university of groningen

The effects of green energy labels on household satisfaction levels

A quantitative study towards the effects and stability of green labels as a parameter on housing, residential and life satisfaction over time.

Abstract.

Household satisfaction levels have been examined extensively, though none of the earlier studies examined the effects of energy labels. This research adds to the literature by examining the effects of green energy labels. Additionally it deviates from existing literature by examining the effects on multiple levels of household satisfaction as well as the stability of the effects of green energy labels over time and between types of tenure. WoON survey data from 2012, 2015 and 2018 is examined using a binary logistic regression analysis. The regression results indicate a significant, positive effect on both housing and residential satisfaction when households live in a home with a green energy label compared to inefficient energy labels. The effects of the most energy efficient labels are increasingly, positively affecting household satisfaction levels throughout time, whereby the least energy efficient green label C is becoming less or even insignificant due to an ever-increasing average energy efficiency of the Dutch housing stock throughout time. Also, the effects of green energy labels on household satisfaction levels are the strongest for owner-occupiers, predominantly driven by label B and C. Contrary to the effects of energy labels on housing and residential satisfaction, energy efficient green labels do not affect the likelihood of households' satisfaction with life in general.

Keywords: [Energy label, Green homes, household satisfaction levels, housing satisfaction, residential satisfaction, life satisfaction]

Roel Bakhuis

9-8-2021

university of groningen

COLOFON

Document	Master thesis
Title	The effects of green labels on household satisfaction levels
Version	Final version
Author	Roel Bakhuis
Student number	4113071
Email RUG	R.bakhuis.1@student.rug.nl
Email private	Roel.bakhuis@hotmail.com
Institution	University of Groningen Faculty of Spatial Sciences MSc in Real Estate Studies Landleven 1, 9747 AD, Groningen
Supervisor	Em. Prof. dr Ed Nozeman
Assessor	Dr. Xiaolong Liu
Date	9-8-2021
Word count	19.912
Research disclaimer	*"Master theses are preliminary materials to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the author and do not indicate concurrence by the supervisor or research staff."*

university of groningen

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Motivation

The recently accepted Climate agreement for The Netherlands ("Klimaatakkoord") of 2019 is part of a policy framework attempting to step-by-step reduce greenhouse gas emissions. Some of the principles were already part of the in 2013 implemented Energy Agreement ("Energie akkoord") (Sociaal Economische Raad[2], n.d.), aimed to increase energy savings and stimulate the use of renewable energy. The residential market was responsible for approximately 25% of the energy consumption (Olaniyan & Evans, 2014), thus the Energy Agreement is unsurprisingly consisting of sustainability measures and policy reforms to improve the energetic performance and towards an energy neutral built environment by 2050 (SER[1], n.d.).

The energetic performance of a dwelling in the Netherlands is measured via the Energy Performance Advice (EPA-certificate) since 2008. The labels range from very energy inefficient (label G) towards more energy efficient green homes labelled as A/B/C (label A+++ being most energy efficient) providing the energetic performance transparently. Additionally, it increases awareness on the energy performance of a home and provides insights towards potential energy efficiency measures (Energielabelvoorwoningen, n.d.). On average, the government aims to establish a minimum energy label B in the public rental sector and label C in the private rental sector, while newly constructed residential real estate should be energy neutral (SER[1], n.d.). In order to achieve that goal, legislation on energy labels is getting stricter. Before 2015, households' were able to mutually agree the exclusion of an energy label at a residential transaction. However, since January 2015 the intensified legislation prohibited transactions without energy labels, therefore enforcing the application of energy labels. Even more recently in 2021 the regulation on energy labels got even stricter as labels must be established by a licensed energy advisor (Energielabelvoorwoningen, n.d.).

The effects of the intensified legislation on energy labels is already visible in the residential market, as the share of green homes labelled A or B grew from 16% in 2013 till 33% in 2019. In the same period, houses labelled D or worse decreased from 54% till 37% in 2019 (Rijksoverheid, 2019). However, this is not completely due to stricter legislation as, according to BPD (2019), nowadays 90% of home seekers mentions energy efficiency as one of the housing requirements. It is therefore not surprising that due to the increased urgency of energy labels, the academic debate on its implications continued as well. Nearly 1 million articles were published on google scholar between 2011 and 2020, almost three times as much published articles compared to 2001-2010.

Most academic research on energy labels is related to financial aspects of the residential market. The perceived impact of energy labels on households' satisfaction remains somewhat underexposed. To the best knowledge of the author, only a peer student conducted a master thesis towards the impact of energy labels on household satisfaction levels. The work of fellow peers however, is seen as grey literature which does not meet the strict requirements of published academic articles needed for academic research. More on this research conducted by Dolunay Olgun (2020) later on. Despite the lack of similar published scientific articles, some published academic studies examined the impact of specific sustainability measures (Goodwin, 2011; Johnson, 2014) or green features (Tan, 2014) towards satisfaction, indicating that certain sustainability measures and green features enhance residential satisfaction. Residents want a green home to increase their quality of life in terms of comfort and health (Ebrahimigharehbaghi et al, 2019), as green energy labels also represent lower energy costs and higher indoor comfort levels (Wong, 2020). This might explain why sustainability and green homes are increasingly popular on the residential market (Goodwin, 2011).

Despite the impact of specific green features on residential satisfaction and the increased attention for sustainable green homes on the residential market, no published scientific research has been conducted towards the relationship between energy labels and household satisfaction levels[1]. The only similar research is an unpublished master thesis conducted by peer student Olgun (2020), which solely assess residential satisfaction. Nonetheless, it does assess the effects of green energy labels on residential satisfaction levels of households in 2018. Whether green energy labels contribute towards an increasingly higher satisfaction level throughout time remains unclear as solely one cycle of data has been taken into consideration. This research however, assess multiple cycles of WoON data surveys in order to assess the stability of the effects over time. Moreover, several levels of household satisfaction have been included, so that the study also provides more insight into the effect of green energy labels on the different levels of household satisfaction, namely housing, residential and life satisfaction. This research thus further extends our existing literature, by assessing the effects of green energy labels with multiple data cycles throughout time, while distinguishing the effects on the various levels of household satisfaction.

[1] Examined search results of Google scholar, Scopus and SmartCat using key search words such as: household satisfaction, housing satisfaction, residential satisfaction, life satisfaction and satisfaction levels combined with key words relatable to energy labels such as sustainability, energy efficient homes, energy labels, green labels, green homes etc.

1.2 Literature review

Earlier literature on household satisfaction is extensive, as it is a complex, multi-dimensional phenomenon that has been debated throughout various research disciplines. However, their theoretical notions are strikingly similar, as most studies are based on one or multiple of the same underlying theoretical mechanisms, namely the housing needs theory, housing deficit theory and psychological construct theory. All of these mechanisms hinge upon the same core principle that the difference between a households' actual and desired dwelling conditions affects the level of satisfaction (Galster & Hesseler, 1981). The current housing situation compared to the desired or reference average condition in society thus effects household satisfaction levels. With an ever-increasing average green energy label in the Dutch housing stock, this might lead to an increasing effect of energy labels on household satisfaction levels.

What deviates however, is the level of household' satisfaction these studies take into consideration. Generally, household satisfaction is considered in either one of the following three levels. Firstly, housing satisfaction which controls for dwelling characteristics. Additionally, as is the case for every level of household satisfaction analysis, the specific socio-demographic composition of the household at hand has been taken into consideration (Grigolon et al, 2014; Golant, 1982) as Grigolon et al (2014) established that households socio-demographic's composition and housing characteristics affects satisfaction levels. Residential satisfaction basically is an extension of the analysis of housing satisfaction, as it also takes the household' socio-demographic composition into consideration as well as the specific dwelling characteristics. However, it also includes classical variables regarding the neighbourhood characteristics (Crull, 1991). Lastly, life satisfaction further extends the analysis by taking certain general features of life into consideration, like health and employment status (Balestra & Sultan, 2013; Elsinga et al, 2015; Ren, Folmer & Van der Vlist, 2018). Extensive academic debates have been going on about the various levels of household satisfaction. Depending on their level of household satisfaction, those studies take classical attributes of the socio-demographic composition, dwelling characteristics, neighbourhood characteristics and general features of life into consideration. Most of these studies solely examine the effects on one level of household satisfaction, ignoring either housing, residential and or life satisfaction. A more comprehensive approach would consider the effects on all levels of household satisfaction.

The impact of energy labels on the residential market has, as well as household satisfaction been debated over extensively. Most of these studies however, examine the financial implications of energy labels. Brounen & Kok (2011) for example, examined the impact of a green energy label on housing prices. Their research indicates that green energy labels impact future housing values, as these

diminish the information asymmetry regarding the energy performance of a building. Similar findings are made by Hyland et al (2013) stating that green homes result in higher sales prices or rent levels. households are therefore willing to pay a premium for dwellings with green energy labels, as the reduced energy costs get capitalised in the housing prices or rent. More importantly, the increased information transparency due to the introduction of energy labels enables consumers to set off their sustainability level of the home relatively to the other households. Apart from altruistic and environmentalist arguments, based on the underlying mechanisms of household satisfaction, reveals the discrepancies between the current and average or reference situation, thus affecting household satisfaction levels.

However, few studies synthesized both themes and take energy labels into consideration when assessing household satisfaction levels. Some studies touched upon both household satisfaction as green homes, as they examine the impact of certain sustainability measures on households satisfaction. Johnson (2014) for example, examined the difference of residential satisfaction between green homes and congenital homes, regarding thermal quality, air quality, sound quality and cooling/heating system. Another related research has been conducted by Tan (2014), who examined the satisfaction of specific energy efficiency measures in Malaysia and concluded that specific measures such as applying double glazing and installing solar panels positively affects the residential satisfaction of homeowners. These studies indicated that certain sustainability measures and green features enhance residential satisfaction. Residents want to live in a green home mainly to increase their quality of life in terms of comfort and health (Ebrahimigharehbaghi et al, 2019, Wong, 2020), potentially declaring the increased importance of sustainably and green homes as factors on the residential market (Goodwin, 2011).

The introduction of the energy label revealed discrepancies in energy efficiency of households homes, affecting household satisfaction levels. With ever increasing attention to energy labels and increasingly tight legislation and policies towards stimulating the sustainability of the energy label, discrepancies are likely to increase. It is not without reason that consumers are actively looking for housing with a sustainable energy label. However, the limited existing research that combined both energy labels and household satisfaction levels as it confined itself to the impact of specific green features on housing or life satisfaction specifically, at one moment in time. No published academic research has been conducted towards the impact of green energy labels on various levels of household satisfaction over time. The only comparable research is grey literature, and is limited to examining the effects of green energy labels on one level or household satisfaction at one moment in time.

1.3. Research problem statement and research questions

This research aims to fill specific gaps in the scientific debate on housing satisfaction in three different ways. Firstly, this research adds by explicitly examining the effects of green energy labels on household satisfaction. Secondly, existing literature mostly examines household satisfaction on one level, whereas this research examines the effects of green energy labels on housing satisfaction and residential satisfaction while controlling for certain general features of life related to life satisfaction. Thirdly, literature examined household satisfaction in a cross-sectional setting using data on one moment in time, whereas this research includes multiple cycles of WoON survey data to examine the stability of parameters over time. Bearing in mind the amended legislation on energy labels within this period, this research is able to examine whether there is a structural break in the impact of green energy labels on satisfaction levels over time. As increasing sustainability may lead to a discrepancy in the housing market, affecting satisfaction levels. Currently there is no insight in the relationship between green energy labels and household satisfaction levels. Therefore, this research aims to answer the following research question: *What is the relationship between green energy labels and household satisfaction levels in the Netherlands?* In order to examine the research problem and answer the main research question, the following sub-questions have been drafted:

1) *What are the determinants of the various levels of household satisfaction?*
2) *What is the relationship between green energy labels and various household satisfaction levels throughout time?*
3) *Are the effects of green energy labels on various levels of household satisfaction different per tenure group?*

1.4 Conceptual model

In order to answer the research questions stated in the previous section, this research will draw upon multiple cycles of the WoON housing survey. The WoON housing survey is carried out by the Dutch Central Bureau of Statistics (CBS) in consultation with the Ministry of Internal Affairs (BZK). These data will be used to examine the impact of energy labels on various household satisfaction levels as displayed in the conceptual model, figure 1.1. Based on theory, the classical control variables in research towards household satisfaction levels are established. There is some overlay in the classical control variables used between these levels, this research examines all three levels of satisfaction as displayed in figure 1.1. These contain various variables on the socio-demographic status of households, dwelling characteristics, neighbourhood characteristics and some general features of life. With the use of a pooled logistic regression analysis, that considers data of the WoON 2012, 2015 and 2018 cycles

as displayed in the arrows in figure 1.1, the effects of green energy labels on the various levels of household satisfaction will be examined, while controlled for the affiliated Z variables.

Previously conducted research on housing satisfaction is based on cross-sectional data using one cycle of WoON data, whilst this research examines three cycles of WoON surveys. Due to data constraints, the effects of energy labels life satisfaction is solely examined for 2015, more on this in chapter 4.

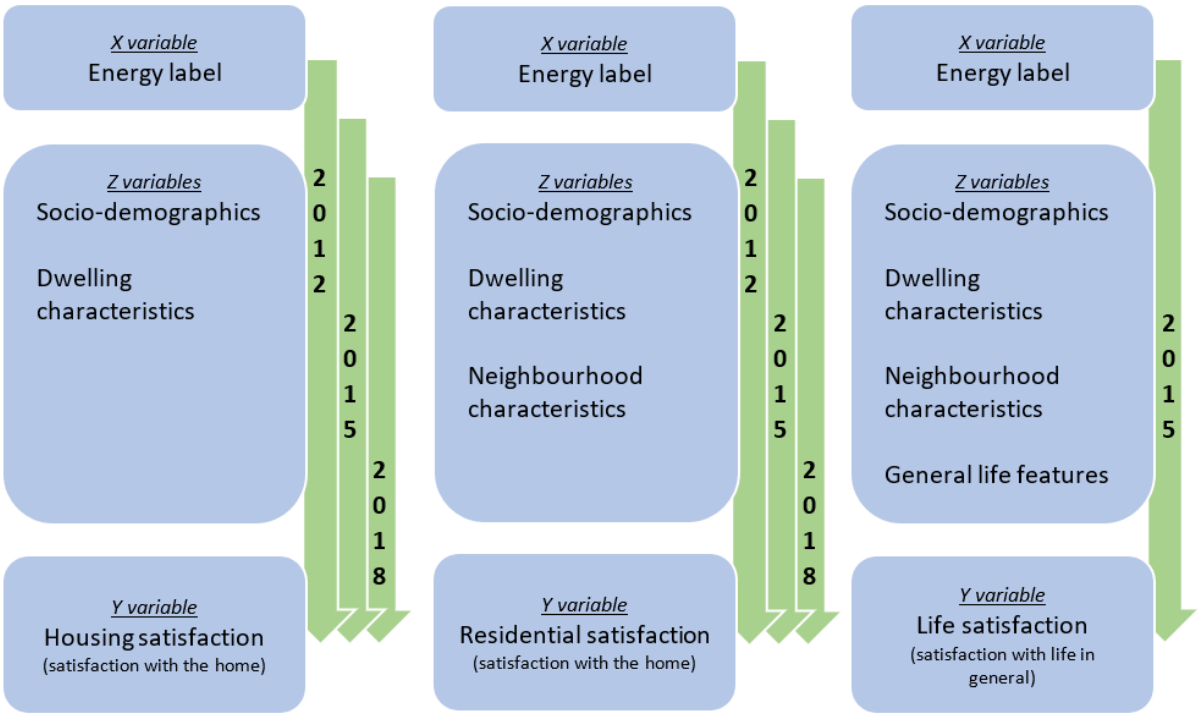


Figure 1.1: Conceptual model (own work)

The remainder of this paper is organized as follows: Chapter 2 places this research within the context of the current housing market developments and energy labels. Chapter 3 examines the existing literature and provides the relevant methodology and variables. In chapter 4 the dataset is introduced, described and prepared for the empirical research. The empirical methodology will be explained and the variables operationalized. Chapter 5 will display the regression results of the robustness checks and the housing, residential and life satisfaction. Chapter 6 contains discussion on the research results followed by the main conclusions, research limitations and research recommendations.

2. CONTEXTUAL FRAMEWORK

The introduction of the research in chapter 1 briefly touched upon the context within which this research is carried out. However, governmental policies towards a more sustainable society are in full development. The relative recent introduction of the energy label and the increasingly stringent legislation towards energy labels, make it important to outline the context of energy policies in the Netherlands as it iterates the importance of this study on the effects throughout time. This chapter outlines the most influential developments on energy label legislation and developments regarding sustainability of the overall housing stock will be described.

2.1 The energy label

The Dutch government introduced the mandatory Energetic Performance Advice (EPA-certificate) at any delivery, sale or rental agreement of a dwelling in 2008. This so-called EPA-certificate is referred to as the energy label and got introduced in order to stimulate the improvement of energetic performances of the built environment (Energiedeskundig, n.d.). The label ranges from very energy inefficient (label D/E/F/G) towards energy efficient green homes, labelled as A/B/C (label A+++ being the most energy efficient). In this way, the energetic performance of a home is displayed transparently. Additionally it actively increases awareness of the energy performance of a home and provides insights towards potential energy efficiency measures (Energielabelvoorwoningen, n.d.), while implicitly referring to comfort in the home, service energy costs and the residential quality (AgentschapNL, 2012).

The energy labels are based on the energetic performance of a dwelling determined via the energy-index. The energy performance calculation measures building-related energy use for heating, ventilation, lighting and hot water usage. In order to make comparisons between dwellings possible, the energy usage is corrected for standard resident behaviour and climate conditions. Therefore, the eventual label is solely based upon the construction properties of the dwelling such as insulation combined with the existing installations such as boilers and solar panels (AgentschapNL, 2012). Table 2.1 displays the classification of energy labels based on the energy-index score.

Energy Label Class	Energy-index score
A++	EI ≤ 0,50
A+	0,51 < EI ≤ 0,7
A	0,71 < EI ≤ 1,05
B	1,06 < EI ≤ 1,30
C	1,31 < EI ≤ 1,60
D	1,61 < EI ≤ 2,00
E	2,01 < EI ≤ 2,40
F	2,41 < EI ≤ 2,90
G	EI > 2,90

Table 2.1: Energy label classification (own work)

2.2 Governmental policies and energy label legislation

The Dutch government aims to create a more sustainable housing stock in the Netherlands as part of an extensive package of measures undertaken to reduce CO2 emissions (SER, n.d.). The energy label got introduced in 2008 as control instrument to stimulate sustainability measures by displaying the energetic performances transparently, while increasing awareness on current energetic performances (Energielabelvoorwoningen, n.d.). Energy labels stimulate the development of price premiums on sustainable homes (Brounen & Kok, 2011), stimulating sustainability efforts.

