I* as original factors, $f_q^{selected}$ to interact with other factors.

The output found out from the *Stage I* for those 5 recycled factors is demonstrated in Table 3-4. The detailed interacting socio-economic factor groups are demonstrated in the Table C-2 in the Appendix C. Due to only the partial customers, who choose level A or B for each factor, could be taken into account for the further interaction-aware sensitivity analysis, this reduces the data size to a certain extent. Thus, all those 5 recycled factors show an enhanced effect on β_c in four-way interactions than their own.

Table 3- 4: Original factors and their P-values detected in Stage II by KL-divergence

Number	*Content of Question*	*Level A*	*Level B*	*Original P-value*	*P-value for the interaction term*
Qu102	How often you use the electric shower	1 (< 5mins)	3 (10-20mins)	$1.74 * 10^{-3}$	$\mathbf{7.70 \times 10^{-4}}$
Qu 95	The number of TVs bigger than 21 inch	0 (None)	2 (Two)	$4.85 * 10^{-3}$	$\mathbf{2.83 \times 10^{-3}}$
Qu 29	You cannot get the people you live with to reduce their usage	1(Agree)	2(Less agree)	$2.33 * 10^{-2}$	$\mathbf{1.21 \times 10^{-2}}$
Qu 7	Do you regularly use the internet	1(Yes)	2 (No)	$2.66 * 10^{-2}$	$\mathbf{2.16 \times 10^{-2}}$
Qu 32	Do you agree with reduction of usage would not make a change to your bill	2(Agree)	3(Less agree)	$2.83 * 10^{-2}$	$\mathbf{1.82 \times 10^{-2}}$

2) *Significant results detected by GMM*

The aim of GMM is to reveal the significant treatment of two different socio-economic factors whose significance has been weakened by the overlapped distributions of other treatments and finally result in the pruning at Stage I. By setting the $\Gamma_\rho$equals to 70%, there are 8 pairs of socio-economic factors which has significant treatments have been detected by GMM.

Due to the space limitation, this section only demonstration the interaction of one pair of factors (Qu. 64 and Qu. 85) instead of all eight pairs. The Qu. 64 surveys whether the households had to go without heating due to lack of money during the last 12 months. Qu. 85 is about the number of the electric heaters the households own. From (a) in Figure 3-6, it is visible that the distributions of treatment 1 (the red one) and 2 (the blue one) are distinct from each other. Their significant test is passed with p-value equals to 0.021. The result exposes that even the households have already stop heating because of lack of money, they would still experience an energy bill increasing if they own electric heaters. However, as Figure 3-6(b) displays, the overlapped distributions of other treatments have weakened its significance.

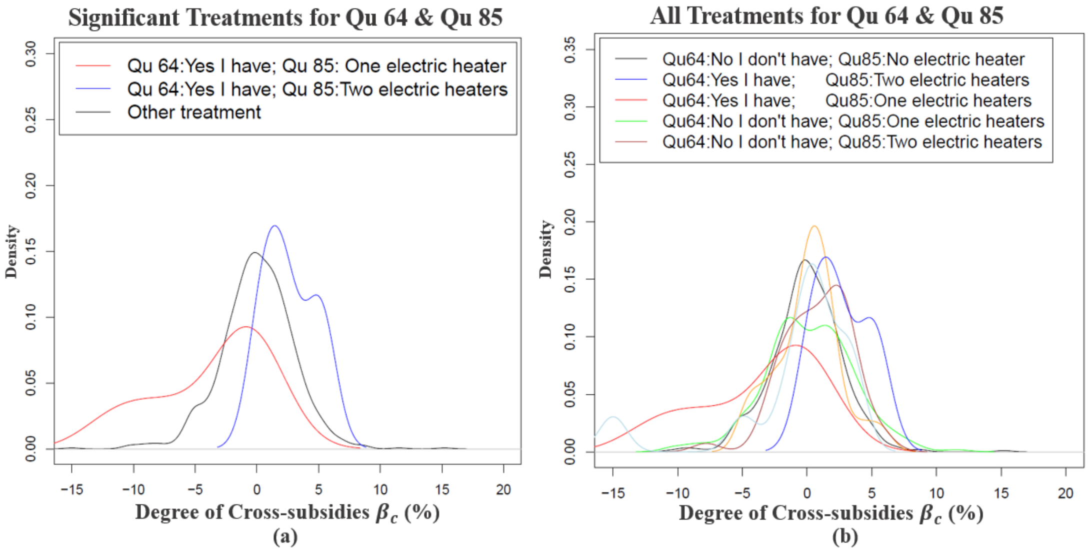


Figure 3- 6: Interaction effect of one pair of factors found in Stage II

3.2.6 Analysis and Discussion

After the proposed KLAM algorithm found out the significant interacted socio-economic factors, further analysis is required to understand which treatments would have a positive influence on the energy bill increasing under the new half-hourly settlement process.

In the real database, for diverse factors' combinations, the number of households in different treatments is dissimilar. Thereupon, this paper sets the confidence level as 95% to calculate the upper and lower bounds for the mean value of the cross-subsidiesβ_c for every treatment. The emphasis of this analysis is not on the treatments with a narrow bound, but on the treatments whose upper and lower bounds are the same sign. For example, if the upper and lower bounds of β_c for one treatment are both bigger than zero, we have 95% confidence that the future energy bills will increase for the households whose socio-economic makeups conform to this treatment.

Due to the limitation of space, this paper demonstrates a partial of interacted-factor combinations in Table 3-5. Meanwhile, by comparing the bill increasing and decreasing group, Table 3-5 also displays the characteristic of every socio-economic factor in each combination. The reason why those factor-combinations are selected to be shown is that they perform some contrary conclusions to common sense in the literature.

In previous studies [40], dwelling type factors, such as the number of rooms and the age of the house, and appliance factors have a significant positive effect on electricity usage. However, from Table 3-5 it can be observed that, with different interacting factors, the impact of one socio-economic factor on customers' bill change is inconsistent. Using a dwelling age factor (Qu 36) as an example, by interacting with different factors, both of the newer houses group and the older houses group could have a positive influence on customers' bill growth. Therefore, only with a particular treatment for a factor combination, there would be a credible result of the future bill changing. Moreover, the factors related to the employment status of the CIEs (Qu 3) and environment awareness (Qu 15) have been found a significant effect in this paper which is contrary to the conclusions in [27, 35, 40] that reported they are insignificant. Those opposite characteristics for the same factor have been highlighted in the same colour in Table 3-5.

The inconsistent conclusions of previous literature and this research proves that 1) the bill difference caused by removing the cross-subsidies is different and more complicated than the amount of electricity usage. 2) The influence of single factors is not robust enough to provide a clear relationship with this problem. Hence, the high-dimensional interaction-aware sensitivity analysis between socio-economic factors and the bill change is necessary.

Finally, based on the outcomes found in this paper we highlight several contrary conclusions:

1) The electricity appliance factors and dwelling related factors, unlike that they influence the consumption, do not always have a positive effect on the energy bill increasing.
2) The factors related to the CIEs' employ status and their environmental awareness

ought to gain attention because they can influence the future bill variation significantly by interacting with other factors.

Table 3- 5: Socio-economic factors show different impact on the cross-subsides

Combination 1	*Qu 97 : The number of the Lap-tops you own*	*Qu 90 : The number of TVs greater than 21 inch you won*	*Qu 74 : Do you have the electric cooker*	*Qu 15 : Do you interested in changing the way you use electricity if it helps the environment*	*Qu 86: The number of the stand-alone freezers you own*
Bill Increasing Group	**Less**	**More**	**More households have**	**Less households interested**	**More**
Combination 2	*Qu 3: The employment status of the CIE*	*Qu 77: Do you have a water pump/electric well pump / pressurised water system*	*Qu 90: The number of TVs greater than 21 inch you won*	*Qu 13: How many adults and children under 15 years old are typically in the house during the day*	*Qu 39: How many bedrooms are there in your home*
Bill Increasing Group	**More CIEs get a job**	**Same as the decreasing group**	**Less**	**Less people**	**More bedrooms**
Combination 3	*Qu 2: Age of the CIE*	*Qu 74: Do you have the electric cooker*	*Qu 92: Do you have Lap-top*	*Qu 100: How often would you use the Tumble Dryer*	*Qu 36: How old is your home*
Bill Increasing Group	**Elder**	**More households have**	**Less households have**	**More frequently and longer**	**Older house**
Combination 4	*Qu 4: Social Class of CIEs*	*Qu 75: Do you have the plug-in electric heater*	*Qu 9: The description the people you live with*	*Qu 100: How often would you use the Tumble Dryer*	*Qu 36: How old is your home*
Bill Increasing Group	**Most belong to Working/non-working class**	**More households have**	**More households have children younger than 15**	**Less frequently and shorter**	**Newer house**

Combination 5	*Qu 41: Do you have plug in electricity heaters*	*Qu 57: When heating is switched off, do you use your immersion*	*Qu 74: Do you have the electric cooker*	*Qu 3: The employment status of the CIE*
Bill Increasing Group	**More households have**	**More households use**	**More households have**	**More CIEs are retired**

Combination 6*	*Qu 102: How often you use the electric shower*	*Qu 4: Social Class of CIEs*	*Qu 15: Do you interested in changing the way you use electricity if it helps the environment*	*Qu 80: The number of Tumble dryers you own*
Bill Increasing Group	**More frequently**	**Most belong to Working/non-working class**	**More households interested**	**More**
Combination 7*	*Qu 29: You cannot get the people you live with to reduce their usage*	*Qu 39: How many bedrooms are there in your home*	*Qu 2: Age of the CIE*	*Qu 43: Do you use oil to heat your room*
Bill Increasing Group	**More households disagree**	**Less bedrooms**	**Younger**	**More households do**

*From KL-Divergence results

3.2.7 Conclusions

This paper investigates the socio-economic makeups of the customer groups that would be adversely impacted under the half-hourly settlement through a sensitivity analysis. To achieve this aim, a novel high-dimensional interaction-aware searching algorithm, the KLAM beam searching algorithm, has been proposed. After validating the proposed algorithm on a real dataset, this paper finds 14 five-way interacting factor-combinations, 8 four-way interacting factor-combinations and 8 pair of factors with significant treatment. By analysing the effect of the interacted factors, there are some characteristics for the bill-increasing group are unusual and even contrary to the common sense in the literature.

The paper contributes to a better understanding of the impact of transforming to the half-hourly settlement on customer groups with different socio-economic status. Furthermore, in the future, the algorithm proposed in this paper could be extended to apply on a larger dataset. This can support the regulators and policymakers both in the accurate vulnerable customer identification with a more comprehensive description of the customers' socio-economic condition and in the active policy implementation.

3.3 Chapter Summary

In this chapter, the socio-economic status of the customers who would be adversely affected in the half-hourly settlement process has been identified and analysed. There are two main contributions for this research, which are: 1) This research is the first time to assess the impact of the HHS reform launched by the Ofgem; 2) A novel interaction-aware searching method has been proposed which can find a significant and comprehensive socio-economic condition.

Taking the advantage of the effect of interaction among the socio-economic factors, there are some key findings which are contrary to previous studies:

1) The electricity appliance factors and dwelling related factors could have negative effect on customers' energy bill changing by interacting with other specific factors. However, those two kinds of factors are commonly found have positive effect on energy consumption.

2) The CIE's employ status, their age and environmental awareness can influence the bill variation significantly in specific factor-interacting-combinations. However, those factors are rarely caught the attention in the previous literature, which is because of the effect of the single of them is inconclusive (mixing effect) or even no significant effect.

By analysis the significant interacting socio-economic factor-combination, it can be found that in the new HHS process, the elder customers who owning more electric appliances or who living in an older house are most likely to be the new vulnerable customers who may need help from the government.

Chapter 4

The Impact of Network Cost Variation on Customers' Electricity Bills and Socio-Economic Status

THIS chapter proposes a novel distribution UHE network pricing method which can accurately allocate customers' network cost based on their usage data. Then, the significant social-economic criteria for the network bill variation will be given.

4.1 Introduction

Chapter 3 has developed a high-dimensional interaction-aware KLAM algorithm, which can identify the impact of directly removing the cross-subsidies for the wholesale market cost on the domestic customers through the significant and highly-interacted social-economic criteria. Identically, the cross-subsidies are also existing in the customers' network charges.

In practice, this cost-reflective network charging methodology only applies to retailers (i.e. suppliers in the UK) and large customers. The vast small customers (i.e. domestic homes and small business) are paying bills which mix up energy-based generation cost and capacity-based network cost. There is a clear mismatch when retailers pay the DUoS by peak power in kW but later charge small customers by volumes in kWh. As a result, the network price signal cannot be transmitted through the retail market and reach the end customers. Massive domestic customers are charged based on their energy consumption with the same unit cost throughout the year.

This limitation has introduced major cross-subsidies across customers, providing perverse incentives in the use of electricity and overstated incentive to the uptake of low carbon technologies regardless of time, location and operation approach. An example of this would be a customer with photovoltaic (PV) uses the network to export electricity in the daytime and import overnight. Over time, he may have zero net energy in terms of kWh and thus avoid the network cost completely, as Figure 4-1 demonstrated. Other customers will quietly bear this cost with higher electricity bills.

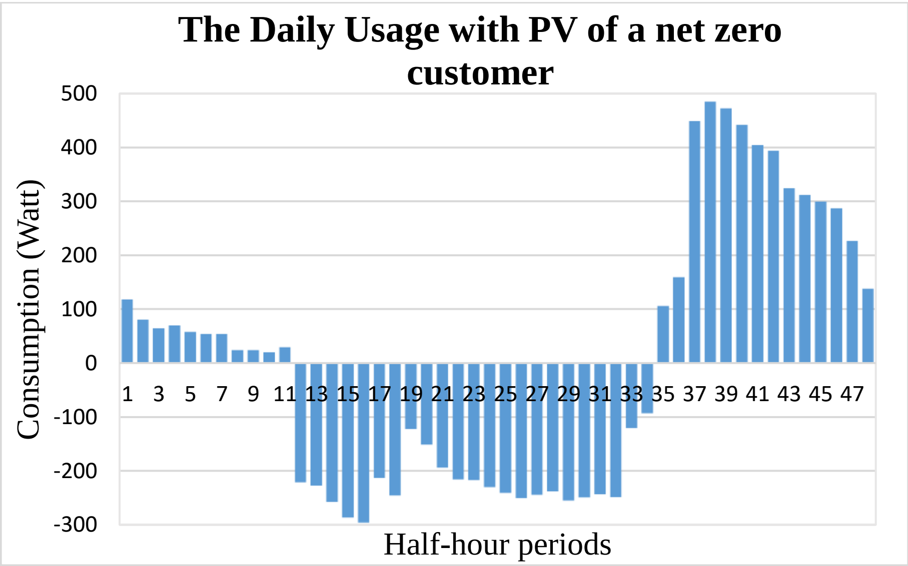


Figure 4- 1: Daily load profile for a PV net zero customer

The research question is how to design a distribution network pricing for small customers in the retail market. Different pricing methodologies have been proposed in the past for retailers and large customers. For example, to encourage the utilisation of energy storage, the concept locational marginal pricing (LMP) at the transmission system is adapted to distribution networks by formulating a distribution LMP [79]. An integrated distribution LMP method is presented in [80], which aims to mitigate the network congestion caused by the growth of the electric vehicles. In [81], the potential for employing the locational pricing to encourage tariff change is investigated, and an extensive review of network pricing is reported in [82]. In [83], nodal pricing in the distribution network has been proposed to reward the distribution generations for reducing line losses. The Long-run Incremental Cost Pricing (LRIC) [84, 85] provides locational cost-effective price signals to reflect customers' impact on network investment. Paper [86] proposes a novel fuzzy logic based network pricing to accommodate the future flexible load. At the low voltage networks, Distribution Reinforcement Model [87] is widely used, which is based on the yard-stick principle. The approaches in these papers mainly focus on network operation rather than the customs' characteristics. Although methodologies in [88, 89] passively react to a set of projected profiles of future demand and generation, they cannot proactively influence their behaviours on the network. Given the increasing penetration of responsive technologies at the household level, it is critical to extend network charging methodologies for small customers.

However, although different pricing methodologies had been proposed in many researches which are mentioned above, directly applying the existing methodologies to domestic customers will be extremely challenging due to high uncertainties of the network cost estimation. The current DUoS charges calculate the reinforcement cost by only considering the customers' contribution at the historical annual peak point. However, due to the exceedingly volatile usage behaviour of domestic customers, there is a high uncertainty between the historical peak point ant the real network peak point in the future. Customers with low or even negative consumption at the historical peak point can evade the network charges, even though they might create a new peak at a different time point in the future.

Therefore, the research objectives of this chapter are:

i) Extending the DUoS charging methodologies to small customers by proposing a novel network charging method which removes the cross-subsidies of network costs.

ii) Investigating the impact of the bill changing on customer with different socio-economic status.

This rest of this chapter is organised as follows. Section 4.2 introduces the new network charging methodology for an individual domestic customer. Section 4.3 represents the details of the case study for the propose distribution network pricing method. The performance of the case study has been demonstrated in Section 4.4. Then, Section 4.5 presents the significant social-economic characteristics for the identification of the impact of removing the network cost cross-subsidies on domestic customers. The conclusions are drawn in Section 4.6.

The structure of this chapter is written in an alternative-based format. The content of the novel distribution network pricing is prepared to submit to the IEEE Transaction on Smart Grid. The author is the third author of this work and collaborated with Dr Ran Li, Dr Shuangyuan Wang, Dr Chenghong Gu and Professor Furong Li. The contributions the author made in this work are: 1) writing the whole article; 2) writing the Matlab code to validate the proposed method in the case study and modifying the programming code of the proposed methodology; 3) participating in the primal discussion of the proposed methodology.

4.2 The Unit Home Equivalent Distribution Network Pricing Method

This chapter, for the first time, proposes a Unit Home Equivalent (UHE) network pricing structure for domestic customers that will move away from the current energy-based pricing to a new position where both energy and capacity components will be factored to reflect the network cost. Inspired by the transport economics, a novel UHE value is proposed to measure the additional number of the same customer can be connected to the network without triggering reinforcement relative to a base customer (a unit constant load customer). Then the network investment cost for every smart metering customer is allocated based on its UHE value.

The novelty of the proposed method lies in two aspects:

i) The proposed method provides a forward-looking signal. Instead of only considering the historical peak point, the proposed method evaluates the likelihood of future peak created by different customers at different time points.

ii) The proposed method provides a behavioural incentive signal, enabling existing customers to reduce their network cost by modifying their usage behaviour. Without changing geographical locations, customers will be rewarded by reducing their likelihood of creating new peaks according to the network's headroom profile.

4.2.1 The Unit Home Equivalent

Inspired by the Passenger Car Equivalent (PCE) value used in transport economics to allocate motorways cost, we propose a new DUoS charging method which allocates network investment to smart metering customers based on its equivalent value to a base home.

For the same road capacity, the utilisation (maximum number of vehicles per unit time that can flow past the point) will be reduced by the introduction of longer and slower vehicles. In transport economics, such reduction is evaluated by PCE values, which are measured relative to a small passenger car as the base vehicle. For example, if a road can flow 100 base car or 50 trucks, then the truck will have a PCE value of 2.

Likewise, a network will have different capacities for different customer classes. A UHE value is introduced here to describe the maximum number of a type of customer that can be accommodated to an existing network before triggering the network reinforcement. A base home is defined as a control group to represents a home with a constant unity demand (1kW). The base capacity of the network, N_{base}, is evaluated as the number of additional base homes that can be connected to the network. Similarly, the capacity of the network for customer$i$, N_i, is defined as the number of customer i that can be connected before reaching the capacity of lines and transformers. The UHE value of customer i is then defined as (4-1) shown:

$$UHE_i = \frac{N_{base}}{N_i} \tag{4-1}$$

The word *capacity* here indicates the number of new customers that can be accommodated in the network. For example, Figure 4-2 shows the load profiles of two customers, which have the same energy consumption but different patterns. The blue bars illustrate the load profile of the substation. Customer-I has an evening peak coincident with the substation peak while customer-II has a morning peak and a trough during the substation peak.

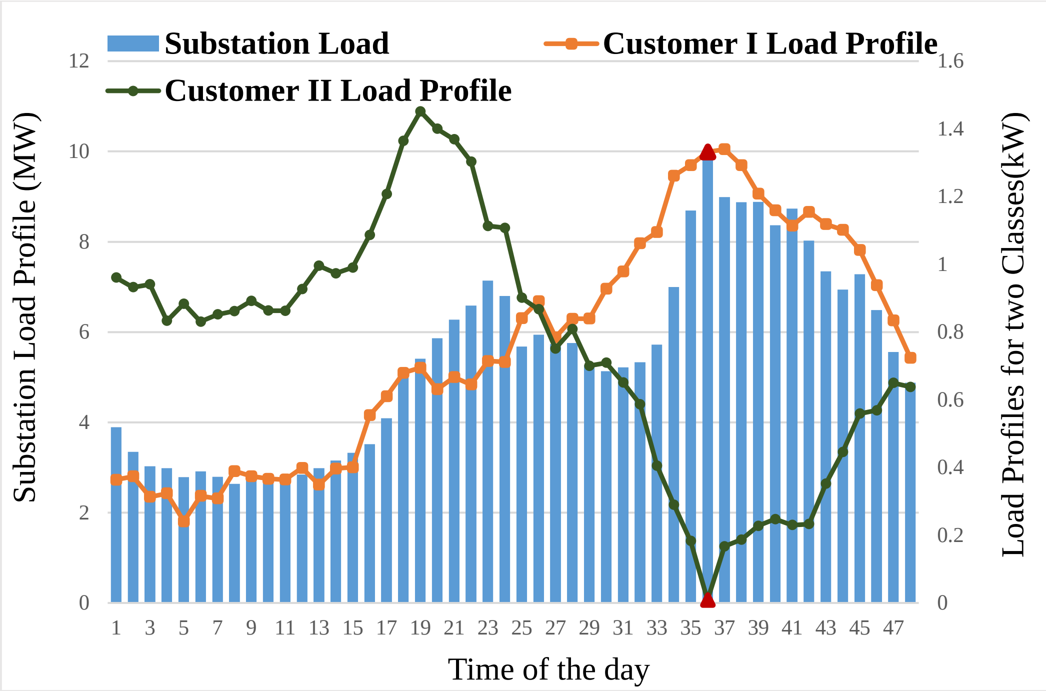


Figure 4- 2: Examples for substation and home load profiles

The two types of customers have different impacts on future network investment. Customer-I type will accelerate the existing peak growth and bring forward network reinforcement; by contrast, customer-II type will flatten the substation load profile and increase the utilisation rate. It is noted that a large number of customer-II might create a new peak on the substation in the morning. This is particularly important for customers with emerging technologies. For example, night-charging EV customers may have little contribution to existing network peak and thus paying no network charges. However, massive EV customers might trigger a new peak at night, which brings forward the reinforcement.

However, existing pricing methods either ignore or overemphasise such different impacts. The flat tariff used in the retail market will charge the two customers equally in that they have the same energy consumption. The network charges currently used for suppliers and large customers only focus on peak contribution. As a result, customer-I will be responsible for all network reinforcement while customer II will get away with any network charges, despite the fact that customer II still utilise the network and may create a new peak at a different time point.

In order to design a pricing method that reflects the incremental likelihood of network reinforcement caused by the usage behaviour of customers, the first step is to identify the Typical Headroom Profile (THP) of a circuit or a network component. THP represents the distribution of the network's spare capacity over time rather than a single annual peak point.

THP could be depleted quickly by introducing more low-utilisation customers. The Figure 4-3 explains this point by relating UHE value to the slope of investment projections. The N_{base} is the number of unit homes can be connected where UHE=1. For customer-I type, only N1 customers can be accommodated in the network before reaching the capacity. The slope of the investment projection line is proportional to the UHE value reflecting relative network peak growth caused by the customer.

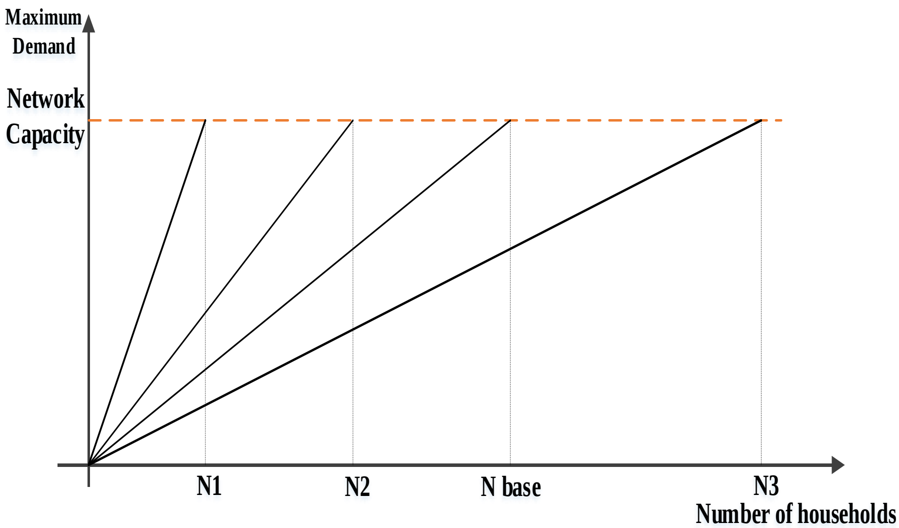


Figure 4- 3: Investment projection: number of additional base homes can be connected to the network without triggering reinforcement

4.2.2 The Unit Home Equivalent in the Network

The proposed UHE network pricing method could be implemented by two main steps. The first step is calculating the UHE values for each household. The second step is the Long-Run Incremental Network Incremental cost allocation based on customers' UHE values. The detailed implementation of the two steps are discussed as follows.

i) Step 1: UHE calculation

The UHE value should reflect the impact of connecting a new customer to the network. By adding an additional customer i at one node, the power flow change on every branch needs to be monitored quantified by the Power Transfer Distribution Factor (PTDF), which is derived from the Jacobian matrix of power flow [90]. Assuming the change of real power transfer between two nodes, m and n, is ΔP_m and the incremental power on line l is $\Delta \mathrm{P}_l$, then the DC-_PTDF can be denoted as (4-2):

$$(DC\_PTDF)_{l-m} = \frac{\Delta P_l}{\Delta P_m} \tag{4-2}$$

Assuming a network with K lines and J nodes (substations), the PTDF of each line and node can be expressed as a matrix T with K rows and J columns. Then, with load or generation change P_j, the resulted power flow variation on line k, $\Delta \mathrm{P}_k$ is calculated by (4-3):

$$\Delta P_k = T_{kj} \times P_j \tag{4-3}$$

The substation load profile is defined as $\boldsymbol{S}$. In our case, $\boldsymbol{S}$ is a vector with 48 variables indicating half-hourly record over a day. Likewise, the customer load profile is defined as a vector $\boldsymbol{H}$. If the maximum rated capacities of all lines are $\boldsymbol{C}$ and the base power running on all lines is $\boldsymbol{B}$, the equivalent value of customer i, i.e. the maximum number of new customers i can be connected to substation j is as (4-4) displayed:

$$\boldsymbol{N}_i^j = \left(\boldsymbol{C} - \boldsymbol{B} - [\boldsymbol{S}_{\lambda_i}]^{K\times 1}\right) / \left(\boldsymbol{T}^j [\boldsymbol{H}_{\lambda_i}]^{K\times 1}\right) \tag{4-4}$$

Here, $\lambda \in [1,48]$is the index of the half-hour period when the maximum value in combined demand of the substation j and the household i, which is defined in (4-5):

$$\begin{gathered} \lambda_i = \underset{0\le\lambda_i\le 48}{argmax}\ (\boldsymbol{S}^{\boldsymbol{j}}_{\lambda_i} + \boldsymbol{H}^{\boldsymbol{i}}_{\lambda_i}), \\ s.t. \left[\boldsymbol{S}^{\boldsymbol{j}}_{\lambda_i}\right]^{K\times 1} + \boldsymbol{T}^j \left[\boldsymbol{H}^{\boldsymbol{i}}_{\lambda_i}\right]^{K\times 1} \le \boldsymbol{C} - \boldsymbol{B} \end{gathered} \tag{4-5}$$

The unit home equivalent value $\boldsymbol{N}^j_{base}$ can be calculated similarly. The only difference is the load profile for the base home is set as a unit load profile, $\mathbb{1}$, which has equal value 1kW for all 48 points.

Therefore, a set of k UHE values for a given household i when connected at substation j as in (4-6).

$$UHE_i^{j,k} = \frac{N_{base}^{j,k}}{N_i^{j,k}} = \frac{\left(\boldsymbol{C_k} - \boldsymbol{B_k} - \boldsymbol{S}_{\lambda_{base}}^{j}\right)/\left(\boldsymbol{T}_k^j\, \mathbb{I}_{\lambda_{base}}\right)}{\left(\boldsymbol{C_k} - \boldsymbol{B_k} - \boldsymbol{S}_{\lambda_i}^{j}\right)/\left(\boldsymbol{T}_k^j \boldsymbol{H}_{\lambda_i}\right)} \tag{4-6}$$

ii) Step 2: Long-run incremental cost calculation

This paper implements the proposed UHE network pricing method with the Long-run Incremental Cost Pricing (LRIC), which is also adopted by the Ofgem as the one of the official charging methodologies in the UK [91].

In LRIC, it firstly derives the network cost of a component l, such as a line, for supporting the existing customers. With the capacity C_l, the load growth rate r and the present power flow of component l, D_l, the number of years, y_l, which indicates how far into the future the investment will be made can be determined from:

$$C_l = D_l \times (1+r)^{y_l} \tag{4-7}$$

Rearranging (4-7) can calculate the value of y_l as shown in (4-8):

$$y_l = \frac{\log C_l - \log D_l}{\log(1+\mathrm{r})} \tag{4-8}$$

Based on the $y_l$value and a discount rate d, the present value for component l can be discount back from its future investment through (4-9):

$$PV_l = \frac{Asset_l}{(1+d)^{y_l}} \tag{4-9}$$

where $Asset_l$is the modern equivalent asset cost of component l.

Then, if the power flow changing along the component l is ΔP_l as the result of the connection of customer i at substation j, this will accelerate the future investment from y_l to y_{lnew}.

$$y_{lnew} = \frac{\log C_l - \log(D_l + \Delta P_l)}{\log(1 + \mathrm{r})} \tag{4-10}$$

In the proposed UHE network pricing method, the UHE value evaluates the impact of the household load upon a line. Customers may have different contributions to the incremental load on lines in the network. The ΔP_l^icaused by injection of a home i on line l is consistent with the UHE's definition as in (4-11):

$$\Delta P_l^i = \frac{UHE_i^l}{\boldsymbol{H}_{\max\_i}} \times \boldsymbol{T_l} \tag{4-11}$$

The $\boldsymbol{H}_{\max\ \_i}$ indicates the maximum demand for household i. $\frac{UHE_i^l}{\boldsymbol{H}_{\max\_i}}$ is a normalisation process for UHE value. Therefore, due to the utilisation of household i, the network investment has been accelerated from y_l to y_{lnew}^i. The value of y_{lnew}^i can be calculated based on (4-12)

$$y_{lnew}^i = \frac{\log C_l - \log(D_l + \frac{UHE_i^l}{\boldsymbol{H}_{\max\_i}} \times \boldsymbol{T_l})}{\log(1 + \mathrm{r})} \tag{4-12}$$

The new present value PV_{lnew}^i could be calculated by (4-13).

$$PV_{lnew}^i = \frac{Asset_l}{(1 + d)^{y_{lnew}^i}} \tag{4-13}$$

Finally, the difference between those two present values, PV_{lnew}^i and PV_l, is the network incremental cost of network component l caused by the household i, just as (4-14) shown.

$$IC_l^i = \left(PV_{lnew}^i - PV_l\right) \times \alpha \tag{4-14}$$

Where the α is the annuity factor which is a constant.

4.2.3 Theoretical Improvement of the UHE Pricing Method

In theory, the proposed UHE network pricing improves the existing ratemaking methods in retail market by providing forward-looking signal to new customers and behaviour incentives to existing customers.

i) Forward-looking signal

With the uncertainty from emerging technologies and behaviour changes, network peak time could vary significantly over time and location. Traditional methods only rely on customers' contribution to systems' historical peak. The proposed UHE pricing provides a forward-looking signal by considering the likelihood of new peaks at different time points. This is achieved by introducing the value of λ_i in (4-5), where λ_i represents the half-hour period number of new peaks. The UHE value in (4-6) represents the likelihood of the future peak at the point λ_i.

The UHE values are dynamic, reflecting the uncertainties of peak over time and location. New customers will have different UHE values when connected at different nodes, reflecting the load profile compatibility between the customer and local networks. The UHE value will also be updated over a period when the states of local network changes.

The proposed method also prevents the over-incentives of "net-zero" customers. Under the proposed pricing, a "net-zero" household still needs to pay for the network cost despite its reduced or even negative energy consumption during system peak time. However, the price will generally be lower because of reduced likelihood of creating new peaks.

ii) Behavioural Incentives

For existing customers who have limited mobility, the UHE pricing provide appropriate behavioural incentives to guide customers modifying their load profiles to maximise the utilisation of existing networks. Three scenarios are discussed below to demonstrate how the value of UHE network price influences customers' energy usage behaviours.

