The Impact of Demographics on Residential Electricity Usage

THE IMPACT OF DEMOGRAPHICS ON RESIDENTIAL ELECTRICITY USAGE

BY

TIANYAO ZHOU, B. A. Finance; Master of Finance

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS AND STATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Master of Science (2016) McMaster University
(Department of Mathematics and Statistics) Hamilton, Ontario, Canada

TITLE: The Impact of Demographics on Residential Electricity Usage

AUTHOR: Tianyao Zhou
B. A. Finance
Master of Finance

SUPERVISOR: Dr. Dean Mountain

NUMBER OF PAGES: x, 170

Abstract

The electricity consumption efficiency in the residential sector is commonly discussed in previous studies. Over the previous studies, different factors influencing electricity consumption have been covered, including economic factors, lifestyle and demographic factors, climate and environmental factors and technological development. With respect to estimation methodologies in these studies, there are three methods existing—conditional demand analysis, neural network and engineering method.

A significant amount of information for my thesis is drawn from the collaborative project involving McMaster University and Hydro One. My thesis mainly focuses on residential electricity consumption efficiency and the relationship between the total electricity consumption and a number of variables, including dwelling information, time-of-use prices, weather data and demographic factors. I am particularly interested in the influence of demographics. The data sources of variables include four categories---dwelling and household information, consumption data, weather data and price data in 2013. In my regression estimation, I include four systems components—heating system, water heating system, cooling system and other appliances system. Each system has its own error term. I discuss and estimate models where the error terms are correlated and uncorrelated. Seven versions of models are discussed with different combinations of variables in the model and variables in the variance model of the errors. I choose a final model after conducting Wald hypothesis tests. Finally, I list a table of illustrative examples explaining the influence of demographic factors---education distribution level, age distribution level and number of residents on electricity usage. From the results, I can conclude that education distribution level exerts a very significant impact on total electricity consumption.

Acknowledgements

I take this opportunity to express my deepest gratitude to my extraordinary supervisor Dr. Dean Mountain. He gives me help, guidance and constructive advice with his patience and kindness in my thesis. I really appreciate his time and energy he has spent on me and my thesis.

Also I want to thank my friends Feng Tian and So HonYiu for their generous support and positive energy when I need help regarding study and life.

My final acknowledgement is to my dear parents, for their love and understanding over these years. They provide me more confidence and opportunities to experience more in my academic and personal life.

Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Overview

There is no denying that energy is a very important aspect for Canadian society. Residential energy occupies a significant part in all energy consumption. Canadian households use energy for space heating, water heating, cooling and other appliances such as computers and lighting. Since the residential energy consumption is increasing gradually, the greenhouse gas emissions and air pollutants are also increasing, which will lead to a significant impact on the environment. Households may choose to reduce their energy use by adjusting their conservation measures. The residential energy consumption depends on different factors, such as household, climate, fuel price, house characteristics, appliance ownership, customer behaviour and household demographics. Here, I consider heating degree hours and cooling degree hours as the climate in my model. Regarding fuel price, I will introduce a concept called time-of-use price in the later chapter. House characteristics include the house age, living space and the number of rooms. Appliance ownership refers to appliances such as lights or computers consumption. With respect to customer behavior, households may change their electricity usage habits due to price changes. For example, they may improve their house equipment and dwelling changes like installing flow restricting shower heads. Demographic factors focus on education level distribution, age level distribution and number of residents per household. Explicit explanation of demographic factors' impacts in my studies has played a significant role in other researches' model planning and variables discussion.

This thesis draws from the project titled "Integrating Dynamic Pricing & Customer Feedback on Electricity Usage to Stimulate Residential Conservation & Demand Response", which concentrates on evaluating participants' behavioral response to different electricity pricing plans and various levels of customer-specific electricity usage data categorized by various equipment and appliances. Field experimental data is provided by Hydro One. According to the customers' list provided by Hydro One, McMaster's energy research group has constructed a pilot questionnaire titled "Residential Energy Pilot Questionnaire". For my thesis, this pilot questionnaire involving 978 households in Ontario focuses on four sections, home, household equipment, conservation actions and household demographics and provides a component in the statistical basis for input into future rate design options and feedback mechanisms. This thesis concentrates on research about the customers’ behavior, electricity usage and analysis of demographic factors influencing residential energy consumption.

1.2 Objectives

The main objective of this type of research is to provide reliable information for management departments in the energy companies to plan sustainable electricity supply to residential sector properly. In the meantime, the research of residential electricity demand is meaningful for the government to make the policy to control green gas emission, balance the ecological development and promote the economic growth. With this as background, there are four important goals for my research. Firstly, it is based on the survey data collected from random customers’ response. In this way, this thesis possesses the property of realistic meaning for Ontario and Canada. Secondly, I conduct research on discovering the relationship between households’ electricity consumption and four sections, including house characteristics, house equipment, conservation actions and home demographics respectively. Thirdly, from a statistical perspective, I consider separating a big model into four sub-models and include error terms between sub-models. Also, I will discuss correlation and no correlation exists among error terms in my model. The last but not least, my thesis emphasizes on exploring the relationship between the electricity consumption and demographic factors, especially the age distribution, education level within a household, one household’s annual income and number of residents. The results will help shed light on improving energy efficiency problems related to social and economic aspects.

Residential electricity consumption in one household is effected by various elements such as living space size, the number of finished rooms, the age of heating or cooling system, number of residents in one household, time-of-use price, heating or cooling degree hours. Considering the role of demographic shifts, income change, education distribution changes, age distribution changes and number of residents change to reduce electricity consumption, my thesis model will help predict the corresponding electricity consumption. In this way, it is helpful for companies and government to make proper policies contributing to energy savings.

1.3 Organization

The organization of my thesis is categorized into 6 chapters. Chapter 2 includes a literature review of previous theoretical studies and works related to my thesis background. Chapter 3 introduces the general modeling framework, comprising the theoretical background and general structure of my model. Chapter 4 discusses the collection of data sources, including dwelling information, consumption data, weather data and price data for my model. Chapter 5 presents the model specification and estimation methodology. In this part, I will explain the variables, model estimation and different versions of statistical assumptions. Chapter 6 describes the results based on corresponding versions of statistical assumptions discussed in Chapter 5. I conduct χ^2 tests on demographic factors as groups, compare illustrative examples using my final model results regarding demographic factors, and compare my study with previous studies in the literature review from the perspective of theoretical background, explanatory factors, analysis methodology and error term specifications.

Chapter 2 Literature Review

2.1 Overview

Regarding the residential electricity demand, many studies have been conducted by researchers from different countries, where the objectives of the studies also vary by different methodologies and structure of models. In this chapter, I will review some previous studies related to this field. Section 2.2 introduces the topics of previous studies on the residential electricity consumption—economics, demographic, climate and technological factors. In section 2.3, I focus on discussing different methodologies on electricity consumption—conditional demand analysis, neural network and engineering method. Section 2.4 will give a summary for chapter 2.

2.2 Topics of previous studies

In this section, I will introduce topics of previous studies. After summarizing the papers reviewed, I categorize these papers into 4 different groups—Economic factors, demographic factors, climate factors and technical development.

Some studies have illustrated that the residential electricity is related to economic factors. In Great Britain, Houthakker (1951) studied electricity consumption and he discovered that electricity consumption was influenced by electricity price and household's annual income using a statistical model. In addition, he included the cost per kilowatt as the price into the demand function. In 1980, Hsiao and Mountain (1985) conducted a study to estimate the short-run income electricity in a conditional demand for electricity model. In their study, they took account of the special nature of the data by recording income variable in categorical form. A similar study to Houthakker (1951) was conducted by Nesbakken (2001) who found that the annual income, the cost of space heating, the capital expense and operation charges exerted an influence on residential electricity demand In the 1980s and 1990s, Nässén, Sprei and Holmberg (2008) implemented a model based on econometrical analysis and interviews to the residential sector. They obtained the conclusion that the energy price exerted an important impact on the energy usage. Nair et al. (2010) conducted an investigation study of effects on energy consumption efficiency based on 3000 households in 2010. The residential electricity consumption savings measures were determined by individual elements, including income, demographic factors, education level of household, the age of house, comfortable thermal energy and cost of electricity. Another significant factor related to economics is the price of electricity referred by Leighty and Meier (2011) .

A second group of studies mainly focused on the lifestyle and demographic factors. Hass (1997) focused on a methodological research of energy efficiency indexes in order to investigate and produce a series of key factors used to compare the indicators of multi-country energy usage efficiency in 1997. The key factors were life-style and demographic elements (Zhang, 2004) The type of energy use, house's address and the house's attitude to the electricity consumption were also considered as elements influencing energy consumption in the residential sector in Jordan in 2002 by Jaber. The conclusion of that paper is that improving the heat preservation and encouraging the usage of reproducible energy may exert a huge effect on the usage technology of energy (Jaber, 2002) . Through the study by Brounen et al. in 2012, they found that electricity consumption was determined by the structural characteristics of the house, such as house type, house characteristics (Brounen et al., 2012). Social psychology and individual behavioral changes may stimulate the changes of residential electricity consumption (McMakin, 2002) . McMakin proceeded a study of incentive factors on residential electricity consumption by people in 2002. This study showed that a model based on social psychology may exert emphasis on individual behavioral changes. The behavior changes of residents were emphasized on the papers from Gyberg and Palm in 2009 (Gyberg, 2009) and Ouyang and Hokao in 2009. In 2011, Kelly conducted a research of residential energy consumption. He illustrated that is a complicated problem to combine society and technology, including physical, demographic and behavior's characteristics of house and house owner (Kelly, 2011). Yu et al. proceeded a study about the effect of consumers' behavior changes towards residential energy consumption by data mining technique. The results show that the residents' behavior changes could exert an influence on residential electricity consumption savings (Yu et al., 2011). In 2012, Hiller conducted a study on energy consumption in the residential sector, especially the relationship between the consumers' behaviors and the electricity consumption, focusing on 57 single houses in Sweden (Hiller, 2012). More specifically, Wall and Crosbie focused on the lighting energy consumption of 18 residential households over one week in UK in 2009. They summarized that the households could save their electricity consumption by changing their choice of lighting bulbs to a great degree (Wall, 2009).

A third group of studies focused on the influence of climate diversity and environmental elements on the energy consumption. Yang et al. (2010) implemented an evaluation of residential buildings' energy efficiency based on climate diversity and relevant indicators of energy efficiency. In this study, the author considered the following comprehensive elements: the design of the building, the property of the building structure, the facilities of energy-savings, the operation management of the building, and comfortable and healthy environment. Chedid and Ghjar (2004) focused on study of emphasizing on the thermal characteristics and energy consumption equipment.

The fourth group of study focusing on technological development was conducted by Sadineni et al. (2011). Their study indicates that by installing effective upgrades, one household may consume less 42.5% electricity of total amount.

2.3 Methodology of previous studies

In this section, the methodology of previous studies is introduced. Regarding methodology, I will discuss 3 different categories—Conditional Demand Analysis, Neural Network and Engineering Method.

Conditional demand analysis (CDA) model is a combination of energy consumption and appliance and demographic survey, consumption and weather data. This method aims to explore the relationship between the residential total electricity consumption and the factors of influencing energy efficiency by statistical regression methodology. Here, the factors can be weather data, house characteristics, equipment characteristics or demographic characteristics. This method's advantages are reflected on the veracity of regression results and direction of research analysis. Nowadays, this method has been widely accepted for the research on residential electricity consumption. The first paper was written by Parti M. and C. Parti (1980). They collected the monthly and annual household electricity consumption and usage of each appliance applied in model. Then, the statistical regression model was built between these variables. In 1984, a study was conducted by Aigner, Sorooshian and Kerwin, (1984) with 24 equations in a regression model explaining the relationship between variation in the time-averaged load (averaged over days) and the size of their house, the indoor temperature and several binary indexes. Similar studies were also implemented by Goldfarb and Huss (1987), by Rosa (1989) and by Newsham and Donnelly (2013) respectively. In their construction process, the researchers considered elements from different areas: society, economics, technology and physical resources, even the electricity consumption for each appliance. In addition to the fixed factors in model, Fiebig and other two researchers (1991) focused on making the coefficients of the dummy variables random instead of fixed. In this way, a structure for the heteroskedasticity has been considered in the data. In order to estimate residential electricity end-use load profiles, Hsiao, Mountain, Illman (1995) used conditional demand analysis approach to model specific load profiles using a larger database. Bartels and Fiebig (1996) sought information on 16 different end uses and eight metering channels for each household. They used the econometric model to estimate the end-use load curves by applying CDA with the readings on all the metered end uses. In most of the studies, the authors considered the combination or the comparison of two or three methods experimenting on the effects of different factors on the residential electricity consumption. In this way, the various variables could be taken into account the aggregate electricity consumption model in a more comprehensive scope. Hsiao, Mountain and Illman used a Bayesian approach where priors are based on specific end-use mentoring and taken account of household-specific information. Caves, Herriges, Train and Windle (1987) used the conditional demand analysis model to obtain the end-use electricity load profile firstly and used the Bayesian analysis approach to obtain the posterior distribution by modifying the engineering priors. This is a good example of combination of the engineering method and conditional demand analysis models. In 1990 and 2000, Bartels and Fiebig combined data-mining and conditional demand analysis to transfer the traditional conditional demand analysis into a new pattern of conditional demand analysis model (Bartels, 1990) (Bartels, 2000). Some papers used the conditional demand analysis models to focus on technological development and customers' behavior changes (LaFrance, 1994). When Bernard and Lacroix used the conditional demand analysis model, they obtained some negative coefficients. They wanted to check if the model was too limited to the residential uniform end-use consumption. Through the data they obtained from study in 1995, they found that the hypothesis of uniformity was accepted excluding the electricity water heating. In this way, they could conclude that households regard the electricity as their main source for their space heating (Bernard and Lacroix, 2005). This method is simple to build, apply and does not require a lot of information with details. It can also include the influence of social and economic factors on electricity consumption in the model. However, this method needs a large database in order to minimize multicollinearity of explanatory factors and provides limited flexibility in analysis of impacts of conservation measures on electricity consumption.

Neural network (NN) is a simplified mathematical model originating from biological neural networks. This is a method to estimate the relationship between the residential electricity consumption and a large number of parameters in residential sector. Due to the simple programming and accurate estimation, neural network has been applied into estimation problems. It is generally used to estimate the load profiles and forecast the consumption of individual architecture in non-linear models. D.Datta and S.A. Tasson (D.Datta and S.A. Tasson) focus on the performance of a neural network in the prediction of electricity demand in a market. This study shows that neural network is a technological tool to forecast the overall consumption of the store with respect to time of the day and environmental conditions. Aydinalp and two persons (2004) used neural network applied into energy consumption level model for end-users who resided in Canada. They conclude that the scope of their neural network model in variables was very restricted. Another study (2002) focused on an estimation model of developing the residential equipment, lighting and space-cooling by neural network models. Even though neural network method can be applied simply to obtain the accurate estimates of social and economic factors, it is not flexible in estimating the impacts of conservation measures. Also, precise interactions and forms of causality are difficult to be covered with neural networks.

The Engineering method is the unique one that can develop a model of energy consumption with no historical consumption information. For example, if we need to estimate the space heating, we can use the heating degree days or the heat transfer on the end-uses based on the climate. Since this method is highly dependent on the physical characteristics (such as level of insulation and fuel use for water heating, which are normally considered long-term investments), it is regarded as the model which has the highest flexibility and ability under no historical information. Farahbakhsh (1998) conducted a study on residential electricity consumption for Canada. He introduced the engineering method to estimate the total energy consumption model. In his paper, he presented the model, the features of energy consumption in Canada, and the influence of energy’s reduction on aggregate energy consumption. Swan and Ugursal (2009) provided an investigation of various models applied into residential energy consumption, including top-down and bottom-up approaches. Top-down method emphasized on historic total electricity consumption emphasizing the indexes such as gross domestic product, inflation and energy price. However, the bottom-up method explored the individual electricity consumption at the regional and national levels. This method requires a large database with household information and usage situation. The disadvantage of this method is that it is hard to estimate the customers’ behavior changes affecting the electricity consumption. Often the source of the engineering method is based on small samples. Furthermore, engineering estimated often don’t take account of behavioral and economic factors.

Another group of authors mainly focused on estimating the residential energy model at the national level by illustrating three methods: engineering method, neural network and conditional demand analysis (Aydinalp, 2003) (Aydinalp-Koksal, 2008). The comparison of these three methods stated that the conditional demand analysis obtained a more accurate result. The conditional demand analysis and neural network method could estimate the effects of social and economic elements.

The studies of different factors influencing on residential electricity consumption and different methods are introduced. Based on the above analysis, we can now summarize - the advantages and disadvantages as follows (Table 2.1):

Methods Ad/Disad	EM (Engineering Method)	NN (Neural Network)	CDA (Conditional Demand Analysis)
Advantages	(1) Develop model without historical consumption data.	(2) Simple application; (3) Accurate estimation of end-use energy consumption and impact of social and economic measure.	(1) Easier to develop and use; (2) No requirements of large details of data; (3) Consideration of social and economic factors.
Disadvantages	(1) Require much information of house description data; (2) Difficult to consider the effects of consumers' behavior changes.	(1) Not flexible in estimating the impacts of conservation measures. (2) precise interactions and forms of causality are difficult to be uncovered	(3) Database needs to be larger; (4) Not flexible in estimating the influence of conservation measures.

Table 2.1: Comparison of the Engineering Method, Neural Network and Conditional Demand Analysis

2.4 Summary

This chapter provided the list of previous studies on the residential electricity consumption. Section 2.2 focused on the introduction of different topics discussed regarding residential electricity consumption. Different topics mainly included economic elements, demographic elements and customer behavior changes, climate and environmental elements and technological development. In section 2.3, three methodologies were summarized based on previous studies, including conditional demand analysis, neural network and engineering method. At the end, a table of comparison of advantages and disadvantages of these three methodologies was listed. Eventually, I will provide a comparison between my model and previous literature in Chapter 6.

Chapter 3 General Modeling Framework

3.1 Overview

In this chapter, I will state the drawbacks and limitations of existing models as I discussed in previous chapter in section 3.2. In section 3.3, I will build my own model based on the advantages and disadvantages of existing models. The introduction of structure and related theoretical background will be brought into later. Section 3.4 will illustrate a summary of this chapter.

3.2 Limitation of existing models

For the purpose of this thesis, I need to build a model that demonstrates a relationship between residential electricity consumption and a number of variables with focus on how demographic influence electricity consumption. Existing models have both benefits and limitations.

(1) Partial explanation:
Some studies have assumed that the residential electricity consumption is related to economic factors, such as the price of the electricity and annual income for one household [Houthakker (1951), Nesbakken (2001), Nässén(2008), Nair(2010), Leighty(2011)]. Undoubtedly, residential electricity consumption is significantly related to the price change. However, other factors cannot be ignored. Similarly, a second group of studies only focused on the lifestyle and demographic factors [Haas(1997), Zhang(2004), Jaber(2002), Brounen(2012), McMakin(2002), Gyberg(2009), Ouyang(2009), Kelly(2011), Yu(2011), Hiller(2012), Wall(2009)]. These studies mainly concentrate on discussing how the customers' behavior changes influence aggregate electricity consumption. A third group of studies refer to climate diversity and environmental elements on the energy consumption [Yang(2010),Chedid(2004)]. Another study conducted by Sadneni et al. in 2011 states that the technological development may influence residential electricity consumption and energy savings.

(2) Too many information requirements:
Although the engineering method is the only one that can develop a model of energy consumption without requiring historical consumption information, this method requires a large amount of information of households and the usage of each

appliance. These models are usually based on small samples and do not account for behavioral responsiveness [Farahbakhsh(1998)].

(3) Not flexible in considering conservation measures:
Neural network is a method to determine the relationship between the residential electricity consumption and a large number of variables [D.Datta, Aydinalp(2004), Aydinalp(2002)]. However, the problems with this method are the lack of consideration of conservation factors, which are important to the variation of residential electricity consumption, and the difficulty in identifying structured causes.

3.3 Setup of my model

In this section, I will describe my model by combing the advantages of existing models and avoiding the shortcomings of previous studies. I will emphasize the important points regarding to following aspects:

(1) Combination of more variables: I will not only account for climate and environmental factors, usage of each appliance, the price of electricity and annual income for households, but also include consumers' adoption of conservation measures;

(2) Based on annual electricity consumption instead of daily: This is a straightforward way to collect information because we only need the whole year's consumption in 2013 rather than request for monthly or even daily. This model is much easier to develop because it does not depend on details of distribution of aggregate electricity consumption;

(3) Flexibility in analysis of conservation measures: Adoption of conservation actions related to energy usage are a significant part which cannot be ignored in my model. I will add this part into my model and analyze if the conservation measures are significant to the whole model and specify the details in the chapter 5.

(4) Explicit accounting for Demographics: For household electricity usage, I will allow for demographic factors to influence electricity consumption.

3.3.1 Background and Theory

(1) General Conditional Demand Analysis Model:

This model is based on conditional demand analysis model illustrating that the total electricity consumption for one household is the sum of end-use consumption for each appliance plus a residual or error term for each appliance. The whole model is a regression model. As Caves and Herriges [Caves(1987)] mentioned, If y_{it} represents the total electricity usage for consumer i during time t, then we can obtain by conditional demand analysis definition:

$$y_{it} = \sum_{j=1}^{M} f_{jt}(Z_{ijt})\, D_{ij} + \epsilon_{it}$$

where t =1, 2,…, T and i =1, 2, …,N.

D_{ij} = 1 if customer i owns appliance j or 0 otherwise;

Z_{ijt} represents the variables that affect customer i's utilization of appliance j at time t;

M and ε_{it} shows the number of appliances and random variation respectively;

$f_{jt}(Z_{ijt})$ provides the contribution of appliance j to total usage at time t for a given set of conditions represented by the Z_{ijt}.

(2) Simplified Conditional Demand Analysis Model:

The general conditional demand analysis model has been introduced above, it may be hard to find the contribution of appliance j to total usage at time t for a given set of conditions represented by the Z_{ijt}. Here, we simplify the $f_{jt}(Z_{ijt})$ to be a constant, that is β_{jt}. Therefore, the model is shown as below:

$$y_{it} = \sum_{j=1}^{M} \beta_{jt} D_{ij} + \epsilon_{it}$$

where t =1, 2,…, *T* and i =1, 2, …,*N*.

D_{ij} = 1 if customer i owns appliance j or 0 otherwise;

The ϵ_{it} ‘s are assumed to have the following distributional properties:

E(ϵ_{it}) = 0 , t =1, 2,…, *T* and i =1, 2, …,*N.*

Cov($\epsilon_{it}, \epsilon_{js}$) = σ_{ts} if $i = j$, 0 otherwise.

β_{jt} represents the average electricity usage of appliance j during time period t .

(3)My model without considering the time period:

For my model, I focus on the total electricity consumption for the whole year of 2013, which does not describe the period (hourly, daily or monthly). I do not need the time t comparing the previous model as discussed above. In addition, due to the large value of annual electricity consumption, I choose Iny_i as the dependent variable instead of y_i. (the base model is assumed to be $y_i = e^{\sum_{j=1}^{4} f_j(Z_{ij}) D_{ij} + \epsilon_i}$)

Thus, my model can be written as follows:

$$Iny_i = \sum_{j=1}^{4} f_j(Z_{ij})D_{ij} + \epsilon_i$$

Where Iny_i represents the aggregate electricity consumption by household i

for a period of time (here, yearly);

D_{ij} is an indicator representing the electricity-using system j by household i;

for example, D_{1j} could represent the dummy variable corresponding to ownership of an electric heating system;

$f_j(Z_{ij})$ represents the function of sub-model j by household i ;

ϵ_i is the random noise or error term by household.

For more details regarding the model, see Figure 3.1. This model illustrates that heating or cooling will be influenced by weather conditions and square footage of house. Behavioral responsiveness can be influenced by prices, demographic factors and adoption of conservation measures.

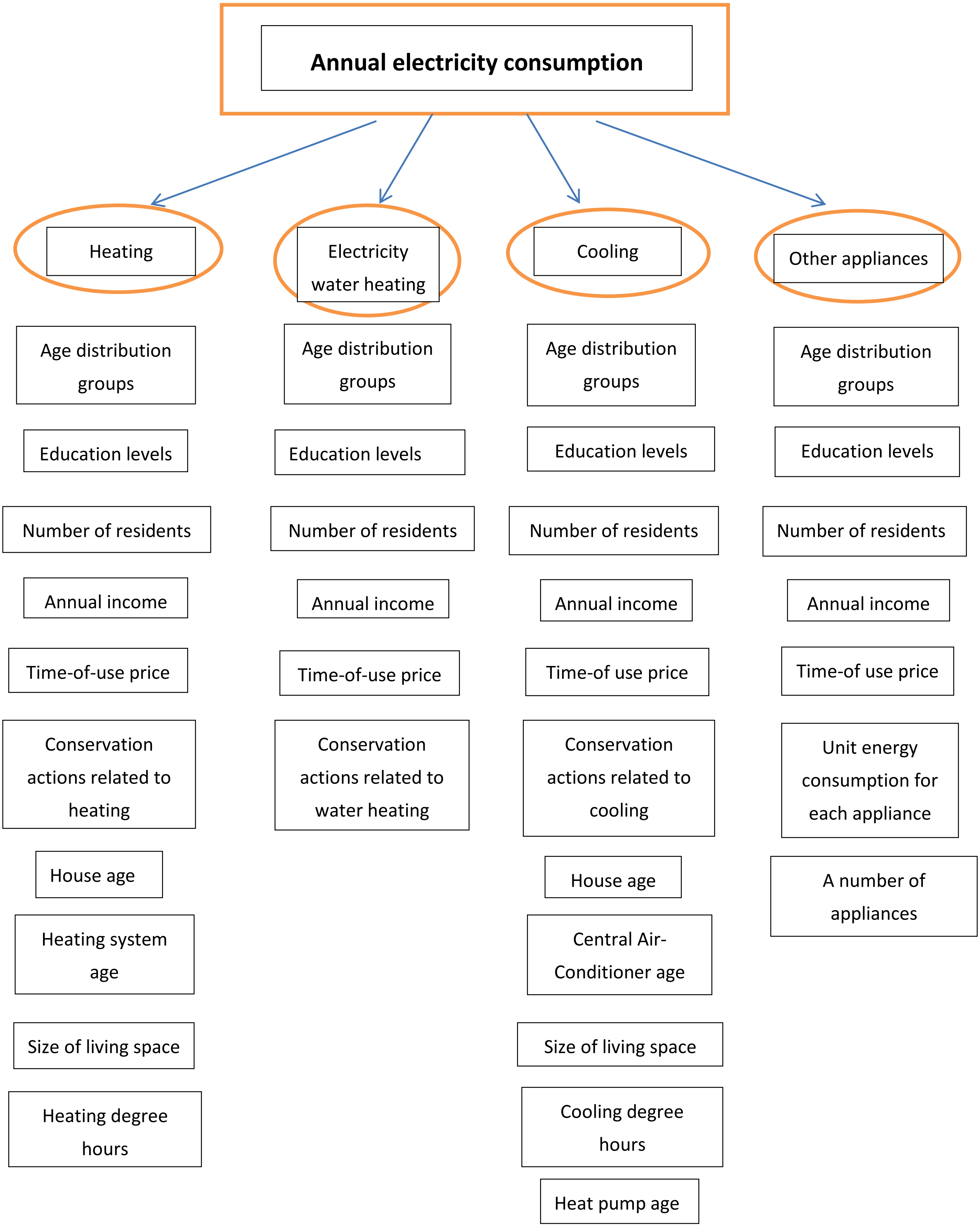


Figure 3.1 Structure for model

From the figure above, I can further describe variables in five parts, including common explanatory variables, heating explanatory variables, water heating explanatory variables, cooling explanatory variables and other appliances explanatory variables.

First of all, the common variables are listed in the following table: (Table 3.1) I also introduce abbreviation or symbols representing groups of variables.

Abbreviation or symbol	Description
PEL	A sum of percentages of each category in education level for one household
AGEDIST	A sum of percentages of each category in age distribution level for one household
NR	Total number of residents in one household
INC	Annual household income before taxes
TUP	Time-of-use prices

Table 3.1Common variables for heating, water heating, cooling and other appliances systems

Secondly, variables under heating system: (Table 3.2)

Abbreviation or symbol	Description
HDH	Heating degree hours
SIZE	Size of living space
AGE	Age of house
HA	Age of heating system
NHS	Number of conservation measure related to heating system, i.e. sum of D_{UW}, D_{IC}, D_{RAS}, D_{PB}.
D_{UW}	Dummy variable if upgrading the windows to reduce the electricity. Equals 1 if done, 0 otherwise.
D_{IC}	Dummy variable if insulating the ceilings, floors or walls. Equals 1 if done, 0 otherwise.
D_{RAS}	Dummy variable if retrofitting air-sealing. Equals 1 if done, 0 otherwise.
D_{PB}	Dummy variable if installing programmable thermostat for baseboards. Equals 1 if done, 0 otherwise.

Table 3.2 Variables for heating system

Thirdly, variables under water heating system: (Table 3.3)

Abbreviation or symbol	Description
NWH	Number of conservation measure related to electric water heating, sum of D_{FRSH},D_{WHI},D_{RDOF},D_{CWL}, and D_{HLO}.
D_{FRSH}	Dummy variable for the installing flow restricting shower heads. Equals 1 if there is one, 0 otherwise.
D_{WHI}	Dummy variable for installing water heater insulation. Equals 1 if there is one, 0 otherwise.
D_{RDOF}	Dummy variable for running dishwasher only when full. Equals 1 if there is one, 0 otherwise.
D_{CWL}	Dummy variable for using cold water for laundry. Equals 1 if there is one, 0 otherwise.
D_{HLO}	Dummy variable for hanging laundry outside or on a rack to dry. Equals 1 if there is one, 0 otherwise.
DTF	The wattage sum of the difference between “with Electricity Water Heating” and “without Electricity Water Heating” for electric appliances Dishwasher, Top Load Washing and Front Load Washing

Table 3.3 Variables for water heating system

Fourthly, variables under cooling system: (Table 3.4)

Abbreviation or symbol	Description
CDH	Cooling degree hours
SIZE	Size of living space
AGE	Age of house
ACHP	Age of central air-conditioner or heat pump or windows or other
NCS	Number of conservation measure related to cooling system, sum of D_{UW}, D_{IC} and D_{RAS}.
D_{UW}	Dummy variable if upgrading the windows to reduce the electricity. Equals 1 if done, 0 otherwise.
D_{IC}	Dummy variable if insulating the ceilings, floors or walls. Equals 1 if done, 0 otherwise.
D_{RAS}	Dummy variable if retrofitting air-sealing. Equals 1 if done, 0 otherwise.

Table 3.4 Variables for cooling system

Fifthly, variables under other appliances system: (Table 3.5)

Abbreviation or symbol	Description
NOA	Number of conservation measure related to other appliances system, sum of D_{RDOF},D_{RUA},D_{CEA},D_{TOL}, and D_{HLO}.
D_{RDOF}	Dummy variable if running dishwasher only when full. Equals 1 if done, 0 otherwise.
D_{RUA}	Dummy variable if reducing use of appliances. Equals 1 if done, 0 otherwise.
D_{CEA}	Dummy variable if controlling any equipment or appliances. Equals 1 if done, 0 otherwise.
D_{TOL}	Dummy variable if turning off lights when not in use. Equals 1 if done, 0 otherwise.
D_{HLO}	Dummy variable if hanging laundry outside or on a rack to dry Equals 1 if done, 0 otherwise.

Table 3.5 Variables for other appliances system

3.3.2 Brief introduction for four sub-models

(1) Heating system:

Combining the theoretical background and information from the pilot questionnaire will be discussed in next chapter. Here, we start with a heating model without error terms. The sub-model for heating is:

$$y_1 * \ D_1 * \sqrt{HDH} * \sqrt{SIZE}$$

where D_1 represents a parameter of percentage of heating with electricity and $0 \ \leq \ D_1 \ \leq 1$. And

$$y_1 = \ \alpha_{10} + AGE + \mathrm{HA} + \mathrm{PEL} + \mathrm{PD} + \alpha_{15} * NR + \ \alpha_{16} * LOG(INC) + \ \alpha_{17} * TUP + \alpha_{18} * NHS \qquad (3.1)$$

where y_1 is a regression model based on a set of elements as above.

α_{10}, α_{15}, α_{16}, α_{17} and α_{18} represent the intercept, coefficients for number of residents, logarithm of income, time-of use price and number of conservation measures in heating system respectively.

NHS =D_{UW} + D_{IC} + D_{RAS} + D_{PB}.

LOG(INC) is used instead of INC due to its relative magnitude.

AGE, HA, PD and PEL are functions of the age of the house, age of the heating system and the percentage of household with various age and education levels, respectively.

(2) Electricity water heating:

Similarly, a sub-model for electricity water heating:

$$y_2 * \ D_2$$

where D_2 = 1 if the water heating system of the house is electricity, D_2 = 0 if otherwise. And

$$y_2 = \ \alpha_{20} + PD + \mathrm{PEL} \ + \ \alpha_{23} * LOG(INC) + \ \alpha_{24} * TUP + \alpha_{25} * NR + \ \alpha_{26} * NWH + \alpha_{27} * DTF \qquad (3.2)$$

where y_2 is a regression model based on a set of elements as above.

α_{20}, α_{23}, α_{24}, α_{25} and α_{26} represent the intercept, coefficients for logarithm of income, time-of use price, number of residents, number of conservation measures in water heating system respectively.

NWH = D_{FRSH} + D_{WHI} + D_{RDOF} + D_{CWL} + D_{HLO}.

PD and PEL are functions of the percentage of household with various age and education levels, respectively.

(3)Cooling system:

Similarly, a sub-model for cooling system:

$$y_3 * \ D_3 * \sqrt{CDH} * \sqrt{SIZE}$$

where D_3 represents a parameter related to the fuel source of air-conditioner and number of rooms and $0 \ \leq \ D_3 \ \leq 1$. And

$$y_3 = \ \alpha_{30} + AC\mathrm{HP} + \mathrm{AGE} + \mathrm{PEL} + PD + \ \alpha_{36} * LOG(INC) + \ \alpha_{37} * TUP + \ \alpha_{38} * NR + \alpha_{39} * NCS \qquad (3.3)$$

Where y_3 is a regression model based on a set of elements as above.

α_{30}, α_{36}, α_{37}, α_{38} and α_{39} represent the intercept, coefficients for logarithm of income, time-of use price, number of residents, number of adoption of conservation measures related to the cooling system respectively.

NCS =D_{UW} + D_{IC}+D_{RAS}.

ACHP, AGE, PEL and PD are functions of the age of the air-conditioning or heat pump system, age of the house and the percentage of household with various age and education levels, respectively.

(4)Other appliances:

Firstly, a list of appliances included in my model will be shown on the left, and The characters on the right indicate the number of corresponding appliance owned: (Table 3.6)

Appliances	Abbreviation
Refrigerator 1	RE1
Refrigerator 2	RE2
Freezer 1	FR1
Freezer 2	FR2
Mini Bar 1	MB1
Mini Bar 2	MB2
Top Load Washing Machine(without EWH)	TWM
Front Load Washing Machine(without EWH)	FWM
Dishwasher(without EWH)	DISH
Laptop Computer	LC
Desktop Computer	DC
CRT Computer Monitor	CCM
Flat Screen Computer Monitor	FSCM
Printer	PR
Fax Machine	FM
Copier Machine	CM
Printer/Fax/Copier Combo	PFCC
CRT Television	CT
Plasma Television	PT
LED/LCD Television	LLT
Stereo or Home Entertainment	SHE
Game Console	GC
DVD player/Recorder	DR
Digital Cable Box	DCB
Microwave Oven	MO
Whirlpool Bathtub	WB
Dehumidifier	DE
Electric Air Filter	EAF
Pool Pump	PP
Hot tub	HT
Range	RA
Pool Heater	PH
Incandescent Bulb	IB
Compact Fluorescent Bulb	CFB
Halogen Bulb	HB
Fluorescent Bulb	FB
LED Light Bulb	LLB

Table 3.6 List of other appliances

The sub-model for other appliances system is:

$$y_4 * \sqrt{D_4}$$

where D_4 has calculation equation as follows:

$$\begin{aligned}\mathrm{D_4} = {} & \mathrm{UEC_{RE1} * RE1 + UEC_{RE2} * RE2 + UEC_{FR1} * FR1 + UEC_{FR2} * FR2} \\ & \mathrm{+ UEC_{MB1} * MB1 + UEC_{MB2} * MB2 + UEC_{TWM} * TWM + UEC_{FWM}} \\ & \mathrm{* FWM + UEC_{DISH} * DISH + UEC_{LC} * LC + UEC_{DC} * DC + UEC_{CCM}} \\ & \mathrm{* CCM + UEC_{FSCM} * FSCM + UEC_{PR} * PR + UEC_{FM} * FM + UEC_{CM}} \\ & \mathrm{* CM + UEC_{PFCC} * PFCC + UEC_{CT} * CT + UEC_{PT} * PT + UEC_{LLT}} \\ & \mathrm{* LLT + UEC_{SHE} * SHE + UEC_{GC} * GC + UEC_{DR} * DR + UEC_{DCB}} \\ & \mathrm{* DCB + UEC_{MO} * MO + UEC_{WB} * WB + UEC_{DE} * DE + UEC_{EAF}} \\ & \mathrm{* EAF + UEC_{PP} * PP + UEC_{HT} * HT + UEC_{RA} * RA + UEC_{PH} * PH} \\ & \mathrm{+ UEC_{IB} * IB + UEC_{CFB} * CFB + UEC_{HB} * HB + UEC_{FB} * FB} \\ & \mathrm{+ UEC_{LLB} * LLB}\end{aligned}$$

And

$$\mathrm{y_4 = \alpha_{40} + PEL + PD + \alpha_{43} * NR + \alpha_{44} * log(INC) + \alpha_{45} * TUP + \alpha_{46} * NOA} \quad (3.4)$$

where y_4 is based on a set of elements as above.

$\mathrm{UEC_{RE1}}$ represents unit energy consumption, namely, wattage for one Refrigerator 1. Other appliances represent corresponding wattage usage for one unit.

α_{40}, α_{43}, α_{44}, α_{45} and α_{46} represent the intercept, coefficients for number of residents, logarithm of income and time-of-use price respectively.

NOA $= D_{RDOF} + D_{RUA} + D_{CEA} + D_{TOL} + D_{HLO}$.

(5) Final model:

The final model is as follows:

$$\mathrm{Iny = \alpha_0 + y_1 * D_1 * \sqrt{HDH} * \sqrt{SIZE} + y_2 * D_2 + y_3 * D_3 * \sqrt{CDH} * \sqrt{SIZE}}$$

$$\mathrm{+ y_4 * \sqrt{D_4}} \quad (3.5)$$

Substituting (3.1), (3.2), (3.3) and (3.4) into (3.5), we observe that:

$$
\begin{aligned}
\mathrm{Iny} = \alpha_0 + (\alpha_{10} + \mathrm{AGE} + \mathrm{HA} + \mathrm{PEL} + \mathrm{PD} + \alpha_{15} * \mathrm{NR} + \alpha_{16} * \log(\mathrm{INC}) + \alpha_{17} \\
* \mathrm{TUP} + \alpha_{18} * \mathrm{NHS}) * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} \\
+ (\alpha_{20} + \mathrm{PD} + \mathrm{PEL} + \alpha_{23} * \mathrm{logINC} + \alpha_{24} * \mathrm{TUP} + \alpha_{25} * \mathrm{NR} + \alpha_{26} \\
* \mathrm{NWH} + \alpha_{27} * \mathrm{DTF}) * \mathrm{D}_2 \\
+(\alpha_{30} + \mathrm{ACHP} + \mathrm{AGE} + \mathrm{PEL} + \mathrm{PD} + \alpha_{36} * \log\mathrm{INC} + \alpha_{37} * \mathrm{TUP} \\
+ \alpha_{38} * \mathrm{NR} + \alpha_{39} * \mathrm{NCS}) * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} \\
+(\alpha_{40} + \mathrm{PEL} + \mathrm{PD} + \alpha_{43} * \mathrm{NR} + \alpha_{44} * \log(\mathrm{INC}) + \alpha_{45} * \mathrm{TUP} + \alpha_{46} \\
* \mathrm{NOA}) * \sqrt{D_4}
\end{aligned}
\tag{3.6}
$$

3.4 Summary

This chapter illustrated the general framework of my model. More specifically, in section 3.2, the drawbacks and limitations of existing models have been discussed. Section 3.3 gave a description of two parts. In the first part, the theoretical background was explained and a general structure of my model was shown in Figure 3.1. In addition, a list of abbreviations or symbols and description related to my model was set according to the different sub-models. The second part introduced brief information for the sub-models of four systems corresponding to electricity heating, electricity water heating, cooling and other appliances. Finally, a final model was followed by four separate sub-models. I will introduce more details about the model specification and estimation later in chapter 5.

Chapter 4 Data Sources, Collection and Limitations

4.1 Overview

This chapter focuses on the data sources, their collection and their limitations. A pilot questionnaire was conducted across a random set of households. My final model consists of 978 households. This survey covers a range of factors for the 978 households influencing end-use electricity consumption for one household, including household demographics, appliance information and adoption of conservation measures. In section 4.2, I will discuss how I collect the data, check and estimate some variables if necessary, including dwelling and household information, electricity consumption, weather data and price data.

4.2 Data Sources

This pilot questionnaire (See Appendix 1) is designed by McMaster University, with a total number of 978 responses. In the survey, four sources of data are discussed, including dwelling and household information, consumption data, weather data and price data. I will introduce each section in details.

4.2.1 Dwelling and Household Information (Residential Energy Pilot Questionnaire)

In this section, dwelling information includes the type of building, the age of house, the size of house, the number of rooms and a list of demographic factors such as age distribution, education distribution and number of residents. I checked the correctness of age distribution, living size of house and main fuel source. In addition, for some households those with missing living space size and income data, I will estimate the missing values by building a linear regression model on standardized households.

(1) Check within the year

Regarding dwelling information, I had to consider changes during the year. Because we are estimating a relationship for 2013, if the change happens before December of 2013, I should consider the current situation, otherwise I will take account of the change.

(1) Size of living space: if there is renovation of the house in the last three years. I had to take account of the change in square footage;
(2) Main heating system: I need to account for the year of the change;
(3) Education level distribution;
(4) Age distribution;
(5) Electricity water heating system.

(2) Check regarding age distribution, number of residents and education level distribution

A check regarding age distribution, number of residents and education level distribution helped verify the reasonability of the demographic elements. Firstly, I obtained the number of residents (NR) by adding each number from the age distribution tables. Secondly, I get the education distribution over 15 years old (NEDUC(≥15)) by adding every number from the education level table.

Combined table is shown as follows: (Table 4.1)

Age distribution table	Education distribution table
(a) 0-10	(a) < high school
(b)11-18	(b) High school or post-secondary
(c) 19-30	(c)College
(d) 31-50	(d) Bachelor
(e) 51-60	(e) Post graduate
(f) 61-64	(f) Other
(g) 65-74	
(h) 75 and over	
NR=(a)+(b)+(c)+(d)+(e)+(f)+(g)+(h)	NEDUC(≥15)=(a)+(b)+(c)+(d)+(e)+(f)
NR(≤10)=(a)	
NR(≥18)=(c)+(d)+(e)+(f)+(g)+(h)	

Table 4.1 Age distribution table and education level table

From the table 4.1, I have to process two checks to make sense.

Check 1:

$$\mathbf{NR - NR(\leq 10) \geq NEDUC(\geq 15)}$$

Check 2:

NR(≥18) ≤ NEDUC(≥15)

Simply, I need to make these two checks true at the same time. For the violation of check 1, it is easy to make corrections. Firstly, I should make the total number from education level table (≥**15**) is less than or equal to the number of residents. Secondly, because I mainly focus on the highest education level of residents in household, I make an adjustment from the lower level to higher level.

With regard to check 2, I couldn't figure out the appropriate adjustments. In total, 18 households had this problem. I sent an email to request for their response for correction. Only 7 households gave their corrections for education level table. Therefore, I decided to drop other households' information with education issues.
(See Appendix 2 3)

(3) Estimation for size of living space

With regard to the square footage of one house, a relationship between square footage and number of rooms exists. Generally speaking, I consider that the ratio of square footage over the number of rooms should be from 50 to 250. In this way, I keep the households within this range and then build a regression model to estimate other households' missing size. I built a regression model by combining the households with a satisfactory range. In total, 772 households are considered as satisfactory households. I built a regression model based on these 772 households' information and make an assumption that the size of living space is relevant to number of rooms (RM), the square of number of rooms (RM^2) and the square root of number of rooms ($\sqrt{RM}$). The regression model shows as follows: (See Appendix 4)

$$\text{Size} = \alpha + \beta * \text{RM} + \theta * RM^2 + \gamma * \sqrt{RM}$$

Using R software's "lm" package, I input the lists of variables of living space size, RM, RM^2 and $\sqrt{RM}$ and build a regression linear model. Then I apply the "summary" order to find the coefficients α, θ and γ: (attached R code: see Appendix 5)

$$\text{Size} = -3892.980 - 894.915 * \text{RM} + 25.322 * RM^2 + 3975.605 * \sqrt{RM}$$

Here, the adjusted R-squared equals to `0.9961`. Such a high R-squared explains that this model fits my data very well.

Using this regression result, I can estimate the households out of range. Final living space results are obtained. (See Appendix 6)

(4) Check living space, heating system and air-conditioner system with “installers’ estimates”

In order to make the data of any space, heating and cooling system more reasonable, installers re-visited some households to double check with the size of house living space, main fuel source of heating system and air-conditioner system. Here, in total, 511 households are re-visited. (See Appendix 7)

For the house size, I compare the information here with original answer from survey. I make some adjustments according to degree of reasonability. I will choose the one with more reliability among them. (See Appendix 8)

Concerning about the main fuel source of heating system and air-conditioner system, I would like to believe in the installers’ re-visiting results. I will modify the source of main heating system and air-conditioner system by replacing original survey answer with results from installers. (See Appendix 9 10)

(5) Check between the main fuel source and electricity water heating system

From a reasonableness perspective, a relationship between the main fuel source of heating system and electricity water heating system should exist. Normally, if the main fuel source is natural gas, corresponding water heating source should be natural gas. And if the main heating source is not natural gas, corresponding water heating source should be electricity. However, I do not exclude the exceptional case. Firstly, I will illustrate the list of main fuel source and water heating source in table 4.2:

Final Main Fuel Source	Water Heating source
Electric baseboards	Electric
Electric furnace	Natural gas-power vented
Electric heat pump	Natural gas-non-power vented
Natural gas	Propane
Propane	Oil
Wood	Air source electric heat pump
Oil	Geothermal electric heat pump
Other	Solar
	Other

Table 4.2 Main fuel source versus Water heating source

Except for the two normal situations above, I have collected two groups of violations: (Figure 4.1 & Figure 4.2)

Violation 1:

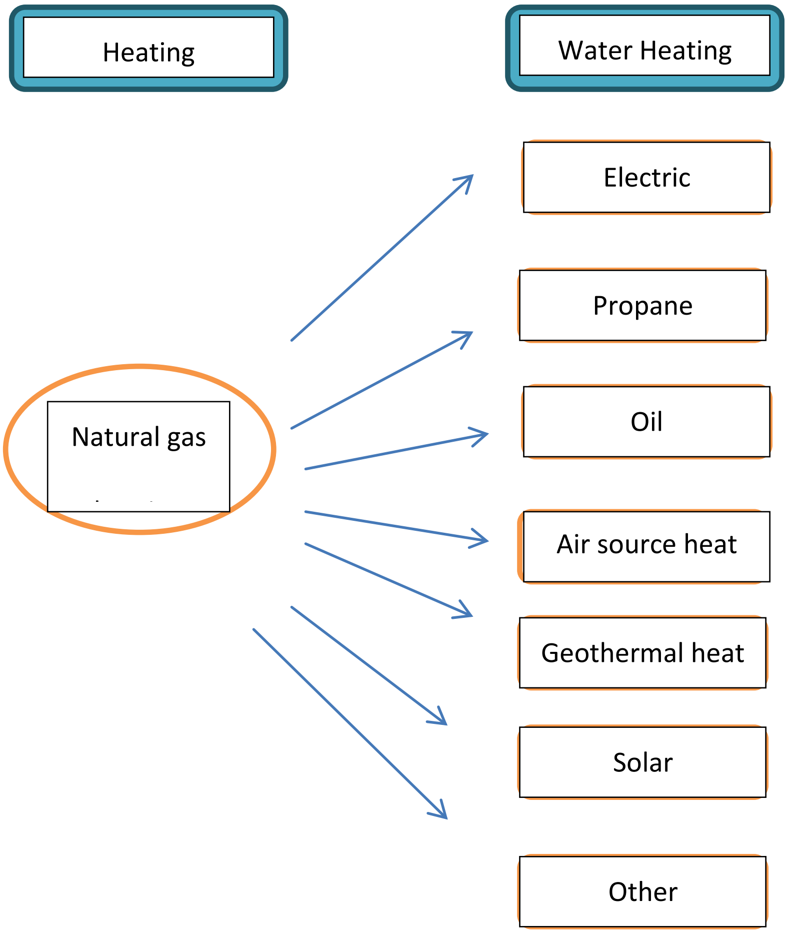


Figure 4.1 Natural gas to any other water heating source (water heating system)

In total, 88 households satisfy with group 1 situation. However, comparing with original survey, I changed 7 households' main fuel source to be the same as the original survey answer instead of results from installers and kept the 81 households with installers' results. In this way, I tried my best to adjust any fuel source matching to electricity for water heating. (See Appendix 11)

Violation 2:

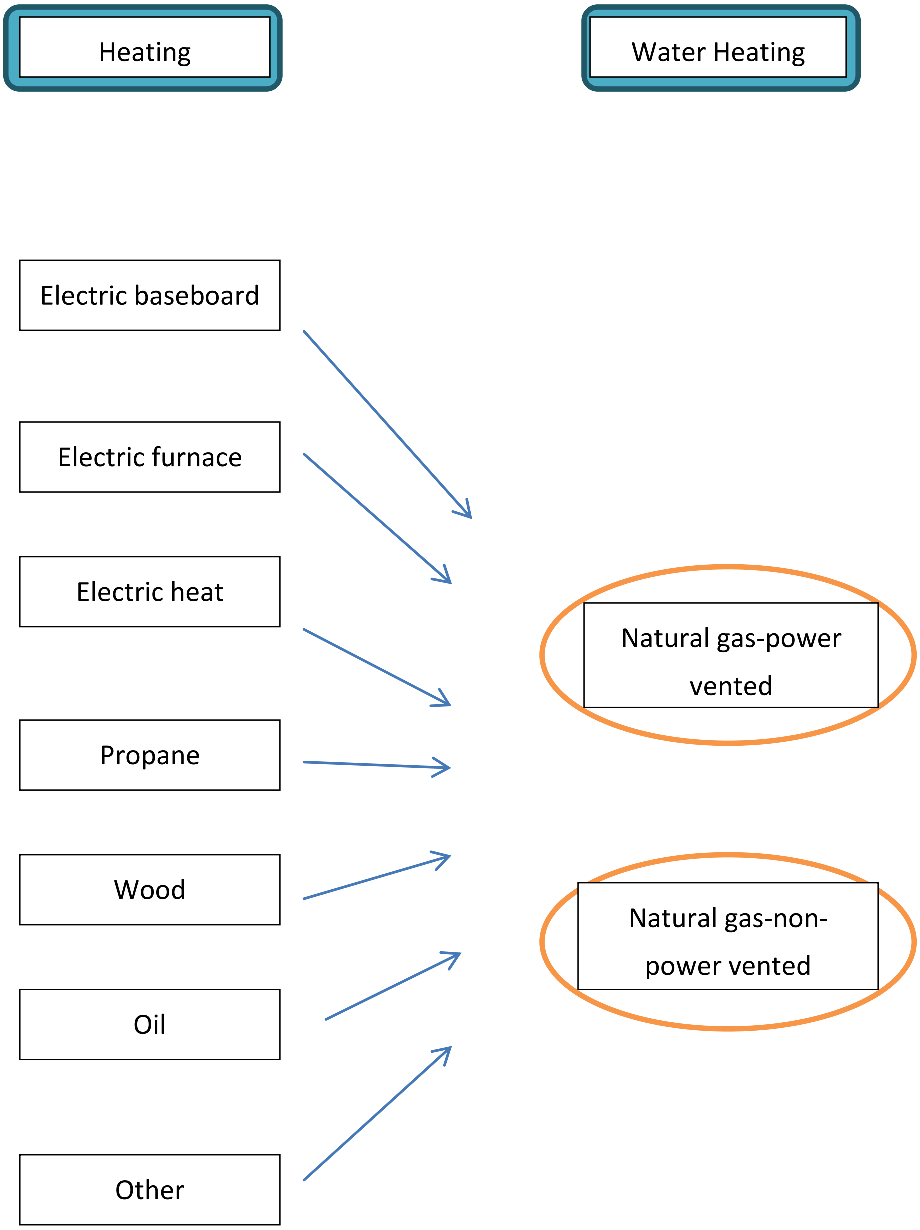


Figure 4.2 Any other main fuel source to natural gas

In total, 24 households satisfy with violation 2. However, comparing with original survey, I changed 7 households' main fuel source same with original survey answer instead of results from installers and kept the 17 households with installers' results. In this way, I try my best to adjust natural gas for main fuel source matching to natural gas for water heating.

(6) Missing house age

Strictly speaking, house age is a significant issue to estimate the aggregate electricity consumption for one household in my model. Originally, I decide to accept 978 households as my database. However, I found 22 households with missing house age. This is a question that couldn't be estimated. I have to request for their information and I obtained 9 replies. Similarly with education issues, I have to delete the households' information with missing house age. (See Appendix 12)

(7) Estimation of Income

Generally speaking, I assume that annual income has a relationship with square footage of house, the education level of household and the age distribution of household. According to the information from living space, education level table and age distribution table, I also build regression model based on existing database. Out of the 978 households, 18 households have missing income information. 18 households have education issues. (section (2) check 2) And 3 households have "0" for the total number of residents over 18 years old. Except for these 39 households (18+18+3), I build a regression model on following factors: (See Appendix 13)

$$\text{Income} = \alpha_0 + \alpha_1 * \text{Size} + \alpha_2 * \text{Size}^2 + \alpha_3 * \frac{\#\, of\ College\ level\ and\ above}{\#\, of\ residents > 18}$$

$$+\alpha_4 * \frac{\#\, of\ residents > 65}{\#\, of\ residents}$$

Using R software's "lm" package, I input the lists of variables of living space size, square footage, the ratio of number of at least college and residents over 18 and the ratio of number of residents over 65 and total number into a regression linear model. Then I apply the "summary" order to find the coefficients α_0, α_1, α_2, α_3 and α_4: (attached R code: Appendix 14)

$$\text{Income} = 3.199 * \ 10^4 + 1.332 * 10 * Size + 5.838 * 10^{-3} * Size^2$$

$$+2.908 * \ 10^4 * \frac{\#\, of\ College\ level\ and\ above}{\#\, of\ residents > 18}$$

$$-3.161 * 10^4 * \frac{\#\, of\ residents > 65}{\#\, of\ residents}$$

Here, the adjusted R-squared is 0.2166. This means that the model fits the data to an extent.

Using this regression result, I can estimate the living size for households. Final living space results are obtained.

4.2.2 Consumption Data

I collect and select the kWh electricity consumption information of 978 households for my database from Hydro One. (See Appendix 15)

4.2.3 Weather Data

An Environment Canada weather station is usually a facility built near airports to record the atmospheric conditions in order to provide the information for weather forecasts and climate study. The information includes temperature, dew point temperature, relative humidity, wind direction, wind speed, visibility, station pressure, humidex, wind chill and occurrence of weather. Given a summary customers by weather stations, I can connect the weather factor with each household. In my thesis, the weather stations are only within Ontario of Canada.

(1) Customers by Postal Code

I was provided a file with postal codes and corresponding abbreviations of weather stations. (See Appendix 16) We must math weather station to each customer. In this file, four parts are contained, including central, east, north and west. In each part, lists of postal codes those may be applied in my database are shown. In addition, their corresponding weather stations marked as abbreviations. This is an important term for the calculation of heating and cooling degree hours in later section.

(2) Sample & Clusters in Each Region

This is helpful to match the postal codes to correct regions and weather stations.(See Appendix 17)

(3) Weather Station Directory

This is a list of abbreviations of weather station name and their full name. It is convenient to search the past weather information online. (See Appendix 18)

(4) Heating and Cooling Degree Hours for weather Stations

Heating degree hours (HDH) is a measurement design to reflect the demand of energy for heating. It comes from the measurement of outdoor air temperature. For example, if the temperature of outside is over a reference temperature, the heating indoor is no longer needed.

Similarly, cooling degree hours (CDH) is a measurement to reflect the demand of energy to cool a building. If the temperature is below a reference temperature, the cooling is not needed any more.

It is possible to use heating all around the year particularly in some parts of Ontario, so that weather data of twelve months need to be considered. Two steps are applied to calculate HDH. Firstly, I need make a reference temperature for demand of heating. By comparing the temperature outside and wind chill, I choose the minimum of these two elements to determine the temperature for demand for heating (TDH) Secondly, I apply a formula to obtain the HDH, taking the maximum of the result of 18 minus TDH and 0. I assume that heating is needed if the temperature for HDH is below 18℃.

Generally, cooling indoor is required from May to September. In this way, I only consider the sum of CDH among these months. Similarly, two steps are projected to CDH. In the first step, I need to define the temperature for demand of cooling (TDC) by taking the maximum of temperature outside and humidex at that time. In the second step, using the result of TDC minus 22, I take the maximum by comparing this result with 0.

Temperature for demand of heating: (TDH)

$$\min(\mathrm{temperature}, \mathrm{wind\ chill})$$

Heating degree hours: (HDH)

$$\max(18 - \mathrm{TDH}, 0)$$

Temperature for demand of cooling: (TDC)

$$\max(\mathrm{temperature}, \mathrm{humidex})$$

Cooling degree hours: (CDH)

$$\max(\mathrm{TDC} - 22, 0)$$

4.2.4 Price Data

For the calculation of the average time-of-use prices, I have classified the households into three groups according to their geographic location provided by Hydro One, namely R1 (medium density), R2 (low density) and UR (high density). In addition, I sort the days into summer and winter through whole year and three time periods according to weekdays, weekends and holidays, namely on peak, mid-peak and off peak. I have attached the calculation formula for each category as endnotes. (See Appendix 19)

After I obtain the calculation, I need to find corresponding categories each household during each time period. Firstly, I need to use the “if statement” in Excel to obtain the logical category for each household during each period and “vlookup” the corresponding price data for this category. I import all the data into R and export all the result as an Excel file. (See Appendix 20)

4.3 Summary

This chapter mainly focused on the data resources, including dwelling information, consumption data, weather data and price data. The dwelling information required data checks regarding household information and demographic factors. Consumption data comprises total electricity consumption for 978 households in 2013. The introduction of HDH and CDH was brought in my model by collecting temperature information online and regional postal information. Regarding price data, in order to calculate the average time-of-use price, I needed classify each household's electricity consumption during each time period into corresponding time-of-use periods. This chapter mainly focused on explaining the resources and checks of data under these four topics.