In an attempt to kickstart the usage of energy labels, the government decided to direct a pioneering role to Dutch housing associations. Hybrid housing associations are private companies that work on social tasks without a profit objective (AEDES, n.d.). In 2008, housing associations covered approximately 80% of all rental properties in the Netherlands. The imposed obligations on the housing associations, to establish an energy label for the entire housing stock as of 1-1-2009 (AgentschapNL, 2012). On average, the government aims to establish a housing stock with green, energy efficient homes labelled B in the public rental sector and C in the private rental sector. Newly constructed homes must be built energy neutral (SER, n.d.). However, to achieve these sustainability goals, policy reforms and tighter legislation was needed for further stimulation of the introduction of the energy label (Energielabelvoorwoningen, n.d.). The following timeline provides an overview of the increasingly stricter legislation and policies on energy-labels (Energiedeskundig, n.d.; Lente-akkoord, n.d.):

2008: Introduction of the energy label in the residential market, however buyers can sign a waiver to buy a dwelling exempt from an energy label;

2009: Housing associations are obliged to establish an energy label for their entire housing stock;

2012: Various influential organisations in the Dutch construction and living world, sign the renewed ‘Lente-akkoord’ (Spring agreement) with the ministry of Internal Affairs to strive newly built dwellings in 2015 to be 50% more energy efficient than in 2007 and energy neural by 2020. Parliament turns down new legislation towards the introduction of a mandatory energy label for newly constructed homes per 2013;

2013: At the request of market parties, the introduction of voluntarily energy label of newly constructed residential properties;

2015: The entire housing stock gets a provisional, simplified energy label that can be altered into a definitive label. Definitive labels are mandatory and enforced with every housing transaction by order of a fine. Consequently, the option to mutually agree to exempt the seller or renter of a dwelling from the presence of an energy label is expelled;

2021: Introduction of the new simplified energy label (in Dutch; 'Vereenvoudigde energielabel'), that needs to be established by a licensed energy advisor. Additionally, the underlying calculation of the energy-index is based on the new NTA 8800 method[2]. Consequently, the costs of obtaining and the penalty for the lacking a label have been increased significantly.

2.3 The Dutch housing stock

The previous sections discussed the energy label in general and the increasingly stricter legislation towards the application and enforcement of energy labels in the residential market. The effects of the policy changes and increasingly stricter legislation are also visible in the development of the Dutch housing stock.

The Dutch residential market is characterised by the presence of a large social rental sector. In 2009 the entire housing stock consisted of approximately 7.1 million homes, whereof 54% was owner-occupied, 30% consisted of social housing and 11% was owned by private investors. In a decade, the entire housing stock grew with approximately 750.000 homes, see figure 2.1. Relatively speaking, the share of social housing remained approximately 29% whereas the share of owner-occupied housing grew until 57% and private investors owned 13% (CBS, 2020[1]).

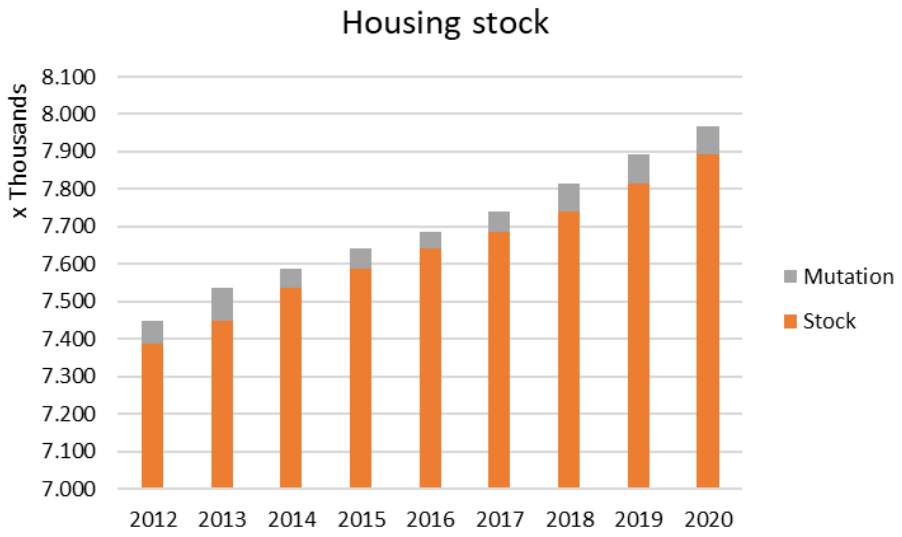


Figure 2.1: Housing stock development (CBS, 2021; edited by author)

Figure 2.2 displays an overview of the newly provided energy labels throughout time. The pioneering role of housing associations in relation to the implementation of the energy labels, combined with the fact that nearly a third of the total housing stock consisted of social housing in 2009, explains the

[2] This research uses data ranging from 2012 until 2018. Consequently, every time that energy labels are referred to, it is logically based on the now recently outdated assessment system.

relatively large amount of newly provided energy labels in 2009. It also indicates that approximately 50% of those newly provided labels are energy inefficient with label D or less, while the other 50% is label C or better. The clear majority consists of energy label C, generally considered as the minimum energy efficiency score to be referred to as a green, energy efficient home.

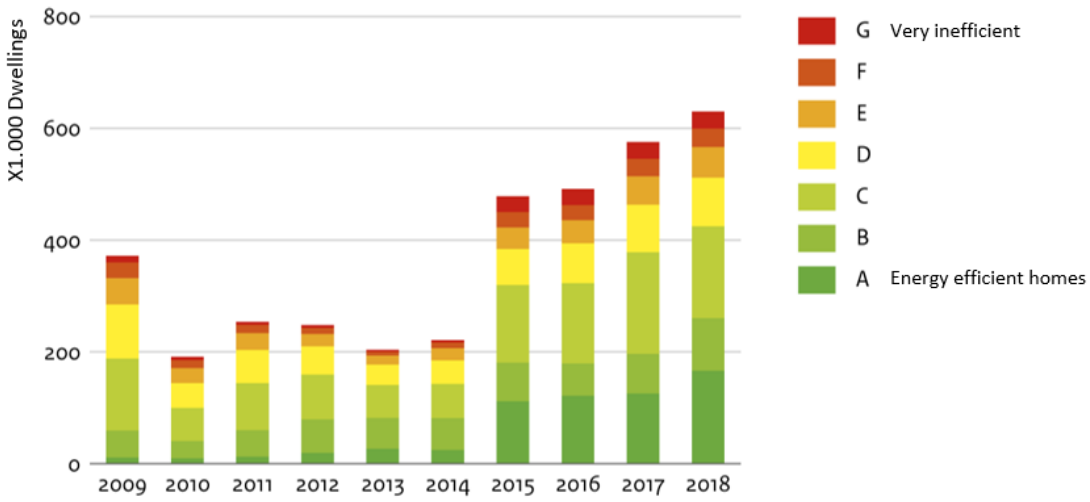


Figure 2.2: Provided energy label distribution of the Dutch housing stock per year (CLO, 2019; edited by author)

The second major disruptive event in terms of legislation occurred with the implementation of provisional energy label in 2015. Figure 2.2, clearly displays a strong increase in new energy labels compared to previous years. However, the structure of the newly provided labels has shifted as approximately 75% consists of labels C or better. Notable, the share of energy efficient homes labelled A is seemingly even to the share of label C.

Lastly, figure 2.3 displays the increase of newly provided energy labels over the last years. Indicating that an increasing share of the housing stock has applied for an energy label. That is also visible in figure 2.2, indicating the share of housing stock with a definitive energy label. It clearly indicates an increasing slope, though also provides an insight in the amount of dwellings in the Netherlands, that does not yet have a definitive energy label. Approximately 50% of the entire housing stock in 2020, therefore still only has the provisional label provided back in 2015.

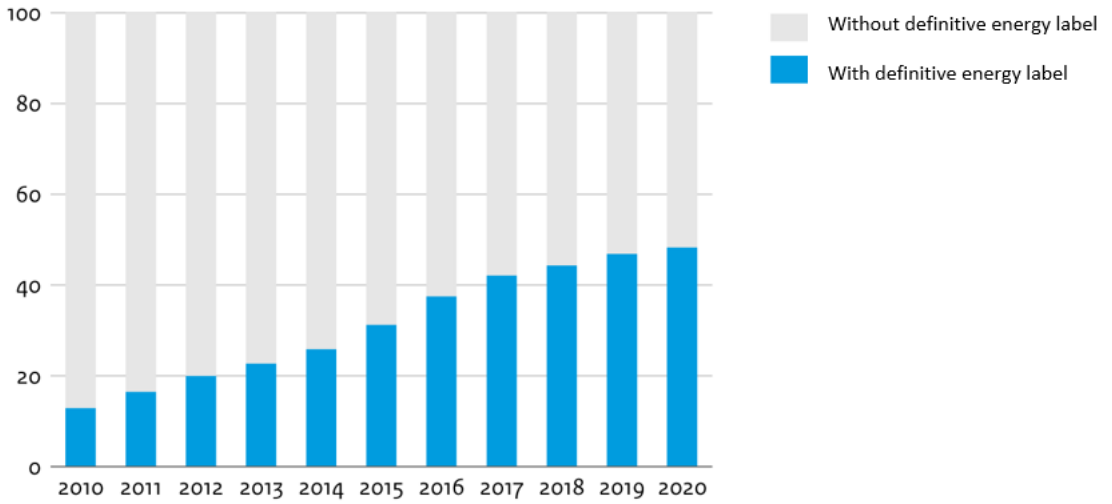


Figure 2.2: Share of housing stock with a definitive energy label (CLO, 2020; edited by author)

The increasing amount of residential properties with an energy label is also affected by the increasing supply of newly constructed dwellings. In 2014 approximately 45,200 newly constructed dwellings were added to the Dutch housing stock, a number that gradually increased towards 69.000 newly constructed homes in 2020. As a consequence of the new earlier discussed 'Lente akkoord', these newly constructed homes were often provided with a voluntary, temporary though green energy label. Combined with the increasing new construction production, the amount of energy labels grew as displayed in figure 2.2 (CBS, 2020; CLO, 2020; BZK, 2020).

3. THEORY

3.1 Mechanisms of household satisfaction

Empirical research on household satisfaction in the residential market revolves around three main theories, namely the housing needs theory, housing deficit theory and psychological construct theory (Mohit, 2014). The housing needs theory got introduced by Rossi (1955), who supposed that housing needs change throughout the lifecycle of a household. Whenever there is a discrepancy between the current and desired housing situation, dissatisfaction will occur. Life cycle changes are especially influential on this discrepancy, as various stages in life often result in changed space requirements (Mohit, 2014). Whenever dissatisfaction occurs, residents are likely to increase space utilisation, or migrate in an attempt to increase the level of satisfaction (Brown and Moore, 1970). Secondly, The housing deficit theory has been introduced in 1978 by Morris and Winter. They identify that housing satisfaction is based on households' judgement between their current housing conditions and societal norms or standards. The trade-off between the actual housing situation and the cultural norms could result in housing deficits and thus in an increased housing (dis)satisfaction. As a result, households are likely to adjust their home, revising their needs and aspirations or remodel the dwelling (Mohit, 2014). The last theory got introduced by Galster (1985). His psychological theory argues that satisfaction is a cognitive construct based on a reference condition. If the current housing conditions is perceived as superior to the reference condition, households tend to be satisfied. Whenever the opposite is the case and households are dissatisfied, a number of housing adjustments might occur. Either the housing aspirations get reduced, alterations on the dwelling are made in an attempt to lower the dissatisfaction, or the household moves towards a more suitable home (Galster, 1981; Mohit, 2014).

3.2 Levels of household satisfaction

In general, research towards household satisfaction in the residential market distinguishes three levels of satisfaction. Firstly housing satisfaction, also referred to as dwelling satisfaction and solely reflects the perceived quality of the dwelling (Elsinga and Hoekstra, 2005; Golant, 1982). To identify the effects of dwelling characteristics on housing satisfaction, the models control for the socio-demographic characteristics of the households when asked to what extent households are satisfied with their dwelling. However, housing satisfaction and neighbourhood satisfaction are closely related as housing assessments take the immediate surroundings into consideration as well (Lu, 1999; Crull, 1991). Therefore, most research towards household satisfaction on the residential market synthesises them[3],

[3] Throughout literature housing and residential satisfaction is used interchangeably due to their close relation. In this research housing satisfaction is solely based on dwelling characteristics, whereas residential satisfaction includes dwelling and neighbourhood characteristics. Both satisfaction levels are controlled for socio-demographics' of the household.

resulting in the second level of household satisfaction referred to as residential satisfaction. In addition to the socio-demographic composition and dwelling characteristics, residential satisfaction includes neighbourhood characteristics. Thirdly, the broadest approach towards household satisfaction levels is life satisfaction[4]. Contrary to both housing and residential satisfaction, it does not examine the perceived satisfaction of the dwelling, but rather households satisfaction levels of life in general. However, various studies established the positive relationship between residential satisfaction and life satisfaction (Balestra & Sultan, 2013; Golant, 1982; Fernández et al, 2017; Quang Tran et al, 2017), therefore self-evidently including similar variables relating the socio-demographics, dwelling and neighbourhood characteristics. Additionally, life satisfaction incorporates general features out of the lives of households, such as self-proclaimed health status and employment status.

The various measurement levels of household satisfaction levels are somewhat interrelated. Crull (1991) for instance recognised the interlinkages between housing satisfaction and residential satisfaction, due to the strong relation of housing assessments and its immediate surroundings. Further literature also identified the positive relationship between housing satisfaction and residential satisfaction (Galster, 1987) and life satisfaction (Quang Tran et al, 2018; Golant, 1982). The opposite effects are noticed as well, where general features of life are normally associated to life satisfaction, affect housing or residential satisfaction. Some of these effects are potentially traceable to self-reported health status, as it has far-reaching consequences on residents' perception of their homes according to Dunn and Hayes (2000). Ren et al (2018) also note that poor health negatively affects tenure, which on its turn is related to lower household satisfaction levels (Hoekstra et al, 2005; Balestra et al, 2013, Ren et al, 2018), thus potentially affecting housing, residential and life satisfaction.

3.3 Empirical overview

Most empirical research towards household satisfaction uses one or multiple theories discussed in section 3.1. These studies take the socio-demographic composition of the household into consideration, as well as the characteristics of the house. Depending on the level of household satisfaction, some neighbourhood characteristics and general features of life are considered as well (Lu, 1999). This section establishes an overview of empirical research on housing satisfaction, residential satisfaction and life satisfaction.

[4] The terms life satisfaction, subjective wellbeing and happiness are used interchangeably throughout literature (Ren, Folmer & Van der Vlist (2018), Quang Trann & Van Vu (2018)). This research uses life satisfaction.

A pan European research conducted by Elsinga et al (2005), examined data from the European Community Household panel to identify potential effects of tenure on housing satisfaction. While controlling for housing quality, household characteristics and housing costs, their results indicate that across Europe home-ownership tends to increase housing satisfaction, compared to renting. Similar findings have been made by Golant (1982), who examined over 400 US citizens aged 60 years and older and identified that homeownership and longer length of residence turned out to be indicators of higher housing satisfaction. Notably, Golant (1982) also identified the positive impact housing satisfaction has on life satisfaction, though simultaneously critically noticed the complex phenomenon of satisfaction levels, because (dis)satisfaction levels are not likely to be the result of solely objective defined housing deficiencies. As subjectivity is also part of household satisfaction, the results should be interpreted with caution. Another research on housing satisfaction has been conducted by Galster (1987), who ran a multiple regression analysis using data on homeowners in Minneapolis. Adequate interior space and plumbing facilities turned out to be strongly related to housing satisfaction, though the extent of it differs per type of household. However, Galster (1987) identified that the effects of interhousehold differences as well as the potential impact of the site, neighbourhood and social contacts have not been taken into consideration.

Previously, Galster conducted a research that did include contextual factors such as neighbourhood characteristics, home and compositional characteristics of the household are taken into consideration (Galster & Hesser, 1981). They used a stratified regression model to examine the determinants of residential satisfaction in Ohio, based on data from 1975. They conclude that objective compositional characteristics of individuals and objective contextual characteristics of the individual's homes and surrounding positively correlate to their housing satisfaction. Earlier research of residential satisfaction in the Netherlands has been conducted by Grigolon et al (2014) applying a logistic regression model to assess the housing satisfaction in the Netherlands. Various socio-demographic, urban setting, housing attributes, accessibility and frequency of social contacts are drawn from the WoON survey data from 2012. They conclude that housing satisfaction decreases in areas that have a higher urbanity rate, while a good social environment increases housing satisfaction. Additionally, increased household satisfaction levels are reducing residential mobility. Huang & Du (2015) examined residential satisfaction in public housing in China. They examine residential satisfaction via various impact factors including; housing characteristics, neighbourhood characteristics, public facilities, social environment, household characteristics and housing allocation institutions. Using an ordered logit model, they examined housing satisfaction in Hangzhou (China) and concluded that neighbourhood environment, public facilities and housing characteristics are the prominent influencers of residential satisfaction.

Ren, Folmer and van der Vlist (2016) examined life satisfaction in Urban China. They specifically examined the impact of tenure, while controlling for income, household size, age, education, health, employment status and region in their ordered probit regression. Ren et al (2016) place a critical note towards some methodological issues found in previous academic research towards satisfaction levels, as households tend to overestimate their satisfaction levels. Ren et al (2016) solve this overestimation issue by using a Propensity Score Matching (PSM) method. Consistent with existing literature their research indicates that home ownership increases households' life satisfaction (Balestra et al, 2013), though the actual implications differ per income group. The middle and upper class have the strongest positive impact of home ownership on life satisfaction. They assume that the limited impact within the lower income groups, might be the result of the financial burdens related to home ownership.

Balestra et al (2013) confirmed these findings but also identified that home ownership affects satisfaction levels due to other positive externalities, such as investments made in social relationship with the neighbours and local politics. Balestra et al (2013) conclude that household characteristics regarding socio-demographics such as age, gender and education play a secondary role whenever the dwelling and neighbourhood features are controlled for. Tenure, shortage of space, housing quality and affordability, neighbourhood crime rates and access to public transport are noted as the most influential factors of household satisfaction levels. Fernandez' research (2017) on satisfaction levels of older people in urban areas, confirmed that both housing and neighbourhood quality strongly affects satisfaction levels. Also, the subjective extent to which the house meets their demands, potentially affects residential satisfaction positively thus affecting satisfaction with life in general. Quang Tran & Van Vu (2018) examined the drivers of housing satisfaction and their relationship to life satisfaction among Vietnamese seniors. They confirmed that housing satisfaction is a strong indicator for life satisfaction, especially for those who live in permanent houses. Though the effects of housing satisfaction on life satisfaction is likely to be underestimated due to endogeneity issues. Their results indicate that housing quality and amenities, income and health positively affect life satisfaction. Rural inhabitants are more satisfied in general and widowed households experience lower satisfaction levels with life in general. Appendix A provides a full overview of the variables considered in the discussed literature towards household satisfaction.