- *Scenario 1*: The new peak is coincident with substation's existing peak time. ($\boldsymbol{S}_{max}^{j} = \boldsymbol{S}_{\lambda_i}^{j}$);

The UHE network cost for household i for line l could be evaluated as (4-15) shown:

$$\begin{aligned} IC_l^i \propto \Delta P_l^i \propto \frac{UHE_i^l}{\boldsymbol{H}_{\max\_i}} \\ = \frac{(\boldsymbol{C_l} - \boldsymbol{B_l} - \boldsymbol{S}_{\lambda_{base}}^{j})}{\boldsymbol{H}_{max\_i} \cdot \boldsymbol{T}_l^{j}\, \mathbb{1}_{\lambda_{base}}} \cdot \frac{\boldsymbol{T}_l^{j} \boldsymbol{H}_{\lambda_i}}{(\boldsymbol{C_l} - \boldsymbol{B_l} - \boldsymbol{S}_{\lambda_i}^{j})} \end{aligned} \tag{4-15}$$

where due to $\mathbb{1}_{\lambda_{base}}$ is a unit value, and $\boldsymbol{S}_{\lambda_{base}}^{j} = \boldsymbol{S}_{max}^{j} = \boldsymbol{S}_{\lambda_i}^{j}$, the (4-15) is simplified as (4-16):

$$IC_l^i \propto \frac{\boldsymbol{H}_{\lambda_i}}{\boldsymbol{H}_{max\_i}} \cdot 1 \tag{4-16}$$

From (4-16), it is evident that the UHE pricing value will be smaller for customer i if this customer moves more power consumption from the ${\lambda_i}^{th}$half-hour to the other half hours. The worst condition is that the customer does not attempt to reduce any consumption at ${\lambda_i}^{th}$half-hour, which leads $\boldsymbol{H}_{\lambda_i} = \boldsymbol{H}_{max\_i}$. It shows customers can always benefit from UHE pricing by shifting their peak demand even the network peak remains unchanged.

- *Scenario 2:* The new peak shifts away from substation's existing peak time, but is coincident with the peak demand of customer i. ($\mathbf{S}_{\mathrm{max}}^{\mathrm{j}} \neq \mathbf{S}_{\lambda_\mathrm{i}}^{\mathrm{j}}, \mathbf{H}_{\lambda_\mathrm{i}} = \mathbf{H}_{\mathrm{max\_i}}$);

It is the scenario that customers create a new peak at a different time point. Firstly, we define $\boldsymbol{Headroom}_{\lambda_i}^{j}$and $\boldsymbol{Headroom}_{min}^{j}$in (4-17) and (4-18):

$$\boldsymbol{Headroom}_{\lambda_i}^{j} = (\boldsymbol{C}_l - \boldsymbol{B}_l - \boldsymbol{S}_{\lambda_i}^{j}) \tag{4-17}$$

$$\boldsymbol{Headroom}_{min}^{j} = (\boldsymbol{C}_l - \boldsymbol{B}_l - \ \boldsymbol{S}_{\lambda_{base}}^{j}) \tag{4-18}$$

Then, the (4-15) can be rewritten as (4-19):

$$IC_l^i \propto \frac{\boldsymbol{Headroom}_{min}^{j}}{\boldsymbol{H}_{max\_i}} \cdot \frac{\boldsymbol{H}_{\lambda_i}}{\boldsymbol{Headroom}_{\lambda_i}^{j}} \tag{4-19}$$

where $\boldsymbol{H}_{\lambda_i} = \boldsymbol{H}_{max\_i}$ and therefore IC_l^i is proportional to the ratio of $\boldsymbol{Headroom}_{min}^{j}/\boldsymbol{Headroom}_{\lambda_i}^{j}$. The UHE price will be lower when new peak occurs at the time when the network has larger spare capacity. It encourages customers to shift demand to the idlest periods of the system.

- *Scenario 3:* The new peak shifts away from substation's existing peak time, and is not coincident with the peak demand of customer i. ($\mathbf{S}_{\mathrm{max}}^{\mathrm{j}} \neq \mathbf{S}_{\lambda_\mathrm{i}}^{\mathrm{j}}, \mathbf{H}_{\lambda_\mathrm{i}} \neq \mathbf{H}_{\mathrm{max\_i}}$);

In this case, the value of $\boldsymbol{Headroom}_{min}^{j}/\boldsymbol{H}_{max\_i}$ will be a constant and $IC_l^i$is related to the ratio of $\boldsymbol{H}_{\lambda_i}/\boldsymbol{Headroom}_{\lambda_i}^{j}$. The UHE price will be lower if the customer contributes less (smaller $\boldsymbol{H}_{\lambda_i}$) to the new peak. It prevents simultaneous responses from all customers that create higher peaks at other periods. It is worth to note that the UHE network price for Scenario 3 is always lower than the price in *Scenario 2.*

4.3 The Case Study for the UHE Method

The proposed UHE pricing methodology is validated on a real distribution network within a Grid Supply Point area in the UK. The structure of the test system is depicted in Figure 4-4 including 20 lines and 7 nodes. The parameters of the network are given in Table 4-1 and the DC-PTDF are calculated accordingly. For the long-run incremental cost, the discount rate d is set to 6.9% as the commonly accepted Minimum Acceptable Rate of Return by the UK's DNOs in setting network charges [85]. The load growth rate r takes the value of 1.0% per annum based on the long-term projected load grow rate in the UK. The annuity factor α is set at 0.0741.

Table 4- 1: Parameters of the Test Network and DC-PTDF

Line	*PTDF at Substation*							*Capacity (MW)*	*Base Power (MW)*	*Security Factor*	*Asset Cost*
	I	*II*	*III*	*IV*	*V*	*VI*	*VII*				
1	0.504	0.000	0.000	0.000	0.000	0.000	0.000	55.73	-12.59	1.99	£1,001,401
2	0.000	0.487	0.148	0.458	0.148	0.000	0.000	78.87	-23.43	2.05	£1,845,674
3	0.000	0.478	0.164	0.505	0.164	0.000	0.000	78.87	-24.18	1.98	£1,482,909
4	0.000	-0.513	0.148	0.458	0.148	0.000	0.000	88.16	8.23	3.77	£324,708
5	-0.496	0.000	0.000	0.000	0.000	0.000	0.000	52.70	12.48	2.01	£1,006,791
6	0.000	0.000	0.000	0.000	0.000	-0.249	-0.246	54.87	6.41	2.04	£1,748,654
7	0.000	0.000	0.000	0.000	0.000	-0.247	-0.244	35.67	6.54	1.93	£2,162,542
8	0.000	0.000	0.000	0.000	0.000	-0.247	-0.244	54.30	6.31	2.07	£446,882
9	0.000	0.000	0.000	0.000	0.000	0.252	0.255	57.27	-1.79	2.05	£597,966
10	0.000	0.000	0.000	0.000	0.000	0.252	0.255	57.27	-1.79	2.05	£1,165,715
11	-0.500	0.000	0.000	0.000	0.000	0.000	0.000	50.75	12.51	2.05	£500,000
12	-0.500	0.000	0.000	0.000	0.000	0.000	0.000	50.75	12.51	2.05	£500,000
13	0.000	-0.503	0.000	0.000	0.000	0.000	0.000	65.00	15.61	2.04	£500,000
14	0.000	-0.497	0.000	0.000	0.000	0.000	0.000	65.00	15.42	2.07	£500,000
15	0.000	-0.018	-0.358	-0.019	-0.358	0.000	0.000	51.25	12.19	1.94	£500,000
16	0.000	-0.017	-0.330	-0.018	-0.330	0.000	0.000	51.25	11.23	2.11	£500,000
17	0.000	0.000	0.000	0.000	0.000	-0.249	-0.246	40.00	6.41	2.00	£500,000
18	0.000	0.000	0.000	0.000	0.000	-0.247	-0.244	40.00	6.31	2.04	£500,000
19	0.000	0.000	0.000	0.000	0.000	-0.252	-0.255	50.00	1.79	2.02	£500,000
20	0.000	0.000	0.000	0.000	0.000	-0.252	-0.255	50.00	1.79	2.03	£500,000

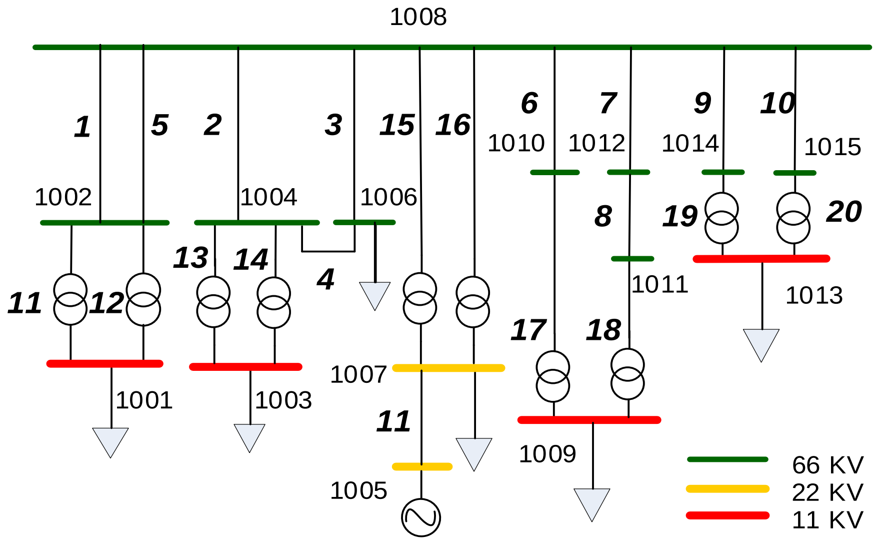


Figure 4- 4: The network structure of the test system

The smart metering data is taken from the Low Carbon London project [92]. 10% of them own the Electric Vehicle (EV). The annually averaged load profile of each household is extracted. The peak demand of each substation is derived from metering data and its load profile is estimated using the network templates and classification tool developed in [93]. Figure 4-5 depicts the estimated load profiles of the seven substations in the test network. Substation-I (Sub-I) and Sub-V are substations with a high proportion of commercial customers. Sub-V is in urban areas and thus has a higher proportion of domestic customers. Sub-II is dominated by Economy 7 customers who have a night peak around 1:00 a.m. due to the lower electricity rate. Sub-III is dominated by industrial customers with a consistent load throughout a day. Sub-IV serves motorway lighting with sharp edges in the morning and evening. Sub-VI and Sub-VII have a mix of domestic and small commercial customers. Sub-VI is located in a suburban area with only one evening peak while Sub-VII feeds a rural area with two peaks around noon and evening.

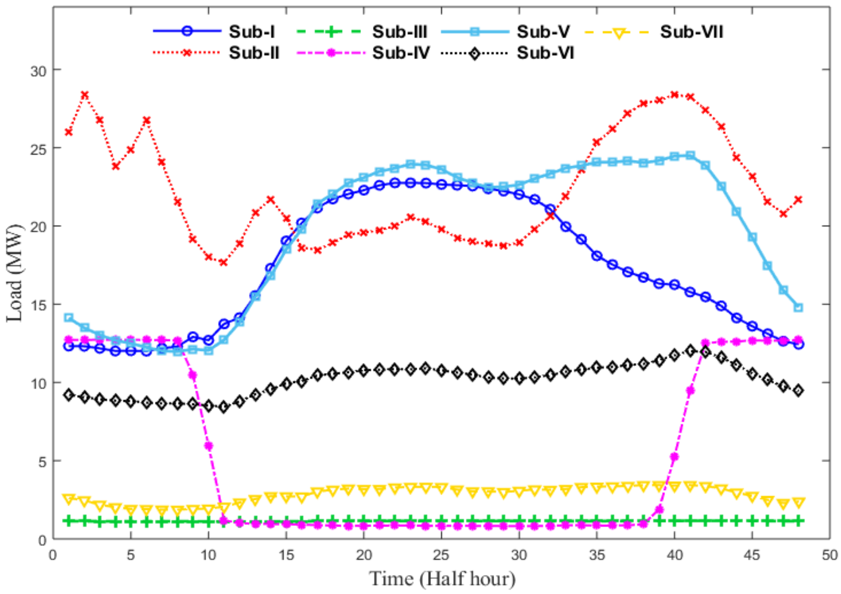


Figure 4- 5: Load profiles of seven substations

A long-term impact (10 years) of the UHE pricing method is accessed in this case study and compared with the result of directly employing the LRIC charging model on the domestic customer. Through repeating the usage data of 1000 customers in the Low Carbon London project, we extend the data size to 10000 households to better observe the customer's injection impact on the network. In each year of the decade, the same 10000 homes will be seen as new customers who require to be connected into the local network.

4.3.1 The LRIC Charging Model

In the LRIC charging model, all of the new customers will be connected to the cheapest substation following the locational signals. For example, as Figure 4-6 shown, Sub-VII is the cheapest location for all customers' injection. Then, all customers share the network cost based on their contribution to the peak period of Sub-VII.

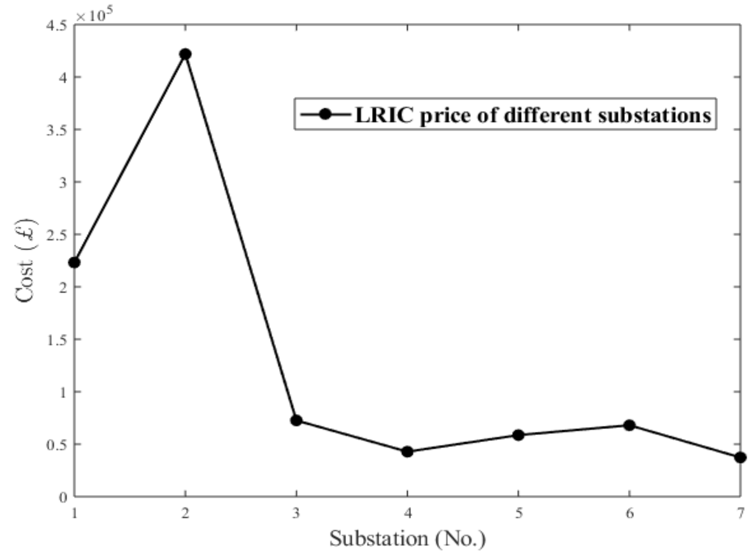


Figure 4- 6: The LRIC cost for all of new customers under every substation

By repeating this process 10 times, the 10-year impact of directly applying LRIC charging model could be accessed.

4.3.2 The Proposed UHE Model

To access the 10-year impact of the proposed UHE pricing method, there are two processes for the new 10000 customers in each year: 1) Choosing substation based on the forward-looking signals; 2) Responding the behavioural incentives after connection.

i) Choosing substation based on the forward-looking signals

In the UHE pricing model, the forward-looking signals consider the likelihood of new peaks at different time points caused by a household's usage pattern. Therefore, the UHE price will guide every customer to the cheapest substation for its own, even it is not the most underutilised substation. For example, Figure 4-7 demonstrates the UHE price of ten typical customers under each substation. For the simplicity, the selected 10 customers are the typical sample for 10 load profile classes of the 10000 new customers, which clustered through K-means. From Figure 4-7, it can be observed that the 10000 new customers are dispersed to different substations.

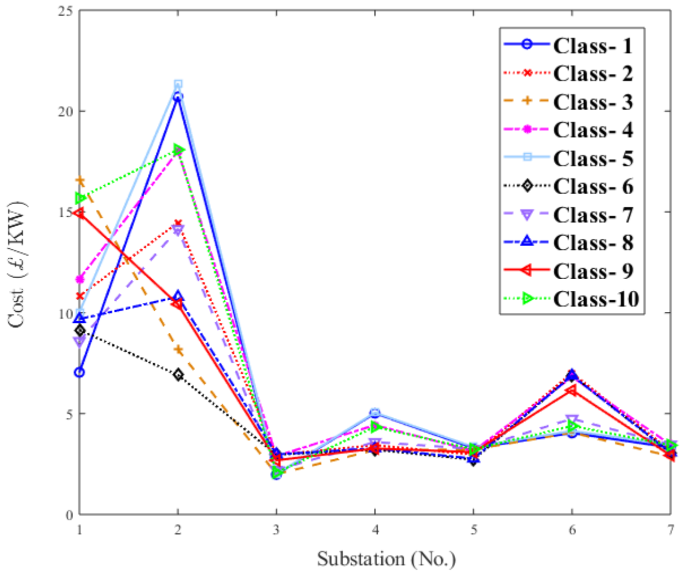


Figure 4- 7: The UHE price for 10 customers under every substation

ii) Responding the behavioural incentives after connection

The second process focus on the customers who have already been connected at the same substation. Although the substation is the cheapest node for the new injected customers, the UHE price among them still could be different due to the behavioural incentive signals.

For example, Figure 4-8 illustrates the normalised UHE price of customers connected to the Sub-V. The black dots represent customers' normalised UHE price. The x-axis indicates the time point of the most likely happened new peak. We extract the three home load profiles for customers whose normalised UHE price value from low to high. The results are displayed in Figure 4-9.

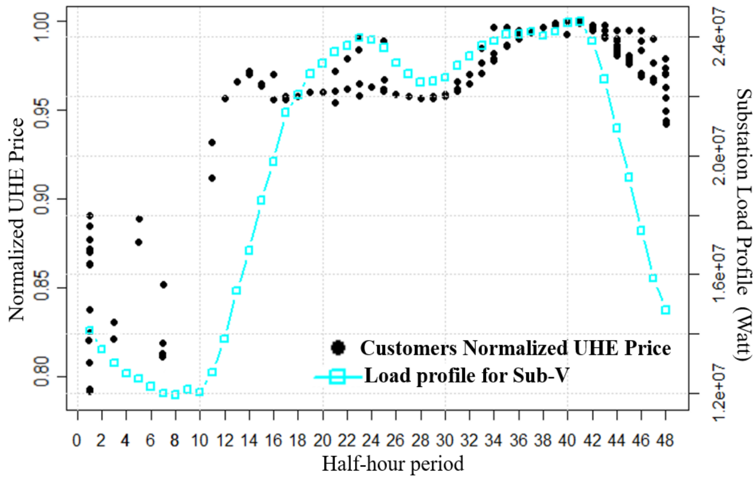


Figure 4- 8: The normalized UHE price of customers connected in Sub-V

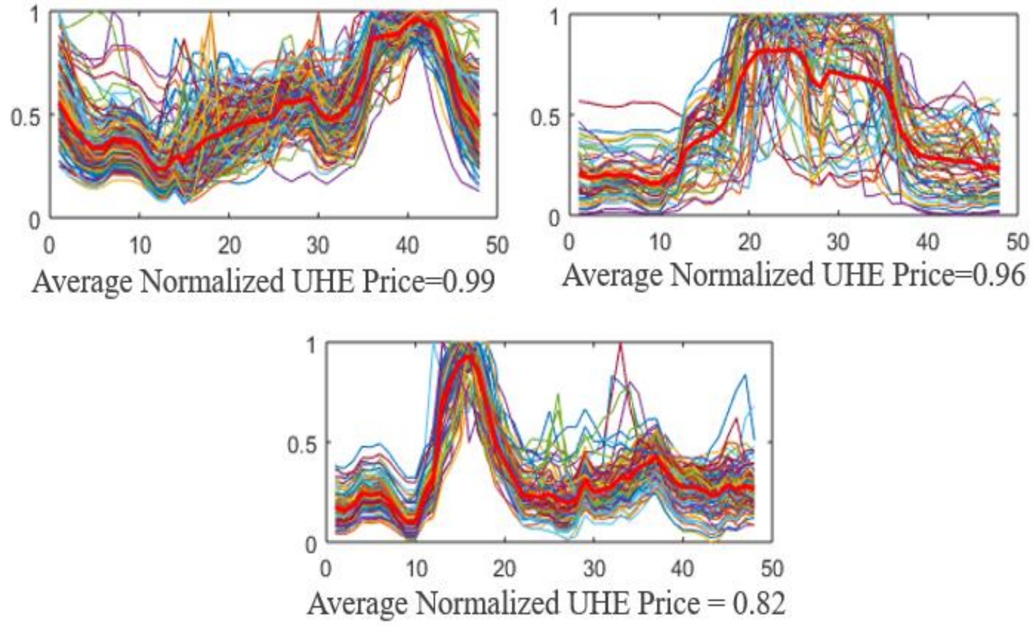


Figure 4- 9: Home load profiles for the different normalized UHE price values

It is evident that the home load curve, which is most complementary to the substation's load profile, would receive the lowest UHE price. This encourages other customers to respond to the incentive signals by modifying their usage pattern to reduce their network bill. In this case study, two scenarios are built up to examine the effectiveness of the behaviour incentive signals:

- Scenario 1:

  It assumes that all of the customers will respond to the behavioural incentive signals after connection. The home load patterns for the existing customers in the substation will change to the home curve who receives the cheapest price.

- Scenario 2:

  This scenario supposes that none of the customers will respond to the behavioural incentive signals. The new injected customers would keep their home load patterns

Finally, the long-term impact of both the UHE pricing model Scenario 1 (with 100% responding) and Scenario 2 (with no responding) can be accessed by repeating those two processes to the new customers each year.

4.4 Results and Discussion of the Case Study

In this section, the long-term (10 years) impact on local network caused by the LRIC charging model and the two scenarios of the UHE pricing model will be analysed and compared.

4.4.1 The Long-Term Impact on the Local Network Cost Change

The investment cost of the 10 years for the LRIC model and two UHE scenarios has been demonstrated in Table 4-2. The accumulated investment cost during the ten years for each substation is displayed. Table 4-3 illustrates the total number of customers injected into each substation during the 10 years. The highest cost among three models is marked in red bold in the tables.

Table 4- 2: The network investment cost in 10 years

INVESTMENT COST IN 10 YEARS (£)	UHE SCENARIO 1: 100% RESONPD	UHE SCENARIO 2: NO RESPOND	LRIC
SUB- I	*£1,000,000*	*£1,000,000*	*£1,000,000*
SUB- II	*£2,170,382*	*£2,170,382*	*£2,170,382*
SUB- III	£145,994.58	£130,396.08	**£161,567.22**
SUB- IV	£26,705.85	£51,132.39	**£118,182.14**
SUB- V	£28,763.49	**£96,886.32**	£76,497.68
SUB- VI	£85,467.30	£80,209.82	**£110,364.60**
SUB- VII	£69,243.22	**£103,056.89**	£55,141.47
Total SUM	**£3,526,556.43**	**£3,632,063.50**	**£3,692,135.11**

Table 4- 3: The number of customers injected into each substation in the 10-years period

NUMBER OF CUSTOMERS INJUCTED INTO	UHE SCENARIO 1: 100% RESONPD	UHE SCENARIO 2: NO RESPOND	LRIC
SUB- I	0	0	0
SUB- II	0	0	0
SUB- III	32670	**35540**	30000
SUB- IV	8690	**10400**	10000
SUB- V	**18750**	13570	10000
SUB- VI	11130	15990	**30000**
SUB- VII	**28760**	24500	20000

Firstly, it can be noticed that both two UHE scenarios achieve a lower investment than the LRIC model. The effectiveness of the behaviour incentive signals is remarkable between the two scenarios. By responding to the incentive signals, the UHE Scenario 1 saves £105,507 comparing to UHE Scenario 2.

Secondly, the merits of the forward-looking signals also can be observed. Using Sub-III as an example, both two UHE scenarios guide more new customers connected under Sub-III than LRIC model. However, even the no-respond scenario model (which means the behavioural incentive signals do not affect) achieves lower investment cost than the LRIC model. It is evident that the forward-looking signals more effectively guide suitable customers to a substation than the pure locational signals. The forward-looking signals can effectively improve the utilisation of the network.

Furthermore, the UHE price also can reflect the locational signals. Such as Sub-I and Sub-II, those two substations are highly-utilised from the beginning. The UHE pricing does not lead any new customers to be connected to those two substations. In the 10th year, the Sub-I triggered the investment of line-11and line-12, meanwhile, Sub-II achieve the capacity of line-2 and line-4 due to the load growth rate. Therefore, the network cost of Sub-I and Sub-II is the same for all of the three models.

4.4.2 The Long-Term Impact on the Local Network Consumption Change

The substations' load profiles in the final year resulted by UHE Scenario 1, Scenario 2 and the LRIC model are illustrated by the Figure 4-10, Figure 4-11and Figure 4-12 respectively. It is evident that the behaviour incentive signals can avoid boosting the original substation peak. More consumption is modified to other periods. Therefore, most of the substations' load profiles in UHE Scenario 1 trend to have a dual peak.

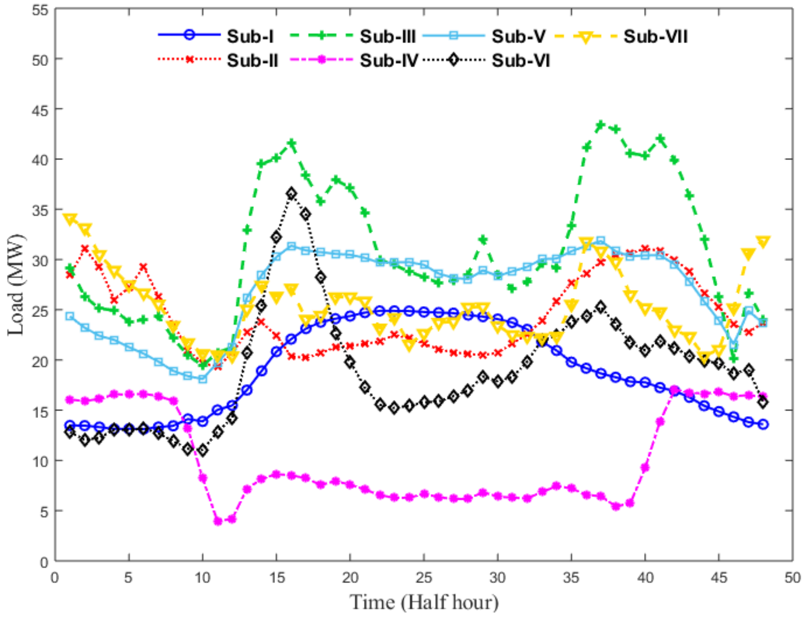


Figure 4- 10: Substations' load profiles for 10th year in UHE Scenario 1

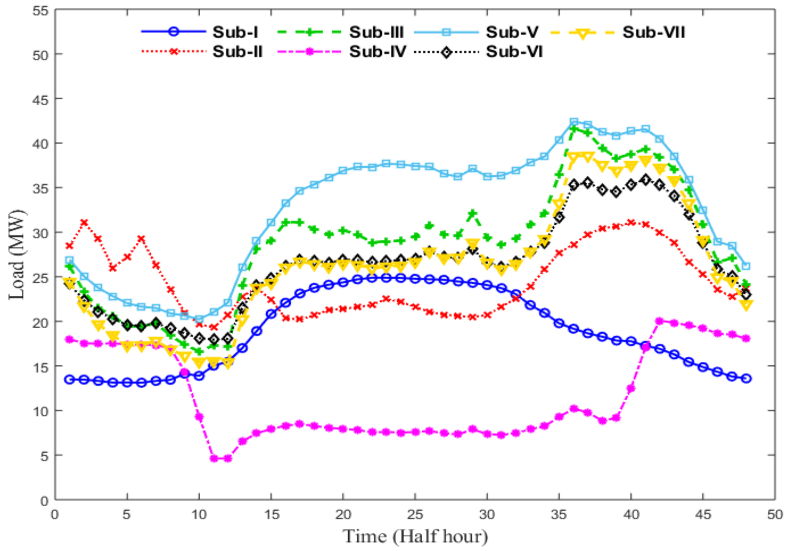


Figure 4- 11: Substations' load profiles for 10th year in UHE Scenario 2

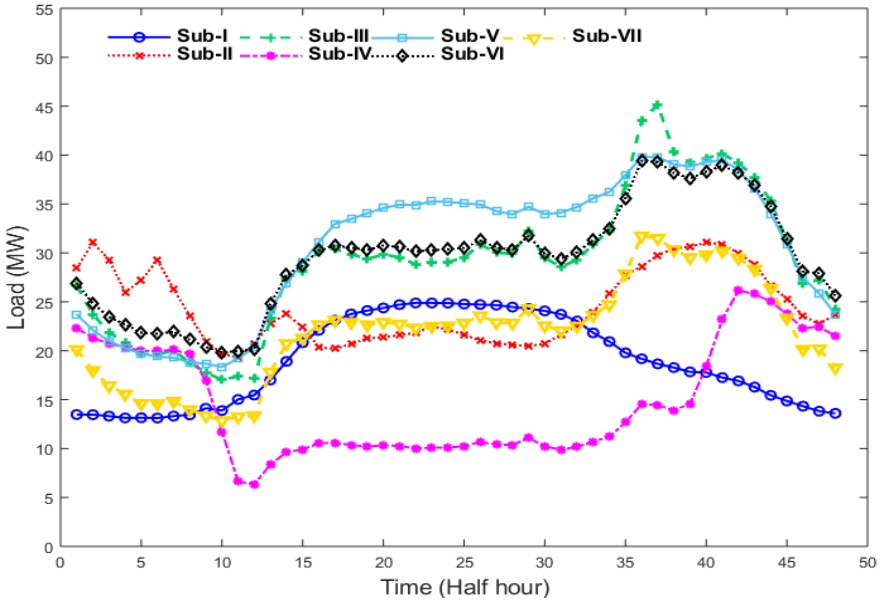


Figure 4- 12: Substations' load profiles for 10th year in LRIC model

In Table 4-3, the consumption difference between the maximum and minimum values for each substation has been demonstrated. The difference value can represent the smooth of the substation load profile. As Table 4-3 demonstrated, the UHE Scenario 1 achieves the smoothest load profile for most of the substations, except Sub-VI. The main reason is displayed in Figure 4-13. In Figure 4-13, the original load profiles for Sub-III and Sub-VI are demonstrated. It is clear that two substations have similar profile-shape. However, the headroom capacity of Sub-III is higher than the Sub-VI. Hence, the Sub-III can provide cheaper UHE price to attract more customers who have complementary home load profiles to smooth its load profile. Until 9th year, Sub-VI gradually gets the chance to smooth its load profile by injected more suitable customers due to the increasing UHE pricing of Sub-III. Therefore, due to the case study only analyse 10 years, it is not long enough to smooth the load profile of Sub-VI.

The UHE Scenario 2 model only relies on the forward-looking signals to disperse new customers to suitable substations. Such as the Sub-III and Sub-IV in Scenario 2 are still better flattened by injected in suitable customers, even they connected more new customers than the LRIC model. However, for Sub-V and Sub-VII, the substation load profiles resulted in Scenario 2 achieve larger gap between peak and bottom values, comparing with the LRIC model. That is due to more customers are assigned to Sub-V and Sub-VII in Scenario 2 than LRIC.

Table 4- 4: The difference between peak load and valley load

LOAD DIFFERENCE (MW)	UHE SCENARIO 1: 100% RESONPD	UHE SCENARIO 2: NO RESPOND	LRIC
SUB- I	11.75	11.75	11.75
SUB- II	11.76	11.76	11.76
SUB- III	24.04	24.98	**28.11**
SUB- IV	13.07	15.45	**19.82**
SUB- V	13.73	**22.18**	21.47
SUB- VI	**25.55**	17.98	19.59
SUB- VII	13.75	**23.17**	18.87

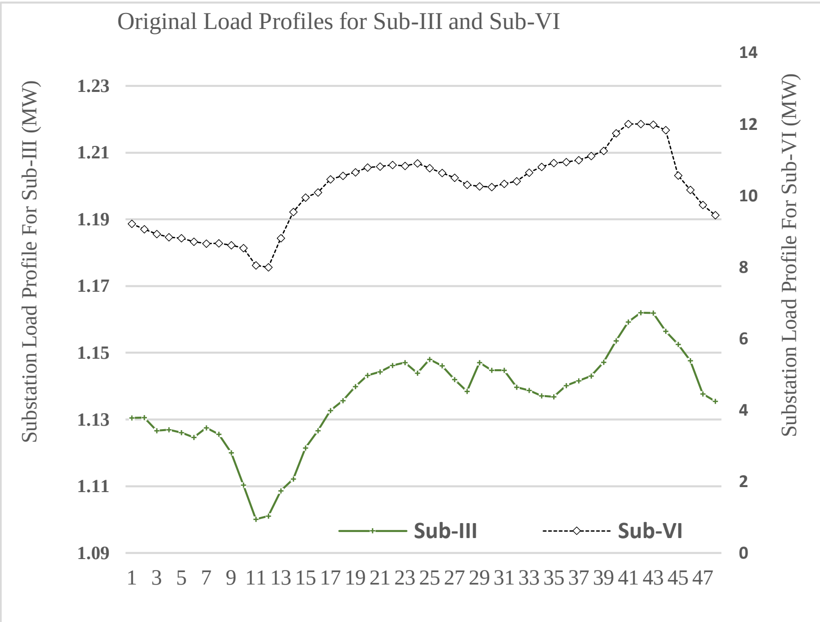


Figure 4- 13: The original load profiles for Sub-III and Sub-VI

By comparing the performance of two scenarios of UHE pricing method and the traditional LRIC pricing model, the advantages of the proposed UHE pricing method can be found. The performance demonstrates that the looking forward signals could accommodate new users to the most appropriate substation to improving the utilisation and deferring the reinforcement. Alongside, the behavioural incentive signals can smooth the load profile of the substation with the response of the customers.

4.5 Identification the Impact of UHE Energy Bill

The previous section has illustrated the UHE pricing method, which can evaluate the network cost for every individual customer. However, by removing the network cost cross-subsidies, there will be an impact in different degree on customers' energy bill. Therefore, in this section, the high-dimensional interaction-aware KLAM algorithm which proposed in chapter 3 has been adopted to identify the social-economic characteristics for the impact of network bill change.