Chapter 5 Model Specification & Estimation

5.1 Overview

With the discussion of general modeling framework in chapter 3 and data sources in chapter 4, I have provided a brief introduction to my model and data processing. In this chapter, I will introduce in detail the combination of my model and data sources from pilot questionnaire in 5.2. In section 5.3, I will give a brief introduction of the final model framework. I discuss model estimation in 5.4, including without error terms and with error terms under the Generalized Least Squares Method.

5.2 Model Specification

In this section, I will introduce the definition for each variable related to heating, water heating, cooling and other appliances electricity usage. Firstly, I set y_1, y_2, y_3 and y_4 as functions of explanatory variables for the heating, water heating, cooling and other appliances systems.

$$\mathrm{Iny} = \alpha_0 + \mathrm{y}_1 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} + \mathrm{y}_2 * \mathrm{D}_2 + \mathrm{y}_3 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}}$$

$$+\mathrm{y}_4 * \sqrt{\mathrm{D}_4}$$

Where

$$y_1 = \alpha_{10} + \mathrm{AGE1} + \mathrm{HA1} + \mathrm{PEL1} + \mathrm{AGEDIST1} + \alpha_{15} * \mathrm{NR} + \alpha_{16} * \log(\mathrm{INC}) + \alpha_{17} * \mathrm{TUP} + \alpha_{18} * \mathrm{NHS}$$

$$y_2 = \alpha_{20} + \mathrm{AGEDIST2} + \mathrm{PEL2} + \alpha_{23} * \log(\mathrm{INC}) + \alpha_{24} * \mathrm{TUP} + \alpha_{25} * \mathrm{NR} + \alpha_{26} * \mathrm{NWH} + \alpha_{27} * \mathrm{DTF}$$

$$y_3 = \alpha_{30} + \mathrm{ACHP3} + \mathrm{AGE3} + \mathrm{PEL3} + \mathrm{AGEDIST3} + \alpha_{36} * \log(\mathrm{INC}) + \alpha_{37} * \mathrm{TUP} + \alpha_{38} * \mathrm{NR} + \alpha_{39} * \mathrm{NCS}$$

$$y_4 = \alpha_{40} + \mathrm{PEL4} + \mathrm{AGEDIST4} + \alpha_{43} * \mathrm{NR} + \alpha_{44} * \log(\mathrm{INC}) + \alpha_{45} * \mathrm{TUP} + \alpha_{46} * \mathrm{NOA}$$

(5.1)

5.2.1 AGE (Age of house)

In the pilot questionnaire, 8 categories are considered. In addition, I collapse these 7 categories (excluding unknown-age category) into 5 groups. Applying dummy variables, I build a model including dummy variables as follows:

before 1965	
1965-1986	DA_2
1987-1993	DA_3
1994-2005	DA_4
2006 or later	DA_5

Therefore,

$$\mathrm{AGE1} = \alpha_{112} * \mathrm{DA}_2 + \alpha_{113} * \mathrm{DA}_3 + \alpha_{114} * \mathrm{DA}_4 + \alpha_{115} * \mathrm{DA}_5$$

$$\mathrm{AGE3} = \alpha_{332} * \mathrm{DA}_2 + \alpha_{333} * \mathrm{DA}_3 + \alpha_{334} * \mathrm{DA}_4 + \alpha_{335} * \mathrm{DA}_5$$

where DA_x represents a number 0 or 1.

5.2.2 HA (Age of heating system)

< 10 Ys	DHA_1
11-20 Ys	DHA_2
>20 Ys	

$$\mathrm{HA} = \alpha_{121} * \mathrm{DHA}_1 + \alpha_{122} * \mathrm{DHA}_2$$

where DHA_x represents a number 0 or 1.

5.2.3 PEL (Percentage of education level)

Less than high school and high school	PEL_1
College and trades	PEL_2
Bachelor’s degree	PEL_3

Post graduates and other

$$\mathrm{PEL1} = \alpha_{131} * \mathrm{PEL}_1 + \alpha_{132} * \mathrm{PEL}_2 + \alpha_{133} * \mathrm{PEL}_3$$

$$\mathrm{PEL2} = \alpha_{221} * \mathrm{PEL}_1 + \alpha_{222} * \mathrm{PEL}_2 + \alpha_{223} * \mathrm{PEL}_3$$

$$\mathrm{PEL3} = \alpha_{341} * \mathrm{PEL}_1 + \alpha_{342} * \mathrm{PEL}_2 + \alpha_{343} * \mathrm{PEL}_3$$

$$\mathrm{PEL4} = \alpha_{411} * \mathrm{PEL}_1 + \alpha_{412} * \mathrm{PEL}_2 + \alpha_{413} * \mathrm{PEL}_3$$

where PEL_x represents the percentage of household in group x, between 0 and 1.

5.2.4 AGEDIST (Percentage of age distribution)

0-18 Ys	PD_1
19-50 Ys	PD_2
51-64 Ys	PD_3
64-75 Ys	PD_4
> 75 Ys	

$$\mathrm{AGEDIST1} = \alpha_{141} * \mathrm{PD}_1 + \alpha_{142} * \mathrm{PD}_2 + \alpha_{143} * \mathrm{PD}_3 + \alpha_{144} * \mathrm{PD}_4$$

$$\mathrm{AGEDIST2} = \alpha_{211} * \mathrm{PD}_1 + \alpha_{212} * \mathrm{PD}_2 + \alpha_{213} * \mathrm{PD}_3 + \alpha_{214} * \mathrm{PD}_4$$

$$\mathrm{AGEDIST3} = \alpha_{351} * \mathrm{PD}_1 + \alpha_{352} * \mathrm{PD}_2 + \alpha_{353} * \mathrm{PD}_3 + \alpha_{354} * \mathrm{PD}_4$$

$$\mathrm{AGEDIST4} = \alpha_{421} * \mathrm{PD}_1 + \alpha_{422} * \mathrm{PD}_2 + \alpha_{423} * \mathrm{PD}_3 + \alpha_{424} * \mathrm{PD}_4$$

where PD_x represents the percentage of household in group x, between 0 and 1.

5.2.5 NR (Number of residents)

By adding the numbers in the age distribution table of the questionnaire, I calculate the NR as the total number of residents.

5.2.6 log(INC) (Logarithm of income)

In the pilot questionnaire, 8 categories are included. For each group, I have set the midpoint number as the final income for groups between $ 20,000 and $ 137,500. For the first group, I selected $ 15,000 (< $ 20,000). And for the final group (> $ 150,000), I selected $ 200,000. The details are shown as follows:

<$20,000 -----set as 15,000

$20,000-$39,999 -----set as 30,000

$40,000-$59,999 -----set as 50,000

$60,000-$79,999 -----set as 70,000

$80,000-$79,999 -----set as 99,999

$100,000-$124,999 -----set as 112,500

$125,000-$150,000 -----set as 137,500

>$150,000 -----set as 200,000

5.2.7 TUP (Time-of-use prices)

As mentioned in 4.2.4, three groups of households have been classified according to geographic location, recognized as R1, R2 and UR. What’s more, summer and winter times are recorded under three geographic locations respectively. In addition, three time periods are separated under different locations and seasons. First of all, formulas of PTPF (on-peak), PTMF (mid-peak) and PTOF (off-peak) are defined as follows:

$$PTPF= (PC+PTP*LF)*1.13*0.9$$

$$PTMF= (PC+PTM*LF)*1.13*0.9$$

$$PTOF= (PC+PTO*LF)*1.13*0.9$$

Where PTPF, PTMF and PTOF represent end-use electricity price of consumer at on-peak, mid-peak and off-peak;

PC represents delivery charge;

PTP, PTM and PTO stand for commodity charges for peak, mid-peak and off-peak;

LF represents Loss factor adjustment.

Secondly, according to the basic formulas above, price ($/kWh) under peak time, mid-peak and off-peak under different seasons and geographic areas have been shown as following table: (Table 5.1)

May 1st-Oct 31st								
	PTP	PTM	PTO	LF	PC	PTPF	PTMF	PTOF
R1	0.124	0.104	0.067	1.085	0.0599796	0.197826433	0.175757533	0.134930068
UR	0.124	0.104	0.067	1.078	0.05120968	0.188024669	0.166098149	0.125534087
R2	0.124	0.104	0.067	1.092	0.0634516	0.202240213	0.180028933	0.138938065
Jan 1st-Apr 30th								
	PTP	PTM	PTO	LF	PC	PTPF	PTMF	PTOF
R1	0.118	0.099	0.063	1.085	0.0607391	0.191978175	0.17101272	0.1312887
UR	0.118	0.099	0.063	1.078	0.05196428	0.182214141	0.161383947	0.121916211
R2	0.118	0.099	0.063	1.092	0.064216	0.196354224	0.175253508	0.135273204
Nov 1st-Dec 31st								
	PTP	PTM	PTO	LF	PC	PTPF	PTMF	PTOF
R1	0.129	0.109	0.072	1.085	0.0599796	0.203343658	0.181274758	0.140447293
UR	0.129	0.109	0.072	1.078	0.05120968	0.193506299	0.171579779	0.131015717
R2	0.129	0.109	0.072	1.092	0.0634516	0.207793033	0.185581753	0.144490885

Table 5.1 Price under different time periods, seasons and geographic areas

Thirdly, according to price under different seasons, PA (annual price) is calculated by following formulas:

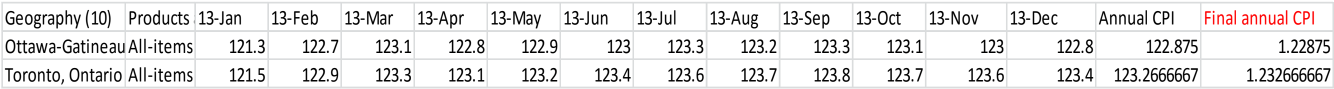
PA(winter)=(sum(PTPF*KWh)+sum(PTMF*KWh)+sum(PTOF*KWh))/all of the electricity usage

PA(summer)=(sum(PTPF*KWh)+sum(PTMF*KWh)+sum(PTOF*KWh))/all of the electricity usage

PA=(PA(winter)*KWh for winter + PA(summer) *KWh for summer)/all of the electricity usage

The sum is a summation across times in the respective periods.

Fourthly, CPI (consumer price index) is also considered in the final price. CPI (monthly, 2013) is collected from Statistics Canada website, listed as following table: (Table 5.2)



Geography (10)	Products	13-Jan	13-Feb	13-Mar	13-Apr	13-May	13-Jun	13-Jul	13-Aug	13-Sep	13-Oct	13-Nov	13-Dec	Annual CPI	Final annual CPI
Ottawa-Gatineau	All-items	121.3	122.7	123.1	122.8	122.9	123	123.3	123.2	123.3	123.1	123	122.8	122.875	1.22875
Toronto, Ontario	All-items	121.5	122.9	123.3	123.1	123.2	123.4	123.6	123.7	123.8	123.7	123.6	123.4	123.2666667	1.232666667

Table 5.2 Final annual CPI

Fifthly, final annual price can be calculated by:

Final price = PA / corresponding final annual CPI (Ottawa or Toronto area)

5.2.8 NHS (Number of conservation measures related to heating system)

With respect to conservation measures related to the heating system, I selected 4 behavior measures. As I mentioned in 3.3.3, each measure represents a number of 0 or 1 depending on whether it happens or not. I take the final sum of these measures into my model as one variable related to behavioral changes.

5.2.9 D_1

In 3.3.3, I have referred to D_1, a parameter of percentage of heating with electricity. Here, I will give an explicit explanation of calculation of D_1: (Table 5.3)

Main fuel source	Secondary fuel source	D_1
Electricity	No secondary or electricity	1
Electricity	Other than electricity (%: k)	1-k%
Other than electricity	Electricity(%: h)	h%
Otherwise	Otherwise	0

Table 5.3 Calculation of D_1

5.2.10 HDH

In 4.3.1, I have mentioned the source of weather stations. After collecting data for 9 weather stations, the calculation of HDH has been described. However, this is only for hourly data. For HDH, I consider to apply the hourly data for 365 days and add them up. In this way, I could obtain the total HDHs for 2013 for the 9 weather stations. Firstly, I should find their corresponding weather station according to their postal code for the all households. Secondly, I apply the corresponding HDH I calculated previously to each customer.

5.2.11 Size

With respect to living space, there are 8 categories in the pilot questionnaire. For the last category, missing information for size of living space has been considered in 4.2.1. I have built a linear regression model to estimate the size according to a series of related factors, especially the number of rooms. After estimation, 7 categories of size are involved in the final consideration. Similarly to income problem, I also set the medium number for size of space in each group as follows:

<1,000 ------set as 700

1,000-1,499 -----set as 1,250
1,500-1,999 -----set as 1,750
2,000-2,999 -----set as 2,500
3,000-3,999 -----set as 3,500
4,000-4,999 -----set as 4,500
5000 or more -----set as 5,500

5.2.12 NWH (Number of conservation measures related to electric water heating)

With respect conservation measures related to the water heating, I selected 5 behavior measures. As I mentioned in 3.3.3, each measure represents a number of 0 or 1 depending on whether it happens or not. I add these numbers up to obtain the final result for NWH.

5.2.13 D_2

Compared with D_1, D_2 is much easier to understand. D_2 is a dummy variable representing that the household's water heating system is electricity.

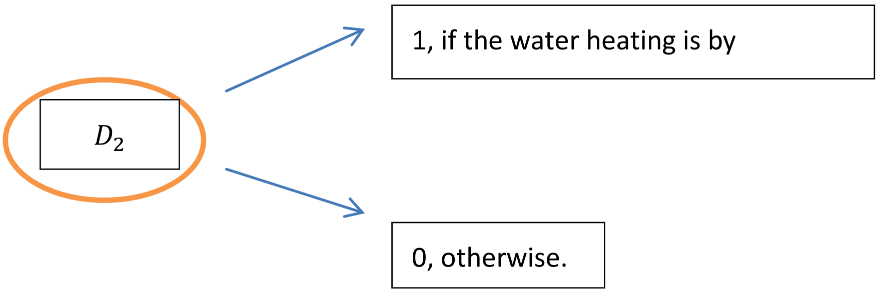


Figure 5.1 D_2

5.2.14 DTF

For the electricity water heating system, I consider the wattage of electric appliances Dishwasher, Top Load Washing and Front Load Washing (recognized as "DTF") for their electricity consumption drawing on the electricity water heating system. What I need to emphasize here is that DTF is the sum of the difference between "with Electricity Water Heating" and "without Electricity Water Heating" for these three appliances. I assume that one household who uses electricity to provide energy from these appliances when I calculate "DTF". Regarding non-water heating electricity provided to these appliances, I consider the energy consumption in "$\boldsymbol{D_4}$" I will discuss later.

5.2.15 ACHP (Age of central air-conditioner or heat pump or windows or other)

For central air-conditioner, heat pump, window air conditioning and other sources to cool the house, I choose through dummy variables for the age of these four sources:

<5 Ys	DACHP_1
5-15 Ys	DACHP_2
>15 Ys	

$$\mathrm{ACHP3} = \alpha_{331} * \mathrm{DACHP}_1 + \alpha_{332} * \mathrm{DACHP}_2$$

where $DACHP_x$ represents a number 0 or 1.

5.2.16 NCS (Number of conservation measures related to cooling system)

With respect to the conversation measures related to the cooling system, I selected 3 behavioral measures. As I mentioned in 3.3.3, each measure represents a number of 0 or 1 depending on whether it happens or not. The final result of NCS is calculated by addition of each number.

5.2.17 $\boldsymbol{D_3}$

In 3.3.3, I have mentioned D_3, a parameter of percentage of cooling with central air-conditioner, heat pump or window air-conditioner. Especially for the window air-conditioner, I give a definition of calculating the D_3. I will give an explicit explanation of calculation of D_3: (Table 5.4)

Cooling system	D_3
Central air-conditioner, air source electric heat pump or geothermal electric heat pump	1
Window air-conditioner	Min($\frac{number\ of\ windows*1.5}{number\ of\ rooms}$,1)
otherwise	0

Table 5.4 Calculation of D_3

5.2.18 CDH

Similar to HDH, I have collected weather data from 9 weather stations and calculated the hourly CDH from May to October in 2013. In the next step, I add the CDH up for these 6 months and the total CDH in 2013 for these 9 stations. For the all households, I find their corresponding weather station according to their postal code.

5.2.19 D_4

As I mentioned in 3.3.2, D_4 represents the total electricity wattage for all other appliances. Here, I add the wattage of each appliance multiplying by quantity for each appliance.

5.3 Final model framework

We can build the final model through adding the dummy variables respectively:

$0 \le D_1 \le 1,\ 0 \le D_3 \le 1$ and $D_2 = 1$ if the water heating system is by electricity, $D_2 = 0$ otherwise.

Recall for equation (5.1) at the beginning of this chapter. To expand this equation very explicitly:

$$
\begin{aligned}
\mathrm{Iny} = \alpha_0 &+ (\alpha_{10} + \mathrm{AGE1} + \mathrm{HA1} + \mathrm{PEL1} + \mathrm{AGEDIST1} + \alpha_{15} * \mathrm{NR} + \alpha_{16} * \log(\mathrm{INC}) \\
&+ \alpha_{17} * \mathrm{TUP} + \alpha_{18} * \mathrm{NHS}) * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} \\
&+ (\alpha_{20} + \mathrm{AGEDIST2} + \mathrm{PEL2} + \alpha_{23} * \log(\mathrm{INC}) + \alpha_{24} * \mathrm{TUP} + \alpha_{25} \\
&* \mathrm{NR} + \alpha_{26} * \mathrm{NWH} + \alpha_{27} * \mathrm{DTF}) * \mathrm{D}_2 \\
&+(\alpha_{30} + \mathrm{ACHP3} + \mathrm{AGE3} + \mathrm{PEL3} + \mathrm{AGEDIST3} + \alpha_{36} * \log(\mathrm{INC}) \\
&+ \alpha_{37} * \mathrm{TUP} + \alpha_{38} * \mathrm{NR} + \alpha_{39} * \mathrm{NCS}) * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} \\
&+(\alpha_{40} + \mathrm{PEL4} + \mathrm{AGEDIST4} + \alpha_{43} * \mathrm{NR} + \alpha_{44} * \log(\mathrm{INC}) + \alpha_{45} \\
&* \mathrm{TUP} + \alpha_{46} * \mathrm{NOA}) * \sqrt{D_4}
\end{aligned}
\tag{5.1}
$$

Where

$$
\begin{aligned}
\mathrm{AGE1} &= \mathrm{AGE} * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} \\
&= \alpha_{112} * \mathrm{DA}_2 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{113} * \mathrm{DA}_3 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} \\
&* \sqrt{\mathrm{SIZE}} + \alpha_{114} * \mathrm{DA}_4 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{115} * \mathrm{DA}_5 * \mathrm{D}_1 \\
&* \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}}
\end{aligned}
$$

$$
\mathrm{HA1} = \mathrm{HA} * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} = \alpha_{121} * \mathrm{DHA}_1 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{122} * \mathrm{DHA}_2 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}}
$$

$$
\mathrm{PEL1} = \mathrm{PEL} * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} = \alpha_{131} * \mathrm{PEL}_1 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{132} * \mathrm{PEL}_2 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{133} * \mathrm{PEL}_3 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}}
$$

$$\mathrm{AGEDIST1} = \mathrm{PD} * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}}$$

$$= \alpha_{141} * \mathrm{PD}_1 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{142} * \mathrm{PD}_2 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}}$$

$$* \sqrt{\mathrm{SIZE}} + \alpha_{143} * \mathrm{PD}_3 * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{144} * \mathrm{PD}_4 * \mathrm{D}_1$$

$$* \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}}$$

$$\mathrm{NHS} = (\mathrm{D}_{\mathrm{UW}} + \mathrm{D}_{\mathrm{IC}} + \mathrm{D}_{\mathrm{RAS}} + \mathrm{D}_{\mathrm{PB}}) * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}}$$

$$\mathrm{AGEDIST2} = \mathrm{PD} * \mathrm{D}_2$$

$$= \alpha_{211} * \mathrm{PD}_1 * \mathrm{D}_2 + \alpha_{212} * \mathrm{PD}_2 * \mathrm{D}_2 + \alpha_{213} * \mathrm{PD}_3 * \mathrm{D}_2 + \alpha_{214}$$

$$* \mathrm{PD}_4 * \mathrm{D}_2$$

$$\mathrm{PEL2} = \mathrm{PEL} * \mathrm{D}_2$$

$$= \alpha_{221} * \mathrm{PEL}_1 * \mathrm{D}_2 + \alpha_{222} * \mathrm{PEL}_2 * \mathrm{D}_2 + \alpha_{223} * \mathrm{PEL}_3 * \mathrm{D}_2$$

$$\mathrm{NWH} = (\mathrm{D}_{\mathrm{FRSH}} + \mathrm{D}_{\mathrm{WHI}} + \mathrm{D}_{\mathrm{RDOF}} + \mathrm{D}_{\mathrm{CWL}} + \mathrm{D}_{\mathrm{HLO}}) * \mathrm{D}_2$$

$$\mathrm{ACHP3} = \mathrm{ACHP} * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} = \alpha_{331} * \mathrm{DACHP}_1 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{332} * \mathrm{DACHP}_2 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}}$$

$$\mathrm{AGE3} = \mathrm{AGE} * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}}$$

$$= \alpha_{332} * \mathrm{DA}_2 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{333} * \mathrm{DA}_3 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}}$$

$$* \sqrt{\mathrm{SIZE}} + \alpha_{334} * \mathrm{DA}_4 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{335} * \mathrm{DA}_5 * \mathrm{D}_3$$

$$* \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}}$$

$$\mathrm{PEL3} = \mathrm{PEL} * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} = \alpha_{341} * \mathrm{PEL}_1 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{342} * \mathrm{PEL}_2 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{343} * \mathrm{PEL}_3 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}}$$

$$\mathrm{AGEDIST3} = \mathrm{PD} * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} = \alpha_{351} * \mathrm{PD}_1 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{352} * \mathrm{PD}_2 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{353} * \mathrm{PD}_3 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} + \alpha_{354} * \mathrm{PD}_4 * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}}$$

$$\mathrm{NCS} = (\mathrm{D_{UW}} + \mathrm{D_{IC}} + \mathrm{D_{RAS}}) * \mathrm{D_3} * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}}$$

$$\mathrm{PEL4} = PEL * \sqrt{\mathrm{D_4}} = \alpha_{411} * \mathrm{PEL_1} * \sqrt{\mathrm{D_4}} + \alpha_{412} * \mathrm{PEL_2} * \sqrt{\mathrm{D_4}} + \alpha_{413} * \mathrm{PEL_3} * \sqrt{\mathrm{D_4}}$$

$$\mathrm{AGEDIST4} = \mathrm{PD} * \sqrt{\mathrm{D_4}} = \alpha_{421} * \mathrm{PD_1} * \sqrt{\mathrm{D_4}} + \alpha_{422} * \mathrm{PD_2} * \sqrt{\mathrm{D_4}} + \alpha_{423} * \mathrm{PD_3} * \sqrt{\mathrm{D_4}} + \alpha_{424} * \mathrm{PD_4} * \sqrt{\mathrm{D_4}}$$

$$\mathrm{NOA} = (D_{RDOF} + D_{RUA} + D_{CEA} + D_{TOL} + D_{HLO}) * \sqrt{\mathrm{D_4}}$$

5.4 Model Estimation

To make my model more reasonable, some statistical assumptions need to be added. With respect to the true regression model, a collection of assumptions need to be taken into account. In this section, I will focus on explaining my model by comparing three situations, without error terms, without interactive error terms (only with ε_0) and with interactive error terms. Furthermore, I will illustrate the theoretical background combined with my model under these situations.

5.4.1 Model without error terms

Firstly, I classify my model into four parts for electricity, heating system, electric water heating, cooling system and other appliances. I will combine these four parts together by grouping elements. Recall equation (3.1) in 3.3.3 and I obtain the final sub-model for heating:

$$(\alpha_{10} + \alpha_{11} * AGE1 + \alpha_{12} * \mathrm{HA} + \alpha_{13} * \mathrm{PEL1} + \alpha_{14} * AGEDIST1 + \alpha_{15} * NR1 + \alpha_{16} * LOG(INC)1 + \alpha_{17} * TUP1 + \alpha_{18} * NHS) * D_1 * \sqrt{HDH} * \sqrt{SIZE} \tag{5.2}$$

Similarly, I obtain the final sub-model for electric water heating:

$$(\alpha_{20} + \alpha_{21} * AGEDIST2 + \alpha_{22} * \mathrm{PEL2} + \alpha_{23} * LOG(INC)2 + \alpha_{24} * TUP2 + \alpha_{25} * NR2 + \alpha_{26} * NWH + \alpha_{27} * DTF) * D_2 \tag{5.3}$$

For the cooling system, the sub-model is:

$$(\alpha_{30} + \alpha_{31} * ACHP + \alpha_{32} * \mathrm{AGE3} + \alpha_{33} * \mathrm{PEL3} + \alpha_{34} * AGEDIST3 + \alpha_{35} * LOG(INC)3 + \alpha_{36} * TUP3 + \alpha_{37} * NR3 + \alpha_{38} * NCS) * D_3 * \sqrt{CDH} * \sqrt{SIZE} \tag{5.4}$$

For the other appliances system, the sub-model is:

$$(\alpha_{40} + \alpha_{41} * \mathrm{PEL4} + \alpha_{42} * \mathrm{AGEDIST4} + \alpha_{43} * \mathrm{NR4} + \alpha_{44} * \log(\mathrm{INC})4 + \alpha_{45} * \mathrm{TUP4} + \alpha_{46} * NOA) * \sqrt{D_4} \tag{5.5}$$

After merging the factors from equation (5.2), (5.3), (5.4) and (5.5) above, we could obtain the equation (5.5) below:

$$\mathrm{Iny} = (\alpha_{10} * D_1 * \sqrt{HDH} * \sqrt{SIZE} + \alpha_{20} * D_2 + \alpha_{30} * D_3 * \sqrt{CDH} * \sqrt{SIZE} + \alpha_{40} * \sqrt{D_4})$$

$$+\alpha_{11} * AGE1 * (D_1 * \sqrt{HDH} * \sqrt{SIZE}) + \alpha_{32} * AGE3 * (D_3 * \sqrt{CDH} * \sqrt{SIZE})$$

$$+ \alpha_{13} * \mathrm{PEL1} * (D_1 * \sqrt{HDH} * \sqrt{SIZE}) + \alpha_{22} * \mathrm{PEL2} * D_2 + \alpha_{33} * \mathrm{PEL3} * (D_3 * \sqrt{CDH} * \sqrt{SIZE}) + \alpha_{41} * \mathrm{PEL4} * \sqrt{D_4}$$

$$+ \alpha_{14} * AGEDIST1 * \left(D_1 * \sqrt{HDH} * \sqrt{SIZE}\right) + \alpha_{21} * AGEDIST2 * D_2 + \alpha_{34} * AGEDIST3 * (D_3 * \sqrt{CDH} * \sqrt{SIZE}) + \alpha_{42} * \mathrm{AGEDIST4} * \sqrt{D_4}$$

$$+ NR1 * (\alpha_{15} * D_1 * \sqrt{HDH} * \sqrt{SIZE}) + NR2 * (\alpha_{25} * D_2) + NR3 * (\alpha_{37} * D_3 * \sqrt{CDH} * \sqrt{SIZE}) + \mathrm{NR4} * (\alpha_{43} * \sqrt{D_4})$$

$$+ LOG(INC)1 * (\alpha_{16} * D_1 * \sqrt{HDH}*\sqrt{SIZE}) + LOG(INC)2 * (\alpha_{23} * D_2)+LOG(INC)3 * (\alpha_{35} * D_3 * \sqrt{CDH} *\sqrt{SIZE})+\mathrm{log(INC)}4 * (\alpha_{44} * \sqrt{D_4})$$

$$+ TUP1 *(\alpha_{17} * D_1 * \sqrt{HDH} *\sqrt{SIZE}) + TUP2 * (\alpha_{24} * D_2) + TUP3 * (\alpha_{36} * D_3 * \sqrt{CDH} *\sqrt{SIZE}) + \mathrm{TUP4} * (\alpha_{45} * \sqrt{D_4})$$

$$+\alpha_{12} * \mathrm{HA} * (D_1 * \sqrt{HDH} *\sqrt{SIZE})$$

$$+ NHS * (\alpha_{18} * D_1 * \sqrt{HDH} *\sqrt{SIZE})$$

$$+(NWH * (\alpha_{26} * D_2)$$

$$+ \mathrm{DTF} * (\alpha_{27} * D_2)$$

$$+ \alpha_{31} * ACHP * (D_3 * \sqrt{CDH} *\sqrt{SIZE})$$

$$+NCS * (\alpha_{38} * D_3 * \sqrt{CDH} *\sqrt{SIZE})$$

$$+ \mathrm{NOA} * (\alpha_{46} * D_3 * \sqrt{CDH} *\sqrt{SIZE})$$

(5.6)

5.4.2 Model without interactive error terms (only with ε_0)

Compared with the model without error terms in section 5.4.1, I will introduce an error term ε_0 at the beginning of equation (5.6). This is often the approach in conditional demand analysis. Here are the details:

$$\mathrm{Iny} = \varepsilon_0 +$$

$$(\alpha_{10} * D_1 * \sqrt{HDH} *\sqrt{SIZE}+\alpha_{20} * D_2 + \alpha_{30} * D_3 * \sqrt{CDH} *\sqrt{SIZE}+\alpha_{40} * \sqrt{D_4})$$

$$+\alpha_{11} * \ AGE1 * (D_1 * \sqrt{HDH} *\sqrt{SIZE}) + \alpha_{32} * AGE3 * (D_3 * \sqrt{CDH} *\sqrt{SIZE})$$

$$+ \alpha_{13} * \mathrm{PEL1} * (D_1 * \sqrt{HDH} *\sqrt{SIZE})+\alpha_{22} * \mathrm{PEL2} * D_2+\alpha_{33} * \mathrm{PEL3} * (D_3 * \sqrt{CDH} *\sqrt{SIZE})+\alpha_{41} * \mathrm{PEL4} * \sqrt{D_4}$$

$$+ \alpha_{14} * AGEDIST1 * \left(D_1 * \sqrt{HDH} \ * \sqrt{SIZE}\right) + \alpha_{21} * AGEDIST2 * D_2 + \alpha_{34} * AGEDIST3 * (D_3 * \sqrt{CDH} * \sqrt{SIZE}) + \alpha_{42} * \mathrm{AGEDIST4} * \sqrt{D_4}$$

$$+ NR1 * (\alpha_{15} * D_1 * \sqrt{HDH} * \sqrt{SIZE}) + NR2 * (\alpha_{25} * D_2) + NR3 * (\alpha_{37} * D_3 * \sqrt{CDH} * \sqrt{SIZE}) + \mathrm{NR4} * (\alpha_{43} * \sqrt{D_4})$$

$$+ LOG(INC)1 * (\alpha_{16} * D_1 * \sqrt{HDH} * \sqrt{SIZE}) + LOG(INC)2 * (\alpha_{23} * D_2) + LOG(INC)3 * (\alpha_{35} * D_3 * \sqrt{CDH} * \sqrt{SIZE}) + \mathrm{log(INC)4} * (\alpha_{44} * \sqrt{D_4})$$

$$+ TUP1 * (\alpha_{17} * D_1 * \sqrt{HDH} * \sqrt{SIZE}) + TUP2 * (\alpha_{24} * D_2) + TUP3 * (\alpha_{36} * D_3 * \sqrt{CDH} * \sqrt{SIZE}) + \mathrm{TUP4} * (\alpha_{45} * \sqrt{D_4})$$

$$+\alpha_{12} * \mathrm{HA} * (D_1 * \sqrt{HDH} * \sqrt{SIZE})$$

$$+ NHS * (\alpha_{18} * D_1 * \sqrt{HDH} * \sqrt{SIZE})$$

$$+(NWH * (\alpha_{26} * D_2)$$

$$+ \mathrm{DTF} * (\alpha_{27} * D_2)$$

$$+ \alpha_{31} * ACHP * (D_3 * \sqrt{CDH} * \sqrt{SIZE})$$

$$+NCS * (\alpha_{38} * D_3 * \sqrt{CDH} * \sqrt{SIZE})$$

$$+ \mathrm{NOA} * (\alpha_{46} * D_3 * \sqrt{CDH} * \sqrt{SIZE})$$

(5.7)

5.4.3 Model with interactive error terms

In section 5.4.2, I have introduced the four parts of sub-models without interactive error terms (only with ε_0). In this section, I will bring an explicit explanation of sub-models and final model with error terms (ε_1, ε_2, ε_3, ε_4). By adding error terms, the sub-model for heating system becomes:

$$(\alpha_{10} + \alpha_{11} * \mathrm{AGE1} + \alpha_{12} * \mathrm{HA} + \alpha_{13} * \mathrm{PEL1} + \alpha_{14} * \mathrm{AGEDIST1} + \alpha_{15} * \mathrm{NR1} + \alpha_{16} * \mathrm{LOG(INC)1} + \alpha_{17} * \mathrm{TUP1} + \alpha_{18} * \mathrm{NHS} + \varepsilon_1) * \mathrm{D_1} * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}}$$

(5.8)

Similarly, the sub-model for electric water heating becomes:

$$(\alpha_{20} + \alpha_{21} * AGEDIST2 + \alpha_{22} * \mathrm{PEL2} + \alpha_{23} * LOG(INC)2 + \alpha_{24} * TUP2 + \alpha_{25} * NR2 + \alpha_{26} * NWH + \alpha_{27} * DTF + \varepsilon_2) * D_2$$

(5.9)

Correspondingly, the cooling system's sub-model is:

$$(\alpha_{30} + \alpha_{31} * ACHP + \alpha_{32} * \mathrm{AGE3} + \alpha_{33} * \mathrm{PEL3} + \alpha_{34} * AGEDIST3 + \alpha_{35} * LOG(INC)3 + \alpha_{36} * TUP3 + \alpha_{37} * NR3 + \alpha_{38} * NCS + \varepsilon_3) * D_3 * \sqrt{CDH} * \sqrt{SIZE}$$

(5.10)

The other appliances system:

$$(\mathrm{\alpha}_{40} + \alpha_{41} * \mathrm{PEL4} + \alpha_{42} * \mathrm{AGEDIST4} + \mathrm{\alpha}_{43} * \mathrm{NR4} + \mathrm{\alpha}_{44} * \log(\mathrm{INC})4 + \mathrm{\alpha}_{45} * \mathrm{TUP4} + \alpha_{46} * NOA + \varepsilon_4) * \sqrt{D_4}$$

(5.11)

In addition to ε_0, I combine equation (5.8), (5.9), (5.10) and (5.11) and combining error terms, I obtain equation (5.12) as follows:

$$\mathrm{In}y' =$$

$$(\alpha_{10} * D_1 * \sqrt{HDH} * \sqrt{SIZE} + \alpha_{20} * D_2 + \alpha_{30} * D_3 * \sqrt{CDH} * \sqrt{SIZE} + \mathrm{\alpha}_{40} * \sqrt{D_4})$$

$$+ \alpha_{11} * AGE1 * (D_1 * \sqrt{HDH} * \sqrt{SIZE}) + \alpha_{32} * AGE3 * (D_3 * \sqrt{CDH} * \sqrt{SIZE})$$

$$+ \alpha_{13} * \mathrm{PEL1} * (D_1 * \sqrt{HDH} * \sqrt{SIZE}) + \alpha_{22} * \mathrm{PEL2} * D_2 + \alpha_{33} * \mathrm{PEL3} * (D_3 * \sqrt{CDH} * \sqrt{SIZE}) + \alpha_{41} * \mathrm{PEL4} * \sqrt{D_4}$$

$$+ \alpha_{14} * AGEDIST1 * \left(D_1 * \sqrt{HDH} * \sqrt{SIZE}\right) + \alpha_{21} * AGEDIST2 * D_2 + \alpha_{34} * AGEDIST3 * (D_3 * \sqrt{CDH} * \sqrt{SIZE}) + \alpha_{42} * \mathrm{AGEDIST4} * \sqrt{D_4}$$

$$+ NR1 * (\alpha_{15} * D_1 * \sqrt{HDH} * \sqrt{SIZE}) + NR2 * (\alpha_{25} * D_2) + NR3 * (\alpha_{37} * D_3 * \sqrt{CDH} * \sqrt{SIZE}) + \mathrm{NR4} * (\mathrm{\alpha}_{43} * \sqrt{D_4})$$

$$
\begin{aligned}
&+LOG(INC)1*(\alpha_{16}*D_1*\sqrt{HDH}*\sqrt{SIZE})+LOG(INC)2*(\,\alpha_{23}*\\
&\quad D_2)+LOG(INC)3*(\alpha_{35}*D_3*\sqrt{CDH}\,*\sqrt{SIZE})+\log(\mathrm{INC})4*(\alpha_{44}*\sqrt{D_4})\\
&+TUP1\,*(\alpha_{17}*D_1*\sqrt{HDH}\,*\sqrt{SIZE})+TUP2*(\alpha_{24}*D_2)+TUP3*(\alpha_{36}*D_3*\\
&\quad \sqrt{CDH}\,*\sqrt{SIZE})+\mathrm{TUP4}*(\alpha_{45}*\sqrt{D_4})\\
&+\alpha_{12}*\mathrm{HA}*(D_1*\sqrt{HDH}\,*\sqrt{SIZE})\\
&\quad +NHS*(\alpha_{18}*D_1*\sqrt{HDH}\,*\sqrt{SIZE})\\
&+(NWH*(\alpha_{26}*D_2)\\
&\quad +\mathrm{DTF}*(\alpha_{27}*D_2)\\
&+\alpha_{31}*ACHP*(D_3*\sqrt{CDH}\,*\sqrt{SIZE})\\
&\quad +NCS*(\alpha_{38}*D_3*\sqrt{CDH}\,*\sqrt{SIZE})\\
&\quad +\mathrm{NOA}*(\alpha_{46}*D_3*\sqrt{CDH}\,*\sqrt{SIZE})\\
&+(\varepsilon_0+\varepsilon_1*D_1*\sqrt{HDH}\,*\sqrt{SIZE}+\varepsilon_2*D_2+\varepsilon_3*D_3*\sqrt{CDH}\,*\sqrt{SIZE}+\varepsilon_4*\sqrt{D_4})
\end{aligned}
\tag{5.12}
$$

5.4.4 Statistical assumptions related to the errors in my model

In reality, there is possible correlation between heating, water heating, cooling and other appliances systems. For example, if someone has a shower, an increase occurs in electric water heating consumption. Meanwhile, the shower may lead to increased indoor temperature. In this way, less heating consumption may be consumed. What's more, if people work inside the house by using computers, heat generated by computers may require an increased electricity to cool the house in the summer. Therefore, a situation that correlation exists between error terms will be introduced.

Firstly, I assume that:

$$\mathrm{Y} = \mathrm{X} * \beta + \vartheta_1$$

where Y represents the logarithm of total annual electricity consumption for one household,

X represents the list of independent variables in my model,

β represents the list of coefficients corresponding to the independent variables in my model.

$$\vartheta_1 = \varepsilon_0 + \varepsilon_1 * D_1 * \sqrt{HDH} * \sqrt{SIZE} + \varepsilon_2 * D_2 + \varepsilon_3 * D_3 * \sqrt{CDH} * \sqrt{SIZE} + \varepsilon_4 * \sqrt{D_4}$$

[1] Assumptions

Here, I discuss the situation that correlation exits among $\varepsilon_{0,}\ \varepsilon_1, \varepsilon_2, \varepsilon_3\ and\ \varepsilon_4$. For one special case, I also assume no correlation exits among these five error terms. Suppose that the variance-covariance matrix of error shows as follows:

$$\Sigma^* = Z\Sigma Z'$$

The variance-covariance matrix of ϑ_1 becomes:

$$\Sigma^* \quad = \quad \begin{pmatrix} 1 & z_1 & z_2 & z_3 & z_4 \end{pmatrix} \otimes \begin{pmatrix} \sigma_0^{\ 2} & \sigma_{01} & \sigma_{02} & \sigma_{03} & \sigma_{04} \\ \sigma_{10} & \sigma_1^{\ 2} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{20} & \sigma_{21} & \sigma_2^{\ 2} & \sigma_{23} & \sigma_{24} \\ \sigma_{30} & \sigma_{31} & \sigma_{32} & \sigma_3^{\ 2} & \sigma_{34} \\ \sigma_{40} & \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^{\ 2} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_0^{\ 2} & z_1\sigma_{01} & z_2\sigma_{02} & z_3\sigma_{03} & z_4\sigma_{04} \\ \sigma_{10} & z_1\sigma_1^{\ 2} & z_2\sigma_{12} & z_3\sigma_{13} & z_4\sigma_{14} \\ \sigma_{20} & z_1\sigma_{21} & z_2\sigma_2^{\ 2} & z_3\sigma_{23} & z_4\sigma_{24} \\ \sigma_{30} & z_1\sigma_{31} & z_2\sigma_{32} & z_3\sigma_3^{\ 2} & z_4\sigma_{34} \\ \sigma_{40} & z_1\sigma_{41} & z_2\sigma_{42} & z_3\sigma_{43} & z_4\sigma_4^{\ 2} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}$$

$$= \begin{pmatrix} {\sigma_0}^2 & z_1\sigma_{01} & z_2\sigma_{02} & z_3\sigma_{03} & z_4\sigma_{04} \\ z_1\sigma_{10} & {z_1}^2{\sigma_1}^2 & z_1z_2\sigma_{12} & z_1z_3\sigma_{13} & z_1z_4\sigma_{14} \\ z_2\sigma_{20} & z_1z_2\sigma_{21} & {z_2}^2{\sigma_2}^2 & z_2z_3\sigma_{23} & z_2z_4\sigma_{24} \\ z_3\sigma_{30} & z_1z_3\sigma_{31} & z_2z_3\sigma_{32} & {z_3}^2{\sigma_3}^2 & z_3z_4\sigma_{34} \\ z_4\sigma_{40} & z_1z_4\sigma_{41} & z_2z_4\sigma_{42} & z_3z_4\sigma_{43} & {z_4}^2{\sigma_4}^2 \end{pmatrix}$$

If I assume that

$$\Sigma = \begin{pmatrix} {\sigma_0}^2 & \sigma_{01} & \sigma_{02} & \sigma_{03} & \sigma_{04} \\ \sigma_{10} & {\sigma_1}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{20} & \sigma_{21} & {\sigma_2}^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{30} & \sigma_{31} & \sigma_{32} & {\sigma_3}^2 & \sigma_{34} \\ \sigma_{40} & \sigma_{41} & \sigma_{42} & \sigma_{43} & {\sigma_4}^2 \end{pmatrix} = \mathrm{L} * L'$$

we can obtain

$$\Sigma^* = \begin{pmatrix} {\sigma_0}^2 & z_1\sigma_{01} & z_2\sigma_{02} & z_3\sigma_{03} & z_4\sigma_{04} \\ z_1\sigma_{10} & {z_1}^2{\sigma_1}^2 & z_1z_2\sigma_{12} & z_1z_3\sigma_{13} & z_1z_4\sigma_{14} \\ z_2\sigma_{20} & z_1z_2\sigma_{21} & {z_2}^2{\sigma_2}^2 & z_2z_3\sigma_{23} & z_2z_4\sigma_{24} \\ z_3\sigma_{30} & z_1z_3\sigma_{31} & z_2z_3\sigma_{32} & {z_3}^2{\sigma_3}^2 & z_3z_4\sigma_{34} \\ z_4\sigma_{40} & z_1z_4\sigma_{41} & z_2z_4\sigma_{42} & z_3z_4\sigma_{43} & {z_4}^2{\sigma_4}^2 \end{pmatrix} = (\mathrm{Z} * \mathrm{L}) * (L' * Z') = \mathrm{P} * P' \tag{5.13}$$

where $\mathrm{P} = \mathrm{Z} * \mathrm{L}$ and $\mathrm{Z} = \begin{pmatrix} 1 & z_1 & z_2 & z_3 & z_4 \end{pmatrix}$.

For the variance, we have:

$$\sigma_{\vartheta_1}{}^2 = (1\ 1\ 1\ 1\ 1) * \Sigma^* * \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\sigma_{\vartheta_1}{}^2 = \widehat{\sigma_{\varepsilon_0}}^2 + \widehat{\sigma_{\varepsilon_1}}^2 z_1{}^2 + \widehat{\sigma_{\varepsilon_2}}^2 z_2{}^2 + \widehat{\sigma_{\varepsilon_3}}^2 z_3{}^2 + \widehat{\sigma_{\varepsilon_4}}^2 z_4{}^2 + 2 * \widehat{\sigma_{\varepsilon_{01}}} z_1 + 2 * \widehat{\sigma_{\varepsilon_{02}}} z_2 + 2 * \widehat{\sigma_{\varepsilon_{03}}} z_3 + 2 * \widehat{\sigma_{\varepsilon_{04}}} z_4 + 2 * \widehat{\sigma_{\varepsilon_{12}}} z_1 z_2 + 2 * \widehat{\sigma_{\varepsilon_{13}}} z_1 z_3 + 2 * \widehat{\sigma_{\varepsilon_{14}}} z_1 z_4 + 2 * \widehat{\sigma_{\varepsilon_{23}}} z_2 z_3 + 2 * \widehat{\sigma_{\varepsilon_{24}}} z_2 z_4 + 2 * \widehat{\sigma_{\varepsilon_{34}}} z_3 z_4. \tag{5.14}$$

A special case with **no correlation** among error terms, where σ_{ij} = 0, for *i, j = 0, 1, 2, 3, 4, $i \neq j$.* The details are shown as follows:

$$\mathrm{Y} = \mathrm{X} * \beta + \vartheta_2$$

$$\vartheta_2 = \varepsilon_0 z_0 + \varepsilon_1 z_1 + \varepsilon_2 z_2 + \varepsilon_3 z_3 + \varepsilon_4 z_4$$

Where $z_0 = 1,\ z_1 = D_1 * \sqrt{HDH} * \sqrt{SIZE}, z_2 = D_2,\ z_3 = D_3 * \sqrt{CDH} * \sqrt{SIZE}, z_4 = \sqrt{D_4}$.

The variance-covariance matrix of ϑ_2 becomes:

$$\Sigma_2{}^* = \begin{pmatrix} 1 & z_1 & z_2 & z_3 & z_4 \end{pmatrix} \otimes \begin{pmatrix} \sigma_0{}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_1{}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_2{}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_3{}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_4{}^2 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_0{}^2 & 0 & 0 & 0 & 0 \\ 0 & z_1\sigma_1{}^2 & 0 & 0 & 0 \\ 0 & 0 & z_2\sigma_2{}^2 & 0 & 0 \\ 0 & 0 & 0 & z_3\sigma_3{}^2 & 0 \\ 0 & 0 & 0 & 0 & z_4\sigma_4{}^2 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_0{}^2 & 0 & 0 & 0 & 0 \\ 0 & z_1{}^2\sigma_1{}^2 & 0 & 0 & 0 \\ 0 & 0 & z_2{}^2\sigma_2{}^2 & 0 & 0 \\ 0 & 0 & 0 & z_3{}^2\sigma_3{}^2 & 0 \\ 0 & 0 & 0 & 0 & z_4{}^2\sigma_4{}^2 \end{pmatrix}$$

For the variance, we have:

$$\sigma_{\vartheta_2}{}^2 = \widehat{\sigma_{\varepsilon_0}}{}^2 + \widehat{\sigma_{\varepsilon_1}}{}^2 z_1{}^2 + \widehat{\sigma_{\varepsilon_2}}{}^2 z_2{}^2 + \widehat{\sigma_{\varepsilon_3}}{}^2 z_3{}^2 + \widehat{\sigma_{\varepsilon_4}}{}^2 z_4{}^2 \quad (5.15)$$

[2]Generalized least squares estimates for coefficients

The sum of squares becomes:

S(β) =ϑ'$* \Sigma^{-1} * \vartheta$

$=(\mathrm{Y} - \mathrm{X} * \beta)' * \Sigma^{-1} * (\mathrm{Y} - \mathrm{X} * \beta)$

$=Y'\Sigma^{-1}Y - \beta'X'\Sigma^{-1}Y - Y'\Sigma^{-1}X\beta + \beta'X'\Sigma^{-1}X\beta$

Where $\vartheta = \varepsilon_0 + \varepsilon_1 * D_1 * \sqrt{HDH}*\sqrt{SIZE}+\varepsilon_2 * D_2+\varepsilon_3 * D_3 * \sqrt{CDH} *\sqrt{SIZE} + \varepsilon_4 * \sqrt{D_4}$

By differentiating the sum of squares on β, I can obtain:

$$\frac{\partial(\mathrm{S}(\beta))}{\partial\beta} = -X'\Sigma^{-1}\mathrm{Y} - Y'\Sigma^{-1}X + 2\beta'X'\Sigma^{-1}X = 0$$

$$-2(X'\Sigma^{-1}\mathrm{Y})' + 2\beta'X'\Sigma^{-1}X = 0$$

$$X'\Sigma^{-1}Y = X'\Sigma^{-1}X\beta$$

$$\hat{\beta}^* = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$$

[3] Transformed model

(1) Based on equation (5.13), I can easily calculate the P^{-1}, which can be recognized as W:

$$\Sigma^{*-1} = P'^{-1} * P^{-1} = W' * W$$

Continuing, if I want to make the new regression model as simple as the previous situation, I have to design a new style model:

$$Y^* = X^*\beta + \vartheta^*$$

Where $Y^* = WY$, $X^* = WX$ and $\vartheta^* = W\vartheta$ respectively,

and $\mathrm{Var}(\vartheta^*) = \sigma^2$.

(2) Expectation and variance of new GLS estimates

Correspondingly, I have following properties:

$$\mathrm{E}(\vartheta^*|X^*) = \mathrm{W} * \mathrm{E}(\vartheta|X^*) = 0$$

$$\mathrm{E}\left(\hat{\beta}^*\middle|X^*\right) = \mathrm{E}((X'\Sigma^{*-1}X)^{-1}X'\Sigma^{*-1}Y|X^*)$$

$$= \mathrm{E}((X'W'WX)^{-1}X'W'WY|X^*)$$

$$= \mathrm{E}(\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}Y^*|X^*)$$

$$= \beta + \mathrm{E}(\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\varepsilon_n{}^*|X^*)$$

$$= \beta$$

Therefore, $\hat{\beta}^*$ is an GLS estimate under $\mathrm{E}(\vartheta^*|X^*) = 0$.

Similarly, the variance of $\hat{\beta}^*$ becomes:

$$\mathrm{Var}\left(\hat{\beta}^*\middle|X^*\right) = \sigma^2 * \left(X^{*\prime}X^*\right)^{-1}$$

$$= \sigma^2((WX)'(WX))^{-1}$$

$$= \sigma^2(X'W'WX)^{-1}$$

$$= \sigma^2(X'\Sigma^{*-1}X)^{-1}$$

Where $\mathrm{Var}(\vartheta^*) = \sigma^2$.

(3) Variance-covariance matrix of new disturbance

$$\mathrm{Var}(\vartheta^*|\mathrm{X}^*) = \mathrm{V}(\mathrm{W}\vartheta|\mathrm{X}^*)$$

$$= \mathrm{E}(\mathrm{W}\vartheta\vartheta'W'|\mathrm{X}^*)$$

$$= \mathrm{W} * \mathrm{E}(\vartheta\vartheta'|X^*) * W'$$

$$= \sigma^2 * W\Sigma^*W'$$

$$= \sigma^2 * W * W^{-1} * W'^{-1}W'$$

$$= \sigma^2 I_{4\times4}$$

5.4.5 Application of my model in Eviews

Recall my final model framework in 5.3, I can conclude that my model is a regression with two or more explanatory variables, which is a multiple regression model. Several software packages can solve this kind of linear regression problem, including matlab, python, Eviews and R. Here, I will introduce how I combine my model and survey results by Eviews.

With regard to conducting a linear regression model, "*lm*" function is properly applied into model in Eviews to fit the linear regression model. For example, I will give an introduction in a general case (A multiple regression model with *n* explanatory variables):

lm response constant variable_1 variable_2 variable_3 … variable_n

Here, response and variable_1, variable_2, variable_3… variable_n should be replaced by the y, $\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3 \ldots \mathrm{x}_n$ in the general theoretical model respectively.

5.4.5.1 Variables

I have mentioned that many factors are considered into my model, including house age, heating system age, number of residents, time-of-use prices, annual income, demographic factors, weather factors and conservations actions.

To make these variables simple to express, I will simplify each variable as follows: (Table 5.5)

Original variable	Simplified variable
In(Total electricity consumption)	Iny
$\alpha_{10} * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	$HEAT10$
$\alpha_{112} * DA_2 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	$HOUSEAGE12$
$\alpha_{113} * DA_3 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	$HOUSEAGE13$
$\alpha_{114} * DA_4 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	$HOUSEAGE14$
$\alpha_{115} * DA_5 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	$HOUSEAGE15$
$\alpha_{121} * DHA_1 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*HA11*
$\alpha_{122} * DHA_2 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*HA12*
$\alpha_{131} * PEL_1 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*PEL11*
$\alpha_{132} * PEL_2 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*PEL12*
$\alpha_{133} * PEL_3 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*PEL13*
$\alpha_{141} * PD_1 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*AGEDIST11*
$\alpha_{142} * PD_2 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*AGEDIST12*
$\alpha_{143} * PD_3 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*AGEDIST13*
$\alpha_{144} * PD_4 * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*AGEDIST14*
$\alpha_{15} * NR * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*NR11*
$\alpha_{16} * \log(\mathrm{INC}) * \mathrm{D}_1 * \sqrt{HDH} * \sqrt{\mathrm{SIZE}}$	*INCOME11*
$\alpha_{17} * TUP * D_1 * \sqrt{HDH} * \sqrt{SIZE}$	*TUP11*
$\alpha_{18} * \mathrm{NHS} * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}}$	*NHS11*
$\alpha_{20} * D_2$	$WATER20$
$\alpha_{211} * PD_1 * D_2$	*AGEDIST21*
$\alpha_{212} * PD_2 * D_2$	*AGEDIST22*
$\alpha_{213} * PD_3 * D_2$	*AGEDIST23*
$\alpha_{214} * PD_4 * D_2$	*AGEDIST24*
$\alpha_{221} * \mathrm{PEL}_1 * \mathrm{D}_2$	$PEL21$
$\alpha_{222} * \mathrm{PEL}_2 * \mathrm{D}_2$	$PEL22$
$\alpha_{223} * \mathrm{PEL}_3 * \mathrm{D}_2$	$PEL23$
$\alpha_{23} * log(INC) * D_2$	$INCOME21$
$\alpha_{24} * TUP * D_2$	$TUP21$
$\alpha_{25} * NR * D_2$	$NR21$
$\alpha_{26} * NWH * D_2$	$NWH21$
$\alpha_{27} * DTF * D_2$	$DTF21$
$\alpha_{30} * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$COOL30$
$\alpha_{331} * DACHP_1 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$ACHP31$
$\alpha_{332} * DACHP_2 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$ACHP32$
$\alpha_{332} * DA_2 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$HOUSEAGE32$
$\alpha_{333} * DA_3 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$HOUSEAGE33$
$\alpha_{334} * DA_4 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$HOUSEAGE34$
$\alpha_{335} * DA_5 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$HOUSEAGE35$
$\alpha_{341} * PEL_1 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$PEL31$
$\alpha_{342} * PEL_2 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$PEL32$
$\alpha_{343} * PEL_3 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$PEL33$
$\alpha_{351} * PD_1 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$AGEDIST31$
$\alpha_{352} * PD_2 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$AGEDIST32$
$\alpha_{353} * PD_3 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	AGEDIST33
$\alpha_{354} * PD_4 * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$AGEDIST34$
$\alpha_{37} * TUP * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$TUP31$
$\alpha_{38} * NR * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$NR31$
$\alpha_{39} * NCS * D_3 * \sqrt{CDH} * \sqrt{SIZE}$	$NCS31$
$\alpha_{40} * \sqrt{\mathrm{D}_4}$	$OTHER40$
$\alpha_{411} * PEL_1 * \sqrt{D_4}$	$PEL41$
$\alpha_{412} * PEL_2 * \sqrt{D_4}$	*PEL42*
$\alpha_{413} * PEL_3 * \sqrt{D_4}$	*PEL43*
$\alpha_{421} * PD_1 * \sqrt{D_4}$	*AGEDIST41*
$\alpha_{421} * PD_1 * \sqrt{D_4}$	*AGEDIST42*
$\alpha_{421} * PD_1 * \sqrt{D_4}$	*AGEDIST43*
$\alpha_{421} * PD_1 * \sqrt{D_4}$	*AGEDIST44*
$\alpha_{43} * NR * \sqrt{D_4}$	*NR41*
$\alpha_{44} * log(INC) * \sqrt{D_4}$	*INCOME41*
$\alpha_{45} * TUP * \sqrt{D_4}$	*TUP41*
$\alpha_{46} * \mathrm{NOA} * \sqrt{\mathrm{D}_4}$	*NOA41*

Table 5.5 Simplified variables for significant model

I will use all of these variables in my discussion in Chapter 6.

5.5 Summary

In this chapter, I introduced the combination of the model and pilot questionnaire, including a list of definitions and calculations of variables of my model in 5.2. In section 5.3, the final model framework was determined. Furthermore, I discussed details of model estimation in 5.4. In this part, I explained the theoretical background for statistical assumptions behind my model in preparation for corresponding versions in Chapter 6.

Chapter 6 Results and Discussion

6.1 Overview

According to statistical assumptions in 5.4.4 in Chapter 5, I have applied these to my database, including OLS on the full model, GLS with correlation between error terms on the full model and GLS without correlation between error terms on full and reduced models. In section 6.2.1, I will discuss the OLS model results. From 6.2.2 to 6.2.3, I will talk about the GLS method with correlation between error terms. In part 6.2.4, I will discuss the result from GLS method without correlation between error terms. In the following section 6.2.5, I will discuss two situations by changing the variable combinations based on model in 6.2.3. I will talk about the new results for model 6.2.5 by using the estimated variance based on new residuals in 6.2.6 and dropped some variables in 6.2.7. In section 6.3, I explain why I choose the model 6.2.7 as the preferred model and conduct χ^2 tests on different combinations of demographic group factors. The results are obtained from software Eviews. The statistical comparison focusing on age and education distribution is explained in 6.4. In 6.5, I discuss the comparison of previous studies and my model, including background, factors under consideration, methodology, error term specification and limitations of my model. In this chapter, I will discuss all of results from each full or reduced model along with the attached software code as appendices.