3.4 Hypotheses

Research towards the effects of green energy labels on household satisfaction levels is new. However, extensive research has been done on household satisfaction. Multiple studies indicate that owner-occupancy positively affects household satisfaction levels (Elsinga et al, 2015; Ren et al, 2018, Grigolon,

2014; Balestra et al, 2013). Also, literature suggests that residential satisfaction is positively related to satisfaction with life in general (Balestra & Sultan, 2013; Golant, 1982; Fernández et al, 2017; Quang Tran et al, 2017). Based on the research direction and existing literature discussed in the previous sections, the following research hypothesis have been drawn up:

Hypothesis 1: Green energy labels are positively affecting household satisfaction.

Hypothesis 2: Green energy labels are increasingly positively affecting household satisfaction levels throughout time.

Hypothesis 3: Homeownership positively affects the impact of green energy labels on household satisfaction;

Hypothesis 4: The impact of green energy labels is equal for residential and life satisfaction;

4. DATA & METHODOLOGY

4.1 Data set

Previously published academic research towards household satisfaction levels in the Netherlands. Grigolon et al (2014) for instance made use of one recent WoON survey data, containing data on satisfaction, energy labels and nearly all control variables. This research however, examines data from various WoON household surveys in order to assess the effects throughout time. The WoON survey is carried out every three years since 2006 in consultation between the Dutch Central Bureau of Statistics (CBS) and the Ministry of Internal Affairs (BZK). The aim of these surveys is to gather statistical information on the living situation of households in the Netherlands. A minimum of 60,000 households is selected using a stratified sample with national coverage of municipalities to provide representative, reliable statistical information similar to previous years of surveying. The actual household sample varies per survey cycle, assuring the representativity and it ensures that, despite the use of multiple data cycles, the research remains a cross-sectional setting. The surveys contain information on households based on individuals of 18 years and older, regarding their household composition, dwelling characteristics, neighbourhood characteristics, housing expenses and desires and moves (CBS, n.d.). As it stands for a representative sample of Dutch households from a reliable and respected source these datasets are very suitable for conducting this research while it contains nearly all relevant variables. Moreover its suitability is endorsed by the fact that previous research towards housing satisfaction within the Netherlands, such as Grigolon et al (2014), also made use of WoON survey data.

As the WoON data are available in cycles of three years, the data are suitable for running a Likelihood ratio test, dividing the data in three groups: 2012, 2015 and 2018. Earlier cycles lacked information on the independent variable of interest, as in 2009 energy labels were not yet implemented. Also, the 2021 WoON survey has not yet been published at the time of research, so it is not possible to include more recent data. Appendix A displays the data limitations. The effects of green energy labels on life satisfaction is solely examined using the WoON 2015 cycle.

As established in chapter two, the policy intervention of the government by implementing the energy label and its increasingly strict legislation did result in increasing numbers of energy labels since 2008 but hardly enforced until 2015. Unsurprisingly, the data towards energy labels in the various WoON surveys varies.

The WoON survey 2012 (conducted in 2011) is the first survey to include data on energy labels, consequently having less rich data due to the recent introduction of the label compared to the WoON

2015 or 2018 cycles. Nonetheless, the implications of the usable cases in the 2012 survey are still of interest. As this enables the research to establish the stability of the effects of green energy labels on household satisfaction levels through time. From a research perspective, this is of interest as it potentially offers insight in the relative importance of green energy labels on household satisfaction levels. The WoON 2012 data should be interpreted with caution, as the share of homes with energy labels is still limited and social housing is overrepresented due its pioneering role.

After dropping the cases without energy labels in all WoON surveys, it becomes clear that the implications of this research regarding the data of WoON 2012 should be taken into consideration with caution. Based on the data clearing described in appendix B, figure 4.1 compares the relative (over)representation of each type of homeownership between the WoON survey to the national average distribution based on CBS. The tenure of some cases in the dataset is unknown, therefore these cases cannot be attributed to owner-occupiers, social tenants or private tenants thus got excluded from the analysis. The pioneering role of the housing associations results in a strong representation of housing associations in the sample of 2012 and generally spoken, this is likely to affect the type of households as well as housing associations are intended to provide housing for those with less income. More on this issue will be discussed in section 4.5.

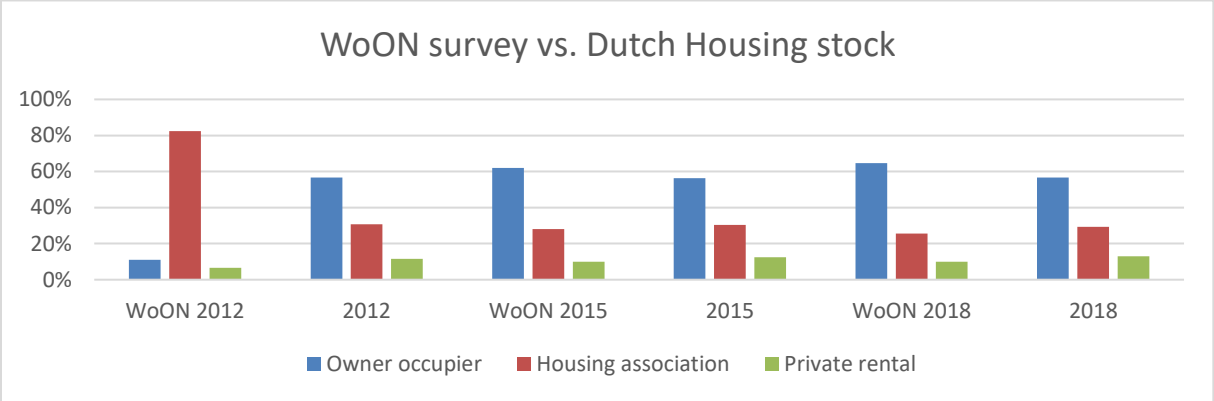


Figure 4.1: Comparison of relative shares of ownership (unknown tenure excluded) (own work, CBS, 2020[1]).

Despite the deviation from the total housing stock, the sample of WoON 2012 survey data may be considered as representative when the distribution is compared against the distribution of preliminary energy labels in 2011[5] as displayed in figure 4.2:

[5] Note that the WoON 2012 survey data is collected at the end of 2011 until the beginning of 2012.

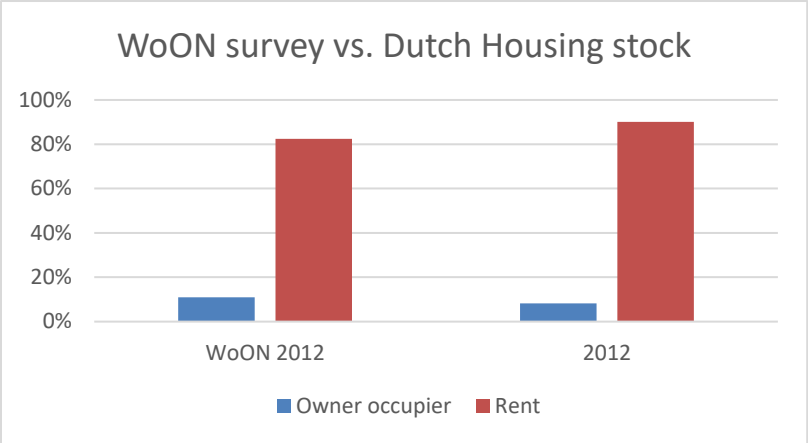


Figure 4.2: Comparison of relative share of ownership after accounting for energy labels (Own work, CBS, 2012).

Other possible options for dealing with this not representative issue are considered less desirable. For instance, solely examining the effects of green labels on levels of household satisfaction in housing associations could potentially cause bias, since it has already been extensively described in the literature and it is empirically substantiated that homeownership often correlates positively with household satisfaction levels. Weights in order to regain representativity is deemed undesirable, as the interpretation becomes more complicated and the representation with reality of that time becomes obscured. Also, excluding the 2012 WoON sample limits the consideration period too much, while the WoON 2021 is not published yet.

4.3 Operationalization of variables

Appendix A provides an overview of the presence of the variables derived from literature. The dependent variable in this research examines the various levels of household satisfaction. Consistent with literature, housing satisfaction and residential satisfaction is measured using the variable that indicates the extent of satisfaction with the house. That extent is examined in the WoON surveys using the following question: *"How satisfied are you with your current home?"* [6] The respondent is able to answer this question on a five-point Likert-type scale, indicating either to be very satisfied (1) satisfied (2) neither satisfied nor dissatisfied (3), dissatisfied (4) and very dissatisfied (5). Table 4.1 provides an overview of the satisfaction levels of the household per year, clearly indicating that the overwhelming majority is either satisfied or very satisfied. Very few observations indicate dissatisfaction or high dissatisfaction, which is consistent with arguments from Amerigo et al. (1990) and Ren et al. (2018) regarding the overestimation of satisfaction levels comparing their 'genuine feeling on satisfaction levels'. Whether or not these figures are representative is impossible to examine, as it would require

[6] See variable Twoning (16.1) in WoON 2012, Twoning (13.1) in WoON 2015 and Twoning (12.1) in WoON2018.

the examination of the satisfaction levels of each and every household in the Netherlands. WoON surveys have been conducting large-scale surveys of the Dutch population for years in consultation with the central statistical office, thus these surveys can be regarded as the best possible approximation of the true reflection. The satisfaction levels shown below in table 4.1 refer to the variable satisfaction with the home, which functions as a dependent variable for housing and residential satisfaction.

SATISFACTION LEVEL	**WOON 2012**			**WOON 2015**			**WOON 2018**		
	Freq.	*Percent*	*Cum.*	*Freq.*	*Percent*	*Cum.*	*Freq.*	*Percent*	*Cum.*
VERY SATISFIED	26.750	44,83	44,83	22.242	41,15	41,15	22.913	40,05	40,05
SATISFIED	26.564	44,52	89,35	25.306	46,82	87,96	26.512	46,34	86,39
NEITHER SATISFIED NOR DISSATISFIED	4.436	7,43	96,78	4.769	8,82	96,79	5.739	10,03	96,42
DISSATISFIED	1.419	2,38	99,16	1.277	2,36	99,15	1.520	2,66	99,08
VERY DISSATISFIED	501	0,84	100,00	460	0,85	100,00	527	0,92	100,00
TOTAL	59.670	100,00		54.054	100,00		57.211	100,00	

Table 4.0-1: Frequency table dependent variable: Satisfaction level

In line with earlier research towards household satisfaction in the Netherlands conducted by Grigolon et al (2014), this research examines the effects of green, energy efficient homes on household satisfaction using a binary logistic regression analysis. Due to the strong majority of the respondents in the WoON survey indicates either very satisfied or satisfied, the variable is transformed into a dichotomized variable. The binary variable splits the data into 'satisfied' (1) vs. 'not satisfied' (0), the latter being the base category. The dichotomization of the variable satisfaction is in line with research conducted by Dunn (2000), resulting in a category 'satisfied' containing all 'satisfied' cases that either scored 'very satisfied' or 'satisfied'. The base category 'not satisfied' contains all cases that not explicitly scored 'satisfied', thus 'neither satisfied nor dissatisfied, 'dissatisfied' and 'very dissatisfied'. The category 'not satisfied' thus includes the neutral category 'Neither satisfied nor dissatisfied' as well as the dissatisfied cases as all of these cases have not explicitly stated being satisfied. Exclusion of the neutral group is deemed undesirable as it would result in a too unbalanced distribution with over 96% scoring satisfied and less than 4% scoring not satisfied. The new distribution after merging the categories into 'satisfied' and 'not satisfied is represented in table 4.2:

SATISFACTION LEVEL	**WOON 2012**		**WOON 2015**		**WOON 2018**	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
SATISFIED (1)	53.314	89,35	47.548	87,97	49.425	86,39
NOT SATISFIED (0)	6.356	10,65	6.506	12,03	7.786	13,61
TOTAL	59.670	100,00	54.054	100,00	57.211	100,00

Table 4.0-2: Frequency table transformed binary dependent variable: Satisfaction level

Life satisfaction deviates from both housing and residential satisfaction as it is generally measured in regard to the overall satisfaction with life. Recent WoON surveys measure life satisfaction using the following question: "*All considered, how satisfied are you with your life in general?*" The respondent is asked to fill in a mark with what grade he or she scores on life at that point, ranging from 1 till 10 where values over 6 are considered as satisfied. However, data constraints prohibit from estimating the effects of green energy labels on life satisfaction, as the dependent variable is just recently added to the WoON survey. As a result, life satisfaction as a variable is not included in the WoON 2012 survey, but solely in the WoON 2015 and WoON 2018[7].

The lack of the life satisfaction variable in the WoON 2012 survey and the absence of more recent WoON surveys at the time of research, limits the consideration period on life satisfaction to two cycles. For similar reasons as discussed on the energy label data constraints, this is deemed insufficient to examine the effects and stability of green energy labels on life satisfaction over time. However, theory indicates the interconnection of the various levels of household satisfaction. Despite data constraints regarding the dependent variable of life satisfaction, the WoON 2012 survey data provides some insight in the most influential factors of life satisfaction namely health and residential satisfaction (Fernández et al, 2017). The main body of this research therefore focusses on the effects of green energy labels on housing and residential satisfaction levels, while controlling for health and employment status. To some extent, the inclusion of these variables in the socio-demographic control variables might control for the effects of life satisfaction on housing and residential satisfaction. Despite the inability to include these effects over time within the context of this research, these are still valid questions. Nonetheless, as far as the author knows no other studies have examined the effects of green energy labels on life satisfaction. Therefore, life satisfaction will be assessed separately solely using the WoON 2015 dataset, as it provides all the variables needed to assess life satisfaction.

The independent X-variable of interest relates to the energy labels. The WoON data surveys differ on the actual variables relating to the energy label, as they are conducted at the very first stages after its introduction. The WoON 2012 survey contains a variable regarding the preliminary energy label disclosed by AgentschapNL, whereas the surveys from 2015 and 2018 contain a preliminary energy label from the Governmental Service for Enterprising Netherlands (RVO) and a definitive energy label. As solely the provisional energy label is included in all three datasets, this research measures the effects of preliminary green energy labels. To enhance compatibility of the various WoON datasets, the energy label classification of the AgentschapNL is transformed, combining the labels A++ A+ and

[7] See variable Leven (30.1) in WoON 2015 and Leven (26.1) in WoON 2018.

An into one classification denoted as A. The ordinal variable regarding the energy labels in the various WoON data cycles all include 7 classifications, ranging from (very)energy efficient A till energy inefficient label G. The labels A, B and C are hereby regarded as green energy efficient labels, whereas labels D and worse are viewed as energy inefficient.

Appendix A&D provide an extensive overview of the variables present in the dataset as well as all relevant information regarding variable transformations to make the various datasets compatible. Most transformations have been processed in order to make the various cycles of WoON datasets compatible, as variable and label names deviate slightly between the surveys and categorical variables are grouped differently. In addition to the dependent y variable and independent variable x, approximately 21 independent Z-variables are included as control variables. These variables are derived from theory and can be grouped as one year control variables, socio-demographic variables, dwelling characteristics and neighbourhood characteristics. Year dummies are created in order to pool the data of the various WoON cycles, while being able to establish the effects of the variables per year.

The variables regarding the socio-demographic composition of the household include relevant control variables related to life satisfaction, as discussed earlier. The socio-demographic variables include variables on the age of the respondent, ethnicity, educational level, disposable household income, household type, number of persons in the household, number of years living at the current address, whether or not the respondent is recently moved or has desires to move and his/her self-reported health status.

The housing characteristics controls include variables on tenure, construction year, housing type, housing expenses, WOZ housing value, housing quality and living area. Theory also indicates the relevance of including a control variable regarding housing costs as a proportion to the household income. The WoONsurvey includes a relevant variable ‘woonquote’, however this variable is not present in each survey. Therefore, a self-constructed variable based on the housing expenses and the disposable household income is created, referred to as cost-to-income ratio and describes the monthly housing costs multiplied by twelve divided by the annual disposable household income. For similar reasons the variable person-space ratio is constructed referring to the amount of persons in the household divided by the living area. However, as both variables on the amount of persons and the living area are ordinal of scale, the outcome of such a constructed variable would therefore be meaningless in terms of interpretation. From a research perspective it would be of interest to examine the effects of familiarity with the energy label on satisfaction levels. The WoON survey provides a suitable variable for this as it asks respondents’ the following question: “*Are you familiar with the*

energy label"? However, it turns out that very few respondents filled out this question, as 2.093, 1.094 and 2.506 respectively answered the question in the WoON 2012, 2015 and 2018 survey. As each survey consists of roughly 60.000 respondents, the benefit of the inclusion of this variable were deemed insufficient in proportion to the loss of observations in the regression analysis.

Lastly the neighbourhood characteristics include control variables on neighbourhood satisfaction, neighbourhood attachment, relationship with neighbours, region, urbanity and satisfaction with the population composition of the neighbourhood.

Appendix C provides an overview of the data cleaning. It includes the data loss due to transformed variables , missing values and outliers. The initial datasets of the WoON 2012, 2015 and 2018 contained 69.339, 62.668 and 67.523 observations respectively. After the data cleaning process, the combined dataset used for the assessment of housing and residential satisfaction consists of 116.920 observations. Life satisfaction is, after data clearing, assessed with 50.165 cases dating from 2015.

4.4 Methodology

In order to examine the relationship between energy labels and household satisfaction levels, this research makes use of a binary logistic regression analysis. The binary logistic regression analysis is a suitable method whenever the dependent variable is dichotomous (DeTrain, 2009). Similar studies towards household satisfaction in the Netherlands also used the logistic regression (Grigolon et al, 2014). However, as satisfaction is measured as a Likert scale variable, this entails transforming the variable into a dichotomous, binary variable resulting in the loss of information. Other methodologies such as the ordered logistic regression were explored, but the proportional odds assumption got violated. The usage of the binary logistic regression is thus despite its information loss a necessary but accepted limitation of the study.

Other studies towards household satisfaction levels used simple linear or multiple regression analysis or the ordered logistic regression. As the dependent variables are all Likert-scale variables, the latter methodology is preferred (Lu, 1999). However, some drawbacks of this approach have been identified as well as some existing literature criticizes the use of scale-dependent variables Likert-type scales with regard to household satisfaction levels. Amerigo & Aragonés (1990) for example, identified potentially skewed distribution of satisfaction levels when used to examine household levels, as respondents tend to overestimate their satisfaction levels. This potentially overestimation might cause upwards biased results as noted by Ren, Folmer and Van der Vlist (2018). They correct the potentially upwards biased

results by using a propensity score matching (PSM) system to correct their models with the 'genuine' satisfaction levels. However, this ordered logistic regression analysis is not suitable for this analysis, as it turned out that the proportional odds assumption (brant test) got violated. More on this in the reflection on this research in chapter 6.