To demonstrate the assessment of the impact, the UHE pricing method is applied on the same smart metering dataset as used in Chapter 3, which is the Irish households' dataset in the Smart Metering Electricity CBTs [94]. Unlike assessing the long-term impact in Section 4.4, this section focuses on the difference of allocation the network cost between evenly or through UHE pricing method. Therefore, all of the 1000 Irish residential households are supposed to be connected under the Sub-6 whose substation load profile is shown in Figure 4-14.

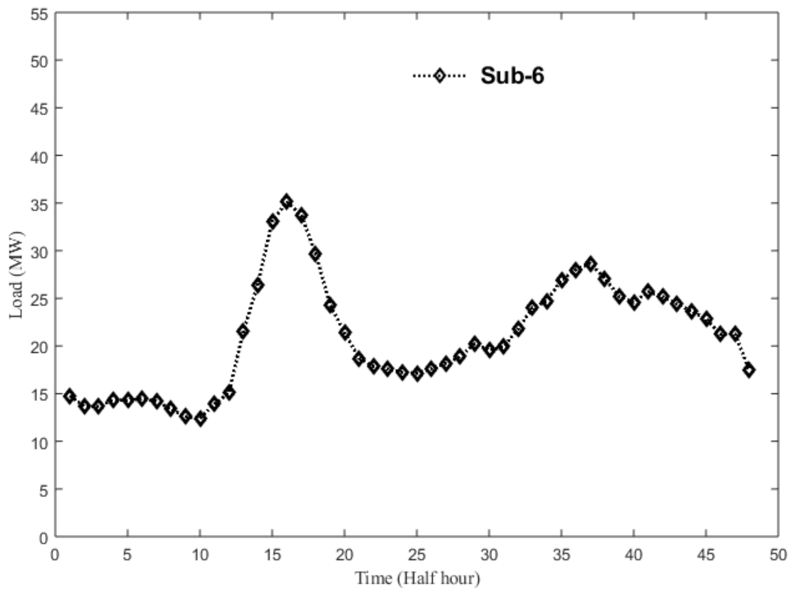


Figure 4- 14: The substation load profile for Sub-VI

4.5.1 The Network Cost Resulted by Using UHE Pricing Method

Every individual residential customer would have a specific UHE price according to its yearly mean load profile. For the simplicity of showing the UHE price for Irish customers, the normalized load profiles for all of 1000 customers have been clustered into 10 groups by the K-means method, which shown in Figure 4-15. The typical load profile for each group is represented by the bold red line. The UHE price for different typical load patterns is shown in Table 4-5. It is evident that the different load pattern could result in different unit UHE price. The cheapest load pattern is marked in red in Table 4-5, which is the most complementary load shape with the Substation load profile.

Then, the cross-subsidies of network cost can be quantified by compared the UHE bill with the traditionally averaged allocated bill, shown in (4-20)

$$\beta_{Net\_c} = \frac{(\delta_{UHEc} - \delta_{ac}) \times 100\%}{\delta_{ac}} \tag{4-20}$$

where the β_{Net_c} represents the degree of network cost cross-subsidies for customer c. The δ_{UHEc} is the network cost calculated by the UHE price and the δ_{ac} is the averaged network bill based on the consumption of the customer c.

Table 4- 5:The UHE price for different load pattern customers

Load pattern class number	*UHE price (£/kW)*
Class 1	6.95
Class 2	5.04
Class 3	4.39
Class 4	5.28
Class 5	5.15
Class 6	5.28
Class 7	4.82
Class 8	**3.96**
Class 9	4.86
Class 10	4.76

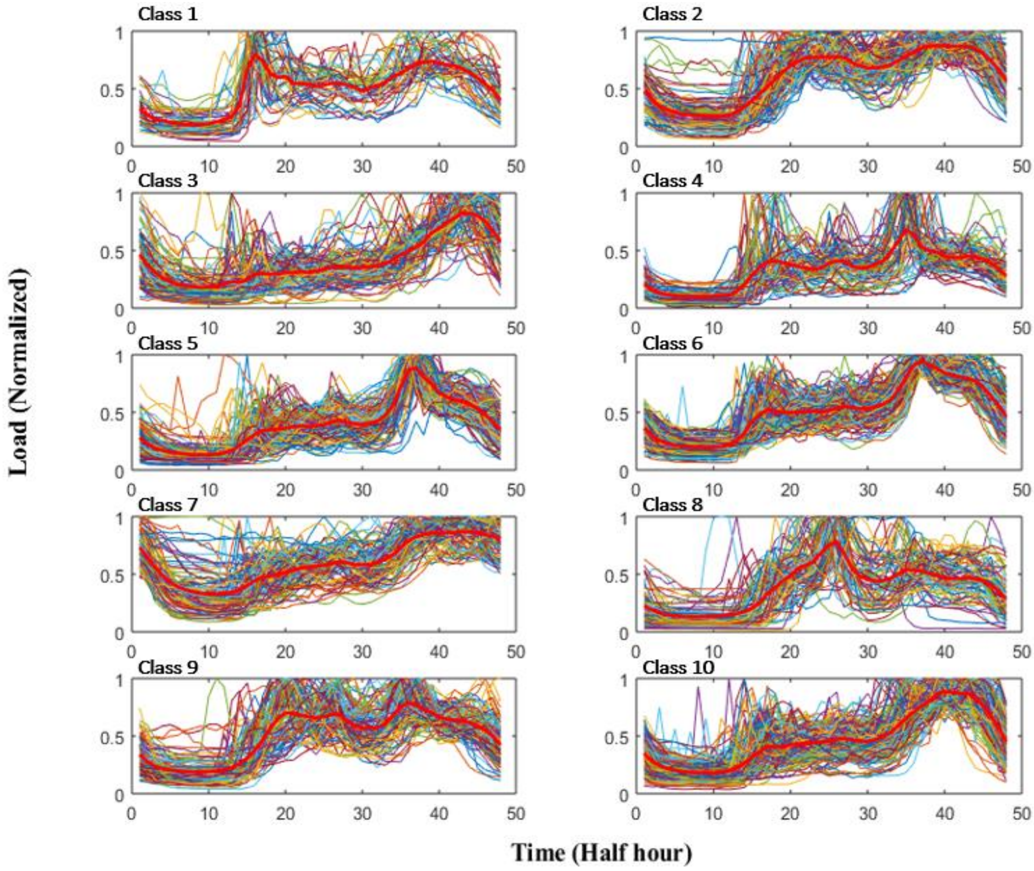


Figure 4- 15: The typical load profile of ten classes

4.5.2 The Significant Socio-Economic Criteria for the Network Bill Variation

After quantifying the degree of the cross-subsidies of the network cost, β_{Net_c}, for every customer, the KLAM algorithm is applied to find the significant interacted socio-economic factors.

There are totally 10 socio-economic factors detected through the one-way ANOVA, which have significant influence on the amount of network cross-subsidies (p-value < 0.05). Then, those ten original factors marked as the selected factors, $f_q^{selected}$, and start to interact with other socio-economic factors. Table 4-6 lists the content of the ten selected factors with the p-value before and after interacting with other socio-economic factors. For every originally selected factor, the p-value has been ameliorated by interacting with other factors. It is evident that the interaction effect among several socio-economic factors has a more significant influence on the amount of network cost cross-subsidies.

Table 4- 6:The P-value of the original significant socio-economic factors

Number	*Content of Question*	*Individually P-value*	*P-value for the interaction term*
Qu 86	The Number of the Stand-alone freezer you own	1.01$\times 10^{-5}$	**2.99**$\times \mathbf{10^{-6}}$
Qu 39	How many bedrooms in your home	8.01$\times 10^{-4}$	**4.74**$\times \mathbf{10^{-4}}$
Qu 3	The employment status of the CIE*	9.26$\times 10^{-4}$	**8.46**$\times \mathbf{10^{-4}}$
Qu 116	The proportion of energy-saving light bulbs in your home	1.21$\times 10^{-3}$	**6.01**$\times \mathbf{10^{-4}}$
Qu 97	Number of the Lap-top you own	1.45$\times 10^{-3}$	**1.29**$\times \mathbf{10^{-3}}$
Qu 35	Do you own or rent your home	5.32$\times 10^{-3}$	**4.45**$\times \mathbf{10^{-3}}$
Qu 128	Will you decide to choose an appliance with a higher energy rating in the future	8.05$\times 10^{-3}$	**5.56**$\times \mathbf{10^{-3}}$
Qu 138	The percentage of electricity being generated from renewable sources	9.61$\times 10^{-3}$	**8.70**$\times \mathbf{10^{-3}}$
Qu 37	The approximate floor area of your home	1.22$\times 10^{-2}$	**1.10**$\times \mathbf{10^{-2}}$
Qu 90	Number of TV greater than 21 inch you won	1.36$\times 10^{-2}$	**1.20**$\times \mathbf{10^{-2}}$

By setting the beam width W as 3, only the top 3 factors' combination with the smallest p-value for each originally selected factor would be output by the Stage I of the KLAM beam searching algorithm. To ensure the number of customers can be involved in each treatment (level's combination of interacted factors), all of the ten original factors are interacted with up to 4 other socio-economic questions.

The abandoned factors are recycled in Stage II through the KL-divergence and GMM. There are two pruned-off socio-economic factors whose two specific levels have high KL-divergence value and significantly affect the cross-subsidies value. Table 4-7 displays those two recycled factors with their significant test results before and after the interaction. The GMM aims to reveal the significant treatments whose significance may be weakened by the overlapped other treatments. However, for this case study, the GMM does not find any pair of socio-economic factors whose treatment effect can pass the significant test.

Table 4- 7: The P-value of the socio-economic factors detected in *Stage II*

Number	*Content of Question*	*Level A*	*Level B*	*Individually P-value*	*P-value for the interaction term*
Qu 125	Do you decide to make minor changes	1 (Strongly agree)	4 (Disagree)	2.68$\times 10^{-2}$	**2.41**$\times \mathbf{10^{-2}}$

	to the way you use electricity				
Qu 36	How old is your house	1 (< 15 years)	3 (> 40 years)	3.28×10^{-2}	$\mathbf{2.36\times 10^{-2}}$

In this section, due to the limitation of space, the most significant interacted-factor combinations are demonstrated in Table 4-8. The detailed information of every involved socio-economic factor can be found.

Additionally, the confidence level is set as 95% for the bill changing influenced by a specific treatment of an interacted-factor combination. In other words, there is 95% confidence that the future network cost will increase (or decrease) as long as the household with the socio-economic makeups conforms to the treatment. Therefore, in Table 4-8, the socio-economic characteristics for the bill increasing customers for each interacted-factor combination are summarized by comparing the treatment chosen by the bill-increasing group and bill-decreasing customer group.

Table 4- 8: The significant socio-economic factor combinations and their impact on cross-subsidies

No.1		**Qu 86**	**Qu 133**	**Qu 44**	**Qu 17**	**Qu 69**
		Number of the Stand-alone freezer you own	*The approximate income of all adults in the household*	*Do you heat your home by solid fuel*	*Have you already done a lot to reduce your electricity usage*	*Do you own watching machine in your home*
	Bill Increasing Group	**Own less freezer in home**	**Earn less income**	**More households heat home by solid fuel**	**Fewer households have reduced their usage**	**All households choose yes**
No.2		**Qu 39**	**Qu 62**	**Qu 11**	**Qu 85**	**Qu 69**
		How many bedrooms are there in your home	*Your home is hard to keep warm due to it is not well insulated*	*How many adults are typically in the house during the day*	*Number of the plug-in electric heater you own*	*Do you own watching machine in your home*
	Bill Increasing Group	**Have less bedrooms in home**	**All households choose yes**	**All households have one adult in home during the day**	**Own more heaters in home**	**All households choose yes**
No.3		**Qu 3**	**Qu 97**	**Qu 14**	**Qu 15**	**Qu 69**
		The employment status of the chief income earner (CIE)	*Number of the Lap-top you own*	*Do you interested in changing the way you use electricity if it can reduce the bill*	*Do you interested in changing the way you use electricity if it helps the environment*	*Do you own watching machine in your home*
	Bill Increasing Group	**More of them are retired or unemployed**	**Own less Lap-top in home**	**More households interested in**	**Fewer households interested in**	**All households choose yes**
No.4		**Qu 116**	**Qu 25**	**Qu 14**	**Qu 19**	**Qu 69**
		The proportion of energy-saving light bulbs in your home	*Did you take any energy reduction activities to reduce your bills last year*	*Do you interested in changing the way you use electricity if it can reduce the bill*	*Would you like to do more to reduce electricity usage*	*Do you own watching machine in your home*
	Bill Increasing Group	**Have less in home**	**Fewer households take activities**	**More households interested in**	**More households like to**	**All households choose yes**

No.5		**Qu 97**	**Qu 12**	**Qu 13**	**Qu 34**	**Qu 69**
		Number of the Lap-top you own	*How many children (<15) live in your home*	*How many children (<15) are in the house during the day*	*The type of your house*	*Do you own watching machine in your home*
	Bill Increasing Group	**Own more Lap-top in home**	**Have more children living with**	**Stay longer in home during the day**	**More households live in detached/ semi-detached house**	**All households choose yes**
No.6		**Qu 35**	**Qu 33**	**Qu 31**	**Qu 59**	**Qu 128**
		Do you own or rent your home	*How much do you believe you could reduce your usage*	*Do you want to be told how much electricity you can use*	*Is your home kept adequately warm*	*Will you decide to choose higher energy rating appliances*
	Bill Increasing Group	**More households rent house from a local authority**	**Less reduction than bill decreasing group**	**Fewer households want to know**	**More households choose yes**	**More households choose yes**
No.7		**Qu 128**	**Qu 5**	**Qu 62**	**Qu 132**	**Qu 58**
		Will you decide to choose higher energy rating appliances	*Do you have internet access in your home*	*Your home is hard to keep warm due to it is not well insulated*	*The level of education of the chief income earner (CIE)*	*Describes how you cook*
	Bill Increasing Group	**All households choose yes**	**More households have**	**More households choose no**	**Have higher education level**	**More households use electricity instead of gas**
No.8		**Qu 138**	**Qu 6**	**Qu 23**	**Qu 124**	**Qu 127**
		The percentage of electricity being generated from renewable sources	*Do you have broadband in your home*	*Would you like to do more to reduce electricity usage*	*Do you want to help the Ireland environment by your participation in a trial*	*Would you like to know the electricity amount of the appliances*
	Bill Increasing Group	**Less percentage**	**More households have**	**Fewer households will**	**All households choose yes**	**All households choose yes**

No.9		**Qu 37**	**Qu 44**	**Qu 124**	**Qu 94**	**Qu 121**
		The approximate floor area of your home	*Do you feat your home by solid fuel*	*Do you want to help the Ireland environment by your participation in a trial*	*How many TV (less than 21 inch) do you own*	*Do you want to learn how to reduce the energy usage by your participation in a trial*
	Bill Increasing Group	**Smaller house**	**All households choose no**	**More households want to help**	**Have more TV in home**	**All households choose yes**
No.10		**Qu 90**	**Qu 10**	**Qu 100**	**Qu 89**	**Qu 54**
		Number of TV greater than 21 inch you won	*How many adults live in your home*	*How often would you use the Tumble Dryer*	*Do you have TV (less than 21 inch)*	*Do you use renewable energy to heat water*
	Bill Increasing Group	**More TV owned**	**More adults live together**	**Less frequently and shorter**	**More households have more than one TV**	**Less households use renewable energy**

No.11*		**Qu 125**	**Qu 10**	**Qu 13**	**Qu 75**
		Would you like to make minor changes to the way you use electricity	*How many adults live in your home*	*How many children (<15) are typically in the house during the day*	*Do you have the plug-in electric heater*
	Bill Increasing Group	**Fewer households will**	**More adults live together**	**Stay longer in home during the day**	**More households have**
No.12*		**Qu 36**	**Qu 56**	**Qu 7**	**Qu 83**
		How old is your house	*Do you have timers to control when your heater comes on and off*	*Do you use the internet regularly*	*How many electric shower (pumped from hot tank) do you own*
	Bill Increasing Group	**Newer house**	**Fewer households have**	**Less frequently and shorter**	**Own less in home**

** represents the original selected factor is detected by the KL-Divergence results*

From Table 4-8, it is also can be observed that the impact of a single socio-economic factor on customers' network bill change could be different by interacting with various factors. For example, Qu.44 and Qu.97, which are highlighted in the same colour in Table 4-8, demonstrate the contrary characteristics for the bill increasing customer group. To summaries the socio-economic characteristics, the socio-economic questions are categorized into four groups which regard to 1) the awareness of energy-saving (marked in purple); 2) personal information (marked in yellow); 3) dwelling information (marked in green) and 4) appliances information (marked in blue). Based on the results found by the KLAM algorithm, several conclusions of the socio-economic characteristics of the high network cost customers can be resulted:

- The awareness of energy-saving is weak for the high network cost customers. They are willing to do more to reduce their bill but not for the environment. Less of them do the energy reduction activities in reality or use renewable energy.
- People in the high network cost group live with a large family. More people are retired or unemployed with lower income.
- The houses for the high network cost customer group are relatively newer and smaller than the lower network cost group. Most high cost customers rent a house and always feeling not warm enough due to the poor insulation of the house.
- The television, electric heater and the washing machine are the significant electricity appliances for the high network cost customer group, which applies the positive effect on the network bill growth. On the opposite side, the Stand-alone freezer, electric shower and tumble dryer have the negative impact on the network cost.

Comparing to the socio-economic characteristics for the high wholesale market cost customers, the network cost for individuals is more depended on the factor related to the energy-saving awareness and the lifestyle of the customers instead of the appliances owned in the home.

4.6 Chapter Summary

The aim of this chapter is to assess the impact of removing the cross-subsidies in the network cost for customers in different socio-economic status. The contributions for this research can be introduced from two aspects.

Firstly, this research proposes a novel Unit Home Equivalent (UHE) distribution network pricing method for individual customers. By removing the cross-subsidies in the network cost, the main contribution of this proposed pricing is that the network signal can be sent to the end-users. The cost-reflective price signals can guide the customers to modify their usage behaviour and achieve higher utilisation of the network.

The proposed UHE pricing moves away from the current energy-based pricing to a new position where both energy and capacity components will be factored to reflect the long-run network cost. The proposed method has two fundamental breakthroughs:

i) Forward-looking signal: instead of only considering customers' contribution to historical peaks, the proposed method evaluates the likelihood of future peaks created by different customers at different time points.

ii) Behavioural incentives: the proposed method encourages not only new customers to under-utilised locations but also existing customers to change energy usage behaviours according to the network's headroom profile.

The result shows the proposed pricing will encourage existing customers to adjust energy usage behaviours to defer network reinforcement and guide the connection of new customer to the location with maximum usage of spare capacity.

Secondly, the impact of the network cost variation caused by applying the UHE pricing on customers in different socio-economic status has been assessed in this research. The socio-economic characteristics for the higher network cost customers are valuable for the suppliers and policymakers to design further interventions and tailored services.

The socio-economic characteristics of the high network cost customers more are more relating to energy-saving awareness, which are summarized as:

- The awareness of energy-saving is weak for the high network cost customers. They are willing to do more to reduce their bill but not for the environment. Less of them do the energy reduction activities in reality or use renewable energy.
- People in the high network cost group live with a large family. More people are retired or unemployed with lower income.
- The houses for the high network cost customer group are relatively newer and smaller than the lower network cost group. Most high cost customers rent a house and always feeling not warm enough due to the poor insulation of the house.
- The television, electric heater and the washing machine are the significant electricity appliances for the high network cost customer group, which applies the positive effect on the network bill growth. On the opposite side, the Stand-alone freezer, electric shower and tumble dryer have the negative impact on the network cost.

Chapter 5

The Impact of Socio-Economic Features on Cost-Reflective Customer Classification

THIS chapter investigates the application of collaborating socio-economic data with the load data to establish a cost-reflective customer classification framework for customer with different available input data.

5.1 Introduction

As mentioned in previous chapters, after the privatisation of the energy market in the UK, a large influx of new energy suppliers emerges in the energy retail market which significantly increases the competition in the market. The suppliers purchase the energy from the wholesale market with a half-hourly changed price and sell them back to their customers. With accurate customers' load forecasting, the suppliers set a flat unit price for residential customers for simplicity. However, due to the boosted renewable energy sources in households, the large uncertainty of the renewable energy output makes the customers' load profiles more volatile. Under the pressure of surviving in the ever-competitive electricity retail market, suppliers need to provide a cost-reflective electricity bill to the individual customer instead of roughly averaged allocate the cost in traditional to gain a competitive edge.

The accurate estimation of customers' supply cost can assist suppliers in attracting and profiting those low-cost customers with a lower price, meanwhile, it also allows suppliers to launch timely interventions to help the high-cost customers to reduce their supply costs, such as the DSR and more tailored pricing schemes. Thus, a cost-reflective customer classification becomes a critical method for suppliers to manage millions of customers into a manageable number of supply-cost groups.

Comparing with the proposed classification methodology, the load profile-based approaches face two deficiencies when fulfilling this object:

1) The historical load data for the customers is unavailable.

In the UK, the historical smart metering data for a new switch-in customer is inaccessible for the new supplier. Besides, there are some households failing to provide their usage data due to the privacy issues or the absence of smart meters. Therefore, adopting load profile-based approaches to segment customers' cost level could be impractical.

2) The influential features are more interpretable for further analysis

At the perspective of suppliers, gaining the influential load features is more intuitive than the load profiles for further analysis about the intervention designing (e.g. the demand side response).

For the two reasons, a cost-reflective customer classification framework has been proposed in this chapter by collaborating customers' usage data with socio-economic factors. Three scenarios are built based on the available data for customers to estimate the energy cost level for new switched-in customers. The novelties for the novel classification framework are:

- It is applicable for the different input data type, e.g. only load data input, only socio-economic information input or both of the load and socio-economic data input.

- It can result in interpretable features which have a significant impact on customers' energy cost. This is convenient for further intervention designing

The structure of this chapter is written in an alternative-based format. The content of the proposed cost-reflective classification framework is prepared to submit to the Energy Policy. The author is the first author of this work and collaborated with Dr Minghao Xu.

This rest of this chapter is organised as follows. Section 5.2 introduces the background of this research. Section 5.3 the proposed cost-reflective classification framework. The implementation of the case study has been demonstrated in Section 5.4. Then, Section 5.5 presents the analysis of the results of the case study. The conclusions are drawn in Section 5.6.

5.2 Background of the Research

With a large influx of new energy suppliers in the UK, the competition in the electricity retail market has significantly increased. The number of electricity suppliers has increased from 6 to 66 [95] after the privatisation of the energy market. The market share of the "big six" has dropped from 98% in 2012 to 82% up to the second quarter of 2017 [96]. To gain a competitive edge in this market, suppliers are developing tailored tariffs and services for different customers. A forefront challenge is how to accurately estimate the supply cost of individual customers. Customer classification has been an effective method to divide millions of individuals into a manageable number of groups, where customers share some similar characteristics within each of the group.

For traditional electricity customer classification methods, the common characteristic is defined as the load shape or load profile. The overall methodology consists of two steps: 1) An unsupervised learning to cluster customers with similar load profile into the same group; 2) A supervised learning to classify the new customers into the clusters based on his load profile.

Although this load-profile based method has been widely used by previous works [97-99] to design different tariff bands associated with different load profile classes to promote electricity business, it essentially reflects the characteristic of load profiles rather than supply cost. In fact, these two characteristics can hardly conform to each other for individual customers due to the following two reasons:

1) **The average load profile cannot represent the daily energy usage of individual customers:**

   A load profile is the average energy usage pattern of a group of customers over a time period like a season or a year [98]. However, the individual supply cost is dependent on his hourly energy usage because the price of energy market varies in real time (e.g. half-hourly in the UK). A case study has been reported in [100][6] that even two customers with same load profile actually have very different daily load profiles.

2) **The variation of individual load profile does not always conform with the energy price**

   The variation of the wholesale price only reflects the demand and supply equilibrium at the aggregated level. For individual customers, their daily load profiles are volatile and can be inconsistent with the wholesale price. It is therefore inaccurate to use load-profile based customer classification to represent the energy supply cost.

This chapter proposes a framework of cost-reflective customer classification for suppliers to identify the actual supply cost of their customers. It enables direct classification of customers into different cost levels using cost-related features. For wider applicability, the framework consists of three models to cope with different scenarios of available data: Scenario 1) customers who only provide smart metering data; Scenario 2) customers who only provide the socio-economic information and Scenario 3) customers providing both two types of data. Each model has four stages of supply cost quantification, feature design, feature selection and classification. Feature selection techniques are adopted over feature extraction owing to the simplicity of the original features which are interpretable for suppliers to identify targeted customers to provide the tailored service.

5.3 The Cost-Reflective Customer Classification Framework

The structure of the proposed cost-reflective customer classification framework is presented in Figure 5-1. It consists of four stages:

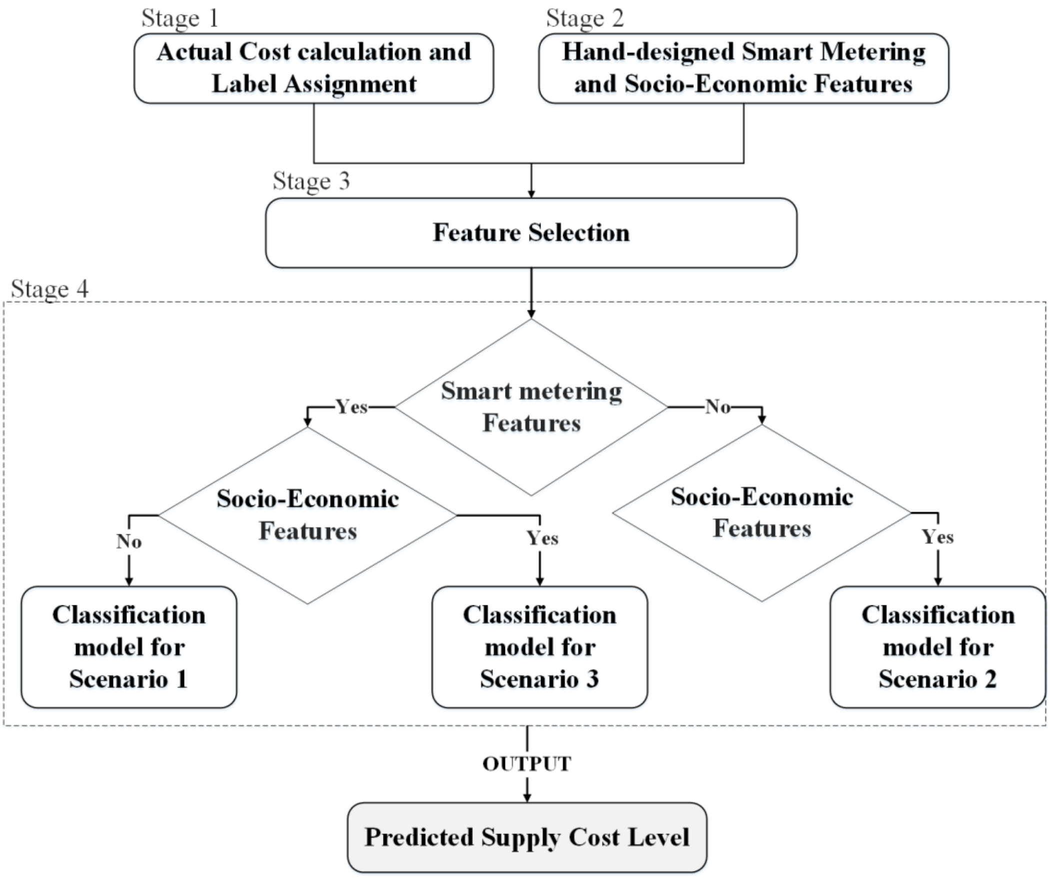


Figure 5- 1: The structure of the proposed customer classification framework

Stage 1: Supply Cost Quantification

On the supply side, wholesale energy cost varies at half-hourly basis. With the advanced smart meters, suppliers can accurately quantify the actual supply cost of individual customers. However, the current retail price is fixed, representing the average over time and customers. This leads to cross-subsidies within customers. Some customers pay less than their actual cost and the deficit will be shouldered by the rest of customers. Customers need to be segmented into different actual supply cost groups so as to be treated with tailored services. Three cost groups are identified (high-cost, medium-cost and low-cost) by calculating the difference between the actual supply cost and the retail price on customers' bills.

Stage 2: Hand-designed Features

Features are generated from the raw smart metering data at this stage for two reasons:

i) To avoid the curse of dimensionality. Specifically speaking, the raw smart metering data are sampled half-hourly and add up to 17520 features over a year. Direct use of the high-dimensional raw data will lead to the over-fitting of the classification model as the number of observations is much smaller than the number of features.

ii) To convert questionnaires to categorical variables which can be used in the classification model. The socio-economic data are usually collected through the questionnaire. It is essential to convert the survey answers to ordinary and dummy variable.

Stage 3: Feature Selection

Feature selection aims to identify a subset of features which are relevant to customers' supply cost. Features with low relevance will be removed. Another method for dimensionality reduction is feature extraction. It creates new features as functions of the original features to be informative and non-redundant. However, instead of adopting feature extraction methods such as PCA, the feature selection method is employed at this stage to retain high interpretability of the selected features.

Stage 4: Classification

The objective of the final stage is to develop a classification model to predict customers' supply cost level based on features selected at stage 3. A number of smart metering and socio-economic features would be used to describe one each observation (customer). In practice, the smart metering data and socio-economic data may not be available for all customers. Hence, the classification algorithm is chosen to build three classification models for different types of input data. Scenario 1 and Scenario 2 are designed for the input data which only contains smart meter data or socio-economic information respectively. Classification model in Scenario 3 is built for the input dataset which includes both of the smart metering data and socio-economic data.

The specific adopted methods and their detailed operation for each stage are presented in the following sub-sections.

5.3.1 Supply Cost Quantification of Individual Customers Based on Smart Metering Data

In reality, the supply cost is made up of several aspects as illustrated in Figure 5-2 [101]. From the breakdown of both gas and electricity bills, nearly two-thirds of the bill are contributed by wholesale supply cost.

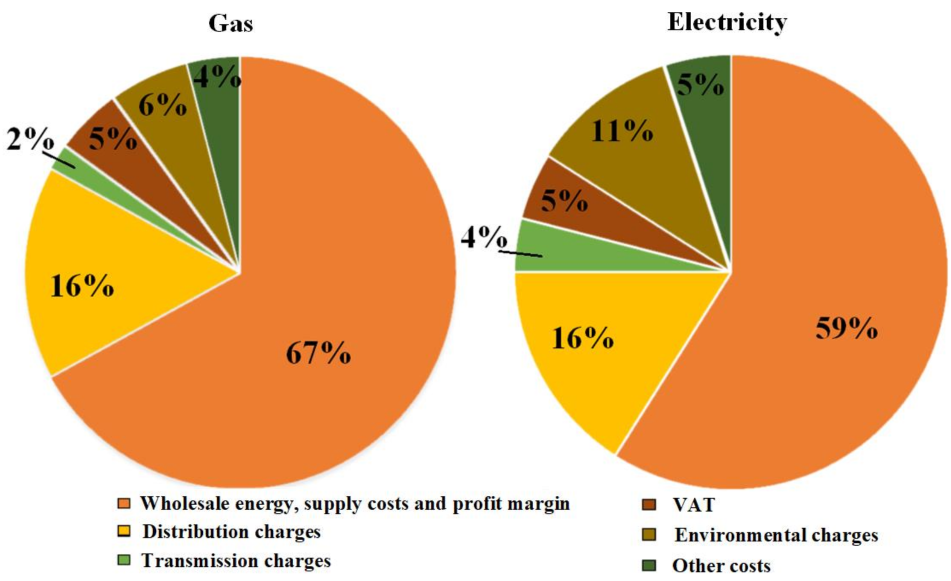


Figure 5- 2: The breakdown for gas and electricity bills

Therefore, in this paper, the supply cost for each customer is scaled up based on their wholesale market cost, which is quantified through (5-1):

$$\delta_n = \frac{\sum_{i=t_{start}}^{i=t_{end}} \sum_{j=1}^{48} (S_{nij} \times p_{ij})}{\gamma} \quad (5\text{-}1)$$

where δ_n indicates the supply cost of customer n over the period from t_{start} to t_{end}. The smart metering data are half-hourly collected and denoted by S. p_{ij} represents the electricity price over the j^{th} half-hourly period on the i^{th} day. The coefficient γ indicates the share of the wholesale market cost in the total supply cost (e.g. γ equals to 59% for electricity as Figure 5-2 shows).

Instead of paying the supply cost δ_n, the majority of domestic customers are charged against a fixed price p_f regardless of the time of use. The difference is calculated in (5-2)

$$\varepsilon_n = \frac{\delta_n - (p_f \times \sum_{i=t_{start}}^{i=t_{end}} \sum_{j=1}^{48} S_{nij})}{p_f \times \sum_{i=t_{start}}^{i=t_{end}} \sum_{j=1}^{48} S_{nij}} \times 100\% \tag{5-2}$$

According to the value of ε_n, customers are segmented into three groups: high-cost, medium-cost and low-cost as demonstrated in (3):

$$L_n = \begin{cases} 1 & if\ \varepsilon < -1\% \quad Low\ Cost \\ 0 & if -1\% \le \varepsilon \le +1\ \% \quad Meduim\ Cost \\ 2 & if\ \varepsilon < +1\% \quad High\ Cost \end{cases} \tag{5-3}$$

where L_n represents the label assigned to customer n.