6.2 Discussion and Analysis of Different Model Results

Figure 6.1 shows the path for my discussion and the analysis of the different model results. I classify this part into 7 sections, including OLS (ordinary least squares model), GLS (generalized least squares) with correlation between error terms by combining ${z_2}^2$ and z_2, special case with GLS by only keeping some of the statistically significant correlation, GLS without correlation between error terms, comparison between models with changes in combination and reduction of variables, discussion of results by using the chosen error correlation and discussion of results about the final reduced model.

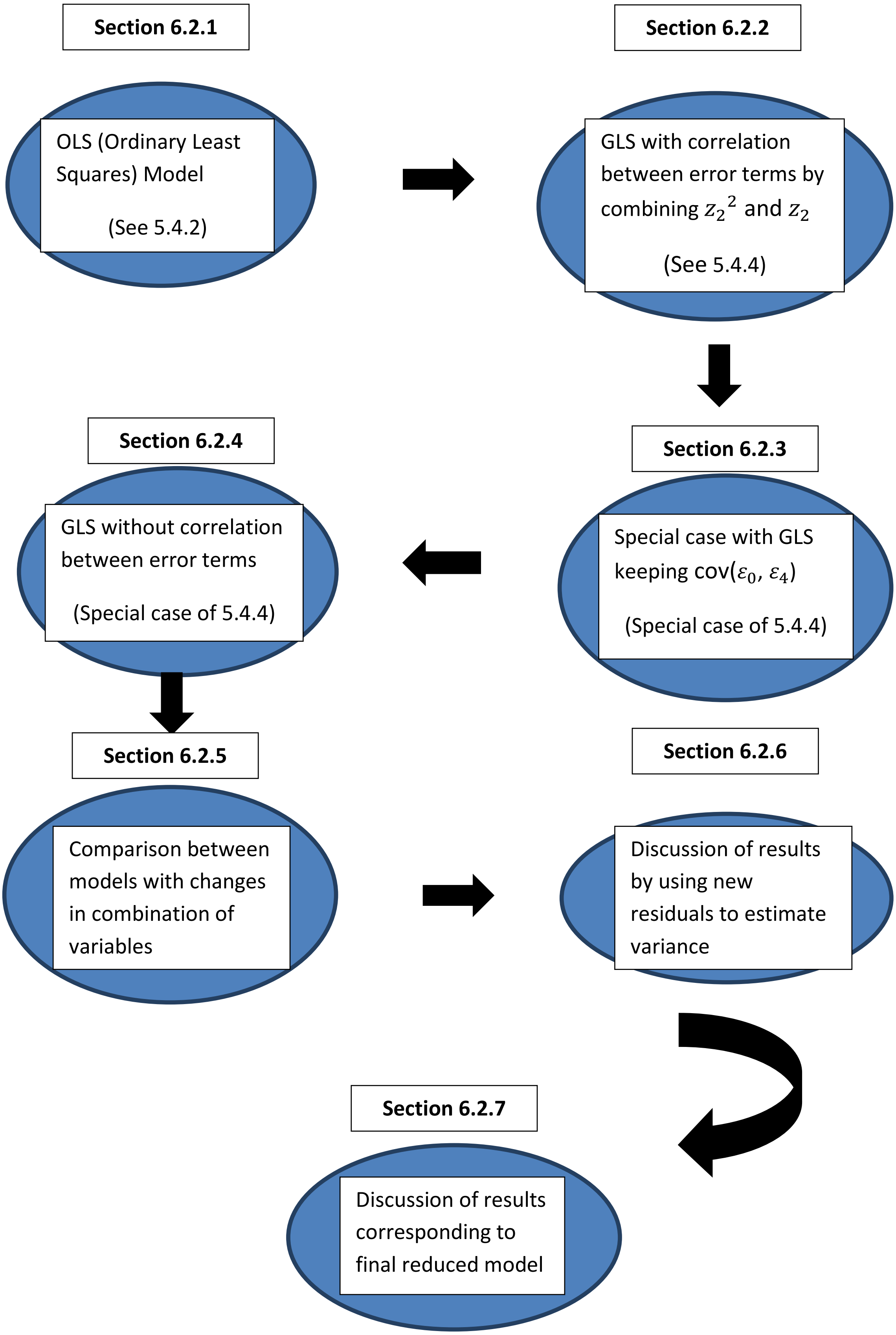


Figure 6.1 Path for my discussion and the analysis of the different model results

6.2.1 OLS (Ordinary Least Squares) Model

As discussed in the final model framework of section 5.3, I have defined each variable by multiplying itself by corresponding parameter under four systems. With respect to the OLS (Ordinary Least Squares) method, I assume that the total electricity consumption and variables follow by linear regression rule.

In this section, I add four constant coefficients in each system and NOA variable under other appliances system and combine CAC and HP as one group variable recognized as “ACHP”. (by code: see Appendix 21)

When I conduct the OLS code in Eviews, I obtain the following table of results: (Table 6.1)

	Variable	Coefficient	Std. Error	t-Statistic	Prob.
	C	8.451780	0.063458	133.1865	0.0000
HEATIN	HEAT10	9.06E-05	0.000109	0.829474	0.4071
	HOUSEAGE12	-1.30E-06	6.02E-06	-0.215342	0.8295
	HOUSEAGE13	6.38E-08	6.49E-06	0.009837	0.9922
	HOUSEAGE14	-2.74E-06	7.30E-06	-0.375258	0.7076
	HOUSEAGE15	9.16E-06	1.13E-05	0.807568	0.4195
	HA11	-3.39E-06	4.94E-06	-0.685437	0.4932
	HA12	-1.05E-05	6.29E-06	-1.671301	0.0950
	PEL11	-1.72E-05	8.73E-06	-1.968167	0.0494
	PEL12	-1.84E-05	9.05E-06	-2.035508	0.0421
	PEL13	-2.89E-05	1.14E-05	-2.530110	0.0116
	AGEDIST11	-1.91E-05	1.77E-05	-1.080482	0.2802
	AGEDIST12	-2.80E-06	1.28E-05	-0.218400	0.8272
	AGEDIST13	1.25E-06	1.14E-05	0.109473	0.9129
	AGEDIST14	1.21E-05	1.19E-05	1.011188	0.3122
	NR11	2.36E-06	2.68E-06	0.880601	0.3788
	INCOME11	-1.37E-05	1.00E-05	-1.372508	0.1702
	TUP11	0.000148	0.000817	0.181239	0.8562
	NHS11	1.99E-06	1.89E-06	1.049838	0.2941
WATER HEATIN	WATER20	-1.029711	1.252608	-0.822054	0.4113
	AGEDIST21	0.282250	0.195560	1.443287	0.1493
	AGEDIST22	0.017580	0.151575	0.115985	0.9077
	AGEDIST23	0.073857	0.141851	0.520664	0.6027
	AGEDIST24	0.172521	0.148871	1.158865	0.2468
	PEL21	-0.024859	0.099633	-0.249506	0.8030
	PEL22	-0.074963	0.100270	-0.747613	0.4549
	PEL23	-0.132136	0.124671	-1.059875	0.2895
	INCOME21	0.233358	0.110322	2.115252	0.0347
	TUP21	0.752506	8.787970	0.085629	0.9318
	NR21	0.001649	0.027287	0.060431	0.9518
	NWH21	-0.004474	0.016837	-0.265718	0.7905
	DTF21	9.84E-06	3.90E-05	0.252115	0.8010
COOLING	COOL30	-0.000780	0.000275	-2.833091	0.0047
	ACHP31	8.06E-06	9.78E-06	0.823765	0.4103
	ACHP32	-3.95E-06	8.66E-06	-0.455823	0.6486
	HOUSEAGE32	-1.30E-05	1.04E-05	-1.248547	0.2121
	HOUSEAGE33	-1.98E-06	1.25E-05	-0.157571	0.8748
	HOUSEAGE34	-2.11E-05	1.10E-05	-1.927179	0.0543
	HOUSEAGE35	8.70E-06	1.60E-05	0.544652	0.5861
	PEL31	8.75E-06	2.27E-05	0.385487	0.7000
	PEL32	-3.62E-05	2.30E-05	-1.571656	0.1164
	PEL33	-8.07E-05	2.71E-05	-2.976219	0.0030
	AGEDIST31	2.86E-05	4.35E-05	0.657790	0.5108
	AGEDIST32	-5.68E-06	3.37E-05	-0.168395	0.8663
	AGEDIST33	3.27E-05	3.18E-05	1.029126	0.3037
	AGEDIST34	3.10E-05	3.43E-05	0.906166	0.3651
	INCOME31	6.30E-05	2.69E-05	2.339619	0.0195
	TUP31	0.003967	0.001897	2.091325	0.0368
	NR31	3.57E-06	6.00E-06	0.595574	0.5516
	NCS31	-6.35E-06	4.24E-06	-1.497948	0.1345
OTHER APPLIANCE	OTHER40	0.017729	0.014001	1.266285	0.2057
	PEL41	0.000490	0.001086	0.451073	0.6520
	PEL42	0.002076	0.001133	1.832054	0.0673
	PEL43	0.003215	0.001381	2.327832	0.0201
	AGEDIST41	-0.002971	0.002149	-1.382473	0.1672
	AGEDIST42	-0.000586	0.001696	-0.345428	0.7299
	AGEDIST43	-0.001371	0.001618	-0.847052	0.3972
	AGEDIST44	-0.001609	0.001790	-0.899176	0.3688
	NR41	0.000727	0.000297	2.444272	0.0147
	INCOME41	-0.000486	0.001336	-0.364100	0.7159
	TUP41	-0.082815	0.097223	-0.851797	0.3945
	NOA41	-6.64E-05	0.000105	-0.630123	0.5288

Table 6.1 OLS coefficients

6.2.2 GLS (Generalized Linear Squares) with correlation between error terms (fifteen coefficients by combining z_2^2 and z_2)

Firstly, I need to obtain the residuals for each household from Eviews. From the OLS result table obtained from 6.2.1, I choose as follows: "View"→"Actual, Fitted, Residual"→" Actual, Fitted, Residual Table". I will obtain the residual table. (See Appendix 22)

Step 1:

Recall for equation (5.13) in chapter 5, I will discuss following situations based on equation (5.13).

Actually, I have done another version for the coefficients distribution before this version:

It included15 coefficients, I separate the coefficients of z_2^2 and z_2. However, I find that singularities exist between coefficients when I estimate the coefficients model.

After rejecting this situation above, I combine the coefficients of z_2^2 and z_2 and the model becomes:

$$\widehat{\vartheta_1}^2 = \alpha_{00}^2 + (\alpha_{01}^2 + \alpha_{11}^2)z_1^2 + (\alpha_{02}^2 + \alpha_{12}^2 + \alpha_{22}^2 + 2\alpha_{00}\alpha_{02})z_2^2 + (\alpha_{03}^2 + \alpha_{13}^2 + \alpha_{23}^2 + \alpha_{33}^2)z_3^2 + (\alpha_{04}^2 + \alpha_{14}^2 + \alpha_{24}^2 + \alpha_{34}^2 + \alpha_{44}^2)z_4^2 + 2\alpha_{00}\alpha_{01} * z_1 + 2\alpha_{00}\alpha_{03} * z_3 + 2\alpha_{00}\alpha_{04} * z_4 + 2(\alpha_{01}\alpha_{02} + \alpha_{11}\alpha_{12}) * (z_1 z_2) + 2(\alpha_{01}\alpha_{03} + \alpha_{11}\alpha_{13}) * (z_1 z_3) + 2(\alpha_{01}\alpha_{04} + \alpha_{11}\alpha_{14}) * (z_1 z_4) + 2(\alpha_{02}\alpha_{03} + \alpha_{12}\alpha_{13} + \alpha_{22}\alpha_{23}) * (z_2 z_3) + 2(\alpha_{02}\alpha_{04} + \alpha_{12}\alpha_{14} + \alpha_{22}\alpha_{24}) * (z_2 z_4) + 2(\alpha_{03}\alpha_{04} + \alpha_{13}\alpha_{14} + \alpha_{23}\alpha_{24} + \alpha_{33}\alpha_{34}) * (z_3 z_4)$$

In Eviews, I set the α_{00} as c(1), α_{01} as c(2), α_{11} as c(3), α_{02} as c(4), α_{12} as c(5), α_{22} as c(6), α_{03} as c(7), α_{13} as c(8), α_{23} as c(9), α_{33} as c(10), α_{04} as c(11), α_{14} as c(12), α_{24} as c(13), α_{34} as c(14), α_{44} as c(15). In other words, I write the equation in the equation specification:

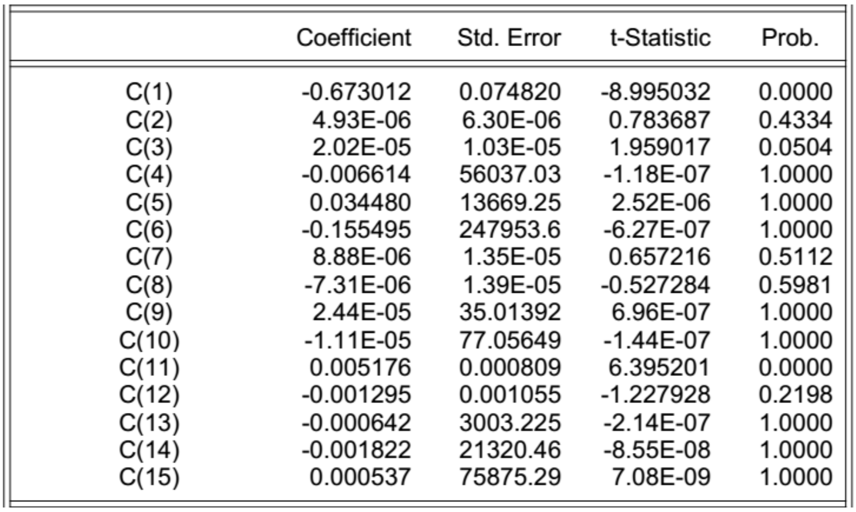
Residual21=c(1)^2+[c(2)^2+c(3)^2]*ZZ1+[c(4)^2+c(5)^2+c(6)^2+2*c(1)*c(4)]*ZZ2+[c(7)^2+c(8)^2+c(9)^2+c(10)^2]*ZZ3+[c(11)^2+c(12)^2+c(13)^2+c(14)^2+c(15)^2]*ZZ4+2*[c(1)*c(2)]*Z1+2*[c(1)*c(7)]*Z3+2*[c(1)*c(11)]*Z4+2*[c(2)*c(4)+c(3)*c(5)]*Z12+2*[c(2)*c(7)+c(3)*c(8)]*Z13+2*[c(2)*c(11)+c(3)*c(12)]*Z14+2*[c(4)*c(7)+c(5)*c(8)+c

(6)*c(9)]*Z23+2*[c(4)*c(11)+c(5)*c(12)+c(6)*c(13)]*Z24+2*[c(7)*c(11)+c(8)*c(12)+c(9)*c(13)+c(10)*c(14)]*Z34.

Where the Z1=z_1, Z2=z_2, Z3=z_3, Z4=z_4, ZZ1=${z_1}^2$, ZZ2=${z_2}^2$, ZZ3=${z_3}^2$, ZZ4=${z_4}^2$, Z12=$z_1 z_2$, Z13=$z_1 z_3$, Z14=$z_1 z_4$, Z23=$z_2 z_3$, Z24=$z_2 z_4$, Z34=$z_3 z_4$.

Residual21 is the square of the OLS residuals.

Here is the regression result: (Table 6.2)



	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.673012	0.074820	-8.995032	0.0000
C(2)	4.93E-06	6.30E-06	0.783687	0.4334
C(3)	2.02E-05	1.03E-05	1.959017	0.0504
C(4)	-0.006614	56037.03	-1.18E-07	1.0000
C(5)	0.034480	13669.25	2.52E-06	1.0000
C(6)	-0.155495	247953.6	-6.27E-07	1.0000
C(7)	8.88E-06	1.35E-05	0.657216	0.5112
C(8)	-7.31E-06	1.39E-05	-0.527284	0.5981
C(9)	2.44E-05	35.01392	6.96E-07	1.0000
C(10)	-1.11E-05	77.05649	-1.44E-07	1.0000
C(11)	0.005176	0.000809	6.395201	0.0000
C(12)	-0.001295	0.001055	-1.227928	0.2198
C(13)	-0.000642	3003.225	-2.14E-07	1.0000
C(14)	-0.001822	21320.46	-8.55E-08	1.0000
C(15)	0.000537	75875.29	7.08E-09	1.0000

Table 6.2 Variance coefficients with correlation between error terms

Step 2:

Replacing the coefficients into equation obtained in the step 1, I can obtain:

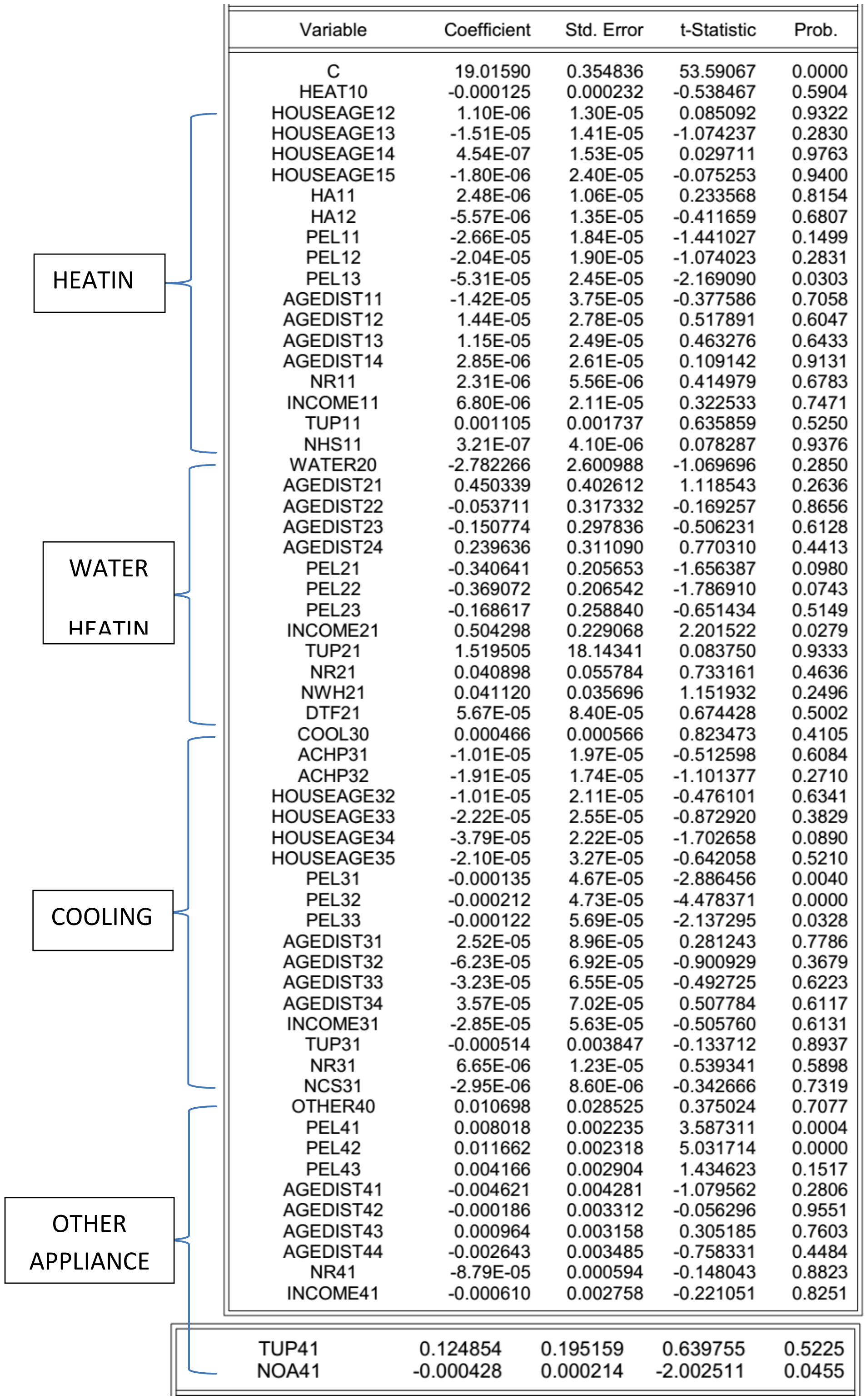
$$
\begin{aligned}
\widehat{\vartheta_1}^2 = {} & (-0.673012)^2 + (4.93\mathrm{E}-06^2 + 2.02\mathrm{E}-05^2)z_1{}^2 \\
& + ((-0.006667)^2 + 0.034493^2 + (-0.155259)^2)z_2{}^2 \\
& + (8.88\mathrm{E}-06^2 + (-7.31\mathrm{E}-06)^2 + 2.44\mathrm{E}-05^2 + 1.10\mathrm{E}-05^2)z_3{}^2 \\
& + (0.005176^2 + (-0.001295)^2 + (-0.000645)^2 + 0.001843^2 \\
& + 0.000458^2)z_4{}^2 + 2 \ast (-0.673012) \ast 4.93E-06 \ast z_1 + 2 \\
& \ast (-0.673012) \ast (-0.006667) \ast z_2 + 2 \ast (-0.673012) \ast 8.88E-06 \ast z_3 \\
& + 2 \ast (-0.673012) \ast 0.005176 \ast z_4 + 2 \ast (4.93E-06 \ast (-0.006667) \\
& + 2.02E-05 \ast 0.034493) \ast (z_1 z_2) + 2 \ast (4.93E-06 \ast 8.88E-06 \\
& + 2.02E-05 \ast (-7.31E-06)) \ast (z_1 z_3) + 2 \ast (4.93E-06 \ast 0.005176 \\
& + 2.02E-05 \ast (-0.001295)) \ast (z_1 z_4) + 2 \ast ((-0.006667) \ast 8.88E-06 \\
& + 0.034493 \ast (-7.31E-06) + (-0.155259) \ast 2.44E-05) \ast (z_2 z_3) + 2 \\
& \ast ((-0.006667) \ast 0.005176 + 0.034493 \ast (-0.001295) + (-0.155259) \\
& \ast (-0.000645)) \ast (z_2 z_4) + 2 \ast (8.88E-06 \ast 0.005176 + (-7.31E-06) \\
& \ast (-0.001295) + 2.44E-05 \ast (-0.000645) + 1.10E-05 \ast 0.001843) \\
& \ast (z_3 z_4)
\end{aligned}
$$

Where $\widehat{\alpha_{00}}$=c(1), $\widehat{\alpha_{01}}$=c(2), $\widehat{\alpha_{11}}$=c(3), $\widehat{\alpha_{02}}$=c(4), $\widehat{\alpha_{12}}$=c(5), $\widehat{\alpha_{22}}$=c(6), $\widehat{\alpha_{03}}$=c(7), $\widehat{\alpha_{13}}$=c(8), $\widehat{\alpha_{23}}$=c(9), $\widehat{\alpha_{33}}$=c(10), $\widehat{\alpha_{04}}$=c(11), $\widehat{\alpha_{14}}$=c(12), $\widehat{\alpha_{24}}$=c(13), $\widehat{\alpha_{34}}$=c(14), $\widehat{\alpha_{44}}$=c(15).

Step 3:

Calculate the $\widehat{\vartheta_1}^{2^{(-1/2)}}$ as P in the assumptions of my model in chapter 5 for 978 households. (P recognized as “hat_residula21__1_2_”, by code Appendix 23)

By multiplying the P*X and P*Y separately, I build a GLS regression model between P*Y and P*X, as Y^* and X^*. The result is as follows: (Table 6.3)



	Variable	Coefficient	Std. Error	t-Statistic	Prob.
	C	19.01590	0.354836	53.59067	0.0000
	HEAT10	-0.000125	0.000232	-0.538467	0.5904
HEATIN	HOUSEAGE12	1.10E-06	1.30E-05	0.085092	0.9322
	HOUSEAGE13	-1.51E-05	1.41E-05	-1.074237	0.2830
	HOUSEAGE14	4.54E-07	1.53E-05	0.029711	0.9763
	HOUSEAGE15	-1.80E-06	2.40E-05	-0.075253	0.9400
	HA11	2.48E-06	1.06E-05	0.233568	0.8154
	HA12	-5.57E-06	1.35E-05	-0.411659	0.6807
	PEL11	-2.66E-05	1.84E-05	-1.441027	0.1499
	PEL12	-2.04E-05	1.90E-05	-1.074023	0.2831
	PEL13	-5.31E-05	2.45E-05	-2.169090	0.0303
	AGEDIST11	-1.42E-05	3.75E-05	-0.377586	0.7058
	AGEDIST12	1.44E-05	2.78E-05	0.517891	0.6047
	AGEDIST13	1.15E-05	2.49E-05	0.463276	0.6433
	AGEDIST14	2.85E-06	2.61E-05	0.109142	0.9131
	NR11	2.31E-06	5.56E-06	0.414979	0.6783
	INCOME11	6.80E-06	2.11E-05	0.322533	0.7471
	TUP11	0.001105	0.001737	0.635859	0.5250
	NHS11	3.21E-07	4.10E-06	0.078287	0.9376
WATER HEATIN	WATER20	-2.782266	2.600988	-1.069696	0.2850
	AGEDIST21	0.450339	0.402612	1.118543	0.2636
	AGEDIST22	-0.053711	0.317332	-0.169257	0.8656
	AGEDIST23	-0.150774	0.297836	-0.506231	0.6128
	AGEDIST24	0.239636	0.311090	0.770310	0.4413
	PEL21	-0.340641	0.205653	-1.656387	0.0980
	PEL22	-0.369072	0.206542	-1.786910	0.0743
	PEL23	-0.168617	0.258840	-0.651434	0.5149
	INCOME21	0.504298	0.229068	2.201522	0.0279
	TUP21	1.519505	18.14341	0.083750	0.9333
	NR21	0.040898	0.055784	0.733161	0.4636
	NWH21	0.041120	0.035696	1.151932	0.2496
	DTF21	5.67E-05	8.40E-05	0.674428	0.5002
COOLING	COOL30	0.000466	0.000566	0.823473	0.4105
	ACHP31	-1.01E-05	1.97E-05	-0.512598	0.6084
	ACHP32	-1.91E-05	1.74E-05	-1.101377	0.2710
	HOUSEAGE32	-1.01E-05	2.11E-05	-0.476101	0.6341
	HOUSEAGE33	-2.22E-05	2.55E-05	-0.872920	0.3829
	HOUSEAGE34	-3.79E-05	2.22E-05	-1.702658	0.0890
	HOUSEAGE35	-2.10E-05	3.27E-05	-0.642058	0.5210
	PEL31	-0.000135	4.67E-05	-2.886456	0.0040
	PEL32	-0.000212	4.73E-05	-4.478371	0.0000
	PEL33	-0.000122	5.69E-05	-2.137295	0.0328
	AGEDIST31	2.52E-05	8.96E-05	0.281243	0.7786
	AGEDIST32	-6.23E-05	6.92E-05	-0.900929	0.3679
	AGEDIST33	-3.23E-05	6.55E-05	-0.492725	0.6223
	AGEDIST34	3.57E-05	7.02E-05	0.507784	0.6117
	INCOME31	-2.85E-05	5.63E-05	-0.505760	0.6131
	TUP31	-0.000514	0.003847	-0.133712	0.8937
	NR31	6.65E-06	1.23E-05	0.539341	0.5898
	NCS31	-2.95E-06	8.60E-06	-0.342666	0.7319
OTHER APPLIANCE	OTHER40	0.010698	0.028525	0.375024	0.7077
	PEL41	0.008018	0.002235	3.587311	0.0004
	PEL42	0.011662	0.002318	5.031714	0.0000
	PEL43	0.004166	0.002904	1.434623	0.1517
	AGEDIST41	-0.004621	0.004281	-1.079562	0.2806
	AGEDIST42	-0.000186	0.003312	-0.056296	0.9551
	AGEDIST43	0.000964	0.003158	0.305185	0.7603
	AGEDIST44	-0.002643	0.003485	-0.758331	0.4484
	NR41	-8.79E-05	0.000594	-0.148043	0.8823
	INCOME41	-0.000610	0.002758	-0.221051	0.8251
	TUP41	0.124854	0.195159	0.639755	0.5225
	NOA41	-0.000428	0.000214	-2.002511	0.0455

Table 6.3 GLS coefficients with correlation between error terms

6.2.3 Model with six coefficients by keeping covariance of $\boldsymbol{\varepsilon_0}$ and $\boldsymbol{\varepsilon_4}$

Based on coefficients model in 6.2.2, I will simplify the coefficients number into 14, the non-linear equation becomes:

$$\widehat{\vartheta_1}^2 = \alpha_0 + \alpha_1 {z_1}^2 + \alpha_2 {z_2}^2 + \alpha_3 {z_3}^2 + \alpha_4 {z_4}^2 + \alpha_5 * z_1 + \alpha_6 * z_3 + \alpha_7 * z_4 + \alpha_8 * (z_1 z_2) + \alpha_9 * (z_1 z_3) + \alpha_{10} * (z_1 z_4) + \alpha_{11} * (z_2 z_3) + \alpha_{12} * (z_2 z_4) + \alpha_{13} * (z_3 z_4)$$

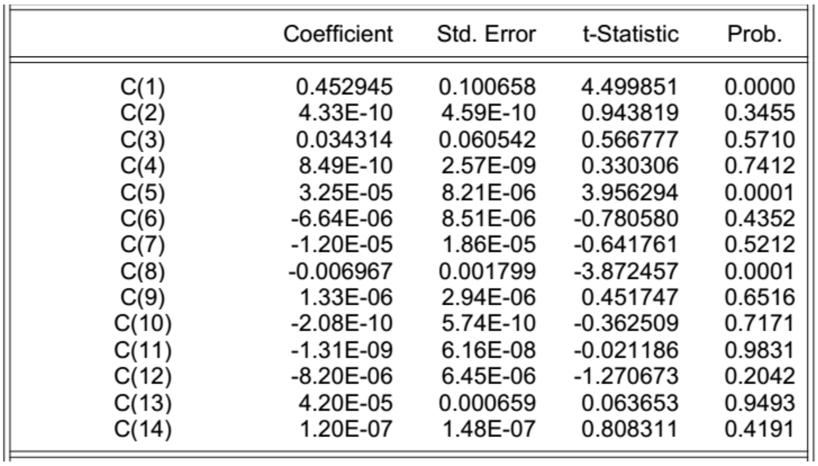
In Eviews, I set the α_0 as c(1), α_1 as c(2), α_2 as c(3), α_3 as c(4), α_4 as c(5), α_5 as c(6), α_6 as c(7), α_7 as c(8), α_8 as c(9), α_9 as c(10), α_{10} as c(11), α_{11} as c(12), α_{12} as c(13), α_{13} as c(14). In other words, I write the equation in the equation specification:

Residual21=c(1)+c(2)*ZZ1+c(3)*ZZ2+c(4)*ZZ3+c(5)*ZZ4+c(6)*Z1+c(7)*Z3+c(8)*Z4+c(9)*Z12+c(10)*Z13+c(11)*Z14+c(12)*Z23+c(13)*Z24+c(14)*Z34.

Where the Z1=z_1, Z2=z_2, Z3=z_3, Z4=z_4, ZZ1=${z_1}^2$, ZZ2=${z_2}^2$, ZZ3=${z_3}^2$, ZZ4=${z_4}^2$, Z12=$z_1 z_2$, Z13=$z_1 z_3$, Z14=$z_1 z_4$, Z23=$z_2 z_3$, Z24=$z_2 z_4$, Z34=$z_3 z_4$.

Residual21 is the square of the OLS residuals.

Here is the regression result:



	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.452945	0.100658	4.499851	0.0000
C(2)	4.33E-10	4.59E-10	0.943819	0.3455
C(3)	0.034314	0.060542	0.566777	0.5710
C(4)	8.49E-10	2.57E-09	0.330306	0.7412
C(5)	3.25E-05	8.21E-06	3.956294	0.0001
C(6)	-6.64E-06	8.51E-06	-0.780580	0.4352
C(7)	-1.20E-05	1.86E-05	-0.641761	0.5212
C(8)	-0.006967	0.001799	-3.872457	0.0001
C(9)	1.33E-06	2.94E-06	0.451747	0.6516
C(10)	-2.08E-10	5.74E-10	-0.362509	0.7171
C(11)	-1.31E-09	6.16E-08	-0.021186	0.9831
C(12)	-8.20E-06	6.45E-06	-1.270673	0.2042
C(13)	4.20E-05	0.000659	0.063653	0.9493
C(14)	1.20E-07	1.48E-07	0.808311	0.4191

Now, I conduct a χ^2 test on $\alpha_5 = \alpha_6 = \alpha_8 = \alpha_9 = \alpha_{10} = \alpha_{11} = \alpha_{12} = \alpha_{13} = 0$, I obtain a p-value of 0.7716. I cannot reject the null hypothesis. Therefore, I consider reducing the non-linear regression model when I estimate the coefficients. I only keep the variables with quadratic terms and first degree related to α_{04} and α_{44}. Now, the equation of variance becomes like this:

Step 1:

Recall for equation (5.13) in chapter 5:

$$\sigma_{\vartheta_1}{}^2 = \widehat{\sigma_{\varepsilon_0}}^2 + \widehat{\sigma_{\varepsilon_1}}^2 z_1{}^2 + \widehat{\sigma_{\varepsilon_2}}^2 z_2{}^2 + \widehat{\sigma_{\varepsilon_3}}^2 z_3{}^2 + \widehat{\sigma_{\varepsilon_4}}^2 z_4{}^2 + 2 * \widehat{\sigma_{\varepsilon_{01}}} z_1 + 2 * \widehat{\sigma_{\varepsilon_{02}}} z_2 + 2 * \widehat{\sigma_{\varepsilon_{03}}} z_3 + 2 * \widehat{\sigma_{\varepsilon_{04}}} z_4 + 2 * \widehat{\sigma_{\varepsilon_{12}}} z_1 z_2 + 2 * \widehat{\sigma_{\varepsilon_{13}}} z_1 z_3 + 2 * \widehat{\sigma_{\varepsilon_{14}}} z_1 z_4 + 2 * \widehat{\sigma_{\varepsilon_{23}}} z_2 z_3 + 2 * \widehat{\sigma_{\varepsilon_{24}}} z_2 z_4 + 2 * \widehat{\sigma_{\varepsilon_{34}}} z_3 z_4.$$

Similarly, I can obtain:

$$\begin{aligned}\widehat{\vartheta_1}^2 = \alpha_{00}{}^2 &+ (\alpha_{01}{}^2 + \alpha_{11}{}^2) z_1{}^2 + (\alpha_{02}{}^2 + \alpha_{12}{}^2 + \alpha_{22}{}^2 + 2\alpha_{00}\alpha_{02}) z_2{}^2 \\ &+ (\alpha_{03}{}^2 + \alpha_{13}{}^2 + \alpha_{23}{}^2 + \alpha_{33}{}^2) z_3{}^2 \\ &+ (\alpha_{04}{}^2 + \alpha_{14}{}^2 + \alpha_{24}{}^2 + \alpha_{34}{}^2 + \alpha_{44}{}^2) z_4{}^2 + 2\alpha_{00}\alpha_{01} * z_1 + 2\alpha_{00}\alpha_{03} * z_3 \\ &+ 2\alpha_{00}\alpha_{04} * z_4 + 2(\alpha_{01}\alpha_{02} + \alpha_{11}\alpha_{12}) * (z_1 z_2) + 2(\alpha_{01}\alpha_{03} + \alpha_{11}\alpha_{13}) \\ &* (z_1 z_3) + 2(\alpha_{01}\alpha_{04} + \alpha_{11}\alpha_{14}) * (z_1 z_4) + 2(\alpha_{02}\alpha_{03} + \alpha_{12}\alpha_{13} + \alpha_{22}\alpha_{23}) \\ &* (z_2 z_3) + 2(\alpha_{02}\alpha_{04} + \alpha_{12}\alpha_{14} + \alpha_{22}\alpha_{24}) * (z_2 z_4) + 2(\alpha_{03}\alpha_{04} + \alpha_{13}\alpha_{14} \\ &+ \alpha_{23}\alpha_{24} + \alpha_{33}\alpha_{34}) * (z_3 z_4)\end{aligned}$$

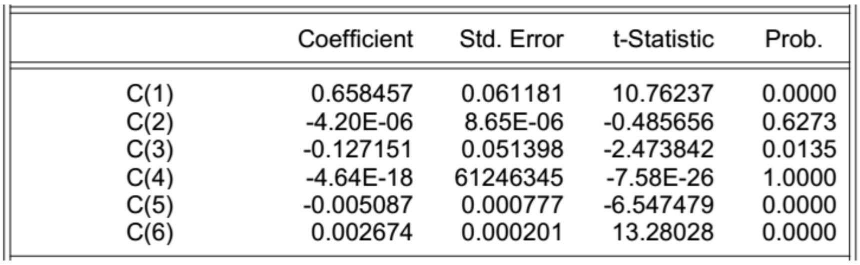
However, I only keep coefficients related to $\boldsymbol{\alpha_{04}}$ and $\boldsymbol{\alpha_{44}}$ and obtain following:

$$\widehat{\vartheta_1}^2 = \alpha_{00}{}^2 + \alpha_{11}{}^2 z_1{}^2 + \alpha_{22}{}^2 z_2{}^2 + \alpha_{33}{}^2 z_3{}^2 + (\alpha_{04}{}^2 + \alpha_{44}{}^2) z_4{}^2 + 2\alpha_{00}\alpha_{04} * z_4$$

In Eviews, I set the α_{00} as c(1), α_{11} as c(2), α_{22} as c(3), α_{33} as c(4), α_{04} as c(5), α_{44} as c(6). In other words, I write the equation in the equation specification:

Residual21=c(1)^2+c(2)^2*ZZ1+c(3)^2*ZZ2+c(4)^2*ZZ3+[c(5)^2+c(6)^2]*ZZ4+ 2*[c(1)*c(5)]*Z4.

Where the Z4=z_4, ZZ1=$z_1{}^2$, ZZ2=$z_2{}^2$, ZZ3=$z_3{}^2$, ZZ4=$z_4{}^2$. (Table 6.4)

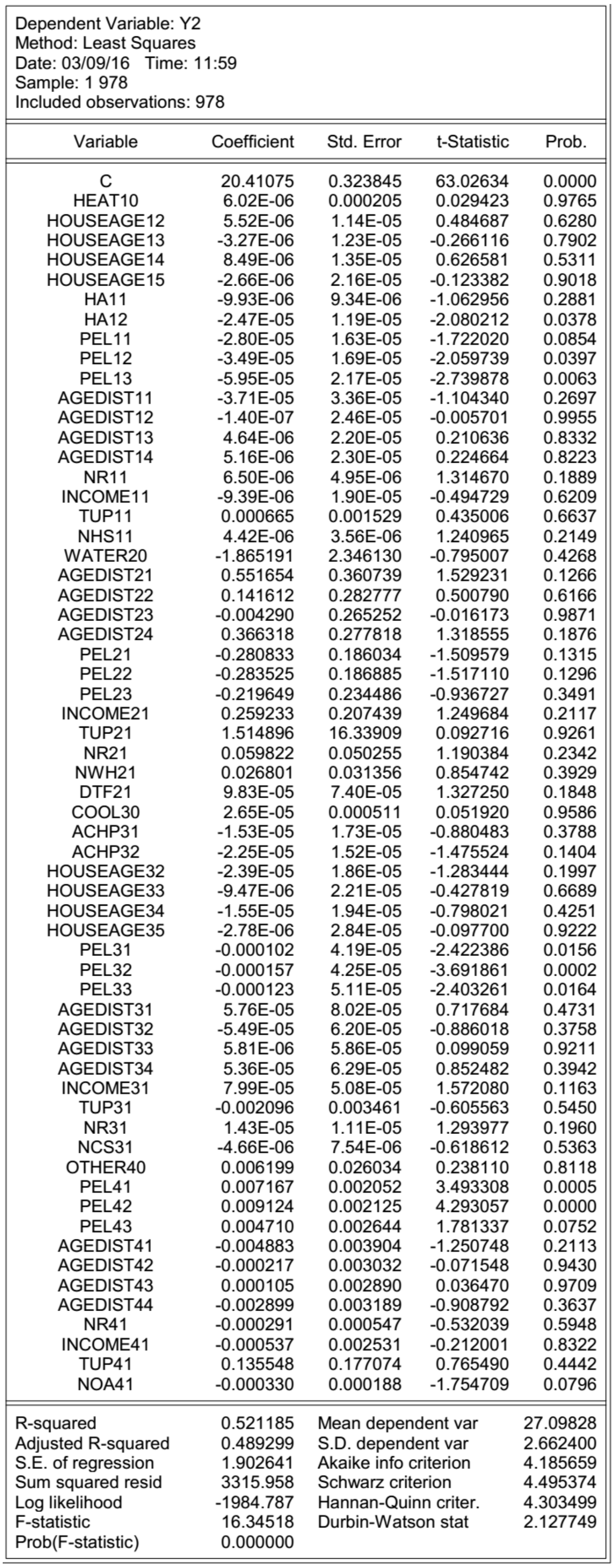


	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.658457	0.061181	10.76237	0.0000
C(2)	-4.20E-06	8.65E-06	-0.485656	0.6273
C(3)	-0.127151	0.051398	-2.473842	0.0135
C(4)	-4.64E-18	61246345	-7.58E-26	1.0000
C(5)	-0.005087	0.000777	-6.547479	0.0000
C(6)	0.002674	0.000201	13.28028	0.0000

Table 6.4 Variance coefficients with covariance of ε_0 and ε_4

Step 2:

After I reduce the number of coefficients in the variance equation into six, the new corresponding results are as follows: (Table 6.5)



```
Dependent Variable: Y2
Method: Least Squares
Date: 03/09/16   Time: 11:59
Sample: 1 978
Included observations: 978

      Variable      Coefficient   Std. Error   t-Statistic   Prob.

         C           20.41075     0.323845     63.02634     0.0000
       HEAT10        6.02E-06     0.000205     0.029423     0.9765
    HOUSEAGE12       5.52E-06     1.14E-05     0.484687     0.6280
    HOUSEAGE13      -3.27E-06     1.23E-05    -0.266116     0.7902
    HOUSEAGE14       8.49E-06     1.35E-05     0.626581     0.5311
    HOUSEAGE15      -2.66E-06     2.16E-05    -0.123382     0.9018
        HA11        -9.93E-06     9.34E-06    -1.062956     0.2881
        HA12        -2.47E-05     1.19E-05    -2.080212     0.0378
       PEL11        -2.80E-05     1.63E-05    -1.722020     0.0854
       PEL12        -3.49E-05     1.69E-05    -2.059739     0.0397
       PEL13        -5.95E-05     2.17E-05    -2.739878     0.0063
     AGEDIST11      -3.71E-05     3.36E-05    -1.104340     0.2697
     AGEDIST12      -1.40E-07     2.46E-05    -0.005701     0.9955
     AGEDIST13       4.64E-06     2.20E-05     0.210636     0.8332
     AGEDIST14       5.16E-06     2.30E-05     0.224664     0.8223
        NR11         6.50E-06     4.95E-06     1.314670     0.1889
      INCOME11      -9.39E-06     1.90E-05    -0.494729     0.6209
       TUP11         0.000665     0.001529     0.435006     0.6637
       NHS11         4.42E-06     3.56E-06     1.240965     0.2149
      WATER20       -1.865191     2.346130    -0.795007     0.4268
     AGEDIST21       0.551654     0.360739     1.529231     0.1266
     AGEDIST22       0.141612     0.282777     0.500790     0.6166
     AGEDIST23      -0.004290     0.265252    -0.016173     0.9871
     AGEDIST24       0.366318     0.277818     1.318555     0.1876
       PEL21        -0.280833     0.186034    -1.509579     0.1315
       PEL22        -0.283525     0.186885    -1.517110     0.1296
       PEL23        -0.219649     0.234486    -0.936727     0.3491
      INCOME21       0.259233     0.207439     1.249684     0.2117
       TUP21         1.514896     16.33909     0.092716     0.9261
        NR21         0.059822     0.050255     1.190384     0.2342
       NWH21         0.026801     0.031356     0.854742     0.3929
       DTF21         9.83E-05     7.40E-05     1.327250     0.1848
       COOL30        2.65E-05     0.000511     0.051920     0.9586
       ACHP31       -1.53E-05     1.73E-05    -0.880483     0.3788
       ACHP32       -2.25E-05     1.52E-05    -1.475524     0.1404
    HOUSEAGE32      -2.39E-05     1.86E-05    -1.283444     0.1997
    HOUSEAGE33      -9.47E-06     2.21E-05    -0.427819     0.6689
    HOUSEAGE34      -1.55E-05     1.94E-05    -0.798021     0.4251
    HOUSEAGE35      -2.78E-06     2.84E-05    -0.097700     0.9222
       PEL31        -0.000102     4.19E-05    -2.422386     0.0156
       PEL32        -0.000157     4.25E-05    -3.691861     0.0002
       PEL33        -0.000123     5.11E-05    -2.403261     0.0164
     AGEDIST31       5.76E-05     8.02E-05     0.717684     0.4731
     AGEDIST32      -5.49E-05     6.20E-05    -0.886018     0.3758
     AGEDIST33       5.81E-06     5.86E-05     0.099059     0.9211
     AGEDIST34       5.36E-05     6.29E-05     0.852482     0.3942
      INCOME31       7.99E-05     5.08E-05     1.572080     0.1163
       TUP31        -0.002096     0.003461    -0.605563     0.5450
        NR31         1.43E-05     1.11E-05     1.293977     0.1960
       NCS31        -4.66E-06     7.54E-06    -0.618612     0.5363
      OTHER40        0.006199     0.026034     0.238110     0.8118
       PEL41         0.007167     0.002052     3.493308     0.0005
       PEL42         0.009124     0.002125     4.293057     0.0000
       PEL43         0.004710     0.002644     1.781337     0.0752
     AGEDIST41      -0.004883     0.003904    -1.250748     0.2113
     AGEDIST42      -0.000217     0.003032    -0.071548     0.9430
     AGEDIST43       0.000105     0.002890     0.036470     0.9709
     AGEDIST44      -0.002899     0.003189    -0.908792     0.3637
        NR41        -0.000291     0.000547    -0.532039     0.5948
      INCOME41      -0.000537     0.002531    -0.212001     0.8322
       TUP41         0.135548     0.177074     0.765490     0.4442
       NOA41        -0.000330     0.000188    -1.754709     0.0796

R-squared            0.521185   Mean dependent var      27.09828
Adjusted R-squared   0.489299   S.D. dependent var      2.662400
S.E. of regression   1.902641   Akaike info criterion   4.185659
Sum squared resid    3315.958   Schwarz criterion       4.495374
Log likelihood      -1984.787   Hannan-Quinn criter.    4.303499
F-statistic          16.34518   Durbin-Watson stat      2.127749
Prob(F-statistic)    0.000000
```

Table 6.5 GLS coefficients with covariance of ε_0 and ε_4

6.2.4 GLS (Generalized Linear Squares) without correlation between error terms (five coefficients)

As the special case discussed in section 5.4.4, one of the assumptions is no correlation between error terms. I assume that there is one error term under each system and no correlation between each error term. In this part, I will assume that α_{04} equal to 0 so that only five quadratic terms are left.

Firstly, I obtain the residuals for each household from Eviews, which is exactly same with 6.2.1.

Step 1:

Recall for equation (5.14) in chapter 5, I use the residuals of 978 households from GLS result and build a non-linear regression model as follows:

$$\widehat{\vartheta_1}^2 = {\beta_0}^2 + {\beta_1}^2{z_1}^2 + {\beta_2}^2{z_2}^2 + {\beta_3}^2{z_3}^2 + {\beta_4}^2{z_4}^2$$

But I can easily find that c(6) is statistically significant (t-Statistic of c(6) is 13.280828). So I keep function connected with equation (5.13).

6.2.5 Models with combinations of House age and Age distribution based on Model in 6.2.3

I have discussed in 6.2.3 GLS with correlation between α_{00} and α_{04} and in 6.2.4 GLS without correlation between residuals. I find that the p-value of c(5) (namely, α_{04}) is 0.0000, which means that this variable is a significant factor in estimating residual in the non-linear regression model. I can't ignore this variable when I estimate the variance. Therefore, I will keep α_{04} in my model and make Model in 6.2.3 (GLS with correlation between α_{00} and α_{04}) as my following model.

From the results in 6.2.3, I find that education factors seem much more important than Age distribution and Houseage to the total electricity consumption. I also feel that I need to combine some variables as one group variable because there are too many variables. My action is mainly focusing on changing the combination of Age distribution and Houseage. I have summarized into two situations I have conducted:

Note: Here, I continue to use the variance I have previously estimated.

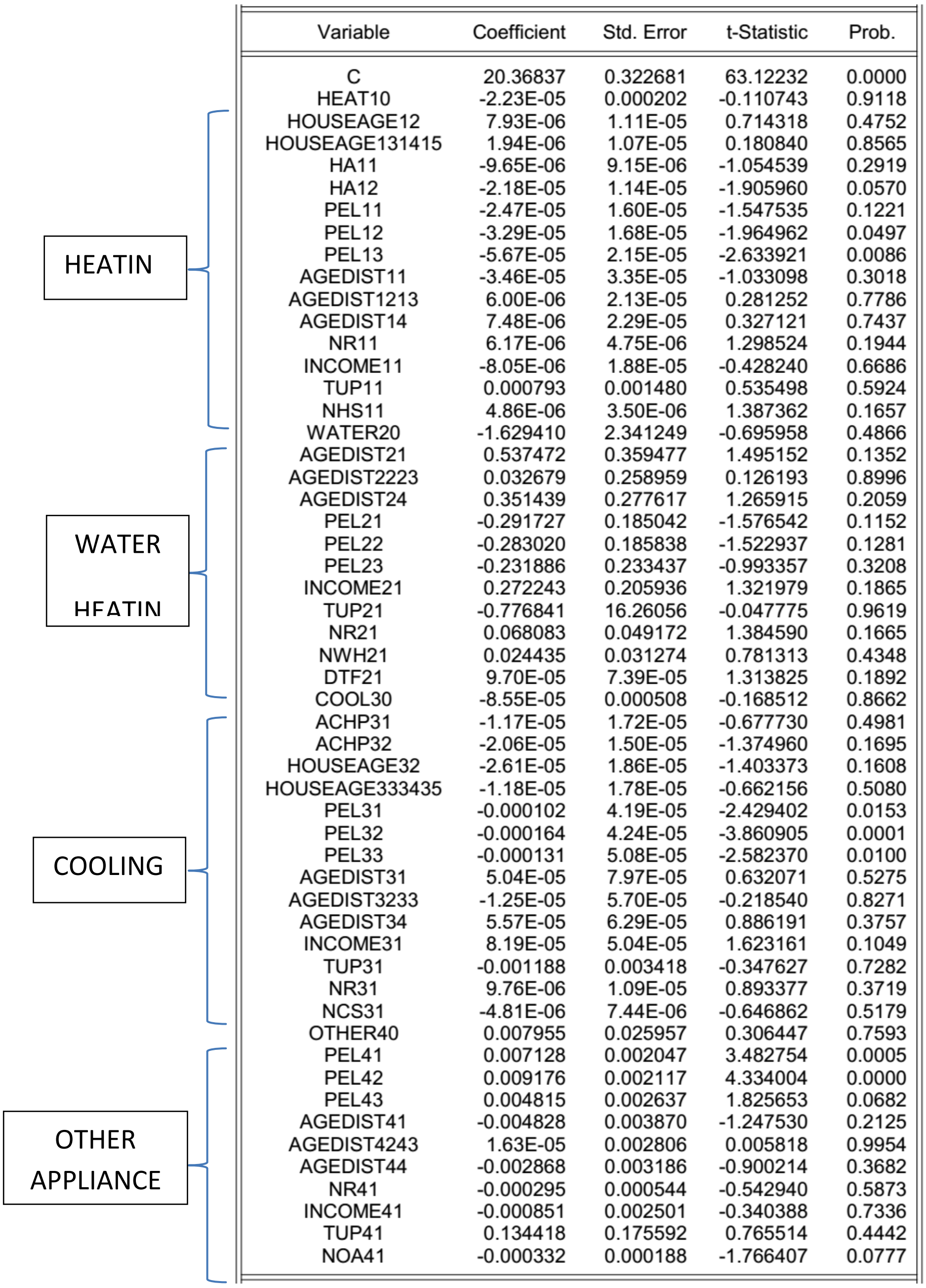
(1) Combine Age Distribution Group 2&3

Based on the result in 6.2.3, I find that the p-value of Age distribution group 2 and 3 variables under four systems are both insignificant to total electricity consumption. I combined these two group variables. Similar to the steps with previous versions, I obtain the following result: (Table 6.6)

	Variable	Coefficient	Std. Error	t-Statistic	Prob.
	C	20.37889	0.323884	62.92026	0.0000
HEATIN	HEAT10	-3.70E-05	0.000203	-0.182454	0.8553
	HOUSEAGE12	8.04E-06	1.11E-05	0.722784	0.4700
	HOUSEAGE13	-3.12E-06	1.23E-05	-0.254266	0.7993
	HOUSEAGE14	9.54E-06	1.34E-05	0.713369	0.4758
	HOUSEAGE15	-1.89E-06	2.16E-05	-0.087389	0.9304
	HA11	-1.00E-05	9.30E-06	-1.080968	0.2800
	HA12	-2.45E-05	1.19E-05	-2.064594	0.0392
	PEL11	-2.71E-05	1.63E-05	-1.662269	0.0968
	PEL12	-3.45E-05	1.70E-05	-2.030753	0.0426
	PEL13	-5.88E-05	2.17E-05	-2.707172	0.0069
	AGEDIST11	-3.47E-05	3.36E-05	-1.033014	0.3019
	AGEDIST1213	3.86E-06	2.16E-05	0.179122	0.8579
	AGEDIST14	5.91E-06	2.30E-05	0.256927	0.7973
	NR11	5.99E-06	4.77E-06	1.254688	0.2099
	INCOME11	-9.07E-06	1.89E-05	-0.480501	0.6310
	TUP11	0.000982	0.001502	0.653739	0.5134
	NHS11	5.08E-06	3.54E-06	1.437910	0.1508
WATER HEATIN	WATER20	-1.765506	2.349244	-0.751521	0.4525
	AGEDIST21	0.553402	0.360616	1.534601	0.1252
	AGEDIST2223	0.042306	0.259689	0.162908	0.8706
	AGEDIST24	0.362310	0.278388	1.301455	0.1934
	PEL21	-0.276266	0.186266	-1.483181	0.1384
	PEL22	-0.271485	0.186912	-1.452479	0.1467
	PEL23	-0.228395	0.234394	-0.974407	0.3301
	INCOME21	0.282915	0.206838	1.367810	0.1717
	TUP21	-0.268613	16.30260	-0.016477	0.9869
	NR21	0.066848	0.049301	1.355926	0.1755
	NWH21	0.025599	0.031379	0.815783	0.4148
	DTF21	9.92E-05	7.42E-05	1.337195	0.1815
COOLING	COOL30	-9.68E-05	0.000509	-0.190092	0.8493
	ACHP31	-1.20E-05	1.73E-05	-0.691395	0.4895
	ACHP32	-2.03E-05	1.52E-05	-1.332489	0.1830
	HOUSEAGE32	-2.60E-05	1.86E-05	-1.396218	0.1630
	HOUSEAGE33	-6.09E-06	2.21E-05	-0.274906	0.7835
	HOUSEAGE34	-1.59E-05	1.95E-05	-0.818316	0.4134
	HOUSEAGE35	-9.47E-06	2.84E-05	-0.333865	0.7386
	PEL31	-0.000101	4.20E-05	-2.412472	0.0160
	PEL32	-0.000162	4.25E-05	-3.817836	0.0001
	PEL33	-0.000130	5.10E-05	-2.548140	0.0110
	AGEDIST31	4.83E-05	7.99E-05	0.604114	0.5459
	AGEDIST3233	-1.41E-05	5.72E-05	-0.246202	0.8056
	AGEDIST34	5.39E-05	6.30E-05	0.854914	0.3928
	INCOME31	8.15E-05	5.06E-05	1.609562	0.1078
	TUP31	-0.001081	0.003428	-0.315216	0.7527
	NR31	9.99E-06	1.10E-05	0.912004	0.3620
	NCS31	-5.35E-06	7.54E-06	-0.709495	0.4782
OTHER APPLIANCE	OTHER40	0.008972	0.026022	0.344788	0.7303
	PEL41	0.007037	0.002053	3.428056	0.0006
	PEL42	0.009059	0.002124	4.264553	0.0000
	PEL43	0.004790	0.002645	1.811405	0.0704
	AGEDIST41	-0.004833	0.003879	-1.246065	0.2131
	AGEDIST4243	-1.69E-05	0.002814	-0.006007	0.9952
	AGEDIST44	-0.002912	0.003196	-0.911199	0.3624
	NR41	-0.000304	0.000545	-0.558909	0.5764
	INCOME41	-0.000843	0.002507	-0.336373	0.7367
	TUP41	0.126764	0.176048	0.720052	0.4717
	NOA41	-0.000328	0.000188	-1.740513	0.0821

Table 6.6 GLS coefficients with combination of Agedist Group 2 & 3

(2) Combine Age distribution Group 2&3 and House age Group 3&4&5
In (1), I have combined Age distribution group 2&3 and find that the combined variable's importance is enhanced compared with separate variables. I also find that the House age variable also shows a very insignificant effect. In this section, I combine the House age group 3&4&5. The following shows the result: (Table 6.7)



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.36837	0.322681	63.12232	0.0000
HEAT10	-2.23E-05	0.000202	-0.110743	0.9118
HOUSEAGE12	7.93E-06	1.11E-05	0.714318	0.4752
HOUSEAGE131415	1.94E-06	1.07E-05	0.180840	0.8565
HA11	-9.65E-06	9.15E-06	-1.054539	0.2919
HA12	-2.18E-05	1.14E-05	-1.905960	0.0570
PEL11	-2.47E-05	1.60E-05	-1.547535	0.1221
PEL12	-3.29E-05	1.68E-05	-1.964962	0.0497
PEL13	-5.67E-05	2.15E-05	-2.633921	0.0086
AGEDIST11	-3.46E-05	3.35E-05	-1.033098	0.3018
AGEDIST1213	6.00E-06	2.13E-05	0.281252	0.7786
AGEDIST14	7.48E-06	2.29E-05	0.327121	0.7437
NR11	6.17E-06	4.75E-06	1.298524	0.1944
INCOME11	-8.05E-06	1.88E-05	-0.428240	0.6686
TUP11	0.000793	0.001480	0.535498	0.5924
NHS11	4.86E-06	3.50E-06	1.387362	0.1657
WATER20	-1.629410	2.341249	-0.695958	0.4866
AGEDIST21	0.537472	0.359477	1.495152	0.1352
AGEDIST2223	0.032679	0.258959	0.126193	0.8996
AGEDIST24	0.351439	0.277617	1.265915	0.2059
PEL21	-0.291727	0.185042	-1.576542	0.1152
PEL22	-0.283020	0.185838	-1.522937	0.1281
PEL23	-0.231886	0.233437	-0.993357	0.3208
INCOME21	0.272243	0.205936	1.321979	0.1865
TUP21	-0.776841	16.26056	-0.047775	0.9619
NR21	0.068083	0.049172	1.384590	0.1665
NWH21	0.024435	0.031274	0.781313	0.4348
DTF21	9.70E-05	7.39E-05	1.313825	0.1892
COOL30	-8.55E-05	0.000508	-0.168512	0.8662
ACHP31	-1.17E-05	1.72E-05	-0.677730	0.4981
ACHP32	-2.06E-05	1.50E-05	-1.374960	0.1695
HOUSEAGE32	-2.61E-05	1.86E-05	-1.403373	0.1608
HOUSEAGE333435	-1.18E-05	1.78E-05	-0.662156	0.5080
PEL31	-0.000102	4.19E-05	-2.429402	0.0153
PEL32	-0.000164	4.24E-05	-3.860905	0.0001
PEL33	-0.000131	5.08E-05	-2.582370	0.0100
AGEDIST31	5.04E-05	7.97E-05	0.632071	0.5275
AGEDIST3233	-1.25E-05	5.70E-05	-0.218540	0.8271
AGEDIST34	5.57E-05	6.29E-05	0.886191	0.3757
INCOME31	8.19E-05	5.04E-05	1.623161	0.1049
TUP31	-0.001188	0.003418	-0.347627	0.7282
NR31	9.76E-06	1.09E-05	0.893377	0.3719
NCS31	-4.81E-06	7.44E-06	-0.646862	0.5179
OTHER40	0.007955	0.025957	0.306447	0.7593
PEL41	0.007128	0.002047	3.482754	0.0005
PEL42	0.009176	0.002117	4.334004	0.0000
PEL43	0.004815	0.002637	1.825653	0.0682
AGEDIST41	-0.004828	0.003870	-1.247530	0.2125
AGEDIST4243	1.63E-05	0.002806	0.005818	0.9954
AGEDIST44	-0.002868	0.003186	-0.900214	0.3682
NR41	-0.000295	0.000544	-0.542940	0.5873
INCOME41	-0.000851	0.002501	-0.340388	0.7336
TUP41	0.134418	0.175592	0.765514	0.4442
NOA41	-0.000332	0.000188	-1.766407	0.0777

Table 6.7 GLS coefficients with combination of Agedist Group 2 & 3 and Houseage Group 3&4&5

6.2.6 Reestimation of model with estimated variance using residuals of Model in 6.2.5 (2)

During the discussion in 6.2.5, I discovered that the combination of Age distribution group 2&3 and Houseage group 3&4&5 (2) exerts relatively important role in affecting electricity consumption than only combination of Age distribution group 2&3 (1). However, I kept the old residuals when I applied the estimated the new variance. In this section, I use the new estimated residual to reestimate the regression model.