The dependent variable assessing the likelihood of household satisfaction, ranges from the base category (0) dissatisfied up to satisfied (1). The regression is estimated using the following equation:

$$y = ln \frac{p_{satisfied}}{p_{not\ satisfied}}$$

$$= \beta_0 + \beta_1 X_{Energy\ label} + \beta_2 X_{Year} + \beta_2 X_{socio-demographics} + \beta_3 X_{housing\ characteristics} + \beta_4 X_{neighbourhood\ characteristics} + \varepsilon$$

Where β_0 is the constant, the error term of the model is denoted as ε. The equation entails the variable of interest Energy label as well as control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics. Table 4.3 below describes the summary of the explanatory variables per group. All variables are derived from literature and adjusted to match between the various datasets of 2012, 2015 and 2018, see appendix A and D for a more extensive overview.

Variables	Operationalization
Energy label	
Energy label	1 if A, A+ or A++, 2 if B, 3 if C, 4 if D, 5 if E, 6 if F, 7 if G (base)
Year	
Year 2012	1 if 2012, 0 if other
Year 2015	1 if 2015, 0 if other
Year 2018	1 if 2018, 0 if other
Socio-demographics	
Age	1 17-34 year, 2 if 35-65 years (base), 3 if Over 65 year
Ethnicity	1 if Native (base), 2 if Non-Western, 3 if Western
Education	1 if Low educated (up to VMBO) (base), 2 if Secondary educated (up to MVO), 3 if Well educated (up to university)
Disposable HH income	1 if up to 35.000, 2 if 35.001 up to 70.000 (base), 3 over 70.000
Household type	1 if Single person household (base), 2 if Couple, 3 if Parent(s) with kid(s), 4 if Non family household
Number of persons in the household	1 if 1 person, 2 if 2 persons, 3 if 3 persons, 4 if 4 persons, 5 if 5 or more persons
Number of years living at this address	Measured in years
Self-reported health-status	1 if Very well (base), 2 if Good, 3 if Ok, 4 if Differs from time to time, 5 if Bad
Housing characteristics	
Tenure	1 if Owner-occupier (base), 2 if Social rent, 3 if private rent
Construction year	Building age measured in years
Housing type	1 if Single-family home (base), 2 if multi-family home
Cost-to-income	Housing expenses divided by disposable household income (yearly)

Housing expenses	Total mortgage or rent cost per year in euro's
WOZ-housing value	Tax value measured in euro's
Poor housing quality	1 if Agreed (base), 2 if Disagreed
Person-space ratio	Number of persons in the household divided by living area
Living area	1 if up to 69 m2 (base), 2 if 70 up to 119 m2, 3 if 120 m2 or over
Neighbourhood characteristics	
Neighbourhood satisfaction	1 if Very satisfied (base), 2 if Satisfied, 3 If neither satisfied nor dissatisfied, 4 if Dissatisfied, 5 if Very dissatisfied
Good relationship with neighbours	1 if Not agreed, 2 if Agreed
Region (G4)	1 if Other municipalities (base), 2 if Big 4 (Adam, Rdam, Utrecht, Hague)
Urbanity	1 if Strong (base), 2 if Moderately, 3 if Little to none

4. 0-3: Variable overview (own work)

The model fit of each model is tested to the asses the added value of each set of control variables using the iterative estimation process of the maximum likelihood. The Chi-square test uses the outcome of the likelihood ratio test and the degrees of freedom of the model to estimate if the fitted model is better than the empty model. McFadden's Pseudo R-squared is also determined, as it provides an insight on the amount of variance explained by the independent variables in the model. The binary logistic regression estimates the probability of a certain outcome over its alternative outcome, expressed as the logarithm of the odds. The probability of being either satisfied or not satisfied is mutually exclusive, as each observation falls within either one. In terms of interpretation, this means that an increase of one unit in X1 results in the increase of the ln(odds) with b1.

The binary logistic regression is estimated under certain model assumptions. The dependent variable needs to be binary and it cannot contain multicollinearity or influential outliers. The observations need to be cross sectional, as each observation needs to be independent from each other. The error term needs to be uncorrelated and has a logistic distribution. The data meet the first requirement, as the dependent variable is measured as satisfied (1) or not satisfied (0). The independent variable of interest is based on an ordinal scale, as it describes energy labels ranging from energy efficient green homes labelled A, B or C, towards energy inefficient homes with labels D, E, F and G. Multicollinearity within the dataset is tested using a correlation matrix and Variance Inflation Factors (VIF). Correlations over 0.5 should be interpreted with caution, while correlations over 0.7 are clear indications of multicollinearity. The correlation matrix indicates that the variables number of persons in the household and household type are correlated (0,7581). Also, eight more variables show signs of potential multicollinearity as their values range between 0.5 and 0.7. In order to assess multicollinearity in more debt, a linear regression has been estimated in order to establish the Variance Inflation Factors (VIF). VIF of over 10 are problematic, as these values clearly indicate multicollinearity issues, while more strict researches indicate that VIF over 5 will cause multicollinearity issues. VIF

values indicate multicollinearity issues for categories in the variables household type, number of persons in the household, housing quality and living area. Appendix E presents an overview of both the correlation matrix and the VIF values.

In order to resolve these multicollinearity issues we first rearranged some of the categorical variables. The variable living area is recoded, merging the categories 50 m2 and 50-59 square metres into 'less than 69 m2', merging 70-89m2 and 90-119m2 into '70-119m2' and merging 120-149, 150-199m2 and 200m2 or more into 'more than 120m2'. Housing quality is recoded and merged from a five-point Likert scale variable indicating whether the inhabitant agreed or disagreed with the following statement; "the house is poorly maintained", where completely agreed and agreed merged into 'agreed' (base) and the others into 'disagreed'. Lastly, the categorical variable household type aggregated couple with kid(s) and single parent family into parent(s) with kids. Disposable household income has been categorised into low, middle and high income, referring to less than €35.000, €35.000 up to €70.000 and over €70.000 a year. However, including all 25 variables in the model would result in a lot of parameters. The decision is made to reduce the number of estimated parameters in the control variables, by aggregating some categories. The merged variables will still serve their control purpose but it limits the model size. Relationship with the neighbours is estimated using the statement "I have lots of contact with my neighbours". "Completely agree" and "Agree" are merged into "Agreed", while the other variables neutral, disagree and completely disagree are merged into "Not agreed". Age is aggregated from 7 into 3 categories, being 17 till 34-year-olds, 35 till 65-year-olds and 65 years and older. Urbanity is aggregated from 5 categories into 3, grouping "Strong", "Moderate" and "Little to none urbanity". These changes resulted in the VIF values lower than 5 for nearly all variables. A total of three categories from household type, number of persons in the household and living area exceed the VIF of 5, but are lower than 10 thus acceptable as these are merely controls. All variables of interest have lower VIF values and the mean VIF is 2.37.

The binary dependent variable and the ordinal independent variable of interest do not contain outliers. Some of the independent control variables have been adjusted in order to delete improbable values and outliers. Negative values for housing costs are deleted and values over €60.000,- have been deleted as outliers. Also, observations with a disposable household income below the social assistance rate of 2012 have been removed from the dataset (see appendix C). Another assumption of the binary logistic regression relates to the independence of each observation, as the error terms need to be uncorrelated. This condition is met, as the WoON survey samples differ per cycle, resulting in a cross-sectional research setting. Lastly, the binary logistic regression assumes a logistic distribution of the error term related to the assumption of the uncorrelated error terms DeMaris (1995).

4.5 Descriptive statistics

The descriptive statistics depicted in table 4.4, providing an overview of all the variables included in the analysis. As discussed in section 1.4, housing and residential satisfaction is assessed using three WoON surveys, whereas life satisfaction is assessed solely using the WoON 2015 dataset as a consequence of data constraints. The dependent variables on housing (H) and residential (R) as well as life satisfaction (L) are included, while all independent and control variable relate to the pooled dataset.. A total of 116.920 cases are observed, of whom 15.050 date from 2012, 50.165 from 2015 and 51.705 date from 2018. Table 4.5 relates to the descriptive statistics of the WoON 2015 dataset regarding life satisfaction, containing a total of 50.165 observed cases. The min and max values clearly indicate the binary or ordinal properties of most of the variables included.

Variable	Obs	Mean	Std. Dev.	Min	Max
Satisfaction with the home (H&R)					
Not satisfied	116,920	.1322614	.3387762	0	1 = Yes
Satisfied	116,920	.8677386	.3387762	0	1 = Yes
Satisfaction with life (L)					
Not satisfied	50,165	.9725905	.1632752	0	1 = Yes
Satisfied	50,165	.0274095	.1.632752	0	1 = Yes
Energy label					
A	116,920	.0829371	.2757882	0	1 = Yes
B	116,920	.1481611	.3552612	0	1 = Yes
C	116,920	.3038915	.4599383	0	1 = Yes
D	116,920	.0972374	.2962821	0	1 = Yes
E	116,920	.1394885	.3464571	0	1 = Yes
F	116,920	.1097674	.3126009	0	1 = Yes
G	116,920	.1185169	.3232206	0	1 = Yes
Year					
2012	116,920	.1287205	.3348918	0	1 = Yes
2015	116,920	.4290541	.4949432	0	1 = Yes
2018	116,920	.4422255	.496653	0	1 = Yes
Age					
17-23 years	116,920	.1590917	.3657631	0	1 = Yes
35-64 years	116,920	.5473059	.4977593	0	1 = Yes
65 years and older	116,920	.2936025	.4554139	0	1 = Yes
Ethnicity					
Native	116,920	.8383082	.3681694	0	1 = Yes
Non-Western	116,920	.0713907	.2574774	0	1 = Yes
Western	116,920	.0903011	.2866138	0	1 = Yes
Education					
Low educated	116,920	.3420287	.4743912	0	1 = Yes
Secondary educated	116,920	.32961	.4700735	0	1 = Yes
Well educated	116,920	.3283613	.469619	0	1 = Yes
Disposable HH income					
Low	116,920	.5467157	.497815	0	1 = Yes
Middle	116,920	.3917722	.4881483	0	1 = Yes
High	116,920	.0615121	.2402684	0	1 = Yes
Household type					
Single person HH	116,920	.3208861	.4668191	0	1 = Yes
Couple	116,920	.3320647	.4709561	0	1 = Yes
Parent(s) with kid(s)	116,920	.327027	.4691292	0	1 = Yes
Non family household	116,920	.0200222	.1400768	0	1 = Yes
Number of persons in the HH					
1 person	116,920	.3208861	.4668191	0	1 = Yes
2 persons	116,920	.3785323	.4850233	0	1 = Yes
3 persons	116,920	.1216558	.3268893	0	1 = Yes

4 persons	116,920	.1270527	.3330334	0	1 = Yes
5 or more persons	116,920	.0518731	.2217717	0	1 = Yes
Years at current address	116,920	1.633.194	1.372.076	0	93
Self-reported health status					
Very well	116,920	.1914814	.3934686	0	1 = Yes
Good	116,920	.5437393	.4980853	0	1 = Yes
Ok	116,920	.1512145	.3582594	0	1 = Yes
Differs from time to time	116,920	.0761974	.2653148	0	1 = Yes
Bad	116,920	.0373674	.1896613	0	1 = Yes
Tenure					
Owner-occupier	116,920	.5862385	.4925089	0	1 = Yes
Social rent	116,920	.3320219	.4709408	0	1 = Yes
private rent	116,920	.0817397	.2739688	0	1 = Yes
Age of the building	116,920	4.793.659	4.351.139	0	1013
Housing type					
Single-family home	116,920	.6978276	.4592015	0	1 = Yes
Multi-family home	116,920	.3021724	.4592015	0	1 = Yes
Cost-to-income-ratio	116,920	.2903167	.1608869	.0000575	5.375.046
Housing expenses	116,920	9.226.598	4.845.393	3.2	59930.94
WOZ housing value (tax value)	116,920	213345.4	125237.5	10000	3494000
Poor housing quality					
Agreed	116,920	.096844	.2957465	0	1 = Yes
Disagreed	116,920	.903156	.2957465	0	1 = Yes
Person-space-ratio	116,920	.5898938	.3754629	.1428571	5
Living area					
Up to 69m2	116,920	.142482	.3495454	0	1 = Yes
70 up to 119m2	116,920	.4830226	.4997138	0	1 = Yes
Over 120m2	116,920	.3744954	.4839944	0	1 = Yes
Neighbourhood satisfaction					
Very satisfied	116,920	.3059613	.4608154	0	1 = Yes
Satisfied	116,920	.522742	.4994847	0	1 = Yes
Neither satisfied, nor dissatisfied	116,920	.1139241	.3177204	0	1 = Yes
Dissatisfied	116,920	.0439446	.2049727	0	1 = Yes
Very dissatisfied	116,920	.013428	.115099	0	1 = Yes
Good Relationship neighbours					
Not agreed	116,920	.4897793	.4998977	0	1 = Yes
Agreed	116,920	.5102207	.4998977	0	1 = Yes
Region (G4)					
Other municipalities	116,920	.886401	.317325	0	1 = Yes
Big 4 (Adam, Rdam, Utrecht, Hague)	116,920	.113599	.317325	0	1 = Yes
Urbanity level					
Strong	116,920	.5090575	.4999201	0	1 = Yes
Moderately	116,920	.1856911	.3888588	0	1 = Yes
Little to none	116,920	.3052515	.4605158	0	1 = Yes

4.4: Descriptive statistics

5. RESULTS

The chapter contains the results of the various logistic regression models, as discussed in section 1.4, relating to housing, residential and life satisfaction. Both housing and residential satisfaction will be assessed using multiple datasets, whereas life satisfaction solely examines one WoON survey dataset. Section 5.1 establishes the robustness of the various models and indicates which models will be used for the in dept analysis of housing and residential satisfaction. Both housing (section 5.2) and residential satisfaction (section 5.3) include full pooled model, as well as a segmented and interaction model based on time (WoON 2012, 2015 and 2018) and tenure (owner-occupied, social rent and private rent). Section 5.4 relates to life satisfaction and is assessed separately as it solely assess the WoON 2015 data.

5.1 Robustness models

The results of the binary logistic regression for the pooled analysis are presented in table 5.1. As discussed in the previous chapters, both housing and residential satisfaction use the dependent variable satisfaction with the home. Thus the regression models regarding both housing and residential satisfaction are estimated using the dependent variable Satisfaction dummy and the independent variable energy label. Model I solely contains the effects of the independent variable energy label on the dependent variable satisfaction. Whereas model II, III, IV and V have added controls for the different years, socio-demographics, housing characteristics and neighbourhood characteristics.

The log likelihood and the degrees of freedom in each model are used to conduct the likelihood ratio test in order to assess whether the added variables improve the model fit significantly. The likelihood ratio tests indicates that each model is a significant improvement (appendix G), as can be seen by the increased pseudo R^2 explaining more variance throughout the models. Model IV represents housing satisfaction and explains 20.8% of the variance, whereas model V represents residential satisfaction explaining 28.5% of the variance in the level of household satisfaction of households. Despite the clear interlinkages of energy labels with housing characteristics, model V indicates higher odds of being satisfied with green labels compared to model IV. Thus, energy labels are significantly affecting the likelihood of both housing and residential satisfaction, though the impact of green energy labels is stronger on the latter. Section 5.2 will further elaborate on housing satisfaction based on model IV, whereas section 5.3 assess residential satisfaction based on model V.

Satisfaction dummy (0 = NS, 1 = S)	Model I	Model II	Model III	Model IV	Model V
Energy label					
A	2.100***	2.093***	2.416***	1.506***	1.734***
	(0.0983)	(0.0980)	(0.118)	(0.0870)	(0.105)
B	1.674***	1.720***	1.689***	1.245***	1.335***
	(0.0613)	(0.0631)	(0.0646)	(0.0582)	(0.0656)
C	1.133***	1.182***	1.249***	1.133**	1.219***
	(0.0335)	(0.0351)	(0.0389)	(0.0451)	(0.0511)
D	1.079*	1.267***	1.185***	0.987	1.023
	(0.0403)	(0.0489)	(0.0482)	(0.0474)	(0.0517)
E	0.496***	0.520***	0.696***	0.904**	0.988
	(0.0153)	(0.0161)	(0.0228)	(0.0355)	(0.0413)
F	1.099**	1.143***	1.107**	0.982	1.019
	(0.0399)	(0.0416)	(0.0420)	(0.0432)	(0.0470)
G	BASE	BASE	BASE	BASE	BASE
Year controls	No	Yes	Yes	Yes	Yes
Socio-demographics	No	No	Yes	Yes	Yes
Housing characteristics	No	No	No	Yes	Yes
Neighbourhood characteristics	No	No	No	No	Yes
Constant	6.305***	4.155***	9.899***	3.438***	7.852***
	(0.156)	(0.140)	(0.575)	(0.310)	(0.779)
Observations	116920	116920	116920	116920	116920
Log likelihood	-44575.003	-44382.441	-40984.716	-36153.697	-32671.842
df	6	8	25	36	44
Pseudo R^2	0.024	0.028	0.103	0.208	0.285

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable Energy label with label G as base category. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively. Inclusion of control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics denoted with yes/no. Appendix F provides a full overview of the control regression models including the control variables.

Table 5.1: Robustness logistic regression models

5.2 Housing satisfaction

The results of the pooled model including all control variables on housing satisfaction is displayed in table 5.2 and relates to model IV in table 5.1. The pooled effects of green energy labels on housing satisfaction, as well as the extent to which these regression results are in line with existing literature will be discussed below.