By following (5-3), the actual supply cost levels of customers will be represented by label 0, 1 and 2.

5.3.2 Feature Generation from the Smart Metering and Socio-Economic Data

Feature generation from the raw data will assist in the further classification model establishment. Instead of the massive half-hourly smart metering data, smart metering features effectively avoid the curse of dimensionality of the classification model. Moreover, adopting features can improve the applicability of the classification models. For example, if customers don't get hold of the same length of the smart metering data, the feature-based classification models could still deduce customers' actual electricity usage habits.

For the socio-economic data, feature generation is a conversion process. The information in the questionnaire is transformed into socio-economic features to better support the classification models.

- **Smart metering feature generation**

The 56 smart metering features are constituted by the widely-used features in other researches [56, 102] and features designed by this paper, which are detailed listed in the Table 5-1. Due to the significant seasonal effect on customers' usage, load features are generated specifically across every time horizon, which are winter data, spring data, summer data, and autumn data and whole year data. Meanwhile, within each season or the annual period, the load features are extracted on a finer time horizon, such as the weekday, weekend and the entire season. All the 56 load features describe the original household load profiles from four aspects:

1) Consumption related features, such as the average daily peak demand, average demand for specific periods and so on;
2) Ratio related features, like the average ratio of mean over max demand;
3) Occurrence related features. For instance, the average peak demand time and the hours when the demand is above the mean value.
4) Statistical features, for example the average of correlation coefficient between two adjacent days and the standard deviation of the usage.

- **Socio-Economic feature generation**

The socio-economic information of households was collected through a questionnaire on a survey. Socio-economic questions can be categorized into two variable types: 1) dummy variable; and 2) ordinal variable. The dummy variables only have two options, "1" and "0", which are used to sort data into mutually exclusive categories (such as smoker/non-smoker) [56].

The ordinal data is a categorical, statistical data type where the variables have natural, ordered categories [75]. These data exist on an ordinal scale, for example, the survey question "Is your general health poor, reasonable, good, or excellent?" may have those answers coded as 1, 2, 3, and 4 respectively.

Table 5- 1: The explanation of 56 load features

Consumption related features		*Occurrence time-related features*	*Ratios features*		*Statistical features*
c_bd_ave	c_t_ave_night	o_bd_above_mean	r_ave_bd/wd	r_t_ave_evening/ave_noon	s_bd_ave_sd
c_bd_ave_day	c_t_ave_noon	o_bd_ave_peak_time	r_bd_ave_mean/max	r_t_ave_morning/ave_noon	s_t_ave_corr
c_bd_ave_evening	c_t_ave_min	o_t_above_mean	r_bd_ave_min/max	r_t_ave_night/ave_day	s_t_ave_sd
c_bd_ave_morning	c_t_ave_peak	o_t_ave_peak_time	r_bd_ave_min/mean	r_t_ave_noon/ave_total	s_wd_ave_sd
c_bd_ave_night	c_wd_ave	o_wd_above_mean	r_bd_ave_evening/ave_noon	r_wd_ave_mean/max	
c_bd_ave_noon	c_wd_ave_day	o_wd_ave_peak_time	r_bd_ave_morning/ave_noon	r_wd_ave_min/max	
c_bd_ave_min	c_wd_ave_evening		r_bd_ave_night/ave_day	r_wd_ave_min/mean	
c_bd_ave_peak	c_wd_ave_morning		r_bd_ave_noon/ave_total	r_wd_ave_evening/ave_noon	
c_t_ave	c_wd_ave_night		r_t_ave_mean/max	r_wd_ave_morning/ave_noon	
c_t_ave_day	c_wd_ave_noon		r_t_ave_min/max	r_wd_ave_night/ave_day	
c_t_ave_evening	c_wd_ave_min		r_t_ave_min/mean	r_wd_ave_noon/ave_total	
c_t_ave_morning	c_wd_avepeak				

Where the first bit :{ c= consumption-related features; r= ratio-related features; s: statistic-related features; o= occurrence time-related features ;} Second bit :{ t = total time range; bd = business day; wd = weekend;} Third bit :{ / = division sign; ave = average; sd = standard deviation; corr= correlation}

- **Selecting Features Algorithm**

Classification with all features will result in the over-fitting and high variance problems. Moreover, the irrelevant features could degrade the performance of classification models both in speed (due to the high-dimensionality) and predictive accuracy (due to the irrelevant features). Thus, a feature selection algorithm is required to select the most discriminable features.

From feature generation in stage 2, it can be noticed that there are two characteristics of the features:

1) Containing both discrete features (i.e. the dummy features) and continuous features (i.e. ordinal and numerical features);

2) Strong interaction between features. For example, the socio-economic feature, “how many children in your household”, is not independent with feature “the square meters are your house”.

Hence, the feature selection algorithm employed in this stage is required to be robust to feature interactions and being applicable for discrete and continuous data.

- **Developing the Classification Models**

The fourth stage is classification with the significant features selected. Multiple classification methods can be employed depending on the dataset. The Irish data used in this paper has a small sample size with high in dimension. Hence, the complex classification models with more parameters, such as the neural networks, are not suitable as they require large training samples to avoid overfitting and reduce the variance. While the kernel methods are suitable to operate this kind of dataset because the kernel trick can avoid the computation burden of the product of high-dimensional features by simply computing the inner products.

Base on the data used in this paper, the two classic kernel techniques, such as the Support Vector Machine (SVM) and Kernel Fisher Analysis (KFA) are assessed and compared with the performance achieved by the Artificial Neural Network (ANN).

5.4 Implementation of the Proposed Framework

The proposed cost-reflective customer classification is tested on a publicly available dataset with residential energy consumption data of 836 Irish households. This smart meter dataset is collected from the Smart Metering Electricity Customer Behaviour Trails (CBTs), launched by the Commission for Energy Regulation (CER) [94], with the socio-economic information for each customer who was involved. The smart meter data was recorded at half-hourly basis from 14th July 2009 to 31st December 2010. The socio-economic data are demonstrated in the form of a questionnaire which comprises 142 questions to describe the socio-economic information for each customer.

To find out the most appropriate feature selection and classification algorithms, a series of state-of-art methods are adopted at stage 3 and 4. Their results have been compared with each other within the proposed framework, which are:

1) Support Vector Machine: SVM is a widely-used kernel classification method [103, 104]. This model directly employs SVM without feature selection to be a control group for performance of other algorithms.

2) Principal Component Analysis (PCA)-SVM: PCA [105] is a popular feature extraction method for dimensionality reduction. By comparing with the result of this model, the effectiveness of the feature selection algorithm can be validated.

3) ReliefF-SVM: ReliefF algorithm has been used in many research [106, 107] to reduce the dimension. It is adopted and collaborates with SVM to segment customers through their supply cost.

4) ReliefF-KFA: Like SVM, KFA is another kernel classification technique adopted in other research [108]. The fourth model adopts the ReliefF and KFA at stage 3 and 4 respectively.

5) ReliefF-ANN: The Artificial Neural Network (ANN) displays a strong discriminative ability for residential customers' socio-economic features in many previous works [109]. Hence, the fifth model utilises the neural network after ReliefF to investigate its performance on cost-reflective classification.

To compare the classification performances of different models objectively, the 4-folds cross-validation has been applied to all models. It can avoid the overfitting and selection bias by using all observations for both training and validation. Then, with the 80% of the total residential customers training, the accuracy of each classification model has been presented in Table 5-2.

Table 5- 2: Comparison results between ReliefF-SVM and other methods

	SVM	*PCA-SVM*	*ReliefF-SVM*	*ReliefF-KFA*	*ReliefF-ANN*
Scenario 1	73.92%	69.85%	**74.88%**	65.07%	69.05%
Scenario 2	**54.29%**	37.35%	53.31%	49.75%	53.76%
Scenario 3	71.29%	61.51%	**75.00%**	61.00%	69.05%

Among the tested algorithms, the ReliefF-SVM model reaches the best classification result in Scenario 1 and Scenario 3. Although the performance of the SVM model is better for Scenario 2, the ReliefF-SVM achieves a slightly inferior accuracy with a significantly reduced feature number. The SVM model uses all the geo-demographic. However, ReliefF-SVM reduce the feature number to 70. By comparing its results with the SVM model and PCA-SVM model, the effectiveness of feature selection can be substantiated.

Therefore, the ReliefF is employed to weight and rank the features inputting in each scenario and SVM is used to build the three classification models. To avoid features with wider numeric range dominating those in smaller range, all the extracted features are normalized before selection.

The ReliefF [110] estimates features' weights according to their ability to discriminate between instances which are near to each other. For this purpose, the differences between the values of feature f in instances X and Y, X_f and Y_f, can be defined as (5-4) when X_f and Y_f are discrete features, as (5-5) when X_f and Y_f are continuous features. This also enable the ReliefF to handle both of those two types of data.

$$diff(X_f, Y_f) = \begin{cases} 0 & if\ X_f = Y_f \\ 1 & if\ X_f \neq Y_f \end{cases} \tag{5-4}$$

$$diff\left(X_f, Y_f\right) = \frac{\left|X_f - Y_f\right|}{\max(f) - \min(f)} \tag{5-5}$$

With the given training data set δ, the ReliefF algorithm cycles trough a randomly selected instance set m. For each instance $R_i$which is an element of set m, ReliefF searches for k instances who are its nearest neighbors in the same or different classes respectively, named the nearest hits and the nearest misses. Hits and misses are denoted respectively by $\{H_1, H_2, \cdots, H_k\}$ and $\{M_1(C), M_2(C), \cdots, M_k(C)\}$ where C represents the class where each miss $M_k$belongs to. The weight of each feature can be estimated and iterated based on the value of $diff\left(R_i, H_j\right)$and $diff\left(R_i, M_j(C)\right)$. The update-weight function of the ReliefF algorithm is demonstrated in (5-6):

$$\begin{aligned} &W[f] \\ &= W[f] - \frac{\sum_{j=1}^{k} diff\left(R_i, H_j\right)}{(m \times k)} \\ &+ \frac{\sum_{C \neq class(R_i)}\left[\frac{P(C)}{1 - P(class(R_i))}\sum_{j=1}^{k} diff(R_i, M_j(C))\right]}{(m \times k)} \end{aligned} \tag{5-6}$$

The weight of feature f, $W[f]$, approximates the difference of two probabilities. The first probability is the second element of (5-6) $\frac{\sum_{j=1}^{k} diff(R_i,H_j)}{(m\times k)}$. It represents the probability of the different value of feature f between the selected instance R_i and its nearest instances hits H_j. Another probability is the third element of the (5-6), which describes the different value of feature f between R_i and its nearest instances misses $M_{\mathrm{j}}(C)$.

Due to the "nearest instance" condition, the ReliefF weights are averaged over local estimates in a smaller part of instance subspace instead of the global instances [111]. This enables ReliefF to be aware of the contextual information. Therefore, ReliefF can correctly estimate the quality of features where features have strong interactions [107]. The pseudo-code of the ReliefF algorithm is as follows:

ReliefF Algorithm ($\boldsymbol{\delta}, \boldsymbol{m}$):

Input the training supply cost instance set δ with the cost label of each instance.

Set the initial value of all feature' weight as zero: $W[F] = (0,0,\cdots,0)$

For i from 1 to m **do:**

Randomly select an instance R_i

Find k nearest hits for $R_i,\ H_j$

For each class $C \neq class(R_i)$**do:**

  Find k nearest miss for $R_i$from class $C,\ M_j(C)$

For F from 1 to F **do:**

  $W[F]$ is iterated by the update-weight function in (6)

End;

Output the vector $W[F]$of estimations of the qualities of features

---

Then, according to the weight of each feature, the features could be sorted by the ReliefF algorithm. Equation (5-7) would be used to select the appropriate subsets of features, I, which should be fed into the SVM:

$$\Theta(I) \geq \Theta(\Omega)\ \ where\ \forall\Omega \in F, \Omega \neq I | I \in F \tag{5-7}$$

where F represents the subset of all features and the function $\Theta()$ denotes the classification function.

The primary goal of using SVM is to classify the unseen data by maximizing the distance between the data points who are the closest to the separating hyperplane. The two-class problem shown in Figure 5-3 [112] is an example of adopting SVM to separate the dot marks and the rectangle marks based on a hyperplane (the dashed line).

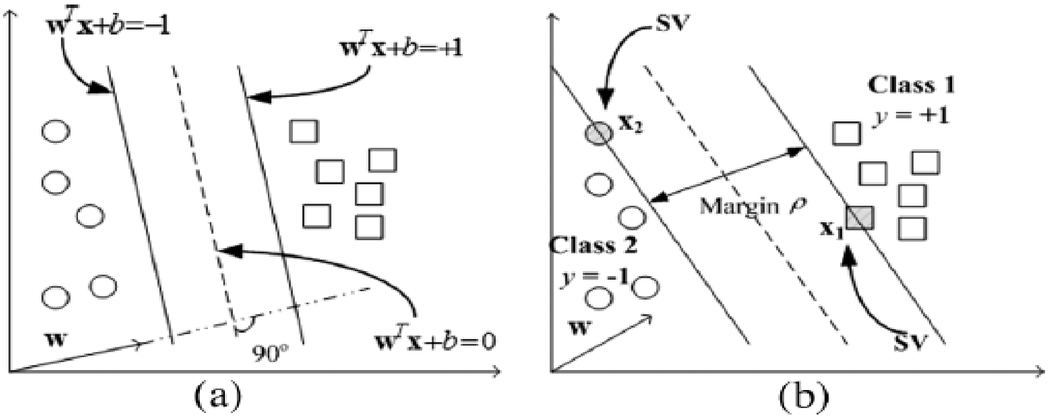


Figure 5- 3: Two-class classification problems example for SVM

The solid lines on both sides of the hyperplane demonstrate a gutter, the optimized goal is to find a hyperplane with the maximum gutter width. By using an orthogonal vector $\vec{\omega}$ and bias b, the point $x_i, i \in (1, l)$, which indicates an n-dimensional input vector, on those three lines satisfies:

$$\begin{cases} \vec{\omega} \cdot \vec{x_i} + b = 0, & where\ x_i\ is\ on\ the\ hyperplane \\ \vec{\omega} \cdot \vec{x_i} + b = +1 & where\ x_i\ is\ on\ the\ right\ bound \\ \vec{\omega} \cdot \vec{x_i} + b = -1, & wherex_i\ is\ on\ the\ left\ bound \end{cases} \quad (5\text{-}8)$$

To formulate the expression of the gutter width D, we set two closest points to the hyperplane $\vec{x_1}, \vec{x_2}$, which also called as Super Vectors (SVs). The margin of the gutter ρ can be expressed as (5-9):

$$\rho = \frac{\vec{\omega}}{\|\vec{\omega}\|} \cdot (\vec{x_1} - \vec{x_2}) = \frac{2}{\|\vec{\omega}\|} \quad (5\text{-}9)$$

From (5-9) it can be found that to maximize the margin ρ, it is equivalent to minimize $\frac{1}{2}\|\vec{\omega}\|^2$. Therefore, it become an optimization problem with constrain (5-10):

$$y \cdot (\vec{\omega} \cdot \vec{x_i} + b) - 1 \geq 0, \qquad y \in \{+1, -1\} \quad (5\text{-}10)$$

To deal with this kind of optimal problem, it is easier to solve in its dual formulation (5-11) in terms of the Lagrange multipliers α_i by maximizing:

$$L = \frac{1}{2}\|\vec{\omega}\|^2 - \sum_{i} \alpha_i \left[y_i(\vec{\omega} \cdot \vec{x_i} + b) - 1 \right] \tag{5-11}$$

where L is the Lagrangian. Hence, (5-11) follows from the saddle point condition constrains by the partial derivatives of L. Finally, we can get (5-12).

$$L = -\frac{1}{2}\sum_{i}\sum_{j} \alpha_i\alpha_j \; y_i y_j(\vec{x_i} \cdot \vec{x_j}) + \sum_{i} \alpha_i \tag{5-12}$$

From (5-12), we find that the performance of maximization of the separable models in SVM only depends on the inner product of each two samples. To handle the nonlinearly separable cases, SVM can extend to map the input vector into higher dimensional space through a kernel function which also depends on the inner product.

In this research, the three classification models are built by SVM with the Radial Basis Function (RBF) kernel function [112], which is defined as (5-13), to segment customers' cost level with different types of input dataset.

$$K_{(\vec{x_i},\ \vec{x_j})} = \exp\left(-\frac{\left\|\vec{x_i} - \vec{x_j}\right\|^2}{2\sigma^2}\right) \tag{5-13}$$

5.5 Results and Analysis

In this section, the classification result achieved by adopting ReliefF and SVM in the proposed classification framework is compared with the load-profile based classification on the same Irish dataset.

Furthermore, the effect of the selected features on actual supply cost has been investigated to summarize the discriminative characteristics for each cost level customers.

5.5.1 Comparison with Load-Profile Based Classification

The load-profile based classification is implemented through two steps:

Step 1: Unsupervised clustering for customers' load profiles

Before clustering, the load profile for each customer is represented by the yearly average energy usage pattern after normalization. Then, two popular clustering methods, the K-means [113] and Euclidean distance based hierarchical algorithm [114, 115], are adopted as the candidate to cluster customers with similar load profile into the same group. The cluster number is set from 3 to 7 for both methods.

After clustering, the largest actual cost group among the customers in the same cluster would be treated as the cost label for this cluster. In other words, no matter how many clusters there are, all clusters would be concluded into three cost groups, which are high, medium and low.

Step 2: Supervised classification for new customers with three different types of input data

In this step, the SVM is adopted to classify customers into different load-profile clusters. Customers who are allocated into each cluster would be represented by the cost label of that cluster. Finally, with 80% of training, the accuracy results after a 4-folds cross-validation for three Scenarios are calculated.

The accuracy results show that the hierarchical clustering performs better than the K-means one, which have been demonstrated in Table 5-3. Considering the performance for all three scenarios, 3 clusters is the best number for the cluster. However, by comparing with the proposed classification framework, the ReliefF-SVM model in Table 5-2 achieves better accuracy than the load-profile based one in Scenario 1 and Scenario 3. Although in Scenario 2 the hierarchical clustering reaches high accuracy, there is still an error between the correct load-profile class and the right cost label. The cost label for each load-profile class is represented by the real cost level for the largest proportion of customers, which cannot represent every customer in that load-profile class. When there are 3 clusters, the error between the load profile groups and the cost labels are 11.32% for high-cost label, 18.03% for medium-cost label and 9.33% for low-cost label. Therefore, by considering the error between load profile groups and the real cost label, the accuracy results of supply cost classification based on hierarchical clustering algorithm are 64.33%, 52.99% and 66.73% for the three scenarios respectively

Hence, the proposed classification framework would be more accurate to identify the actual supply cost for customers. The framework improves the accuracy by 16.40%, 0.60% and 12.40%.

Table 5- 3: Classification accuracy results based on hierarchical clustering algorithm

Cluster Number	*Scenario 1*	*Scenario 2*	*Scenario 3*
Accuracy of ReliefF-SVM model	**74.88%**	**53.31%**	**75.00%**
3 clusters	72.54%	64.64%	73.60%
4 clusters	71.77%	57.78%	68.54%
5 clusters	69.86%	58.73%	66.75%
6 clusters	60.05%	34.21%	50.60%
7 clusters	60.29%	33.85%	51.08%

5.5.2 Results Analysis for the Selected Features

An advantage of the propose cost-reflective customer classification framework is easy-interpretable. The features which are selected by the feature selection algorithm have the potential to provide insights into the key drivers of the difference between supply cost groups. This knowledge can aid suppliers in designing more tailored services to reduce the cost for the high-cost group.

The feature selection algorithm figures out that the best feature-number is 25 for Scenario 1, 70 for Scenario 2, 65 for Scenario 3. To further concentrate on the analysis of the most significant features, this research will analysis the most frequently selected feature in all three Scenarios throughout the cross-validation.

- **Results analysis for Smart Metering Features**

The ability to produce interpretable discriminative features is an advantage for the proposed cost-reflective customer classification framework. Among all the selected features, some of smart metering features show strong discriminative ability in customers' supply cost level by their own. The specific relation between those features and the supply cost should be valuable for energy suppliers to guide them to provide more premium service (tariff plans design, usage recommendation and so on) to survive in this competitive market.

The selected smart metering features are listed in Table 5-4, followed with the information about the selection rate in two Scenarios throughout the cross-validation. The nomenclature for the smart metering features is shown at the bottom of the table. If a smart metering feature is coded like "r_wd_ave_night/ave_day", it represents a ratio-related feature about the ratio between average night consumption and average daytime consumption for the weekend.

Table 5- 4: The smart meter features selected by ReliefF algorithm

No.	***Feature Name***	***Selection Rate in Scenario 1***	***Selection Rate in Scenario 3***
1	*r_wd_ave_night/ave_day*	100.0%	100.0%
2	*c_t_ave_night*	100.0%	88.30%
3	*c_bd_ave_night*	88.2%	71.80%
4	*r_wd_avemorning/avenoon*	90.3%	87.1%
5	*r_t_avemorning/avenoon*	84.3%	59.3%
6	*r_t_avenight/aveday*	79.4%	81.1%
7	*r_bd_avenight/aveday*	82.9%	75.5%
8	*r_bd_avemorning/avenoon*	79.6%	/
9	*s_t_ave_sd*	77.9%	/
10	*c_wd_ave_night*	60.3%	/
11	*s_bd_ave_sd*	57.2%	/

The impacts of the top 7 smart metering features in Table 5-4 are demonstrated in Figure 5-3. The cost levels of all 836 customers are plotted with different colours against the corresponding feature for each column. Customers are ranked by the values of each feature of the corresponding column. The highest value goes to the top of this column while the customer with the lowest value goes to the bottom. Consequently, the y-axis only represents the accumulated number of customers and each row does not necessarily represent the same customer.

The Figure 5-4 shows that all those 7 smart metering features are negatively correlated with the actual supply cost. The high cost label customers always have a high value for these 7 features.

Among them, the 4th feature (*r_wd_avemorning/avenoon*) and the 5th feature (*r_t_avemorning/avenoon*) are related with the ratio of demand happened during the morning (6 a.m.to 10 a.m.) and noon (10 a.m. – 2 p.m.). Additionally, the other five smart metering features describe the usage happened during the night period (1 a.m.-5 a.m.). From these results, the following conclusions can be drawn:

1) Ratio between the consumption during the morning and the noon has negative influence on the supply cost level;

2) Ratio between the consumption during the night and the whole day has negative influence on the supply cost level.

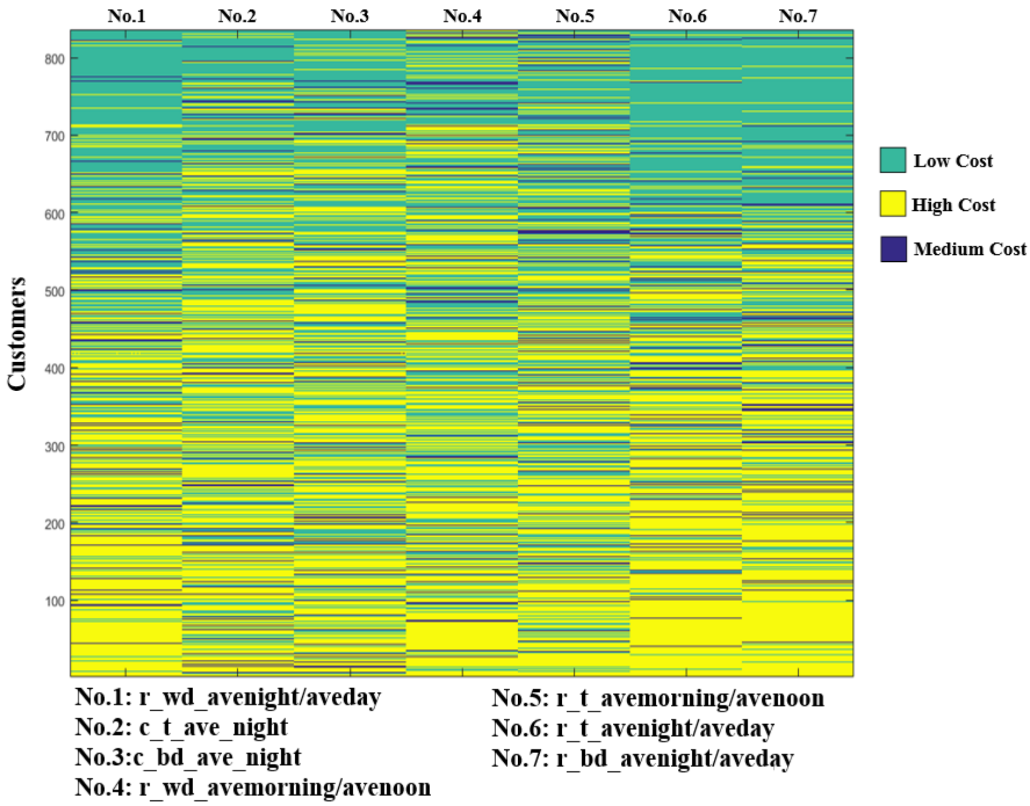


Figure 5- 4: The impact of top 7 smart metering features on cost groups

The last four features in Table 5-5 are only selected in Scenario 3. The relationship between those four features and the actual supply cost is demonstrated in Figure 5-5, which was displayed in the same way as Figure 5-4. In Figure 5-5, the 8th feature (r_bd_avemorning/avenoon) and 10th feature (c_wd_ave_night) show a negative correlation with the customer supply cost which coincides with the findings in Figure 5-4. On the contrary, the 9th (s_t_ave_sd) and 11st (s_bd_ave_sd) smart metering features are positively correlated with supply cost. Both of those two features are related to the standard deviation for the average consumption. Therefore, third conclusion is:

3) The dispersion degree (variance) of the whole consumption has positive influence on the supply cost level.

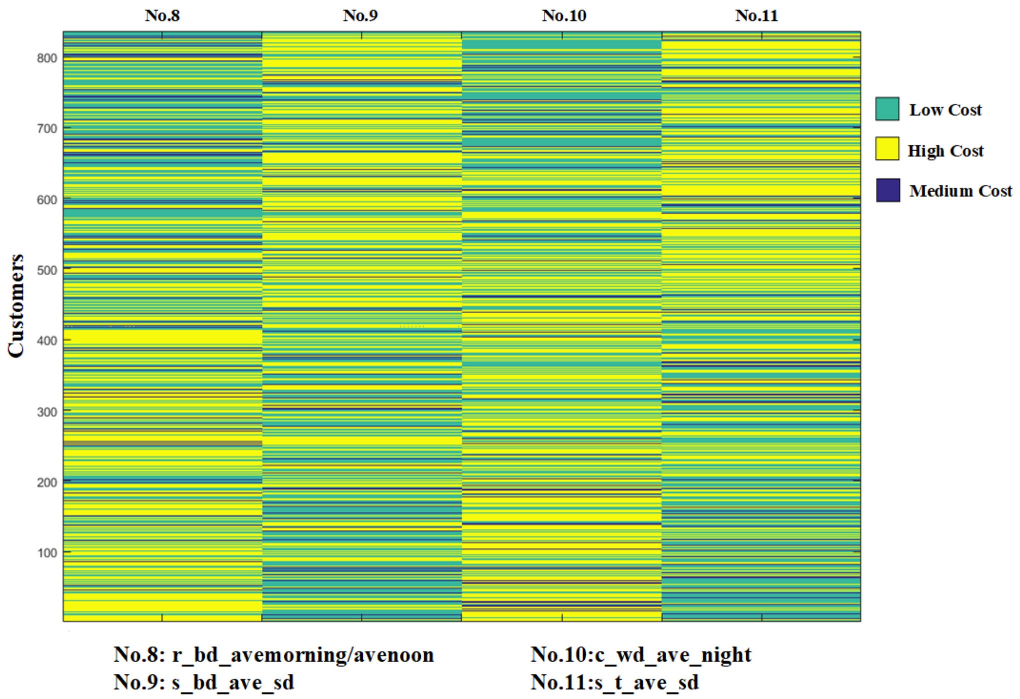


Figure 5- 5: The impact of four smart metering features on cost groups

- ## Results Analysis for Socio-Economic Features

The socio-economic features whose average selection rate are over 50% in both Scenario 2 and Scenario 3 would be chosen to analysis. Table 5-5 presents detailed information and the selection rate in each Scenario throughout the cross-validation process. There are 10 geo-demographic features exhibiting their outstanding discriminative ability on customers' supply cost in both two Scenarios. The options of the features are included in the table as well.

The impacts of different answers for each socio-economic feature are demonstrated in Figure 5-6. Feature 1 to feature 10 represent those ten socio-economic features in Table 5-5 respectively. Each bar in Figure 5-6 represents a subgroup within which customers have the same answer for the corresponding feature. Different cost levels are illustrated by different colours. The length of the area rendered by different colours denotes the percentage of the corresponding cost level. Additionally, at the bottom, a comprehensive data table has been given to display the specific percentage values.

From Figure 5-6, the 7th feature shows that owing multiple TVs has a remarkable influence on the medium cost group. The proportion of medium cost group increase from 5.263% (do not have a TV) to 29.41% (have more than 3 TVs). Additionally, the owning of other electrical appliances, such as tumble dryers, washing machines, game consoles has positively related with the supply cost. However, the lap-top computers (the 8th feature) have contrary impaction on the supply cost. Therefore, it can be concluded as:

1) The electricity appliances (except the lap-top computer) have positive influence on the supply cost level;

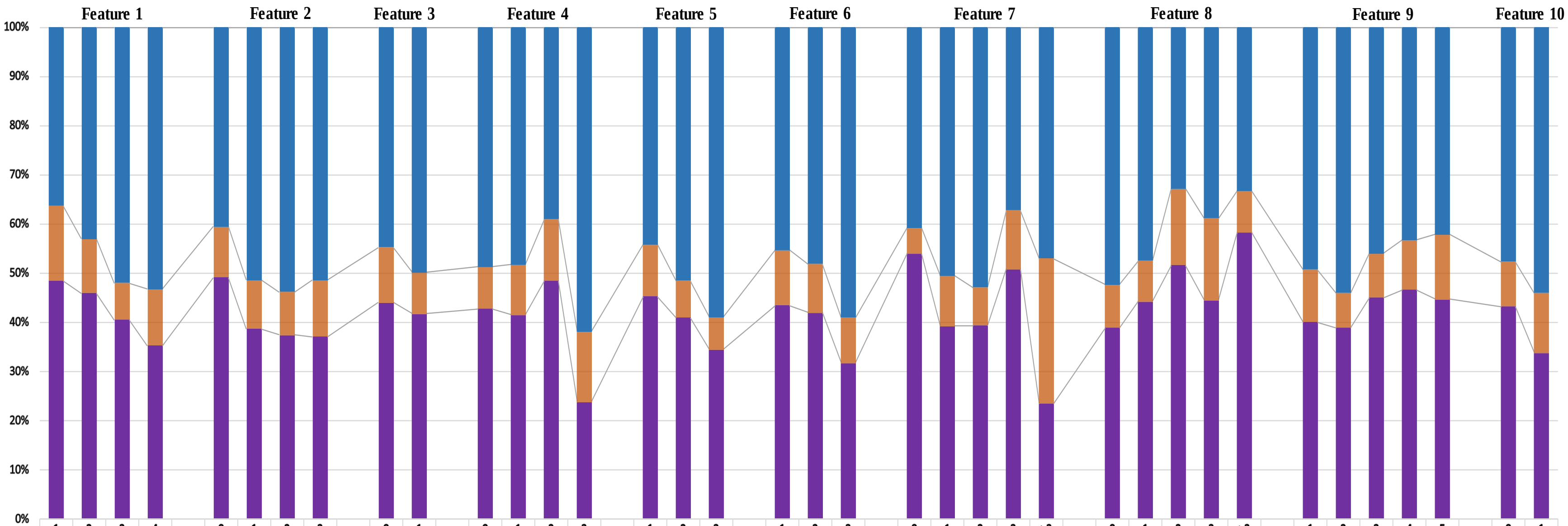
Both 9th and 10th socio-economic features are related the energy saving consciousness of the customers. Using the 9th feature as an example, customers, who are unsatisfied with the opportunity of selling back extra solar power, are highly possible to generate more solar energy than they could consume. From Figure 5-6 it can be found that the proportion of high cost group decreases with the dropping of satisfaction level. However, the willingness to reduce more usage is negatively correlated with the customers' supply cost. Therefore, it can be concluded as:

2) The energy saving consciousness has inconsistent effects on customers' actual supply cost level.