In this part, I obtain the new estimated variance based on the new model’s errors. Now, the equation becomes the following for the estimated variance:

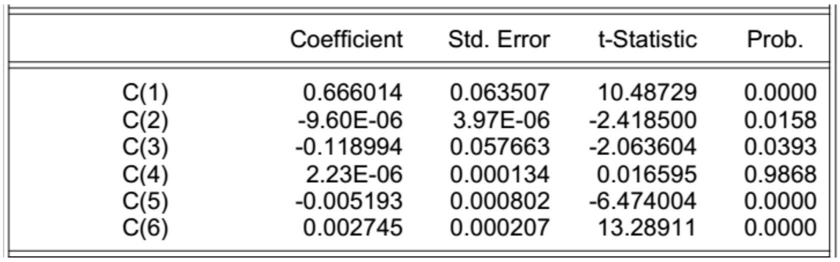
Step 1:

$$\widehat{\vartheta_1}^2 = \alpha_{00}{}^2 + \alpha_{11}{}^2 z_1{}^2 + \alpha_{22}{}^2 z_2{}^2 + \alpha_{33}{}^2 z_3{}^2 + (\alpha_{04}{}^2 + \alpha_{44}{}^2) z_4{}^2 + 2\alpha_{00}\alpha_{04} * z_4$$

In Eviews, I set the α_{00} as c(1), α_{11} as c(2), α_{22} as c(3), α_{33} as c(4), α_{04} as c(5), α_{44} as c(6). In other words, I write the equation in the equation specification:

Residual21=c(1)^2+c(2)^2*ZZ1+c(3)^2*ZZ2+c(4)^2*ZZ3+[c(5)^2+c(6)^2]*ZZ4+2*[c(1)*c(5)]*Z4.

Where the Z4=z_4, ZZ1=$z_1{}^2$, ZZ2=$z_2{}^2$, ZZ3=$z_3{}^2$, ZZ4=$z_4{}^2$. (Table 6.8)

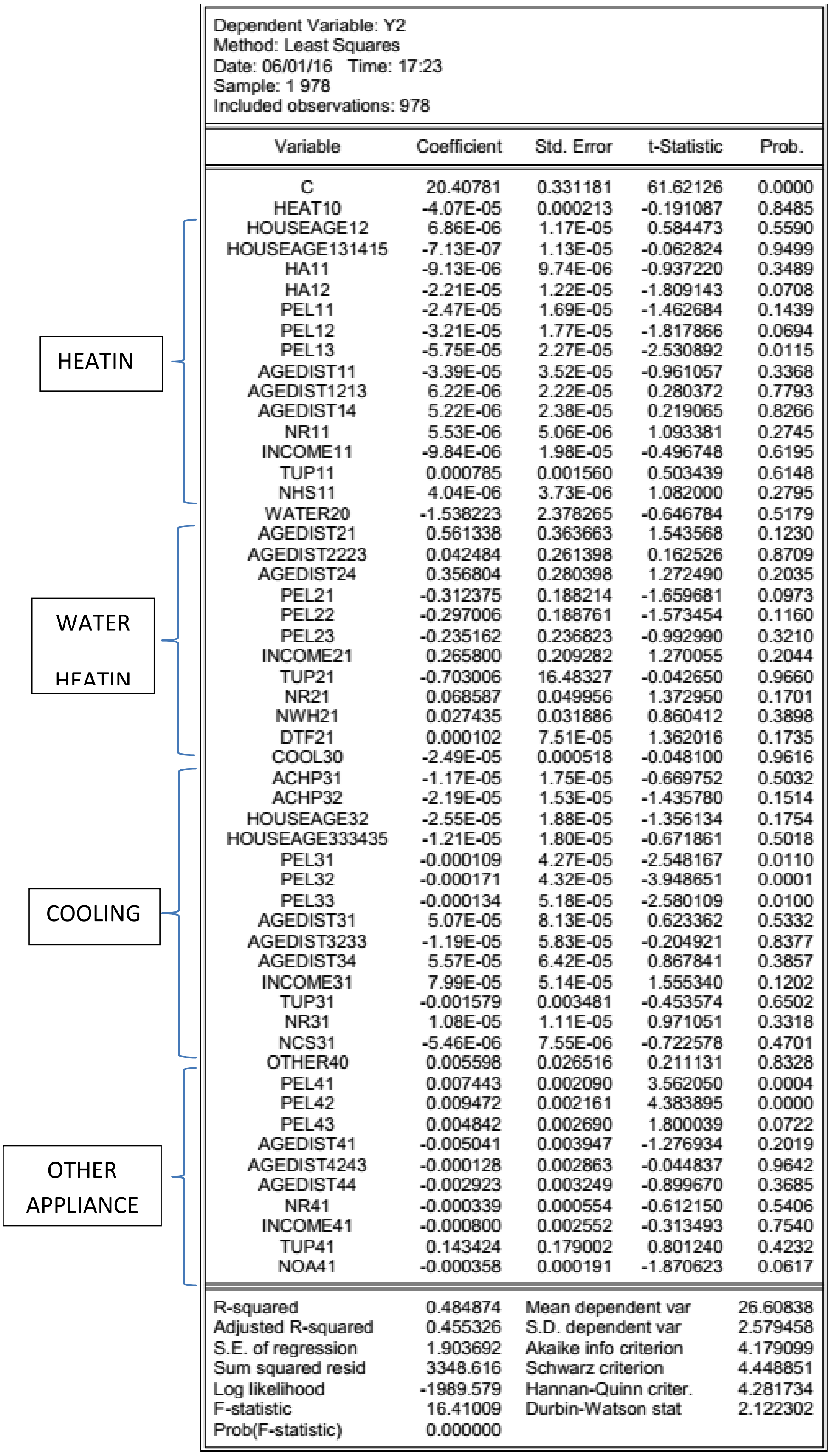


	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.666014	0.063507	10.48729	0.0000
C(2)	-9.60E-06	3.97E-06	-2.418500	0.0158
C(3)	-0.118994	0.057663	-2.063604	0.0393
C(4)	2.23E-06	0.000134	0.016595	0.9868
C(5)	-0.005193	0.000802	-6.474004	0.0000
C(6)	0.002745	0.000207	13.28911	0.0000

Table 6.8 Variance coefficients with combination of Agedist Group 2 & 3 and Houseage Group 3&4&5 with new residuals

Step 2 &3:

Similarly, I use the same methods as in previous sections to apply to this part. I obtain the new GLS result as follows: (Table 6.9)



```
Dependent Variable: Y2
Method: Least Squares
Date: 06/01/16   Time: 17:23
Sample: 1 978
Included observations: 978

        Variable          Coefficient   Std. Error   t-Statistic   Prob.

            C              20.40781     0.331181     61.62126     0.0000
         HEAT10            -4.07E-05    0.000213    -0.191087     0.8485
       HOUSEAGE12           6.86E-06    1.17E-05     0.584473     0.5590
     HOUSEAGE131415        -7.13E-07    1.13E-05    -0.062824     0.9499
          HA11             -9.13E-06    9.74E-06    -0.937220     0.3489
          HA12             -2.21E-05    1.22E-05    -1.809143     0.0708
          PEL11            -2.47E-05    1.69E-05    -1.462684     0.1439
          PEL12            -3.21E-05    1.77E-05    -1.817866     0.0694
          PEL13            -5.75E-05    2.27E-05    -2.530892     0.0115
        AGEDIST11          -3.39E-05    3.52E-05    -0.961057     0.3368
       AGEDIST1213          6.22E-06    2.22E-05     0.280372     0.7793
        AGEDIST14           5.22E-06    2.38E-05     0.219065     0.8266
          NR11              5.53E-06    5.06E-06     1.093381     0.2745
        INCOME11           -9.84E-06    1.98E-05    -0.496748     0.6195
          TUP11             0.000785    0.001560     0.503439     0.6148
          NHS11             4.04E-06    3.73E-06     1.082000     0.2795
         WATER20           -1.538223    2.378265    -0.646784     0.5179
        AGEDIST21           0.561338    0.363663     1.543568     0.1230
       AGEDIST2223          0.042484    0.261398     0.162526     0.8709
        AGEDIST24           0.356804    0.280398     1.272490     0.2035
          PEL21            -0.312375    0.188214    -1.659681     0.0973
          PEL22            -0.297006    0.188761    -1.573454     0.1160
          PEL23            -0.235162    0.236823    -0.992990     0.3210
        INCOME21            0.265800    0.209282     1.270055     0.2044
          TUP21            -0.703006    16.48327    -0.042650     0.9660
          NR21              0.068587    0.049956     1.372950     0.1701
          NWH21             0.027435    0.031886     0.860412     0.3898
          DTF21             0.000102    7.51E-05     1.362016     0.1735
          COOL30           -2.49E-05    0.000518    -0.048100     0.9616
          ACHP31           -1.17E-05    1.75E-05    -0.669752     0.5032
          ACHP32           -2.19E-05    1.53E-05    -1.435780     0.1514
       HOUSEAGE32          -2.55E-05    1.88E-05    -1.356134     0.1754
     HOUSEAGE333435        -1.21E-05    1.80E-05    -0.671861     0.5018
          PEL31            -0.000109    4.27E-05    -2.548167     0.0110
          PEL32            -0.000171    4.32E-05    -3.948651     0.0001
          PEL33            -0.000134    5.18E-05    -2.580109     0.0100
        AGEDIST31           5.07E-05    8.13E-05     0.623362     0.5332
       AGEDIST3233         -1.19E-05    5.83E-05    -0.204921     0.8377
        AGEDIST34           5.57E-05    6.42E-05     0.867841     0.3857
        INCOME31            7.99E-05    5.14E-05     1.555340     0.1202
          TUP31            -0.001579    0.003481    -0.453574     0.6502
          NR31              1.08E-05    1.11E-05     0.971051     0.3318
          NCS31            -5.46E-06    7.55E-06    -0.722578     0.4701
         OTHER40            0.005598    0.026516     0.211131     0.8328
          PEL41             0.007443    0.002090     3.562050     0.0004
          PEL42             0.009472    0.002161     4.383895     0.0000
          PEL43             0.004842    0.002690     1.800039     0.0722
        AGEDIST41          -0.005041    0.003947    -1.276934     0.2019
       AGEDIST4243         -0.000128    0.002863    -0.044837     0.9642
        AGEDIST44          -0.002923    0.003249    -0.899670     0.3685
          NR41             -0.000339    0.000554    -0.612150     0.5406
        INCOME41           -0.000800    0.002552    -0.313493     0.7540
          TUP41             0.143424    0.179002     0.801240     0.4232
          NOA41            -0.000358    0.000191    -1.870623     0.0617

R-squared              0.484874    Mean dependent var      26.60838
Adjusted R-squared     0.455326    S.D. dependent var      2.579458
S.E. of regression     1.903692    Akaike info criterion   4.179099
Sum squared resid      3348.616    Schwarz criterion       4.448851
Log likelihood        -1989.579    Hannan-Quinn criter.    4.281734
F-statistic            16.41009    Durbin-Watson stat      2.122302
Prob(F-statistic)      0.000000
```

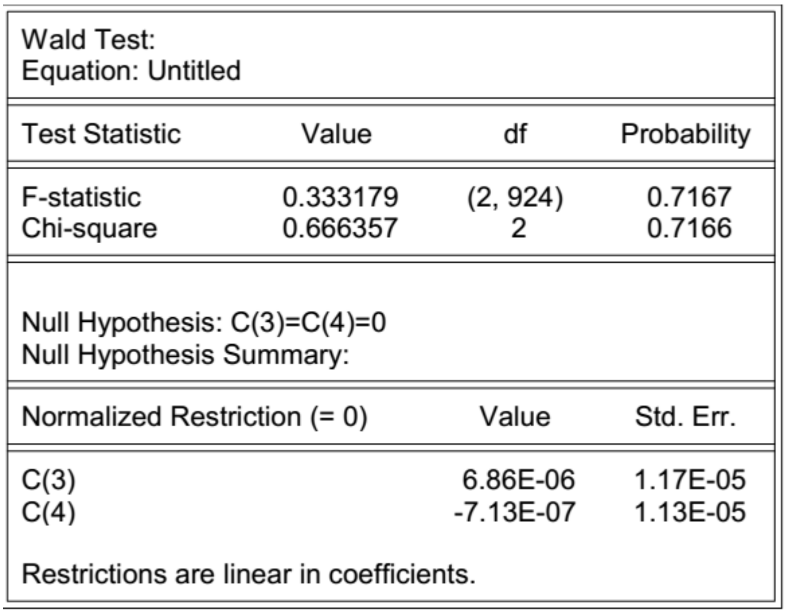
Table 6.9 GLS coefficients with combination of Agedist Group 2 & 3 and Houseage Group 3&4&5 with new residuals

6.2.7 Discussion of results for dropping Houseage 12-15, Agedist Group11-14 and 31—34 and TUPs based on Model in 6.2.6

In 6.2.6, I discussed the results for combining Age distribution group 2&3 and Houseage group 3&4&5 using new estimated residuals. Compared with 6.2.5, model 6.2.6 did not change a lot but improved a little bit significance with respect to demographic factors. Therefore, I decide to use 6.2.6 model as my following model.

In table 6.11, I calculated the p-values of the coefficients of Houseage under the heating system, Age distribution under heating and cooling system and all the time-of-use prices variables are shown much more than 10%, which mean that all of them are relatively unimportant to the energy consumption. In this section, I plan to drop them. But firstly, I conduct a Wald test on these group variables.

Firstly, I test the **Houseage 12-15** and obtain following result: (Table 6.10)



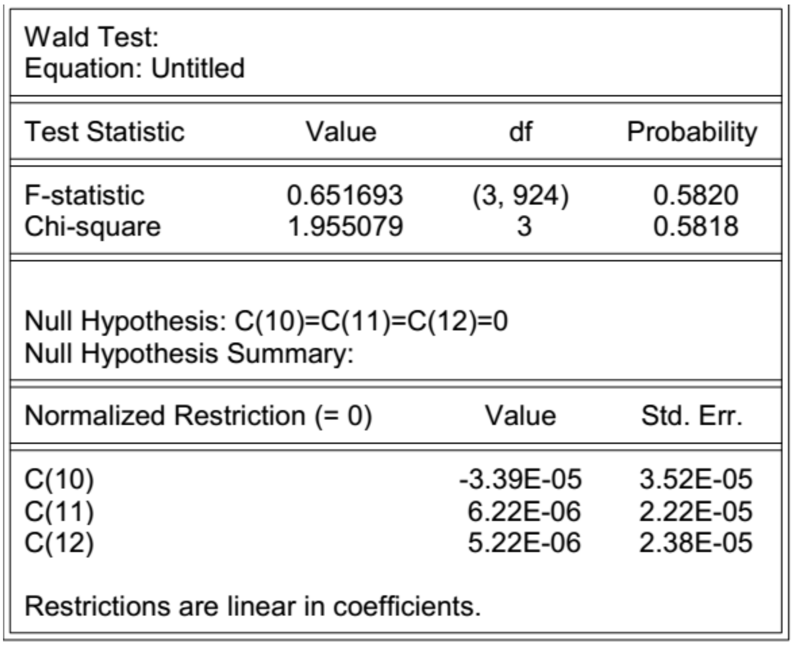
```
Wald Test:
Equation: Untitled

Test Statistic          Value           df        Probability

F-statistic             0.333179     (2, 924)       0.7167
Chi-square              0.666357        2           0.7166


Null Hypothesis: C(3)=C(4)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)          Value       Std. Err.

C(3)                               6.86E-06       1.17E-05
C(4)                              -7.13E-07       1.13E-05

Restrictions are linear in coefficients.
```

Table 6.10 Wald test on Houseage 12-15

With a p-value 0.7166, this means that the group variable of Houseage 12-15 is an insignificant factor for total electricity consumption.

Now, I want to conduct hypothesis testing on Agedist 11-14, Agedist 31-34 and all TUPs respectively: (Table 6.11)

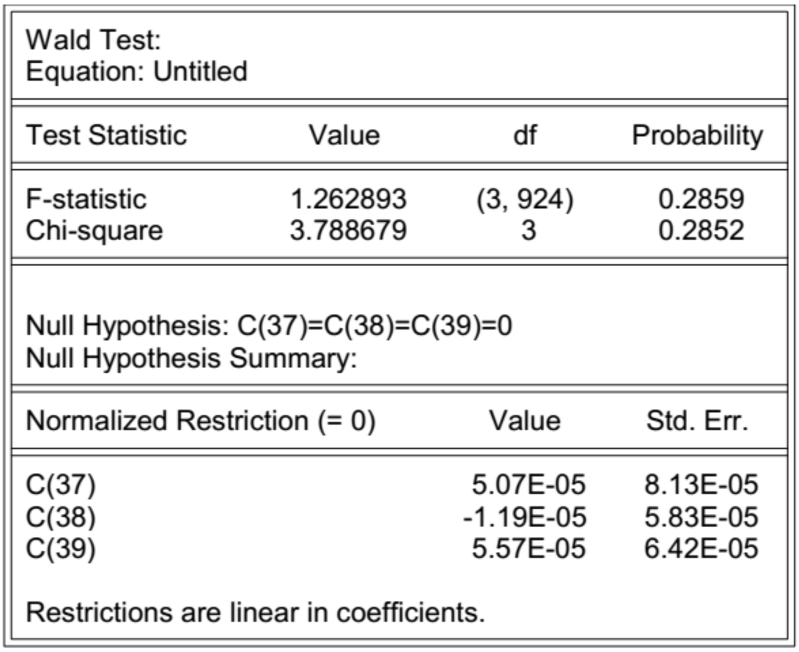
Agedist 11-14



```
Wald Test:
Equation: Untitled

Test Statistic          Value           df         Probability

F-statistic           0.651693       (3, 924)        0.5820
Chi-square            1.955079          3            0.5818

Null Hypothesis: C(10)=C(11)=C(12)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)           Value        Std. Err.

C(10)                                -3.39E-05      3.52E-05
C(11)                                 6.22E-06      2.22E-05
C(12)                                 5.22E-06      2.38E-05

Restrictions are linear in coefficients.
```

Table 6.11 Wald test on Agedist 11-14

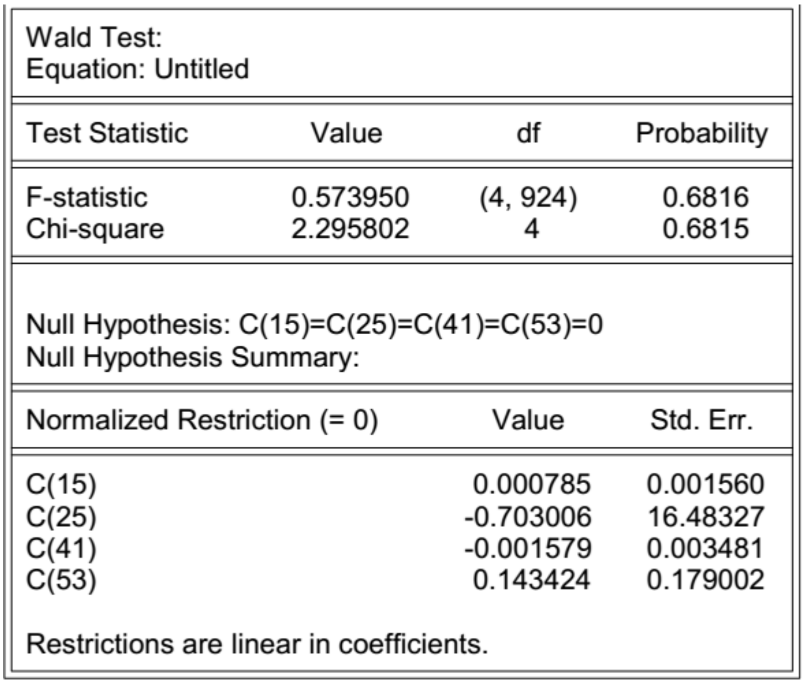
Agedist 31-34 (Table 6.12)



```
Wald Test:
Equation: Untitled

Test Statistic          Value           df         Probability

F-statistic           1.262893       (3, 924)        0.2859
Chi-square            3.788679          3            0.2852

Null Hypothesis: C(37)=C(38)=C(39)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)           Value        Std. Err.

C(37)                                 5.07E-05      8.13E-05
C(38)                                -1.19E-05      5.83E-05
C(39)                                 5.57E-05      6.42E-05

Restrictions are linear in coefficients.
```

Table 6.12 Wald test on Agedist 31-34

TUPs (Table 6.13)



Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	0.573950	(4, 924)	0.6816
Chi-square	2.295802	4	0.6815

Null Hypothesis: C(15)=C(25)=C(41)=C(53)=0
Null Hypothesis Summary:

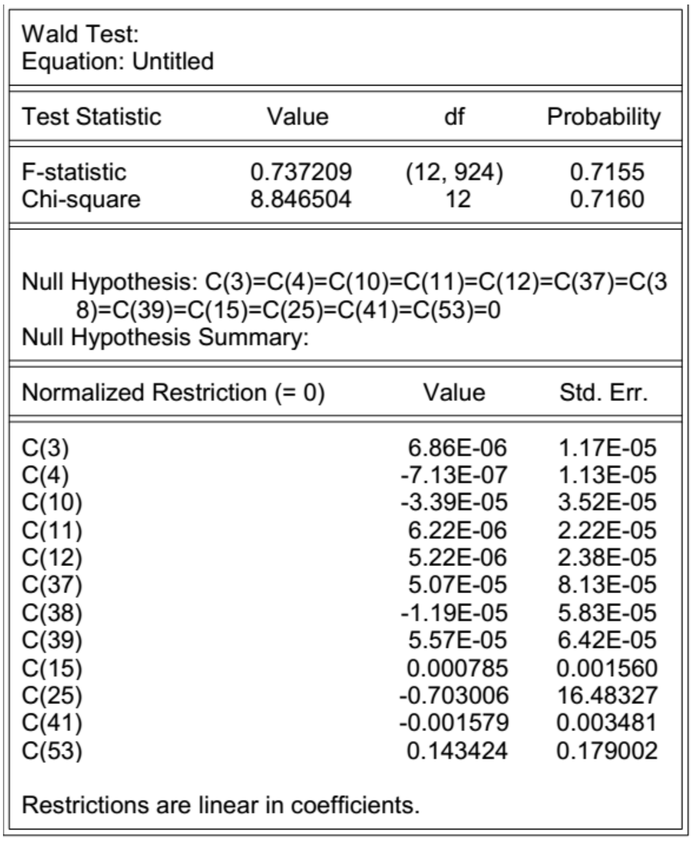
Normalized Restriction (= 0)	Value	Std. Err.
C(15)	0.000785	0.001560
C(25)	-0.703006	16.48327
C(41)	-0.001579	0.003481
C(53)	0.143424	0.179002

Restrictions are linear in coefficients.

Table 6.13 Wald test on TUPs

All of these group variables show that they don't exert a significant influence on electricity consumption. The last but least, I want to test that all of these variables as a group (Houseage 12-15, Agedist 11-14 & 31-34 and TUPs):

Houseage 12-15, Agedist 11-14 & 31-34 and TUPs (Table 6.14)

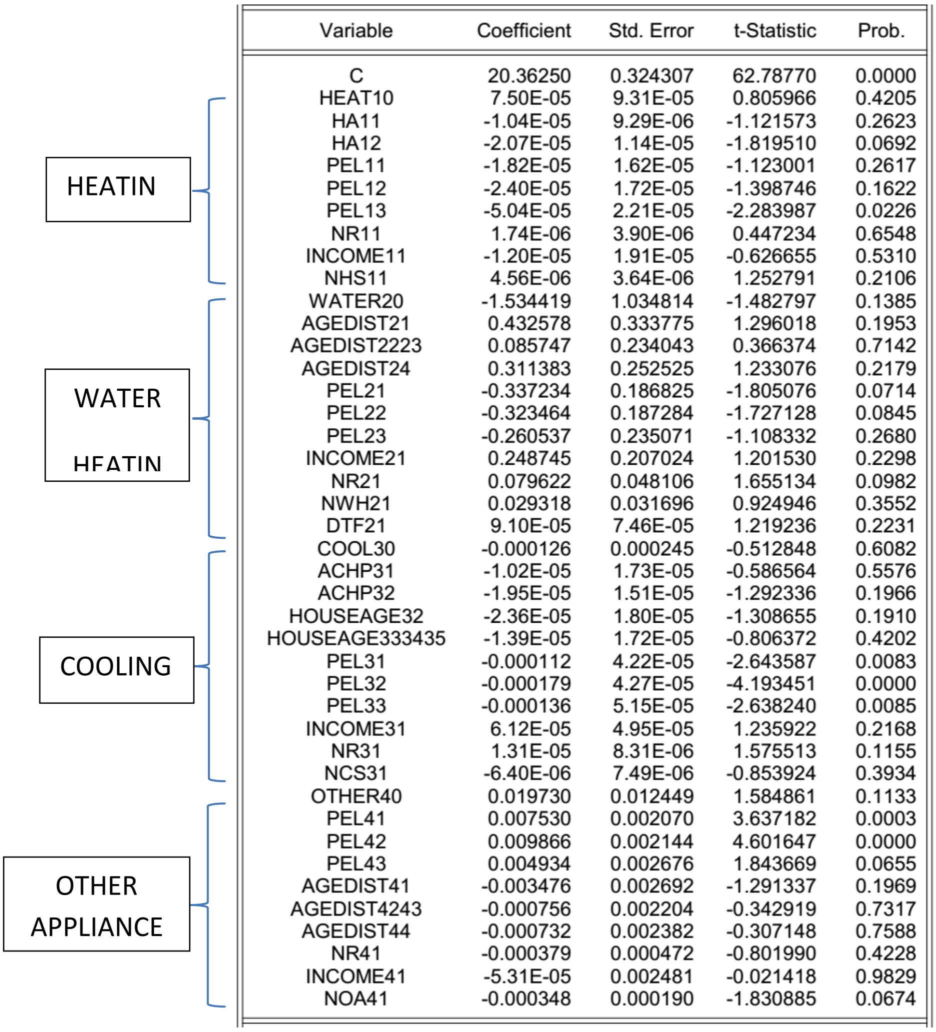


```
Wald Test:
Equation: Untitled

Test Statistic        Value          df        Probability

F-statistic          0.737209      (12, 924)     0.7155
Chi-square           8.846504         12         0.7160


Null Hypothesis: C(3)=C(4)=C(10)=C(11)=C(12)=C(37)=C(3
       8)=C(39)=C(15)=C(25)=C(41)=C(53)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)          Value       Std. Err.

C(3)                                 6.86E-06     1.17E-05
C(4)                                -7.13E-07     1.13E-05
C(10)                               -3.39E-05     3.52E-05
C(11)                                6.22E-06     2.22E-05
C(12)                                5.22E-06     2.38E-05
C(37)                                5.07E-05     8.13E-05
C(38)                               -1.19E-05     5.83E-05
C(39)                                5.57E-05     6.42E-05
C(15)                                0.000785     0.001560
C(25)                               -0.703006     16.48327
C(41)                               -0.001579     0.003481
C(53)                                0.143424     0.179002

Restrictions are linear in coefficients.
```

Table 6.14 Wald test on Houseage 12-15, Agedist 31-34 and TUPs

As a collective group, Houseage under heating system, Age distribution under heating and cooling system and time-of-prices under four systems don’t show an important impact on final electricity consumption.

From the Wald test results above, I therefore drop Houseage 12-15, Agedist 11-14 & 31-34 and TUPs and obtain the new GLS results after fitting as follows: (Table 6.15)



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.36250	0.324307	62.78770	0.0000
HEAT10	7.50E-05	9.31E-05	0.805966	0.4205
HA11	-1.04E-05	9.29E-06	-1.121573	0.2623
HA12	-2.07E-05	1.14E-05	-1.819510	0.0692
PEL11	-1.82E-05	1.62E-05	-1.123001	0.2617
PEL12	-2.40E-05	1.72E-05	-1.398746	0.1622
PEL13	-5.04E-05	2.21E-05	-2.283987	0.0226
NR11	1.74E-06	3.90E-06	0.447234	0.6548
INCOME11	-1.20E-05	1.91E-05	-0.626655	0.5310
NHS11	4.56E-06	3.64E-06	1.252791	0.2106
WATER20	-1.534419	1.034814	-1.482797	0.1385
AGEDIST21	0.432578	0.333775	1.296018	0.1953
AGEDIST2223	0.085747	0.234043	0.366374	0.7142
AGEDIST24	0.311383	0.252525	1.233076	0.2179
PEL21	-0.337234	0.186825	-1.805076	0.0714
PEL22	-0.323464	0.187284	-1.727128	0.0845
PEL23	-0.260537	0.235071	-1.108332	0.2680
INCOME21	0.248745	0.207024	1.201530	0.2298
NR21	0.079622	0.048106	1.655134	0.0982
NWH21	0.029318	0.031696	0.924946	0.3552
DTF21	9.10E-05	7.46E-05	1.219236	0.2231
COOL30	-0.000126	0.000245	-0.512848	0.6082
ACHP31	-1.02E-05	1.73E-05	-0.586564	0.5576
ACHP32	-1.95E-05	1.51E-05	-1.292336	0.1966
HOUSEAGE32	-2.36E-05	1.80E-05	-1.308655	0.1910
HOUSEAGE333435	-1.39E-05	1.72E-05	-0.806372	0.4202
PEL31	-0.000112	4.22E-05	-2.643587	0.0083
PEL32	-0.000179	4.27E-05	-4.193451	0.0000
PEL33	-0.000136	5.15E-05	-2.638240	0.0085
INCOME31	6.12E-05	4.95E-05	1.235922	0.2168
NR31	1.31E-05	8.31E-06	1.575513	0.1155
NCS31	-6.40E-06	7.49E-06	-0.853924	0.3934
OTHER40	0.019730	0.012449	1.584861	0.1133
PEL41	0.007530	0.002070	3.637182	0.0003
PEL42	0.009866	0.002144	4.601647	0.0000
PEL43	0.004934	0.002676	1.843669	0.0655
AGEDIST41	-0.003476	0.002692	-1.291337	0.1969
AGEDIST4243	-0.000756	0.002204	-0.342919	0.7317
AGEDIST44	-0.000732	0.002382	-0.307148	0.7588
NR41	-0.000379	0.000472	-0.801990	0.4228
INCOME41	-5.31E-05	0.002481	-0.021418	0.9829
NOA41	-0.000348	0.000190	-1.830885	0.0674

Table 6.15 Final version of GLS coefficients

6.3 χ^2 test on demographic factors

6.3.1 The final model

In my thesis, I mainly focus on discovering the relationship between total electricity consumption and demographic factors such as Education, Age distribution, Income and Number of Residents. Through the result in 6.2.7, I can reserve that education level under four systems exerts an important role in influencing electricity consumption. Age distribution shows a relatively weak influence on consumption. In addition, number of residents under water heating and cooling system shows an important role in affecting electricity consumption. In this way, I find this model seems appropriate for my aim. This model is the model used for further discussion.

6.3.2 χ^2 test on demographic factors

After determining the final model, I need to investigate the underlying relationship between total electricity consumption and demographic variables. I plan to do χ^2 test on demographic factors by different combination to show the significance degree of these factors by following path. (Figure 6.2) I will conduct Wald test to test the joint significance of a subset of coefficients. These variables are individually insignificant based on t-tests with very high p values. But we should test the joint significance of them using Wald test.

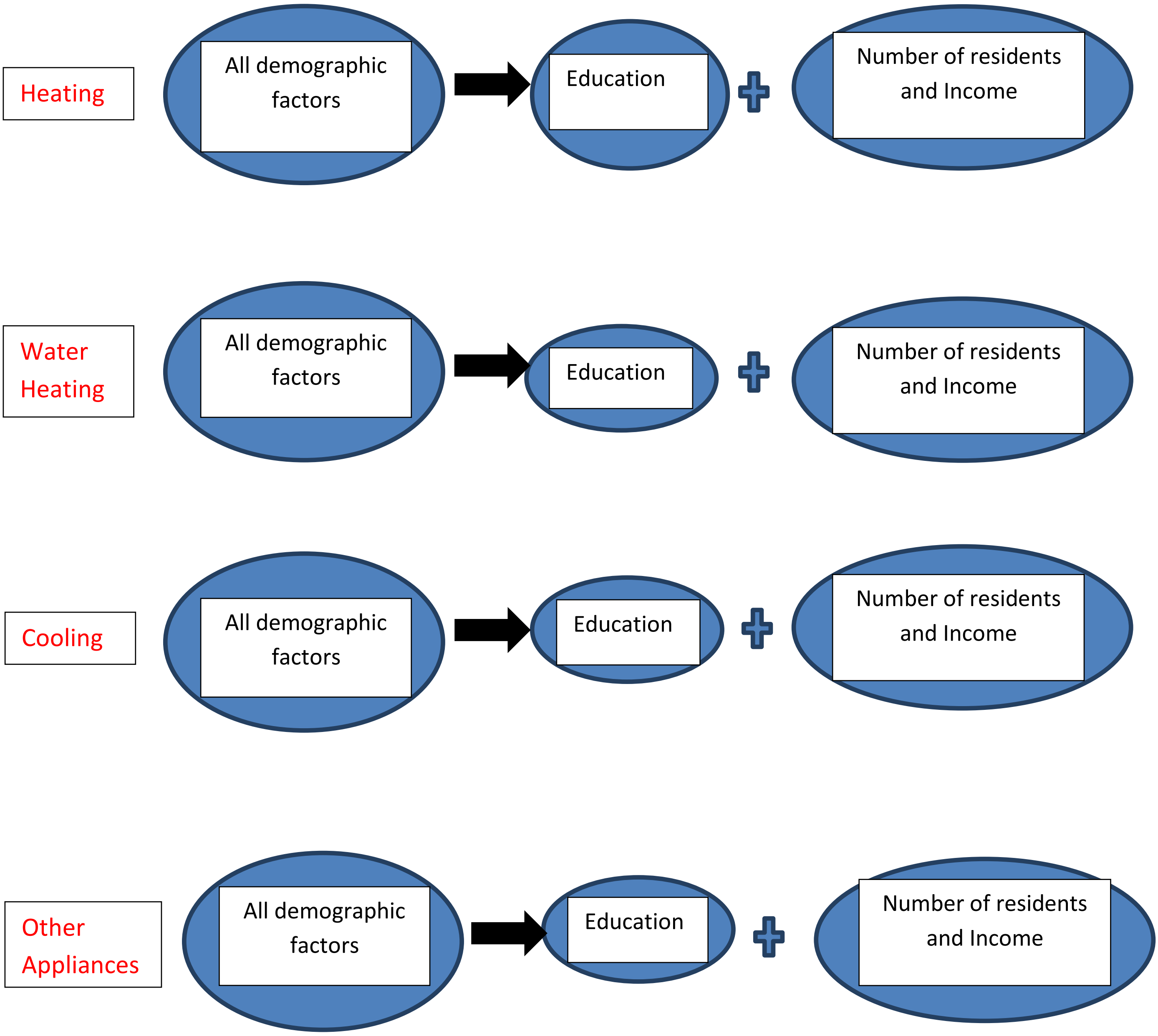


Figure 6.2 Demographic factors by different combination to show the significance degree

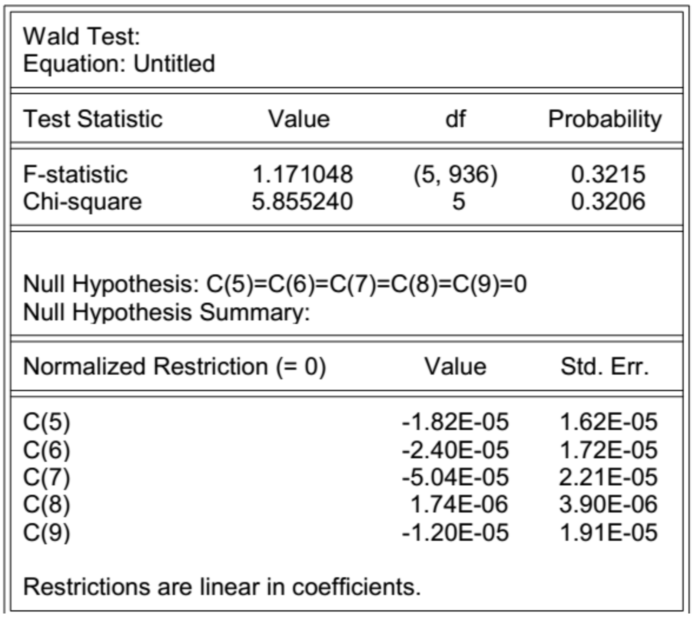
(1) Testing of PEL11-13, NR11 and INCOME11 (Table 6.11)

In this part, I am conducting χ^2 test on factors education, number of residents and income under heating system. Therefore, I go to “View” -> “Coefficients Diagnostics” -> “Wald Test-Coefficient Restrictions”. In this part, I find corresponding number of these factors is 5, 6, 7, 8 and 9 respectively. I set null hypothesis as “C(5)=C(6)=C(7)=C(8)=C(9)=0” as follows:

Null $H_0 : \beta_{PEL11} = \beta_{PEL12} = \beta_{PEL13} = \beta_{NR11} = \beta_{INCOME11} = 0$

Alternative H_A : At least some are non-zero.

I can obtain the following result: (Table 6.16)



```
Wald Test:
Equation: Untitled

Test Statistic        Value          df        Probability

F-statistic          1.171048      (5, 936)      0.3215
Chi-square           5.855240         5          0.3206

Null Hypothesis: C(5)=C(6)=C(7)=C(8)=C(9)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)        Value        Std. Err.

C(5)                              -1.82E-05      1.62E-05
C(6)                              -2.40E-05      1.72E-05
C(7)                              -5.04E-05      2.21E-05
C(8)                               1.74E-06      3.90E-06
C(9)                              -1.20E-05      1.91E-05

Restrictions are linear in coefficients.
```

Table 6.16 Testing of PEL11-13, NR11 and INCOME11

Withχ^2’s p-value is 0.3206, it means that all demographic factors under heating system are not significant for total electricity consumption. Therefore, we cannot reject the null hypothesis.

(2) Testing of PEL11-13

Here, I use the similar method as (1). I replace the input content into "C(5)=C(6)=C(7)=0".

A p-value of 0.1522 represents that the education level under heating system is an insignificant factor to the electricity consumption. We can also not reject null hypothesis.

(3) Testing of NR11 and INCOME11

A p-value of 0.7932 means that age distribution and number of residents under heating system is not a significant factor for households' electricity consumption at 90% significance level.

(4) Testing of PEL21-23, AGEDIST21-24, NR21 and INCOME21

The p-value of χ^2 is 0.0051, this means that demographic factors under water heating system are very significant for total electricity consumption.

(5) Testing of PEL21-23

A p-value of 0.2888 represents that the education level under water heating system is not important to consumption.

(6) Testing of AGEDIST21-24, NR21 and INCOME21

0.0052 means that the age distribution level, number of residents and income under water heating system is significant to households' electricity consumption.

(7) Testing of PEL31-33, NR31 and INCOME31

A p-value of χ^2 test 0.0002 means all demographic factors under cooling system are significant for energy consumption in households.

(8) Testing of PEL31-33

A p-value of 0.0005 means education level under cooling system is very important to total energy consumption.

(9) Testing of NR31 and INCOME31

A p-value of χ^2 test 0.0592 represents that number of residents and age distribution of cooling system in one household is an important factor to its electricity consumption at 90% significance level.

(10) Testing of PEL41-43, AGEDUST 41-44, NR41 and INCOME41

The p-value of χ^2 is 0.0002, which means that the demographic elements are very important to electricity consumption.

(11) Testing of PEL41-43

A p-value of 0.0000 represents that education level under other appliances system is very important to aggregate consumption.

(12) Testing of AGEDIST41-44, NR41 and INCOME41

For age distribution, number of residents and income under other appliances system, p-value of χ^2 is 0.3232. This means that these factors exert an unimportant impact to aggregate consumption.

The following is a summary of the above χ^2 test:

Summary: χ^2 test for group variables (use 10% as the base level)

(Table 6.17)

	Variables	p-value	Significant or not
HEATIN	PEL11-13, NR11 and INCOME11	0.3206	No
	PEL11-13	0.1522	No
	NR11 and INCOME11	0.7932	No
WATER HEATIN	PEL21-23,AGEDIST21-24,NR21 and INCOME21	0.0051	Yes
	PEL21-23	0.2888	No
	AGEDIST21-24, NR21 and INCOME21	0.0052	Yes
COOLING	PEL31-33,NR31 and INCOME31	0.0002	Yes
	PEL31-33	0.0005	Yes
	NR31 and INCOME31	0.0592	Yes
OTHER APPLIANCE	PEl41-43,AGEDIST41-44,NR41 and INCOME41	0.0002	Yes
	PEL41-43	0.0000	Yes
	AGEDIST41-44, NR41 and INCOME41	0.3232	No

Table 6.17 Summary: χ^2 test for group variables

From the results shown above, I obtain the conclusion that there are no obvious important factors under heating system. The education distribution group seems to approach to significance to total electricity consumption. However, number of residents or income does not exert important influence on the final consumption.

In the water heating system, age distribution, number of residents and income are very significant factors to consumption no matter the education level exists or not. As education distribution group itself, it does not show any significance to consumption.

For the cooling system, education, number of residents and income as a group factor, is extremely important to affect electricity usage. Education distribution level itself is also a key group variable to final energy use. Number of residents and income seem relatively weak important group variable to influence household's electricity consumption.

With respect to the other appliances system, education distribution level group variable is an extremely significant factor for electricity consumption. Age distribution, number of residents, income and education level as a group variable is also very significant to households' electricity consumption. Meanwhile, age distribution, number of residents and income group variable seems not that important to dependent variable total electricity consumption.

6.4 Illustrative comparisons regarding demographic factors

In this section, I will compare some demographic group factors by using the final model. I will list the comparison between different group factors and include corresponding conclusions. For the next step, the results will be compared with the conclusions obtained in 6.3. Recall for Table 6.10 in 6.2.7, this is my final model used for discussion in this section.

Here, I will introduce the definition of illustrative groups where I discuss consumption impacts due to variations in the education distribution, age distribution and number of residents: (Table 6.18)

Demographic factor	Number of group	Description
Education distribution	Group 1	2 adults under 65 with high school
	Group 2	2 adults under 65 with college
	Group 3	2 adults under 65 with university
	Group 4	2 adults under 65 with post graduate degree
Age distribution	Group 1	1 adult under 65 & 1 child
	Group 2	2 adults under 65
	Group 3	2 adults over 75
Number of residents	Group 1	1 person
	Group 2	2 people

Table 6.18 Description of scenarios involving variations in the education distribution, age distribution and number of residents

6.4.1 Comparison between different education groups under an electricity heating system

For those households with an electricity heating system, I assume other variables to be the same except for education level distribution. Recall for equation 5.1:

$$
\begin{aligned}
\mathrm{Iny} = {} & \alpha_0 + (\alpha_{10} + \mathrm{AGE1} + \mathrm{HA1} + \mathrm{PEL1} + \mathrm{AGEDIST1} + \alpha_{15} * \mathrm{NR} + \alpha_{16} * \log(\mathrm{INC}) \\
& + \alpha_{17} * \mathrm{TUP} + \alpha_{18} * \mathrm{NHS}) * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} \\
& + (\alpha_{20} + \mathrm{AGEDIST2} + \mathrm{PEL2} + \alpha_{23} * \log(\mathrm{INC}) + \alpha_{24} * \mathrm{TUP} + \alpha_{25} \\
& * \mathrm{NR} + \alpha_{26} * \mathrm{NWH} + \alpha_{27} * \mathrm{DTF}) * \mathrm{D}_2 \\
& + (\alpha_{30} + \mathrm{ACHP3} + \mathrm{AGE3} + \mathrm{PEL3} + \mathrm{AGEDIST3} + \alpha_{36} * \log(\mathrm{INC}) \\
& + \alpha_{37} * \mathrm{TUP} + \alpha_{38} * \mathrm{NR} + \alpha_{39} * \mathrm{NCS}) * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} \\
& + (\alpha_{40} + \mathrm{PEL4} + \mathrm{AGEDIST4} + \alpha_{43} * \mathrm{NR} + \alpha_{44} * \log(\mathrm{INC}) + \alpha_{45} \\
& * \mathrm{TUP} + \alpha_{46} * \mathrm{NOA}) * \sqrt{D_4}
\end{aligned}
$$

To transfer from qualitative to quantitative impacts, the following table is shown:

Variables	2 adults under 65 with high school (group1)	2 adults under 65 with college (group2)	2 adults under 65 with university (group3)	2 adults under 65 with post graduate degree (group 4)
D_1	1	1	1	1
PEL11	1	0	0	0
PEL12	0	1	0	0
PEL13	0	0	1	0
NR11	2	2	2	2

In comparing groups above, I assume that other variables are exactly the same. I am calculating the difference in logarithm of total electricity consumption between these two types of households.

$$
\begin{aligned}
\Delta\mathrm{Iny} = \mathrm{Iny}_2 - \mathrm{Iny}_1 = {} & [\alpha_1 * (PEL11_2 - PEL11_1) + \\
& \alpha_2 * (PEL12_2 - PEL12_1) + \\
& \alpha_3 * (PEL13_2 - PEL13_1) + \\
& \alpha_4 * (NR11_2 - NR11_1)] * \mathrm{D}_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}}
\end{aligned}
\tag{6.1}
$$

where α_1, α_2, α_3 and α_4 represent the coefficients of $PEL11$, $PEL12$, $PEL13$ and $NR11$ in Table 6.10.

I have assumed that D_1 equal to 1 for both groups. Based on my final model, 160 households are with $D_1 = 1$. The mean of $\sqrt{\mathrm{HDH}}$ and $\sqrt{\mathrm{SIZE}}$ are 330.1441249 and 44.73962772 respectively.

(1) group 3 relative to group 1

Note that in that comparison between group 3 and group 1—the only variables that change are PEL11 and PEL13. If I conduct a joint test on PEL11 and PEL13, I obtain a p-value of χ^2 as 0.0736. This means that the education level involving these two coefficients in the heating system is an important factor on electricity consumption.

Replacing corresponding coefficients in (6.1):

$$\Delta \mathrm{Iny} = [(-1.82 * 10\text{^}(-5)) * (0-1) + (-5.04 * 10\text{^}(-5)) * (1-0)] * 1 * 330.1441249 * 44.73962772$$

$$= -0.475610913$$

The standard error is:

$$\sigma(\Delta Iny) = \sqrt{Var((-\mathrm{c(PEL11)} + \mathrm{c(PEL13)}) * 330.1441249 * 44.73962772)}$$

$$= 330.1441249 * 44.73962772 * \sqrt{\left(\sigma_{c(\widehat{PEL11})}\right)^2 + \left(\sigma_{c(\widehat{PEL13})}\right)^2 - 2cov(c(PEL11), c(PEL13))}$$

$$= 330.1441249 * 44.73962772 * \sqrt{(1.62 * 10\text{^}(-5))^2 + (2.21 * 10\text{^}(-5))^2 - 2cov(c(PEL11), c(PEL13))}$$

$$= 330.1441249 * 44.73962772 * \sqrt{(1.62 * 10\text{^}(-5))^2 + (2.21 * 10\text{^}(-5))^2 - 2 * 1.79 * 10\text{^}(-10)}$$

$$= 0.292758368$$

When I take the exponential of $\Delta\mathrm{Iny}$, I obtain:

$$e^{\Delta \mathrm{Iny}} = e^{\mathrm{Iny}_2 - \mathrm{Iny}_1} = e^{\Delta \mathrm{Iny}} = e^{\mathrm{Iny}_2 - \mathrm{Iny}_1} = e^{In\frac{y_2}{y_1}} = \frac{y_2}{y_1}$$

$$\frac{y_2}{y_1} - 1 = e^{\Delta \mathrm{Iny}} - 1 = e^{-0.475610913} - 1 = -0.378494745$$

From the result above, I obtain the conclusion that group 3 reduces electricity consumption by 37.8% compared with group 1. That is to say, adults under 65 with university education level will consume less electricity compared with adults under 65 with high school education level.

I conduct the following hypothesis test:

$$H_0\colon\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta \mathrm{Iny}} - 1 \neq 0$$

Where $\Delta \mathrm{Iny} = [\mathrm{c}(5) * (0-1) + \mathrm{c}(7) * (1-0)] * 330.1441249 * 44.73962772.$

And c(5) and c(7) represent coefficient of PEL11 and PEL13.

Correspondingly, I conduct a Wald test (See Appendix 24) on this hypothesis test and obtain a p-value of χ^2 as 0.0371. I reject the null hypothesis $e^{\Delta \mathrm{Iny}} - 1 = 0$ at a 10% significance level. These two groups are statistically different from each other when other variables are the same. This means that education level in comparing group 3 with group 1 exerts a significant impact on electricity consumption coming from the heating system.

(2) group 4 relative to group 1

I find that the p-value PEL11 is 0.2617, this means that this variable is not that significant to total electricity consumption.

$$\Delta \mathrm{Iny} = [(-1.82 * 10\text{^}(-5)) * (0-1)]*1*330.1441249*44.73962772$$

$$= 0.268823559$$

The standard error is:

$$\sigma(\Delta Iny) = \sqrt{Var((-\mathrm{c(PEL11)}) * 330.1441249 * 44.73962772)}$$

$$= 330.1441249*44.73962772*\sqrt{\left(\sigma_{c(\widehat{PEL11})}\right)^2}$$

$$= 330.1441249*44.73962772*\sqrt{(1.62 * 10\text{^}(-5))^2}$$

$$= 0.239282509$$

$$e^{\Delta \mathrm{Iny}} - 1 = e^{0.268823559} - 1 = 0.308424262$$

I obtain the conclusion that group 4 consumes more electricity consumption by 30.8% compared with group 1. That is to say, adults under 65 with post graduate degree will consume more electricity compared with adults under 65 with high school education level.

I conduct the following hypothesis test:

$$H_0\colon\ e^{\Delta\mathrm{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta\mathrm{Iny}} - 1\ \neq 0$$

Where $\Delta\mathrm{Iny} = [\mathrm{c}(5) * (0-1)] * 330.1441249 * 44.73962772$.

And c(5) represent coefficient of PEL11.

Correspondingly, I conduct a Wald test on this hypothesis test and obtain a p-value of χ^2 as 0.3248. I cannot reject the null hypothesis $e^{\Delta\mathrm{Iny}} - 1 = 0$ at a 10% significance level. These two groups are not statistically different from each other when other variables are the same. This means that education level in comparing group 4 and group 1 doesn't exert a significant impact on electricity consumption coming from the heating system.

(3) group 2 relative to group 1

Similarly, I can compare group 2 and group 1 and obtain:

$\Delta\mathrm{Iny} = [(-1.82 * 10^\wedge(-5)) * (0-1) + (-2.40 * 10^\wedge(-5)) * (1-0)]$*1*330.1441249*44.73962772

= -0.085669046

The standard error is:

$\sigma(\Delta Iny)$= $\sqrt{Var((-\mathrm{c(PEL11)} + \mathrm{c(PEL12)}) * 330.1441249 * 44.73962772)}$

=330.1441249*44.73962772*

$$\sqrt{\left(\sigma_{c(\widehat{PEL11})}\right)^2 + \left(\sigma_{c(\widehat{PEL12})}\right)^2 - 2cov(c(PEL11), c(PEL12))}$$

=330.1441249*44.73962772*

$$\sqrt{(1.62 * 10^\wedge(-5))^2 + (1.72 * 10^\wedge(-5))^2 - 2cov(c(PEL11), c(PEL12))}$$

= 330.1441249*44.73962772*

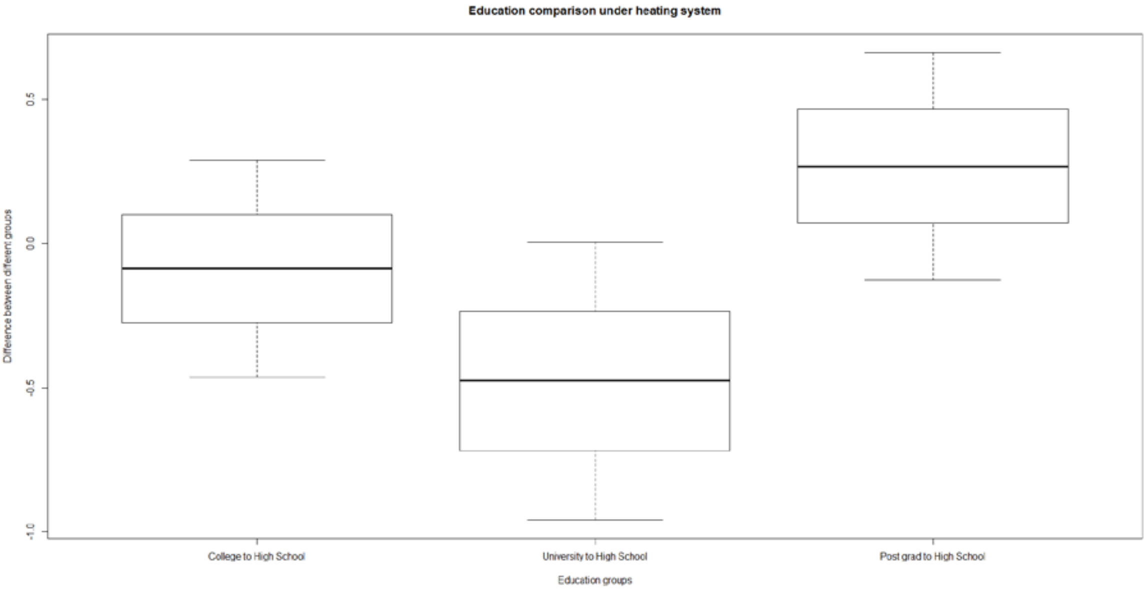
$$\sqrt{(1.62 * 10^\wedge(-5))^2 + (1.72 * 10^\wedge(-5))^2 - 2 * 1.60 * 10^\wedge(-10)}$$

= 0.228002566

From three comparisons between different groups above, I can conclude the following table:

	College relative to **high school**	**University** relative to **high school**	**Post graduate** relative to **high school**
$\Delta \mathrm{Iny}$	-0.085669046	-0.475610913	0.268823559
$\sigma\ (\Delta Iny)$	0.228002566	0.292758368	0.239282509
$\Delta \mathrm{Iny}+1.65*\ \sigma\ (\Delta Iny)$	0.290535188	0.007440395	0.663639699
$\Delta \mathrm{Iny}-1.65*\ \sigma\ (\Delta Iny)$	-0.46187328	-0.95866222	-0.12599258

By R, I can build a boxplot on these comparisons:



For example, the first boxplot represents the comparison between college and high school and includes five boundaries. Starting from the lowest boundary $\Delta \mathrm{Iny}-1.65*\ \sigma\ (\Delta Iny)$ to the highest one represents $\Delta \mathrm{Iny}+1.65*\ \sigma\ (\Delta Iny)$, the second lowest boundary is $\Delta \mathrm{Iny}-1/2*1.65*\ \sigma\ (\Delta Iny)$, the middle one is $\Delta \mathrm{Iny}$ and second largest represents $\Delta \mathrm{Iny}+1/2*1.65*\ \sigma\ (\Delta Iny)$.

I use the high school as the reference group and find that post graduate shows a positive difference relative to high school on residential electricity consumption. In addition, College reduces the consumption a little and University reduces more consumption relative to high school under the heating system.

6.4.2 Comparison between different education groups under an electricity water heating system

Again, the three types of households have the following characteristics:

Variables	2 adults under 65 with high school (group1)	2 adults under 65 with college (group2)	2 adults under 65 with university (group3)	2 adults under 65 with post graduate degree (group 4)
D_2	1	1	1	1
PEL21	1	0	0	0
PEL22	0	1	0	0
PEL23	0	0	1	0
NR21	2	2	2	2

$$\begin{aligned}\Delta \text{Iny} = \text{Iny}_2 - \text{Iny}_1 = [&\gamma_1 * (PEL21_2 - PEL21_1) + \\ &\gamma_2 * (PEL22_2 - PEL22_1) + \\ &\gamma_3 * (PEL23_2 - PEL23_1) + \\ &\gamma_4 * (NR21_2 - NR21_1)] * 1\end{aligned} \tag{6.2}$$

where γ_1, γ_2, γ_3 and γ_4 represent the coefficients of $PEL21$, $PEL22$, $PEL23$ and $NR21$ in Table 6.10.