Housing satisfaction (Model IV)			
Satisfaction dummy (0 = NS, 1 = S)		**Number of persons in the HH**	
A	1.506***	Number of persons in the household	1.031
	(0.0870)		(0.0211)
B	1.245***	**Amount of years lived at this address**	
	(0.0582)	Amount of years lived at this address	1.003**
C	1.133**		(0.000958)
	(0.0451)	**Self-reported health status**	
D	0.987	Very well	BASE
	(0.0474)		
E	0.904**	Good	0.755***
	(0.0355)		(0.0219)
F	0.982	Ok	0.482***
	(0.0432)		(0.0171)
G	BASE	Differs from time to time	0.428***
			(0.0172)
Year		Bad	0.363***
Year=2012	BASE		(0.0175)
		Tenure	
Year=2015	0.830***	Owner-occupier	BASE

	(0.0262)		
Year=2018	0.622***	Social rent	0.403***
	(0.0197)		(0.0112)
Age		private rent	0.387***
17-23 year	0.918**		(0.0137)
	(0.0258)	**Construction year**	
35-64 year	BASE	Age of the building	1.000
			(0.000258)
65 years and older	1.799***	**Housing type**	
	(0.0516)	Single-family home	BASE
Ethnicity			
Native	BASE	Multi-family home	0.939*
			(0.0247)
Non-Western	0.651***	**Cost-to-income ratio**	
	(0.0204)	Yearly housing expenses / Yearly Disposable HH income	1.523***
Western	0.806***		(0.150)
	(0.0256)	**Housing expenses**	
Education		Housing expenses	1.000
Low educated	BASE		(0.00000405)
		Housing value	
Secondary educated	0.862***	Housing value (WOZ tax value)	1.000***
	(0.0212)		(0.000000179)
Well educated	0.846***	**Poor housing quality**	
	(0.0239)	Agreed	BASE
Disposable HH income			
Low	0.880***	Disagreed	5.275***
	(0.0278)		(0.126)
Middle	BASE	**Person-space-ratio**	
		Persons in HH / Living area	0.669***
High	1.144*		(0.0262)
	(0.0742)	**Living space**	
Household type		up to 69m2	BASE
Single person HH	BASE		
		70 up to 119m2	1.042
Couple	1.010		(0.0356)
	(0.0327)	over 120m2	1.188***
Parent(s) with kids	0.744***		(0.0585)
	(0.0360)		
Non family household	1.002	**Constant**	1.633
	(0.0694)		(0.817)
Observations	116920	**df**	36
Log likelihood	-36153.697	**Pseudo R^2**	0.208

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable Energy label with label G as base category. Inclusion of control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics denoted with yes/no. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Table 5.2: Pooled regression model IV: Housing satisfaction

Table 5.2 indicates that the independent variable, energy label are partly significant. Labels A and B are statistically significant at a 1% level and label C at 5%. Whereas the labels D and F are not significantly different from the base category label G. The odds ratio for Label A, B and C indicate that these households are 1.506, 1.245 and 1.133 times more likely to be satisfied compared to households with energy label G, *ceteris paribus*. This translates into a 50.6%, 24.5% and 13.3% higher housing satisfaction with energy label A, B or C respectively. Notably, label E is also statistically significant at a 5% rate, though the odds ratios is smaller than 1, indicating that label E results in a lower likelihood on

satisfaction compared to label G. The variables label D and F are insignificant, thus these households are not significantly more likely to be satisfied (or dissatisfied) compared to households with energy label G. As the odds ratio's are declining with less energy efficiency, the opposite reasoning is examined as well, see appendix H. With label A as a base category, all other labels are significant with odds increasingly lower than one, indicating that les efficient label is also negatively affecting housing satisfaction levels. Households with inefficient energy labels are less likely to be satisfied. In other words, inefficient homes make it more likely to be less satisfied. These results might relate to the fact that energy labels also serve as a representation of energy costs and living comfort (Ebrahimigharehbaghi et al, 2019; Wong, 2020).

These findings are in line with earlier results on research towards the impact of sustainability measures or green features on satisfaction levels. As Goodwin (2011), Johnson (2014), and Tan (2014) found that specific green features might enhance satisfaction levels. Plausibly, a sustainable home with a green energy label will have more of these features compared to properties with an energy insufficient label.

These results indicate that green energy labels (A, B and C) positively affect the likelihood to be satisfied. Also, 'the greener' the energy label, the higher the likelihood of being satisfied becomes. Thus green energy labels are positively affecting household satisfaction levels, therefore the null hypothesis is rejected. Green energy labels positively affecting housing satisfaction.

The control variables included in the model are derived from literature. Appendix A provides an extensive overview of the variables examined in literature. Contrasting literature, the number of persons in the household, construction year and housing expenses are insignificant. Thus, based on this sample, these variables do not affect the likelihood of being satisfied or dissatisfied. The other variables are significantly affecting the likelihood of being satisfied as a household. The socio-demographic variable disposable household income is statistically significant at a 1% level, indicating that lower income results in significant lower likelihood of being satisfied as well as higher than average income resulting in higher likelihood of being satisfied. This finding is consistent with literature (Elsinga et al, 2015; Grigolon, 2014; Ren et al (2018), as nearly all studies towards household satisfaction report a positive relationship between income and satisfaction. Tenure is mostly cited as a significant contributor in assessing household satisfaction levels, Elsinga et al (2015), Grigolon (2014) and Ren et al (2018) state that homeownership increases the likelihood of being satisfied. The model results are in line with these statements, as the odds of being satisfied when living in a social or private rental home are smaller than one. Compared to owner-occupiers these households are less likely to be satisfied.

Using the regression results of model V as robustness check, the stability of this relationship over time will be examined by segmenting the regression into years. Table 5.2 shows the regression results of model IV (pooled model on housing satisfaction) as well as the regression results per year. The likelihood ratio test, used to examine whether the segmented models have a better fit is significant (appendix G), thus indicate an improved model fit by segmenting the regression analysis in years. Table 5.3 provides an overview of the earlier assessed Model IV as well the segmented version Model IV 2012, Model IV 2015 and Model IV 2018.

Housing satisfaction	Pooled model IV	2012	2015	2018
A	1.506***	1.295	1.430***	1.618***
	(0.0870)	(0.283)	(0.122)	(0.141)
B	1.245***	1.159	1.238**	1.282***
	(0.0582)	(0.163)	(0.0882)	(0.0909)
C	1.133**	1.140	1.216**	1.084
	(0.0451)	(0.139)	(0.0740)	(0.0657)
D	0.987	1.001	1.095	0.923
	(0.0474)	(0.120)	(0.0993)	(0.0729)
E	0.904**	0.929	0.932	0.853**
	(0.0355)	(0.115)	(0.0549)	(0.0517)
F	0.982	0.957	1.071	0.962
	(0.0432)	(0.125)	(0.0749)	(0.0634)
G	BASE	BASE	BASE	BASE
Year=2012	BASE			
Year=2015	0.830***			
	(0.0262)			
Year=2018	0.622***			
	(0.0197)			
Year controls	Yes	Yes	Yes	Yes
Socio-demographics	Yes	Yes	Yes	Yes
Housing characteristics	Yes	Yes	Yes	Yes
Neighbourhood characteristics	No	No	No	No
Constant	1.633	0.182	0.359	7.123*
	(0.817)	(0.364)	(0.250)	(5.768)
Observations	116920	15050	50165	51705
Log likelihood	-36153.697	-5998.0262	-14294.581	-15741.724
Df	36	34	34	34
Pseudo R^2	0.208	0.170	0.208	0.219

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable Energy label with label G as base category. Inclusion of control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics denoted with yes/no. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Table 5.3: Pooled regression model IV: Housing satisfaction segmented based on years

Table 5.3 shows the logistic regression results per year. Green energy labels are not significantly affecting the likelihood of a household satisfaction level in 2012, compared to households living in energy inefficient homes. In 2015 however, the results indicate that households with label A, B or C are significantly more likely to be satisfied compared to households with energy label G. The odds ratios van 1.430, 1.238 and 1.216 indicate that the more energy efficient a house is labelled, as it results in a 43%, 23.8% or 21.6% higher likelihood of satisfaction for labels A, B or C respectively. These odds ratio for green labels A and B are in 2018 1.618 and 1.282 respectively. Indicating a stronger effect of

the most energy efficient labels compared to the base category, label G. Label A increasingly has the strongest effect throughout time as it results in 61.8% higher likelihood of satisfaction in 2018, while strikingly the less green labels are becoming not as or even insignificant as time goes by.

The insignificant results for energy labels in 2012 might be explained by the fact that energy labels were just recently introduced and the average sustainability level of the housing stock was still low. As discussed in section 2.3, three years later in 2015 the amount of energy labels nearly doubled, increasing awareness of the average and personal sustainability level of a household's home. The effects of green energy labels throughout time potentially relates to a lagged market adjustment before it affects housing satisfaction levels. A reflection of the own label compared to the market average relates to one of the underlying mechanisms of household satisfaction levels, as households judge their personal housing conditions to norms in society (Mohit et al, 2014). Therefore, increased awareness and average sustainability levels of the housing stock could be instrumental to satisfaction levels for households with green energy labels throughout time. After all, the greenest energy labels are increasingly positively affecting household satisfaction, as the odds ratios for green labels for the greenest energy labels increase. However, the model includes merely three datapoints in nine years, thus the stability of energy label effects or trend indications should be considered carefully. Notably, the less energy efficient label C is insignificant for the 2018 sample, which might indicate that the relative difference with less efficient labels has decreased due to an increased average sustainability level. See section 2.3, indicating that more houses have been provided with energy labels A and B both in relative and absolute terms.

However, some theory suggests that the odds ratios of various models cannot be compared directly. Solely the sign and size of the effects should be taken into consideration. Mood (2010) for instance, states that the comparison of odds ratios across various logistic regression or sample groups might lead to unwarranted results. As the logistic regression model estimates the maximum likelihood method, therefore the likelihood function differs per regression model. However, other authors have reservations about this view. Kuha and Mils (2020) published an article in the leading magazine SMR, countering these statements and indicate that such comparisons are possible. Also, the segmentation method is not uncommon in academic work on household satisfaction, as various studies used segmented models to examine heterogeneity amount groups (Huang & Du, 2015; Olgun, 2020). However, the academic debate on this matter is still ongoing. Due to the lack of consensus in literature on this topic, this research opts for including an additional method enabling the author to examine the effects of green energy labels alternatively while it also contributes to the critical execution of the research. With the use of interaction variables, the impact of green energy labels throughout time as

well as the effects of tenure can be asses in one model without segmentation. However, the usage of interaction models has some drawbacks. Interacting each year with each energy label would result in an additional 21 variables (3 years times 7 labels) in addition to the separate year and energy labels variables themselves. In order remain overview the amount of interaction effects is reduced by aggregating various energy labels into green energy efficient labels and energy inefficient labels. As discussed in chapter 2, labels A, B and C are generally regarded as energy efficient, whereas labels D and worse are energy inefficient. Obviously, the aggregation of the various energy labels into groups results in significant information loss, though is considered more reasonable weighed against the reduction of interaction variables. Table 5.4 provides an overview of the interaction effects between the years and the energy efficiency of homes.

Housing Satisfaction	Pooled model VI HS with interaction effects
Satisfaction dummy (0 = NS, 1 = S)	
Year=2012	BASE
Year=2015	0.805***
	(0.0307)
Year=2018	0.610***
	(0.0234)
Energy efficient green label	1.186***
	(0.0574)
Energy inefficient	BASE
Year=2012 # Energy efficient green label	BASE
Year=2012 # Energy inefficient	BASE
Year=2015 # Energy efficient green label	1.082
	(0.0626)
Year=2015 # Energy inefficient	BASE
Year=2018 # Energy efficient green label	1.052
	(0.0598)
Year=2018 # Energy inefficient	BASE
Socio-demographics	Yes
Housing characteristics	Yes
Neighbourhood characteristics	No
Constant	1.464
	(0.694)
Observations	116920
Log likelihood	-36177.954
Df	33
Pseudo R^2	0.208

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable includes the years, energy labels and the interaction effects between energy efficient homes and years. Energy efficient homes have label A, B or C whereas energy inefficient homes have label D or worse. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Table 5.4: Pooled regression model IV: Housing satisfaction with interaction effects

The model results in table 5.4 indicate that the interaction effects between the energy label and the years are insignificant, thus there is no significantly different relationship between the energy efficient green label per year. The pseudo R2 has not increased compared to model IV and underlines these findings. The likelihood ratio test comparing model V and model VI with the interaction effects however, indicates a significant improvement of the model fit potentially due to significant differences in control variables.

It should be noted that the interaction variable uses the energy efficient green label variable, which is a merged variable consisting of energy labels A, B and C as energy efficient and D and worse as energy inefficient. Consequently, this variable is less specific compared to the segmented model that includes each label separately. The results of the interaction model in table 5.4 indirectly iterate the findings of the segmented model in table 5.3. The segmented model results indicate an increasing effect of the greenest energy labels, as the odds ratios are increasing. However, the 'least green' energy label C became insignificant. The latter is a consequence of the increasing average sustainability character of the housing market over time. The effects of label C compared to label G are thus levelling throughout time, while simultaneously the opposite is occurring for label A, as newly constructed home or built energy neutral (ready) denoted as label A++. it is not for nothing that the valuation of the energy labels is being redesigned per 2021 using the new NTA 8800 method (Energiedeskundig, n.d.; Lente-akkoord, n.d.), in order to recalibrate how 'green' energy labels are. Therefore, the greenest energy labels are increasingly significantly affecting the likelihood of being satisfied throughout time. The null hypothesis of a stable effect of energy labels throughout time can be rejected, as the greenest energy labels are increasingly positively affecting the likelihood of being satisfied as a household.

Multiple studies (Elsinga et al, 2015; Ren et al, 2018, Grigolon, 2014; Balestra et al, 2013) identified that owning a house is positively affecting household satisfaction levels. Though Balestra et al (2013) argue that the effects of tenure on residential satisfaction, depends on the income of the household as well. Households with lower incomes are less satisfied with an owner-occupied house compared to middle- and high-income groups. Nonetheless, the impact of green energy labels on housing satisfaction levels might be different per tenure. An additional logistic regression model is estimated with the pooled model IV and the segregated models per tenure. The results of the likelihood ratio test are significant, see appendix G, indicate that the model fit improves by segmenting the regression into the Owner-occupier, Social rent and Private rent. Table 5.5 provides an overview of the earlier assessed Model IV as well the results on the segmented version Model IV Owner-occupier, Model IV Social rent and model IV Private rent.

Housing satisfaction	Pooled model IV	Owner-occupier	Social rent	Private rent
Satisfaction dummy (0 = NS, 1 = S)				
A	1.506***	1.506***	1.339**	1.753***
	(0.0870)	(0.142)	(0.126)	(0.244)
B	1.245***	1.467***	1.009	1.200
	(0.0582)	(0.113)	(0.0797)	(0.130)
C	1.133**	1.141*	1.067	1.110
	(0.0451)	(0.0718)	(0.0750)	(0.101)
D	0.987	0.998	0.937	0.930
	(0.0474)	(0.0780)	(0.0723)	(0.139)
E	0.904**	0.815**	0.890	0.985
	(0.0355)	(0.0554)	(0.0601)	(0.0839)
F	0.982	1.059	0.889	1.064
	(0.0432)	(0.0688)	(0.0682)	(0.126)
G	Base	Base	Base	Base
Owner-occupier	BASE			
Social rent	0.403***			
	(0.0112)			
private rent	0.387***			
	(0.0137)			
Year	Yes	Yes	Yes	Yes
Socio-demographics	Yes	Yes	Yes	Yes
Housing characteristics	Yes	Yes	Yes	Yes
Neighbourhood characteristics	No	No	No	No
Constant	1.633	0.159*	0.152*	10.05*
	(0.817)	(0.146)	(0.130)	(9.135)
Observations	116920	68543	38820	9557
Log likelihood	-36153.697	-13821.873	-17802.067	-4328.3309
Df	36	34	34	34
Pseudo R^2	0.208	0.114	0.159	0.164

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable includes the years, energy labels and the interaction effects between energy efficient homes and years. Energy efficient homes have label A, B or C whereas energy inefficient homes have label D or worse. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Table 5.5: Pooled regression model IV: Housing satisfaction segmented based on tenure

The pooled model IV indicates significant results for tenure at a 1% level. In line with theory, both social rent and private rent have an odds ratio smaller than one, indicating that living in a rental unit makes it less likely to be satisfied compared to owner-occupiers. The segmented models are of interest, as these results indicate the impact of green energy labels per tenure type. Green energy labels are significantly affecting the likelihood of being satisfied, though the effects differ per tenure.

Label A is statistically significant at a 1% level for owner-occupiers and private renters, whereas social rent is significant at a 5% level. The less significant effects of label A compared to label G in the social rental sector indicates 33.9% higher likelihood of being satisfied. For owner-occupiers, the effect is stronger indicating a 50.6% higher likelihood of being satisfied, whereas the strongest effect occurs for private renters who are 75.3% more likely to be satisfied when living in a house with label A compared to label G. The largest effect for private renters is somewhat surprising as renters generally do not

retain the financial benefits of an energy-efficient home, as lower energy costs often lead to higher rents. It should be noted though, that the standard errors are larger as well. A possible explanation might be the average energy efficiency of the housing stock per tenure. After all, housing associations are forced to make their housing stock more sustainable whereas private renters are lacking the incentive to do so. This might therefore explain the higher odds ratio's for private rent, as relatively speaking a private rental unit with a green energy label might be harder to acquire (see section 2.3).

The fact the owner-occupiers have a lower odds ratio for label A than private renters, appears to be counterintuitive as green energy labels generally result in lower energy costs as well. Notably, the effects for less energy efficient labels B and C are only significantly affecting the likelihood on housing satisfaction for owner-occupiers, whereas social and private renters are insignificant. The effects of label A and B for owner-occupiers differs marginally with 50.6% and 46.7% respectively. The financial burden of achieving a greener energy label might potentially explain these notions. Sustainability measures to increase the energy label towards label A needs large financial investments might be a possible explanation for the lower odds compared to private renters. Whereas lower investments for a slightly less energy efficient label B, might be a better cost-benefit trade-off for owner-occupiers, hence perhaps the significant effects for labels B and C in the owner-occupier sector.

Again, in order to assess the potential effects of tenure on the effects of green energy labels on housing satisfaction more critically, a similar approach is taken as with the effects per years. Table 5.6 provides an overview of the interaction effects between the tenure and the energy efficiency of homes.

Housing satisfaction	Pooled model VII HS with interaction effects
Satisfaction dummy (0 = NS, 1 = S)	
Owner-occupier	Base
Social rent	0.438***
	(0.0150)
private rent	0.449***
	(0.0198)
Energy efficient green label	1.432***
	(0.0506)
Energy inefficient	BASE
Owner-occupier # Energy efficient green label	BASE
Owner-occupier # Energy inefficient	BASE
Social rent # Energy efficient green label	0.835***
	(0.0360)
Social rent # Energy inefficient	BASE
private rent # Energy efficient green label	0.712***
	(0.0463)
private rent # Energy inefficient	BASE

Year controls	Yes
Socio-demographics	Yes
Housing characteristics	Yes
Neighbourhood characteristics	No
Constant	1.249
	(0.587)
Observations	116920
Log likelihood	-36162.564
Df	33
Pseudo R^2	0.208

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable includes the years, energy labels and the interaction effects between energy efficient homes and tenure. Energy efficient homes have label A, B or C whereas energy inefficient homes have label D or worse. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Table 5.6: Pooled regression model IV: Housing satisfaction segmented based on tenure

The model results in table 5.6 show significant interaction effects between the energy label and tenure. The effects of green energy labels on housing satisfaction is significantly different per tenure group. Energy efficient green labels are significant at a 1% level, with an odds ratio of 1.432. An owner-occupier with a green label (label A, B or C) is 1.432 times more likely to be satisfied compared to households with an energy inefficient label (D and worse). As owner-occupier serves as base category, the odds of energy efficient green labels in the social sector is derived by multiplying the odds ratio of Energy efficient green labels with the odds of green energy labels in the social rental sector. Households in the social rental sector are thus (1.432*0.835=) 1.196 times more likely to be satisfied with a green label. Private renters with green labels are (1.432*0.712=) 1.020 times more likely to be satisfied compared to those with an inefficient label. Thus, households with a green energy label are stronger, positively affected when they are owner-occupier compared to households in a social or private rental house.