Table 5- 5: The socio-economic features selected by ReliefF algorithm

No.	*Feature Description*	*Options for each Feature*	*Selection Rate in Scenario 2*	*Selection Rate in Scenario 3*
1	Frequency of using the TV's greater than 21 inches	1~4 = Frequency from low~ strong	98.8%	82.5%

2	Number of housemates are typically in the house during the day (e.g. 5-6 hours/ day)	Real Number	96.3%	96.3%
3	Number of Tumble dryers in your home	Real Number	78.8%	88.5%
4	Number of Games consoles (Xbox, PlayStation or Wii) in your home	Real Number	92.1%	88.3%
5	Frequency of using the Washing machines	1~3 = Frequency from low~ strong	74.2%	47.5
6	Frequency of using the Games consoles (Xbox, PlayStation or Wii)	1~3 = Frequency from low~ strong	87.9%	72.1%
7	Number of the TV's greater than 21 inches in your home	Real Number	65.0%	77.4%
8	Number of the Lap-top computers in hour home	Real Number	70.8%	85.4%
9	Satisfaction of the opportunity to sell back extra electricity you may generate (from solar panels) to your electricity supplier	1~5 = Satisfaction from high~ Low	98.8%	98.8%
10	Would you like to do more to reduce your electricity usage?	0=No 1=Yes	74.2%	70.4%



	Feature 1				Feature 2				Feature 3	
	1	2	3	4	0	1	2	3	0	1
High cost level	36.364	43.066	52.066	53.309	40.517	51.44	53.817	51.429	44.788	49.913
Medium Cost level	15.152	10.949	7.438	11.397	10.345	9.8765	8.7786	11.429	11.197	8.3189
Low cost level	48.485	45.985	40.496	35.294	49.138	38.683	37.405	37.143	44.015	41.768

	Feature 4				Feature 5			Feature 6		
	0	1	2	3	1	2	3	1	2	3
High cost level	48.768	48.315	39.063	61.905	44.206	51.575	58.889	45.283	48.148	59.091
Medium Cost level	8.4507	10.112	12.5	14.286	10.515	7.4803	6.6667	11.321	9.8765	9.0909
Low cost level	42.782	41.573	48.438	23.81	45.279	40.945	34.444	43.396	41.975	31.818

	Feature 7					Feature 8				
	0	1	2	3	>3	0	1	2	3	>3
High cost level	40.789	50.498	52.913	37.288	47.059	52.369	47.507	32.813	38.889	33.333
Medium Cost level	5.2632	10.199	7.767	11.864	29.412	8.7282	8.2111	15.625	16.667	8.3333
Low cost level	53.947	39.303	39.32	50.847	23.529	38.903	44.282	51.563	44.444	58.333

	Feature 9					Feature 10	
	1	2	3	4	5	0	1
High cost level	49.268	54.04	45.97	43.333	42.105	47.6	54
Medium Cost level	10.732	7.0707	8.9552	10	13.158	9.05	12.1
Low cost level	40	38.889	45.075	46.667	44.737	43.2	33.7

Figure 5- 6: The distributions of three cost groups for 10 socio-economic features

5.6 Chapter Summary

This chapter proposed a novel framework to do the customer classification based on customers' actual supply cost. Compared to existing customer classification method, the propose classification framework has two superiorities.

i) The accurate cost level estimation for customers with different types of input data

The proposed framework builds three scenarios for customers with different input data: Scenario 1) who only provide their smart metering data; Scenario 2) who only provide their socio-economic information; Scenario 3) who have socio-economic information with smart metering data. In this way, the suppliers would not be limited by the absence of customers' historical smart metering data when estimation their supply cost. The effectiveness of the proposed methodology is evaluated on the CER Irish dataset. The accuracy of the three scenarios can reach 74.88% and 53.31% and 75.00% respectively, which improves the accuracy by 16.40%, 0.60% and 12.40% compared to the existing classification method.

ii) The interpretability of the significant features which drive the difference between supply cost.

The interpretability of the resulted features is valuable for suppliers and policy makers. Their strong discriminative ability in customers' cost level can provide valuable insights into customers' characteristics in different cost groups. This knowledge can guide the suppliers and policy makers to design more premium and tailored services, such as tariff designs, demand-side responses programs

Based on a case study of Irish smart metering data, the proposed cost-reflective classification framework reveals key findings which were not discovered by the traditional load-profile classification methods. They are summarised as follows:

1) The ratio between the consumption during the morning and the noon has a negative influence on the supply cost level.

2) The ratio between the consumption during the night and the whole day has a negative influence on the supply cost level;

3) The dispersion degree (variance) of the whole consumption has a positive influence on the supply cost level;

4) The electricity appliances (except the lap-top computer) have a positive influence on the supply cost level;

5) The energy saving consciousness has inconsistent effects on customers' actual supply cost level.

These findings can assist the suppliers to launch more tailored tariff plans and services for different cost levels customs, which strengthen the competitiveness of the suppliers in the highly competitive retail market.

Chapter 6

The Impact of Socio-Economic Features on the Responsiveness to Different Tariff Plans

T**HIS** chapter utilised the socio-economic data with the load features and the TOU tariff data to detect the significant features for customers' responsiveness to different tariff plans.

6.1 Introduction

As mentioned in previous chapters, renewable electricity generation capacity evolved very fast over the past decade. This brought a huge challenge in keeping the supply and the demand in synchronous to accommodate the intermittent renewable power into the grid. The demand flexibility in the residential sector, which occupies the biggest portion (30% - 40%) of the total electricity consumption in numerous countries [116, 117], has caught attention as a potential solution to equilibrate the network [118, 119]. In recent decade, various Demand-side Response (DR) schemes emerged to motivate customers via designed incentives to modify their usage pattern.

However, although many literatures investigate the potential demand response of residential customers, there is limited evidence as to how this potential can be fully attained and at what cost [120]. The respond of different households is not equal to the same DR programme. The responsiveness may be related to many factors, such as the load pattern, electrical appliances, and energy-saving awareness of customers and so on. The author in [120] after reviewing a series of literature indicates that the real response may only 1% ~ 10% peak load reduction and 0% ~ 5% total energy consumption reduction for a DR programme which is expected to achieve a reduction around 15% ~ 20 % for peak demand and 10% for the total demand.

Therefore, many researchers have focused on the investigation of potential responsiveness of residential customer, which can be categorised into two types of research:

1) *Analysis based on customers' characteristics:*

   The investigated customers' characteristics include the consumption characteristics and the socio-demographic characteristics, such as building type, appliances and so on. In the early literature, researchers paid more attention to understand the impact of residential customers' load behaviour through data-mining. They aimed to improve the efficiency of DR programmes through more appropriate customer targeting based on their consumption characteristics [14, 100, 114, 121-124].

   Later, in many practical DR experiments [50, 52, 125-128], the researchers found that there are huge deviations among the DR responsiveness results. This difference is mainly caused by diverse socio-economic conditions of customers. Hence, the housing category [36, 127], appliances utilisation [36, 129-131], and socio-demographic features [36, 50] have been taken into consideration which might explain the DR responsiveness.

2) *Simulation based on demand flexibility sources*

This kind of responsiveness research concentrates on the discrete demand flexibility sources. The main objective is to understand the availability and consumption of electrical appliances. Then, the households' responsiveness to the DR could be aggregated up by setting up the optimal control strategies based on the utilisation time and magnitude for every appliance [132-134]. The kind of research is more employed to access the impact of Demand-Side Management (DSM) programmes. For example, in [132] the time-shiftable appliances, such as washing machine, dishwasher and the power-shiftable appliances like the electrical water heaters and storage has been simulated and optimally scheduled based on their consumption and utilisation characteristics to achieve the best result of the DSM.

However, the major challenges for the existing research of responsiveness are:

1) *Ignoring the interaction effect:*

For those researches which attempt to link the responsiveness with the load or socio-economic characteristics of the household, most of them investigate the impact of every single feature. Fewer studies consider the effect among features interaction on customers' response. The impact of features interaction has little be discussed. Although some work, such as [36], have utilised the regression algorithm to identify the determinants for load shifting, which considers the interaction effect. However, the socio-economic status of a household should be depicted by many features, including the dwelling information features (e.g. house type, floor area), appliance features (e.g. ownership, frequency of use, number of the same appliance), customers' personal features (e.g. income, education, age, family number) and psychological features (e.g. energy-saving awareness, willingness-to-participate). Every feature may have a significant effect on customers' responsiveness by interacting with several other features. The regression algorithm is not appropriate to handle the analysis which has a considerable number of inputs.

2) *Simulation based on the assumption of customers' willingness*

For the analysis based on demand flexibility sources, those simulations of appliances' DR potential mentioned above are based on assumptions about the customers' willingness to participate and the ability to fulfil the load schedule. Several studies had surveyed the influence of customers' attitude on the DR final achievement [135-138]. Although [139, 140] indicates that segmenting households based on their willingness-to-participate in DR could support more efficient strategies, the attitude-behaviour gaps [141, 142] still a barrier of the flexibility in the real life.

Therefore, to handle those challenges, this chapter proposes a framework for DR programmes which can pre-evaluate customers' responsiveness for different tariff plans by taking the interaction effect among customers' load characteristics and socio-economic characteristics. The data utilised in this chapter are collected from a smart metering trial conducted in Ireland [143], in which both smart metering data and the socio-economic survey data were provided with 4 different types of Time-Of-Use (TOU) tariffs. The Irish data allows the investigation of what interacted characteristics segments households' responsiveness to a specific TOU tariff.

The remainder of this paper is organized as follow: Section 6.2 displays the experimental data description and the data pre-processing. Section 6.3 introduces the details of the proposed framework of the responsiveness analysis. The results and discussion are demonstrated in Section 6.4. Finally, Section 6.5 illustrates the summary and conclusions.

6.2 Experimental Data Description

The main objective for this chapter is utilising the interacted socio-economic features with other features (such as the load feature and intervention TOU tariff features) to detect the significant determinants of DR potential responsiveness.

6.2.1 The TOU Tariff Types

The data utilised in this work are published by the Commission of Energy Regulation in Ireland [143]. The trials collect the smart metering data at half-hour basis for 4225 residential households with their socio-economic survey answers. The trails lasted from July 2009 to December in 2010. From 1st July to 31st December 2009 is the benchmark period, and all households are charged with the normal Electric Ireland Tariff is 14.1 pence per kWh. Then four types of TOU tariffs were applied to participants from 1st January 2010 to the end of 2010, which name as Tariff A, B, C, and D. The weekday and weekend tariffs for those four types of TOU are displayed in Figure 6-1 and Figure 6-2, respectively.

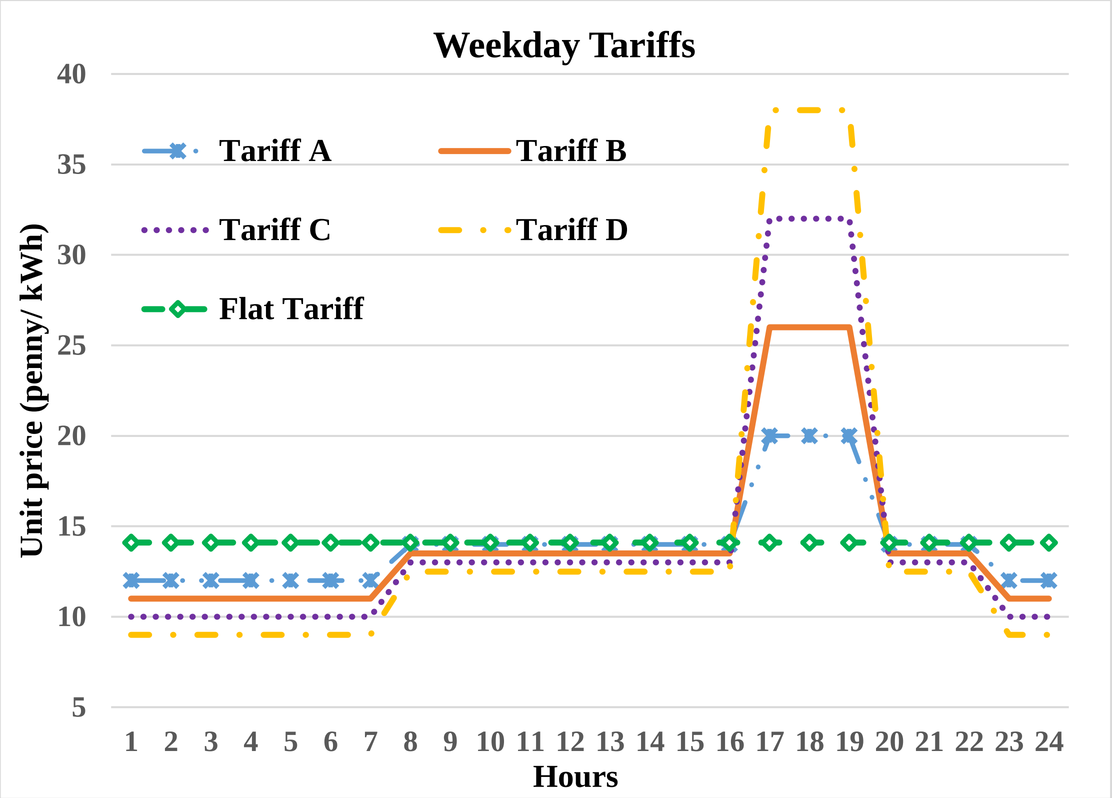


Figure 6- 1: The four types of weekday TOU Tariffs

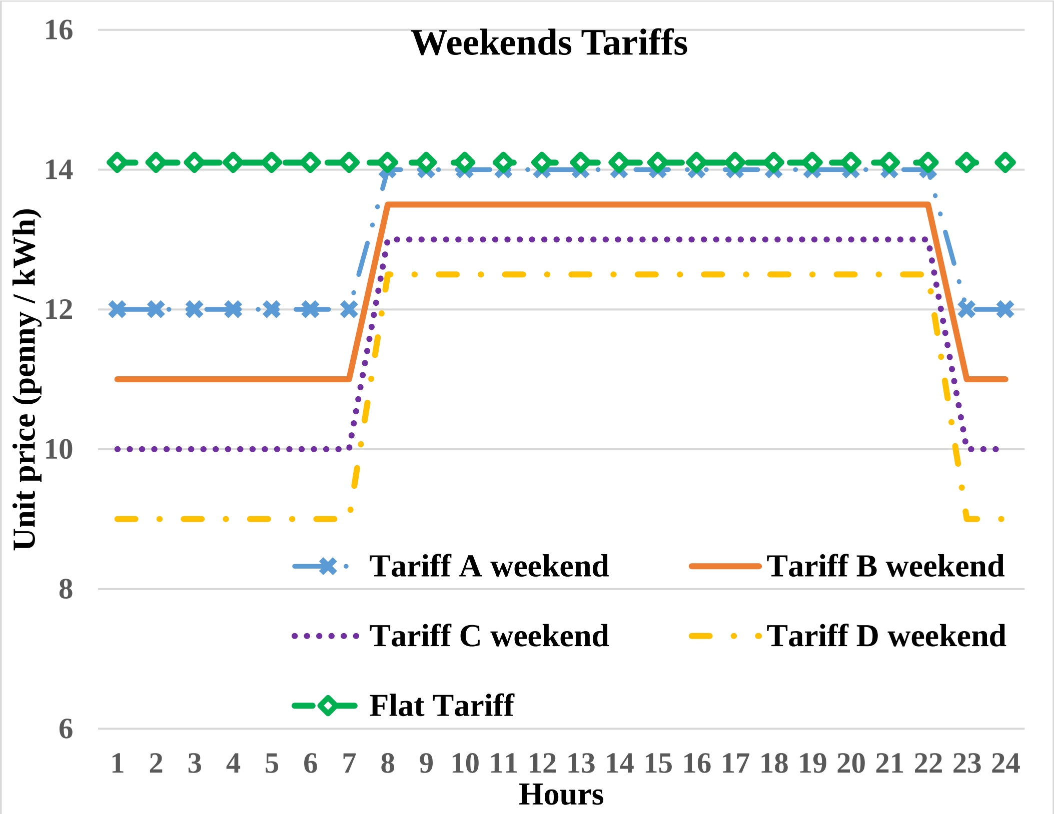


Figure 6- 2: The four types of weekends TOU Tariffs

6.2.2 The Definition of Customers' Responsiveness

In most of the responsiveness analysis studies, the load reduction [36] or the capacity of the shifted load toward the off-peak time period [58] are utilised to quantify the responsiveness of each household. However, by transforming to the smart grid, the considerable challenge faced by the residential customers is the affordability of the time-varying bill after removing the buffer provided by the flat tariff, which has been investigated in Chapter 3 and 4. The DR programmes would be an efficient procedure to assist customers to save their energy bill. Additionally, in the findings report for this Irish customer behaviour trails [94], researchers noticed that the barriers to peak demand reduction are mainly due to the difficulty of linking the bill decrease with the behaviour change. Therefore, in this analysis, the degree of the unit price change before and after TOU intervention would be used to represent the responsiveness label of every customer, which is calculated as (6-1) shown below:

$$\beta_{r\_c} = \frac{\left(U_{vc} - U_{fc}\right) \times 100\%}{U_{fc}} \tag{6-1}$$

where the β_{r_c} represents the responsiveness of customer c. The U_{vc} indicates the unit price based on the bill calculated by the TOU tariff, and the U_{fc} represents the flat tariff before the DR programmes implemented. In this experiment, the U_{fc}equals to 14.1 pence per kWh, but the U_{vc} depends on the usage behaviour of each customer c which may result in different U_{vc} value. For each kind of TOU tariff intervention, the responsiveness can be categorized into three groups based on the β_{r_c}values: 1) The Benefit Group ($\beta_{r\_c} < -1\%$); 2) The Neutral Group ($-1\% < \beta_{r\_c} < 1\%$); 3) The Afflicted Group where $\beta_{r\_c} > 1\%$.

6.2.3 The Load Features and Socio-economic Features

In this chapter, the load features are employed to describe a customer more comprehensively with socio-economic features. Therefore, the input dataset for each customer should contain both 142 socio-economic features and household load information. However, the raw smart metering data are sampled at a granularity of one measurement every 30 minutes. The finer dataset for every customer incredible boost the dimensionality of the input data. There are a variety of methods for the dimensionality reduction, for instance the Principal Component Analysis (PCA) [105], feature selection [144, 145] and clustering algorithms [146, 147]. However, the load features are extracted to keep the interpretability of every feature, which is important for the post analysis. Hence, 56 load features are extracted to describe the consumption behavioural which are listed in Table 5-1.

In this experiment, the load features mainly used to depict the original usage characteristics before taken any DR incentives. Thereupon, the load features are generated based on the smart metering data in 2009.

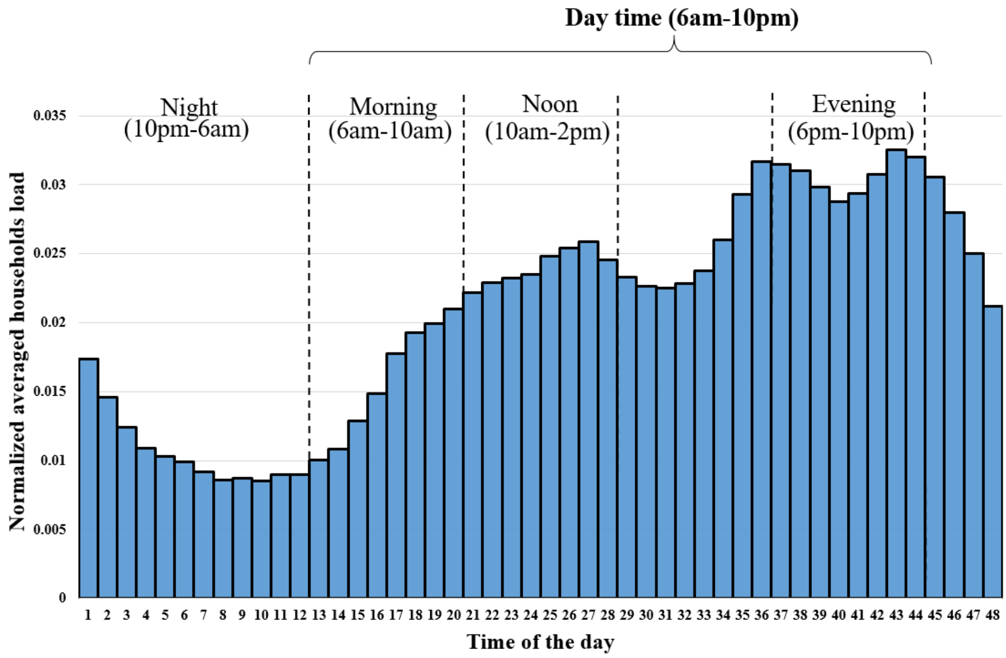


Figure 6- 3: The divided time period for consumption related load feature

6.3 Proposed Responsiveness Analysis Framework

In this chapter, an interaction-aware responsiveness pre-evaluating framework is proposed to identify the significant criteria for customers' responsiveness to different TOU tariffs. The framework analyses the interaction effect among households' intuitionistic and inherent features, such as the psychological, socio-economic and load features, to achieve the appropriate customer-targeting for different TOU tariff plans. The flowchart of the proposed framework is demonstrated in Figure 6-4., which consists of three steps: 1) calculating the customers' responsiveness; 2) the features pre-processing and the searching method; 3) Pre-evaluate new customers based on the significant criteria.

1) *Step 1: Calculating the customers' responsiveness*

The first step calculates the energy bill before and after applying the DR scheme to the customers. In this research, to exclude the seasonal impact on customers' consumption, the smart metering data in 2010 are chosen from the same date period in 2009. The responsiveness of each customer is defined as the bill difference before and after the DR stimulation, which can be calculated by (6-1).

1) In the second step, the significant load and socio-economic features are generated for every customer. Then, the high-dimensional interaction-aware KLAM search methodology is adopted to detect the critical features which influence customers' responsiveness significantly under a specific TOU tariff plan.

2) *Step 3: Pre-evaluation to accommodate new customers*

After the significant socio-economic and load characteristics being detected in the *Step 2*, it can be used to pre-evaluate the possible responsiveness for a new customer under a DR programme. In this way, based on fewer significant feature-combinations, new customers could be better accommodated to the suitable TOU tariff.

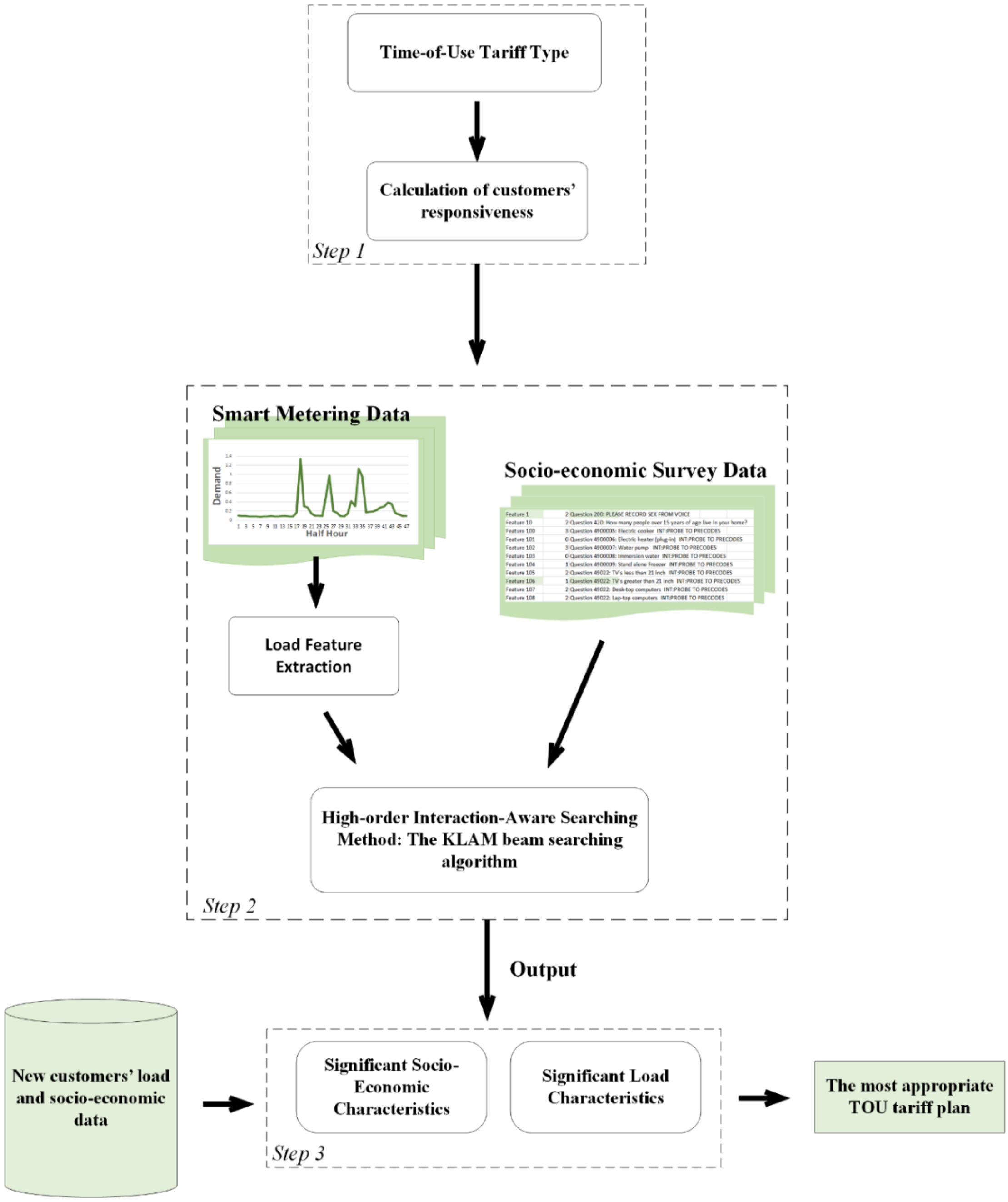


Figure 6- 4: The flowchart of the propose framework

6.4 Results and Discussion

In order to determine what kind of social groups would be benefited under the introduction of a specific TOU tariff, this research investigates the determinate characteristics (include both load and socio-economic characteristics) of the customers who have an effective response to a TOU tariff. The proposed framework is demonstrated against the real data from Ireland.

The number of features for every customer increase to 198, including 56 load features generated based on the usage data in July to September 2009 and 142 socio-economic features collected from the survey. After cleaning the features and responsiveness label for every customer, the total number of customers participated in every TOU tariff plan is shown in Table 6-1.

Table 6- 1: The Number of Participants for Different TOU Tariffs

	Tariff A	***Tariff B***	***Tariff C***	***Tariff D***
Number of participants	580	216	593	217

There are four types of TOU tariff applied to customers, which are drawn in Section 6.2.1. From Tariff A to Tariff D, the price gap between off-peak layer and peak time layer becomes bigger gradually, and the energy bill becomes more sensitive to the way of customers' usage behaviour change. The Figure 6-5 demonstrates the population proportion of three responsiveness groups for those four TOU tariffs. It can be noticed that, following the rise of price stimulate, more customers would benefit from the DR programme by responding to the tariff signals. The occupation of the customers in the benefit group changed from 23.8% to nearly 30.1%. However, meanwhile more customers changed from the neutral group to the afflicted group. The main reason is the high stimulation tariff will have a more severe penalty to those customers who didn't respond or inappropriate response. Hence, it is important to detect which load or socio-economic criteria or their combinations influence the customers' responsiveness under different TOU tariff plans.

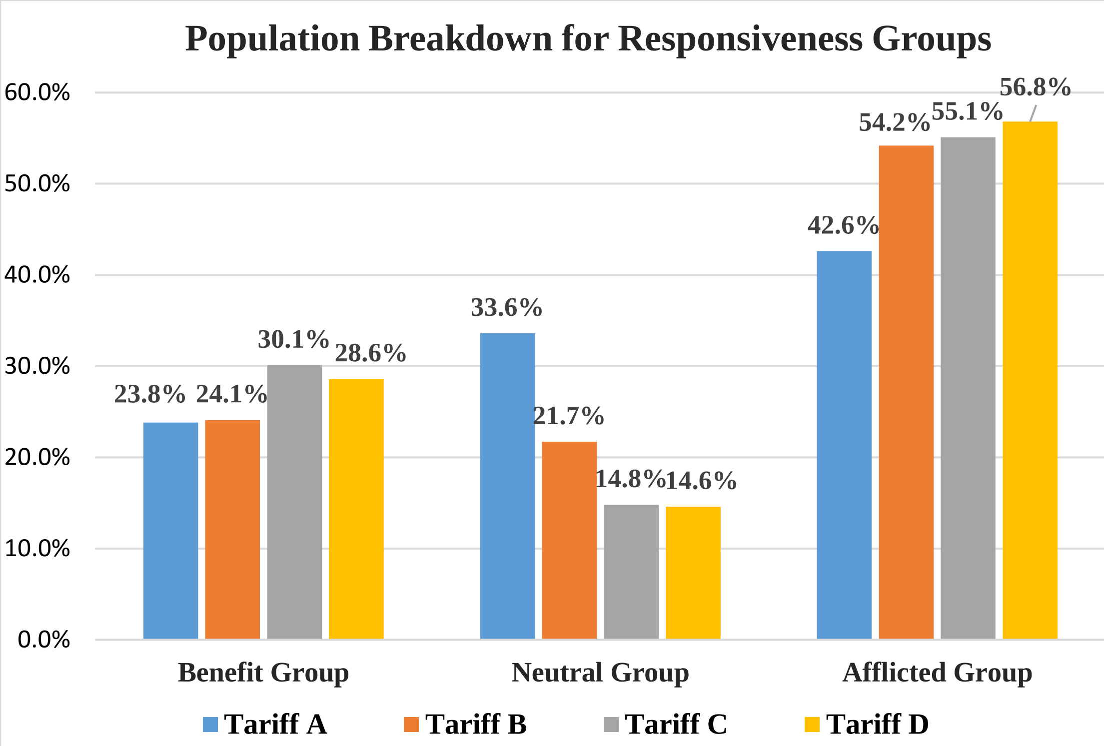


Figure 6- 5: Population breakdown for responsiveness groups

Thence, the KLAM beam searching method is applied to detect the significant load and socio-economic characteristics for different TOU tariffs. Due to the space limitation, this section concentrates on the significant features for all four TOU tariffs and the unique influential features for each TOU plans. A 95% confidence level is employed to all conducted tests.

6.4.1 The Common Significant Features for Different TOU Tariff Plans

In Table 6-2, the significant features, which have significant influence on customers' responsiveness under different tariff plan, are listed. Among them, Feature 16 illustrates its significant impact under all of the four TOU tariffs. Additionally, with the incentive of tariff growth, Feature 38 has the significant influence on customers' responsiveness only under Tariff C and Tariff D. The other four features have been detected in the one-way ANOVA for Tariff A, B, and C.

Besides the explanation of the features, the averaged feature level for the benefit-customer group under every Tariff plan are displayed in Table 6-2. For Feature 161 which is the only socio-economic feature, its average level number represents the using time of electric cooker.

The larger number the longer daily using time. It is obvious that the high incentive TOU tariff plans are more appropriate with the higher feature value households.

The larger values of Feature 10, 11 and 49 indicate that the maximum and minimum values of customers' load profile are closer to the mean demand, and the load profiles are smoother. Feature 38 proof the later noon peak customers are more suitable for high stimulate TOU price. Feature 38 describes the average time when the load peak happened in the business day. This feature only shows its significant impact on Tariff C and D. Furthermore, the more considerable value of Feature 16 and Feature 161 suggest customers with the higher consumption level of night and appliances may be more appropriate for high incentive tariffs.

Table 6- 2: The Common Significant Features for Different TOU Tariffs

Feature Number	*Content of the Feature*	*Tariff A*	*Tariff B*	*Tariff C*	*Tariff D*
16	Averaged Ratio: **Night / Day** consumption for the entire time period	**0.1913**	**0.3901**	**0.6111**	**0.5870**
10	Averaged Ratio: **Mean / Max** consumption for the entire time period	0.2015	0.1570	0.2547	
11	Averaged Ratio: **Min / Mean** consumption for the entire time period	**0.0998**	**0.1894**	**0.3084**	
49	Averaged Ratio: **Min / Mean** consumption for all weekends	**0.0717**	**0.1002**	**0.3314**	
161	How long do you use your **electric cooker** at daily level	**1.714**	**3**	**3**	
38	Average hour numbers of Peak time for all business day			**12:00 a.m.**	**12:37 a.m.**

Therefore, by horizontal comparison for the common features among different tariffs, the characteristics of the benefit customer groups for the TOU tariff who has larger price gap between the lowest and highest value can be summarized as:

- The customers with a smoother original load profile are more suitable.
- The customers whose larger part of demand are consumed during the night time period are more suitable.
- The customers who have a later noon load peak during the business day.
- The customers who utilise the electric cooker longer time are more suitable.

Then, the interacted feature-combinations for those common significant features have been investigated and displayed in Table 6-3. Due to the difference in the number of participated customers for those four tariff plans, some significant features could only do three-way interaction to ensure there are enough customers in every treatment. This research set a 95% confidence level to calculate the upper and lower bounds for the mean value of the customers' responsiveness level resulted in each interacted feature-combination. All the feature combinations shown in Table 6-3 at least have two treatments:

1) The upper bound of the treatment is less than zero. This guarantees that 95% of customers will be benefited by responding to the TOU tariff (unit price decreasing) after opening the trial to all population;

2) The lower bound of treatment is large than zero. Customers in this treatment have 95% confidence to experience a unit price growth in the DR programme (afflicted in the DR).

The characteristics extract for the benefit customer group are compared with the failure to response group under the same TOU plan. By this way, this research can obtain which socio-economic and load characteristics determine the success of the responding to a specific TOU tariff.