(1) group 3 relative to group 1

By comparing households with high school and university education, I conduct a hypothesis testing on PEL21 and PEL23 and obtain a p-value of 0.1947. This means that the education factor involving these two coefficients doesn’t play an important role in the electricity towards water heating system.

Replacing corresponding coefficients into (6.2):

$\Delta \text{Iny} = [(-0.337234) * (0 - 1) + (-0.260537) * (1-0)] * 1$

$= 0.076697$

The standard error is:

$\sigma(\Delta Iny)$= $\sqrt{Var((-\mathrm{c(PEL21)}+\mathrm{c(PEL23)})}$

$$=\sqrt{\left(\sigma_{c(\widehat{PEL21})}\right)^2+\left(\sigma_{c(\widehat{PEL23})}\right)^2-2cov(c(PEL21),c(PEL23))}$$

$$=\sqrt{(0.186825)^2+(0.235071)^2-2cov(c(PEL21),c(PEL23))}$$

$$= \sqrt{(0.186825)^2+(0.235071)^2-2*0.024535}$$

= 0.202711508

$$e^{\Delta \mathrm{Iny}}-1 = 0.079714873$$

For education part, people with high education consume more than people with low education by only 8%.

$$H_0:\ e^{\Delta \mathrm{Iny}}-1=0$$

$$H_a:\ e^{\Delta \mathrm{Iny}}-1\ \neq 0$$

Conducting this hypothesis test, the corresponding p-value is 0.7157. I do not reject null hypothesis that $e^{\Delta \mathrm{Iny}}-1=0$ at a 10% significance level. This means that these two groups are not statistically different from each other.

(2) group 4 relative to group 1

I find that the p-value PEL21 is 0.0714, this means that this variable is

significant to total electricity consumption.

$\Delta \mathrm{Iny}=[(-0.337234)*(0-1)]*1$

= 0.337234

The standard error is:

$\sigma(\Delta Iny)$= $\sqrt{Var(\mathrm{c(PEL21)})}$

$$=\sqrt{\left(\sigma_{c(\widehat{PEL21})}\right)^2}$$

$$=\sqrt{(0.186825)^2}$$

= 0.186825

$$e^{\Delta \mathrm{Iny}}-1=e^{0.337234}-1=0.401066875$$

I obtain the conclusion that group 4 consumes more electricity consumption by 40.1% compared with group 1. That is to say, adults under 65 with post graduate degree will consume more electricity compared with adults under 65 with high school education level.

I conduct the following hypothesis test:

$$H_0\colon\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta \mathrm{Iny}} - 1 \neq 0$$

Where $\Delta \mathrm{Iny} = [\mathrm{c}(15) * (0-1)] * 1$.

And c(15) represent coefficient of PEL21.

Correspondingly, I conduct a Wald test on this hypothesis test and obtain a p-value of χ^2 as 0.1258. This means that these two groups are not statistically different from each other.

(3) group 2 relative to group 1

Similarly, I can compare group 2 and group 1 and obtain:

$$\Delta \mathrm{Iny} = [(-0.337234) * (0-1) + (-0.323464) * (1-0)]*1$$

$$= 0.01377$$

The standard error is:

$$\sigma(\Delta Iny) = \sqrt{Var((-\mathrm{c(PEL21)} + \mathrm{c(PEL22)})}$$

$$= \sqrt{\left(\sigma_{c(\widehat{PEL21})}\right)^2 + \left(\sigma_{c(\widehat{PEL22})}\right)^2 - 2cov(c(PEL21), c(PEL22))}$$

$$= \sqrt{(0.186825)^2 + (0.187284)^2 - 2cov(c(PEL21), c(PEL22))}$$

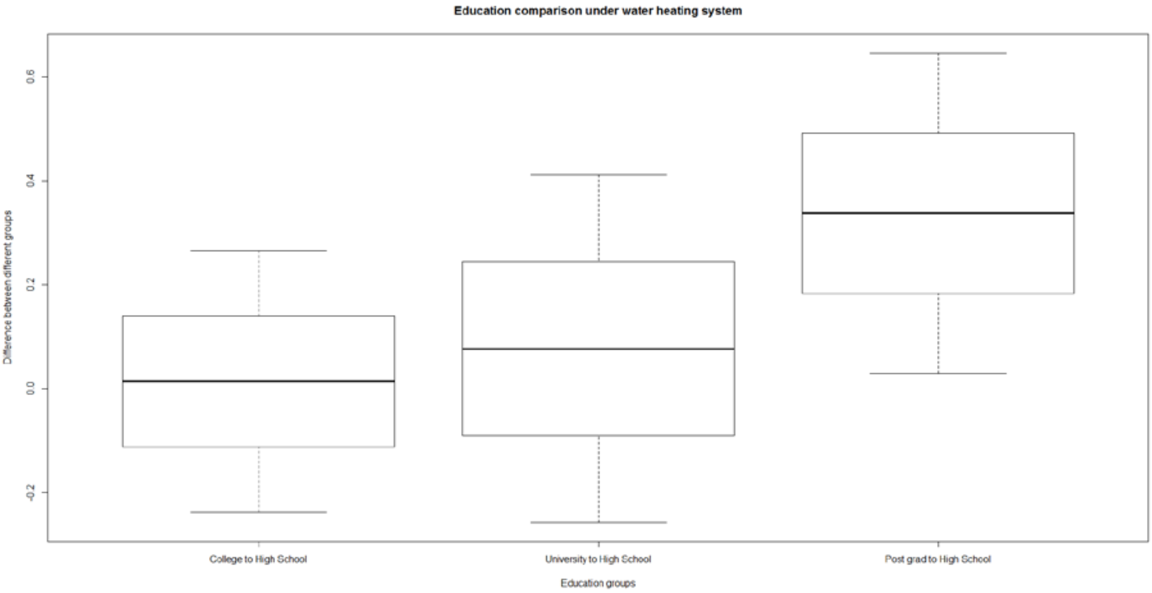
$$= \sqrt{(0.186825)^2 + (0.187284)^2 - 2 * 0.023323}$$

$$= 0.15275103$$

From three comparisons between different groups above, I can conclude the following table:

	College relative to **high school**	**University** relative to **high school**	**Post graduate** relative to **high school**
$\Delta \mathrm{Iny}$	0.01377	0.076697	0.337234
$\sigma\ (\Delta Iny)$	0.15275103	0.202711508	0.186825
$\Delta \mathrm{Iny}+1.65*\ \sigma\ (\Delta Iny)$	0.2658092	0.411170989	0.64549525
$\Delta \mathrm{Iny}-1.65*\ \sigma\ (\Delta Iny)$	-0.2382692	-0.257776989	0.02897275

By R, I can build a boxplot on these comparisons:



As the plot shows, I can conclude that the differences of these three comparisons are positive. The group post graduate relative to high school has the biggest difference and the college relative to high school has the least increase under the water heating system.

6.4.3 Comparison between different education groups under an electricity cooling system

Similar to heating system, I consider the same three groups except for cooling system:

Variables	2 adults under 65 with high school (group1)	2 adults under 65 with college (group2)	2 adults under 65 with university (group3)	2 adults under 65 with post graduate degree (group 4)
D_3	1	1	1	1
PEL31	1	0	0	0
PEL32	0	1	0	0
PEL33	0	0	1	0
NR31	2	2	2	2

Similarly,

$$\Delta \mathrm{Iny} = \mathrm{Iny}_2 - \mathrm{Iny}_1 = [\delta_1 * (PEL31_2 - PEL31_1) + \delta_2 * (PEL32_2 - PEL32_1) + \delta_3 * (PEL33_2 - PEL33_1) + \delta_4 * (NR31_2 - NR31_1)] * \mathrm{D}_3 * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} \quad (6.3)$$

where δ_1, δ_2, δ_3 and δ_4 represent the coefficients of $PEL31$, $PEL32$, $PEL33$ and $NR31$ in Table 6.10.

Assuming $\mathrm{D}_3 = 1$, the mean of $\sqrt{\mathrm{CDH}}$ and $\sqrt{\mathrm{SIZE}}$ are equal to 98.8927854 and 46.12389214 respectively.

(1) group 3 relative to group 1

Note that in that comparison between group 3 and group 1—the only variables that change are PEL31 and PEL33. If I conduct a joint test on PEL31 and PEL33, I obtain a p-value of χ^2 as 0.0228. This means that education level involving these two coefficients in the cooling system is an important factor on electricity consumption.

Replacing corresponding coefficients into (6.3):

$$\Delta \mathrm{Iny} = [(-0.000112) * (0 - 1) + (-0.000136) * (1 - 0)] * 1 * 98.8928854 * 46.12389214$$

= -0.109471795

The standard error is:

$\sigma(\Delta Iny)$= $\sqrt{Var((-\mathrm{c(PEL31)} + \mathrm{c(PEL33)})} * 1 * 98.8928854 * 46.12389214$

=$\sqrt{\left(\sigma_{c(\widehat{PEL}31)}\right)^2 + \left(\sigma_{c(\widehat{PEL}33)}\right)^2 - 2cov(c(PEL31), c(PEL33))} * 1 *$ $98.8928854 * 46.12389214$

=

$\sqrt{(4.22 * 10\hat{}(-5))^2 + (5.15 * 10\hat{}(-5))^2 - 2cov(c(PEL31), c(PEL33))} * 1 *$ $98.8928854 * 46.12389214$

= $\sqrt{(4.22 * 10\hat{}(-5))^2 + (5.15 * 10\hat{}(-5))^2 - 2 * 1.20 * 10\hat{}(-9)} * 1 *$ $98.8928854 * 46.12389214$

= 0.205669219

$$e^{\Delta \mathrm{Iny}} - 1 = \text{-0.103692456}$$

It can be easily obtained that households with higher education consume less electricity on cooling by 10.4%. It is possible that people with higher education have the awareness how to select energy-efficient air-conditioners to save costs on cooling during summer or are mere conscientious about conservation.

Similarly,

$$H_0\text{: } e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\text{: } e^{\Delta \mathrm{Iny}} - 1 \neq 0$$

The p-value of χ^2 is 0.0838 (<10%). I reject the null hypothesis and conclude that these two groups are statistically different from each other.

(2) group 4 relative to group 1

I find that the p-value PEL31 is 0.0083, this means that this variable is

significant with respect to electricity consumption.

$\Delta \mathrm{Iny} = [(-0.000112) * (0 - 1)] * 1 * 98.8928854 * 46.12389214$

= 0.510868375

The standard error is:

$\sigma(\Delta Iny)$= $\sqrt{Var(\mathrm{c(PEL31)})} * 98.8928854 * 46.12389214$

=$\sqrt{\left(\sigma_{c(\widehat{PEL31})}\right)^2}$*$98.8928854 * 46.12389214$

=$\sqrt{(4.22 * 10^\wedge(-5))^2} * 98.8928854 * 46.12389214$

= 0.192487906

$$e^{\Delta \mathrm{Iny}} - 1 = e^{0.510867859} - 1 = 0.66673706$$

I obtain the conclusion that group 4 consumes more electricity consumption by 66.7% compared with group 1. That is to say, adults under 65 with post graduate degree will consume more electricity compared with adults under 65 with high school education level.

I conduct the following hypothesis test:

$$H_0\colon\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta \mathrm{Iny}} - 1\ \neq 0$$

Where $\Delta \mathrm{Iny} = [\mathrm{c}(27) * (0 - 1)] * 1 * 98.8928854 * 46.12389214.$

And c(27) represent coefficient of PEL31.

Correspondingly, I conduct a Wald test on this hypothesis test and obtain a p-value of χ^2 as 0.0492. This means that these two groups are statistically different from each other.

(3) group 2 relative to group 1

Similarly, I can compare group 2 and group 1 and obtain:

$\Delta \mathrm{Iny} = [(-0.000112) * (0 - 1) + (-0.000179) * (1 - 0)]$*1*$98.8928854 * 46.12389214$

= -0.30560876

The standard error is:

$\sigma(\Delta Iny)$= $\sqrt{Var((-\mathrm{c(PEL31)} + \mathrm{c(PEL32)})}$ *1*$98.8928854 * 46.12389214$

=$\sqrt{\left(\sigma_{c(\widehat{PEL31})}\right)^2 + \left(\sigma_{c(\widehat{PEL32})}\right)^2 - 2cov(c(PEL31), c(PEL32))}$*1*$98.8928854 * 46.12389214$

=$\sqrt{(4.22 * 10\hat{}(-5))^2 + (4.27 * 10\hat{}(-5))^2 - 2cov(c(PEL31), c(PEL32))}$*1*$98.8928854 * 46.12389214$

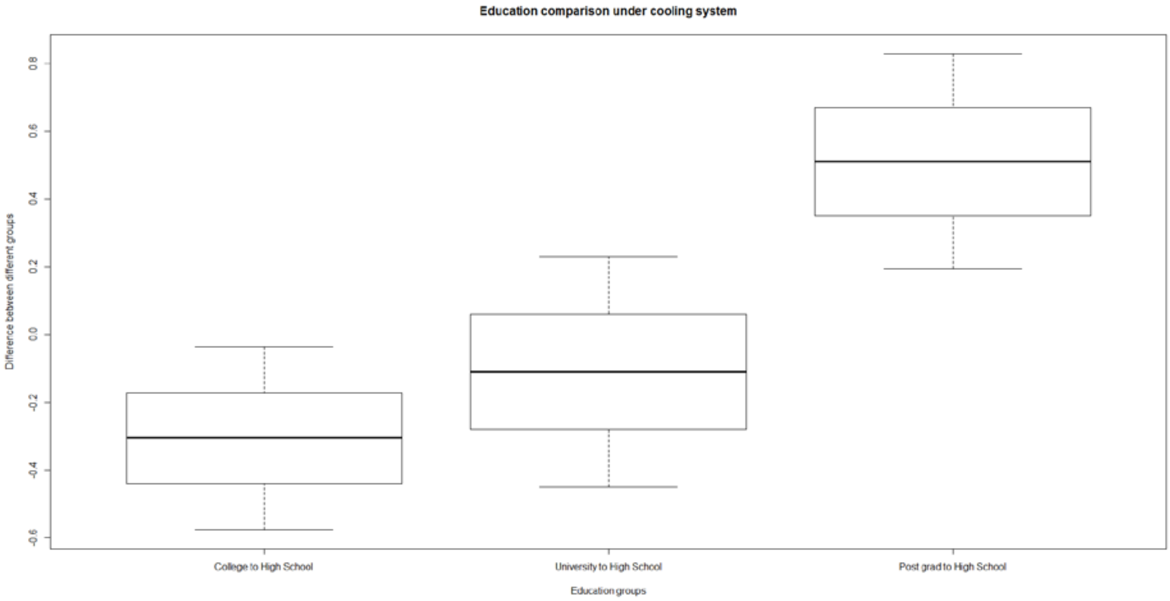
=
$\sqrt{(4.22 * 10\hat{}(-5))^2 + (4.27 * 10\hat{}(-5))^2 - 2 * 1.16 * 10\hat{}(-9)}$*1*$98.8928854 * 46.12389214$

= 0.163453977

From three comparisons between different groups above, I can conclude the following table:

	College relative to **high school**	**University** relative to **high school**	**Post graduate** relative to **high school**
ΔIny	-0.30560876	-0.109471795	0.510868375
σ (ΔIny)	0.163453977	0.205669219	0.192487906
ΔIny+1.65* σ (ΔIny)	-0.035909698	0.229882416	0.82847342
ΔIny-1.65* σ (ΔIny)	-0.575307823	-0.448826006	0.193263331

By R, I can build a boxplot on these comparisons:



From the plot above, I find that post graduate relative to high school shows significant difference from each other. University decreases a little and college reduces more consumption compared to high school under the cooling system.

6.4.4 Comparison between different education groups under other appliances system

The characteristics for these three groups are:

Variables	2 adults under 65 with high school (group1)	2 adults under 65 with college (group2)	2 adults under 65 with university (group3)	2 adults under 65 with post graduate degree (group 4)
$\sqrt{D_4}$	88.47725526	88.47725526	88.47725526	88.47725526
PEL41	1	0	0	0
PEL42	0	1	0	0
PEL43	0	0	1	0
NR41	2	2	2	2

$$\begin{aligned}\Delta \mathrm{Iny} = \mathrm{Iny}_2 - \mathrm{Iny}_1 = [&\mu_1 * (PEL41_2 - PEL41_1) + \\ &\mu_2 * (PEL42_2 - PEL42_1) + \\ &\mu_3 * (PEL43_2 - PEL43_1) + \\ &\mu_4 * (NR41_2 - NR41_1)] * 88.47725526\end{aligned} \tag{6.4}$$

where μ_1, μ_2, μ_3 and μ_4 represent the coefficients of $PEL41$, $PEL42$, $PEL43$ and $NR41$ in Table 6.10.

(1) group 3 relative to group 1

Note that in that comparison between group 3 and group 1—the only variables that change are PEL41 and PEL43. If I conduct a joint test on PEL41 and PEL43, I obtain a p-value of χ^2 as 0.0013. This means that education level involving these two coefficients in the other appliances system is an important factor on electricity consumption.

Replacing corresponding coefficients into (6.4):

$\Delta \mathrm{Iny} = [(0.007530) * (0 - 1) + (0.004934)*(1\text{-}0)]*$ 88.47725526

= -0.229686955

The standard error is:

$\sigma(\Delta Iny) = \sqrt{Var((-\mathrm{c(PEL41)} + \mathrm{c(PEL43)})} *$ 88.47725526

$$=\sqrt{\left(\sigma_{c(\widehat{PEL41})}\right)^2+\left(\sigma_{c(\widehat{PEL43})}\right)^2-2cov(c(PEL41),c(PEL43))}*$$
88.47725526

$$=\sqrt{(0.002070)^2+(0.002676)^2-2cov(c(PEL41),c(PEL43))}*$$
88.47725526

$$=\sqrt{(0.002070)^2+(0.002676)^2-2*2.98*10^\wedge(-6)}*\ 88.47725526$$

= 0.207230958

$$e^{\Delta\text{Iny}}-1=-0.205217634$$

Towards education's influence on appliances system, I find that residents with relatively high level consume 20.5% less electricity on appliances system compared with residents with lower education level.

$$H_0:\ e^{\Delta\text{Iny}}-1=0$$

$$H_a:\ e^{\Delta\text{Iny}}-1\ \neq 0$$

Within this test, I obtain the p-value of χ^2 is 0.2130. I cannot reject the null hypothesis at a 10% significance level. Adults under 65 and elderly adults do not show significantly different electricity consumption under other appliances system.

(2) group 4 relative to group 1

I find that the p-value PEL41 is 0.0003, this means that this variable is significant to total electricity consumption.

$$\Delta\text{Iny}=[(0.007530)*(0-1)]*1*88.47725526$$

= -0.666233732

The standard error is:

$$\sigma(\Delta Iny)=\sqrt{Var(\text{c(PEL41)})}*88.47725526$$

$$=\sqrt{\left(\sigma_{c(\widehat{PEL41})}\right)^2}*88.47725526$$

$$=\sqrt{0.002070^2}*88.47725526$$

= 0.183147918

$$e^{\Delta \mathrm{Iny}} - 1 = e^{-0.666233732} - 1 = \text{-0.486360557}$$

I obtain the conclusion that group 4 consumes less electricity consumption by 48.6% compared with group 1. That is to say, adults under 65 with post graduate degree will consume less electricity compared with adults under 65 with high school education level.

I conduct the following hypothesis test:

$$H_0\text{: } e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\text{: } e^{\Delta \mathrm{Iny}} - 1 \neq 0$$

Where $\Delta \mathrm{Iny} = [\mathrm{c}(34) * (0 - 1)] * 1 * 88.47725526.$

And c(34) represent coefficient of PEL41.

Correspondingly, I conduct a Wald test on this hypothesis test and obtain a p-value of χ^2 as 0.0000. This means that these two groups are very statistically different from each other.

(3) group 2 relative to group 1

Similarly, I can compare group 2 and group 1 and obtain:

$$\Delta \mathrm{Iny} = [0.007530 * (0 - 1) + 0.009866 * (1 - 0)]*1*88.47725526$$

$$= 0.206682868$$

The standard error is:

$$\sigma(\Delta Iny) = \sqrt{Var((-\mathrm{c(PEL41)} + \mathrm{c(PEL42)})}*1*88.47725526$$

$$= \sqrt{\left(\sigma_{c(\widehat{PEL41})}\right)^2 + \left(\sigma_{c(\widehat{PEL42})}\right)^2 - 2cov(c(PEL41), c(PEL42))}$$

$$*1*88.47725526$$

$$= \sqrt{(0.002070)^2 + (0.002144)^2 - 2cov(c(PEL41), c(PEL42))}$$

$$*1*88.47725526$$

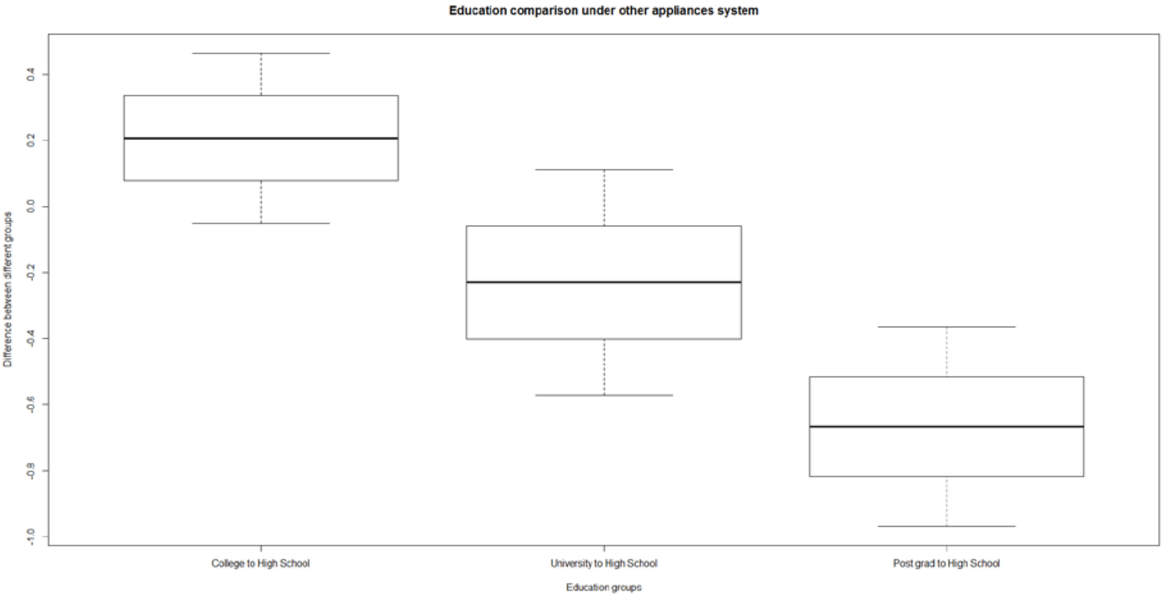
$$= \sqrt{(0.002070)^2 + (0.002144)^2 - 2 * 2.88 * 10\text{^}(-6)}*1*88.47725526$$

$$= 0.156322961$$

From three comparisons between different groups above, I can conclude the following table:

	College relative to **high school**	**University** relative to **high school**	**Post graduate** relative to **high school**
ΔIny	0.206682868	-0.229686955	-0.666233732
$\sigma\ (\Delta Iny)$	0.156322961	0.207230958	0.183147918
$\Delta \text{Iny}+1.65*\ \sigma\ (\Delta Iny)$	0.464615753	0.112244126	-0.364039667
$\Delta \text{Iny}-1.65*\ \sigma\ (\Delta Iny)$	-0.051250017	-0.571618036	-0.968427797

By R, I can build a boxplot on these comparisons:



Compared with high school, post graduate has the larger decrease, university has less decrease and college increases relative to high school in consumption regarding other appliances system.

6.4.5 Comparison between different education groups under the total of the four systems

Here we are assuming a household with electricity heating, electricity water heating and air-conditioning.

Variables	2 adults under 65 with high school (group1)	2 adults under 65 with university (group2)	2 adults under 65 with post graduate degree (group 3)
D_1	1	1	1
$\mathrm{PEL11}$	1	0	0
$\mathrm{PEL12}$	0	0	0
$\mathrm{PEL13}$	0	1	0
D_2	1	1	1
$\mathrm{PEL21}$	1	0	0
$\mathrm{PEL22}$	0	0	0
$\mathrm{PEL23}$	0	1	0
D_3	1	1	1
$\mathrm{PEL31}$	1	0	0
$\mathrm{PEL32}$	0	0	0
$\mathrm{PEL33}$	0	1	0
$\sqrt{D_4}$	85.82331696	85.82331696	85.82331696
$\mathrm{PEL41}$	1	0	0
$\mathrm{PEL42}$	0	0	0
$\mathrm{PEL43}$	0	1	0

(1) group 2 relative to group 1

Note that in that comparison between group 2 and group 1—the variables that change are PEL11, PEL13, PEL21, PEL23, PEL31, PEL33, PEL41 and PEL43. If I conduct a joint test on these variables, I obtain a p-value of χ^2 as 0.0001. This means that education level involving these twelve coefficients in all systems is a very important factor on electricity consumption.

When $\mathrm{D_1} = 1$, $\mathrm{D_2} = 1$, $\mathrm{D_3} = 1$ and $\mathrm{NR} = 2$, the mean of $\sqrt{\mathrm{HDH}}$, $\sqrt{\mathrm{SIZE}}$, $\sqrt{\mathrm{CDH}}$ and $\sqrt{D_4}$ are 330.1519852, 45.79157123, 92.11899055 and 85.82331696 respectively.

Based on corresponding coefficients, I obtain:

$$
\begin{aligned}
\Delta \mathrm{Iny} = \{ & [\mathrm{c}(5) * PEL11_2 + c(6) * PEL12_2 + c(7) * PEL13_2] * D_1 * \sqrt{\mathrm{HDH}} * \sqrt{\mathrm{SIZE}} \\
& + [\mathrm{c}(15) * PEL21_2 + c(16) * PEL22_2 + c(17) * PEL23_2] * D_2 \\
& + [\mathrm{c}(27) * PEL31_2 + c(28) * PEL32_2 + c(29) * PEL33_2] * D_3 \\
& * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} \\
& + [\mathrm{c}(34) * PEL41_2 + c(35) * PEL41_2 + c(36) * PEL41_2] * \sqrt{D_4} \} \\
& - \{ [\mathrm{c}(5) * PEL11_1 + c(6) * PEL12_1 + c(7) * PEL13_1] * D_1 * \sqrt{\mathrm{HDH}} \\
& * \sqrt{\mathrm{SIZE}} + [\mathrm{c}(15) * PEL21_1 + c(16) * PEL22_1 + c(17) * PEL23_1] \\
& * D_2 + [\mathrm{c}(27) * PEL31_1 + c(28) * PEL32_1 + c(29) * PEL33_1] * D_3 \\
& * \sqrt{\mathrm{CDH}} * \sqrt{\mathrm{SIZE}} \\
& + [\mathrm{c}(34) * PEL41_1 + c(35) * PEL41_1 + c(36) * PEL41_1] * \sqrt{D_4} \}
\end{aligned}
$$

= -0.734144227

$$e^{\Delta \mathrm{Iny}} - 1 = -0.520084017$$

Towards education’s influence on four systems, I find that residents with university education level consume 52.01% less electricity on four systems compared with residents with high school education level.

$$H_0\colon\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta \mathrm{Iny}} - 1 \neq 0$$

Within this test, I obtain the p-value of χ^2 is 0.0001. I reject the null hypothesis (<10%). High school and university education level people show very significantly different from each other on electricity consumption under the total of the four systems.

(2) group 3 relative to group 1

Note that in that comparison between group 3 and group 1—the variables that change are PEL11, PEL21, PEL31, PEL41. If I conduct a joint test on these variables, I obtain a p-value of χ^2 as 0.0062. This means that education level involving these four coefficients in all systems is a very important factor on electricity consumption.

Based on corresponding coefficients, I obtain that:

$$\Delta\text{Iny} = \{[\text{c}(5) * PEL11_3] * D_1 * \sqrt{\text{HDH}} * \sqrt{\text{SIZE}} + [\text{c}(15) * PEL21_3] * D_2 + [\text{c}(27) * PEL31_3] * D_3 * \sqrt{\text{CDH}} * \sqrt{\text{SIZE}} + [\text{c}(34) * PEL41_3] * \sqrt{D_4}\} - \{[\text{c}(5) * PEL11_1] * D_1 * \sqrt{\text{HDH}} * \sqrt{\text{SIZE}} + [\text{c}(15) * PEL21_1] * D_2 + [\text{c}(27) * PEL31_1] * D_3 * \sqrt{\text{CDH}} * \sqrt{\text{SIZE}} + [\text{c}(34) * PEL41_1] * \sqrt{D_4}\}$$

= 0.438581877

$$e^{\Delta\text{Iny}} - 1 = 0.550506849$$

Towards education's influence on the total of the four systems, I find that

residents with post graduate education level consume 55.05% more electricity on four systems compared with residents with high school education level.

$$H_0:\ e^{\Delta\text{Iny}} - 1 = 0$$
$$H_a:\ e^{\Delta\text{Iny}} - 1 \neq 0$$

Within this test, I obtain the p-value of χ^2 is 0.1195. I cannot reject the null hypothesis at a 10% significance level. High school and post graduate education level people show not very significantly different towards to electricity consumption for the total of the four systems.

6.4.6 Comparison between different age groups under an electricity water heating system

For water heating system, I consider several situations as follows:

Variables	1 adult under 65 & 1 child (group1)	2 adults under 65 (group2)	2 adults over 75 (group3)
D_2	1	1	1
AGEDIST21	0.5	0	0
AGEDIST2223	0.5	1	0
AGEDIST24	0	0	0
NR21	2	2	2

According to different groups above, I make corresponding comparisons through the following formula:

$$\Delta\text{Iny} = \text{Iny}_2 - \text{Iny}_1 = [\beta_1 * (AGEDIST21_2 - AGEDIST21_1) +$$
$$\beta_2 * (AGEDIST2223_2 - AGEDIST2223_1) +$$
$$\beta_3 * (AGEDIST24_2 - AGEDIST24_1) +$$

$$\beta_4 * (NR21_2 - NR21_1)] * 1 \tag{6.5}$$

where β_1, β_2, β_3 and β_4 represent the coefficients of $AGEDIST21$, $AGEDIST2223$, $AGEDIST24$ and $NR21$ in Table 6.10. Replacing corresponding coefficients into (6.5) by following 2 comparisons:

(1) group 2 relative to group 1

Note that in that comparison between group 2 and group 1—the only variables that change are AGEDIST21 and AGEDIST2223. If I conduct a joint test on AGEDIST21 and AGEDIST2223, I obtain a p-value of χ^2 as 0.3662. This means that age distribution level involving these two coefficients in the water heating system is not an important factor on electricity consumption.

$\Delta \text{Iny} = 0.432578 * (0 - 0.5) + 0.085747 *(1\text{-}0.5)$

$= \text{-}0.1734155$

$$e^{\Delta \text{Iny}} - 1 = \text{-}0.159211805$$

For those households with 1 child, single-parent consumes less energy compared with those with two adults by 15.9%. It is easily understood that more residents will lead less energy consumption.

$$H_0\text{: } e^{\Delta \text{Iny}} - 1 = 0$$

$$H_a\text{: } e^{\Delta \text{Iny}} - 1 \neq 0$$

The p-value of χ^2 here is 0.1468, which means that I cannot reject the null hypothesis (<10%). In another word, these two groups are not statistically different from each other.

(2) group 3 relative to group 2

Note that in that comparison between group 3 and group 2—the only variable that change is AGEDIST2223. I find that a p-value of AGEDIST2223 is 0.7142. This means that age involving these two coefficients in the water heating system is not an important factor on electricity consumption.

$\Delta \text{Iny} = 0.085747 *(0\text{-}1)$

$= \text{-}0.085747$

$$e^{\Delta \text{Iny}} - 1 = \text{-}0.082173588$$

Remaining the same number of residents in one household, the very old people (>75) would consume 8% less electricity on water heating than adults under 65. This means that very old people consume less electricity regarding water heating because that very old people will realize to save energy compared with adults.

$$H_0\colon\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta \mathrm{Iny}} - 1\ \neq 0$$

Within this test, I obtain the p-value of χ^2 is 0.7021. I cannot reject null hypothesis (<10%). Adults under 65 and over 75 show not much difference to electricity consumption coming from water heating system.

6.4.7 Comparison between different age groups under other appliances system

The only difference is $\sqrt{D_4}$, which I will take account for the average of different groups.

Variables	1 adult under 65 & 1 child (group1)	2 adults under 65 (group2)	2 adults over 75 (group3)
$\sqrt{D_4}$	92.54317582	85.6194008	84.49629082
AGEDIST41	0.5	0	0
AGEDIST4243	0.5	1	0
AGEDIST44	0	0	0
NR41	2	2	2

Similarly,

$$\begin{aligned}\Delta \mathrm{Iny} = \mathrm{Iny}_2 - \mathrm{Iny}_1 \\ &= [\theta_1 * \mathrm{AGEDIST41}_2 + \theta_2 * \mathrm{AGEDIST4243}_2 + \theta_3 * \mathrm{AGEDIST44}_2\ )] \\ &* (\sqrt{D_4})_2 - [\theta_1 * \mathrm{AGEDIST41}_1 + \theta_2 * \mathrm{AGEDIST4243}_1 + \theta_3 \\ &* \mathrm{AGEDIST44}_1\ )] * (\sqrt{D_4})_1\end{aligned} \quad (6.6)$$

where θ_1, θ_2, θ_3 and θ_4 represent the coefficients of $AGEDIST41$, $AGEDIST4243$, $AGEDIST44$ and $NR41$ in Table 6.10. Replacing corresponding coefficients into (6.6):

(1) group 2 relative to group 1:

Note that in that comparison between group 2 and group 1—the only variables that change are AGEDIST41 and AGEDIST4243. If I conduct a joint test on AGEDIST41and AGEDIST4243, I obtain a p-value of χ^2 as 0.3032. This means that age distribution level coming from these appliance system coefficients is not an important factor on electricity consumption.

$\Delta \text{Iny} = [(-0.003476) * 0 + (-0.000756) * 1]$* 85.6194008- $[(-0.003476) * 0.5 + (-0.000756) * 0.5]$* 92.54317582

$= 0.136341314$

$$e^{\Delta \text{Iny}} - 1 = 0.146072998$$

Concerning about adults and adults and children, it is normal that households with 2 adults consume 14.6% more energy than those with 1 adult and 1 child in other appliances system.

$$H_0\colon\ e^{\Delta \text{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta \text{Iny}} - 1 \neq 0$$

The p-value of χ^2 here is 0.1549, which means that I cannot reject null hypothesis (<10%). In another word, these two groups are not statistically different from each other towards other appliances electricity consumption.

(2) Group 3 relative to group 2:

Note that in that comparison between group 3 and group 2—the only variable that change is AGEDIST4243. I find that p-value of AGEDIST4243 as 0.7317. This means that age distribution level coming from these appliance system coefficients is not an important factor on electricity consumption.

$\Delta \text{Iny} = [(-0.000756) * 0]$* 84.49629082 - $[(-0.000756) * 1]$* 85.6194008

$= 0.065579584$

$$e^{\Delta \text{Iny}} - 1 = 0.067777712$$

Remaining the same number of residents in one household, the very old adults would consume only 6.8% more electricity on water heating than adults.

$$H_0\!:\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\!:\ e^{\Delta \mathrm{Iny}} - 1 \neq 0$$

Within this test, I obtain the p-value of χ^2 is 0.7398. I cannot reject null hypothesis (<10%). Adults under 65 and elderly adults show not much difference towards to electricity consumption under other appliances system.

6.4.8 Comparison between different age groups under electricity water heating and other appliances system

Variables	1 adult under 65 & 1 child (group1)	2 adults under 65 (group2)	2 adults over 75 (group3)
D_2	1	1	1
AGEDIST21	0.5	0	0
AGEDIST2223	0.5	1	0
AGEDIST24	0	0	0
$\sqrt{D_4}$	92.54317582	85.6194008	84.49629082
AGEDIST41	0.5	0	0
AGEDIST4243	0.5	1	0
AGEDIST44	0	0	0

(1) group 2 relative to group 1

Note that in that comparison between group 2 and group 1—the variables that change are AGEDIST21, AGEDIST2223, AGEDIST41 and AGEDIST4243. If I conduct a joint test on these variables, I obtain a p-value of χ^2 as 0.6103. This means that age involving these four coefficients in water heating and other appliances systems is a not important factor on electricity consumption.

For D_4 , the mean of group1 and group2 are equal to 92.54317582 and 85.6194008 respectively.

Based on corresponding coefficients, I obtain that:

$$\begin{aligned}\Delta \mathrm{Iny} = \Big\{&[\mathrm{c}(12) * AGEDIST21_2 + c(13) * AGEDIST2223_2] * D_2 \\ &+ [\mathrm{c}(37) * AGEDIST41_2 + c(38) * AGEDIST2223_2] * \sqrt{D_{4_2}}\Big\} \\ &- \Big\{[\mathrm{c}(12) * AGEDIST21_1 + c(13) * AGEDIST2223_1] * D_2 \\ &+ [\mathrm{c}(37) * AGEDIST41_1 + c(38) * AGEDIST2223_1] * \sqrt{D_{4_1}}\Big\}\end{aligned}$$

= -0.042322407

$$e^{\Delta \mathrm{Iny}} - 1 = -0.041439316$$

Towards age distribution’s influence on water heating and other appliances systems, I find that residents of 2 adults under 65 consume only 4% less electricity on water heating and other appliances systems compared with residents of 1 adult and 1 child.

$$H_0:\ e^{\Delta \text{Iny}} - 1 = 0$$

$$H_a:\ e^{\Delta \text{Iny}} - 1 \neq 0$$

Within this test, I obtain the p-value of χ^2 is 0.1951. I cannot reject null hypothesis (<10%). 1 adult and 1 child and 2 adults show not significantly different from each other towards to electricity consumption under water heating and other appliances systems.

(2) group 3 relative to group 2

Note that in that comparison between group 2 and group 1—the variables that change are AGEDIST2223 and AGEDIST4243. If I conduct a joint test on these coefficients, I obtain a p-value of χ^2 as 0.9307. This means that age distribution level of water heating and other appliances systems involving these two coefficients is a not important factor on electricity consumption.

For D_4 , the mean of group2 and group3 are equal to 85.6194008 and 84.49629082 respectively.

Based on corresponding coefficients, I obtain that:

$$\begin{aligned}\Delta \text{Iny} = \Big\{&[c(13) * AGEDIST2223_3] * D_2 + [c(38) * AGEDIST2223_3] \\ & * \sqrt{D_{4_3}}\Big\} \\ & - \Big\{[c(13) * AGEDIST2223_2] * D_2 \\ & + [c(38) * AGEDIST2223_2] * \sqrt{D_{4_2}}\Big\}\end{aligned}$$

= -0.021018733

$$e^{\Delta \text{Iny}} - 1 = -0.020799379$$

Towards age distribution’s influence on water heating and other appliances systems, I find that residents of 2 very old adults consume only 2% less electricity on water heating and other appliances systems compared with residents of 1 adult under 65.

$$H_0\colon\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta \mathrm{Iny}} - 1 \neq 0$$

Within this test, I obtain the p-value of χ^2 is 0.8880. I cannot reject the null hypothesis (<10%). 2 adult under 65 and 2 adults over 75 show not significantly different from each other towards to electricity consumption under water heating and other appliances systems.

6.4.9 Comparison between different number of residents groups under an electricity heating system

Variables	1 person (group1)	2 people (group2)
D_1	1	1
$\mathrm{NR11}$	1	2
$\sqrt{\mathrm{HDH}}$	332.2059151	330.4228891
$\sqrt{\mathrm{SIZE}}$	42.6984759	44.2193876

Note that in that comparison between group 2 and group 1—the only variable that changes is NR11. I obtain a p-value of χ^2 as 0.6548 from table 6.10. This means that number of residents involving this coefficient in the heating system is an unimportant factor on electricity consumption.

$\Delta \mathrm{Iny}$ =1.74*10^(-6)* 2*1*330.422889*44.2194 - 1.74*10^(-6)*1*

1*332.205915*42.6985

=0.026165266

When I take the exponential of $\Delta \mathrm{Iny}$, I obtain:

$$e^{\Delta \mathrm{Iny}} = e^{\mathrm{Iny}_2 - \mathrm{Iny}_1} = e^{\Delta \mathrm{Iny}} = e^{\mathrm{Iny}_2 - \mathrm{Iny}_1} = e^{In\frac{y_2}{y_1}} = \frac{y_2}{y_1}$$

$$\frac{y_2}{y_1} - 1 = e^{\Delta \mathrm{Iny}} - 1 = e^{0.026165266} - 1 = 0.026510582$$

From the result above, I obtain the conclusion that group 2 increases electricity consumption by only 2.65% compared with group 1. That is to say, 2 people in one household will consume a little more electricity compared with only 1 person.

I conduct the following hypothesis test:

$$H_0\colon\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta \mathrm{Iny}} - 1\ \neq 0$$

Where $\Delta \mathrm{Iny} = \mathrm{c}(8) * 2 * 1 *\ 330.422889 * 44.2194\ -\ \mathrm{c}(8) * 1 * 1 * 332.205915 * 42.6985$

And c(8) represents coefficient of NR11.

Correspondingly, I conduct a Wald test on this hypothesis test and obtain a p-value of χ^2 as 0.6589. I cannot reject the null hypothesis $e^{\Delta \mathrm{Iny}} - 1 = 0$ at a 10% significance level. These two groups are not statistically significant from each other when other variables are the same.

6.4.10 Comparison between different number of residents groups under an electricity water heating system

Variables	1 person (group1)	2 people (group2)
D_2	1	1
NR21	1	2

Note that in that comparison between group 2 and group 1—the only variable that changes is NR21. I obtain a p-value of χ^2 as 0.0982 from table 6.10. This means that number of residents involving this coefficient in the water heating system is an unimportant factor on electricity consumption at a 5% significance level.

$\Delta \mathrm{Iny}$ =0.079622* 2*1 - 0.079622*1*1

=0.079622

When I take the exponential of $\Delta \mathrm{Iny}$, I obtain:

$$e^{\Delta \mathrm{Iny}} = e^{\mathrm{Iny}_2 - \mathrm{Iny}_1} = e^{\Delta \mathrm{Iny}} = e^{\mathrm{Iny}_2 - \mathrm{Iny}_1} = e^{In\frac{y_2}{y_1}} = \frac{y_2}{y_1}$$

$$\frac{y_2}{y_1} - 1 = e^{\Delta \mathrm{Iny}} - 1 = e^{0.079622} - 1 = 0.082877663$$

From the result above, I obtain the conclusion that group 2 increases electricity consumption by only 8.3% compared with group 1. That is to say, 2 people in one household will consume a little more electricity compared with only 1 person.

I conduct the following hypothesis test:

$$H_0:\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a:\ e^{\Delta \mathrm{Iny}} - 1 \neq 0$$

Where $\Delta \mathrm{Iny} = \mathrm{c}(19) * 2 * 1 - \mathrm{c}(19) * 1 * 1$

And c(19) represents coefficient of NR21.

Correspondingly, I conduct a Wald test on this hypothesis test and obtain a p-value of χ^2 as 0.1116. I cannot reject the null hypothesis $e^{\Delta \mathrm{Iny}} - 1 = 0$ at a 10% significance level. These two groups are not statistically significant from each other when other variables are the same.

6.4.11 Comparison between different number of residents groups under an electricity cooling system

Variables	1 person (group1)	2 people (group2)
D_3	1	1
$\mathrm{NR31}$	1	2
$\sqrt{\mathrm{CDH}}$	97.79469127	97.55999727
$\sqrt{\mathrm{SIZE}}$	43.88480236	45.46789612

Note that in that comparison between group 2 and group 1—the only variable that changes is NR31. I obtain a p-value of χ^2 as 0.1155 from table 6.10. This means that number of residents involving this coefficient in the cooling system is not an important factor on electricity consumption.

Replacing corresponding coefficients into (6.1):

$\Delta \mathrm{Iny}$ =1.31*10^(-5)* 2*1*97.55999727*45.46789612 - 1.31*10^(-5)*1*

1*97.79469127*43.88480236

= 0.059997934

When I take the exponential of $\Delta \mathrm{Iny}$, I obtain:

$$e^{\Delta \mathrm{Iny}} = e^{\mathrm{Iny}_2 - \mathrm{Iny}_1} = e^{\Delta \mathrm{Iny}} = e^{\mathrm{Iny}_2 - \mathrm{Iny}_1} = e^{In\frac{y_2}{y_1}} = \frac{y_2}{y_1}$$

$$\frac{y_2}{y_1} - 1 = e^{\Delta \mathrm{Iny}} - 1 = e^{0.059997934} - 1 = 0.061834353$$

From the result above, I obtain the conclusion that group 2 increases electricity consumption by only 6.18% compared with group 1. That is to say, 2 people in one household will consume a little more electricity compared with only 1 person.

I conduct the following hypothesis test:

$$H_0\colon\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta \mathrm{Iny}} - 1\ \neq 0$$

Where $\Delta \mathrm{Iny} = \mathrm{c}(31) * 2 * 1 *\ 97.55999727 * 45.46789612\ -\ \mathrm{c}(31) * 1 * 1 * 97.79469127 * 43.88480236$

And c(31) represents coefficient of NR31.

Correspondingly, I conduct a Wald test on this hypothesis test and obtain a p-value of χ^2 as 0.1262. I cannot reject the null hypothesis $e^{\Delta \mathrm{Iny}} - 1 = 0$ at a 10% significance level. These two groups are not statistically significant from each other when other variables are the same.

6.4.12 Comparison between different number of residents groups under other appliances system

Variables	1 person (group1)	2 people (group2)
NR41	1	2
$\sqrt{D_4}$	73.84321861	85.9950821

Note that in that comparison between group 2 and group 1—the only variable that changes is NR41. I obtain a p-value of χ^2 as 0.4228 from table 6.10. This means that number of residents involving this coefficient in the other appliances system is not an important factor on electricity consumption.

Replacing corresponding coefficients into (6.1):

$\Delta \mathrm{Iny}$ =(-0.000379)* 2*85.9950821 - (-0.000379)*1*73.84321861

= -0.037197692

When I take the exponential of $\Delta \mathrm{Iny}$, I obtain:

$$e^{\Delta \mathrm{Iny}} = e^{\mathrm{Iny}_2 - \mathrm{Iny}_1} = e^{\Delta \mathrm{Iny}} = e^{\mathrm{Iny}_2 - \mathrm{Iny}_1} = e^{In\frac{y_2}{y_1}} = \frac{y_2}{y_1}$$

$$\frac{y_2}{y_1} - 1 = e^{\Delta \mathrm{Iny}} - 1 = e^{-0.037197692} - 1 = \text{-0.036514357}$$

From the result above, I obtain the conclusion that group 2 reduces electricity consumption by only 3.65% compared with group 1. That is to say, 2 people in one household will consume a little less electricity compared with only 1 person.

I conduct the following hypothesis test:

$$H_0\text{: } e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\text{: } e^{\Delta \mathrm{Iny}} - 1 \neq 0$$

Where $\Delta \mathrm{Iny} = \mathrm{c}(40) * 2 * 85.9950821 - \mathrm{c}(40) * 1 * 73.84321861$

And c(40) represents coefficient of NR41.

Correspondingly, I conduct a Wald test on this hypothesis test and obtain a p-value of χ^2 as 0.4139. I cannot reject the null hypothesis $e^{\Delta \mathrm{Iny}} - 1 = 0$ at a 10% significance level. These two groups are not statistically significant from each other when other variables are the same.

6.4.13 Comparison between different number of residents groups under the total of the four systems

Here we are assuming a household with electricity heating, electricity water heating and air-conditioning.

Variables	1 person (group1)	2 people (group2)
D_1	1	1
NR11	1	2
$\sqrt{\mathrm{HDH}}$	327.5720554	330.1519852
$\sqrt{\mathrm{SIZE}}$	45.11454523	45.79157123
D_2	1	1
NR21	1	2
D_3	1	1
NR31	1	2
$\sqrt{\mathrm{CDH}}$	94.18805597	92.11899055
NR41	1	2
$\sqrt{D_4}$	69.15808851	85.82331696

Group 2 relative to group 1

Note that in that comparison between group 2 and group 1—the variables that change are NR11, NR21, NR31 and NR41. If I conduct a joint test on these variables, I obtain a p-value of χ^2 as 0.0775. This means that number of residents involving these four coefficients is an important factor on electricity consumption at a 10% significance level.

When $\mathrm{D}_1 = 1$, $\mathrm{D}_2 = 1$, $\mathrm{D}_3 = 1$ and $\mathrm{NR} = 2$, the mean of $\sqrt{\mathrm{HDH}}$, $\sqrt{\mathrm{SIZE}}$, $\sqrt{\mathrm{CDH}}$ and $\sqrt{D_4}$ are 330.1519852, 45.79157123, 92.11899055 and 85.82331696 respectively.

Similarly, $\mathrm{D}_1 = 1$, $\mathrm{D}_2 = 1$, $\mathrm{D}_3 = 1$ and $\mathrm{NR} = 1$, the mean of $\sqrt{\mathrm{HDH}}$, $\sqrt{\mathrm{SIZE}}$, $\sqrt{\mathrm{CDH}}$ and $\sqrt{D_4}$ are 327.5720554, 45.11454523, 94.18805597 and 69.15808851 respectively.

Based on corresponding coefficients, I obtain that:

$$\begin{aligned}\Delta \mathrm{Iny} = \Big\{&[\mathrm{c}(8) * NR11_2] * D_1 * \sqrt{\mathrm{HDH}_2} * \sqrt{\mathrm{SIZE}_2} + [\mathrm{c}(19) * NR21_2] * D_2 \\ &+ [\mathrm{c}(31) * NR31_2] * D_3 * \sqrt{\mathrm{CDH}_2} * \sqrt{\mathrm{SIZE}_2} + [\mathrm{c}(40) * NR41_2] \\ &* \sqrt{D_{4_2}}\Big\} \\ &- \Big\{[\mathrm{c}(8) * NR11_1] * D_1 * \sqrt{\mathrm{HDH}_1} * \sqrt{\mathrm{SIZE}_1} + [\mathrm{c}(19) * NR21_1] * D_2 \\ &+ [\mathrm{c}(31) * NR31_1] * D_3 * \sqrt{\mathrm{CDH}_1} * \sqrt{\mathrm{SIZE}_1} + [\mathrm{c}(40) * NR41_1] \\ &* \sqrt{D_{4_1}}\Big\}\end{aligned}$$

= 0.12252949

$$e^{\Delta \mathrm{Iny}} - 1 = 0.130352454$$

Towards number of residents' influence on four systems, I find that 2 people consume 13.04% more electricity on four systems compared with 1 person.

$$H_0\colon\ e^{\Delta \mathrm{Iny}} - 1 = 0$$

$$H_a\colon\ e^{\Delta \mathrm{Iny}} - 1 \neq 0$$

Within this test, I obtain the p-value of χ^2 is 0.0563. I reject null hypothesis (<10%). Different number of residents in one household shows very significantly different from each other towards to electricity consumption under the total of the four systems.

With respect to number of residents, each separate system shows an insignificant effect, but four systems as a whole show an important role in electricity consumption at a 10% significance level but not important at a 5% significance level. There is an interesting problem when I consider the number of residents. For each separate system, number of residents does not show a statistically significant role in electricity consumption, however, it shows statistically significant as the total of the four systems. This could happen when the specific error components are highly correlated. I have mentioned that error terms exist in each system and outside four systems. If the error terms are highly correlated, the regression result could obtain statistically significant overall predictors but statistically non-significant separate predictors. Even there is no multicollinearity, I can still get non-significant predictors and an overall significant model if two or more variables are close to significant. From my case, I find that the p-value of number of residents under electricity water heating and cooling system is 0.1116 and 0.1262 respectively. However, the overall prediction passes the threshold of statistical significance. Also, this could happen that number of residents can't decompose its influence into four sub-models. Therefore, the overall predictors show a statistically significant role in affecting electricity consumption.

6.5 Comparison with Literature

In this section, I will compare my model along with my results with previous studies in the literature review. (Chapter 2) I classify this part into two sections, background industry, factors consideration and analysis methodology. I will also introduce the limitation of my model data.

6.5.1 Comparison with theoretical background

Study	**Study Summary**
Hass, R.(1997)	End-use can be identified with reasonable reliability;
Zhang, Q.(2004)	Considered annual consumption per household (UEC) and Heating degree-days
Nair, G., L. Gustavsson, and K. Mahapatra(2010)	Considered personal attributes such as income, education, age and contextual factors to improve the energy efficiency

These three papers mention that energy efficiency in the residential sector is discussed for many countries around the world. In Hass (1997), it mainly identifies that long-term demand depends on gross life-style and demographic factors, such as house occupancy or family size, saturation effects of energy and electricity services. In this paper, the author conducts a cross-country comparison, including USA, Japan, Germany, UK, Sweden, Denmark, France, Austria, Italy and Norway. Zhang (2004) investigates the relationship between the annual energy consumption per household (UEC) and heating degree-days for China, Japan, Canada and the United States. In the latest paper Nair (2010), it considers more specific factors, such as income, education, age and contextual factors to improve energy efficiency. This paper' database comes from Swedish residential buildings.

Although they mention that there is a relationship between weather situation or personal situations and energy consumption, they don't offer a solution how to find the relationship between energy efficiency and these factors.

In my thesis, in addition to providing the equation between energy consumption and factors, and identifying these factors, there is a quantification of these variables.

6.5.2 Comparison with explanatory factors

Study	**Study Summary**
McMakin, A.H, E.L. Malone, and R.E. Lundgren(2002)	Residents' behaviors into before-and after energy use.
Kelly, S.(2011)	A combination of physical, demographic and behavioral characteristics.
Yu, Z., et al.(2011)	Examined the influences of occupant behavior on building energy consumption.
Brounen, D., N. Kok, and J.M. Quigley(2012)	Energy influence is determined by structural dwelling characteristics, such as the vintage, building type and characteristics of the dwelling.
Hiller, C.(2012)	Clarify that characteristics of load curves differences between weekdays and weekend days.

All of these papers emphasize exploring the relationship between energy use and characteristics factors, including dwelling information, behavior changes, building type and weekdays or weekends. In McMakin (2002), it does a survey on residents' behaviors into before-and after energy use. Kelly (2011) explains that energy consumption from the residential sector is a complex problem including a combination of physical, demographic and behavioral characteristics. Yu (2011) examined the influences of households' behavior changes on total electricity consumption. Brounen (2012) illustrates that structural dwelling characteristics such as building type determine the energy consumption. Hiller (2012) clarifies that electricity load differs from weekdays and weekends.

These papers investigate the exact factors those may exert an influence on total electricity consumption in the residential sector. However, each paper has limited specification detailed factors. Referring to my model, I include four detailed comparisons into the discussion: heating, water heating, cooling and other appliances. Under each system, some common variables are shared and their variables are separately considered. From this perspective, my model covers a wide range when I consider the factors affecting total energy consumption.

6.5.3 Comparison with analysis methodology

Study	**Study Summary**
Bartels, R. and D.G Fiebig(1990)	Introduced conditional demand analysis (CDA) allocating the electricity consumption with a particular appliance.
Fiebig, D.G., R. Bartels, & D.J Aigner(1991)	Considered intensity of use of an appliance will vary from each household; Dummies indicate only absence or presence of the appliance.
LaFrance, G. and D. Perron(1994)	Presented an electricity demand survey with CDA to indicate that decreasing electricity consumption was mainly related to a large decline in net heating system.
Hsiao, C. Mountain, and K.H. Illman(1995)	CDA: the total load data of a house are regressed on appliance ownership dummy variables.
Bartels, R. & D.G. Fiebig(1996)	CDA model to show considerable potential to improve precision of estimates of end-use consumption.
Bartels, R. & D.G. Fiebig(2000)	CDA model explained how total residential load is disaggregated by end uses; For consumption of lighting, it is successfully estimated.

Regarding methodology, the papers above specify a Conditional Demand Analysis (CDA) model to build a statistical model between total end-use electricity consumption and particular appliances. Every paper has its own unique feature. Bartels (1990) introduces CDA as a method for collecting total household electricity consumption with each relevant particular appliance. Bartels and Fiebig (1991) continues to build on a previous study. They include the dummy variables indicating only absence or presence of the appliance. This may lead different electricity consumption from household to household. LaFrance (1994) conducts a large-scale survey in Quebec residential sector to present that decreasing electricity consumption for a dwelling equipped with an electric space heating system is related to a large decline in net heating consumption to a large extent. Hsiao & Mountain

(1995) propose a Bayesian framework to combine end-use information with appliance data. In addition, they consider building a regression model between total load data and appliance ownership dummy variables. Bartels & Fiebig (1996) discusses two ways to improve the statistical precision for estimates of end-use consumptions. One way is to analyse a households' total load data. The other way is to use the principles of optimal sample design to select which end-use in a household are to be metered. Bartels & Fiebig (2000) provides a very detailed picture of how total residential load is disaggregated by end-use. Especially, the consumption of lighting has been successfully estimated.

These papers introduce a significant method called CDA, which stands for a regression model between total electricity consumption and specific appliance usage. In addition, these papers have their own variables of focus. However, they do not consider all variables as a whole model. They only focus on separate specific appliance usage.

What is more important in my model, I build a regression model between total electricity consumption for one year and a large set of factors. I break down my model into four sub-models according to their categories—heating, water heating, cooling and other appliances system. Under these four systems, they share common variables and they have their own unique variables. I use dummy variable to these four systems and one of them--D_4 represents the consumption of all electric appliances.

In summary, I emphasize the advantages but also make up for some of weakness of previous studies. Admittedly, there are still limitations existing in my model. I will introduce this in following section.

6.5.4 Comparison with error term specification

Study	**Error term Description**
H.S. Houthakker(1954)	A random disturbance ε is set with a number of “economic” variables.
Caves, Herriges, Train and Windle(1987)	Random variation ε_{it} is set in the total electricity usage for customer i at time t.
Fiebig, Bartels and Aigner(1991)	A random disturbance v_i in the coefficients for the appliance dummies.
Hsiao, Mountain and Illman(1995)	An error term in the electricity heating and electricity water heating systems along with several variables.
Bartels and Fiebig(2000)	A random disturbance η_{ij} is put into total energy consumption along with economic factors such as income, price and household characteristics
Nesbakken(2001)	A random variable μ_j with zero conditional expectation given that heating system *j* is chosen.
Lins, DaSilva and Rosa(2003)	An error term ε_t is set for each household in CDA.
Aydinalp, Ugursal and Fung(2003)	A random error term e_{ijt} is included in estimating end-use energy consumption with a variety of factors for household *I*’s appliance *j* in period *t*.
Brounen, Kok and Quigley(2012)	Error term ε_i is considered into electricity consumption for dwelling i.
Newsham and Donnelly(2013)	Error term ε_i is set in household *i*’s annual energy use.