The interpretation of the segmented and interaction models is somewhat different, as the segmented regression in table 5.5 displays energy labels A till F compared to label G and the interaction model in table 5.6 displays energy efficient green labels (read A, B or C) compared to inefficient labels (read D or worse). The segmented model indicates the strongest effects for label A in the private rental sector, though the effect in the owner-occupier sector is the strongest with label B and C. Model 5.6 indicates the effects of a specific green energy label compared to an inefficient energy label per tenure. Whereas model 5.6 compares the effects less specifically, as it merges labels A, B and C into energy efficient labels. The segmented model indicates insignificant effects of green energy labels B and C in the social rental sector. Whereas the interaction model results are marginally different between the private and social rent, as the effects are the strongest in the owner-occupied sector. Also, the segmented model indicates that label A results in a higher likelihood of being satisfied for private renters compared to owner-occupiers. However, label B and C has the strongest effect on the likelihood of being satisfied

in the owner-occupied sector. Thus, based on the interaction model green energy labels compared to inefficient labels indeed have the strongest increase in the likelihood of being satisfied as an owner-occupier compared the social and private rent. More specifically, the segmented model in table 5.5 indicates that this is primarily driven by households with energy label B and C. Thus, the null hypothesis can be rejected as the regression results of the segmented model and the interaction model on tenure indicate that green energy labels are stronger affecting the likelihood of being satisfied in the owner-occupied sector compared to social and private renters.

5.3 Residential satisfaction

Multiple studies (Crull, 1991; Galster, 1987) recognised the interlinkages between housing and residential satisfaction, due to the strong relation of housing assessments and its immediate surroundings. Given the similarities and the strong interlinkages, this section explicitly examines the main findings and deviations in residential satisfaction from housing satisfaction.

The results on the pooled model for residential satisfaction is displayed in table 5.7 and relates to model IV in table 5.1 including all control variables. The pooled effects of green energy labels on housing satisfaction, as well as the extent to which these regression results are in line with existing literature will be discussed below.

Residential satisfaction (Model V)			
Satisfaction dummy (0 = NS, 1 = S)		**Tenure**	
A	1.734***	Owner-occupier	BASE
	(0.105)		
B	1.335***	Social rent	0.438***
	(0.0656)		(0.0128)
C	1.219***	private rent	0.367***
	(0.0511)		(0.0136)
D	1.023	**Construction year**	
	(0.0517)	Age of the building	1.000
E	0.988		(0.000267)
	(0.0413)	**Housing type**	
F	1.019	Single-family home	BASE
	(0.0470)		
G	BASE	Multi-family home	0.993
			(0.0292)
Year		**Cost-to-income ratio**	
Year=2012	BASE	Yearly housing expenses / Yearly Disposable HH income	1.588***
			(0.163)
Year=2015	0.796***	**Housing expenses**	
	(0.0268)	Housing expenses	1.000*
Year=2018	0.573***		(0.00000416)
	(0.0194)	**Housing value**	
Age		Housing value (WOZ tax value)	1.000***
17-23 year	0.990		(0.000000177)
	(0.0295)	**Poor housing quality**	
35-64 year	BASE	Agreed	BASE
65 years and older	1.517***	Disagreed	5.128***
	(0.0460)		(0.132)
Ethnicity		**Person-space-ratio**	

Native	BASE	Persons in HH / Living area	0.637***
			(0.0263)
Non-Western	0.601***	**Living space**	
	(0.0205)	up to 69m2	BASE
Western	0.791***		
	(0.0266)	70 up to 119m2	1.070
Education			(0.0387)
Low educated	BASE	over 120m2	1.250***
			(0.0651)
Secondary educated	0.916***	**Neighbourhood satisfaction**	
	(0.0238)	Very satisfied	BASE
Well educated	0.941*		
	(0.0280)	Satisfied	0.548***
Disposable HH income			(0.0171)
Low	0.883***	Neither satisfied, nor dissatisfied	0.146***
	(0.0292)		(0.00516)
Middle	BASE	Dissatisfied	0.105***
			(0.00452)
High	1.193**	Very dissatisfied	0.0661***
	(0.0800)		(0.00442)
Household type		**Good relationship with neighbours**	
Single person HH	BASE	Not agreed	BASE
Couple	1.069	Agreed	1.425***
	(0.0364)		(0.0308)
Parent(s) with kids	0.791***	**Region (G4)**	
	(0.0404)	Other municipalities	BASE
Non family household	1.154		
	(0.0850)	Big four (A'dam, R'dam, Utrecht, the Hague)	1.073*
Number of persons in the HH			(0.0352)
Number of persons in the household	1.030	**Urbanity level**	
	(0.0222)	Strong	BASE
Amount of years lived at this address			
Amount of years lived at this address	1.005***	Moderately	0.921**
	(0.00102)		(0.0278)
Self-reported health status		Little to none	0.929**
Very well	BASE		(0.0256)
Good	0.829***		
	(0.0253)		
Ok	0.565***		
	(0.0212)		
Differs from time to time	0.519***		
	(0.0221)		
Bad	0.465***	**Constant**	7.852***
	(0.0239)		(0.779)
Observations	116920	**df**	44
Log likelihood	-32671.842	**Pseudo R^2**	0.285

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable Energy label with label G as base category. Inclusion of control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics denoted with yes/no. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Table 5.7: Pooled regression model V

Table 5.2 indicates that the independent variables energy label are partly significant. Labels A, B and C are statistically significant at a 1% level, whereas the labels D, E and F are not significantly different from the base category label G. The odds ratio for Label A, B and C indicate that these households are 1.734, 1.335 and 1.219 times more likely to be satisfied compared to households with energy label G,

ceteris paribus. This translates into a 73,4%, 33.5% and 21.9% higher likelihood of being satisfied with energy label A, B or C respectively. These findings deviate somewhat from the pooled regression results on housing satisfaction, as the affect housing satisfaction with 50.6%, 24.5% and 13.3% respectively. The effect of green energy labels is stronger on residential satisfaction compared to housing satisfaction. Households living in a house labelled with energy label D, E or F are not significantly more likely to be satisfied (nor dissatisfied) compared to households with energy label G. Appendix H displays the model with the opposite reasoning as it considers label A as base category assessing whether a lower energy efficiency results in lower residential satisfaction levels. In line with housing satisfaction, all energy labels are significantly, negatively effecting residential satisfaction levels. These effects are again somewhat stronger for residential satisfaction compared to housing satisfaction.

These results indicate that green energy labels (A, B and C) positively affect the likelihood to be satisfied. Also, 'the greener' the energy label, the higher the likelihood of being satisfied becomes. Thus green energy labels are positively affecting household satisfaction levels, therefore the null hypothesis is rejected. Green energy labels positively affect residential satisfaction.

The control variables included in the regression analysis on residential satisfaction are derived from literature, see appendix A for a full overview. In addition to the model on housing satisfaction, some neighbourhood characteristics are included in the analysis. As with housing satisfaction, number of persons in the household, construction year and housing type are not significantly affecting the residential satisfaction levels. All of additional included neighbourhood characteristics affect residential satisfaction levels, which is in line with existing theory. The significant control variables are similar in sign and size compared to the results on housing satisfaction levels and differ marginally. These effects are in line with existing literature. Additional to the housing satisfaction, residential satisfaction includes neighbourhood characteristics. In line with findings by Balestra et al (2013) and Lu (1999), being satisfied with the neighbourhood is significantly contributing to the likelihood of being satisfied as a household. Lastly, in line with literature on life satisfaction (Ren et al, 2018; Balestra et al, 2013) higher levels of self-reported health status is significantly affecting the likelihood of being satisfied. More on life satisfaction in the following section.

In an equal manner to housing satisfaction in section 5.2, the stability of the relationship between energy labels and residential satisfaction over time is examined by segmenting the regression into years. Table 5.7 shows the regression results of the earlier analyse model V (pooled model on residential satisfaction) as well as the regression results per year. The likelihood ratio test, used to examine whether the segmented models have a better fit is significant (appendix G), thus indicate an

improved model fit by segmenting the regression analysis in years. Table 5.8 provides an overview of the earlier assessed Model V as well the segmented version Model V 2012, Model V 2015 and Model V 2018.

Residential satisfaction	Pooled model V	2012	2015	2018
A	1.734***	1.484	1.693***	1.806***
	(0.105)	(0.345)	(0.151)	(0.165)
B	1.335***	1.235	1.316***	1.390***
	(0.0656)	(0.183)	(0.0984)	(0.104)
C	1.219***	1.256	1.302***	1.160*
	(0.0511)	(0.162)	(0.0835)	(0.0743)
D	1.023	1.084	1.097	0.927
	(0.0517)	(0.137)	(0.104)	(0.0766)
E	0.988	1.017	1.018	0.934
	(0.0413)	(0.133)	(0.0642)	(0.0605)
F	1.019	0.977	1.108	1.005
	(0.0470)	(0.135)	(0.0812)	(0.0696)
G	BASE	BASE	BASE	BASE
Year=2012	BASE			
Year=2015	0.796***			
	(0.0268)			
Year=2018	0.573***			
	(0.0194)			
Year controls	Yes	Yes	Yes	Yes
Socio-demographics	Yes	Yes	Yes	Yes
Housing characteristics	Yes	Yes	Yes	Yes
Neighbourhood characteristics	Yes	Yes	Yes	Yes
Constant	7.852***	5.945***	4.647***	5.854***
	(0.779)	(1.637)	(0.684)	(0.861)
Observations	116920	15050	50165	51705
Log likelihood	-32671.842	-5453.4907	-12862.382	-14235.449
Df	44	42	42	42
Pseudo R^2	0.285	0.245	0.288	0.294

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable Energy label with label G as base category. Inclusion of control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics denoted with yes/no. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Table 5.8: Pooled regression model V segmented based on years

Table 5.8 shows the logistic regression results per year. Notably, none of the energy labels is significantly affecting residential satisfaction levels in 2012. In 2015 however, the results indicate that households with label A, B or C are significantly more likely to be satisfied compared to households with energy label G. The odds ratios van 1.693, 1.316 and 1.302 indicate that the more energy efficient a house is labelled, the higher the likelihood is that the household is satisfied. Label C becomes less significant and decreases in size and label B is only marginally stronger. These results deviate somewhat from housing satisfaction, as label C was insignificant. The odds ratio for energy label A increase till 1.806 in 2018. Indicating a stronger effect of the most energy efficient labels compared to the base category, label G as well as a stronger effect on residential satisfaction compared to housing satisfaction levels. Again, like with housing satisfaction the insignificant effects of energy labels in 2012

might be due to its recent introduction. Similar to the observed effects on housing satisfaction, energy efficient green labels are increasingly affecting the odds of being satisfied.

The additional model displayed in table 5.9 examines the effects of energy efficient green labels (read label A, B or C) compared to energy inefficient labels (read D or worse) with interaction variables.

Residential satisfaction	Pooled model VI RS with interaction effects
Satisfaction dummy (0 = NS, 1 = S)	
Year=2012	1
	(.)
Year=2015	0.778***
	(0.0317)
Year=2018	0.570***
	(0.0233)
Energy efficient green label	1.212***
	(0.0625)
Energy inefficient	BASE
Year=2012 # Energy efficient green label	BASE
Year=2012 # Energy inefficient	BASE
Year=2015 # Energy efficient green label	1.090
	(0.0670)
Year=2015 # Energy inefficient	BASE
Year=2018 # Energy efficient green label	1.046
	(0.0631)
Year=2018 # Energy inefficient	BASE
Socio-demographics	Yes
Housing characteristics	Yes
Neighbourhood characteristics	Yes
Constant	7.762***
	(0.705)
Observations	116920
Log likelihood	-32698.274
Df	41
Pseudo R^2	0.284

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable includes the years, energy labels and the interaction effects between energy efficient homes and years. Energy efficient homes have label A, B or C whereas energy inefficient homes have label D or worse. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Table 5.9: Pooled regression model with interaction effects

The regression results iterate the stronger effect of green energy labels on residential satisfaction compared to housing satisfaction, as the odds ratios for energy efficient green labels is 1.212 for residential satisfaction compared to 1.186 for housing satisfaction. However, in line with findings on housing satisfaction, the interaction variables are insignificant. Again, the aggregated use of 'energy efficient green labels' might be insignificant due to levelling effects throughout time. After all, table 5.8 indicated a decreasing significance and effect of less energy efficient green labels. Nonetheless, the most energy efficient labels are increasingly positively affecting residential satisfaction throughout

time. The null hypothesis of a stable effect of energy labels throughout time can be rejected, as the greenest energy labels are increasingly positively affecting the likelihood of being satisfied as a household.

An additional logistic regression model is estimated with the pooled model V and the segregated models per tenure. The results of the likelihood ratio test are significant, see appendix G, indicate that the model fit improves by segmenting the regression into the Owner-occupier, Social rent and Private rent. Table 5.5 provides an overview of the earlier assessed Model V as well the segmented version Model V Owner-occupier, Model V Social rent and model V Private rent.

Residential Satisfaction	Pooled model V	Owner-occupier	Social rent	Private rent
Satisfaction dummy (0 = NS, 1 = S)				
A	1.734***	1.654***	1.611***	2.085***
	(0.105)	(0.163)	(0.161)	(0.304)
B	1.335***	1.507***	1.095	1.410**
	(0.0656)	(0.122)	(0.0916)	(0.162)
C	1.219***	1.173*	1.177*	1.269*
	(0.0511)	(0.0782)	(0.0880)	(0.124)
D	1.023	1.000	0.999	0.947
	(0.0517)	(0.0820)	(0.0819)	(0.149)
E	0.988	0.827**	1.004	1.082
	(0.0413)	(0.0603)	(0.0724)	(0.0981)
F	1.019	1.083	0.923	1.134
	(0.0470)	(0.0741)	(0.0751)	(0.142)
G	Base	Base	Base	Base
Owner-occupier	Base			
Social rent	0.438***			
	(0.0128)			
private rent	0.367***			
	(0.0136)			
Year	Yes	Yes	Yes	Yes
Socio-demographics	Yes	Yes	Yes	Yes
Housing characteristics	Yes	Yes	Yes	Yes
Neighbourhood characteristics	Yes	Yes	Yes	Yes
Constant	7.852***	7.449***	3.176***	2.369***
	(0.779)	(1.491)	(0.493)	(0.580)
Observations	116920	68543	38820	9557
Log likelihood	-32671.842	-12248.972	-16216.403	-3974.4219
Df	44	42	42	42
Pseudo R^2	0.285	0.215	0.234	0.232

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable includes the years, energy labels and the interaction effects between energy efficient homes and years. Energy efficient homes have label A, B or C whereas energy inefficient homes have label D or worse. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Table 5.10: Pooled regression model V segmented based on tenure

Similar to the results on housing satisfaction tenure is significant at a 1% level with odds ratios smaller than one, indicating that living in a rental unit makes it less likely to be satisfied compared to owner-occupiers. However, the effects of energy labels on residential satisfaction levels differ per tenure. Clearly, the greenest energy label A, is statistically significant at a 1% level for all tenures. Deviating

from results on housing satisfaction, the odds ratios of owner-occupier and social rent only differ marginally as they indicate a 65.4% and 61.1% higher likelihood of being satisfied with label A compared to label G. The odds ratio of private rents is higher, as it indicates that private renters with label A are about twice as likely to be satisfied compared to private renters with label G. Which is somewhat surprisingly, but in line with results on housing satisfaction. Overall, the effects of green energy labels per tenure are more significant and stronger for residential satisfaction compared to housing satisfaction.

The effects of green energy labels is the weakest in the social renter sector, as label B is statistically insignificant and label C merely significant at a 10% level. The effects for label B are stronger in the owner-occupier sector than for private renters, as label B resulting in a 50.7% and 41% higher likelihood of being satisfied compared to having label G as an owner-occupier or private renter, respectively. Label C however, indicates that the effects for private renters is stronger, as all tenures are significant at a 10% level and result in a 17.3%, 17.7% and 26.9% higher likelihood compared to label G as an owner-occupier, social renter or private renter respectively. These results deviate from housing satisfaction as label B and C did not affect housing satisfaction significantly, but do significantly positively affect residential satisfaction levels.

An additional model on energy efficient green labels and tenure is estimated. Table 5.6 provides an overview of the interaction effects between the tenure and the energy efficiency of homes.

Residential satisfaction	Pooled model VII RS with interaction effects
Satisfaction dummy (0 = NS, 1 = S)	
Owner-occupier	Base
Social rent	0.477***
	(0.0172)
private rent	0.403***
	(0.0189)
Energy efficient green label	1.432***
	(0.0525)
Energy inefficient	Base
Owner-occupier # Energy efficient green label	Base
Owner-occupier # Energy inefficient	Base
Social rent # Energy efficient green label	0.845***
	(0.0383)
Social rent # Energy inefficient	Base
private rent # Energy efficient green label	0.808**
	(0.0552)
private rent # Energy inefficient	Base
Year controls	Yes

Socio-demographics	Yes
Housing characteristics	Yes
Neighbourhood characteristics	Yes
Constant	7.097***
	(0.649)
Observations	116920
Log likelihood	-32690.761
Df	41
Pseudo R^2	0.284

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable includes the years, energy labels and the interaction effects between energy efficient homes and tenure. Energy efficient homes have label A, B or C whereas energy inefficient homes have label D or worse. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Table 5.11: Pooled regression model V segmented based on tenure

Similar to housing satisfaction, the model results in table 5.6 indicate significant interaction effects between the energy label and the tenure. Compared to housing satisfaction, green energy efficient labels have an equal effect on residential satisfaction for owner-occupiers, as the odds ratio is 1.432. The ratios for the various tenure types are marginally lower. Households in the social rental sector are thus (1.432*0.845=) 1.210 times more likely to be satisfied if the dwelling has an energy efficient green label compared to an energy inefficient label. Private renters with green labels are (1.432*0.808=) 1.157 times more likely to be satisfied compared to a private renter with an energy inefficient home. Thus, households with a green energy label are stronger positively affected when they are owner-occupier compared to households in a social or private rental house. In line with results on housing satisfaction, the segmented and interaction model indicate that green energy labels affect the residential satisfaction levels the strongest in the owner-occupied sector. Deviating from housing satisfaction, these results are primarily driven by energy label B instead of B and C, but the null hypothesis can still be rejected.

5.4 Life satisfaction

Due to data constraints, the regression analysis on life satisfaction is solely conducted on data from the WoON 2015 survey, as two cycles were deemed inefficient to examine the effects on the stability of the parameter energy label throughout time. Again a logistic regression analysis is estimated, though the dependent variable relates to the overall satisfaction with life of households. Based on literature residential satisfaction is included as a control variable as it is a key indicator of life satisfaction. The regression analysis is similarly estimated; however the dependent variable is related to the household satisfaction level with life in general. Again the regression model is built up, by adding control variables and testing whether the model fit improves significantly. Table 5.12 provides an overview of the main results.