Table 6- 3: The significant interacting-feature combinations for common significant features

No.1		**Feature 16**	**Feature 20**	**Feature 162**	
	Tariff A	*Averaged Ratio: **Night / Day** consumption for the entire time period*	*Averaged **correlation coefficient** of current day and the previous day*	*How long do you use your plug-in electric heater at daily level*	
	Benefit group	**Smaller ratio value**	**Smaller Correlation value**	**Using shorter**	
		Feature 16	**Feature 98**	**Feature 168**	
	Tariff B	*Averaged Ratio: **Night / Day** consumption for the entire time period*	*Do you use plug-in heaters to heat your home*	*Frequency of using the Desk-top computers at daily level*	
	Benefit group	**Larger ratio value**	**Both do not use plug-in heaters**	**Using longer**	
		Feature 16	**Feature 79**	**Feature 86**	**Feature 51**
	Tariff C	*Averaged Ratio: **Night / Day** consumption for the entire time period*	*Have you already changed your way to use energy for bill reduction*	*Can you get the people live with you to reduce their electricity usage*	*Averaged Ratio: **Morning / Noon** consumption for all weekends*
	Benefit group	**Larger ratio value**	**More households haven't changed before trials**	**More households choose yes**	**Smaller ratio value**
		Feature 16	**Feature 14**	**Feature 84**	
	Tariff D	*Averaged Ratio: **Night / Day** consumption for the entire time period*	*Averaged Ratio: **Evening / Noon** consumption for the entire time period*	*Do you think it is inconvenient to reduce your usage*	
	Benefit group	**Larger ratio value**	**Smaller ratio value**	**More households think it is inconvenient**	
No.2		**Feature 10**	**Feature 15**	**Feature 162**	
	Tariff A	*Averaged Ratio: **Mean / Max** consumption for the entire time period*	*Averaged Ratio: **Noon / Total** consumption for the entire time period*	*How long do you use your plug-in electric heater at daily level*	

	Benefit group	**Larger ratio value**	**Larger ratio value**	**Using shorter**	
		Feature 10	**Feature 15**	**Feature 170**	
	Tariff B	*Averaged Ratio:* ***Mean / Max*** *consumption for the entire time period*	*Averaged Ratio:* ***Noon / Total*** *consumption for the entire time period*	*How long do you use your Game consoles at daily level*	
	Benefit group	**Smaller ratio value**	**Smaller ratio value**	**Using shorter**	
		Feature 10	**Feature 15**	**Feature 183**	**Feature 29**
	Tariff C	*Averaged Ratio:* ***Mean / Max*** *consumption for the entire time period*	*Averaged Ratio:* ***Noon / Total*** *consumption for the entire time period*	*Would you make major changes to the way you use electricity*	*The averaged minimum consumption in business day*
	Benefit group	**Larger ratio value**	**Smaller ratio value**	**More households choose yes**	**Larger minimum consumption**
		Feature 11	**Feature 15**	**Feature 97**	**Feature 183**
	Tariff A	*Averaged Ratio:* ***Min / Mean*** *consumption for the entire time period*	*Averaged Ratio:* ***Noon / Total*** *consumption for the entire time period*	*Do you use central storage heating to heat your home*	*Would you make major changes to the way you use electricity*
	Benefit group	**Smaller ratio value**	**Smaller ratio value**	**Less customers use central heating**	**More households choose yes**
No.3		**Feature 11**	**Feature 15**	**Feature 69**	
	Tariff B	*Averaged Ratio:* ***Min / Mean*** *consumption for the entire time period*	*Averaged Ratio:* ***Noon / Total*** *consumption for the entire time period*	*How many children under 15 lived in your house*	
	Benefit group	**Larger ratio value**	**Smaller ratio value**	**Lived with more children**	
		Feature 11	**Feature 15**	**Feature 107**	**Feature 149**

	Tariff C	*Averaged Ratio: **Min / Mean** consumption for the entire time period*	*Averaged Ratio: **Noon / Total** consumption for the entire time period*	*Do you use electric instantaneous heater to heat water*	*Do you have Lap-top Computers*
	Benefit group	**Larger ratio value**	**Larger ratio value**	**More households use**	**More households do not have**
No.4		**Feature 49**	**Feature 15**	**Feature 157**	**Feature 188**
	Tariff A	*Averaged Ratio: **Min / Mean** consumption for all weekends*	*Averaged Ratio: **Noon / Total** consumption for the entire time period*	*How long do you use your Tumble dryer at daily level*	*As part of the trial, how much do you think your bill would decrease*
	Benefit group	**Smaller ratio value**	**Larger ratio value**	**Using shorter**	**More households think the bill would reduce less**
		Feature 49	**Feature 15**	**Feature 69**	
	Tariff B	*Averaged Ratio: **Min / Mean** consumption for all weekends*	*Averaged Ratio: **Noon / Total** consumption for the entire time period*	*How many children under 15 lived in your house*	
	Benefit group	**Smaller ratio value**	**Smaller ratio value**	**Lived with more children**	
		Feature 49	**Feature 15**	**Feature 69**	**Feature 96**
	Tariff C	*Averaged Ratio: **Min / Mean** consumption for all weekends*	*Averaged Ratio: **Noon / Total** consumption for the entire time period*	*How many children under 15 lived in your house*	*How many bedrooms in your home*
	Benefit group	**Larger ratio value**	**Larger ratio value**	**Lived with more children**	**Have more bedrooms**
No.5		**Feature 161**	**Feature 35**	**Feature 90**	**Feature 159**
	Tariff A	*How long do you use your electric cooker at daily level*	*Averaged Ratio: **Noon / Total** consumption for all weekends*	*how much do you think you can reduce your usage*	*How long do you use your electric instant shower at daily level*

	Benefit group	**Using shorter**	**Smaller ratio value**	**More households think they would reduce less**	**Using shorter**
		Feature 161	**Feature 35**	**Feature 194**	
	Tariff B	*How long do you use your electric cooker at daily level*	*Averaged Ratio: **Noon** / **Total** consumption for all weekends*	*Do you satisfy with the competition among energy suppliers*	
	Benefit group	**Using longer**	**Smaller ratio value**	**More households unsatisfied with it**	
		Feature 161	**Feature 35**	**Feature 65**	**Feature 195**
	Tariff C	*How long do you use your electric cooker at daily level*	*Averaged Ratio: **Noon** / **Total** consumption for all weekends*	*Do people lived with you use internet regularly*	*Do you satisfy with the renewable generation percentage*
	Benefit group	**Using longer**	**Larger ratio value**	**More households choose No**	**More households unsatisfied with it**
No.6		**Feature 38**	**Feature 35**	**Feature 24**	**Feature 88**
	Tariff C	*Average hour numbers of Peak time for all business day*	*Averaged Ratio: **Noon** / **Total** consumption for all weekends*	*Average evening consumption for all business day*	*Do you want to be told how much electricity you can use*
	Benefit group	**Smaller**	**Larger ratio value**	**Smaller evening consumption value**	**More households do not want to be told**
		Feature 38	**Feature 35**	**Feature 145**	
	Tariff D	*Average hour numbers of Peak time for all business day*	*Averaged Ratio: **Noon** / **Total** consumption for all weekends*	*Number of Immersion*	
	Benefit group	**Averaged peak time is shorter**	**Smaller ratio value**	**Have less**	

From Table 6-3, it is evident that Feature 15 and 35 have a significant interaction effect with the common significant features (except Feature 16). Feature 15 and 35 are the load features related to the ratio between the noon consumption and total demand across the entire time and all weekends respectively. However, the peculiarity of the Feature 15 and 35 for the benefit groups is not consistent under the same TOU tariff, for example, the Tariff A. This may be the main reason for why those two features did not demonstrate their outstanding influence individually. Those two features could only play a driving role by interacting with other specific features.

By analysis the KLAM results of every common significant feature, it is can be summarized from Table 6-3 that:

1) Tariff A:

- The Tariff A can benefit the customers whose load profiles are not relatively smooth. This is due to the benefit group has a relatively larger gap between maximum and mean consumption.
- Tariff A is more suitable for households with relatively smaller consumption level. Due to the using time of plug-in heater, tumble dryer and electric instant shower for the benefit group are shorter at the daily level.
- The psychological characteristic for customers who successfully respond to the Tariff A is contradictory. Feature 188 and 90 both demonstrate that the customer didn't believe they can reduce their usage and bill.

2) Tariff B:

- Tariff B can benefit the customers who have relatively small noon consumption
- Households who can benefit by responding to the Tariff B have more children under 15 years old.

3) Tariff C:

- The customers with smoother load profiles (minimum and maximum load values are both closer to the mean demand) show significant superiority under TOU Tariff C.
- Tariff C prefer the customers with relatively higher consumption level. Due to the benefited customer group reveals high minimum consumption and large noon demand.

The number of children in a household is also bigger. However, the evening consumption of the profited group is relatively small.

4) Tariff D:

- The customers benefited under this Tariff D have more flatten load profiles. The consumption during the night period is large. However, the load during noon and evening are relatively smaller.
- More customers in the benefited group consider energy usage reduction as inconvenient.

6.4.2 The Unique Significant Features for Different TOU Tariff Plans

To analysis what kind of customers are more appropriate to a specific TOU tariff plan in a more comprehensively way, this Section analysis the unique significant features and their interacting feature-combinations for each tariff. Table 6-4 to Table 6-7 display the unique features and its interacted feature-combinations which demonstrated significant impact on customers' responsiveness for each TOU tariff.

Table 6- 4: The unique significant features for Tariff A

		Feature 13	**Feature 20**	**Feature 177**	**Feature 182**
		Averaged Ratio: ***Morning / Noon*** *consumption for the entire time period*	*Averaged* ***correlation coefficient*** *of current day and the previous day*	*Are the external walls of your home insulated*	*Would you make minor changes to the way you use electricity*
	Benefit group	**Smaller ratio value**	**Smaller Correlation value**	**More households have**	**Most households will**
		Feature 23	**Feature 21**	**Feature 97**	**Feature 67**
		Average consumption of business day during the daytime period	*Averaged Ratio:* ***Business Day / Weekends*** *consumption for the entire time period*	*Do you use central storage heating to heat your home*	*How many people (>15) live in your home*
	Benefit group	**Smaller daytime consumption**	**Smaller ratio value**	**More households use**	**Live with less people**
Tariff A		**Feature 72**	**Feature 53**	**Feature 173**	**Feature 71**
		Do you interested in changing the way you use electricity if it helps environment	*Averaged Ratio:* ***Noon / Total*** *consumption for all weekends*	*The approximate proportion of energy saving light bulbs in your home*	*Do you interested in changing the way you use electricity if it reduces the bill*
	Benefit group	**More people choose yes**	**Smaller ratio value**	**Larger proportion**	**More people choose yes**
		Feature 176	**Feature 20**	**Feature 196**	**Feature 73**
		Is your attic insulated, if so when was the insulation fitted	*Averaged* ***correlation coefficient*** *of current day and the previous day*	*Are you satisfied with the overall cost of electricity (Before trials)*	*Can you reduce bill by changing the way the people you live with use electricity*
	Benefit group	**More people do not have an insulated attic**	**Smaller Correlation value**	**More satisfied with the bill**	**Less households can do this**

Table 6- 5: The unique significant features for Tariff B

		Feature 85	**Feature 52**	**Feature 93**
		Do you know enough about the appliances' consumption in order to reduce bill	*Averaged Ratio: **Evening / Noon** consumption for all weekends*	*The age of your house*
Tariff B	Benefit group	**Know less**	**Larger ratio value**	**Elder houses**
		Feature 138	**Feature 53**	**Feature 160**
		How many dishwashers do you own	*Averaged Ratio: **Noon / Total** consumption for all weekends*	*Frequency of using the electric shower at daily level*
	Benefit group	**Both own 1 dishwasher**	**Smaller ratio value**	**Using longer**

Table 6- 6: The unique significant features for Tariff C

Tariff C		**Feature 68**	**Feature 15**	**Feature 27**	**Feature 99**
		How many people (>15) stay at home during the day	*Averaged Ratio:* ***Noon / Total*** *consumption for the entire time period*	*Average consumption of business day noon period*	*Do you use gas to heat your home*
	Benefit group	**Less people in**	**Larger ratio value**	**Smaller consumption value**	**More households choose No**
		Feature 91	**Feature 15**	**Feature 93**	**Feature 199**
		The house type	*Averaged Ratio:* ***Noon / Total*** *consumption for the entire time period*	*How old is your house*	*Do you agree with the environmental damage related to the electricity usage*
	Benefit group	**Bungalow or Semi-detached house**	**Smaller ratio value**	**Elder house**	**More households agree with it**
		Feature 122	**Feature 35**	**Feature 164**	**Feature 37**
		Have you ever had to go without heat on a cold day	*Averaged Ratio:* ***Noon / Total*** *consumption for all weekends*	*Frequency of using the immersion to heat water at daily level*	*Hour numbers when consumption is over the averaged value*
	Benefit group	**Never**	**Smaller ratio value**	**Using shorter**	**Longer time over averaged value**
		Feature 189	**Feature 15**	**Feature 164**	**Feature 60**
		The Education Level of the CIE	*Averaged Ratio:* ***Noon / Total*** *consumption for the entire time period*	*Frequency of using the immersion to heat water at daily level*	*The employment status of the CIE*
	Benefit group	**Lower education level**	**Smaller ratio value**	**Using longer**	**Unemployed or self-employed**
		Feature 190	**Feature 15**	**Feature 61**	**Feature 85**
		The income level of your household	*Averaged Ratio:* ***Noon / Total*** *consumption for the entire time period*	*The social class of the CIE*	*Do you know enough about the appliances' consumption in order to reduce bill*
	Benefit group	**Higher income level**	**Smaller ratio value**	**More AB, $C1^*$ grade**	**Know more**

***: The Social Grade AB represents customers who work on Higher and intermediate managerial, administrative, professional occupations; C1 represents customers who work on Supervisory, clerical & junior managerial, administrative, professional occupations**

Table 6- 7: The unique significant features for Tariff D

Tariff D		**Feature 92**	**Feature 53**	**Feature 195**
		Do you own or rent your home	*Averaged Ratio: **Noon / Total** consumption for all weekends*	*Do you satisfy with the renewable generation percentage*
	Benefit group	**Have some households rent house from a local authority**	**Smaller ratio value**	**More households unsatisfied with it**
		Feature 170	**Feature 35**	**Feature 188**
		How long do you use your Game consoles at daily level	*Averaged Ratio: **Noon / Total** consumption for all weekends*	*If you believe your bill will decrease after the trial, how much do you think it could decrease?*
	Benefit group	**Using shorter**	**Smaller ratio value**	**They believe they can reduce less of the bill**

Through analysis the unique significant feature-combinations, the load and socio-economic characteristics of benefited customers under each TOU tariff can be summarised below:

1) For Tariff A:

- Do not require regular load profile for the customer: due to the low correlation coefficient (Feature 20) value for the benefited customer group.
- The consumption level for the benefit customers is comparatively low and there are fewer people (>15) living together.
- The benefited customers under TOU Tariff A have relatively stronger energy-saving awareness (Feature 71, 182 and 72) and equip more energy-saving device, such as the external walls, central storage heating and energy-saving light bulbs.

2) For Tariff B:

- The evening consumption is comparatively higher and the demand during the noon period is lower for the benefited customer group.
- The benefited customers have less energy-saving awareness and less knowledge of the consumption difference between appliances (Feature 85). Meanwhile, those customers utilise the electric shower for a longer time each day.
- The houses are elder for the profited customers who were accommodated in Tariff B.

3) For Tariff C:

In Table 6-6, the KLAM results for the unique significant features demonstrate a contrary conclusion summarized in Section 6.4.1. In Table 6-3, the larger total consumption level and noon consumption are important characteristics for the benefit group. However, there are some specific conditions for low consumption level or low noon demand households, which makes them also can effectively response to the Tariff and reduce their energy bills.

- Relatively lower education level customers who are unemployed or self-employed can be benefited by responding to Tariff C with a small noon consumption.
- The income level for a household also become significant. The higher income level household who also know relatively more about the energy saving also can be profited under Tariff C with a small noon consumption.
- Housing type and age for customers also become an important feature related to the effective responsiveness for Tariff C. Customers live in the bungalow or semi-detached elder houses, have energy-saving awareness and with lower noon consumption are more likely to be benefited under Tariff C.

4) For Tariff D:

- Most of the customers who successfully response to Tariff D rent house from a local authority instead of owning a house.
- Customers in the benefited group use fewer electric appliances, such as the game consoles.

6.5 Chapter Summary

This chapter proposes a framework to pre-evaluate the customers' responsiveness based on their socio-economic data and load characteristics for different Tariff plans. By considering the interaction effect among features, the significant socio-economic criteria and load characteristics of the effective response group can be detected. Those detected significant features can help various DR schemes to target the most appropriate customers quickly.

The framework is validated on a case study where four different TOU tariffs are provided. From Tariff A to Tariff D, the price stimulation becomes greater gradually. Following the growth of the price stimulation, both the proportions of customers who are benefited and the customers who are afflicted in the DR programme are increasing. The case study resulted that the smoother load profile, higher consumption level, larger proportion of the night demand and more children in the family are the significant characteristics of the successfully respond customers under high price-stimulation TOU tariffs. Additionally, the opposite characteristics of the same feature could have a consistent impact on customers' responsiveness by interacting with different features.

The detailed significant load and socio-economic characteristics of the benefited customers under each TOU tariff are concluded as below:

1）For Tariff A:

The consumption level of the benefited customers is relatively lower. Those customers live with fewer people in their household. The use of appliances such as the plug-in heater, electric instant shower and tumble is comparatively small. Furthermore, the daily load profiles for those customers are not smooth (larger difference between daily max and min demand) and regular (lower correlation value of the consumption of two adjacent days).

The environmental awareness of the benefited customers is stronger and equipped with more energy-saving devices, for instance, the external walls and energy-saving bulbs. However, those customers demonstrated less confidence in how much energy they can reduce in the DR scheme than the afflicted customers.

2）For Tariff B:

Tariff B is suitable for the customers who have higher evening consumption but smaller noon demand.

The age of houses is elder for customers who are benefited under Tariff B. Their households have more children (under 15 years old). However, the benefited customers have less energy-saving awareness.

3）For Tariff C:

The customers with lower education level, who are unemployed or self-employed with a smaller noon consumption are appropriate to the Tariff C.

On the contrary, customers with larger noon consumption but know more about the energy-saving knowledge, higher income level customers also can benefit in Tariff

C. Customers who live with more children have larger noon consumption but lower evening consumption also appropriate to Tariff C.

The customers who live in elder bungalow or semi-detached house with stronger energy-saving awareness are more benefited under Tariff C.

In Tariff C, the customers with smoother load profiles are benefited significantly.

4）For Tariff D:

The load profiles are more flatten for the customers who successfully response to Tariff D. Meanwhile, the night consumption is high, but the noon and evening energy consumption is small.

Tariff D is more suitable for the customers who rent a house and with fewer electric appliances, such as the game consoles.

Most of the customers in the benefit group for Tariff D think to reduce energy usage is inconvenient. This may not prove they have weak energy-saving awareness because the consumption level for those customers is relatively lower. Therefore, energy reduction could be more difficult for them.

Chapter 7

Conclusions

THIS chapter demonstrates the conclusions of the thesis by outlining the significant findings and limitations.

7.1 Conclusions

Nowadays, the increasing of the LCTs in the power system not only introduces an attractive opportunity to transform to a greener energy system but also brings tremendous uncertainty and challenges to the energy suppliers and regulators. Under this circumstance, the smart meters are widely installed to provide the fine-grained usage data for the individual customers, which help suppliers to understand their customers better. On the other hand, the Half-Hourly Settlement reformation is applied to remove the cross-subsidies among customers and bring the market signals to encourage customers to modify their usage behaviour optimally.

In the dynamic smarter energy market, it requires the suppliers and regulators have a deeper understanding in their customers' usage behaviour, not only from the smart metering data but also from the socio-economic data of each household.

This thesis aims to develop a comprehensive investigation of the effect of socio-economic data. Two research explores the effect of high-dimensional interacting socio-economic data on customers' wholesale market cost and distribution network cost respectively. Additionally, the effect of socio-economic data has been validated in two demand-side applications which are a cost-reflective customer classification model and a responsiveness re-evaluation framework to different TOU tariffs. The two conclusions for the thesis could be summarized as follow:

i) Firstly, the influence of the interacted socio-economic factors on customers' bill change has been proved as significant. The ownership and utilisation of electrical appliances have an influence on both wholesale market cost and network cost. The energy-saving awareness factors impact more on the network bill.

ii) Secondly, the socio-economic data facilitates cost-reflective customers' classification. The classification accuracy has been improved when socio-economic data collaboration with the load features. Even when the load features are inaccessible, the socio-economic data could estimate customers' cost level as a remedy. Additionally, the customers' responsiveness to different TOU tariffs can be pre-evaluated by identifying the significant socio-economic criteria and load characteristic. The case study resulted that the smoother load profile, higher consumption level, larger proportion of the night demand and more children in the family are the significant characteristics of the successfully respond customers under high price-stimulation TOU tariffs. With the significant criteria, the assignment of customers to different TOU plans could be more appropriate.

The more detailed conclusions of significant socio-economic data effects carried out from the investigations in this thesis are shown in the following sub-chapters.

7.2 The Socio-Economic Criteria for the Wholesale Market Cost Variation

A high-dimensional interaction-aware search methodology has been proposed, which called the KLAM beam search algorithm, to detect the socio-economic factor-combinations which have a significant effect on customers' wholesale market cost variation in the HHS process. With the novel KLAM methodology, there are some key findings which are contrary to previous studies:

i) The electricity appliance factors and dwelling related factors could have negative effect on customers' energy bill changing by interacting with other specific factors. However, those two kinds of factors are commonly found have positive effect on energy consumption;
ii) The CIE's employ status, their age and environmental awareness can influence the bill variation significantly in specific factor-interacting-combinations. However, those factors are rarely caught the attention in the previous literature, which is because the effect of the single of them is inconclusive (mixing effect) or even no significant effect.

By analysis the significant interacting socio-economic factor-combination, it can be found that in the new HHS process, the elder customers who owning more electric appliances or who living in an older house are most likely to be the new vulnerable customers who may need help from the government.

7.3 The Socio-Economic Criteria for the Network Cost Variation

In order to remove the cross-subsidies of network cost among customers, a novel Unit Home Equivalent pricing method has been proposed to calculate the accurate network cost based on their smart metering data. Comparing to the current energy-based pricing, the proposed method has two fundamental breakthroughs:

i) Forward-looking signal: instead of only considering customers' contribution to historical peaks, the proposed method estimates the possibility of future peaks created by different customers at different time points.

ii) Behavioural incentives: the proposed method encourages not only new customers to under-utilised locations but also existing customers to change energy usage behaviours according to the network's headroom profile.

After moving to the precise network cost, the socio-economic characteristics for the high network cost customers have been detected by the KLAM beam search method. The significant interacted socio-economic characteristics are:

- The awareness of energy-saving is weak for the high network cost customers. They are willing to do more to reduce their bill but not for the environment. Less of them do energy reduction activities in reality or use renewable energy.

- People in the high network cost group live with a large family. More people are retired or unemployed with lower income.

- The houses for the high network cost customer group are relatively newer and smaller than the lower network cost group. Most high-cost customers rent a house and always feeling not warm enough due to the poor insulation of the house.

- The television, electric heater and the washing machine are the significant electricity appliances for the high network cost customer group, which applies the positive effect on the network bill growth. On the opposite side, the Stand-alone freezer, electric shower and tumble dryer have the negative impact on the network cost.

7.4 The Impact of Socio-Economic Features on Cost-Reflective Customer Classification

In this chapter, the socio-economic data cooperate with the smart metering data in a customer classification application.

In the fully competitive retail market, identification of the customers' energy cost level is crucial for suppliers, especially small suppliers. However, due to the challenge of the availability issues of the historical smart metering data for new switched-in customers, the load profile-based approaches for the classification face huge challenge.

Hence, a cost-reflective customer classification framework has been proposed where the socio-economic information can remedy the problem caused by the absent of smart metering data. The framework consists of three models to cope with different scenarios of available data: Scenario 1) customers who only provide smart metering data; Scenario 2) customers who only provide the socio-economic information and Scenario 3) customers providing both two types of data. The proposed framework has three superiorities:

i) The framework can select the interpretable features;

ii) The framework builds several classification models to cope with different scenarios of input data;

iii) The framework improves the accuracy by 16.40%, 0.60% and 12.40%for each scenario of input data.

Based on a case study of Irish smart metering data, the proposed cost-reflective classification framework reveals key findings which were not discovered by the traditional load-profile classification methods. They are summarised as follows:

- The ratio between the consumption during the morning and the noon has a negative influence on the supply cost level.

- The ratio between the consumption during the night and the whole day has a negative influence on the supply cost level;

- The dispersion degree (variance) of the whole consumption has a positive influence on the supply cost level;

- The electricity appliances (except the lap-top computer) have a positive influence on the supply cost level;

- The energy saving consciousness has inconsistent effects on customers' actual supply cost level.

7.5 The Impact of Socio-Economic Features on the Responsiveness to Different Tariff Plans

In this chapter, the socio-economic data cooperates with the smart metering data and the different TOU tariff data to establish a framework to identify the significant criteria for customers' responsiveness under different tariff plans. There are four kinds of TOU tariffs in the investigation. From Tariff A to Tariff D, the difference between the peak price and the minimum price for each TOU tariff becomes greater gradually.

With the increase of the price simulation, the load and socio-economic characteristics for customers who can save their energy bill by responding to the TOU tariff have been identified through the proposed framework:

i) For Tariff A:

- Do not require regular load profile for the customer: due to the low correlation coefficient value for the benefited customer group.
- The consumption level for the benefit customers is comparatively low and there are fewer people (>15) living together.
- The benefited customers under TOU Tariff A have relatively stronger energy-saving awareness and equip more energy-saving device, such as the external walls, central storage heating and energy-saving light bulbs.

- The Tariff A can benefit the customers whose load profiles are not relatively smooth. This is due to the benefit group has a relatively larger gap between maximum and mean consumption.
- Tariff A is more suitable for households with relatively smaller consumption level. Due to the using time of plug-in heater, tumble dryer and electric instant shower for the benefit group are shorter at the daily level.
- The psychological characteristic for customers who successfully respond to the Tariff A is contradictory. Two features demonstrate that the customer didn't believe they can reduce their usage and bill.

ii) For Tariff B:

- The evening consumption is comparatively higher and the demand during the noon period is lower for the benefited customer group.
- The benefited customers have less energy-saving awareness and less knowledge of the consumption difference between appliances. Meanwhile, those customers utilise the electric shower for a longer time each day.
- The houses are elder for the profited customers who were accommodated in Tariff B.
- Tariff B can benefit the customers who have relatively small noon consumption
- Households who can benefit by responding to the Tariff B have more children under 15 years old.

iii) For Tariff C:

- Relatively lower education level customers who are unemployed or self-employed can be benefited by responding to Tariff C with a small noon consumption.
- The income level for a household also become significant. The higher income level household who also know relatively more about the energy saving also can be profited under Tariff C with a small noon consumption.
- Housing type and age for customers also become an important feature related to the effective responsiveness for Tariff C. Customers live in the bungalow or

semi-detached elder houses, have energy-saving awareness and with lower noon consumption are more likely to be benefited under Tariff C.

- The customers with smoother load profiles (minimum and maximum load values are both closer to the mean demand) show significant superiority under TOU Tariff C.
- Tariff C prefer the customers with relatively higher consumption level. Due to the benefited customer group reveals high minimum consumption and large noon demand. The number of children in a household is also bigger. However, the evening consumption of the profited group is relatively small.

iv) For Tariff D:

- Most of the customers who successfully response to Tariff D rent house from a local authority instead of owning a house.
- Customers in the benefited group use fewer electric appliances, such as the game consoles.
- The customers benefited under this Tariff D have more flatten load profiles. The consumption during the night period is large. However, the load during noon and evening are relatively smaller.
- More customers in the benefited group consider energy usage reduction as inconvenient.

7.6 The Limitations of the Research

Although the effect of socio-economic data on customers' bill variation and applications' performances has been validated, there still are two main limitations of this work.

The first limitation is the high requirements for the dataset. This thesis investigates the relationship between socio-economic information and other data sources (such as the smart metering usage data, the responsiveness data). Therefore, it requires that the dataset must collect both smart metering data and socio-economic information for the same households. Due to the high cost of data collection, the datasets seldom recorded the smart metering data and socio-economic information at the same time. The strict requirement blocks massive other datasets to do the validation research for the conclusions resulted in this thesis. The four studies in this thesis are all based on the same Irish dataset, which was collected in the electricity smart metering customer behaviour trails. Although the confidence level has been set as 95% to guarantee the generalization of the findings in the research, the particularity of the conclusions still might exist.

The second limitation is the conciseness of the conclusions resulted from the high-dimensional interacting-aware methodology. Although there are several advantages for exploring the effect of multi interacting socio-economic factors (such as stronger effect, more comprehensive description of the socio-economic status), the complexity of the resulted socio-economic factors makes it difficult to be summarised and spread. To solve this limitation, cooperation with the researcher from social science could be helpful.

Chapter 8

Future Works

T**HIS** chapter draws the future works and potential research topic related to the socio-economic data.

8.1 Future Works

The effect of the socio-economic data has already been proven in this thesis. The proposed interaction-aware KLAM beam searching method allows many applications taking the effect of customers' socio-economic status into consideration. Some potential research topics in the future works can be discussed.

8.1.1 Development of Tariff Design based on Customers' Flexibility

In this thesis, the impact of socio-economic data on customers' responsiveness to different TOU tariff has already been investigated. The responsiveness of customers in different socio-economic status can be estimated.

In the UK, the change in the wholesale market price is mainly due to the increasing demand. During the peak demand period, the more expensive generator will be dispatched to generate electricity to meet the demand. In this condition, the use of different electrical appliances during the system peak time should be charged with different tariff. For example, if there are two households, one of them owns an electric vehicle which will be charged during the system peak period. The other household with little flexibility and maintains its basic energy consumption. Obviously, this two categorize customers need to be stimulated with different TOU tariff. A greater stimulation is needed to encourage the EV users to avoid charging their car during the peak hour. On the contrary side, for the vulnerable customers with little or none flexible demand, applying the high stimulation TOU tariff on them would cause the socio issues

Therefore, the TOU tariff should be designed based on customers' socio-economic characteristics. Basically, it should be designed based on the ability of customers' responsiveness.

8.1.2 Deeper Investigation between Socio-Economic Data and Customers' Usage Behaviour

The usage behaviour of an individual customer is potentially decided by the customers' interacted socio-economic data. By investigating the deeper causality between the customers' socio-economic data and their energy usage behaviour, there are many applications could be better achieved, such as the 1) extraction of more precise typical load profiles for residential customers; 2) load forecasting for individual household; 3) facility establishment of the Peer to Peer renewable energy retail market among customers.

Appendix A

Irish Smart Meter Trial Socio-economic Information Questionnaire

QUESTION 200

PLEASE RECORD SEX FROM VOICE

1 Male
2 Female

QUESTION 300

May I ask what age you were on your last birthday?
INT: IF NECCESSARY, PROMPT WITH AGE BANDS

1 18 - 25
2 26 - 35
3 36 - 45
4 46 - 55
5 56 - 65
6 65+
7 Refused

QUESTION 310

What is the employment status of the chief income earner in your household, is he/she

1 An employee
2 Self-employed (with employees)
3 Self-employed (with no employees)
4 Unemployed (actively seeking work)
5 Unemployed (not actively seeking work)
6 Retired
7 Carer: Looking after relative family

QUESTION 400 **OPEN**

IF [Q310 , 1 , 2 , 3]
SAVE IN CLASS

What is the occupation of the chief income earner in your household?

QUESTION 401

SOCIAL CLASS
Interviewer, Respondent said that occupation of chief income earner was....

<Question 400>
Please code

1 AB
2 C1
3 C2
4 DE
5 F [RECORD ALL FARMERS]
6 Refused

QUESTION 405
Do you have internet access in your home?

1 Yes
2 No

QUESTION 406
IF [Q405 , 1]
Do you have broadband in your home?

1 Yes
2 No

QUESTION 407
Do you use the internet regularly yourself?
1 Yes
2 No

QUESTION 408
Are there other people in your household that use the internet regularly?

1 Yes
2 No

QUESTION 410
What best describes the people you live with?
READ OUT

1 I live alone
2 All people in my home are over 15 years of age
3 Both adults and children under 15 years of age live in my home

QUESTION 420

IF [Q410 , 2 , 3]

How many people over 15 years of age live in your home?

1	1
2	2
3	3
4	4
5	5
6	6
7	7 or more

QUESTION 430

IF [Q410 , 2 , 3]

And how many of these are typically in the house during the day (for example for 5-6 hours during the day)

1	1
2	2
3	3
4	4
5	5
6	6
7	7 or more
8	None

QUESTION 43111

IF [Q410 , 3]

How many people under 15 years of age live in your home?

1	1
2	2
3	3
4	4
5	5
6	6
7	7 or more

QUESTION 4312

IF [Q410 , 3]

And how many of these are typically in the house during the day (for exanmple for 5-6 hours during the day)

1	1
2	2
3	3
4	4
5	5
6	6
7	7 or more
8	None

QUESTION 431

And now, I would like to ask you a few questions about your general attitudes towards energy, electricity use and the
electricity bill.
Please rate each of the following statements on a scale of 1 to 5 where 1 is strongly agree and 5 is strongly disagree

QUESTION 4311

DUMMY QUESTION

3	I we am are interested in changing the way I we use electricity if it reduces the bill
4	I we am are interested in changing the way I we use electricity if it helps the environment
5	I we can reduce my electricity bill by changing the way the people I we live with use electricity

PUT IN state Q4311,1

QUESTION 4331

PUT IN state Q4311,2

QUESTION 4331

PUT IN state Q4311,3

QUESTION 4331

I we am are interested in changing the way I we use electricity if it reduces the bill

1	1 - strongly agree
2	2
3	3
4	4
5	5 - strongly disagree

PUT IN state Q4311,4

QUESTION 4331

I we am are interested in changing the way I we use electricity if it helps the environment

1	1 - strongly agree
2	2
3	3
4	4
5	5 - strongly disagree

PUT IN state Q4311,5

QUESTION 4331

I we can reduce my electricity bill by changing the way the people I we live with use electricity

1	1 - strongly agree
2	2
3	3
4	4
5	5 - strongly disagree

QUESTION 432

And now, I would like to ask you a few questions about your own efforts to date to reduce your electricity usage in your
household.
Please rate each of the following statements on a scale of 1 to 5 where 1 is strongly agree and 5 is strongly disagree

QUESTION 4321 **MULTIPLE**

DUMMY QUESTION

1	I we have already done a lot to reduce the amount of electricity I we use
2	I we have already made changes to the way I we live my life in order to reduce the amount of electricity we use.
3	I we would like to do more to reduce electricity usage
4	I we know what I we need to do in order to reduce electricity usage

PUT IN state1 Q4321,1

QUESTION 4332

I we have already done a lot to reduce the amount of electricity I we use

1 1 - strongly agree
ADD TO Q4321 [1]
2 2
ADD TO Q4321 [1]
3 3
4 4
5 5 - strongly disagree

PUT IN state1 Q4321,2

QUESTION 4332

I we have already made changes to the way I we live my life in order to reduce the amount of electricity we use.