Houthakker (1954) sets a random disturbance ε with a number of “economic” variables (income, marginal price of electricity, marginal price of gas, average holdings of domestic equipment) in the demand equation for electricity consumption. For each household, the error term ε is a normal variable independent of the predetermined variables and of the errors in the corresponding equation for other consumers.

Caves, Herriges, Train and Windle (1987) observe usage data (via conditional demand estimates) and then used the data to modify a set of prior beliefs by engineering approach. The last but not least, the authors transform them into a posterior distribution to describe appliance usage. Random variation ε_{it} is set in the total electricity usage for customer i at time t.

Similarly, Fiebig, Bartels and Aigner (1991) applies CDA model into estimate end-use load curves. However, the only difference from other papers is that coefficients of appliance dummies as random rather than fixed. To achieve this goal, the authors add a random disturbance v_i in the coefficients for the appliance dummies.

Hsiao, Mountain and Illman (1995) propose a better method called Bayesian framework to combine end uses monitoring information with the aggregate-load/appliance data. The authors set an error term in the electricity heating and electricity water heating systems along with several variables, such as education distribution, number of residents, the house age, square footage and number of dishwashers.

Bartels and Fiebig (2000) estimate the consumption of high penetration end uses such as lighting successfully and determine the costs to distributors related to individual end uses. A random disturbance η_{ij} is put into total energy consumption along with economic factors such as income, price and household characteristics.

Nesbakken (2001) emphasizes on the relationship between the choice of heating equipment and energy consumption. Regarding energy demand function, μ_j, a random variable with zero conditional expectation given that heating system j is chosen. μ_j is assumed to follow from a normal distribution with expectation zero and constant variance, given the heating system j.

Lins, DaSilva and Rosa (2003) applies the CDA technique to estimate appliances' consumption. In the basic model, a relationship between electricity consumption and each appliance is set for each household. In addition, there is an error term in each household, ε_t.

Aydinalp, Ugursal and Fung (2003) discuss three methods—engineering method, conditional demand analysis and neural network method and conduct a comparative assessment of them. In developing a CDA model, a random error term e_{ijt} is included in estimating end-use energy consumption with a variety of factors for household *I*'s appliance *j* in period *t*.

Brounen (2012) analyzes how the gas and electricity is determined by the technical specifications of the dwelling and demographic characteristics of the residents. In the gas and electricity consumption model, the author includes a vector of the hedonic characteristics of building, a dummy variable representing existence of building in province and one error term, which is assumed to be independent identically distributed.

Newsham and Donnelly (2013) applies CDA to estimate the average annual energy use of different electrical and natural gas appliances and derives energy reductions related to appliances' upgrades and behaviors. For each household i, an error term ε_i is along with sum product of annual energy use and number of appliance.

Almost all previous papers focus on adding error terms for each household for each appliance at certain time periods, which are based on CDA model. Compared previous studies regarding error terms, my model takes account comprehensively-- four error terms in four systems and a separate constant error term. In addition, correlation exists among these five error terms. After several trials for model selection, I choose the most suitable model with only correlated errors between constant coefficients and other appliances system. (shown in 6.2.7)

6.5.5 Limitations of my model data

Although my model includes all aspects (i.e. explanatory factors and estimation methodology) from previous studies, I have to admit that limitations exist in my model.

Certainly, some limitations exist in the survey results. The biggest problem is the missing information. I have to solve this problem depending on the feature of information. For example, for the house age, it is hard to estimate without any verification. However, for the size of living space, I can make up the missing data by existing data and some possible relevant elements. I will explain how I deal with missing information as follows:

(1) Size of living space

Recall for 4.2.1 (3), I build a regression model on households whose information is complete or satisfied with a ratio range from 50 to 250. The regression equation includes size of living space and number of rooms. In this way, I obtain the complete reasonable information for size of house.

(2) Missing house age

Recall for 4.2.1 (6), households with missing house age could not be estimated. Therefore, those households with missing house age will be deleted through database.

(3) Income

Recall for 4.2.1 (7), I use the same estimation method with size of living space. Now, I build a regression model between income and size of living space and education level.

Even though there is a lack of data in size of living space, house age and income, I have built regression model to estimate them according to existing complete information. I made the data as “perfect” as I can.

(4)Time lag problem

In my model, I consider only 1 year of cross-section data, so there is no opportunity to interpret time in the analysis. As well, there is limited variation in some variables. (e.g. TUP, weather days)

6.6 Summary

In this chapter, I have discussed all the results I obtain from my real database and Eviews software from OLS and GLS with and without correlation between error terms. Model in 6.2.7 are the most “Perfect ”model until I tried ever. From the result shown in χ^2 test in 6.3, I can recognize that PEL (Percentage of education level distribution) under four systems exerts a significant role in influencing the total electricity consumption if I use 10% as the standard significance level. From 6.3 section result, it shows that age distribution, number of residents and income are partly important to energy consumption for households under systems depending on different combination. Statistical analysis focusing on demographic factors such as age and education distribution level in households is discussed in 6.4. In 6.5, I describe the comparison between previous studies and my model, including background, factors consideration, methodology, and error terms. In addition, limitation exists in my model. However, I make regression to estimate them based on relevant factors.

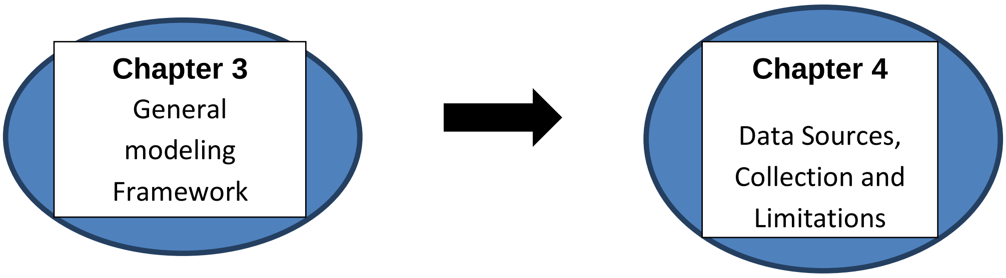
Chapter 7 Summary and Conclusions

7.1 Overview

A significant amount of information for my thesis is drawn from the collaborative project involving McMaster University and Hydro One. My thesis mainly focuses on residential electricity consumption efficiency and the relationship between the total electricity consumption and a number of variables, including dwelling information, time-of-use prices, weather data and demographic factors. In my model, I have successfully estimated the relationship between final electricity consumption and demographic variables, such as age level distribution, education level distribution and number of residents in the household. I am particularly interested in the influence of demographics. Regarding influence of demographic variables, I find that education distribution level exerts a significant impact on total electricity consumption as each separate system and four systems and number of residents as a whole play an important role in electricity consumption. I will give a summary of my research process in the following sections.

7.2 Model Framework and Data Preparation

With respect to the model framework and data preparation, I display it as the following figure:



In chapter 3, firstly, I have introduced the theoretical background about my final model—an example of a Conditional Demand Analysis Model. In my case, I have simplified the conditional demand analysis model. What is more, combining with my database, I did not consider the time variation for the price of my data. Secondly, I have introduced the variables under four systems components —heating, water heating, cooling and other appliances system. Based on the components, I have given explicit explanation of each explanatory variable under the four systems. Thirdly, I have split the four systems into four sub-models and provided a general framework of my final model.

Regarding chapter 4—Data Sources and Collection, I have categorized my data into four sources—dwelling and household information, consumption data, weather data and price data. With respect to dwelling and household information, I check the accuracy of house age, house size, the logical relationship among age distribution, number of residents and education level distribution, main fuel source and income estimation. I have selected households' consumption data from an original Hydro One data set. Regarding weather data, I have combined the location of each household with respective temperature, humidex and wind chill and calculated heating degree hours and cooling degree hours respectively. The last but not least, I have considered the average price through year 2013 for each household according to a specific formula accounting for time-differentiated prices.

7.3 Methodology

Chapter 5 mainly focuses on discussing the details of my model and estimation methodology. I have already introduced abbreviations or symbols of variables in four systems in chapter 3. In chapter 5, I have explained explicitly the variables under each system. Here, the only difference from chapter 3 is that I define the different variables in different systems in detail. Regarding the model estimation part, I have illustrated three versions in a successive order—model without error terms, model without interactive error terms but with one error term ε_0, model with interactive error terms (ε_0 and four error terms in four systems separately) and corresponding specific statistical assumptions in my model.

In chapter 5, I restate the statistical assumptions (methodology) in my model. First of all, I have assumed that one general error term and one error term under each system, in total 5 error terms. In this part, I discuss the general case for assumptions, correlation existing among these five error terms. For this case, variance includes 15 terms. One special case for this situation has been discussed—no correlation between error terms. In this case, five terms are in the variance equation.

In addition to the methodology discussion, I have added the application of my model in Eviews. Including coefficients, I have defined in detail each variable and simplified these explanatory variables into a variable with an abbreviated name, which is easier in reference in model discussion in chapter 6.

7.4 Findings

(1) Model estimation

Results from different versions of models and respective discussion are introduced in chapter 6. I have built a figure (Figure 6.1) illustrating the logical order of all my versions of models and connected them to corresponding theory in chapter 5.

Here, I will include the details of 7 versions of models are shown as follows: (Table 7.1)

Version # of Model	Methodology	Model Variables	Variables to be included in variance model of the error	Results (Significant Variables)
1	Ordinary Least Squares Model with one general error term	Original variables	None	Education level under heating and other appliances system
2	Generalized Linear Squares Model with correlated error terms	Original variables	Combining the coefficients of D_2 and ${D_2}^2$	Education level under cooling and other appliances system
3	Generalized Linear Squares Model with correlated error terms	Original variables	Only keep the coefficients related to ${z_1}^2$, ${z_2}^2$, ${z_3}^2$, ${z_4}^2$, z_4	Education level under heating, cooling and other appliances system
4	Generalized Linear Squares Model with uncorrelated error terms	Original variables	Only keep coefficients ${z_1}^2$, ${z_2}^2$, ${z_3}^2$, ${z_4}^2$	It is obvious that correlation exists between ε_0 and ε_4, so I decide to use version 3 as base model for further discussion.
5(1)	Generalized Linear Squares Model with correlated error terms	Combining age distribution group 2 & 3 as a group variable	Only keep the coefficients related to ${z_1}^2$, ${z_2}^2$, ${z_3}^2$, ${z_4}^2$, z_4 (version 3 variance)	Education level under heating, cooling and other appliances system
5(2)	Generalized Linear Squares Model with correlated error terms	Combine age distribution group 2 & 3 and houseage group 3 & 4 & 5	Only keep the coefficients related to ${z_1}^2$, ${z_2}^2$, ${z_3}^2$, ${z_4}^2$, z_4 (version 3 variance)	Education level under heating, cooling and other appliances system
6	Generalized Linear Squares Model with correlated error terms	Combine age distribution group 2 & 3 and houseage group 3 & 4 & 5	Using residuals from 5(2) reestimate variance and keeping the coefficients related to ${z_1}^2$, ${z_2}^2$, ${z_3}^2$, ${z_4}^2$, z_4	Education level under cooling and other appliances system
7	Generalized Linear Squares Model with correlated error terms	By dropping houseage 12-15, agedist group11-14 and 31-34 and tups	Using version 6 modeling of error variance	Part of education factors in heating and water heating and all education factors in cooling and other appliances system

Table 7.1 Details of 7 versions of models

For my first model where I only have one error (ε_0), I have introduced the classical case—ordinary least squares model. In this version, I keep every variable in my model and find that only education variables under heating and other appliances system are significant to the total electricity consumption.

Regarding the second model where errors are specified for each system, I have discussed generalized linear squares methodology based on the theory in chapter 5. Here, the variance equation includes 15 terms and the coefficients of ${z_2}^2$ and z_2 (coefficient in the water heating system) are separated. However, I have found that singularities exist between ${z_2}^2$ and z_2 and obtained unreasonable results. Due to the perfect correlation that exists between ${z_2}^2$ and z_2, I have combined the coefficients of these two terms. I have obtained the coefficients results and replaced them to get the new residuals for reestimates of the variance. I obtain estimates of the variance of ${\sigma_\vartheta}^2$ (which is non-constant) and then transform variables (by P^{-1}) to achieve new GLS results. In this version, the results continue to show education level distribution under cooling and other appliances system play an important role in total electricity consumption.

From the results in the second model, I have found that many of variables involving z_i when estimating ${\sigma_\vartheta}^2$ are insignificant. Therefore, after conducting some tests, I decided to keep the coefficients related to ${z_1}^2$, ${z_2}^2$, ${z_3}^2$, ${z_4}^2$, z_4 (coefficients of heating, water heating, cooling and other appliances system)--this is the third version of my model. By using similar residual estimation method, the result obtained from this version shows that education level distribution under heating, cooling and other appliances system exert a very significant impact on the total electricity consumption.

I also begin to discuss the situation with no correlation between error terms in the fourth version of model. This model is based on the special case of GLS theory from chapter 5. With regard to coefficients in estimating residuals, I did not consider the coefficient related to z_4. However, I have found that the p-value of coefficient related to z_4 shows a very significant role in residual estimation model. Therefore, I decide to use the model with correlated error terms.

By using version 3 of the model, I have compared models with different combinations of variables. For the first combination, I have combined age distribution group 2&3 as a group variable—5(1) version of the model. The result shows that education level under heating, cooling and other appliances system is very important to electricity consumption. Concerning the second combination, I combine age distribution group 2&3 and houseage group 3&4&5—5(2) version of the model. The conclusion is similar to that in the first change, education distribution level under heating, cooling and other appliances system has an important impact on total electricity consumption.

I decide to focus on the variables in the second combination, the only difference I have made is that I use the new estimated residual to reestimate the variance of the regression model—version 6 of the model. It is evident that education level under cooling and other appliances system is very important.

In the final model, version 7, I have shown that the results after I drop some group variables, such as houseage under heating system, age distribution level under heating and cooling system and all time-of-use prices based on version 6 model. Before I drop these group variables, I have done Wald tests on them and obtained a p-value over 10% significance level. Finally, I have decided my final GLS regression model and concluded that part of education factors in heating and water heating and all education factors in cooling and other appliances system are significant in explaining total electricity consumption.

(2) χ^2 test on demographic factors

In this section, I classify these variables into several groups—all demographic, only education and number of residents and income separately under four systems. I have done Wald test on these group variables and obtained significant group variables as follows: (Table 7.2)

System	System Significant Group Variables
Water heating	All demographic factors (education, age, number of residents and income)
Water heating	Age, number of residents and income
Cooling	All demographic factors (education, number of residents and income)
Cooling	Education
Cooling	Number of residents and income
Other appliances	All demographic factors (education, age, number of residents and income)
Other appliances	Education

Table 7.2 Wald test on group variables and significant group variables

From the χ^2 test results above, I obtain the conclusion that all demographic factors including or excluding education level under water heating system exert a significant role in electricity consumption. All demographic factors, only education and only number of residents and income under cooling system are all important to households' energy consumption. Regarding other appliances system, all demographic factors and only education affect total electricity consumption very significantly.

(3) Illustrative comparisons among demographic factors

The main purpose of my thesis is to explore the relationship between total electricity consumption and education distribution level, age distribution level and number of residents under four systems with some illustrative comparisons. I have made a table to show the estimated influence of different demographic variables on total electricity consumptions for some illustrative comparisons: (Table 7.3)

System	Compared Groups	Demographic variables	Influence to total electricity consumption
Electricity heating	2 adults under 65 with university **relative to** 2 adults under 65 with high school	Education between university and high school	Significant at 10% significance level **
	2 adults under 65 with post graduate **relative to** 2 adults under 65 with high school	Education between post graduate and high school	Not significant at 10% significance level
Electricity water heating	2 adults under 65 with university **relative to** 2 adults under 65 with high school	Education between university and high school	Not significant at 10% significance level
	2 adults under 65 with post graduate **relative to** 2 adults under 65 with high school	Education between post graduate and high school	Not significant at 10% significance level
Electricity cooling	2 adults under 65 with university **relative to** 2 adults under 65 with high school	Education between university and high school	Significant at 10% significance level *
	2 adults under 65 with post graduate **relative to** 2 adults under 65 with high school	Education between post graduate and high school	Significant at 10% significance level **
Other appliances	2 adults under 65 with university **relative to** 2 adults under 65 with high school	Education between university and high school	Not Significant at 10% significance level
	2 adults under 65 with post graduate **relative to** 2 adults under 65 with high school	Education between post graduate and high school	Significant at 10% significance level ***
The total of the four systems	2 adults under 65 with university **relative to** 2 adults under 65 with high school	Education between university and high school	Significant at 10% significance level ***
	2 adults under 65 with post graduate **relative to** 2 adults under 65 with high school	Education between post graduate and high school	Not significant at 10% significance level
Electricity water heating	2 adults under 65 **relative to** 1 adult under 65 & 1 child	Age between adult and child	Not significant at 10% significance level
	2 adults over 75 **relative to** 2 adults under 65	Age between old person and adult	Not significant at 10% significance level
Other appliances	2 adults under 65 **relative to** 1 adult under 65 & 1 child	Age between adult and child	Not significant at 10% significance level
	2 adults over 75 **relative to** 2 adults under 65	Age between old person and adult	Not significant at 10% significance level
Electricity water heating and other appliances	2 adults under 65 **relative to** 1 adult under 65 & 1 child	Age between adult and child	Not significant at 10% significance level
	2 adults over 75 **relative to** 2 adults under 65	Age between old person and adult	Not significant at 10% significance level
Electricity heating	2 people **relative to** 1 person	Number of residents	Not significant at 10% significance level
Electricity water heating	2 people **relative to** 1 person	Number of residents	Not significant at 10% significance level
Electricity cooling	2 people **relative to** 1 person	Number of residents	Not significant at 10% significance level
Other appliances	2 people **relative to** 1 person	Number of residents	Not significant at 10% significance level
The total of the four systems	2 people **relative to** 1 person	Number of residents	Significant at 10% significance level **

Table 7.3 Influence of different demographic variables on total electricity consumptions for some illustrative comparisons

Notes: * represents the p–value ranges from 0.08 to 0.10,

** represents the p–value ranges from 0.03 to 0.0.08,

*** represents the p–value ranges from 0.00 to 0.03.

With regard to influence of factors on residential electricity consumption, each paper in the literature has its limited detailed factors. In my model, I have included four sub-systems, heating, water heating, cooling and other appliances. My model covers a wider range of variables influencing total electricity consumption.

Besides demographic factors such as education, age, number of residents and annual income, I include age of house, age of air-conditioner or heat pump, time-of-use price and conservation measures in four sub-models. For age of house, age of air-conditioner or heat pump, I categorize them into several groups and insert dummy variables. For the time-of-use price, I transfer it from hourly data to yearly data according to different time periods, seasons and geographical areas. Concerning conservation measures, I consider the summation of corresponding dummy variables as the conservation measures variable of four systems.

From the summary table above, I have concluded that education level between university and high school under heating system is significant. For the water heating system, education doesn’t play an important role. In the cooling system, education level exerts significant influence among university, high school and post graduate comparisons. Education plays a non-ignorable role between post graduate and high school in other appliances system. For the total of the four systems, education in university and high school shows significantly different from each other.

With regard to age distribution level, it doesn’t show very significant in water heating, other appliances system and the total of these two systems.

With respect to number of residents, each separate system shows an insignificant effect, but four systems as a whole show an important role in electricity consumption at a 10% significance level but not important at a 5% significance level.

References

Aigner, D.J., C. Sorooshian, and P. Kerwin, (1984). *Conditional demand analysis for estimating residential end-use load profiles,* The Energy Journal, p. 81-97.

Aydinalp M, Ugursal VI, Fung AS, (2002). *Modeling of the appliance, lighting, and space cooling energy consumption in the residential sector using neural networks*. Appl Energy, 71(2):87–110.

Aydinalp, M., V. Ugursal, and a.S. Fung (2003, March). *Modelling of residential energy consumption at the national level*, International Journal of Energy Research 27 (4), 441–453.

Aydinalp, M., V. Ismet Ugursal, and A.S. Fung, (2004). *Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural Networks,* Applied Energy, **79**(2): p. 159-178.

Aydinalp-Koksal, M. and V.I. Ugursal, (2008). *Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector,* Applied Energy, **85**(4): p. 271-296.

Bartels, R. and D.G. Fiebig (1990), *Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Loads*, The Energy Journal, 11:4, 79-97.

Bartels, R. & D. G. Fiebig (1996). *Metering and modelling residential end-use electricity load curves,* Journal of forecasting 15 (6), 415–426.

Bartels, R. & D. G. Fiebig (2000). *Residential end-use electricity demand: results from a designed experiment*, The Energy Journal 21 (2), 51–81.

Bernard, J-T. and G. Lacroix (2005), *Conditional Demand Analysis: Tests for Homoskedasticity and Uniformity*, Mimeo, GREEN, Department of Economics, Laval University, June.

Brounen, D., N. Kok, and J.M. Quigley, (2012). *Residential energy use and conservation: Economics and demographics,* European Economic Review.

Caves, D. W., J. a. Herriges, K. E. Train, and R. J. Windle (1987, August). *A Bayesian Approach to Combining Conditional Demand and Engineering Models of Electricity Usage,* The Review of Economics and Statistics 69 (3), 438.

Chedid, R.B. and R.F. Ghajar, (2004). *Assessment of energy efficiency options in the building sector of Lebanon,* Energy Policy, **32**(5): p. 647-655.

D.Datta and S.A. Tasson, *Application of Neural Networks for the Prediction of the Energy Consumtpion in a Supermarket.*

Farahbakhsh H, Ugursal VI, Fung AS, (1998). *A residential end-use energy consumption model for Canada*. Int J Energy Res, 22(13):1133–43.

Fiebig, D. G., R. Bartels, & D. J. Aigner (1991). *A random coefficient approach*

to the estimation of residential end-use load profiles, Journal of Econometrics

50 (3), 297–327.

Goldfarb, D. L. and R. Huss (1988). *Building Utility Scenarios for an Electric*, Long Range Planning 21 (2), 78–85.

Gyberg, P. and J. Palm, (2009). *Influencing households' energy behaviour-how is this done and on what premises?* Energy Policy, **37**(7): p. 2807-2813.

Haas, R., (1997). *Energy efficiency indicators in the residential sector: What do we know and what has to be ensured?* Energy Policy, **25**(7-9): p. 789-802.

Hiller, C., (2012). *Influence of residents on energy use in 57 Swedish houses measured during four winter days,* Energy and buildings, **54**: p. 376-385.

H. S. Houthakker, (1951). *Some Calculations on Electricity Consumption in Great Britain,* Journal of the Royal Statistical Society, Series A (General),Vol. 114, No. 3, pp. 359-371

Hsiao, C., C. Mountain, and K. H. Illman (1995, July). *A Bayesian Integration of End-Use Metering and Conditional-Demand Analysis*, Journal of Business and Economic Statistics 13 (3), 315.

Hsiao, C., C. Mountain, (Jun. 1985). *Estimating the short-run income elasticity of demand for electricity by using cross-sectional categorized data*, Journal of the American Statistical Association, Vol. 80, No. 390, pp. 259-265.

Jaber, J.O., (2002). *Prospects of energy savings in residential space heating.* Energy and buildings, **34**(4): p. 311-319.

Kelly, S., (2011). *Do homes that are more energy efficient consume less energy? A structural equation model of the English residential sector,* Energy, **36**(9): p. 5610-5620.

LaFrance, G. and D. Perron (1994). *Evolution of Residential Electricity Demand by End-Use in Quebec 1979-1989: A Conditional Demand Analysis Evolution of Residential Electricity Demand by End-Use in Quebec*, Energy Studies Review 6 (2), 164–173.

Leighty, W. and A. Meier, (2011). *Accelerated electricity conservation in Juneau, Alaska: A study of household activities that reduced demand 25%,* Energy Policy, **39**(5): p. 2299-2309.

Lins, M., A. DaSilva, and L. Rosa (2003). *Regional Variations in Energy Consumption of Appliances: Conditional Demand Analysis Applied to Brazilian Households*, Annals of Operations Research, 235–246.

McMakin, A.H., E.L. Malone, and R.E. Lundgren, (2002). *Motivating residents to conserve energy without financial incentives,* Environment and Behavior, **34**(6): p. 848-863.

Nair, G., L. Gustavsson, and K. Mahapatra, (2010). *Factors influencing energy efficiency investments in existing Swedish residential buildings,* Energy Policy, **38**(6): p. 2956-2963.

Nässén, J., F. Sprei, and J. Holmberg, (2008). *Stagnating energy efficiency in the Swedish building sector—Economic and organisational explanations,* Energy Policy, **36**(10): p. 3814-3822.

Nesbakken, (2001). R. *Energy consumption for space heating: a discrete-continuous approach,* The Scandinavian journal of economics, **103**(1): p. 165-184.

Newsham, G. R. & C. L. Donnelly (2013). *A model of residential energy end-use in canada: Using conditional demand analysis to suggest policy options for community energy planners*, Energy Policy 59, 133–142.

Ouyang, J. and K. Hokao, (2009). *Energy-saving potential by improving occupants' behavior in urban residential sector in Hangzhou City, China,* Energy and buildings, **41**(7): p. 711-720.

Parti M. and C. Parti (1980). *The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector*, The Bell Journal of Economics 11(1): 309-321.

Sadineni, S.B., T.M. France, and R.F. Boehm, (2011). *Economic feasibility of energy efficiency measures in residential buildings.* Renewable energy, **36**(11): p.

2925-2931.

Swan, L. G. and V. I. Ugursal (2009, October). *Modeling of end-use energy consumption in the residential sector: A review of modeling techniques*, Renewable and Sustainable Energy Reviews 13 (8), 1819–1835.

Wall, R. and T. Crosbie, (2009). *Potential for reducing electricity demand for lighting in households: An exploratory socio-technical study.* Energy Policy, **37**(3): p.

1021-1031.

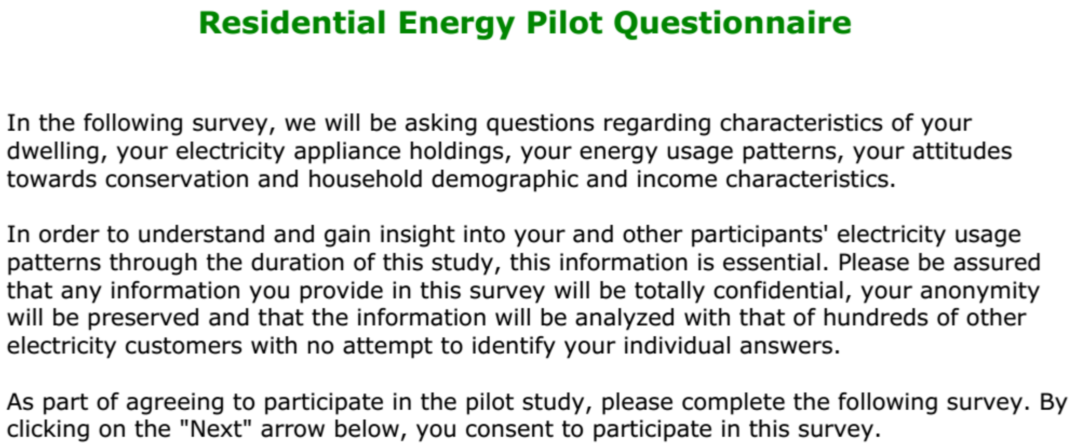
Yang, Y., B. Li, and R. Yao, (2010). *A method of identifying and weighting indicators of energy efficiency assessment in Chinese residential buildings,* Energy Policy, **38**(12): p. 7687-7697.

Yu, Z., et al., (2011). *A systematic procedure to study the influence of occupant behavior on building energy consumption,* Energy and buildings, **43**(6): p. 1409-1417.

Zhang, Q., (2004). *Residential energy consumption in China and its comparison with Japan, Canada, and USA.* Energy and buildings, **36**(12): p. 1217-1225.

Appendices

Appendix 1: Pilot Questionnaire

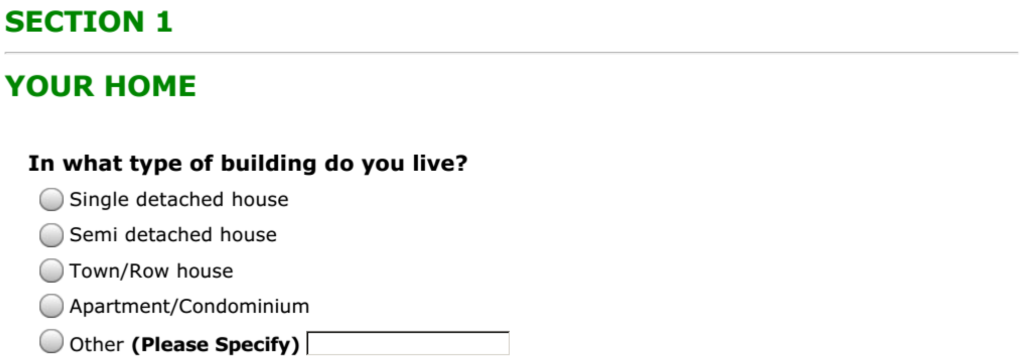


Residential Energy Pilot Questionnaire

In the following survey, we will be asking questions regarding characteristics of your dwelling, your electricity appliance holdings, your energy usage patterns, your attitudes towards conservation and household demographic and income characteristics.

In order to understand and gain insight into your and other participants' electricity usage patterns through the duration of this study, this information is essential. Please be assured that any information you provide in this survey will be totally confidential, your anonymity will be preserved and that the information will be analyzed with that of hundreds of other electricity customers with no attempt to identify your individual answers.

As part of agreeing to participate in the pilot study, please complete the following survey. By clicking on the "Next" arrow below, you consent to participate in this survey.

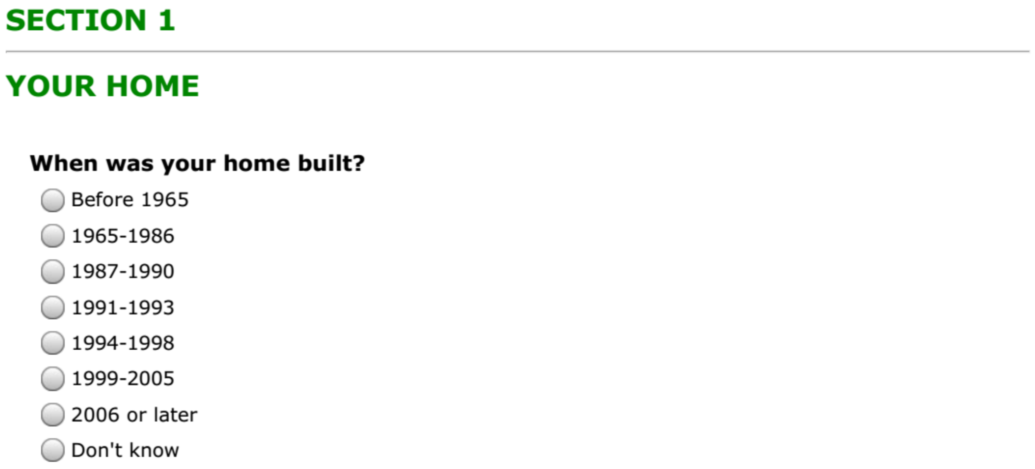


SECTION 1

YOUR HOME

In what type of building do you live?

- ○ Single detached house
- ○ Semi detached house
- ○ Town/Row house
- ○ Apartment/Condominium
- ○ Other **(Please Specify)**

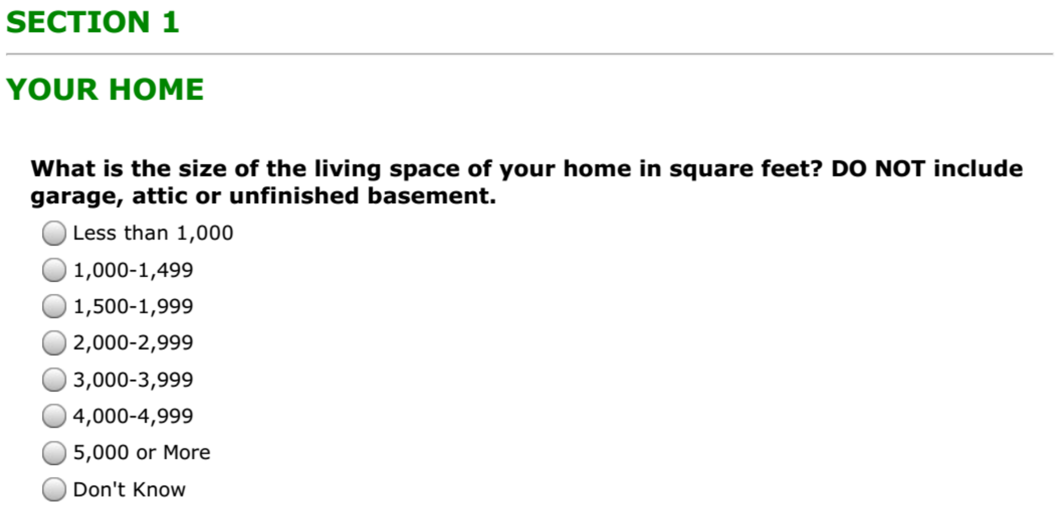


SECTION 1

YOUR HOME

When was your home built?

- ○ Before 1965
- ○ 1965-1986
- ○ 1987-1990
- ○ 1991-1993
- ○ 1994-1998
- ○ 1999-2005
- ○ 2006 or later
- ○ Don't know

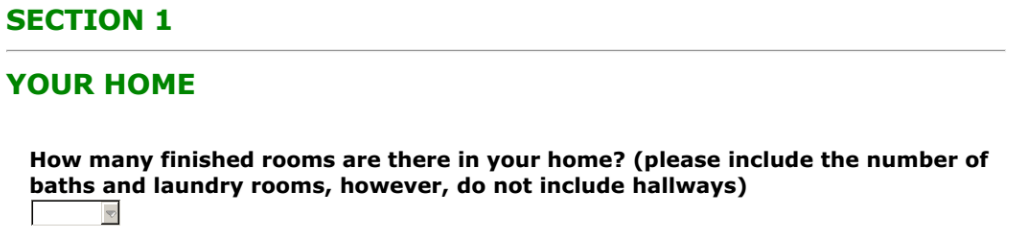


SECTION 1

YOUR HOME

What is the size of the living space of your home in square feet? DO NOT include garage, attic or unfinished basement.

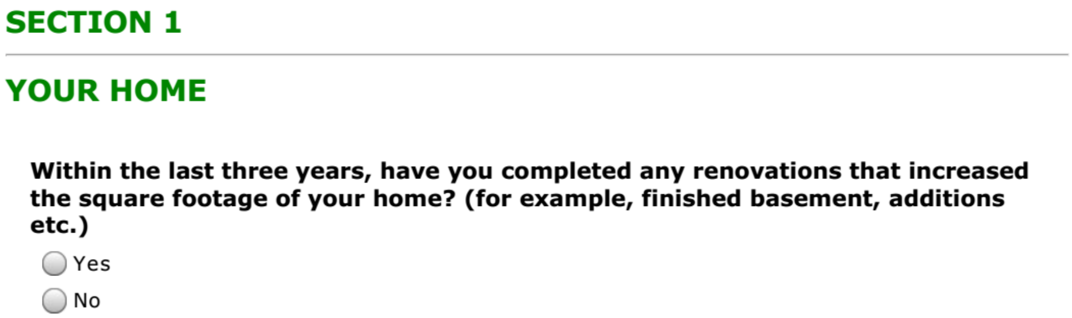
- ○ Less than 1,000
- ○ 1,000-1,499
- ○ 1,500-1,999
- ○ 2,000-2,999
- ○ 3,000-3,999
- ○ 4,000-4,999
- ○ 5,000 or More
- ○ Don't Know



SECTION 1

YOUR HOME

How many finished rooms are there in your home? (please include the number of baths and laundry rooms, however, do not include hallways)

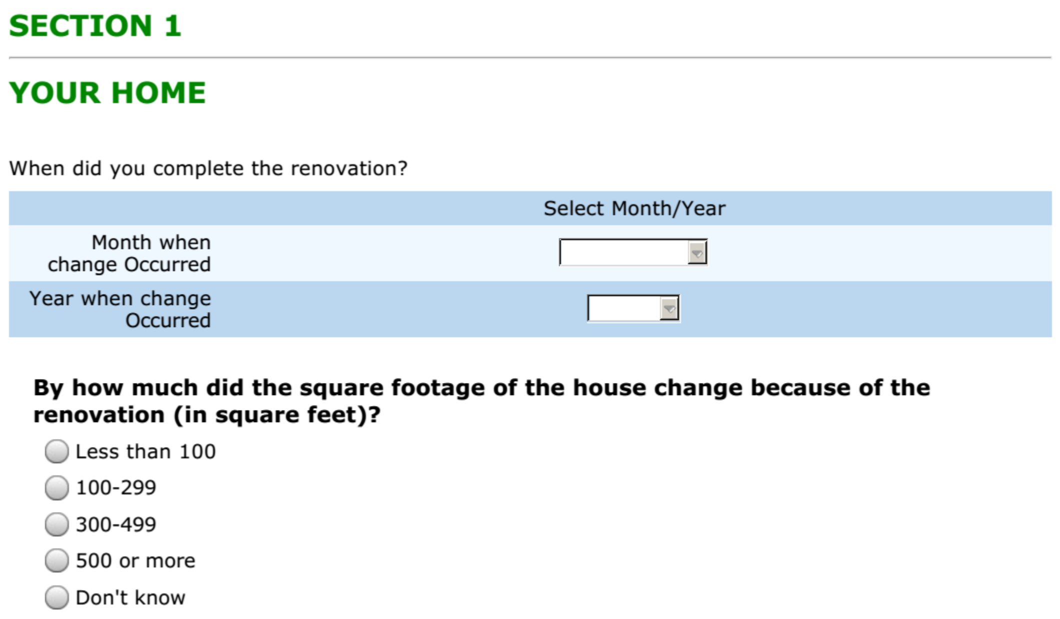


SECTION 1

YOUR HOME

Within the last three years, have you completed any renovations that increased the square footage of your home? (for example, finished basement, additions etc.)

- ○ Yes
- ○ No



SECTION 1

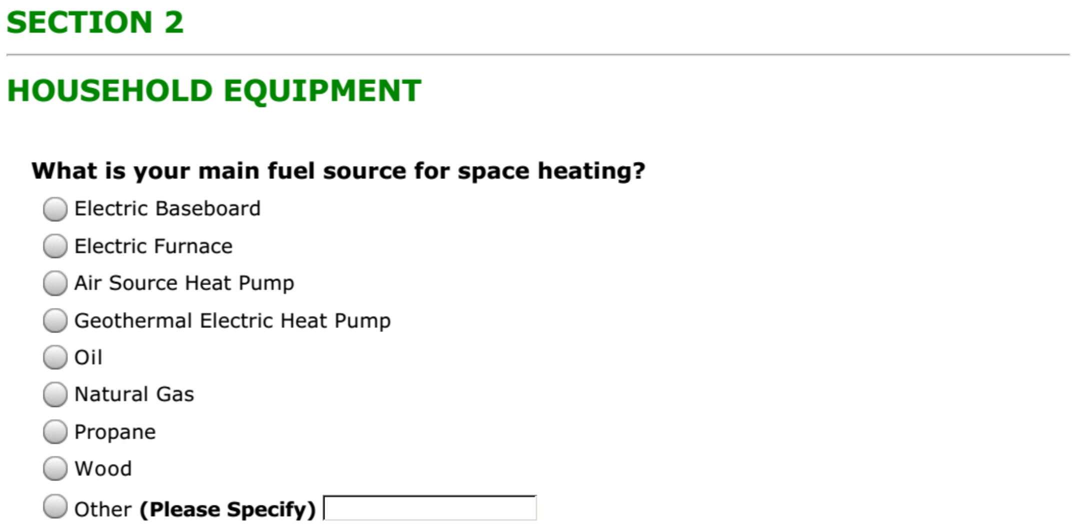
YOUR HOME

When did you complete the renovation?

	Select Month/Year
Month when change Occurred	
Year when change Occurred	

By how much did the square footage of the house change because of the renovation (in square feet)?

- ○ Less than 100
- ○ 100-299
- ○ 300-499
- ○ 500 or more
- ○ Don't know

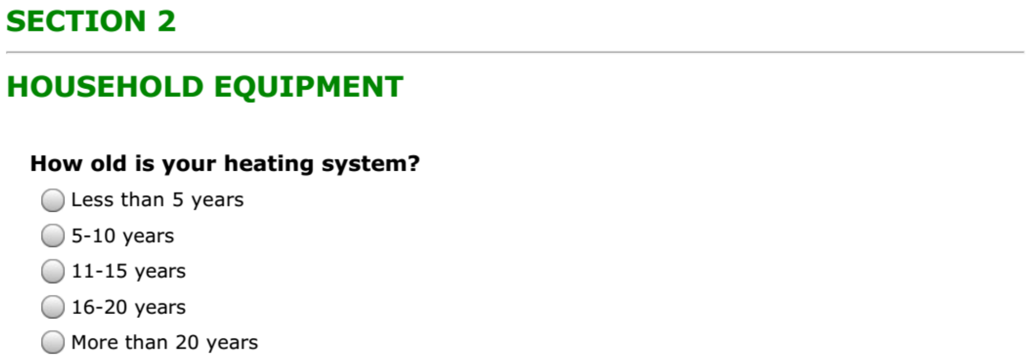


SECTION 2

HOUSEHOLD EQUIPMENT

What is your main fuel source for space heating?

- ○ Electric Baseboard
- ○ Electric Furnace
- ○ Air Source Heat Pump
- ○ Geothermal Electric Heat Pump
- ○ Oil
- ○ Natural Gas
- ○ Propane
- ○ Wood
- ○ Other **(Please Specify)** ☐

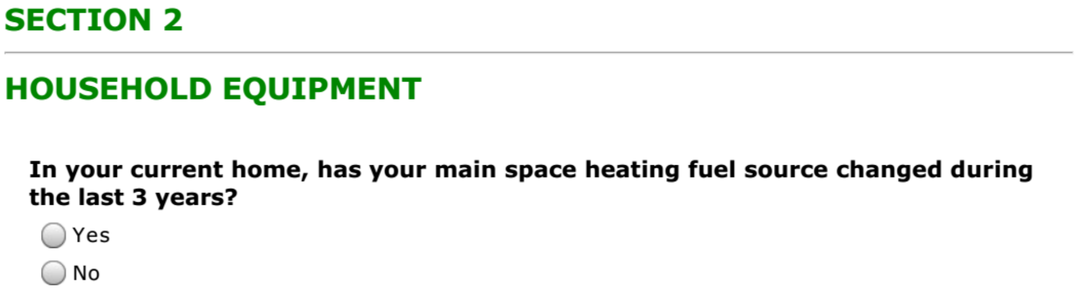


SECTION 2

HOUSEHOLD EQUIPMENT

How old is your heating system?

- ○ Less than 5 years
- ○ 5-10 years
- ○ 11-15 years
- ○ 16-20 years
- ○ More than 20 years



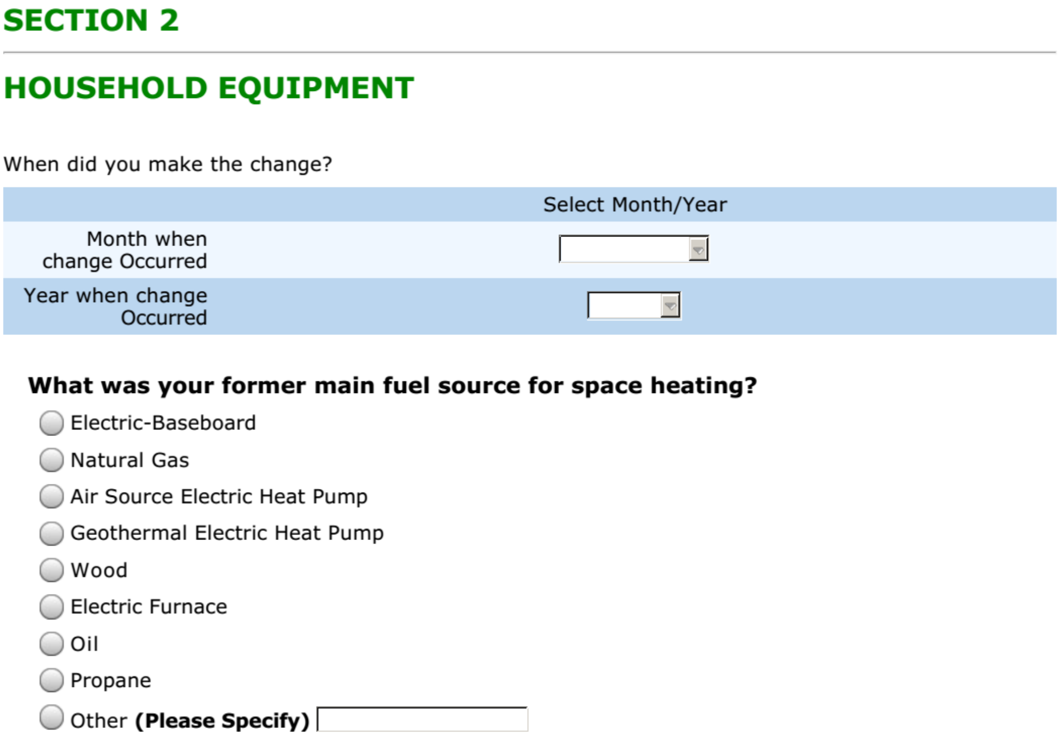
SECTION 2

HOUSEHOLD EQUIPMENT

In your current home, has your main space heating fuel source changed during the last 3 years?

○ Yes

○ No



SECTION 2

HOUSEHOLD EQUIPMENT

When did you make the change?

	Select Month/Year
Month when change Occurred	☐
Year when change Occurred	☐

What was your former main fuel source for space heating?

○ Electric-Baseboard

○ Natural Gas

○ Air Source Electric Heat Pump

○ Geothermal Electric Heat Pump

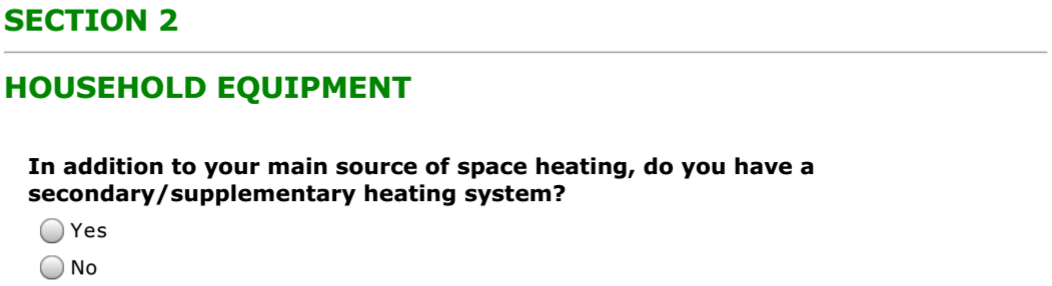
○ Wood

○ Electric Furnace

○ Oil

○ Propane

○ Other **(Please Specify)** ☐



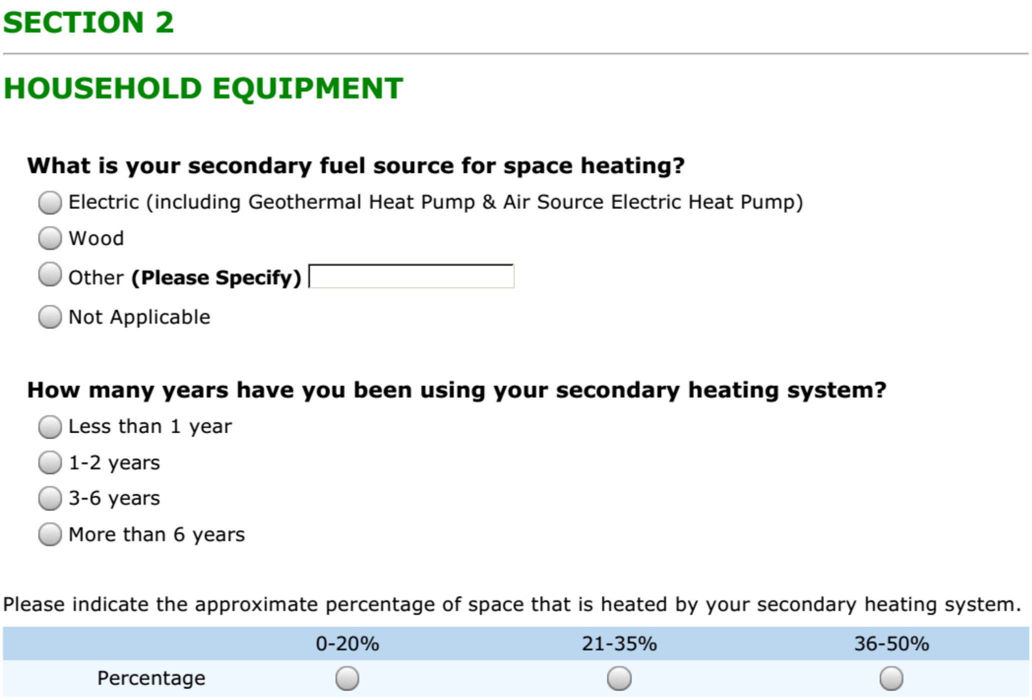
SECTION 2

HOUSEHOLD EQUIPMENT

In addition to your main source of space heating, do you have a secondary/supplementary heating system?

○ Yes

○ No



SECTION 2

HOUSEHOLD EQUIPMENT

What is your secondary fuel source for space heating?

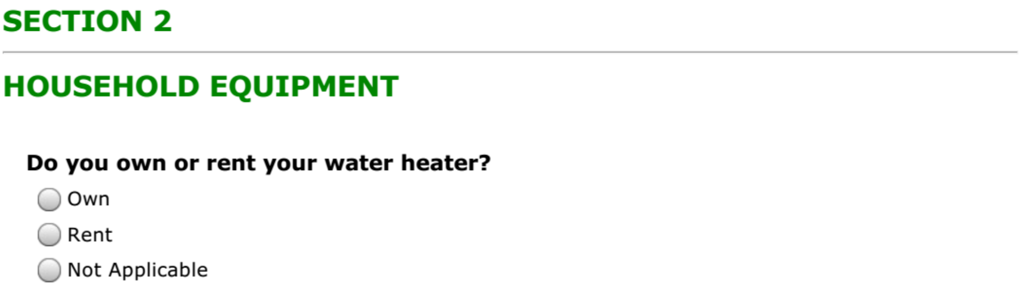
- ○ Electric (including Geothermal Heat Pump & Air Source Electric Heat Pump)
- ○ Wood
- ○ Other **(Please Specify)** ☐
- ○ Not Applicable

How many years have you been using your secondary heating system?

- ○ Less than 1 year
- ○ 1-2 years
- ○ 3-6 years
- ○ More than 6 years

Please indicate the approximate percentage of space that is heated by your secondary heating system.

	0-20%	21-35%	36-50%
Percentage	○	○	○

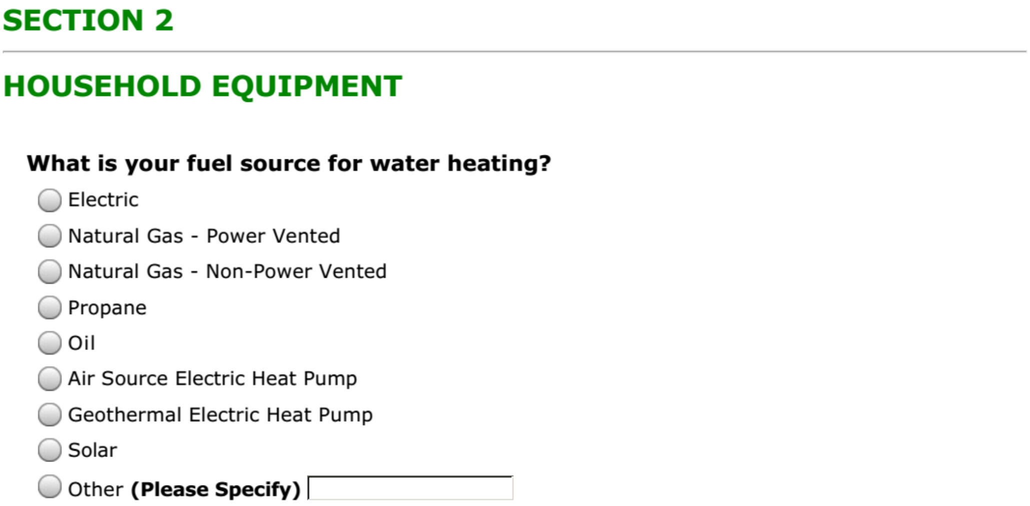


SECTION 2

HOUSEHOLD EQUIPMENT

Do you own or rent your water heater?

- ○ Own
- ○ Rent
- ○ Not Applicable

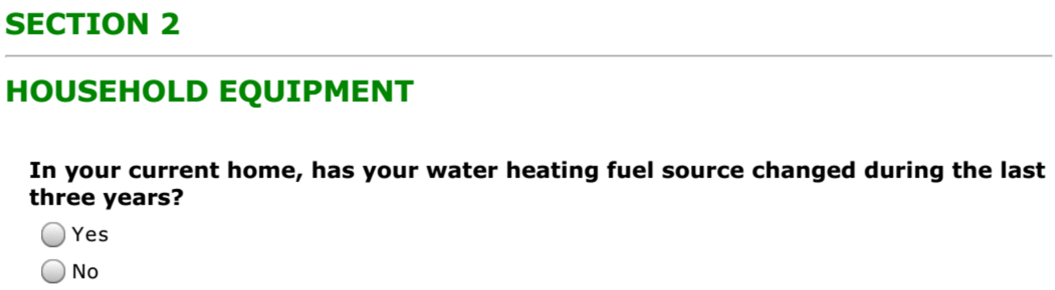


SECTION 2

HOUSEHOLD EQUIPMENT

What is your fuel source for water heating?

- ○ Electric
- ○ Natural Gas - Power Vented
- ○ Natural Gas - Non-Power Vented
- ○ Propane
- ○ Oil
- ○ Air Source Electric Heat Pump
- ○ Geothermal Electric Heat Pump
- ○ Solar
- ○ Other **(Please Specify)** ☐

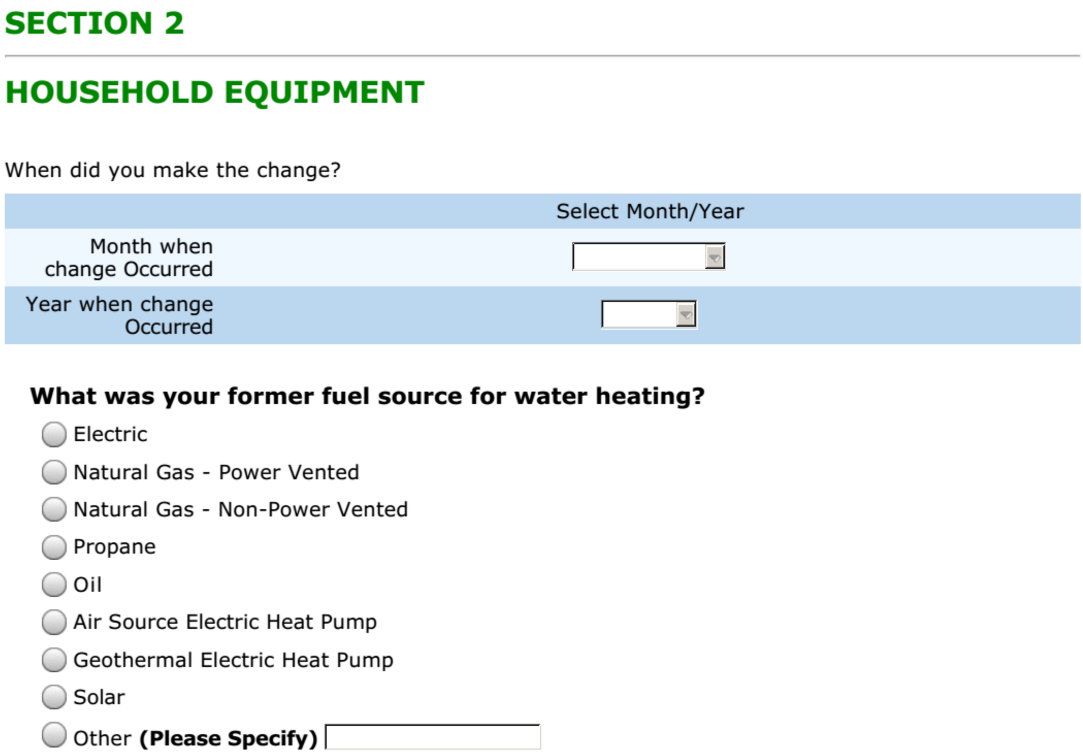


SECTION 2

HOUSEHOLD EQUIPMENT

In your current home, has your water heating fuel source changed during the last three years?

○ Yes
○ No



SECTION 2

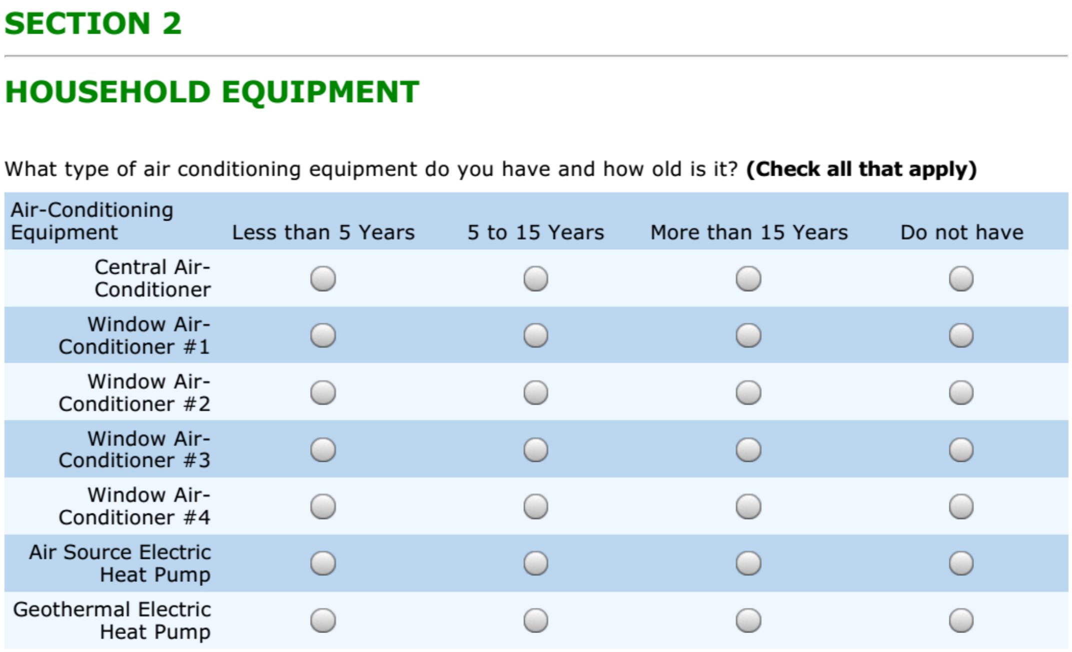
HOUSEHOLD EQUIPMENT

When did you make the change?