Life satisfaction	Model I	Model II	Model III	Model IV
Satisfied with life				
A	1.154	1.340*	0.935	0.961
	(0.144)	(0.183)	(0.149)	(0.154)
B	1.186	1.214	0.978	0.959
	(0.130)	(0.145)	(0.133)	(0.132)
C	0.933	1.138	0.985	0.981
	(0.0858)	(0.113)	(0.117)	(0.118)
D	1.657**	1.329	1.141	1.084
	(0.267)	(0.226)	(0.210)	(0.201)
E	0.549***	1.110	1.205	1.215
	(0.0524)	(0.116)	(0.141)	(0.145)
F	1.186	1.155	1.038	1.032
	(0.143)	(0.149)	(0.146)	(0.146)
G	Base	Base	Base	Base
Year controls	Yes	Yes	Yes	Yes
Socio-demographics	Yes	Yes	Yes	Yes
Housing characteristics	Yes	Yes	Yes	Yes
Neighbourhood characteristics	Yes	Yes	Yes	Yes
Constant	37.58***	86.09***	3.400	3.679
	(2.912)	(17.87)	(4.453)	(4.782)
Observations	50165	50165	50165	50165
Log likelihood	-6241.1827	-4871.2837	-4728.9737	-4671.1131
Df	6	23	35	43
Pseudo R^2	0.010	0.227	0.250	0.259

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Life Satisfaction dummy, independent variable Energy label with label G as base category. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively. Inclusion of control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics denoted with yes/no.

5.0-12: Life satisfaction

The full model (Model IV) has the best model fit as it explains 25.9% of the variance in the model (pseudo R2). Appendix J contains the full model with all control variables, which are consistent with literature as both self-reported health status and satisfaction with the home are statistically significant and positively related to satisfaction with life in general. However, more the energy labels are of interest and deviate from earlier regression models on housing and residential satisfaction.

The energy labels are insignificant. Therefore, energy labels are not affecting the satisfaction of households with life in general. The null hypothesis of an equal impact of green energy labels on residential and life satisfaction cannot be rejected. Deviating from housing and residential satisfaction, energy labels are thus not affecting a households' satisfaction with life in general. Consequently, the segmented and interaction model on tenure is not of any added value.

6. Discussion

6.1 Conclusion

Previous studies towards household satisfaction levels did not examine the effects of living in a house with green energy labels. This research however, aims to address the literature gap on household satisfaction levels by examining the effects of green energy labels in order to answer the following RQ: *What is the relationship between green energy labels and household satisfaction levels in the Netherlands?* In order to assess the effects of energy labels on all levels of household satisfaction, this research includes housing, residential and life satisfaction. Due to data constraints, only housing and residential satisfaction are examined with multiple cycles of WoON survey data, whereas life satisfaction is examined with solely one cycle of WoON data. Both housing and residential satisfaction are examined with the binary variable relating to satisfaction with the home, whereas life satisfaction is examined with a variable relating to satisfaction with life in general. The energy labels, range from energy efficient green labels A, B and C towards inefficient energy labels D, E, F and G.

The pooled regression models on housing and residential satisfaction indicate a positive relationship between green energy labels and satisfaction levels, compared to household with energy inefficient homes. Both models on housing and residential satisfaction indicate that the greenest energy labels have the strongest effect on satisfaction levels. These effects are stronger for residential satisfaction, as the odds ratio's of 73.4%, 33.5% and 21.9% for label A, B and C respectively, are somewhat higher than 50.6%, 24.5% and 13.3% for housing satisfaction levels. Green labels are thus positively affecting housing and residential satisfaction though the impact of green energy labels is stronger on the latter.

The impact of time and tenure on the effects of green energy labels are examined using both a segmented and an interaction model. The segmented model uses exact energy labels and the interaction models aggregates label A, B and C into 'green energy efficient labels' set of against the other 'inefficient energy labels'. The segmented models on both housing and residential satisfaction indicate insignificant results for 2012. Potentially due to the relative recent introduction of energy labels and on average an inefficient housing stock. As green energy labels were not the norm yet, the increased attention and legislation on energy labels might have resulted in an increased awareness and average sustainability levels of the housing stock could be instrumental to satisfaction levels for households with green energy labels throughout time. However, only three datapoints in nine years were considered, so these effects should be considered with caution. Both housing and residential satisfaction are positively affected by green energy labels in 2015 and 2018. The greenest energy labels are increasingly affecting satisfaction levels, whereas the less efficient label C becomes insignificant

for housing satisfaction while losing significance and effect on residential satisfaction. Potentially the increased average sustainability of the housing stock reduced the relative difference of label C with inefficient label. Indirectly these results are iterated with the interaction models, as it examines the energy efficient labels including label A, B and C. Together, these labels are insignificant though the more specific segmented model indicate a significant effect driven by the most energy efficient labels as they are increasingly positively affecting housing and residential satisfaction levels throughout time.

The pooled models on the effect of tenure on both housing and residential satisfaction are in line with literature, indicating a higher likelihood on satisfaction for owner-occupiers. The segmented housing satisfaction model indicate that private renters have the strongest effect for label A, whereas labels B and C are only significantly affecting owner-occupiers. These effects might relate to the average energy efficiency of the housing stock per tenure. Relatively speaking, a private rental unit with a green energy label might be harder to acquire, resulting in higher odds. The interaction model on green energy labels per tenure confirms this, as a green energy efficient label results in a 43.2%, 19.6% and 2% higher likelihood on housing satisfaction for owner occupiers, social renters and private renters respectively. The results on residential satisfaction levels are similar, as green energy labels are resulting in a 43.2%, 21% and 15.7% higher likelihood on residential satisfaction compared to households with inefficient energy labels. Households with a green energy label are thus stronger positively affected as owner-occupier compared to households in the social or private rental sector. More specifically, the segmented model indicates that the positive effects for owner-occupiers is primarily driven by label B and C in housing satisfaction and solely label B in residential satisfaction. The fact that the higher likelihood of both housing and residential satisfaction is not driven by label A, might relate to the large financial investments needed to acquire a label A. Whereas lower investments for a slightly less energy efficient label B (and C), might be a better cost-benefit trade-off for owner-occupiers.

The effects of green energy labels on life satisfaction are insignificant. Consequently, green energy labels do not affect the likelihood of being satisfied with life in general. Due to data constraints, these effects were solely examined at one moment in time eliminating examination on the insignificance throughout time. As a consequence of the insignificant results and limited data cycles, the effects of tenure on green energy label effects are of no interest.

Reflecting on the main research question, there is a positive relation between green energy labels and housing and residential satisfaction levels. The effect of the greenest energy labels is increasing throughout time. The effects are the strongest in the owner-occupied sector. Despite its implications on housing and residential satisfaction, green energy labels are insignificantly affecting life satisfaction.

6.2 Research limitations and recommendations

Firstly, the dependent variable of interest in this research on household satisfaction levels, is estimated using a five-point Likert-scale variable. An ordinal variable is perfectly suitable for an ordered logistic regression analysis. However, in assessing the suitability of this methodological approach it turned out that the regression assumption on proportional odds, also known as the brant (1990) assumption got violated. Therefore, the Likert scale variable had to be reduced to a binary variable in order to make it suitable for a binary logistic regression analysis. As a consequence, data on the differentiation between various levels of satisfaction have been lost, resulting in a research limitation. Secondly, energy labels are fairly recently introduced resulting in a limited amount of available data in the WoON surveys. Unfortunately, at the time of conducting this research the WoON 2021 is not yet available. In addition, due to the recent introduction of the energy label, the WoON survey 2008 has no data and the WoON 2012 has limited data available on energy labels. Due to the pioneering role of social housing associations, the social rental sector is overrepresented. In addition, not all variables are available, including the lack of data on life satisfaction. As a consequence, the effects of energy labels on life satisfaction had to be examined separately, only taken WoON 2015 data into consideration. Lastly, academic debate on household satisfaction research identified the potential effects of overestimation of satisfaction levels due to overestimation of scale-dependent variables Likert-type scales with regard to household satisfaction levels. Therefore the results are potentially positively skewed, resulting in a potential overestimation of satisfaction.

Future research towards household satisfaction in the Netherlands should encounter these research limitations. Firstly by reiterating the research with more cycles of WoON survey data to assess the stability of the effects of green energy labels more carefully. Also, the regression results indicated that the greenest energy labels are becoming increasingly important, whereas the least ‘green’ energy labels are becoming less significant. Thus, future research should entail more detailed grouping of energy labels, as the current research set up aggregated the energy labels A++, A+ and A into label A. Secondly, future research should address household satisfaction research with an ordered logistic regression, in order to retain data information and gain more insight in the extent of the effect of green energy labels on household satisfaction levels. As in the current setup, very satisfied and satisfied are merged into ‘satisfied’ and neither satisfied nor dissatisfied, dissatisfied and very dissatisfied are merged into ‘not satisfied’, therefore losing information on the extent of the impact of green energy labels on household satisfaction levels. Lastly, future research should control for upwards biased results, potentially using the tested approach of Ren et al (2018) by using the PSM method.

Bibliography

References

Agentschap NL (2012). Thema uitgewerkt voor de corporatiesector: Energielabel als sturingsinstrument.

Amerigo, M. & Aragonés, J.I. (1990). Residential Satisfaction in Council Housing. *Journal Of Environmental Psychology*, 10(4), pp. 313-325.

Balestra, C. & Sultan, J. (2013). *Home Sweet Home: the determinants of residential satisfaction and its relation with well-being*. OECD Statistics Working Papers: Organisation for Economic Co-operation and Development.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics, 46, pp. 1171-1178.*

Brounen, D., & Kok, N. (2011). On the economics of energy labels in the housing market . *Journal of Environmental Economics and Management*, 62(11), pp. 166-179.

Brown, L.A. & Moore, E.G. (1970). *The intro-urban migration process: A perspective*. Geografiska Annaler, 52, pp 1-13.

DeMaris, A. (1995). A tutorial in logistic regression. *Journal of marriage and family,* vol 57 (4), pp. 956-968.

Dunn, J. R. & M. V. Hayes (2000). Social inequality, population health, and housing: a study of two Vancouver neighbourhoods. *Social Science and Medicine*, Vol.51 (4), pp. 563-87.

Ebrahimigharehbaghi, S., Qian, Q., Meijer, F. & Vischer, H. (2019). Unravelling Dutch homeowners' behaviour towards energy efficiency renovations: What drives and hinders their decision-making? *Energy policy*, 129, pp. 546-561

Elsinga, M. & Hoekstra, J. (2005). Homeownership and Housing Satisfaction. *Journal of Housing and the Built Environment*, 20(4), pp. 401-424.

Fernàndez-Portero, C., Alarcón, D. & Barrios Padura, Á. (2017). Dwelling conditions and life satisfaction of older people through residential satisfaction. *Journal of Environmental Psychology, 49, pp. 1-7.*

Galster, G. (1987). Identifying the Correlates of Dwelling Satisfaction: An Empirical Critique. *Environment and behaviour, 19 (5), pp. 539-568*

Galster, G. & Hesser, G. (1981). Residential satisfaction. Compositional and contextual correlates. *Environment and behaviour*, 13 (6), pp. 735-758

Golant, S.M. (1982). Individual Differences Underlying the Dwelling Satisfaction of the Elderly. *Journal of Social Issues, 38 (3), pp. 121-133*

Greenwood, K. R. (2011). The demand for green housing amenities. *The Journal of Sustainable Real*

Estate, 3 (1), pp. 127-141.

Grigolon, A., Dane, G., Rasouli, S. & Timmermans, H. (2014). Binomial random parameters logistic regression model of housing satisfaction. *Procedia Environmental Sciences*, 22, pp. 280-287.

Huang, Z. & Du, X. (2015). Assessment and determinants of residential satisfaction with public housing in Hangzhou, China. *Habitat international*, 47, pp. 218-230.

Hyland, M., Lyons, R.C. & Lyons, S. (2013). The Value of Domestic Building Energy Efficiency – Evidence from Ireland. *Energy Economics*, 40, 943-952.

Johnson, A. (2014). Parameters contributing to occupants' satisfaction. Green and conventional residential buildings. *Real Estate and Construction Management,* 32, pp. 411-437.

Kuha, J., Mils, C. (2020). On group comparisons with logistic regression models. *Sociological methods & research*, 49 (2), pp. 498-525.

Lu, M. (1999). Determinants of residential satisfaction: ordered logit vs . regression models. *Journal of growth and change*, 30, pp 264- 287.

Mood, c. (2010). Logistic regression: why we cannot do what we think we can do, and what we can do about it. *European sociological review,* 26(1), pp 67-82.

Olaniyan, M., & Evans, J. (2014). The importance of engaging residential energy customers' hearts and minds. *Energy Policy*, 69, pp. 273-284.

Olgun, D. (2020). Do Residents Appreciate Green Labels? – A Quantitative Approach into the Relationship between Energy Labels and Residential Satisfaction in the Netherlands. Master thesis, Rijksuniversiteit Groningen.

Ren, H., Folmer, H., van der Vlist, A.J. (2016). The Impact of Home Ownership on Life Satisfaction in Urban China: A Propensity Score Matching Analysis. *Journal of Happiness Studies,* 19, pp. 397-422.

Quang Tran, T. & Van Vu, H. (2017). A microeconometric analysis of housing and life satisfaction among the Vietnamese elderly. *Qual Quant, 52, pp. 849-867.*

Rossi, P. H. (1955). Why Families Move. Glenco, IL: The Free Press.

Tan, T. (2014). Assessing Green Home Performance: A Case Study of Iskandar Malaysia. International *Journal of Property Sciences*, 4 (1).

Train, K. (2009). Discrete choice methods with simulation. Chapter 7 Variations on a theme. *Cambridge university press pp. 151-182.*

Wong, G. k. m. (2002). A Conceptual Model of the Household's Housing Decision–Making Process: The Economic Perspective. *Review of Urban & Regional Development Studies*, Vol. 14(3), pp. 217–234.

Sources

AEDES (n.d.). About Aedes. Accessed on 27-4-2021, available at: https://www.aedes.nl/algemeen/over-aedes

Bijstandsnormen, (n.d.). *Bijstandsnormen, uitkeringen en wwb normen; uw bijstandsnorm voor 2012.* Accessed on 26-5-2021, available at: https://www.bijstandsnormen.nl/2012/

BPD Bouwfonds gebiedsontwikkeling (2019). *90% van de consumenten heeft belangstelling voor duurzaam wonen.* Accessed on 20-4-2021, available at: https://www.bpd.nl/actueel/onderzoeken/90-procent-van-de-consumenten-heeft-belangstelling-voor-duurzaam-wonen/

CBS (2012). *Energieklasse bekend van ruim 2 miljoen woningen*. Accessed on 28-4-2021, available at: https://www.cbs.nl/nl-nl/nieuws/2012/43/energieklasse-bekend-van-ruim-2-miljoen-woningen

CBS (n.d.). *WoonOnderzoek Nederland (WoON).* Accessed on 5-5-2021, available at: https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/woononderzoek-nederland--woon--

CBS (2020). *Ruim 69 duizend nieuwbouwwoningen in 2020*. Accesed on 28-6-2021, available at: https://www.cbs.nl/nl-nl/nieuws/2021/04/ruim-69-duizend-nieuwbouwwoningen-in-2020

CBS (2020)[1]. *Voorraad woningen; eigendom, type verhuurder, bewoning, regio.* Accessed on 28-4-2021, available at: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82900NED/table?fromstatweb

CBS. (2021). *Voorraad woningen en niet-woningen; mutaties, gebruiksfunctie, regio*. Assessed on 14-6-2021, available at: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81955NED/table?fromstatweb

CLO (2019). *Energielabels van woningen, 2008 – 2018.* Accessed on 28-4-2021, available at: https://www.clo.nl/indicatoren/nl055607-energielabels-woningen

CLO (2020). *Energielabels van woningen, 2010-2019.* Accessed on 28-4-2021, available at: https://www.clo.nl/indicatoren/nl0556-energielabels-woningen

Energiedeskundig (n.d.). *Geschiedenis van het energielabel voor woningen.* Accessed on 26-4-2021, available at: https://www.energiedeskundig.nl/energielabel-voor-woningen-uitleg/geschiedenis-van-het-energielabel-voor-woningen/?doing_wp_cron=1619440217.4730169773101806640625

Energie label voor woningen (n.d.) *Energie label voor woningen.* Accessed on 3-3-2021, available at: https://www.energielabelvoorwoningen.nl/?cookieCheck=true

LenteAkkoord (n.d.) *Zoekresultaat voor: Energielabel 2012.* Accessed on 5-5-2021, available at: https://www.lente-akkoord.nl/page/4/?s=energielabel+2012&sa

Rijksoverheid (2019*). Staat van de woningmarkt. Jaarrapportage 2019*. Accessed on 3-3-2021, available at: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2019/11/07/staat-van-de-woningmarkt-2019/20191107+Staat+van+de+Woningmarkt+2019.pdf

SER[1] (n.d.) *Gebouwde omgeving*. Accessed on 3-3-2021, available at: https://www.ser.nl/nl/thema/energie-en-duurzaamheid/energieakkoord/~/link.aspx?_id=D88E14B4D54B43808C7D4A92AB689AAF&_z=z

SER[2] (n.d.) *Wat is het Energieakkoord*? Accessed on 3-3-2021, available at: https://www.ser.nl/nl/thema/energie-en-duurzaamheid/energieakkoord/wat