1 1 - strongly agree
ADD TO Q4321 [2]
2 2
ADD TO Q4321 [2]
3 3
4 4
5 5 - strongly disagree

PUT IN state1 Q4321,3

QUESTION 4332

I we would like to do more to reduce electricity usage

1 1 - strongly agree
ADD TO Q4321 [3]
2 2
ADD TO Q4321 [3]
3 3
4 4
5 5 - strongly disagree

PUT IN state1 Q4321,4

QUESTION 4332

I we know what I we need to do in order to reduce electricity usage

1 1 - strongly agree
ADD TO Q4321 [4]
2 2
ADD TO Q4321 [4]
3 3
4 4
5 5 - strongly disagree

PUT IN state1 Q4321,5

QUESTION 4332

QUESTION 433

IF [Q4321 , 1 , 2]

Thinking about the energy reduction activities undertaken by you or your family/household, in the last year, did your
efforts reduce your bills?

1 Yes
2 No
3 Don't know

QUESTION 434 **NUMBER**

IF [Q433 , 1]

Approximately what % savings on average did you achieve on the average bill?

QUESTION 435

IF [Q4321 , 3 , 4 , 5 \ # Q4321 , 1 , 2]

Please rate each of the following statements on a scale of 1 to 5 where 1 is strongly agree and 5 is strongly disagree

QUESTION 4351

DUMMY QUESTION

1 It is too inconvenient to reduce our usage of electricity
2 I do not know enough about how much electricity different appliances use in order to reduce my usage
3 I am not be able to get the people I live with to reduce their electricity usage
4 I do not have enough time to reduce my electricity usage
5 I do not want to be told how much electricity I can use
6 Reducing my usage would not make enough of a difference to my bill

PUT IN state2 Q4351,1

QUESTION 4352

IF [Q4321 , 3 , 4 , 5]

It is too inconvenient to reduce our usage of electricity

1 1 - strongly agree
2 2
3 3
4 4
5 5 - strongly disagree

PUT IN state2 Q4351,2

QUESTION 4352

IF [Q4321 , 3 , 4 , 5]

I do not know enough about how much electricity different appliances use in order to reduce my usage

1	1 - strongly agree
2	2
3	3
4	4
5	5 - strongly disagree

PUT IN state2 Q4351,3

QUESTION 4352

IF [Q4321 , 3 , 4 , 5]

I am not be able to get the people I live with to reduce their electricity usage

1	1 - strongly agree
2	2
3	3
4	4
5	5 - strongly disagree

PUT IN state2 Q4351,4

QUESTION 4352

IF [Q4321 , 3 , 4 , 5]

I do not have enough time to reduce my electricity usage

1	1 - strongly agree
2	2
3	3
4	4
5	5 - strongly disagree

PUT IN state2 Q4351,5

QUESTION 4352

IF [Q4321 , 3 , 4 , 5]

I do not want to be told how much electricity I can use

1	1 - strongly agree
2	2
3	3
4	4
5	5 - strongly disagree

PUT IN state2 Q4351,6

QUESTION 4352

IF [Q4321 , 3 , 4 , 5]

Reducing my usage would not make enough of a difference to my bill

1	1 - strongly agree
2	2
3	3
4	4
5	5 - strongly disagree

QUESTION 43521

IF [Q4321 , 3 , 4 , 5]

If you were to make changes to the way you and people you live with use electricity, how much do you believe you could
reduce your usage by?

1	Nothing
2	less than 5%
3	Between 5% and 10%
4	Between 10% and 20%
5	Between 20% and 30%
6	More than 30%

QUESTION 450

I would now like to ask some questions about your home. Which best describes your home?

1	Apartment
2	Semi-detached house
3	Detached house
4	Terraced house
5	Bungalow
6	Refused

QUESTION 452

Do you own or rent your home?

1	Rent (from a private landlord)
2	Rent (from a local authority)
3	Own Outright (not mortgaged)
4	Own with mortgage etc
5	Other

QUESTION 453 **NUMBER**

What year was your house built
INT ENTER FOR EXAMPLE: 1981- CAPTURE THE FOUR DIGITS

QUESTION 4531

IF [Q453 , 9999]

Approximately how old is your home?

1 Less than 5 years old
2 Less than 10 years old
3 Less than 30
4 Less than 75
5 More than 75 years old

QUESTION 6103 **NUMBER**

What is the approximate floor area of your home?

QUESTION 61031

IF [Q6103 < 999999999]

Is that

1 square meters
2 or square feet

QUESTION 460

How many bedrooms are there in your home

1 1
2 2
3 3
4 4
5 5+
6 Refused

QUESTION 470 **MULTIPLE**

Which of the following best describes how you heat your home?

1 Electricity (electric central heating storage heating)
2 Electricity (plug in heaters)
3 Gas
4 Oil
5 Solid fuel
6 Renewable (e.g. solar)
7 Other

QUESTION 47001

Do you have a timer to control when your heating comes on and goes off?

1 Yes
2 No

QUESTION 4701 **MULTIPLE**

Which of the following best describes how you heat water in your home?

1 Central heating system
2 Electric (immersion)
3 Electric (instantaneous heater)
4 Gas
5 Oil
6 Solid fuel boiler
7 Renewable (e.g. solar)
8 Other

QUESTION 47011

Do you have a timer to control when your hot water/immersion heater comes on and goes off?

1 Yes
2 No

QUESTION 4801

IF [Q4701 , 2]

Do you use your immersion when your heating is not switched on?

1 Yes
2 No

QUESTION 4704

Which of the following best describes how you cook in your home

1 Electric cooker
2 Gas cooker
3 Oil fired cooker
4 Solid fuel cooker (stove aga)

QUESTION 471

Returning to heating your home, in your opinion, is your home kept adequately warm?

1 Yes
2 No

QUESTION 472 **MULTIPLE**

IF [Q471 , 2]

Do any of the following reasons apply?

1 I prefer cooler temperature
2 I cannot afford to have the home as warm as I would like
3 It is hard to keep the home warm because it is not well insulated
4 None of these

QUESTION 473

Have you had to go without heating during the last 12 months through lack of money?

1 Yes
2 No

QUESTION 474 **MULTIPLE**

IF [Q473 , 1]

Have any of the following ever applied to you?

1 I had to go without heat on a cold day
2 I had to go to bed to keep warm
3 I lit the fire late or switched on the heat late because I did not have enough fuel or money for fuel
4 None of these

QUESTION 490

Please indicate how many of the following appliances you have in your home?

QUESTION 49001 **MULTIPLE**

DUMMY QUESTION

1 Washing machine
2 Tumble dryer
3 Dishwasher
4 Electric shower (instant)
5 Electric shower (electric pumped from hot tank)
6 Electric cooker
7 Electric heater (plug-in convector heaters)
8 Stand alone freezer
9 A water pump or electric well pump or pressurised water system
10 Immersion

PUT IN state3 Q49001,1

QUESTION 49002

Washing machine

1 None
2 1
ADD TO Q49001 [1]
3 2
ADD TO Q49001 [1]
4 More than 2
ADD TO Q49001 [1]

PUT IN state3 Q49001,2

QUESTION 49002

Tumble dryer

1	None
2	1 *ADD TO Q49001 [2]*
3	2 *ADD TO Q49001 [2]*
4	More than 2 *ADD TO Q49001 [2]*

PUT IN state3 Q49001,3

QUESTION 49002

Dishwasher

1	None
2	1 *ADD TO Q49001 [3]*
3	2 *ADD TO Q49001 [3]*
4	More than 2 *ADD TO Q49001 [3]*

PUT IN state3 Q49001,4

QUESTION 49002

Electric shower (instant)

1	None
2	1 *ADD TO Q49001 [4]*
3	2 *ADD TO Q49001 [4]*
4	More than 2 *ADD TO Q49001 [4]*

PUT IN state3 Q49001,5

QUESTION 49002

Electric shower (electric pumped from hot tank)

1	None
2	1 *ADD TO Q49001 [5]*
3	2 *ADD TO Q49001 [5]*
4	More than 2 *ADD TO Q49001 [5]*

PUT IN state3 Q49001,6

QUESTION 49002

Electric cooker

1 None
2 1
ADD TO Q49001 [6]
3 2
ADD TO Q49001 [6]
4 More than 2
ADD TO Q49001 [6]

IF [1 & Q4704 , 1 & Q49002 , 1] ADD TO Q49002 [2]
PUT IN state3 Q49001,7

QUESTION 49002
Electric heater (plug-in convector heaters)

1 None
2 1
ADD TO Q49001 [7]
3 2
ADD TO Q49001 [7]
4 More than 2
ADD TO Q49001 [7]

PUT IN state3 Q49001,8
QUESTION 49002
Stand alone freezer

1 None
2 1
ADD TO Q49001 [8]
3 2
ADD TO Q49001 [8]
4 More than 2
ADD TO Q49001 [8]

PUT IN state3 Q49001,9
QUESTION 49002
A water pump or electric well pump or pressurised water system

1 None
2 1
ADD TO Q49001 [9]
3 2
ADD TO Q49001 [9]
4 More than 2
ADD TO Q49001 [9]

PUT IN state3 Q49001,10
QUESTION 49002

Immersion

1 None
2 1
ADD TO Q49001 [10]
3 2
ADD TO Q49001 [10]
4 More than 2
ADD TO Q49001 [10]

QUESTION 4901
And how many of the following entertainment appliances do you have? Only those that are actually used should be mentioned?

QUESTION 4902 **MULTIPLE**
DUMMY QUESTION

1 TV's less than 21 inch
2 TV's greater than 21 inch
3 Desk-top computers
4 Lap-top computers
5 Games consoles, such as xbox, playstation or Wii

PUT IN state4 Q4902,1
QUESTION 490002
TV's less than 21 inch

1 None
2 1
ADD TO Q4902 [1]
3 2
ADD TO Q4902 [1]
4 3
ADD TO Q4902 [1]
5 More than 3
ADD TO Q4902 [1]

PUT IN state4 Q4902,2
QUESTION 490002

TV's greater than 21 inch

1 None
2 1
ADD TO Q4902 [2]
3 2
ADD TO Q4902 [2]
4 3
ADD TO Q4902 [2]
5 More than 3
ADD TO Q4902 [2]

PUT IN state4 Q4902,3

QUESTION 490002

Desk-top computers

1 None
2 1
ADD TO Q4902 [3]
3 2
ADD TO Q4902 [3]
4 3
ADD TO Q4902 [3]
5 More than 3
ADD TO Q4902 [3]

PUT IN state4 Q4902,4

QUESTION 490002

Lap-top computers

1 None
2 1
ADD TO Q4902 [4]
3 2
ADD TO Q4902 [4]
4 3
ADD TO Q4902 [4]
5 More than 3
ADD TO Q4902 [4]

PUT IN state4 Q4902,5

QUESTION 490002

Games consoles, such as xbox, playstation or Wii

1 None
2 1
ADD TO Q4902 [5]
3 2
ADD TO Q4902 [5]
4 3
ADD TO Q4902 [5]
5 More than 3
ADD TO Q4902 [5]

QUESTION 4903
In a typical day, how often would you or your family/household use each appliance - please think of the total use by all
household/family members

QUESTION 49003
DUMMY QUESTION

1 Washing machine
2 Tumble dryer
3 Dishwasher

PUT IN state5 Q49003,1
QUESTION 49004
Washing machine
INT:PROBE TO PRECODES

1 Less than 1 load a day typically
2 1 load typically
3 2 to 3 loads
4 More than 3 loads

PUT IN state5 Q49003,2
QUESTION 49004
Tumble dryer
INT:PROBE TO PRECODES

1 Less than 1 load a day typically
2 1 load typically
3 2 to 3 loads
4 More than 3 loads

PUT IN state5 Q49003,3
QUESTION 49004

Dishwasher
INT:PROBE TO PRECODES

1 Less than 1 load a day typically
2 1 load typically
3 2 to 3 loads
4 More than 3 loads

QUESTION 490004

IF [Q49001 , 4]

Electric shower (instant)
INT:PROBE TO PRECODES

1 Less than 5 mins
2 5-10 mins
3 10-20 mins
4 Over 20 mins

QUESTION 4900004

IF [Q49001 , 5]

Electric shower (pumped from hot tank)
INT:PROBE TO PRECODES

1 Less than 5 mins
2 5-10 mins
3 10-20 mins
4 Over 20 mins

QUESTION 4900005

IF [Q49001 , 6]

Electric cooker
INT:PROBE TO PRECODES

1 Less than 30 mins
2 30-60 mins
3 1-2 hours
4 Over 2 hours

QUESTION 4900006

IF [Q49001 , 7]

Electric heater (plug-in)
INT:PROBE TO PRECODES

1 Less than 30 mins
2 30-60 mins
3 1-2 hours
4 Over 2 hours

QUESTION 4900007

IF [Q49001 , 9]

Water pump
INT:PROBE TO PRECODES

1 Less than 30 mins
2 30-60 mins
3 1-2 hours
4 Over 2 hours

QUESTION 4900008

IF [Q49001 , 10]

Immersion water
INT:PROBE TO PRECODES

1 Less than 30 mins
2 30-60 mins
3 1-2 hours
4 Over 2 hours

QUESTION 4900009

IF [Q49001 , 8]

Stand alone Freezer
INT:PROBE TO PRECODES

1 For part of the year (4-6 months)
2 All year

QUESTION 49011

And considering the following appliances - please indicate the daily level of total use by all household/family members

QUESTION 49021

DUMMY QUESTION

1 TV's less than 21 inch
2 TV's greater than 21 inch
3 Desk-top computers
4 Lap-top computers
5 Games consoles, such as xbox, playstation or Wii

PUT IN state6 Q49021,1

QUESTION 49022

TV's less than 21 inch
INT:PROBE TO PRECODES

1 Less than 1 hour a day or a few hours a week typically
2 1 - 3 hours per day typically
3 3-5 hours per day typically
4 More than 5 hours per day typically

PUT IN state6 Q49021,2
QUESTION 49022
TV's greater than 21 inch
INT:PROBE TO PRECODES

1 Less than 1 hour a day or a few hours a week typically
2 1 - 3 hours per day typically
3 3-5 hours per day typically
4 More than 5 hours per day typically

PUT IN state6 Q49021,3
QUESTION 49022
Desk-top computers
INT:PROBE TO PRECODES

1 Less than 1 hour a day or a few hours a week typically
2 1 - 3 hours per day typically
3 3-5 hours per day typically
4 More than 5 hours per day typically

PUT IN state6 Q49021,4
QUESTION 49022
Lap-top computers
INT:PROBE TO PRECODES

1 Less than 1 hour a day or a few hours a week typically
2 1 - 3 hours per day typically
3 3-5 hours per day typically
4 More than 5 hours per day typically

PUT IN state6 Q49021,5
QUESTION 49022
Games consoles, such as xbox, playstation or Wii
INT:PROBE TO PRECODES

1 Less than 1 hour a day or a few hours a week typically
2 1 - 3 hours per day typically
3 3-5 hours per day typically
4 More than 5 hours per day typically

QUESTION 455

Does your home have a Building Energy Rating (BER) - a recently introduced scheme for rating the energy efficiency of your
home?

1 Yes
2 No
3 Don’t know

QUESTION 4551

IF [Q455 , 1]

What rating did your house achieve?

1 A
2 B
3 C
4 D
5 E
6 F
7 G

QUESTION 4905

And now considering energy reduction in your home please indicate the approximate proportion of light bulbs which are
energy saving (or CFL)?
INT:READ OUT

1 None
2 About a quarter
3 About half
4 About three quarters
5 All

QUESTION 4906

Please indicate the approximate proportion of windows in your home which are double glazed?
INT:READ OUT

1 None
2 About a quarter
3 About half
4 About three quarters
5 All

QUESTION 4907

Does your hot water tank have a lagging jacket?

1 Yes
2 No

QUESTION 4908

Is your attic insulated and if so when was the insulation fitted?
INT:PROBE TO PRECODES

1 Yes, within the last 5 years
2 Yes, more than 5 years ago
3 No
4 Don’t know

QUESTION 4909
Are the external walls of your home insulated?

1 Yes
2 No
3 Don’t know

QUESTION 1060
I would now like to ask a few questions about your decision to participate in the national smart meter trial.
Thinking about the reasons why you chose to participate, please rate each of the following potential reasons on a scale of
1 to 5 where 1 is very close to your reason and 5 is not at all a reason.

QUESTION 1061 **MULTIPLE**
DUMMY QUESTION

QUESTION 1062

QUESTION 5000

QUESTION 5511 **MULTIPLE**
I would now like to ask you about your expectations about your participation in the trial. I must stress that you should
not interpret these questions as meaning that any of these will happen as a part of the trial
Which of the following do you think will be benefits?
READ OUT

3 Learn how to reduce my energy usage
4 Learn how to reduce my electricity bill
5 Do my part to help the environment by my participation
6 Do my part to make Ireland become more up to date

QUESTION 5512
Thinking of what will be the main consequences of your participation in the trial, for each of the following statements,
please state whether you agree or disagree using a scale of 1 to 5 where 1 in Strongly agree and 5 is strongly disagree?

QUESTION 55122 **MULTIPLE**

DUMMY QUESTION

1 My household may decide to make minor changes to the way we use electricity
2 My household may decide to make major changes to the way we use electricity
3 My household may decide to be more aware of the amount of electricity used by appliances we own or buy.
5 In future, when replacing an appliance, my household may decide to choose one with a higher energy rating

PUT IN state10 Q55122,1

QUESTION 55123

My household may decide to make minor changes to the way we use electricity

1 1-Strongly agree
2 2
3 3
4 4
5 5-Strongly disagree.

PUT IN state10 Q55122,2

QUESTION 55123

My household may decide to make major changes to the way we use electricity

1 1-Strongly agree
2 2
3 3
4 4
5 5-Strongly disagree.

PUT IN state10 Q55122,3

QUESTION 55123

My household may decide to be more aware of the amount of electricity used by appliances we own or buy.

1 1-Strongly agree
2 2
3 3
4 4
5 5-Strongly disagree.

PUT IN state10 Q55122,4

PUT IN state10 Q55122,5

QUESTION 55123

In future, when replacing an appliance, my household may decide to choose one with a higher energy rating

1 1-Strongly agree
2 2
3 3
4 4
5 5-Strongly disagree.

PUT IN state10 Q55122,6

QUESTION 5414
How do you think that your electricity bills will change as part of the trial?

1 No change
2 Increase
3 Decrease

QUESTION 5415

IF [Q5414 , 2]

By what amount?

1 less than 5%
2 between 5% and 10%
3 between 10% and 20%
4 between 20% and 30%
5 more than 30%
6 don’t know

QUESTION 54155

IF [Q5414 , 3]

By what amount?

1 less than 5%
2 between 5% and 10%
3 between 10% and 20%
4 between 20% and 30%
5 more than 30%
6 don’t know

QUESTION 5418

Moving on to education, which of the following best describes the level of education of the chief income earner

1 No formal education
2 Primary
3 Secondary to Intermediate Cert Junior Cert level
4 Secondary to Leaving Cert level
5 Third level
6 Refused

QUESTION 402 **NUMBER**

And considering income, what is the approximate income of your household - this should be before tax, you should include
the income of all adults in the household? Please note that this figure will remain completely confidential and will not be
reported at an individual level.
[ATTEMPT TO CAPTURE ANNUAL]
INTERVIEWER: IF THE RESPONDENT SAYS THEIR INCOME IS 50 GRAND or THOUSAND PLEASE ENTER 50000 DO NOT ENTER JUST 50

QUESTION 4021

IF [Q402 = 9999999]

Can you state which of the following broad categories best represents the yearly household income BEFORE TAX?

1 Less than 15,000 Euros
2 15,000 to 30,000 Euros
3 30,000 to 50,000 Euros
4 50,000 to 75,000 Euros
5 75,000 or more Euros
6 Refused

QUESTION 403

IF [Q402 < 9999999]

Is that figure

1 Per week
2 Per month
3 Per year

QUESTION 404

IF [Q402 < 9999999]

Can I just double check is that figure..

1 Before tax
2 or after tax

QUESTION 55101

Thinking about electricity and its use, generation and sale in the Irish context, please indicate your level of satisfaction with each of the following were 1 is very satisfied and 5 is very dissatisfied:

QUESTION 55111

DUMMY QUESTION

1	The number of suppliers competing in the market
3	The percentage of electricity being generated from renewable sources
5	The overall cost of electricity
6	The number of estimated bills received by customers
7	The opportunity to sell back extra electricity you may generate (from solar panels etc) to your electricity supplier
8	The environmental damage associated with the amount of electricity used

PUT IN state8 Q55111,1

QUESTION 55112

The number of suppliers competing in the market

1	1-Very Satisfied
2	2
3	3
4	4
5	5-Very Dissatisfied

PUT IN state8 Q55111,2

PUT IN state8 Q55111,3

QUESTION 55112

The percentage of electricity being generated from renewable sources

1	1-Very Satisfied
2	2
3	3
4	4
5	5-Very Dissatisfied

PUT IN state8 Q55111,4

QUESTION 55112

PUT IN state8 Q55111,5

QUESTION 55112

The overall cost of electricity

1	1-Very Satisfied
2	2
3	3
4	4
5	5-Very Dissatisfied

PUT IN state8 Q55111,6

QUESTION 55112

The number of estimated bills received by customers

1	1-Very Satisfied
2	2
3	3
4	4
5	5-Very Dissatisfied

PUT IN state8 Q55111,7

QUESTION 55112

The opportunity to sell back extra electricity you may generate (from solar panels etc) to your electricity supplier

1	1-Very Satisfied
2	2
3	3
4	4
5	5-Very Dissatisfied

PUT IN state8 Q55111,8

QUESTION 55112

The environmental damage associated with the amount of electricity used

1	1-Very Satisfied
2	2
3	3
4	4
5	5-Very Dissatisfied

PUT IN state8 Q55111,9

QUESTION 55112

PUT IN state8 Q55111,10

QUESTION 55112

QUESTION 30000

This was my last question.

Appendix B

- **Gaussian Mixture Model (GMM)**

In statistics, the mixture model is a probabilistic model whose mixture probability density function (PDF) for all observations can be described as weighted sums of PDFs of the finite number of sub-populations (the sub-populations is also called mixture component). In GMM, the PDF of every sub-population is a normal distribution.

Suppose the PDF of the overall observations is $p(x)$, which is consisted by K Gaussian mixture component. Then, the PDF of all observations can be written as (B-1) shown below:

$$p(x) = \sum_{k=1}^{K} p(k) \cdot p(x|k) \tag{B-1}$$

where x represents each observation, $p(k)$ is the probability of observation x is belong to k^{th} mixture component. Due to in GMM, the PDF of every mixture component follow the normal Gaussian distribution, (B-1) can be rewritten as (B-2).

$$p(x) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x\,|\mu_k,\ \Sigma_k) \tag{B-2}$$

In (B-2), the π_k is the weight of the k^{th} mixture component. Since the total probability of every mixture component density equals to 1, the sum of π_k subject to (B-3). $\mathcal{N}(x\,|\mu_k,\ \Sigma_k)$ represents the normal distributions of k^{th} mixture component where the μ_k and Σ_k are the mean value and variance of the distribution.

$$\sum_{k=1}^{K} \pi_k = 1 \ \ (0 < \pi_k < 1) \tag{B-3}$$

Then, it becomes a parameter estimation problem where the parameters of π_k, μ_k, and Σ_k need to be estimated for each mixture component.

The Expectation Maximum (EM) algorithm is applied to estimate those three parameters. The EM is an iteration method which alternates the parameters values between performing an Expectation (E) step and a Maximisation (M) step.

In the E step, the initial parameters values are used to calculate the expectation of the log-likelihood function. For given total J observations, the log-likelihood function, $L(x_j, K)$, is calculated by (B-4):

$$\log L(x_j, K) = \sum_{j=1}^{J} \log\left\{\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(x_j \,\middle|\, \mu_k,\ \Sigma_k\right)\right\} \tag{B-4}$$

Once the parameters are estimated, for each observation x_j, the probability of it can be assign to mixture component k can be calculated by (B-5).

$$p\left(k \middle| x_j\right) = \frac{\pi_k \cdot \mathcal{N}\left(x_j \,\middle|\, \mu_k,\ \Sigma_k\right)}{\sum_{g=1}^{K} \pi_g \cdot \mathcal{N}\left(x_j \,\middle|\, \mu_g,\ \Sigma_g\right)} \tag{B-5}$$

Then, the M step will estimate parameters π_k, μ_k, and Σ_k by maximising equation (B-4) under the new probability $p\left(k \middle| x_j\right)$. The new parameters' values can be obtained by (B-6) –(B-8)

$$\mu_k = \frac{1}{J_k} \sum_{j=1}^{J} p\left(k \middle| x_j\right) \cdot x_j \tag{B-6}$$

$$\Sigma_k = \frac{1}{J_k} \sum_{j=1}^{J} p\left(k \middle| x_j\right) \cdot (x_j - \mu_k) \cdot (x_j - \mu_k)^{T} \tag{B-7}$$

$$\pi_k = \frac{J_k}{J} \tag{B-8}$$

The final parameters can be figured out by repeating E step and M step until the difference of max ($\log L(x_j, K)$) values between two adjacent iterations converge.

• Principal Component Analysis (PCA)

The principal component analysis is defined as an orthogonal linear transformation, which is a widely used algorithm in dimensionality reduction. High-dimensional data could be transformed into a new coordinate system with smaller dimensions. The new feature which has the greatest variance after projection on the first coordinate is defined as the first principal component, the second greatest variance on the second coordinate is the second principal component and so on.

Suppose there is a dataset X whose size is $n \times p$. Therefore, p is the dimension number of the original dataset. Then, to transform each row vector to a l-dimensional vector (where $l < p$), a coefficient vector $\boldsymbol{\omega}_{(r)} = (\omega_1, \omega_2, \dots, \omega_l)_{(r)}$ will multiply the original vector to obtain a new vector of principal component scores$t_{(i)} = (t_1, t_2, \dots, t_l)_{(i)}$, just as equation (B-9) shown.

$$t_{r(i)} = \boldsymbol{x}_{(i)} \cdot \boldsymbol{\omega}_{(r)} \quad \text{for } i = 1, \dots, n; \;\; r = 1, \dots, l \tag{B-9}$$

Due to first component is the projection with the greatest variance, the coefficient $\boldsymbol{\omega}_{(1)}$ must satisfy the (B-10):

$$\boldsymbol{\omega}_{(1)} = \arg\, max\left\{\textstyle\sum_i (t_1)^2_{(i)}\right\}, \text{ where } \|\boldsymbol{\omega}\|=1 \tag{B-10}$$

Equivalently, (B-10) can be writing in matrix version, just as (B-11) demonstrated.

$$\boldsymbol{\omega}_{(1)} = \underset{\|\boldsymbol{\omega}\|=1}{argmax}\{\|\boldsymbol{X}\boldsymbol{\omega}\|^2\} = \underset{\|\boldsymbol{\omega}\|=1}{argmax}\{\boldsymbol{\omega}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\omega}\}, \tag{B-10}$$

Since the $\boldsymbol{\omega}_{(1)}$ is defined as a unit vector, it will be equal to:

$$\boldsymbol{\omega}_{(1)} = \underset{\|\boldsymbol{\omega}\| = 1}{argmax}\left\{\frac{\boldsymbol{\omega}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\omega}}{\boldsymbol{\omega}^T\boldsymbol{\omega}}\right\} \tag{B-11}$$

To calculate the coefficient vector for the further component, it needs to subtract the previous principal components. For example, the k^{th} component $\hat{\mathbf{X}}_k$ is equal to (B-12).

$$\hat{\mathbf{X}}_k = \boldsymbol{X} - \sum_{s=1}^{k-1} \boldsymbol{X}\,\boldsymbol{\omega}_{(\mathrm{s})}\boldsymbol{\omega}_{(s)}^T \tag{B-12}$$

Then, the coefficient vector $\boldsymbol{\omega}_{(\mathrm{k})}$ for the k^{th} component $\hat{\mathbf{X}}_k$ can be calculated as (B-13).

$$\boldsymbol{\omega}_{(\mathrm{k})} = \arg max \left\{ \frac{\boldsymbol{\omega}^T \hat{\mathbf{X}}_k{}^T \hat{\mathbf{X}}_k \boldsymbol{\omega}}{\boldsymbol{\omega}^T \boldsymbol{\omega}} \right\} \tag{B-13}$$

With the coefficient matrix, the new features with lower dimensionality can be calculated. The vector $\boldsymbol{\omega}$ is the projection dimension found in the PCA where you can maximize the variance.

- **Kernel Fisher Analysis (KFA)**

The kernel fisher analysis, also known as kernel fisher discriminant analysis, is a kernelised version of Linear Discriminant Analysis (LDA). The intuitive idea of LDA is to project data to a new space where class separation is maximised. The degree of class separation in LDA is defined as the ratio of class means difference over the sum of within-class variance. Suppose there are two classes of data, C_1 and C_2, l_i is the number of examples $\boldsymbol{x}_n{}^i$ in class C_i(n=1,…, l_i), m_i is the mean value of in class C_i. LDA aims to find a project dimension $\boldsymbol{w}$ which can maximum the ratio in (B-14)

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w}} \tag{B-14}$$

The $\boldsymbol{S}_B$ is the between-class covariance matrix, which can be calculated as (B-15). The $\boldsymbol{S}_W$ is the within-class covariance matrix which can be calculated as (B-16)

$$\boldsymbol{S}_B = (\boldsymbol{m}_2 - \boldsymbol{m}_1)(\boldsymbol{m}_2 - \boldsymbol{m}_1)^T \tag{B-15}$$

$$\boldsymbol{S}_W = \sum_{i=1,2} \sum_{n=1}^{l_i} (\boldsymbol{x}_n^i - \boldsymbol{m}_i)\,(\boldsymbol{x}_n^i - \boldsymbol{m}_i)^T \tag{B-16}$$

To maximising of formula (B-14) is equivalent to maximising $\boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w}$, subjected to $\boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w} = \mathbf{1}$ by introduce the Lagrange multiplier λ. Therefore, maximising (B-14) becomes equivalent to maximizing (B-17).

$$I(\boldsymbol{w},\lambda) = \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w} - \lambda(\boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w} - \mathbf{1}) \tag{B-17}$$

The derivations of $I(\boldsymbol{w},\lambda)$ with respect to $\boldsymbol{w}$ and λ must be zero. Therefore,

$$\frac{dI}{d\boldsymbol{w}} = \boldsymbol{S}_B \boldsymbol{w} - \lambda \boldsymbol{S}_W \boldsymbol{w} = 0 \tag{B-18}$$

To satisfied (B-18), the vector $\boldsymbol{w}$ will equal to (B-19) and λ will equal to (B-20).

$$\boldsymbol{w} = c\boldsymbol{S}_W{}^{-1}(\boldsymbol{m}_2 - \boldsymbol{m}_1) \tag{B-18}$$

$$\lambda = (\boldsymbol{m}_2 - \boldsymbol{m}_1)^T \boldsymbol{S}_W{}^{-1}(\boldsymbol{m}_2 - \boldsymbol{m}_1) \tag{B-20}$$

The KFA extendS LDA to non-linear mapping, the data x_n would be mapped to a new feature space via the function $\phi(x_n)$. Then, equations (B-14) to (B-16) will be written as (B-21) – (B-23)

$$J(\boldsymbol{W}) = \frac{\left|\boldsymbol{W}^T \boldsymbol{S}_B{}^{\phi} \boldsymbol{W}\right|}{\left|\boldsymbol{W}^T \boldsymbol{S}_W{}^{\phi} \boldsymbol{W}\right|} \tag{B-21}$$

With

$$\boldsymbol{S}_B{}^{\phi} = \sum_{i=1}^{c} l_i (\boldsymbol{m}_i{}^{\phi} - \boldsymbol{m}^{\phi})(\boldsymbol{m}_i{}^{\phi} - \boldsymbol{m}^{\phi})^T \tag{B-22}$$

$$\boldsymbol{S}_W{}^{\phi} = \sum_{i=1}^{c} \sum_{n=1}^{l_i} (\phi(\boldsymbol{x}_n{}^i) - \boldsymbol{m}_i{}^{\phi})(\phi(\boldsymbol{x}_n{}^i) - \boldsymbol{m}_i{}^{\phi})^T \tag{B-23}$$

In terms of dot products and using the kernel trick where $k(\boldsymbol{x},\boldsymbol{y}) = \phi(\boldsymbol{x}) \cdot \phi(\boldsymbol{y})$, the maximisation of (B-21) becomes to the goal to achieve the maximum value of (B-24).

$$\boldsymbol{A}^* = \underset{A}{argmax} \frac{|\boldsymbol{A}^T M \boldsymbol{A}|}{|\boldsymbol{A}^T N \boldsymbol{A}|} \tag{B-24}$$

where $\boldsymbol{A}$ is a constant matrix and the $\boldsymbol{M}$ and $\boldsymbol{N}$ which are defined in (B-25) and (B-26) will decide the maximum value.

$$\boldsymbol{M} = \sum_{j=1}^{c} l_j \{\boldsymbol{M}_j - \frac{1}{l}\sum_{n=1}^{l} k(\boldsymbol{x}_j, \boldsymbol{x}_n)\} \{(\boldsymbol{M}_j - \frac{1}{l}\sum_{n=1}^{l} k(\boldsymbol{x}_j, \boldsymbol{x}_n)\}^T \tag{B-25}$$

$$\boldsymbol{N} = \sum_{j=1}^{c} \boldsymbol{K}_j \, (\boldsymbol{I} - \boldsymbol{1}_{l_j})\boldsymbol{K}_j{}^T \tag{B-26}$$

Therefore, the goal of maximisation of $J(\boldsymbol{W})$ could be replaced by finding the dot product value which satisfies the maximum value of $\boldsymbol{A}^*$.