	Select Month/Year
Month when change Occurred	
Year when change Occurred	

What was your former fuel source for water heating?

○ Electric
○ Natural Gas - Power Vented
○ Natural Gas - Non-Power Vented
○ Propane
○ Oil
○ Air Source Electric Heat Pump
○ Geothermal Electric Heat Pump
○ Solar
○ Other **(Please Specify)**

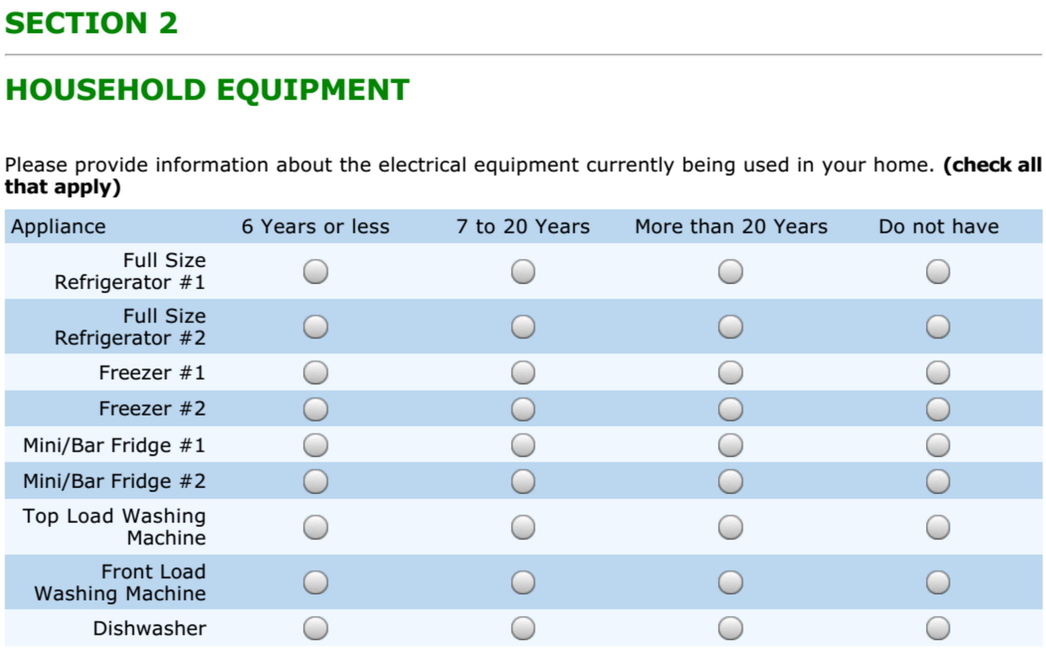


SECTION 2

HOUSEHOLD EQUIPMENT

What type of air conditioning equipment do you have and how old is it? **(Check all that apply)**

Air-Conditioning Equipment	Less than 5 Years	5 to 15 Years	More than 15 Years	Do not have
Central Air-Conditioner	○	○	○	○
Window Air-Conditioner #1	○	○	○	○
Window Air-Conditioner #2	○	○	○	○
Window Air-Conditioner #3	○	○	○	○
Window Air-Conditioner #4	○	○	○	○
Air Source Electric Heat Pump	○	○	○	○
Geothermal Electric Heat Pump	○	○	○	○

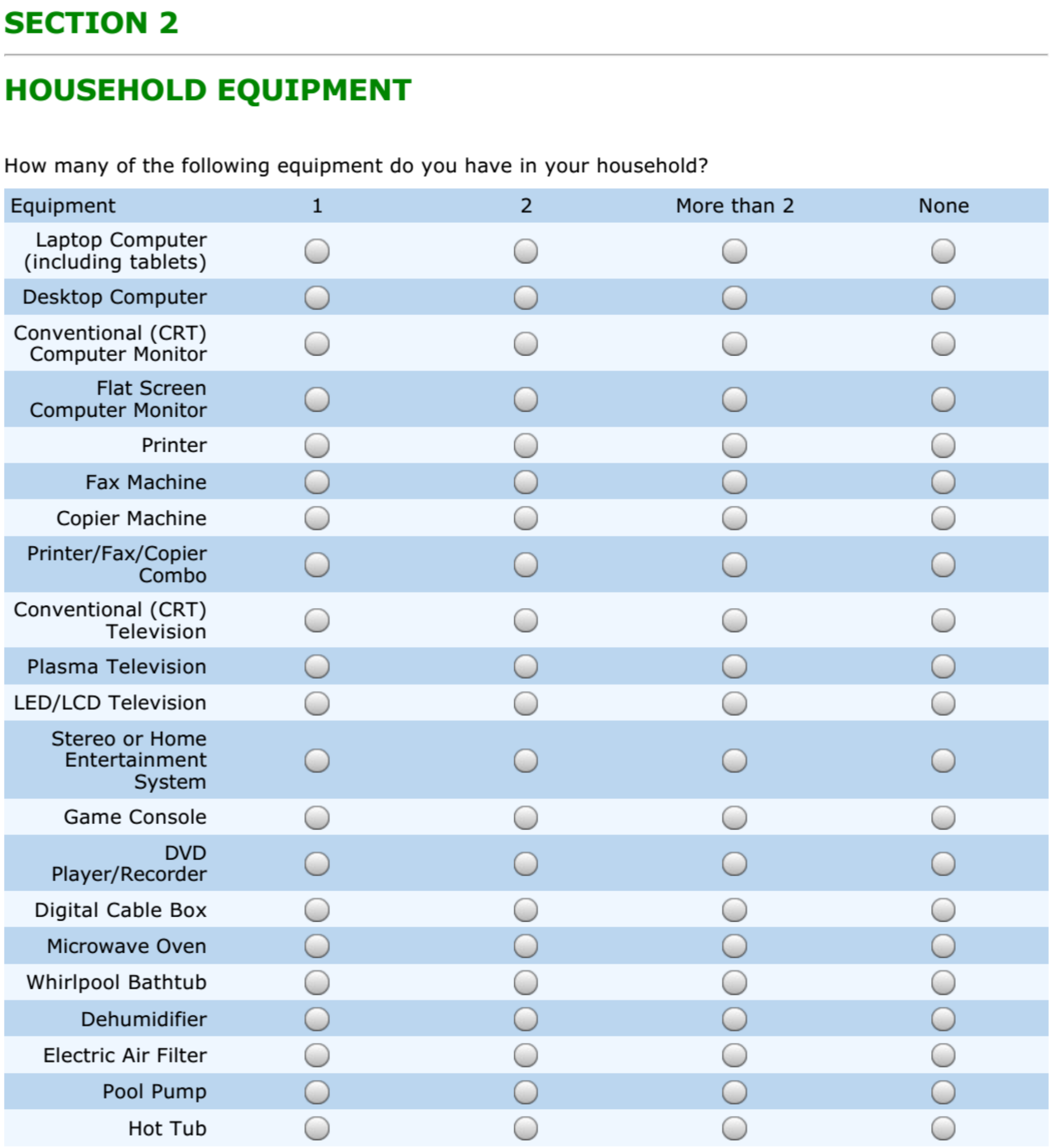


SECTION 2

HOUSEHOLD EQUIPMENT

Please provide information about the electrical equipment currently being used in your home. **(check all that apply)**

Appliance	6 Years or less	7 to 20 Years	More than 20 Years	Do not have
Full Size Refrigerator #1	○	○	○	○
Full Size Refrigerator #2	○	○	○	○
Freezer #1	○	○	○	○
Freezer #2	○	○	○	○
Mini/Bar Fridge #1	○	○	○	○
Mini/Bar Fridge #2	○	○	○	○
Top Load Washing Machine	○	○	○	○
Front Load Washing Machine	○	○	○	○
Dishwasher	○	○	○	○

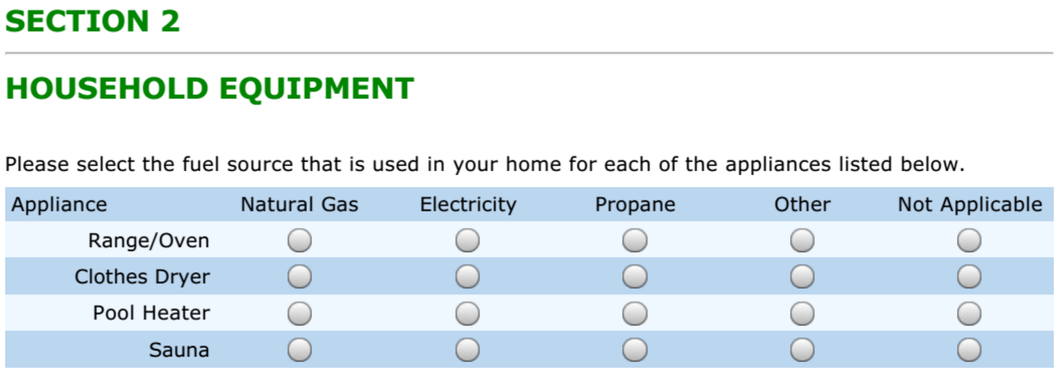


SECTION 2

HOUSEHOLD EQUIPMENT

How many of the following equipment do you have in your household?

Equipment	1	2	More than 2	None
Laptop Computer (including tablets)	○	○	○	○
Desktop Computer	○	○	○	○
Conventional (CRT) Computer Monitor	○	○	○	○
Flat Screen Computer Monitor	○	○	○	○
Printer	○	○	○	○
Fax Machine	○	○	○	○
Copier Machine	○	○	○	○
Printer/Fax/Copier Combo	○	○	○	○
Conventional (CRT) Television	○	○	○	○
Plasma Television	○	○	○	○
LED/LCD Television	○	○	○	○
Stereo or Home Entertainment System	○	○	○	○
Game Console	○	○	○	○
DVD Player/Recorder	○	○	○	○
Digital Cable Box	○	○	○	○
Microwave Oven	○	○	○	○
Whirlpool Bathtub	○	○	○	○
Dehumidifier	○	○	○	○
Electric Air Filter	○	○	○	○
Pool Pump	○	○	○	○
Hot Tub	○	○	○	○

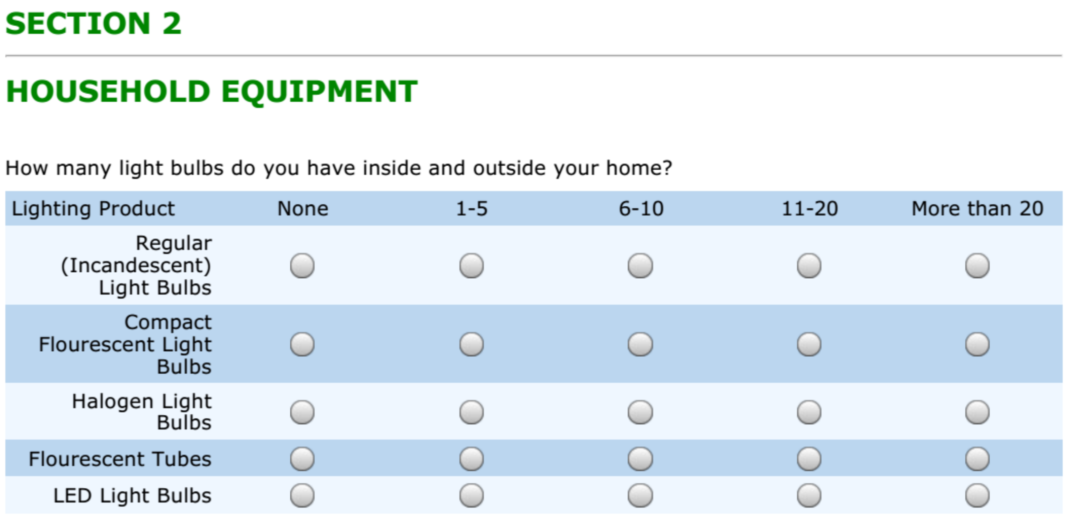


SECTION 2

HOUSEHOLD EQUIPMENT

Please select the fuel source that is used in your home for each of the appliances listed below.

Appliance	Natural Gas	Electricity	Propane	Other	Not Applicable
Range/Oven	○	○	○	○	○
Clothes Dryer	○	○	○	○	○
Pool Heater	○	○	○	○	○
Sauna	○	○	○	○	○

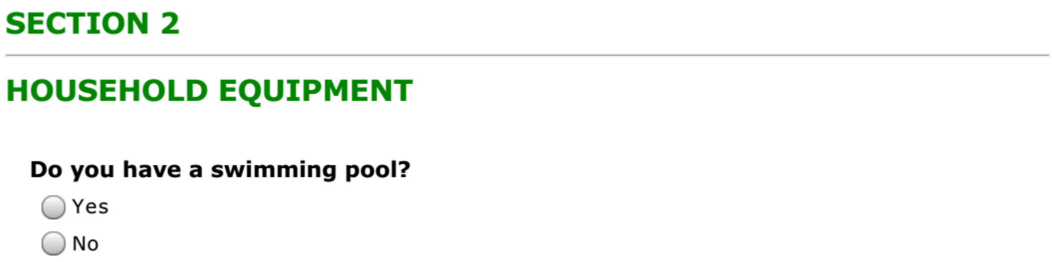


SECTION 2

HOUSEHOLD EQUIPMENT

How many light bulbs do you have inside and outside your home?

Lighting Product	None	1-5	6-10	11-20	More than 20
Regular (Incandescent) Light Bulbs	○	○	○	○	○
Compact Flourescent Light Bulbs	○	○	○	○	○
Halogen Light Bulbs	○	○	○	○	○
Flourescent Tubes	○	○	○	○	○
LED Light Bulbs	○	○	○	○	○

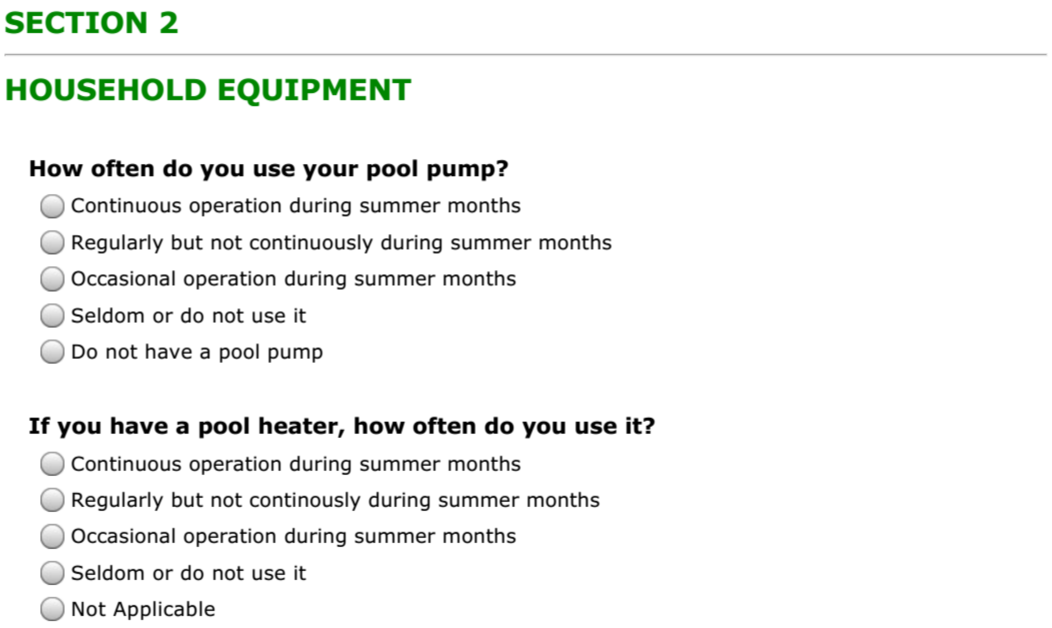


SECTION 2

HOUSEHOLD EQUIPMENT

Do you have a swimming pool?

○ Yes
○ No



SECTION 2

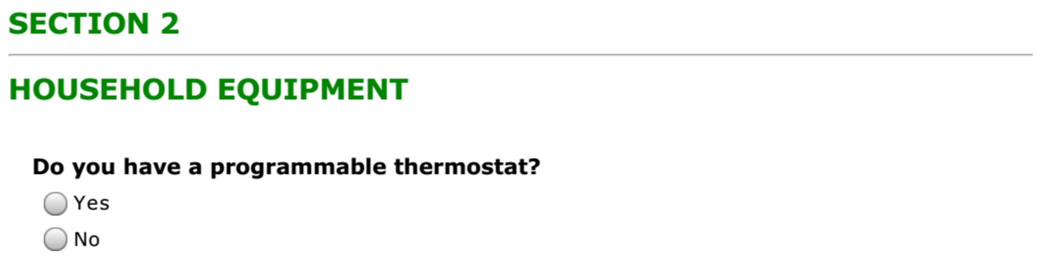
HOUSEHOLD EQUIPMENT

How often do you use your pool pump?

○ Continuous operation during summer months
○ Regularly but not continuously during summer months
○ Occasional operation during summer months
○ Seldom or do not use it
○ Do not have a pool pump

If you have a pool heater, how often do you use it?

○ Continuous operation during summer months
○ Regularly but not continously during summer months
○ Occasional operation during summer months
○ Seldom or do not use it
○ Not Applicable

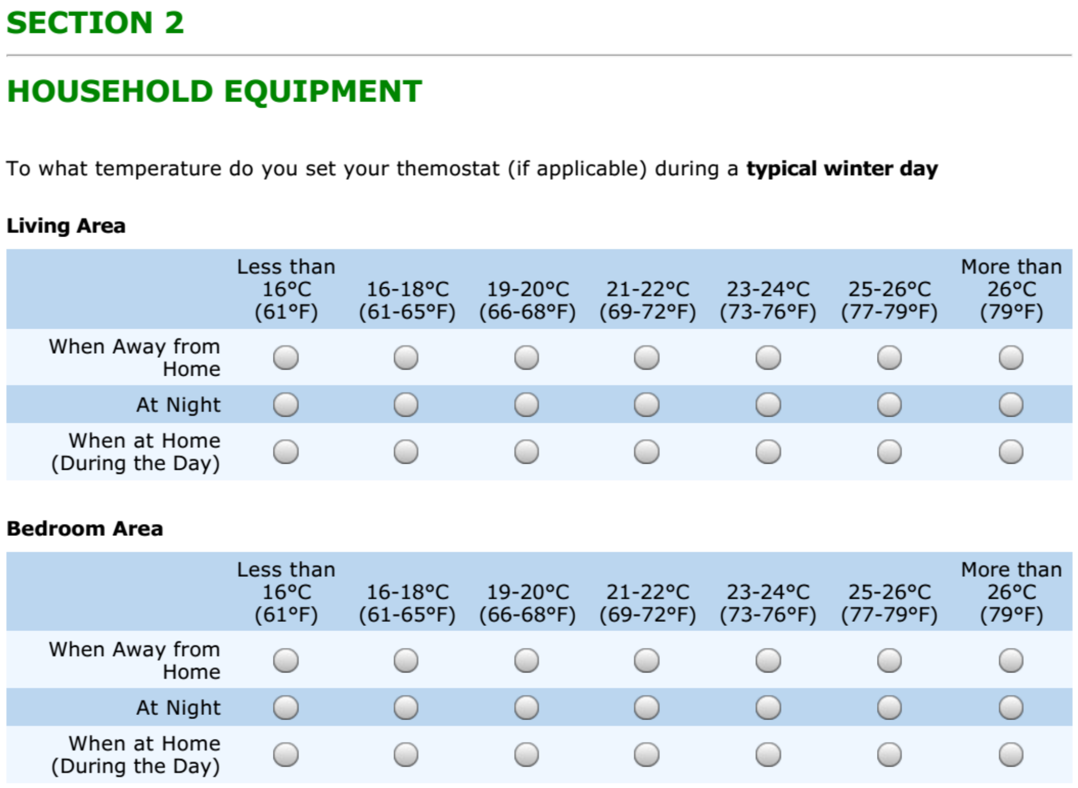


SECTION 2

HOUSEHOLD EQUIPMENT

Do you have a programmable thermostat?

○ Yes
○ No



SECTION 2

HOUSEHOLD EQUIPMENT

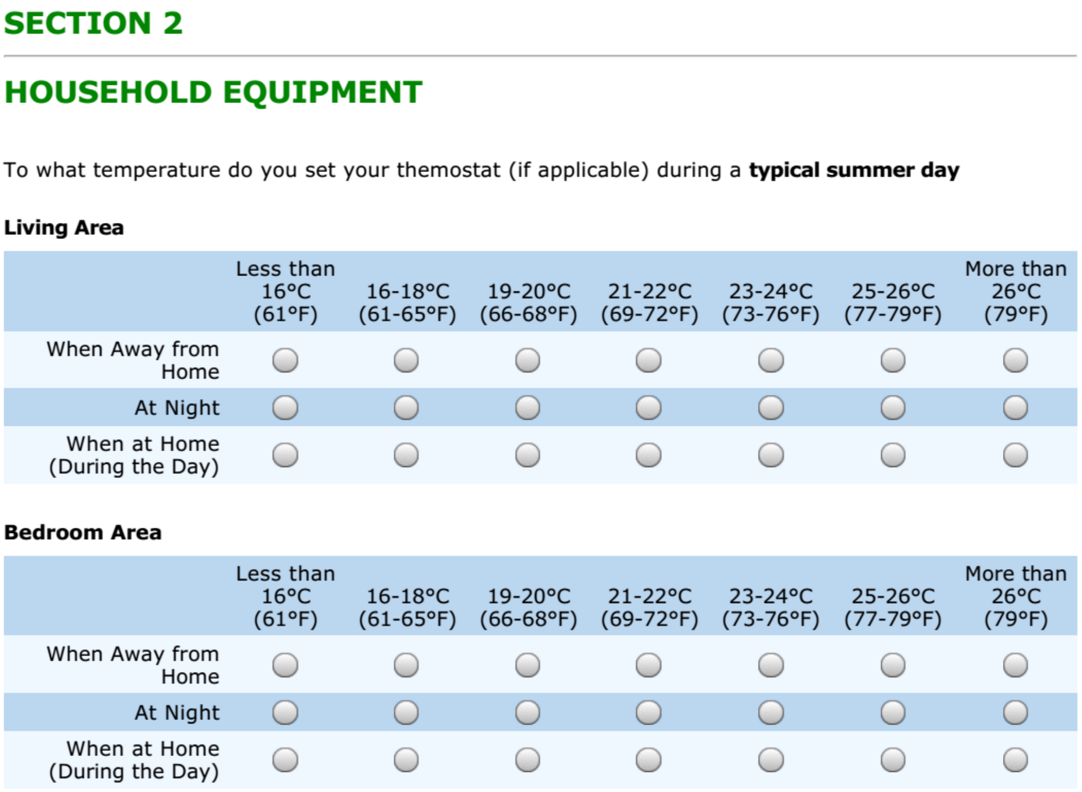
To what temperature do you set your themostat (if applicable) during a **typical winter day**

Living Area

	Less than 16°C (61°F)	16-18°C (61-65°F)	19-20°C (66-68°F)	21-22°C (69-72°F)	23-24°C (73-76°F)	25-26°C (77-79°F)	More than 26°C (79°F)
When Away from Home	○	○	○	○	○	○	○
At Night	○	○	○	○	○	○	○
When at Home (During the Day)	○	○	○	○	○	○	○

Bedroom Area

	Less than 16°C (61°F)	16-18°C (61-65°F)	19-20°C (66-68°F)	21-22°C (69-72°F)	23-24°C (73-76°F)	25-26°C (77-79°F)	More than 26°C (79°F)
When Away from Home	○	○	○	○	○	○	○
At Night	○	○	○	○	○	○	○
When at Home (During the Day)	○	○	○	○	○	○	○



SECTION 2

HOUSEHOLD EQUIPMENT

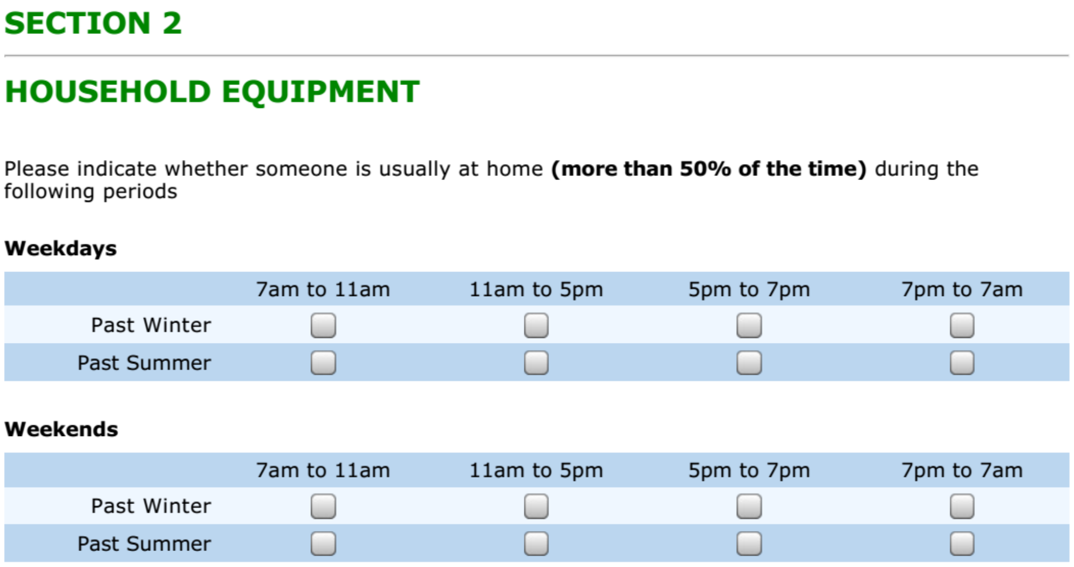
To what temperature do you set your themostat (if applicable) during a **typical summer day**

Living Area

	Less than 16°C (61°F)	16-18°C (61-65°F)	19-20°C (66-68°F)	21-22°C (69-72°F)	23-24°C (73-76°F)	25-26°C (77-79°F)	More than 26°C (79°F)
When Away from Home	○	○	○	○	○	○	○
At Night	○	○	○	○	○	○	○
When at Home (During the Day)	○	○	○	○	○	○	○

Bedroom Area

	Less than 16°C (61°F)	16-18°C (61-65°F)	19-20°C (66-68°F)	21-22°C (69-72°F)	23-24°C (73-76°F)	25-26°C (77-79°F)	More than 26°C (79°F)
When Away from Home	○	○	○	○	○	○	○
At Night	○	○	○	○	○	○	○
When at Home (During the Day)	○	○	○	○	○	○	○



SECTION 2

HOUSEHOLD EQUIPMENT

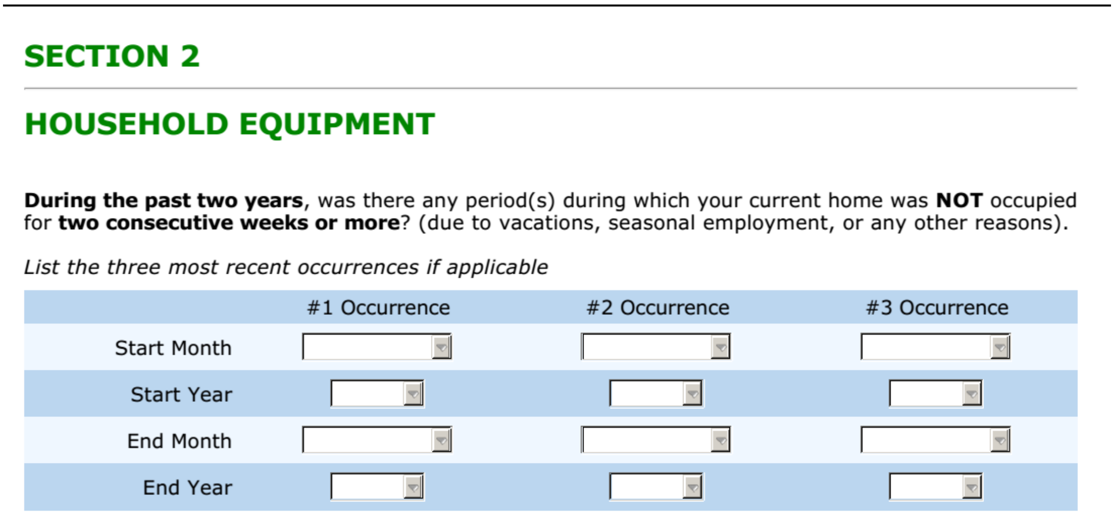
Please indicate whether someone is usually at home **(more than 50% of the time)** during the following periods

Weekdays

	7am to 11am	11am to 5pm	5pm to 7pm	7pm to 7am
Past Winter	☐	☐	☐	☐
Past Summer	☐	☐	☐	☐

Weekends

	7am to 11am	11am to 5pm	5pm to 7pm	7pm to 7am
Past Winter	☐	☐	☐	☐
Past Summer	☐	☐	☐	☐



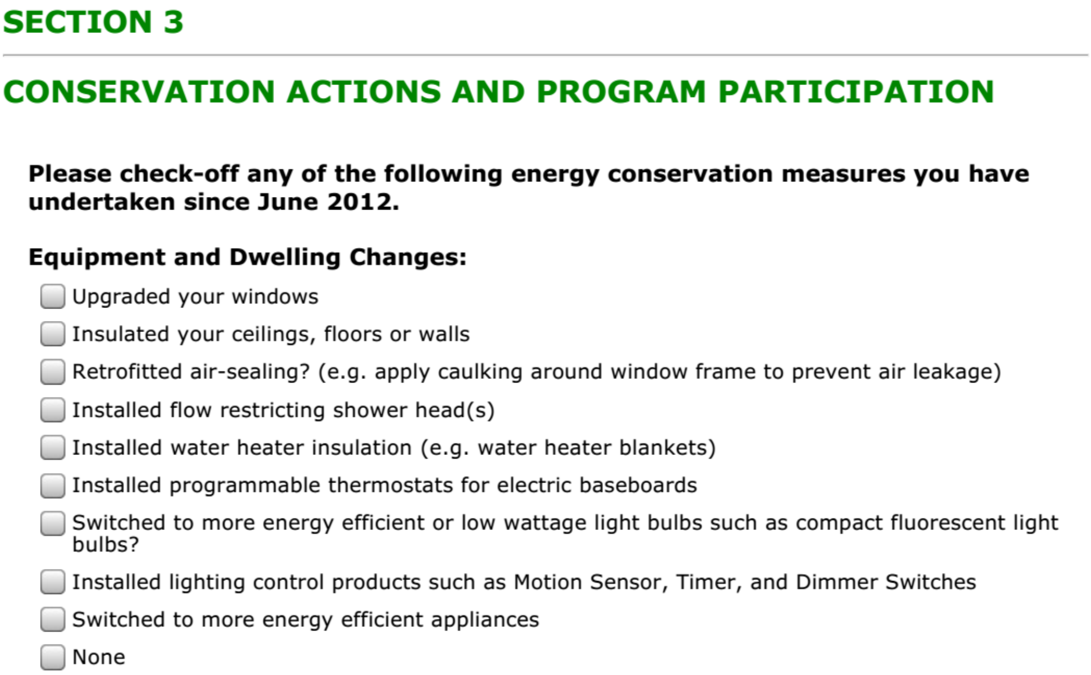
SECTION 2

HOUSEHOLD EQUIPMENT

During the past two years, was there any period(s) during which your current home was **NOT** occupied for **two consecutive weeks or more**? (due to vacations, seasonal employment, or any other reasons).

List the three most recent occurrences if applicable

	#1 Occurrence	#2 Occurrence	#3 Occurrence
Start Month			
Start Year			
End Month			
End Year			



SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

Please check-off any of the following energy conservation measures you have undertaken since June 2012.

Equipment and Dwelling Changes:

☐ Upgraded your windows

☐ Insulated your ceilings, floors or walls

☐ Retrofitted air-sealing? (e.g. apply caulking around window frame to prevent air leakage)

☐ Installed flow restricting shower head(s)

☐ Installed water heater insulation (e.g. water heater blankets)

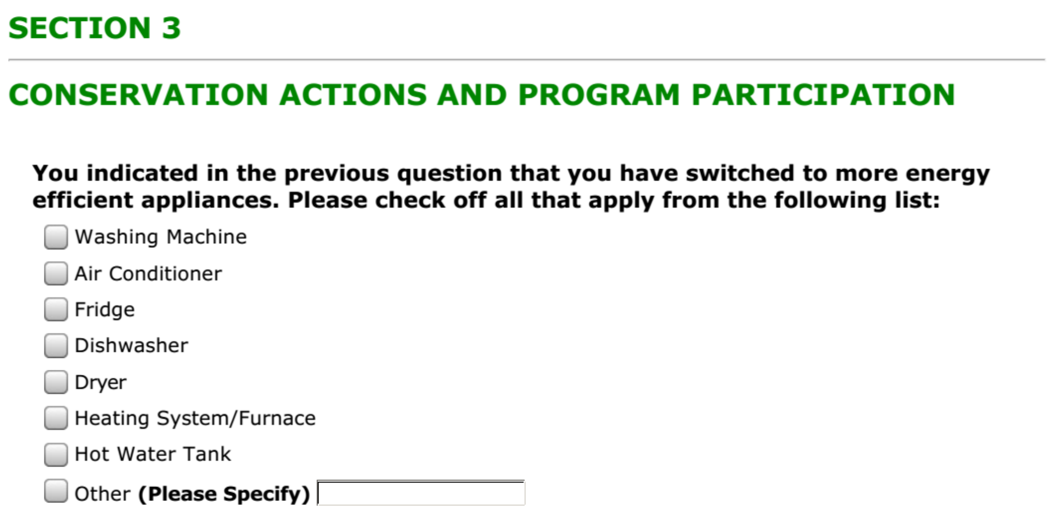
☐ Installed programmable thermostats for electric baseboards

☐ Switched to more energy efficient or low wattage light bulbs such as compact fluorescent light bulbs?

☐ Installed lighting control products such as Motion Sensor, Timer, and Dimmer Switches

☐ Switched to more energy efficient appliances

☐ None

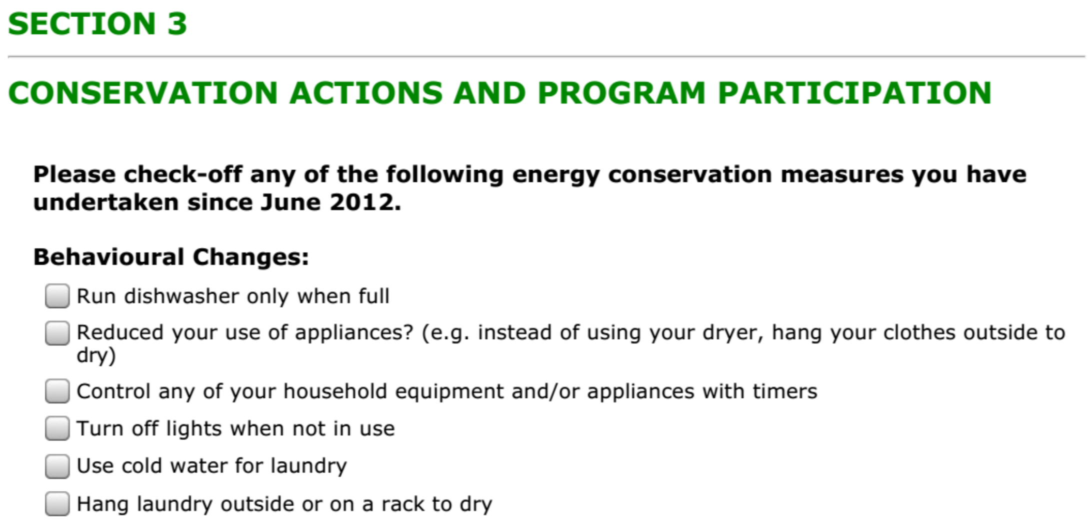


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

You indicated in the previous question that you have switched to more energy efficient appliances. Please check off all that apply from the following list:

☐ Washing Machine
☐ Air Conditioner
☐ Fridge
☐ Dishwasher
☐ Dryer
☐ Heating System/Furnace
☐ Hot Water Tank
☐ Other **(Please Specify)** ______



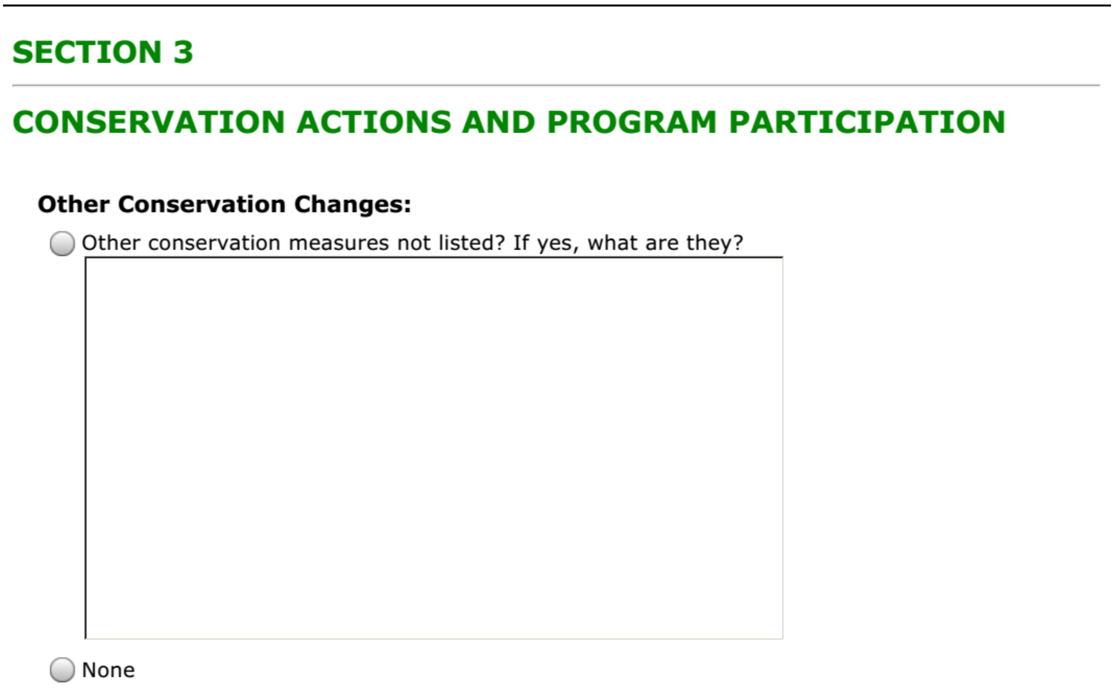
SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

Please check-off any of the following energy conservation measures you have undertaken since June 2012.

Behavioural Changes:

☐ Run dishwasher only when full
☐ Reduced your use of appliances? (e.g. instead of using your dryer, hang your clothes outside to dry)
☐ Control any of your household equipment and/or appliances with timers
☐ Turn off lights when not in use
☐ Use cold water for laundry
☐ Hang laundry outside or on a rack to dry



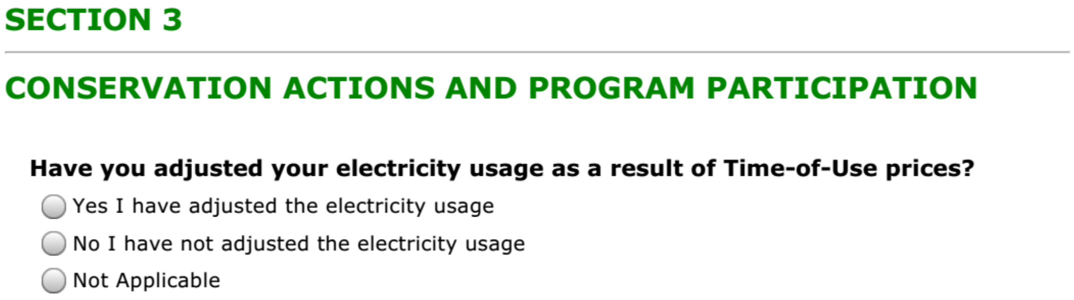
SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

Other Conservation Changes:

○ Other conservation measures not listed? If yes, what are they?

○ None



SECTION 3

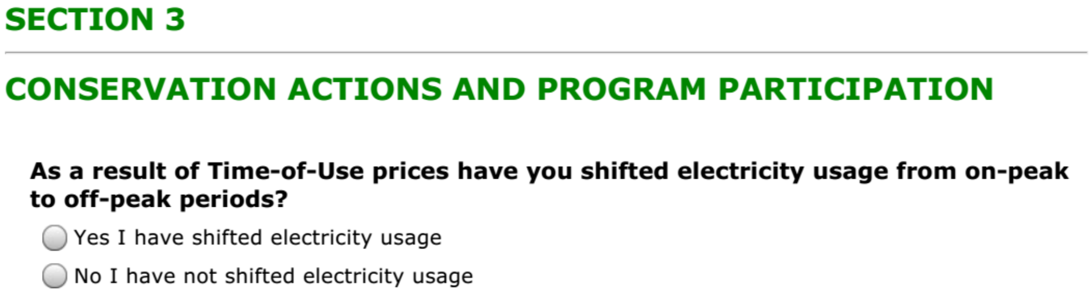
CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

Have you adjusted your electricity usage as a result of Time-of-Use prices?

○ Yes I have adjusted the electricity usage

○ No I have not adjusted the electricity usage

○ Not Applicable



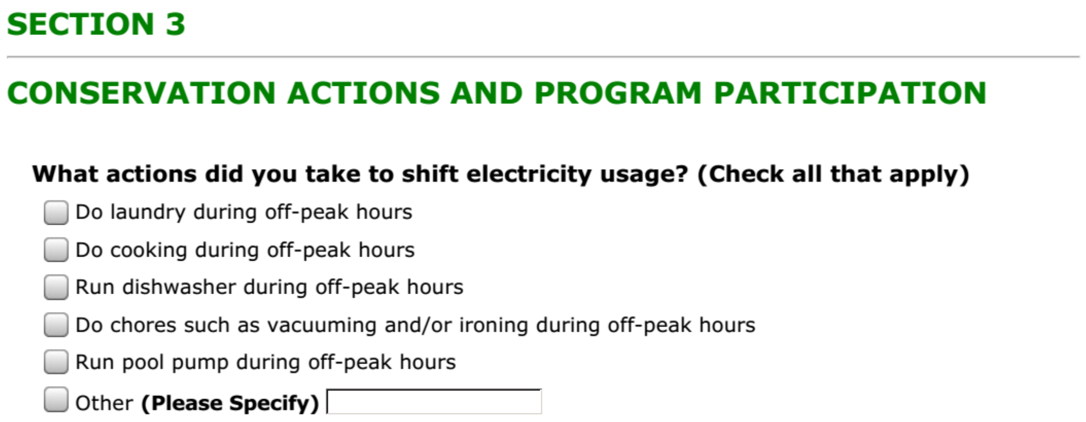
SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

As a result of Time-of-Use prices have you shifted electricity usage from on-peak to off-peak periods?

○ Yes I have shifted electricity usage

○ No I have not shifted electricity usage

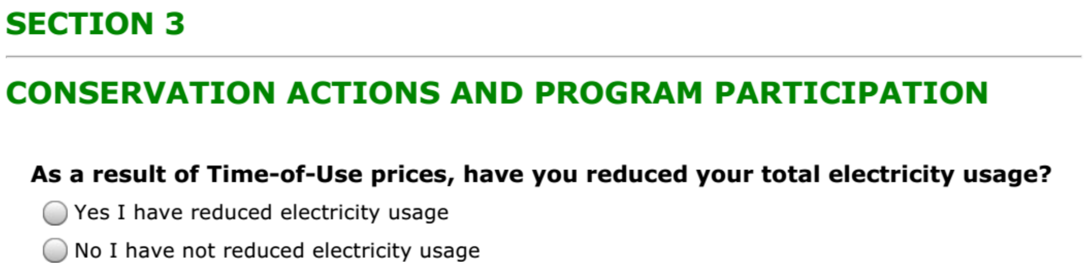


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

What actions did you take to shift electricity usage? (Check all that apply)

☐ Do laundry during off-peak hours
☐ Do cooking during off-peak hours
☐ Run dishwasher during off-peak hours
☐ Do chores such as vacuuming and/or ironing during off-peak hours
☐ Run pool pump during off-peak hours
☐ Other **(Please Specify)** ______

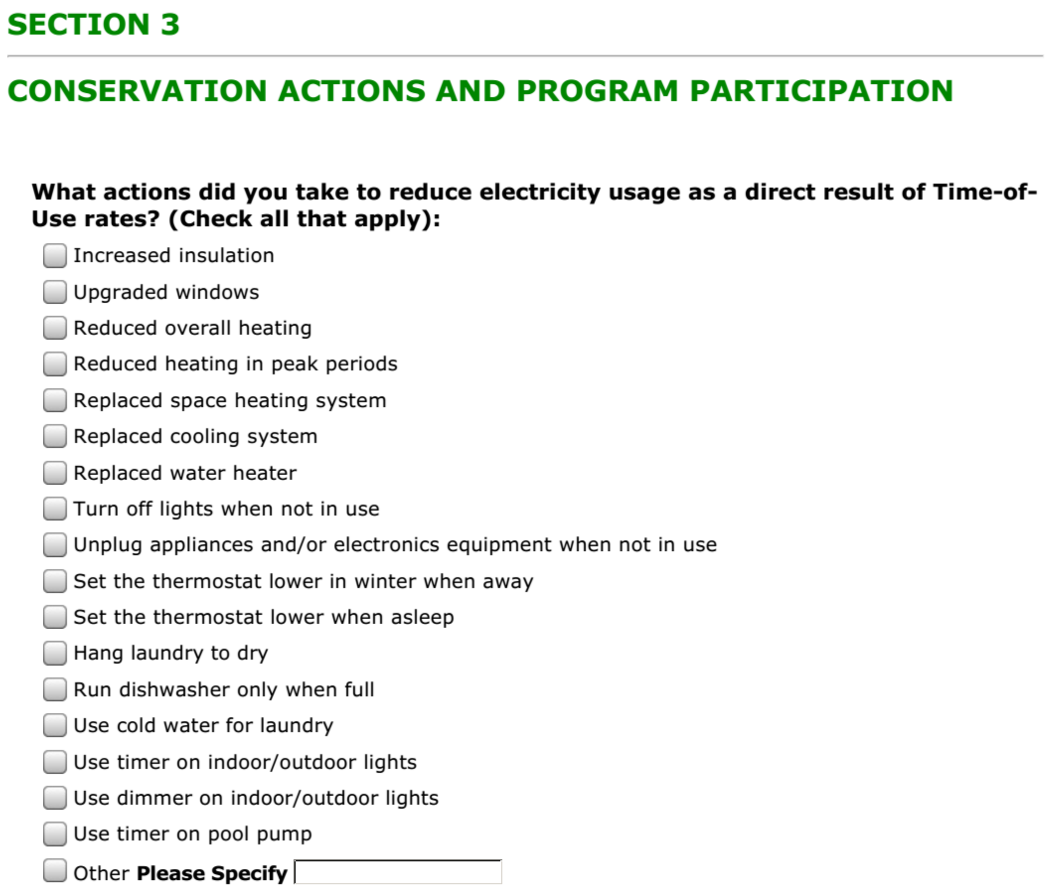


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

As a result of Time-of-Use prices, have you reduced your total electricity usage?

○ Yes I have reduced electricity usage
○ No I have not reduced electricity usage

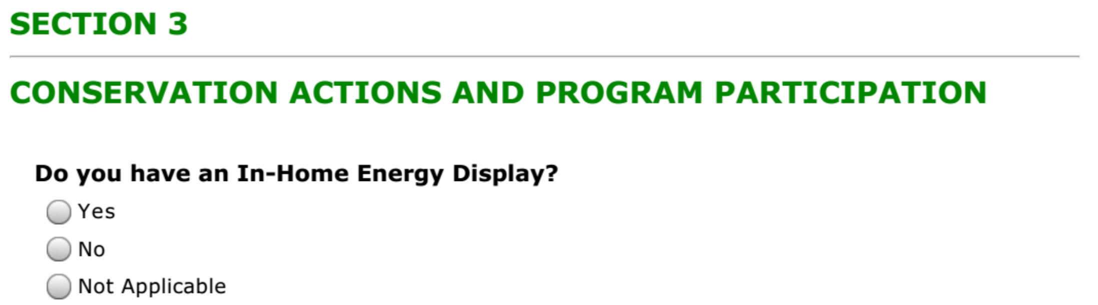


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

What actions did you take to reduce electricity usage as a direct result of Time-of-Use rates? (Check all that apply):

☐ Increased insulation
☐ Upgraded windows
☐ Reduced overall heating
☐ Reduced heating in peak periods
☐ Replaced space heating system
☐ Replaced cooling system
☐ Replaced water heater
☐ Turn off lights when not in use
☐ Unplug appliances and/or electronics equipment when not in use
☐ Set the thermostat lower in winter when away
☐ Set the thermostat lower when asleep
☐ Hang laundry to dry
☐ Run dishwasher only when full
☐ Use cold water for laundry
☐ Use timer on indoor/outdoor lights
☐ Use dimmer on indoor/outdoor lights
☐ Use timer on pool pump
☐ Other **Please Specify** ______

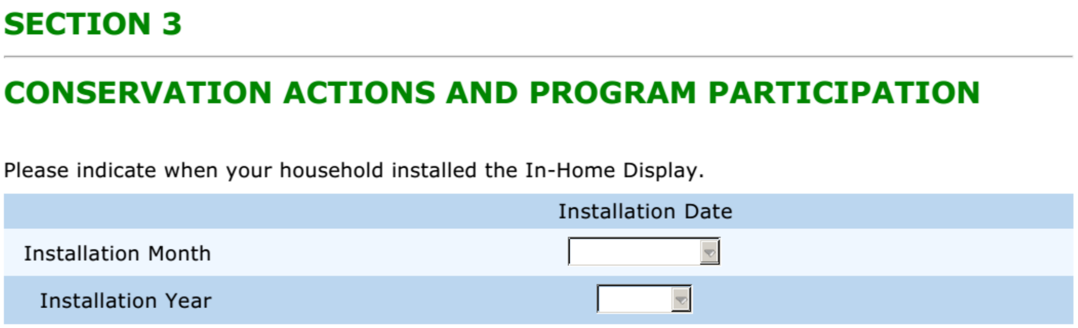


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

Do you have an In-Home Energy Display?

○ Yes
○ No
○ Not Applicable

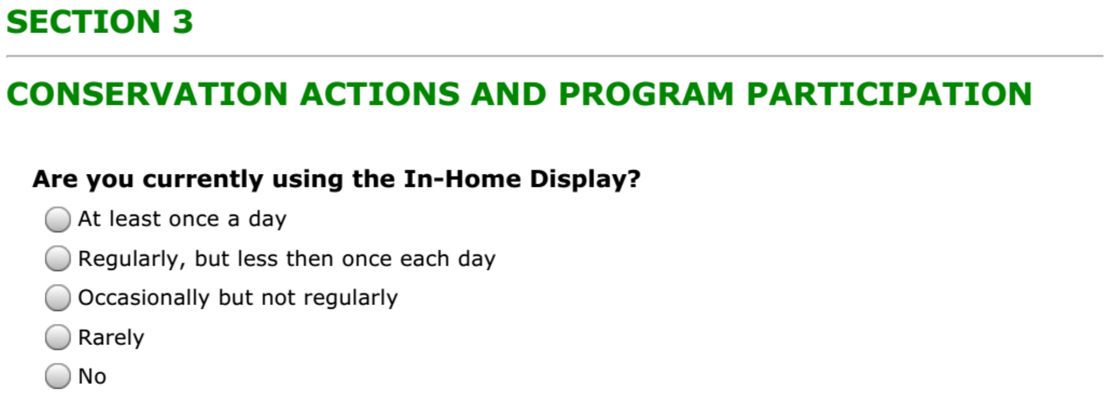


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

Please indicate when your household installed the In-Home Display.

	Installation Date
Installation Month	
Installation Year	

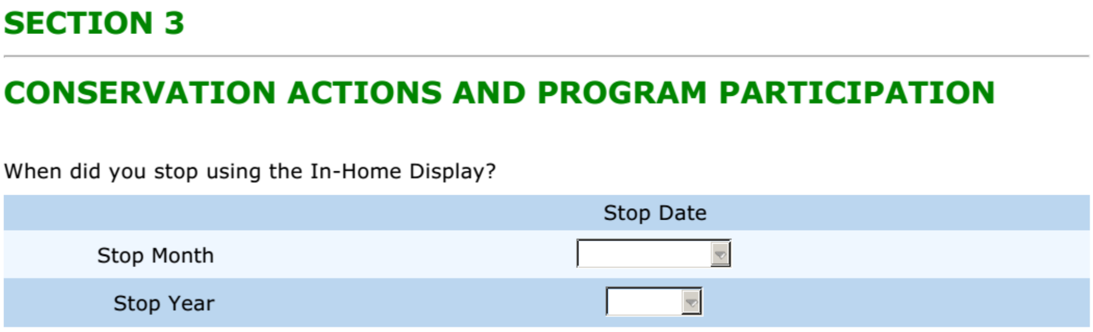


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

Are you currently using the In-Home Display?

○ At least once a day
○ Regularly, but less then once each day
○ Occasionally but not regularly
○ Rarely
○ No

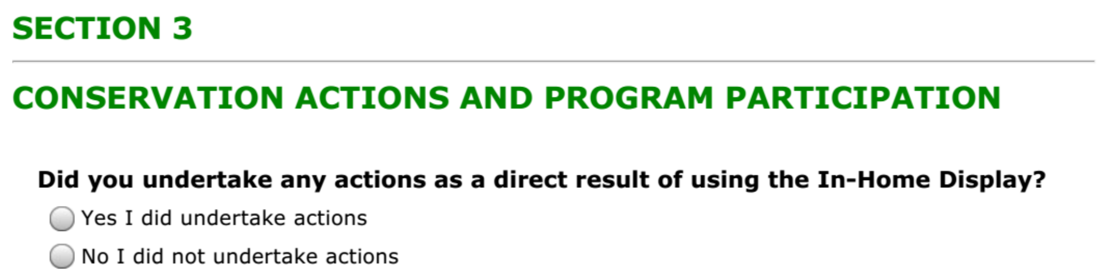


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

When did you stop using the In-Home Display?

	Stop Date
Stop Month	
Stop Year	



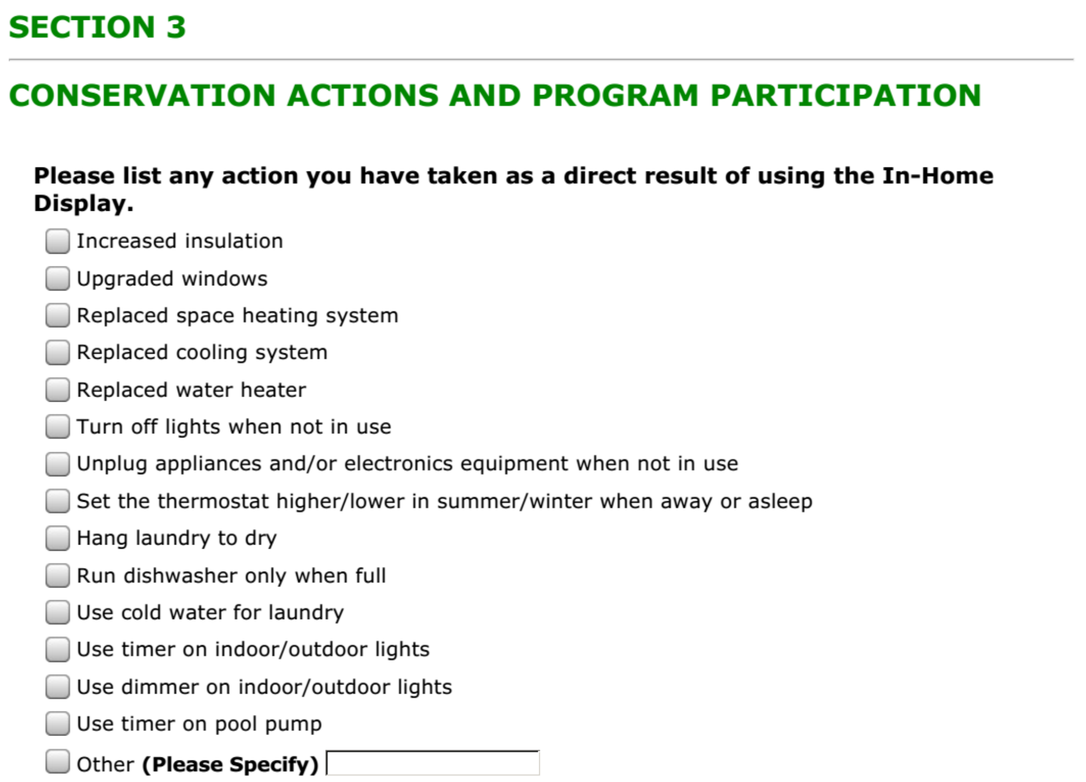
SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

Did you undertake any actions as a direct result of using the In-Home Display?

○ Yes I did undertake actions

○ No I did not undertake actions

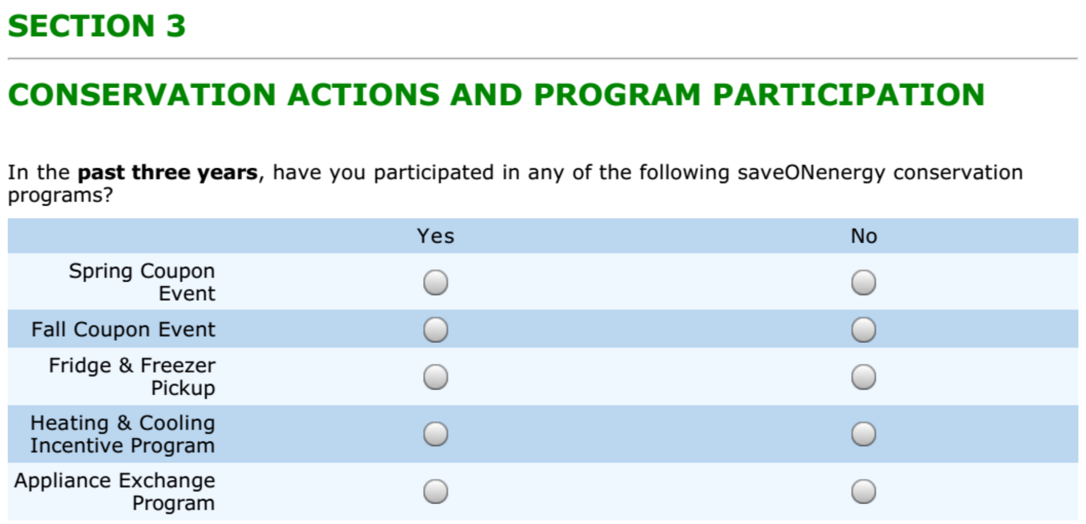


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

Please list any action you have taken as a direct result of using the In-Home Display.

- ☐ Increased insulation
- ☐ Upgraded windows
- ☐ Replaced space heating system
- ☐ Replaced cooling system
- ☐ Replaced water heater
- ☐ Turn off lights when not in use
- ☐ Unplug appliances and/or electronics equipment when not in use
- ☐ Set the thermostat higher/lower in summer/winter when away or asleep
- ☐ Hang laundry to dry
- ☐ Run dishwasher only when full
- ☐ Use cold water for laundry
- ☐ Use timer on indoor/outdoor lights
- ☐ Use dimmer on indoor/outdoor lights
- ☐ Use timer on pool pump
- ☐ Other **(Please Specify)** ____________

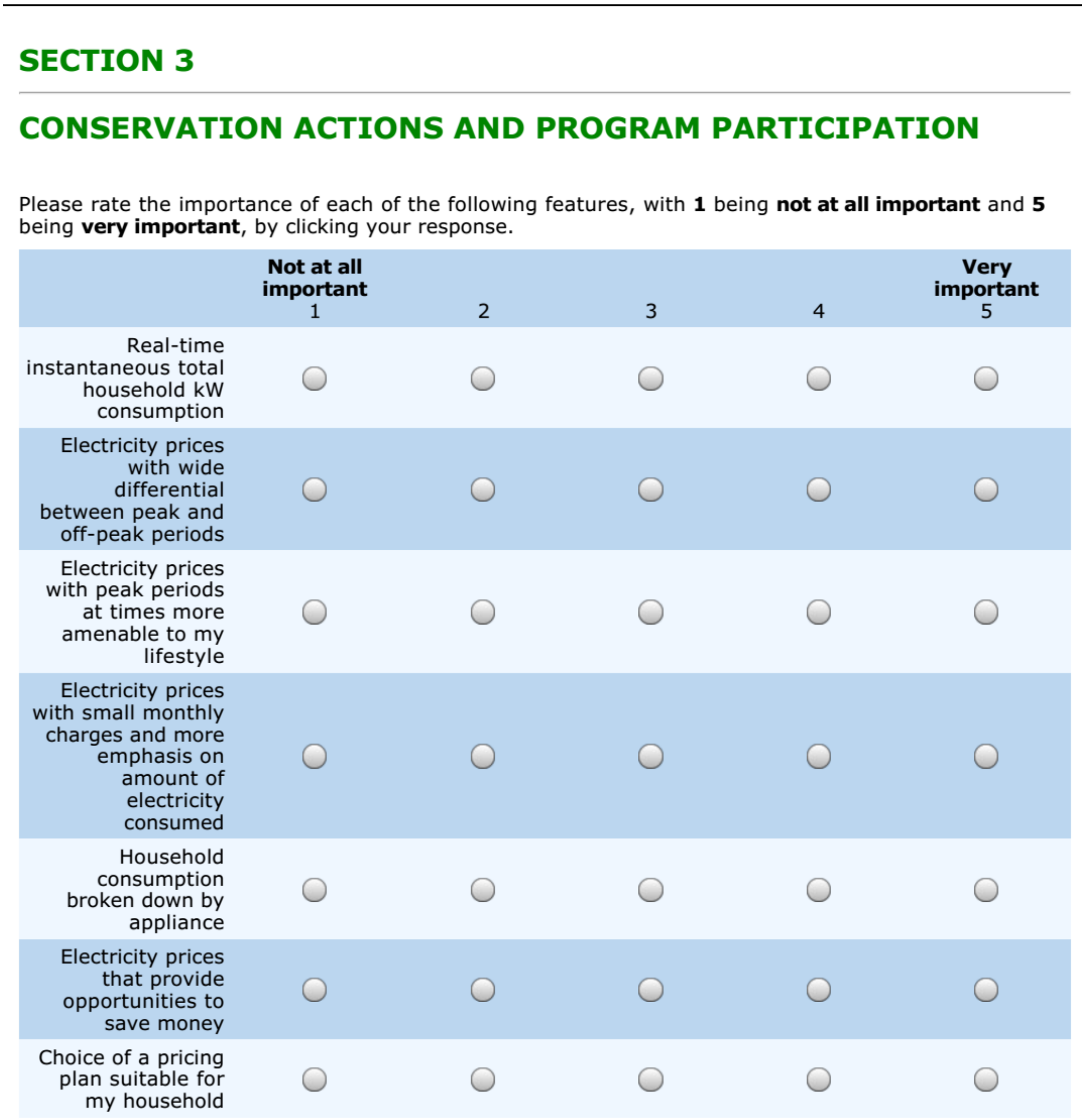


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

In the **past three years**, have you participated in any of the following saveONenergy conservation programs?

	Yes	No
Spring Coupon Event	○	○
Fall Coupon Event	○	○
Fridge & Freezer Pickup	○	○
Heating & Cooling Incentive Program	○	○
Appliance Exchange Program	○	○

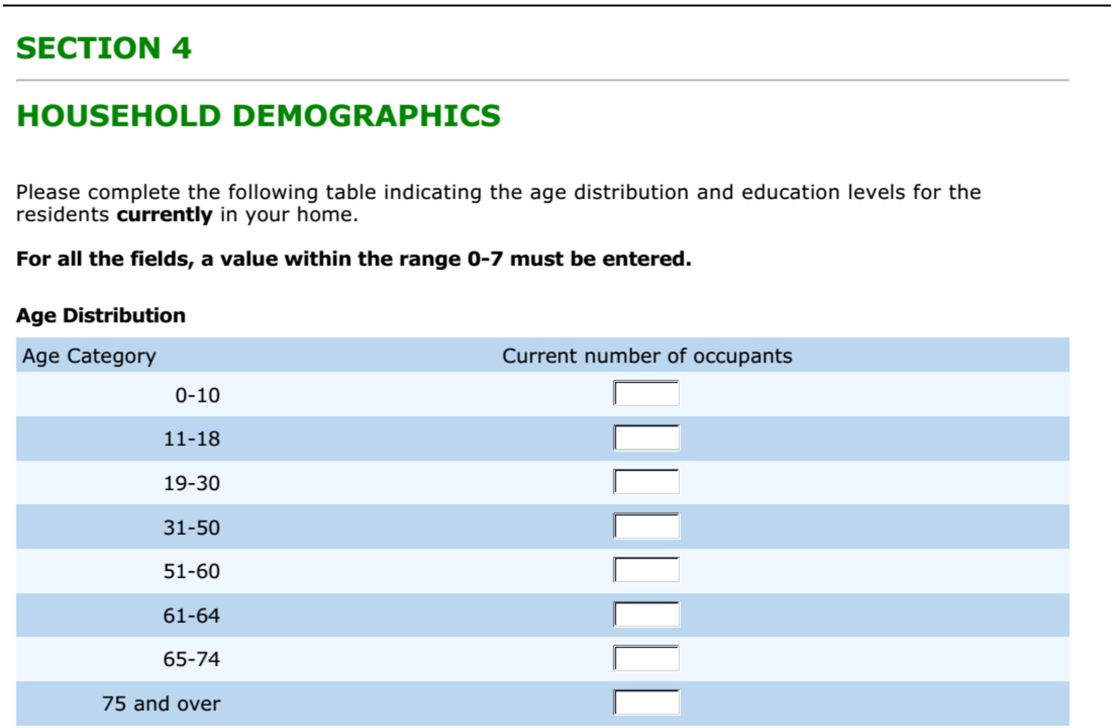


SECTION 3

CONSERVATION ACTIONS AND PROGRAM PARTICIPATION

Please rate the importance of each of the following features, with **1** being **not at all important** and **5** being **very important**, by clicking your response.

	Not at all important 1	2	3	4	**Very important** 5
Real-time instantaneous total household kW consumption	○	○	○	○	○
Electricity prices with wide differential between peak and off-peak periods	○	○	○	○	○
Electricity prices with peak periods at times more amenable to my lifestyle	○	○	○	○	○
Electricity prices with small monthly charges and more emphasis on amount of electricity consumed	○	○	○	○	○
Household consumption broken down by appliance	○	○	○	○	○
Electricity prices that provide opportunities to save money	○	○	○	○	○
Choice of a pricing plan suitable for my household	○	○	○	○	○



SECTION 4

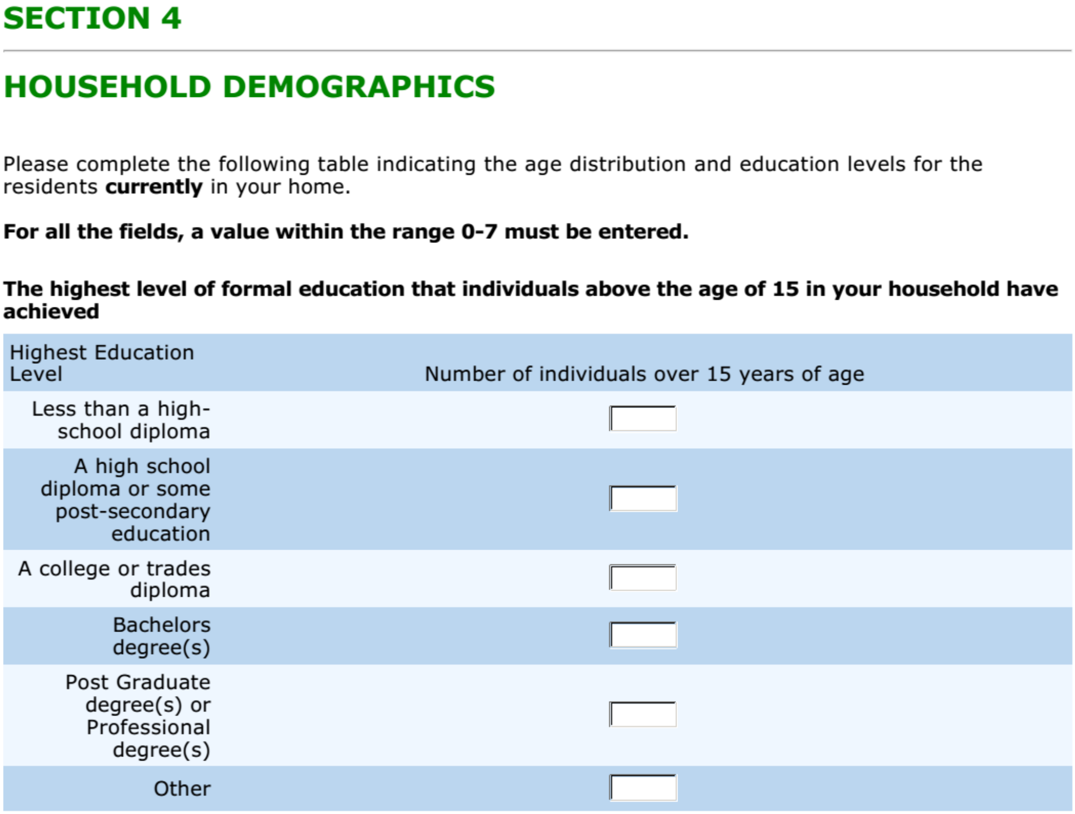
HOUSEHOLD DEMOGRAPHICS

Please complete the following table indicating the age distribution and education levels for the residents **currently** in your home.

For all the fields, a value within the range 0-7 must be entered.

Age Distribution

Age Category	Current number of occupants
0-10	
11-18	
19-30	
31-50	
51-60	
61-64	
65-74	
75 and over	



SECTION 4

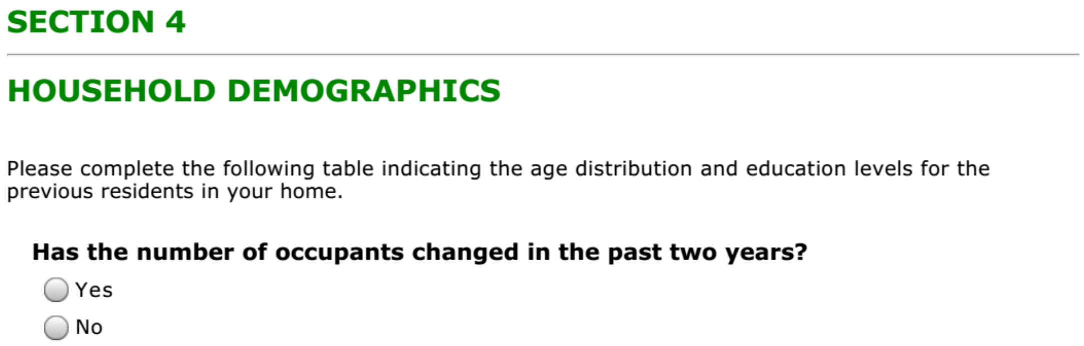
HOUSEHOLD DEMOGRAPHICS

Please complete the following table indicating the age distribution and education levels for the residents **currently** in your home.

For all the fields, a value within the range 0-7 must be entered.

The highest level of formal education that individuals above the age of 15 in your household have achieved

Highest Education Level	Number of individuals over 15 years of age
Less than a high-school diploma	
A high school diploma or some post-secondary education	
A college or trades diploma	
Bachelors degree(s)	
Post Graduate degree(s) or Professional degree(s)	
Other	



SECTION 4

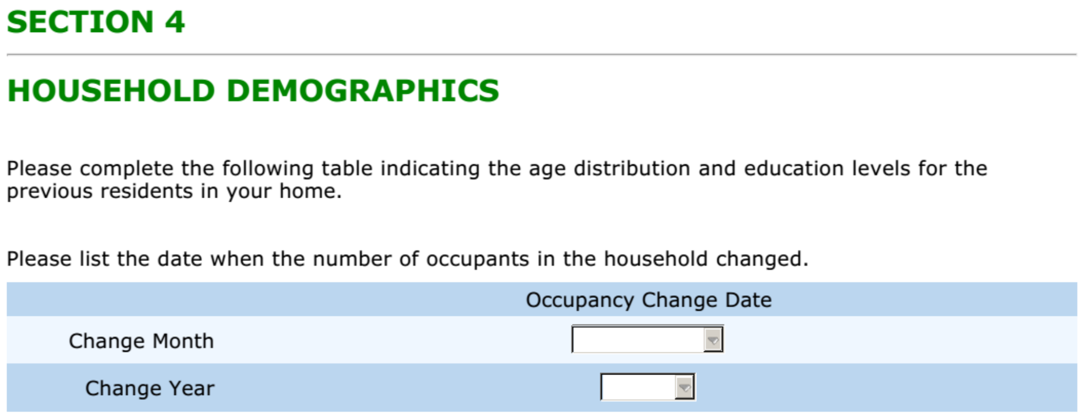
HOUSEHOLD DEMOGRAPHICS

Please complete the following table indicating the age distribution and education levels for the previous residents in your home.

Has the number of occupants changed in the past two years?

○ Yes

○ No



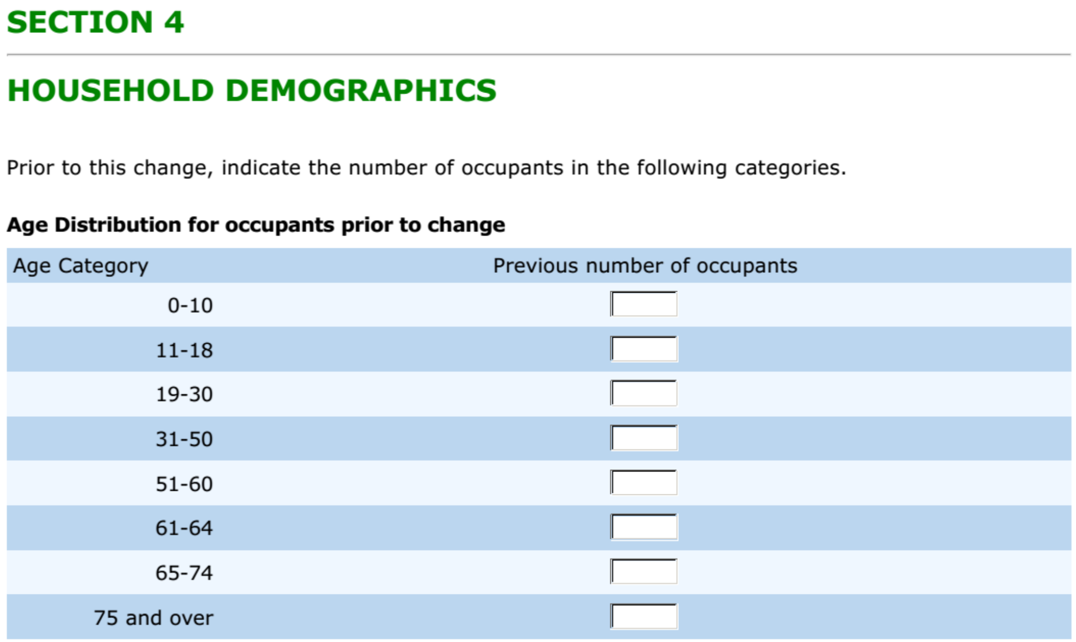
SECTION 4

HOUSEHOLD DEMOGRAPHICS

Please complete the following table indicating the age distribution and education levels for the previous residents in your home.

Please list the date when the number of occupants in the household changed.

	Occupancy Change Date
Change Month	
Change Year	



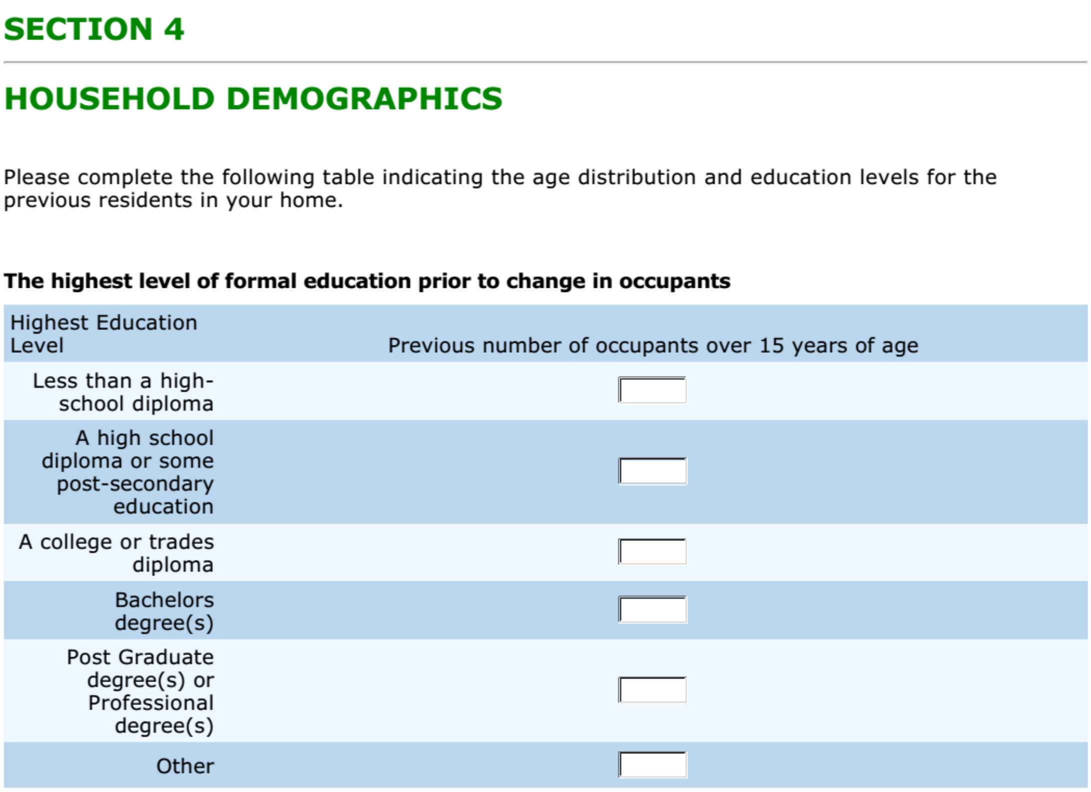
SECTION 4

HOUSEHOLD DEMOGRAPHICS

Prior to this change, indicate the number of occupants in the following categories.

Age Distribution for occupants prior to change

Age Category	Previous number of occupants
0-10	
11-18	
19-30	
31-50	
51-60	
61-64	
65-74	
75 and over	



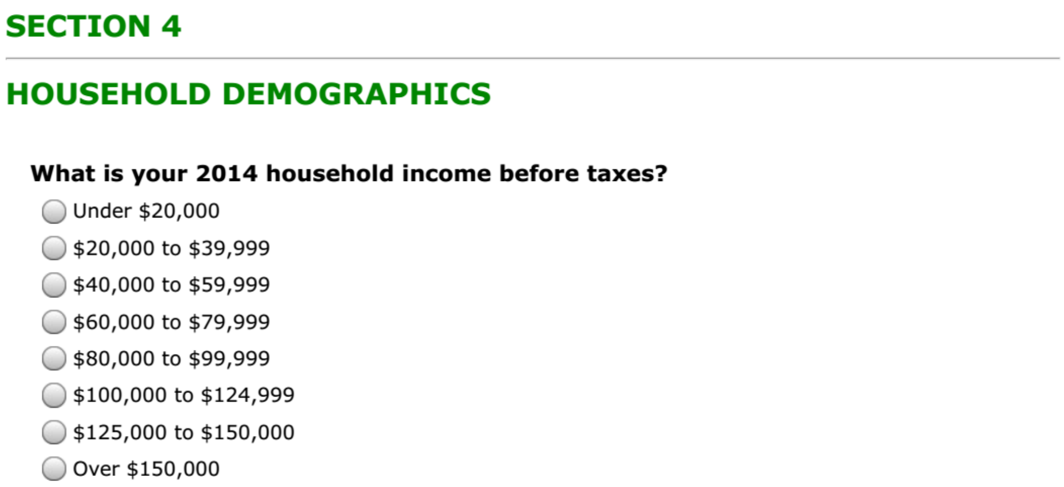
SECTION 4

HOUSEHOLD DEMOGRAPHICS

Please complete the following table indicating the age distribution and education levels for the previous residents in your home.

The highest level of formal education prior to change in occupants

Highest Education Level	Previous number of occupants over 15 years of age
Less than a high-school diploma	
A high school diploma or some post-secondary education	
A college or trades diploma	
Bachelors degree(s)	
Post Graduate degree(s) or Professional degree(s)	
Other	



SECTION 4

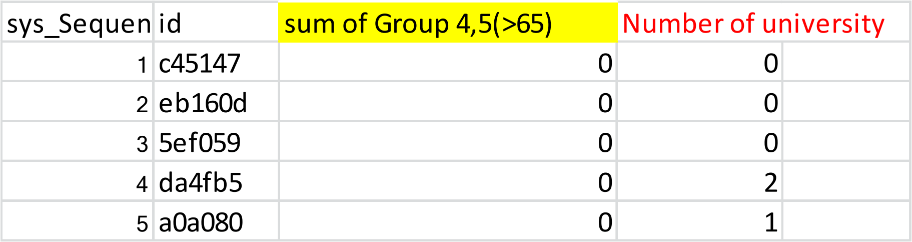
HOUSEHOLD DEMOGRAPHICS

What is your 2014 household income before taxes?

- ○ Under $20,000
- ○ $20,000 to $39,999
- ○ $40,000 to $59,999
- ○ $60,000 to $79,999
- ○ $80,000 to $99,999
- ○ $100,000 to $124,999
- ○ $125,000 to $150,000
- ○ Over $150,000

Appendix 2: Samples of Percentage of Demography

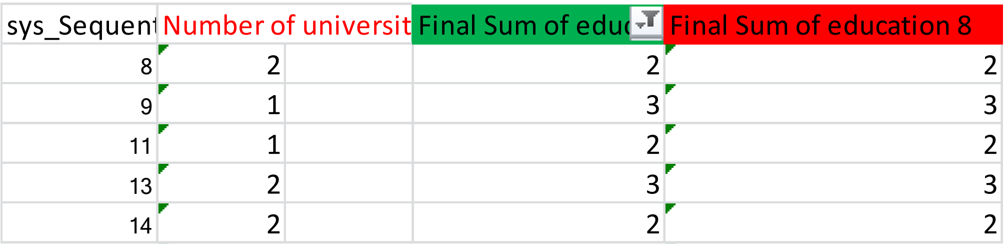
(..\Step 8\Heating System\Percentage of Demograpy--8.xlsx)



sys_Sequen	id	sum of Group 4,5(>65)	Number of university
1	c45147	0	0
2	eb160d	0	0
3	5ef059	0	0
4	da4fb5	0	2
5	a0a080	0	1

Appendix 3: Samples of Percentage of Education

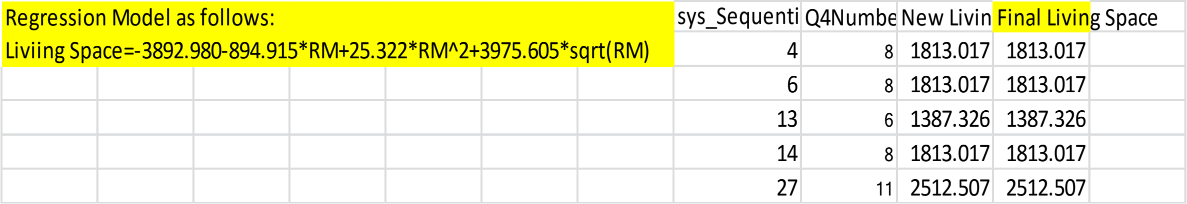
(..\Step 8\Heating System\Percentage of Education--8.xlsx)



sys_Sequent	Number of universit	Final Sum of educ	Final Sum of education 8
8	2	2	2
9	1	3	3
11	1	2	2
13	2	3	3
14	2	2	2

Appendix 4: 775 Living Space (Sample)

..\Step 8\Heating System\775 Living Space--8.xlsx



Regression Model as follows:
Liviing Space=-3892.980-894.915*RM+25.322*RM^2+3975.605*sqrt(RM)

sys_Sequenti	Q4Numbe	New Livin	Final Living Space
4	8	1813.017	1813.017
6	8	1813.017	1813.017
13	6	1387.326	1387.326
14	8	1813.017	1813.017
27	11	2512.507	2512.507

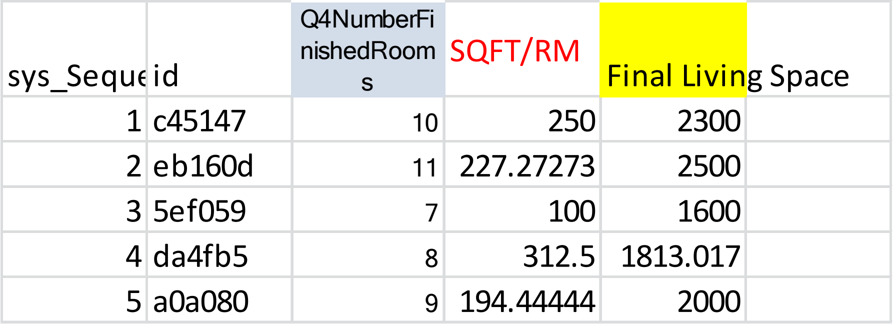
Appendix 5: 775 Living Space—R code

..\Step 8\Heating System\775 Living Space--8.R

```
data<-read.csv('775 Living Space--8.csv',nrows=775)

y<-data[,2]
y
length(y)

x1<-data[,3]
x1
length(x1)

x2<-data[,4]
x2
length(x2)

x3<-data[,5]
```

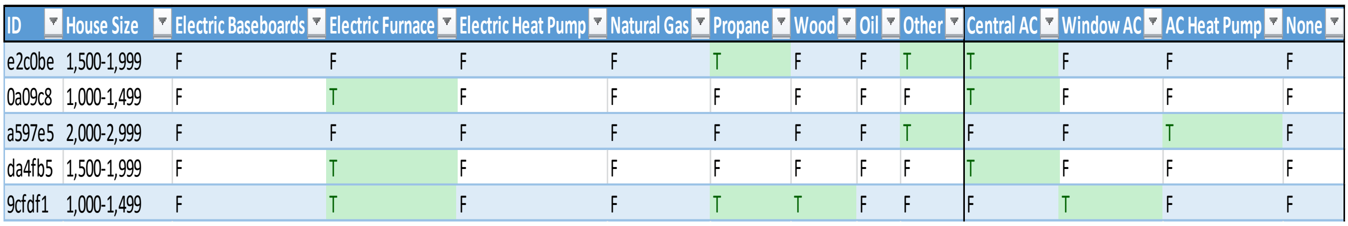
```
x3
length(x3)

model2<-lm(y~x1+x2+x3)
summary(model2)
```

Appendix 6: Living Space Estimation (Sample)
..\Step 8\Heating System\Living Space--8.xlsx



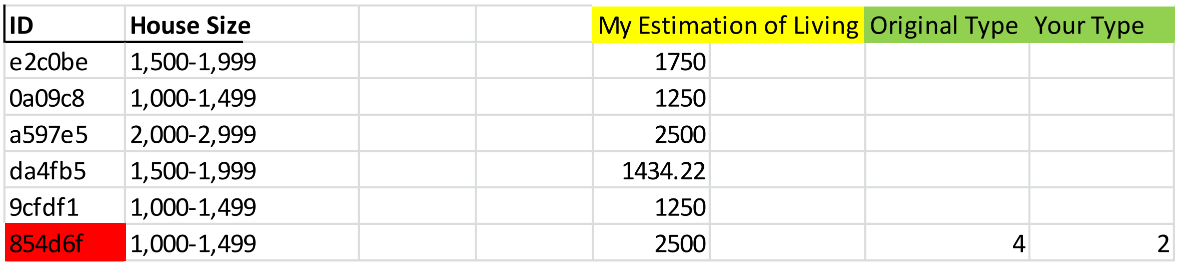
sys_Seque	id	Q4NumberFinishedRooms	SQFT/RM	Final Living Space	
1	c45147	10	250	2300	
2	eb160d	11	227.27273	2500	
3	5ef059	7	100	1600	
4	da4fb5	8	312.5	1813.017	
5	a0a080	9	194.44444	2000	

Appendix 7: Installers Customer House Estimates (Sample)
..\Step 8\SummerHill Customer House Estimates 20150923.xlsx



ID	House Size	Electric Baseboards	Electric Furnace	Electric Heat Pump	Natural Gas	Propane	Wood	Oil	Other	Central AC	Window AC	AC Heat Pump	None
e2c0be	1,500-1,999	F	F	F	F	T	F	F	T	T	F	F	F
0a09c8	1,000-1,499	F	T	F	F	F	F	F	F	T	F	F	F
a597e5	2,000-2,999	F	F	F	F	F	F	F	T	F	F	T	F
da4fb5	1,500-1,999	F	T	F	F	F	F	F	F	T	F	F	F
9cfdf1	1,000-1,499	F	T	F	F	T	T	F	F	F	T	F	F

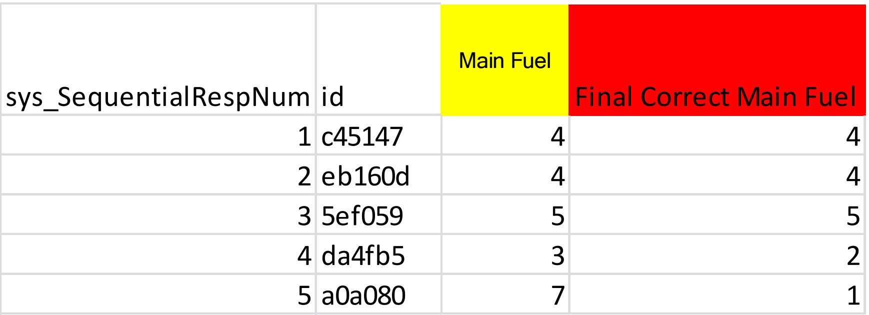
Appendix 8: Check Living Space with Installers (Sample)
..\Step 8\Check with Living Space.xlsx



ID	House Size			My Estimation of Living	Original Type	Your Type
e2c0be	1,500-1,999			1750		
0a09c8	1,000-1,499			1250		
a597e5	2,000-2,999			2500		
da4fb5	1,500-1,999			1434.22		
9cfdf1	1,000-1,499			1250		
854d6f	1,000-1,499			2500	4	2

Appendix 9: Check Heating System with Installers (Sample)

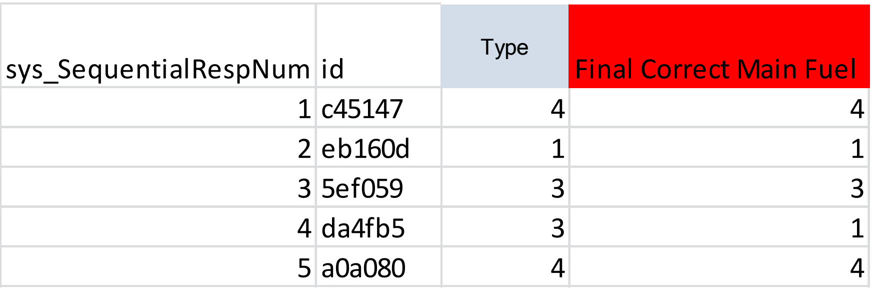
..\Step 8\Check with SummerHills Heating.xlsx



sys_SequentialRespNum	id	Main Fuel	Final Correct Main Fuel
1	c45147	4	4
2	eb160d	4	4
3	5ef059	5	5
4	da4fb5	3	2
5	a0a080	7	1

Appendix 10: Check Air-Conditioning with Installers

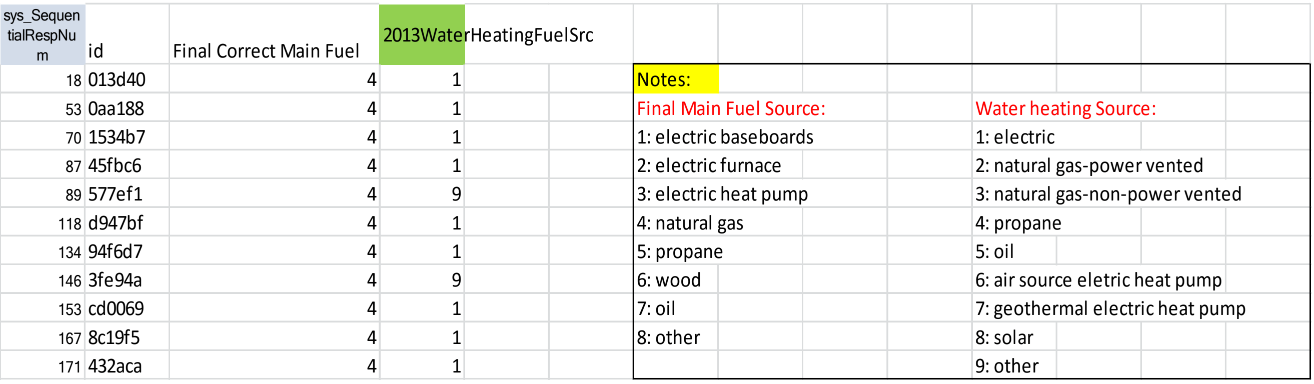
..\Step 8\Check with SummerHills AC.xlsx



sys_SequentialRespNum	id	Type	Final Correct Main Fuel
1	c45147	4	4
2	eb160d	1	1
3	5ef059	3	3
4	da4fb5	3	1
5	a0a080	4	4

Appendix 11: Check between Main fuel source and Electricity Water Heating (Sample)

..\Step 8\Check between Main fuel source and Electricity Water heating.xlsx



sys_SequentialRespNum	id	Final Correct Main Fuel	2013WaterHeatingFuelSrc
18	013d40	4	1
53	0aa188	4	1
70	1534b7	4	1
87	45fbc6	4	1
89	577ef1	4	9
118	d947bf	4	1
134	94f6d7	4	1
146	3fe94a	4	9
153	cd0069	4	1
167	8c19f5	4	1
171	432aca	4	1

Notes:

Final Main Fuel Source:
1: electric baseboards
2: electric furnace
3: electric heat pump
4: natural gas
5: propane
6: wood
7: oil
8: other

Water heating Source:
1: electric
2: natural gas-power vented
3: natural gas-non-power vented
4: propane
5: oil
6: air source eletric heat pump
7: geothermal electric heat pump
8: solar
9: other

Appendix 12: Follow-up corrections regarding to Missing House Age

..\Step 9\follow-up corrections.xlsx

id	corrected house age (years old)
c3c59e	60
bdb106	25
05049e	28
eddea8	55
7f975a	41
1c9ac0	53
b8c37e	17
fb60d4	15
f3f1b7	58

id	correction
6c1da8	2 occupants have bachelor's degrees
db85e2	for high school or some post-secondary, change 1 to 2
814481	change # of occupants (for ages 31 to 50) to 1 (used to be 2)
bb7946	for high school or some post-secondary, change 1 to 2
eb160d	for high school or some post-secondary, should be 2
af1fb9	for high school or some post-secondary, change to 3
5e3881	for less than high school diploma, should be 1; for college trades diploma, should be 1

Appendix 13: 974 Households' Income sample

..\Step 8\974 Income Households.xlsx

sys_SequentialRespNum	Annual Income
1	137500
3	50000
4	70000
5	90000
6	200000

Appendix 14: Income Estimation---R code

..\Step 8\Estimation of Income.R

```
data<-read.csv('974 Income Households.csv',nrows=974)

y2<-data[,2]
y2
length(y2)

x11<-data[,3]
x11
length(x11)
```

```
x12<-data[,4]
x12
length(x12)

x13<-data[,9]
x13
length(x13)

x14<-data[,10]
x14
length(x14)

model3<-lm(y2~x11+x12+x13+x14-1)
summary(model3)
```

Appendix 15: Annual KWh for 978 households sample

..\..\Step 11-Jan 19\KWh for 978 hhs.xlsx

id	sys_SequentialRespNum	total_kWh_2013	LN(kWh)
c45147	1	10591.014	9.267761185
eb160d	2	8402.064998	9.036232788
5ef059	3	25263.099	10.13710007
da4fb5	4	14614.396	9.589762349
a0a080	5	10967.21	9.302665191

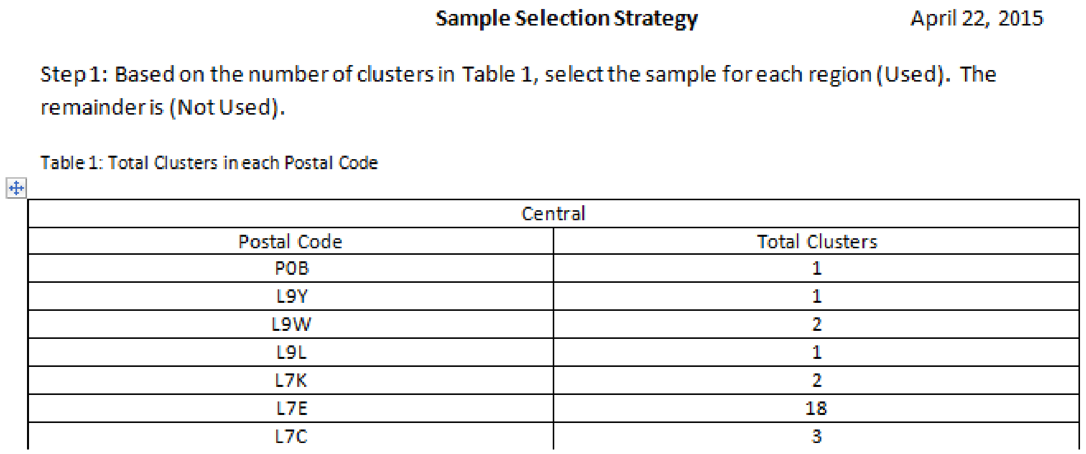
Appendix 16:Weather Station by Postal Code Sample

..\Step 2---HDH & CDH\Customers by Postal Code 10 Mar 13 Sorted.xlsx

12-三月-13	**Central**	# cust	mean kwh (2009)	WS		**East**	# cust	mean kwh (2009)	WS
Residential	N0A	1		HAM		K8P	197	11,357	KIN
Residential	N3B	1		HAM		K6T	218	14,535	KIN
Residential	N1G	7		HAM		K6K	255	11,917	KIN
Residential	N1T	7		HAM		K8R	438	11,303	KIN
Residential	N1K	35	14,225	HAM		K7G	648	17,437	KIN
Residential	N1E	56	11,722	HAM		K7L	1209	16,021	KIN

Appendix 17: Sample & Clusters in Each Region

..\Step 2---HDH & CDH\Sample & Clusters in Each Region_additional_rev_Jul9.docx



Sample Selection Strategy April 22, 2015

Step 1: Based on the number of clusters in Table 1, select the sample for each region (Used). The remainder is (Not Used).

Table 1: Total Clusters in each Postal Code

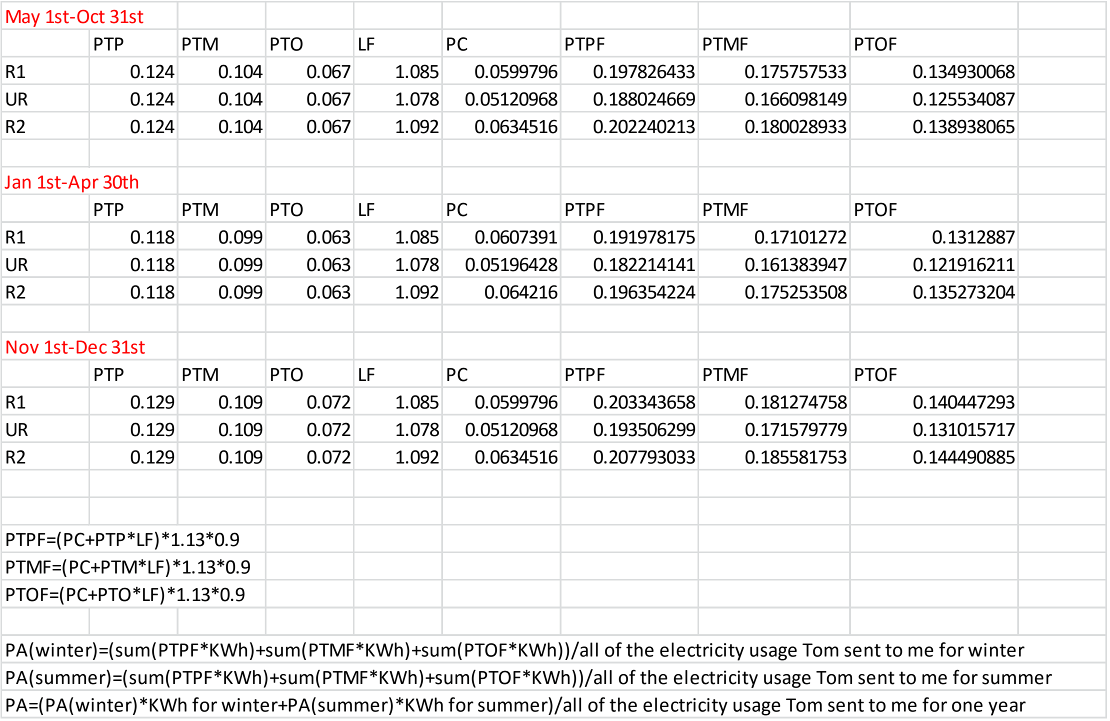
Central	
Postal Code	Total Clusters
P0B	1
L9Y	1
L9W	2
L9L	1
L7K	2
L7E	18
L7C	3

Appendix 18: Weather data download website

http://climate.weather.gc.ca/

Appendix 19: Price formula

..\..\Step 11-Jan 19\Formula for Prices data.xlsx



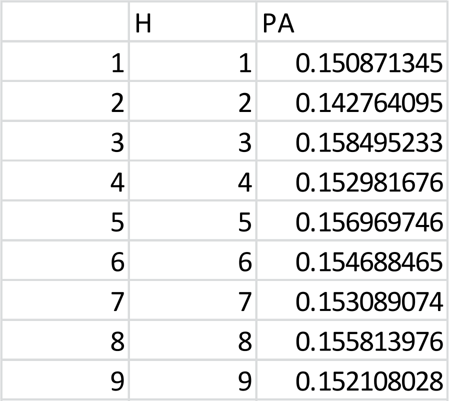
May 1st-Oct 31st								
	PTP	PTM	PTO	LF	PC	PTPF	PTMF	PTOF
R1	0.124	0.104	0.067	1.085	0.0599796	0.197826433	0.175757533	0.134930068
UR	0.124	0.104	0.067	1.078	0.05120968	0.188024669	0.166098149	0.125534087
R2	0.124	0.104	0.067	1.092	0.0634516	0.202240213	0.180028933	0.138938065
Jan 1st-Apr 30th								
	PTP	PTM	PTO	LF	PC	PTPF	PTMF	PTOF
R1	0.118	0.099	0.063	1.085	0.0607391	0.191978175	0.17101272	0.1312887
UR	0.118	0.099	0.063	1.078	0.05196428	0.182214141	0.161383947	0.121916211
R2	0.118	0.099	0.063	1.092	0.064216	0.196354224	0.175253508	0.135273204
Nov 1st-Dec 31st								
	PTP	PTM	PTO	LF	PC	PTPF	PTMF	PTOF
R1	0.129	0.109	0.072	1.085	0.0599796	0.203343658	0.181274758	0.140447293
UR	0.129	0.109	0.072	1.078	0.05120968	0.193506299	0.171579779	0.131015717
R2	0.129	0.109	0.072	1.092	0.0634516	0.207793033	0.185581753	0.144490885

PTPF=(PC+PTP*LF)*1.13*0.9
PTMF=(PC+PTM*LF)*1.13*0.9
PTOF=(PC+PTO*LF)*1.13*0.9

PA(winter)=(sum(PTPF*KWh)+sum(PTMF*KWh)+sum(PTOF*KWh))/all of the electricity usage Tom sent to me for winter
PA(summer)=(sum(PTPF*KWh)+sum(PTMF*KWh)+sum(PTOF*KWh))/all of the electricity usage Tom sent to me for summer
PA=(PA(winter)*KWh for winter+PA(summer)*KWh for summer)/all of the electricity usage Tom sent to me for one year

Appendix 20 :Time-of-Use Prices result sample

..\..\Step 11-Jan 19\Calculate corresponding product of PTPF and KWH\PA list--worksheet.xlsx



	H	PA
1	1	0.150871345
2	2	0.142764095
3	3	0.158495233
4	4	0.152981676
5	5	0.156969746
6	6	0.154688465
7	7	0.153089074
8	8	0.155813976
9	9	0.152108028

Appendix 21: **OLS Model**

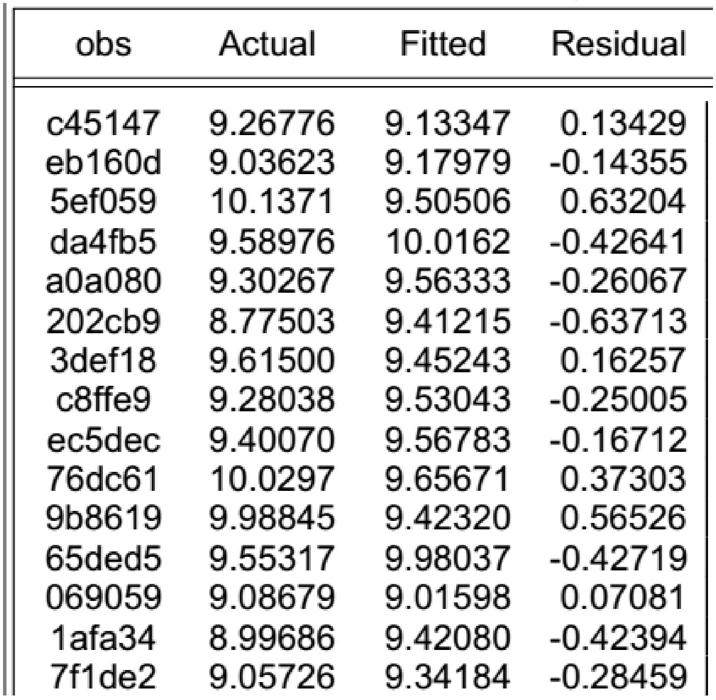
```
series heat10= d1*sqrt_hdh_*sqrt_size_


series houseage12=agegroup2*d1*sqrt_hdh_*sqrt_size_

series houseage13=agegroup3*d1*sqrt_hdh_*sqrt_size_

series houseage14=agegroup4*d1*sqrt_hdh_*sqrt_size_

series houseage15=agegroup5*d1*sqrt_hdh_*sqrt_size_


series ha11=hagroup1*d1*sqrt_hdh_*sqrt_size_

series ha12=hagroup2*d1*sqrt_hdh_*sqrt_size_


series pel11=_edugroup1*d1*sqrt_hdh_*sqrt_size_

series pel12=_edugroup2*d1*sqrt_hdh_*sqrt_size_

series pel13=_edugroup3*d1*sqrt_hdh_*sqrt_size_


series agedist11=_agegroup1*d1*sqrt_hdh_*sqrt_size_

series agedist12=_agegroup2*d1*sqrt_hdh_*sqrt_size_

series agedist13=_agegroup3*d1*sqrt_hdh_*sqrt_size_

series agedist14=_agegroup4*d1*sqrt_hdh_*sqrt_size_


series nr11=number_of_residents*d1*sqrt_hdh_*sqrt_size_
```

```
series income11=final_log_income*d1*sqrt_hdh_*sqrt_size_

series tup11=final_pa*d1*sqrt_hdh_*sqrt_size_

series nhs11=nhs*d1*sqrt_hdh_*sqrt_size_


series water20=d2


series agedist21=_agegroup1*d2

series agedist22=_agegroup2*d2

series agedist23=_agegroup3*d2

series agedist24=_agegroup4*d2


series pel21=_edugroup1*d2

series pel22=_edugroup2*d2

series pel23=_edugroup3*d2


series income21=final_log_income*d2

series tup21=final_pa*d2

series nr21=number_of_residents*d2

series nwh21=nwh*d2

series dtf21=dtf*d2


series cool30= d3*sqrt_cdh_*sqrt_size_


series achp31=achpgroup1* d3*sqrt_cdh_*sqrt_size_

series achp32=achpgroup2* d3*sqrt_cdh_*sqrt_size_


series houseage32=agegroup2*d3*sqrt_cdh_*sqrt_size_

series houseage33=agegroup3*d3*sqrt_cdh_*sqrt_size_

series houseage34=agegroup4*d3*sqrt_cdh_*sqrt_size_

series houseage35=agegroup5*d3*sqrt_cdh_*sqrt_size_
```

```
series pel31=_edugroup1*d3*sqrt_cdh_*sqrt_size_

series pel32=_edugroup2*d3*sqrt_cdh_*sqrt_size_

series pel33=_edugroup3*d3*sqrt_cdh_*sqrt_size_


series agedist31=_agegroup1*d3*sqrt_cdh_*sqrt_size_

series agedist32=_agegroup2*d3*sqrt_cdh_*sqrt_size_

series agedist33=_agegroup3*d3*sqrt_cdh_*sqrt_size_

series agedist34=_agegroup4*d3*sqrt_cdh_*sqrt_size_


series income31=final_log_income*d3*sqrt_cdh_*sqrt_size_

series tup31=final_pa*d3*sqrt_cdh_*sqrt_size_

series nr31=number_of_residents*d3*sqrt_cdh_*sqrt_size_

series ncs31=ncs*d3*sqrt_cdh_*sqrt_size_


series other40= sqrt_d4_


series pel41=_edugroup1*sqrt_d4_

series pel42=_edugroup2*sqrt_d4_

series pel43=_edugroup3*sqrt_d4_


series agedist41=_agegroup1* sqrt_d4_

series agedist42=_agegroup2* sqrt_d4_

series agedist43=_agegroup3* sqrt_d4_

series agedist44=_agegroup4* sqrt_d4_


series nr41=number_of_residents* sqrt_d4_

series income41=final_log_income* sqrt_d4_

series tup41=final_pa* sqrt_d4_

series noa41=noa* sqrt_d4_
```

```
ls ln_kwh_ c heat10 houseage12 houseage13 houseage14 houseage15 ha11 ha12 pel11 pel12 pel13 agedist11 agedist12 agedist13 agedist14 nr11 income11 tup11 nhs11 water20 agedist21 agedist22 agedist23 agedist24 pel21 pel22 pel23 income21 tup21 nr21 nwh21 dtf21 cool30 achp31 achp32 houseage32 houseage33 houseage34 houseage35 pel31 pel32 pel33 agedist31 agedist32 agedist33 agedist34 income31 tup31 nr31 ncs31 other40 pel41 pel42 pel43 agedist41 agedist42 agedist43 agedist44 nr41 income41 tup41 noa41
```

Appendix 22: **Residuals of each household from Eviews (From the OLS result table: ..\..\Step 15-Feb 16\residuals from OLS.pdf)**



obs	Actual	Fitted	Residual
c45147	9.26776	9.13347	0.13429
eb160d	9.03623	9.17979	-0.14355
5ef059	10.1371	9.50506	0.63204
da4fb5	9.58976	10.0162	-0.42641
a0a080	9.30267	9.56333	-0.26067
202cb9	8.77503	9.41215	-0.63713
3def18	9.61500	9.45243	0.16257
c8ffe9	9.28038	9.53043	-0.25005
ec5dec	9.40070	9.56783	-0.16712
76dc61	10.0297	9.65671	0.37303
9b8619	9.98845	9.42320	0.56526
65ded5	9.55317	9.98037	-0.42719
069059	9.08679	9.01598	0.07081
1afa34	8.99686	9.42080	-0.42394
7f1de2	9.05726	9.34184	-0.28459

Appendix 23: **GLS Model**

```
series heat10= d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_


series houseage12=agegroup2*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_

series houseage13=agegroup3*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_

series houseage14=agegroup4*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_

series houseage15=agegroup5*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_


series ha11=hagroup1*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_

series ha12=hagroup2*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_


series pel11=_edugroup1*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_

series pel12=_edugroup2*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_
```

```
series pel13=_edugroup3*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_


series agedist11=_agegroup1*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_
series agedist12=_agegroup2*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_
series agedist13=_agegroup3*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_
series agedist14=_agegroup4*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_


series nr11=number_of_residents*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_
series income11=final_log_income*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_
series tup11=final_pa*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_
series nhs11=nhs*d1*sqrt_hdh_*sqrt_size_*hat_residual21___1_2_



series water20=d2*hat_residual21___1_2_


series agedist21=_agegroup1*d2*hat_residual21___1_2_
series agedist22=_agegroup2*d2*hat_residual21___1_2_
series agedist23=_agegroup3*d2 *hat_residual21___1_2_
series agedist24=_agegroup4*d2*hat_residual21___1_2_


series pel21=_edugroup1*d2*hat_residual21___1_2_
series pel22=_edugroup2*d2*hat_residual21___1_2_
series pel23=_edugroup3*d2*hat_residual21___1_2_


series income21=final_log_income*d2*hat_residual21___1_2_
series tup21=final_pa*d2*hat_residual21___1_2_
series nr21=number_of_residents*d2*hat_residual21___1_2_
series nwh21=nwh*d2*hat_residual21___1_2_
series dtf21=dtf*d2*hat_residual21___1_2_
```

```
series cool30= d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_


series achp31=achpgroup1* d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series achp32=achpgroup2* d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_


series houseage32=agegroup2*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series houseage33=agegroup3*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series houseage34=agegroup4*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series houseage35=agegroup5*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_


series pel31=_edugroup1*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series pel32=_edugroup2*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series pel33=_edugroup3*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_


series agedist31=_agegroup1*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series agedist32=_agegroup2*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series agedist33=_agegroup3*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series agedist34=_agegroup4*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_


series income31=final_log_income*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series tup31=final_pa*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series nr31=number_of_residents*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_

series ncs31=ncs*d3*sqrt_cdh_*sqrt_size_*hat_residual21___1_2_


series other40= sqrt_d4_*hat_residual21___1_2_


series pel41=_edugroup1*sqrt_d4_*hat_residual21___1_2_

series pel42=_edugroup2*sqrt_d4_*hat_residual21___1_2_

series pel43=_edugroup3*sqrt_d4_*hat_residual21___1_2_


series agedist41=_agegroup1* sqrt_d4_*hat_residual21___1_2_
```

```
series agedist42=_agegroup2* sqrt_d4_*hat_residual21___1_2_

series agedist43=_agegroup3* sqrt_d4_*hat_residual21___1_2_

series agedist44=_agegroup4* sqrt_d4_*hat_residual21___1_2_


series nr41=number_of_residents* sqrt_d4_*hat_residual21___1_2_

series income41=final_log_income* sqrt_d4_*hat_residual21___1_2_

series tup41=final_pa* sqrt_d4_*hat_residual21___1_2_

series noa41=noa* sqrt_d4_*hat_residual21___1_2_


series y2= ln_kwh_*hat_residual21___1_2_


ls y2 c heat10 houseage12 houseage13 houseage14 houseage15 ha11 ha12 pel11 pel12 pel13
agedist11 agedist12 agedist13 agedist14 nr11 income11 tup11 nhs11 water20 agedist21 agedist22
agedist23 agedist24 pel21 pel22 pel23 income21 tup21 nr21 nwh21 dtf21 cool30 achp31 achp32
houseage32 houseage33 houseage34 houseage35 pel31 pel32 pel33 agedist31 agedist32 agedist33
agedist34 income31 tup31 nr31 ncs31 other40 pel41 pel42 pel43 agedist41 agedist42 agedist43
agedist44 nr41 income41 tup41 noa41
```

Appendix 24: **Wald test details**

Consider a general nonlinear regression model

$$y = x(\beta) + \epsilon$$

where β is a *k* vector of parameters to estimate. Any restrictions on the parameters can be written as

$$H_0: g(\beta) = 0$$

In this example, $g(\beta) = e^{(c(5)*(0-1)+c(7)*(1-0))*1*330.1441249*44.73962772} - 1.$

where *g* is a smooth *q* dimensional vector imposing *q* restrictions on β. The Wald statistic is then computed as

$$W = ng(\beta)'(\frac{\partial g}{\partial \beta} V(b) \frac{\partial g}{\partial \beta'})^{-1} g(\beta)|_{\beta=b}$$

where *n* is the number of observations and *b* is the vector of unrestricted parameter estimates. *V* is the estimated variance of *b* given by

$$V(b) = ns^2(\frac{\partial x}{\partial \beta}\frac{\partial x}{\partial \beta'})^{-1}|_{\beta=b}\ , \quad s^2 = \frac{u'u}{n-k}$$

Where *u* is the unrestricted residuals.

More formally, under the null hypothesis H_0 , the Wald statistic has an asymptotic $\mathcal{X}^2$(*q*) distribution, where *q* is the number of restrictions under H_0.