Appendix C

Table C- 1: The interacting socio-economic groups for the original factors detected by *Stage I*

Qu 84	**Qu 98**	**Qu 99**	**Qu 2**	**Qu 142**
Number of the electric cookers you own	Number of the Game consoles you own	Number of the Wash Machine you own	Age of the CIE	Do you agree the environmental damage associated with the amount of electricity used
Qu 58	**Qu 98**	**Qu 99**	**Qu 5**	**Qu 85**
Describes how you cook	Number of the Game consoles you own	Number of the Wash Machine you own	Do you have internet access in your home	Number of the plug-in electric heater you own
Qu 97	**Qu 90**	**Qu 74**	**Qu 15**	**Qu 86**
Number of the Lap-top you own	Number of TV greater than 21 inch you won	Do you have the electric cooker	Do you interested in changing the way you use electricity if it helps the environment	Number of the Standalone freezer you own
Qu 92	**Qu 74**	**Qu 41**	**Qu 60**	**Qu 86**
Do you have Lap-top	Do you have the electric cooker	Do you have plug in electricity heaters	Your home did not keep adequate warm due to you prefer cooler temperature	Do you have enough time to reduce usage
Qu 104	**Qu 98**	**Qu 74**	**Qu 128**	
How often would you use electricity cooker	Number of the Game consoles you own	Do you have the electric cooker	Do you agree that you will decide to choose an appliance with a higher energy rating	
Qu 41	**Qu 57**	**Qu 74**	**Qu 3**	
Do you have plug in electricity heaters	When heating is not switched on, do you use your immersion	Do you have the electric cooker	The employment status of the CIE	

Qu 3	**Qu 77**	**Qu 90**	**Qu 13**	**Qu 39**
The employment status of the CIE	Do you have a water pump or electric well pump or pressurised water system	Number of TV greater than 21 inch you won	How many adults and children under 15 years old are typically in the house during the day	How many bedrooms are there in your home
Qu 90	**Qu 96**	**Qu 142**	**Qu 113**	
Number of TV greater than 21 inch you won	Number of the Desk-top computers you own	Do you agree the environmental damage associated with the amount of electricity used	How often would you use the Games consoles	
Qu 9	**Qu 84**	**Qu 2**	**Qu 117**	**Qu 102**
The description the people you live with	Number of the electric cookers you own	Age of the HRP	The proportion of windows in your home which are double glazed	How often would you use the electric shower (instant)
Qu 105	**Qu 73**	**Qu 74**	**Qu 139**	**Qu 88**
How often would you use the plug-in electricity heater	Do you have the electric shower pumped from hot tank	Do you have the electric cooker	Do you satisfy with the overall cost of electricity	Number of immersions you own
Qu 2	**Qu 74**	**Qu 92**	**Qu 100**	**Qu 36**
Age of the CIE	Do you have the electric cooker	Do you have Lap-top	How often would you use the Tumble Dryer	How old is your home
Qu 5	**Qu 89**	**Qu 75**	**Qu 139**	**Qu 112**
Do you have internet access in your home	Do you have TV less than 21 inch	Do you have the plug-in electric heater	Do you satisfy with the overall cost of electricity	How often would you use the Lap-top computers
Qu 11	**Qu 98**	**Qu 75**	**Qu 5**	**Qu 36**
How many people typically in the house during the day	Number of the Game consoles you own	Do you have the plug-in electric heater	Do you have internet access in your home	How old is your home

Qu 53	**Qu 89**	**Qu 69**	**Qu 100**	**Qu 4**
Do you heat water by the solid fuel	Do you have TV less than 21 inch	Do you have washing machines	How often would you use the Tumble Dryer	SOCIAL CLASS of CIEs
Qu 4	**Qu 75**	**Qu 9**	**Qu 100**	**Qu 36**
SOCIAL CLASS of CIEs	Do you have the plug-in electric heater	The description the people you live with	How often would you use the Tumble Dryer	How old is your home
Qu 91	**Qu 74**	**Qu 97**	**Qu 139**	**Qu 48**
Do you have the Desk-top computer	Do you have the electric cooker	Number of the Lap-top you own	Do you satisfy with the overall cost of electricity	Do you heat water by the central heating
Qu 99	**Qu 83**	**Qu 54**	**Qu 20**	**Qu 116**
Number of the Wash Machine you own	How many electric shower (pumped from hot tank) do you have	Do you heat water by the solar energy	Do you know what you need to do in order to reduce electricity usage	The approximate proportion of light bulbs

Table C- 2: The interacting socio-economic groups for the original factors detected in *Stage II* by KL-divergence

Qu 102	**Qu 4**	**Qu 15**	**Qu 80**
How often you use the electric shower	SOCIAL CLASS of CIEs	Do you interested in changing the way you use electricity if it helps the environment	Number of Tumble dryer you own
Qu 95	**Qu 117**	**Qu 54**	**Qu 13**
Number of TVs greater than 21 inch you have	The proportion of windows in your home which are double glazed	Do you heat water by the solar energy	How many adults and children under 15 years old are typically in the house during the day
Qu 29	**Qu 39**	**Qu 2**	**Qu 43**
You cannot get the people you live with to reduce their usage	How many bedrooms are there in your home	Age of the CIE	Do you use oil to heat your room
Qu 7	**Qu 84**	**Qu 104**	**Qu 79**
Do you use the internet regularly yourself	Number of the electric cookers you own	How often would you use electricity cooker	Do you have washing machine
Qu 32	**Qu 76**	**Qu 25**	**Qu 35**
Do you think that reducing usage would not make enough difference to your bill	Do you have a Standalone freezer	Thinking about the energy reduction activities undertaken by you or your family household in the last year did your efforts reduce your bills.	Do you own or rent your home

Publications

Journal Publications

Qiuyang. Ma; Minghao. Xu; Ran Li; Furong Li; " Impact Assessment of Smart Meters on Electricity Cross-subsidies A High-dimensional Interaction-Aware KLAM Algorithm," *IEEE Transaction on power system*, (under review).

Qiuyang. Ma; Minghao. Xu; Ran Li; Furong Li; Youbo Liu; Yue Xiang; "A Cost-reflective Customer Segmentation Method for Electricity Suppliers," *Energy policy*, (in submitting).

Shuangyuan Wang; Ran Li; Qiuyang Ma; Furong Li; "Unit Home Equivalent Distribution Network Pricing for Electricity Retail Market," *IEEE Transaction on power system*, (in submitting).

Heng Shi; Qiuyang Ma; Nathan Smith; Furong Li; "Data-driven Uncertainty Quantification and Characterization for Household Energy Demand Across Multiple Time-scales*." IEEE Transactions on Smart Grid 10.3 (2018): 3092-3102*.

Shi H; Xu M; Ma Q, et al. “A Whole System Assessment of Novel Deep Learning Approach on Short-Term Load Forecasting”, *Energy Procedia, 2017, 142: 2791-2796*

Conference Publications

Ma Q; Xu M; Li R and Li F. “Quantitative benefit assessment of electricity settlement using smart meters*” European Energy Market (EEM), 2016 13th International Conference on the. IEEE*, 2016: 1-4.

Yan X, Ma Q, Wyman-Pain H, et al. “Mitigating load forecast errors for suppliers by utilizing energy storage at a substation level” *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2017 IEEE PES. IEEE*, 2017: 1-5.

References

[1] (2012). *Climate Change* [Online]. Available: http://royalsociety.org/policy/climate-change/.

[2] (2010). *2050 Pathways Analysis*.

[3] C. Breidenich, D. Magraw, A. Rowley, and J. W. Rubin, "The Kyoto protocol to the United Nations framework convention on climate change," *American Journal of International Law*, vol. 92, no. 2, pp. 315-331, 1998.

[4] (December 2012). *UK Greenhouse Gas Emission Trading Scheme Regulations*.

[5] (2008). *Climate Change Act 2008* [Online]. Available: http://www.legislation.gov.uk/ukpga/2008/27/pdfs/ukpga_20080027_en.pdf.

[6] DECC. (2018). *2017 UK GREENHOUSE GAS EMISSIONS, PROVISIONAL FIGURES* [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/695930/2017_Provisional_Emissions_statistics_2.pdf.

[7] (July 2011). *UK Renewable Energy Roadmap*.

[8] (2019). *Digest of United Kingdom energy statistics*.

[9] (2010). *Creating Britain's low carbon future. Today* [Online]. Available: https://www.ofgem.gov.uk/ofgem-publications/84840/2013innovationcompetitionbrochure-pdf

[10] N. Pearsall, K. Hynes, C. Martin, and M. Munzinger, "Analysis of performance parameters for UK domestic PV systems," in *2006 IEEE 4th World Conference on Photovoltaic Energy Conference*, 2006, vol. 2: IEEE, pp. 2300-2303.

[11] C. W. Price, "The effect of liberalizing UK retail energy markets on consumers," *Oxford Review of Economic Policy*, vol. 21, no. 1, pp. 128-144, 2005.

[12] (January 2013). *Smart meters: a guide*.

[13] (2014). *Electricity settlement reform – moving to half-hourly settlement*.

[14] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911-918, 2014.

[15] P. Zhang, X. Wu, X. Wang, and S. Bi, "Short-term load forecasting based on big data technologies," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 3, pp. 59-67, 2015.

[16] C. Keerthisinghe, G. Verbič, and A. C. Chapman, "A fast technique for smart home management: ADP with temporal difference learning," *IEEE Transactions on smart grid*, vol. 9, no. 4, pp. 3291-3303, 2016.

[17] M. Bedir, E. Hasselaar, and L. Itard, "Determinants of electricity consumption in Dutch dwellings," *Energy and buildings*, vol. 58, pp. 194-207, 2013.

[18] F. McLoughlin, A. Duffy, and M. Conlon, "Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study," *Energy and buildings*, vol. 48, pp. 240-248, 2012.

[19] D. Wiesmann, I. L. Azevedo, P. Ferrão, and J. E. Fernández, "Residential electricity consumption in Portugal: Findings from top-down and bottom-up models," *Energy Policy*, vol. 39, no. 5, pp. 2772-2779, 2011.

[20] F. Bartiaux and K. Gram-Hanssen, "Socio-political factors influencing household electricity consumption: A comparison between Denmark and Belgium," in *ECEEE Summer Study Proceedings*, 2005, vol. 3, pp. 1313-1325.
[21] D. Brounen, N. Kok, and J. M. Quigley, "Residential energy use and conservation: Economics and demographics," *European Economic Review,* vol. 56, no. 5, pp. 931-945, 2012.
[22] E. Leahy and S. Lyons, "Energy use and appliance ownership in Ireland," *Energy Policy,* vol. 38, no. 8, pp. 4265-4279, 2010.
[23] K. Gram-Hanssen, C. Kofod, and K. N. Petersen, "Different everyday lives: different patterns of electricity use," in *2004 ACEEE Summer study on energy efficiency in buildings*, 2004, pp. 1-13.
[24] L. Nielsen, "How to get the birds in the bush into your hand: results from a Danish research project on electricity savings," *Energy policy,* vol. 21, no. 11, pp. 1133-1144, 1993.
[25] K. J. Baker and R. M. Rylatt, "Improving the prediction of UK domestic energy-demand using annual consumption-data," *Applied Energy,* vol. 85, no. 6, pp. 475-482, 2008.
[26] G. K. Tso and K. K. Yau, "A study of domestic energy usage patterns in Hong Kong," *Energy,* vol. 28, no. 15, pp. 1671-1682, 2003.
[27] Y. G. Yohanis, J. D. Mondol, A. Wright, and B. Norton, "Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use," *Energy and Buildings,* vol. 40, no. 6, pp. 1053-1059, 2008.
[28] P. Wyatt, "A dwelling-level investigation into the physical and socio-economic drivers of domestic energy consumption in England," *Energy Policy,* vol. 60, pp. 540-549, 2013.
[29] H. Chong, "Building vintage and electricity use: Old homes use less electricity in hot weather," *European Economic Review,* vol. 56, no. 5, pp. 906-930, 2012.
[30] P. Tiwari, "Architectural, demographic, and economic causes of electricity consumption in Bombay," *Journal of Policy Modeling,* vol. 22, no. 1, pp. 81-98, 2000.
[31] B. Halvorsen and B. M. Larsen, "Norwegian residential electricity demand—a microeconomic assessment of the growth from 1976 to 1993," *Energy policy,* vol. 29, no. 3, pp. 227-236, 2001.
[32] J. O'Doherty, S. Lyons, and R. S. Tol, "Energy-using appliances and energy-saving features: Determinants of ownership in Ireland," *Applied energy,* vol. 85, no. 7, pp. 650-662, 2008.
[33] W. Abrahamse and L. Steg, "How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings?," *Journal of economic psychology,* vol. 30, no. 5, pp. 711-720, 2009.
[34] D. R. Carlson, H. S. Matthews, and M. Bergés, "One size does not fit all: Averaged data on household electricity is inadequate for residential energy policy and decisions," *Energy and Buildings,* vol. 64, pp. 132-144, 2013.
[35] J. C. Cramer *et al.*, "Social and engineering determinants and their equity implications in residential electricity use," *Energy,* vol. 10, no. 12, pp. 1283-1291, 1985.
[36] S. Yilmaz, S. Weber, and M. Patel, "Who is sensitive to DSM? Understanding the determinants of the shape of electricity load curves and demand shifting: Socio-demographic characteristics, appliance use and attitudes," *Energy Policy,* vol. 133, p. 110909, 2019.

[37] L. M. Platchkov, M. G. Pollitt, D. Reiner, and I. Shaorshadze, "2010 EPRG Public opinion survey: policy preferences and energy saving measures," 2011.

[38] M. O. Oseni, M. G. Pollitt, D. Retner, L.-L. Richter, K. Chyong, and M. Baddeley, "2013 EPRG Public Opinion Survey: Smart Energy Survey–Attitudes and Behaviours," 2013.

[39] A. Spence, C. Demski, C. Butler, K. Parkhill, and N. Pidgeon, "Public perceptions of demand-side management and a smarter energy future," *Nature Climate Change*, vol. 5, no. 6, p. 550, 2015.

[40] R. V. Jones, A. Fuertes, and K. J. Lomas, "The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings," *Renewable and Sustainable Energy Reviews,* vol. 43, pp. 901-917, 2015.

[41] K. Genjo, S.-i. Tanabe, S.-i. Matsumoto, K.-i. Hasegawa, and H. Yoshino, "Relationship between possession of electric appliances and electricity for lighting and others in Japanese households," *Energy and Buildings*, vol. 37, no. 3, pp. 259-272, 2005.

[42] C. Bartusch, M. Odlare, F. Wallin, and L. Wester, "Exploring variance in residential electricity consumption: Household features and building properties," *Applied Energy,* vol. 92, pp. 637-643, 2012.

[43] V. G. Munley, L. W. Taylor, and J. P. Formby, "Electricity Demand in Multi-Family, Renter-Occupied Residences," *Southern Economic Journal,* vol. 57, no. 1, 1990.

[44] S. H. Schwartz, "Normative influences on altruism," in *Advances in experimental social psychology*, vol. 10: Elsevier, 1977, pp. 221-279.

[45] S. H. Schwartz and J. A. Howard, "A normative decision-making model of altruism," *Altruism and helping behavior,* pp. 189-211, 1981.

[46] G. Brandon and A. Lewis, "Reducing household energy consumption: a qualitative and quantitative field study," *Journal of environmental Psychology,* vol. 19, no. 1, pp. 75-85, 1999.

[47] J. S. Black, P. C. Stern, and J. T. Elworth, "Personal and contextual influences on househould energy adaptations," *Journal of applied psychology*, vol. 70, no. 1, p. 3, 1985.

[48] I. Mansouri, M. Newborough, and D. Probert, "Energy consumption in UK households: impact of domestic electrical appliances," *Applied Energy*, vol. 54, no. 3, pp. 211-285, 1996.

[49] S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE transactions on smart grid,* vol. 7, no. 1, pp. 136-144, 2015.

[50] J. Schofield, R. Carmichael, S. Tindemans, M. Woolf, M. Bilton, and G. Strbac, "Residential consumer responsiveness to time-varying pricing," 2014.

[51] Caci, "The ACORN user guide," ed: CACI Marketing Services London, 1989.

[52] L. Blázquez, N. Boogen, and M. Filippini, "Residential electricity demand in Spain: new empirical evidence using aggregate data," *Energy economics,* vol. 36, pp. 648-657, 2013.

[53] A. Kavousian, R. Rajagopal, and M. Fischer, "Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior," *Energy,* vol. 55, pp. 184-194, 2013.

[54] K. Louw, B. Conradie, M. Howells, and M. Dekenah, "Determinants of electricity demand for newly electrified low-income African households," *Energy policy,* vol. 36, no. 8, pp. 2812-2818, 2008.

[55] G. Sun, Y. Cong, D. Hou, H. Fan, X. Xu, and H. Yu, "Joint household characteristic prediction via smart meter data," *IEEE Transactions on Smart Grid,* vol. 10, no. 2, pp. 1834-1844, 2017.

[56] C. Beckel, L. Sadamori, T. Staake, and S. Santini, "Revealing household characteristics from smart meter data," *Energy,* vol. 78, pp. 397-410, 2014.

[57] Y. Wang, Q. Chen, D. Gan, J. Yang, D. S. Kirschen, and C. Kang, "Deep learning-based socio-demographic information identification from smart meter data," *IEEE Transactions on Smart Grid,* vol. 10, no. 3, pp. 2593-2602, 2018.

[58] E. Klaassen, C. Kobus, J. Frunt, and J. Slootweg, "Responsiveness of residential electricity demand to dynamic tariffs: Experiences from a large field test in the Netherlands," *Applied Energy,* vol. 183, pp. 1065-1074, 2016.

[59] S. Gottwalt, W. Ketter, C. Block, J. Collins, and C. Weinhardt, "Demand side management—A simulation of household behavior under variable prices," *Energy policy,* vol. 39, no. 12, pp. 8163-8174, 2011.

[60] B. Sütterlin, T. A. Brunner, and M. Siegrist, "Who puts the most energy into energy conservation? A segmentation of energy consumers based on energy-related behavioral characteristics," *Energy Policy,* vol. 39, no. 12, pp. 8137-8152, 2011.

[61] H. Allcott, "Social norms and energy conservation," *Journal of public Economics,* vol. 95, no. 9-10, pp. 1082-1095, 2011.

[62] J.-M. Cayla, N. Maizi, and C. Marchand, "The role of income in energy consumption behaviour: Evidence from French households data," *Energy policy,* vol. 39, no. 12, pp. 7874-7883, 2011.

[63] G. K. Tso and K. K. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy,* vol. 32, no. 9, pp. 1761-1768, 2007.

[64] J. P. Gouveia and J. Seixas, "Unraveling electricity consumption profiles in households through clusters: Combining smart meters and door-to-door surveys," *Energy and Buildings,* vol. 116, pp. 666-676, 2016.

[65] M. Santamouris, K. Kapsis, D. Korres, I. Livada, C. Pavlou, and M. N. Assimakopoulos, "On the relation between the energy and social characteristics of the residential sector," *Energy and Buildings,* vol. 39, no. 8, pp. 893-905, 2007.

[66] (30 July 2013). *Supplier reporting to Ofgem during the smart meter roll-out.*

[67] Ofgem, "The Retail Market Review -Implementation of Simpler Tariff Choices and Clearer Information," 27 August 2013.

[68] X. Tong, R. Li, F. Li, and C. Kang, "Cross-domain feature selection and coding for household energy behavior," *Energy,* vol. 107, pp. 9-16, 2016.

[69] D. Vercamer, B. Steurtewagen, D. Van den Poel, and F. Vermeulen, "Predicting consumer load profiles using commercial and open data," *IEEE Transactions on Power Systems,* vol. 31, no. 5, pp. 3693-3701, 2015.

[70] M. Aydinalp, V. I. Ugursal, and A. S. Fung, "Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks," *Applied energy,* vol. 71, no. 2, pp. 87-110, 2002.

[71] H. Christopher Frey and S. R. Patil, "Identification and review of sensitivity analysis methods," *Risk analysis,* vol. 22, no. 3, pp. 553-578, 2002.

[72] A. Saltelli, K. Chan, and M. Scott, "Sensitivity Analysis John Wiley & Sons publishers," *Probability and Statistics series*, 2000.

[73] S. Szymczak *et al.*, "Genetic association studies for gene expressions: permutation-based mutual information in a comparison with standard ANOVA and as a novel approach for feature selection," in *BMC proceedings*, 2007, vol. 1, no. 1: BioMed Central, p. S9.

[74] D. N. Gujarati, *Basic econometrics*. Tata McGraw-Hill Education, 2009.

[75] A. Agresti and M. Kateri, *Categorical data analysis*. Springer, 2011.

[76] H. R. Lindman, *Analysis of variance in complex experimental designs*. WH Freeman & Co, 1974.

[77] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.

[78] D. Johnson and S. Sinanovic, "Symmetrizing the kullback-leibler distance," *IEEE Transactions on Information Theory*, 2001.

[79] B. R. Sathyanarayana and G. T. Heydt, "Sensitivity-based pricing and optimal storage utilization in distribution systems," *IEEE Transactions on Power Delivery*, vol. 28, no. 2, pp. 1073-1082, 2012.

[80] R. Li, Q. Wu, and S. S. Oren, "Distribution locational marginal pricing for optimal electric vehicle charging management," *IEEE Transactions on power Systems*, vol. 29, no. 1, pp. 203-211, 2013.

[81] C. Brandstätt, G. Brunekreeft, and N. Friedrichsen, "Locational distribution network pricing in Germany," in *2011 IEEE Power and Energy Society General Meeting*, 2011: IEEE, pp. 1-3.

[82] F. Economic. (2014). *Regulatory arrangements for electricity network pricing.* [Online]. Available: https://www.nsw.gov.au/sites/default/files/miscellaneous/frontier_report_network_pricing_oct_14.pdf.

[83] P. M. Sotkiewicz and J. M. Vignolo, "Nodal pricing for distribution networks: efficient pricing for efficiency enhancing DG," *IEEE transactions on power systems*, vol. 21, no. 2, pp. 1013-1014, 2006.

[84] F. Li and D. L. Tolley, "Long-run incremental cost pricing based on unused capacity," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1683-1689, 2007.

[85] C. Gu, F. Li, and Y. Song, "Long-run network pricing to facilitate users' different security preference," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2408-2416, 2011.

[86] C. Gu, W. Yang, Y. Song, and F. Li, "Distribution network pricing for uncertain load growth using fuzzy set theory," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1932-1940, 2016.

[87] C. Gu and F. Li, "New development in distribution network pricing for revenue recovery in the Great Britain," in *2011 IEEE Power and Energy Society General Meeting*, 2011: IEEE, pp. 1-5.

[88] J. Pan, Y. Teklu, S. Rahman, and K. Jun, "Review of usage-based transmission cost allocation methods under open access," *IEEE transactions on power systems*, vol. 15, no. 4, pp. 1218-1224, 2000.

[89] Z. Jing, X. Duan, F. Wen, Y. Ni, and F. F. Wu, "Review of transmission fixed costs allocation methods," in *2003 IEEE Power Engineering Society General Meeting (IEEE Cat. No. 03CH37491)*, 2003, vol. 4: IEEE, pp. 2585-2592.

[90] J. Kumar and A. Kumar, "ACPTDF for Multi-transactions and ATC Determination in Deregulated Markets," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 1, no. 1, pp. 71-84, 2011.

[91] Ofgem. (2009). *Delivering the electricity distribution structure of charges project: decision on extra high voltage charging and governance arrangements* [Online]. Available: http://www.ofgem.gov.uk/Networks/ElecDist/Policy/DistChrgs/Pages/D.

[92] U. P. Networks. (2015). *Low Carbon London Programme website* [Online]. Available: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households.

[93] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of low voltage network templates—Part I: Substation clustering and classification," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3036-3044, 2014.

[94] C.f.E, "Electricity Smart Metering Customer Behaviour Trails (CBT) Findings Report," 16 May 2011.

[95] (2017). *Competition in UK electricity markets.*

[96] (2017). *State of the energy market.*

[97] E. D. Varga, S. F. Beretka, C. Noce, and G. Sapienza, "Robust real-time load profile encoding and classification framework for efficient power systems operation," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1897-1904, 2014.

[98] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 381-387, 2003.

[99] C. Chen, J. Hwang, and C. Huang, "Application of load survey systems to proper tariff design," *IEEE Transactions on Power Systems*, vol. 12, no. 4, pp. 1746-1751, 1997.

[100] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420-430, 2014.

[101] Ofgem, "Updated Household energy bills explained," 2013.

[102] C. Beckel, L. Sadamori, and S. Santini, "Towards automatic classification of private households using electricity consumption data," in *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2012: ACM, pp. 169-176.

[103] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft," in *2011 IEEE/PES Power Systems Conference and Exposition*, 2011: IEEE, pp. 1-8.

[104] P. S. Bradley and O. L. Mangasarian, "Feature selection via concave minimization and support vector machines," in *ICML*, 1998, vol. 98, pp. 82-90.

[105] I. Jolliffe, *Principal component analysis*. Springer, 2011.

[106] I. Kononenko, "Estimating attributes: analysis and extensions of RELIEF," in *European conference on machine learning*, 1994: Springer, pp. 171-182.

[107] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine learning*, vol. 53, no. 1-2, pp. 23-69, 2003.

[108] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE transactions on neural networks*, vol. 12, no. 2, pp. 181-201, 2001.

[109] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of electricity consumption behavior dynamics toward big data applications," *IEEE transactions on smart grid*, vol. 7, no. 5, pp. 2437-2447, 2016.

[110] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Aaai*, 1992, vol. 2, pp. 129-134.

[111] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of biomedical informatics,* vol. 85, pp. 189-203, 2018.

[112] T. B. Trafalis and H. Ince, "Support vector machine for regression and applications to financial forecasting," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, 2000, vol. 6: IEEE, pp. 348-353.

[113] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Applied energy,* vol. 141, pp. 190-199, 2015.

[114] G. J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, "Two-stage pattern recognition of load curves for classification of electricity customers," *IEEE Transactions on Power Systems,* vol. 22, no. 3, pp. 1120-1128, 2007.

[115] S. Ramos, Z. Vale, J. Santana, and J. Duarte, "Data mining contributions to characterize MV consumers and to improve the suppliers-consumers settlements," in *2007 IEEE Power Engineering Society General Meeting*, 2007: IEEE, pp. 1-8.

[116] Eurostat. (2017, 10 July 2018). *Household consumption by purpose* [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php/Household_consumption_by_purpose.

[117] Z. Wu, S. Zhou, J. Li, and X.-P. Zhang, "Real-time scheduling of residential appliances via conditional risk-at-value," *IEEE Transactions on Smart Grid,* vol. 5, no. 3, pp. 1282-1291, 2014.

[118] (2010). *Demand side response: a discussion paper*.

[119] IEA, "Implementing Agreement on Demand-Side Management Technologies and Programmes-2017 Annual Report," 2018.

[120] C. Belgium, "Impact Assessment Study on Downstream Flexibility, Price Flexibility, Demand Response & Smart Metering," ed: COWI Belgium, 2016.

[121] I. Benítez, A. Quijano, J.-L. Díez, and I. Delgado, "Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers," *International Journal of Electrical Power & Energy Systems,* vol. 55, pp. 437-448, 2014.

[122] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Load pattern-based classification of electricity customers," *IEEE Transactions on Power Systems,* vol. 19, no. 2, pp. 1232-1239, 2004.

[123] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *IEEE Transactions on power systems,* vol. 20, no. 2, pp. 596-602, 2005.

[124] S. V. Verdú, M. O. Garcia, C. Senabre, A. G. Marin, and F. G. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," *IEEE Transactions on Power Systems,* vol. 21, no. 4, pp. 1672-1682, 2006.

[125] C.-K. Woo, R. Li, A. Shiu, and I. Horowitz, "Residential winter kW h responsiveness under optional time-varying pricing in British Columbia," *Applied energy,* vol. 108, pp. 288-297, 2013.

[126] S. Fan and R. J. Hyndman, "The price elasticity of electricity demand in South Australia," *Energy Policy,* vol. 39, no. 6, pp. 3709-3719, 2011.

[127] C. Bartusch and K. Alvehag, "Further exploring the potential of residential demand response programs in electricity distribution," *Applied Energy*, vol. 125, pp. 39-59, 2014.

[128] J. Torriti, "Price-based demand side management: Assessing the impacts of time-of-use tariffs on residential electricity demand and peak shifting in Northern Italy," *Energy*, vol. 44, no. 1, pp. 576-583, 2012.

[129] G. R. Newsham, B. J. Birt, and I. H. Rowlands, "A comparison of four methods to evaluate the effect of a utility residential air-conditioner load control program on peak electricity use," *Energy Policy*, vol. 39, no. 10, pp. 6376-6389, 2011.

[130] Y. He, B. Wang, J. Wang, W. Xiong, and T. Xia, "Residential demand response behavior analysis based on Monte Carlo simulation: The case of Yinchuan in China," *Energy*, vol. 47, no. 1, pp. 230-236, 2012.

[131] R. D'hulst, W. Labeeuw, B. Beusen, S. Claessens, G. Deconinck, and K. Vanthournout, "Demand response flexibility and flexibility potential of residential smart appliances: Experiences from large pilot test in Belgium," *Applied Energy*, vol. 155, pp. 79-90, 2015.

[132] T. Yu, D. S. Kim, and S.-Y. Son, "Optimization of scheduling for home appliances in conjunction with renewable and energy storage resources," *Int. J. Smart Home*, vol. 7, no. 4, pp. 261-272, 2013.

[133] W. Labeeuw, J. Stragier, and G. Deconinck, "Potential of active demand reduction with residential wet appliances: A case study for Belgium," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 315-323, 2014.

[134] P. Finn, M. O'connell, and C. Fitzpatrick, "Demand side management of a domestic dishwasher: Wind energy gains, financial savings and peak-time load reduction," *Applied energy*, vol. 101, pp. 678-685, 2013.

[135] A. Nilsson, C. J. Bergstad, L. Thuvander, D. Andersson, K. Andersson, and P. Meiling, "Effects of continuous feedback on households' electricity consumption: Potentials and barriers," *Applied Energy*, vol. 122, pp. 17-23, 2014.

[136] J. S.-B. Wilma Mert, Wibke Tritthart, "Consumer acceptance of smart appliances. D5.5. of WP5 from the EIE smart-A project," 2008.

[137] E. Dütschke and A.-G. Paetz, "Dynamic electricity pricing—Which programs do consumers prefer?," *Energy Policy*, vol. 59, pp. 226-234, 2013.

[138] S. Gyamfi and S. Krumdieck, "Price, environment and security: Exploring multi-modal motivation in voluntary residential peak demand response," *Energy Policy*, vol. 39, no. 5, pp. 2993-3004, 2011.

[139] S. Hille, S. Weber, and T. Brosch, "Consumers' preferences for electricity-saving programs: Evidence from a choice-based conjoint study," *Journal of Cleaner Production*, vol. 220, pp. 800-815, 2019.

[140] M. Nicolson, G. Huebner, and D. Shipworth, "Are consumers willing to switch to smart time of use electricity tariffs? The importance of loss-aversion and electric vehicle ownership," *Energy research & social science*, vol. 23, pp. 82-96, 2017.

[141] A. Kollmuss and J. Agyeman, "Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior?," *Environmental education research*, vol. 8, no. 3, pp. 239-260, 2002.

[142] F. Friis and T. H. Christensen, "The challenge of time shifting energy demand practices: Insights from Denmark," *Energy Research & Social Science*, vol. 19, pp. 124-133, 2016.

[143] SEAI. (2012). *Press Release – Full Data from National Smart Meter Trial* [Online]. Available: http://www.seai.ie/News_Events/Press_Releases/2012/National_Smart_Meter_Trial_Data_ Release.pdf.

[144] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157-1182, 2003.

[145] M. Dash and H. Liu, "Feature selection for classification," *Intelligent data analysis,* vol. 1, no. 1-4, pp. 131-156, 1997.

[146] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR),* vol. 31, no. 3, pp. 264-323, 1999.

[147] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Annals of Data Science,* vol. 2, no. 2, pp. 165-193, 2015.