Appendices:

university of groningen

Appendix A: Variable overview

Inventory of the presence of variables WoON Survey Cycles					
Dataset:	**WoON 2012**	**WoON 2015**	**WoON 2018**	**Source:**	**Remarks:**
Cases:	**(N = 69.339)**	**(N = 62.668)**	**(N = 67.523)**		
Dependent variable					
Satisfaction	✓	✓	✓	Elsinga et al, 2015; Golant, 1982; Galster, 1987; Grigolon, 2014; Huang et al, 2015; Balestra et al, 2013; Fernandez, 2017; Quang Tran et al, 2018; Lu, 1999; Mohit et al,	
Satisfaction with life in general	✗	✓	✓	Ren et al, 2018; Fernandez et al, 2017; Quang Tran et al, 2018;	Data is not available in all WoON cycles, no suitable alternative either.
Key independent variable (model 1)					
Energy label	✓	✓	✓		Deviating sources, though all provided an overview with preliminary energy labels
Personal socio-demographic characteristics households' (model 2)					
Age	✓	✓	✓	Elsinga et al, 2015; Galster, 1981; Grigolon, 2014; Huang et al, 2015; Ren et al, 2018; Balestra et al, 2013; Fernandez, 2017; Quang Tran et al, 2018; Lu, 1999; Mohit et al,	Included
Gender	✓	✓	✗	Golant, 1982; Galster, 1987; Grigolon, 2014; Huang et al, 2015; Balestra et al, 2013; Fernandez, 2017; Quang Tran et al, 2018; Lu, 1999;	Data is not available in all WoON cycles, no suitable alternative either.
Ethnicity	✓	✓	✓	Golant, 1982;Galster, 1981; Quang Tran et al, 2018; Lu, 1999;	Included
Education	✓	✓	✓	Golant, 1982; Galster, 1987;Galster, 1981; Huang et al, 2015; Ren et al, 2018; Balestra et al, 2013; Fernandez, 2017; Quang Tran et al, 2018; Lu, 1999; Mohit et al, 2014;	Measurement scale deviates so transformation of variables needed
Disposable household income	✓	✓	✓	Elsinga et al, 2015; Golant, 1982; Galster, 1987;Galster, 1981; Grigolon, 2014; Huang et al, 2015; Ren et al, 2018; Balestra et al, 2013; Quang Tran et al, 2018; Lu, 1999; Mohit	See Cost-to-income as well
Subjective economic status	✗	✓	✗	Golant, 1982;	Data is not available in all WoON cycles, no suitable alternative either.
Household type	✓	✓	✓	Galster, 1987;Galster, 1981; Grigolon, 2014; Balestra et al, 2013; Fernandez, 2017; Quang Tran et al, 2018; Lu, 1999; Mohit et al, 2014;	Included
Number of persons in the household	✓	✓	✓	Elsinga et al, 2015; Huang et al, 2015; Ren et al, 2018; Quang Tran et al, 2018;	Included
Number of years living at this adress	✓	✓	✓	Golant, 1982; Huang et al, 2015; Quang Tran et al, 2018; Lu, 1999; Mohit et al, 2014; Mohit et al, 2014;	Included
Religion	✗	✗	✗	Quang Tran et al, 2018;	Data is not available in all WoON cycles, no suitable alternative either.
Contact with family	✓	✓	✗	Grigolon, 2014;	Data is not available in all WoON cycles, no suitable alternative either.
Contact with friends	✓	✓	✗	Grigolon, 2014;	Data is not available in all WoON cycles, no suitable alternative either.
Social isolation	✗	✗	✗	Huang et al, 2015; Quang Tran et al, 2018;	Data is not available in all WoON cycles, no suitable alternative either.
Forced move	✗	✗	✗	Galster, 1981;	Data is not available in all WoON cycles, no suitable alternative either.
Travel time to work	✓	✗	✗	Galster, 1981; Grigolon, 2014;	Data is not available in all WoON cycles, no suitable alternative either.
Level of activity	✗	✗	✗	Golant, 1982;	Data is not available in all WoON cycles, no suitable alternative either.
Psychological well-being	✗	✗	✗	Golant, 1982;	Data is not available in all WoON cycles, no suitable alternative either.
Employment status	✓	✓	✗	Golant, 1982;Galster, 1981; Huang et al, 2015; Ren et al, 2018; Quang Tran et al, 2018; Mohit et al, 2014;	Data is not available in all WoON cycles, no suitable alternative either.
Marital status	✗	✗	✗	Golant, 1982;Galster, 1981; Quang Tran et al, 2018; Quang Tran et al, 2018;	Data is not available in all WoON cycles, no suitable alternative either.
Self-reported health	✓	✓	✓	Golant, 1982; Ren et al, 2018; Balestra et al, 2013; Quang Tran et al, 2018;	Included

university of groningen

Housing characteristics (model 3)					
Tenure	✓	✓	✓	Elsinga et al, 2015; Golant, 1982;Galster, 1981; Grigolon, 2014; Huang et al, 2015; Balestra et al, 2013; Fernandez, 2017; Lu, 1999; Mohit et al, 2014;	Included
Construction year	✓	✓	✓	Galster, 1987; Mohit et al, 2014;	Included
Housing type	✓	✓	✓	Elsinga et al, 2015; Galster, 1981; Grigolon, 2014; Balestra et al, 2013; Quang Tran et al. 2018: Mohit et al. 2014:	Included
Scale of housing adaptation	✗	✗	✗	Fernandez, 2017;	Data is not available in all WoON cycles, no suitable alternative either.
Scale of habitability	✗	✗	✗	Fernandez, 2017;	Data is not available in all WoON cycles, no suitable alternative either.
Lot size	✗	✗	✗	Galster, 1987; Galster, 1981;	Aanwezige variabele geeft enkel de aanwezigheid van een buitenruimte aan
Garage	✗	✗	✗	Galster, 1987;	Data is not available in all WoON cycles, no suitable alternative either.
Floor	✓	✗	✗	Huang et al, 2015;	Data is not available in all WoON cycles, no suitable alternative either.
Orientation	✗	✗	✗	Huang et al, 2015;	Data is not available in all WoON cycles, no suitable alternative either.
Financial burden	✗	✗	✗	Balestra et al, 2013;	Data is not available in all WoON cycles, no suitable alternative either.
Cost-to-income	✓	✓	✓	Lu, 1999;	Self constructed variable (Housing expenses / Disposable household income)
Housing expenses	✓	✓	✓	Elsinga et al, 2015; Balestra et al, 2013; Lu, 1999;	Included
WOZ housing value	✓	✓	✓	Lu, 1999; Mohit et al, 2014;	Included, as alternative for housing value
Housing quality	✓	✓	✓	Elsinga et al, 2015; Galster, 1987; Galster, 1981; Balestra et al, 2013; Mohit et al, 2014:	Included
Person-space ratio	✓	✓	✓	Elsinga et al, 2015; Galster, 1987; Balestra et al, 2013; Lu, 1999; Mohit et al, 2014;	Self constructed variabel (Persons in household / Living area)
Living area	✓	✓	✓	Grigolon, 2014; Huang et al, 2015; Quang Tran et al, 2018; Mohit et al, 2014;	See Person-space ratio
Neighbourhood characteristics (model 4)					
Neighbourhood satisfaction	✓	✓	✓	Galster, 1981; Balestra et al, 2013; Lu, 1999; Mohit et al, 2014;	Included
Relationship with neighbours	✓	✓	✓	Galster, 1981; Balestra et al, 2013; Lu, 1999;	Included
Sound nuisance	✓	✓	✗	Galster, 1981; Huang et al, 2015; Balestra et al, 2013;	Data is not available in all WoON cycles, no suitable alternative either.
Safety	✓	✓	✓	Galster, 1981; Huang et al, 2015; Balestra et al, 2013; Mohit et al, 2014;	Included
Commercial amenities	✓	✗	✓	Golant, 1982; Mohit et al, 2014;	Data is not available in all WoON cycles, no suitable alternative either.
Green amenities	✓	✗	✗	Huang et al, 2015;	Data is not available in all WoON cycles, no suitable alternative either.
Public (transport) facilities	✓	✗	✓	Huang et al, 2015; Balestra et al, 2013; Mohit et al, 2014;	Data is not available in all WoON cycles, no suitable alternative either.
Distance to CBD	✗	✗	✗	Huang et al, 2015; Mohit et al, 2014;	Solely desired distance to CBD
Region (G4)	✓	✓	✓	Grigolon, 2014; Huang et al, 2015; Ren et al, 2018; Quang Tran et al, 2018; Lu, 1999;	Included
Urbanity	✓	✓	✓	Golant, 1982;Galster, 1981; Grigolon, 2014; Balestra et al, 2013; Lu, 1999;	Included
Proportion ethnicities	✗	✗	✗	Galster, 1981;	Data is not available in all WoON cycles, no suitable alternative either.
Year controls					
Year	✓	✓	✓		Self constructed dummy variable to distinguish WoON cylces in the Pooled model

Legenda: ✓ Present ✗ Missing ✓ Self-constructed

university of groningen

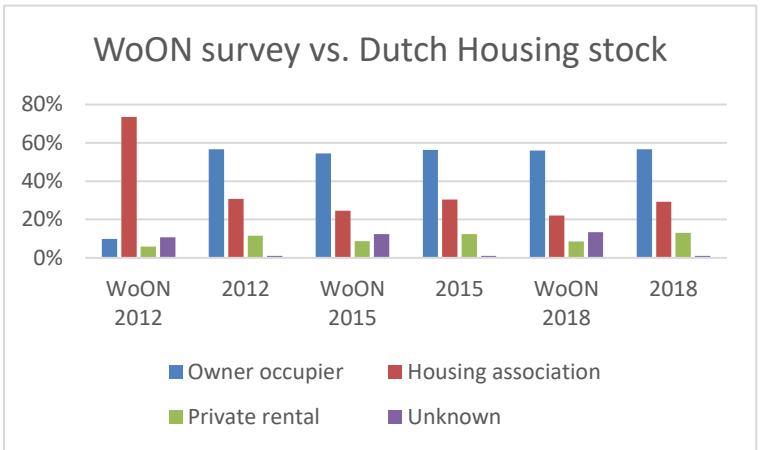
APPENDIX B: Representativity of the WoON surveys

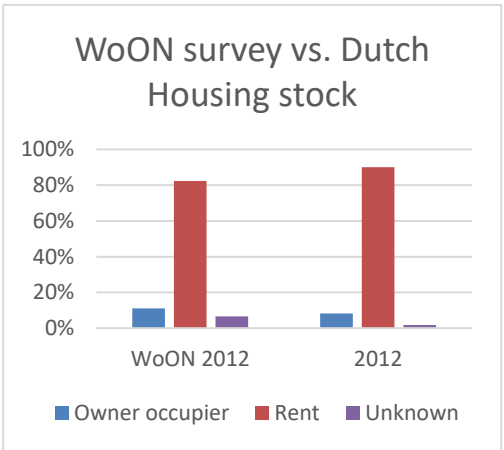
Tenure 2012	Frequency	Percent	Cumulative percentage
Owner-occupier	1738	9,82%	9,82%
Social rent	13024	73,56%	83,38%
Private rent	1039	5,87%	89,25%
Unknown	1905	10,76%	100%
	17706	100%	

Tenure 2015	Frequency	Percent	Cumulative percentage
Owner-occupier	33269	54,40%	54,40%
Social rent	15040	24,59%	78,99%
Private rent	5317	8,69%	87,68%
Unknown	7532	12,32%	100%
	61158	100%	

Tenure 2018	Frequency	Percent	Cumulative percentage
Owner-occupier	36024	55,92%	55,92%
Social rent	14224	22,08%	78,00%
Private rent	5521	8,57%	86,57%
Unknown	8653	13,43%	100%
	64422	100%	

Observations deleted	Observations	Deleted	Percent
2012	69.369	51.663	26%
2015	19.216	1.510	92%
2018	20.807	3.101	85%





APPENDIX C: Data cleaning

Dataset	Observations
WoON2012	69.339
WoON2015	62.668
WoON2018	67.523
SUM N pooled dataset	**199.530**

Data cleaning pooled dataset			**N = 195.798**
		Deleted observations	**Remaining observations**
Transformation	Recode education variables into 3 similar categories 2012	593	198.937
	Recode education variables into 3 similar categories 2015	1.090	197.847
	Recode education variables into 3 similar categories 2018	2.049	195.798
Missing values	Drop if Satisfaction==.	24.863	170.935
	Drop if Energy_lable==.	48.201	122.734
	Drop if Housing_type==.	3.569	119.165
	Drop if Living_area==.	0	119.165
	Drop if Houisng_expenses==.	0	119.165
	Drop if Cost_to_income==.	0	119.165
	Drop if Tenure==unkown	0	119.165
Odd values	Drop if Housing_expenses<0	263	118.902
	Drop if Housing_expenses>60000	67	118.835
	Drop if Disposable_HH_income<8018*	1.915	116.920
Remaining observations N in pooled dataset:			**116.920**

* Based on minimum yearly income of single person household depending on social assistance benefits of the government in 2012 (Bijstandsnormen, n.d.).

university of groningen

APPENDIX D: Variable transformations

Dataset: WoON 2012						N= 69.339
Variables	**Variable code WoON 2012**	**Label**	**Transformation**	**New variable name**	**Variable type**	**Missing values**
Dependent variable						
Satisfaction	TWoning	(16.1) Tevredenheid met huidige woning	Duplicated and transformed into dummy variable (0 Not Satisfied 1 Satisfied)	Satisfaction_dummy	Ordinal	
Key independent variable (model 1)						
Energy label	e_label	Energielabel (registratie)	Regroup energy lables to combine the classes A++ A+ and A into class A.	Energy_label	Ordinal	
Personal socio-demographic characteristics households' (model 2)						
Age	Leeftijd	leeftijd OP in 7 klassen		Age	Ordinal	
Ethnicity	etniop3	etniciteit OP naar herkomst (3 klassen)(registratie)		Ethnicity	Ordinal	
Education	VOplOP	(33.6) Niveau voltooide opleiding respondent	Recode ordinal variaible highest obtained education level into Low educated (up to VMBO/MAVO) secondary educated (up to MBO) and well educated (up to university). Drop cases without education labeled as 'anders'.	Education	Ordinal	593
Disposable household income	vromhh	Besteedbaar huishoudensinkomen, def WoON	Transformed into categories low (up to €35.000 yearly, middle (€35.001-70.000) and high (€70.000 and over).	Disposable_HH__inc	Continious	
Household type	SAMHH5	Samenstelling huishouden (5 klassen)	Redefine labels to match other WoON datasets in terms of spelling	Household_type	Ordinal	
Number of persons in the household	AantalPP	Aantal personen in huishouden	Recode continouse variable into categorical variable matchitng dataset 2018	Number_of_persons_in_HH	Ordinal	
Number of years living at this adress	JrKomWon	(20.1) Sinds welk jaar woont u op dit adres	Variable displays year of arrival, whereas the intended variable displays the amount of years living at the current adress. Therefore replace Years_at_current_adress = 2012-Years_at_current_adress.	Years_at_current_adress	Continious	
Self-reported health	Gezond	(9.1) Oordeel gezondheid	Redefine labels to match datasets 2015&2018 in terms of spelling	Self_reported_health	Ordinal	

Housing characteristics (model 3)						
Tenure	eigendom	Eigendom van de huidige woning	Redefine labels to match datasets 2015&2018 in terms of spelling	Tenure	Ordinal	
Construction year	BJaar	(8.1) Bouwjaar	Variable displays construction year, though the intended variable displays the age of the building in years. Therefore: replace Construction_year = 2012-Construction_year	Construction_year	Continious	
Housing type	vorm	een-/meergezins huidige woning (2 klassen)	Recode dummy into 0/1 and define 'eengezinswoning' as base	Housing_type	Dummy	
Cost-to-income	CONSTRUCED		Self constructed variables based on new variable Housing_expenses, to examine the housing quote. Housing expenses (totwl) / disposable household income (vromhh)	Cost_to_income	Continious	
Housing expenses	totwl	Totale woonlasten **per maand**	Transform to yearly costs for interpretation sake, as income is depicted as yearly income as well. Relabel to Housing expenses **per year and incorporate in Cost-to-income variable.**	Housing_expenses	Continious	
WOZ housing value	wozwaarde	WOZ waarde (registratie peildatum 1 januari 2011)		WOZ_housing_value	Continious	
Housing quality	Tonderho	(16.6) Woning is slecht onderhouden	Redefine labels to match datasets 2015&2018 in terms of spelling	Housing_quality	Ordinal	
Person-space ratio	CONSTRUCED		Number of persons in the household (AantalPP) / Living area (oppwon7)	Person_space_ratio	Continious	
Living area	oppwon7	Woon Oppervlakte in 7 klassen		Living_area	Ordinal	
Neighbourhood characteristics (model 4)						
Neighbourhood satisfaction	TWoonOmg	(17.1) Tevredenheid met huidige woonomgeving	Redefine labels to match datasets 2015&2018 in terms of spelling	Neighbourhood_satisfaction	Ordinal	
Relationship with neighbours	ConBuur1	(17.26) Ik heb veel contact met mijn directe buren	Redefine labels to match datasets 2015&2018 in terms of spelling	Relationship_neighbours	Ordinal	
Safety	Brtveilg	(17.34) Bang in deze buurt om lastiggevallen of beroofd te		Safety	Ordinal	
Region (G4)	g4_2	G4 (ja/nee) (2)	Recode dummy into 0/1 and define other municipalites as base	Region_G4	Dummy	
Urbanity	stedgem	stedelijkheid gemeente (5)		Urbanity	Ordinal	
Year controls						
Year	Year_2012		Construced dummy for each cycle (2012 / 2015 / 2018)	Year	Dummy	
				remaining cases (N):		**68.746**

university of groningen

Dataset: WoON 2015 (N = 62.668)					N= 69.339	
Variables	**Variable code WoON 2015**	**Label**	**Transformation**	**New variable name**	**Variable type**	**Missing values**
Dependent variable						
Satisfaction	TWoning	(13.1) Tevredenheid met huidige woning	Duplicated and transformed into dummy variable (0 Not Satisfied 1 Satisfied)	Satisfaction_dummy	Ordinal	
Key independent variable (model 1)						
Energy label	Energieklasse_vlp	Energielabel voorlopig (bron: rvo)	Recode energy labels into capital letters to match other WoON datasets	Energy_label	Ordinal	
Personal socio-demographic characteristics households' (model 2)						
Age	Leeftijd	Leeftijd op in 7 klassen		Age	Ordinal	
Ethnicity	etniop3	Etniciteit onderzoekspersoon naar herkomst (3 klassen) (bron: basis registratie		Ethnicity	Ordinal	
Education	NivBehOP	(26.7) Niveau voltooide opleiding respondent	Recode ordinal variaible highest obtained education level into Low educated (up to VMBO/MAVO) secondary educated (up to MBO) and well educated (up to university). Drop if education is 'anders'	Education	Ordinal	1.090
Disposable household income	vromhh	Besteedbaar huishoudinkomen, definitie vrom (persoon geaggregeerd naar huishoude	Transformed into categories low (up to €35.000 yearly, middle (€35.001-70.000) and high (€70.000 and over).	Disposable_HH__inc	Continious	
Household type	SamHH5	Samenstelling huishouden (5 klassen)	Redefine labels to match other WoON datasets in terms of spelling	Household_type	Ordinal	
Number of persons in the household	AantalPP	Aantal personen in huishouden	Recode continouse variable into categorical variable matchitng dataset 2018	Number_of_persons_in_HH	Ordinal	
Number of years living at this adress	JrKomWon	(15.1) Sinds welk jaar woont u op dit adres	Variable displays year of arrival, whereas the intended variable displays the amount of years living at the current adress. Therefore replace Years_at_current_adress = 2015-Years_at_current_adress.	Years_at_current_adress	Continious	
Self-reported health	Gezond	(29.1) Oordeel gezondheid		h	Ordinal	

Housing characteristics (model 3)						
Tenure	Eigendom	Eigenaar, sociale of overige huurder		Tenure	Ordinal	
Construction year	bjaarbag	Bouwjaar van de huidige woning (bron: bag peildatum 1 januari 2015)	Variable displays construction year, though the intended variable displays the age of the building in years. Therefore: replace Construction_year = 2015-Construction_year	Construction_year	Continious	
Housing type	vorm	Een-/meergezins huidige woning (2 klassen)	Recode dummy into 0/1 and define 'eengezinswoning' as base	Housing_type	Dummy	
Cost-to-income	CONSTRUCED		Self constructed variables based on new variable Housing_expenses, to examine the housing quote. Housing expenses (totwl) / disposable household income (vromhh)	Cost_to_income	Continious	
Housing expenses	totwl	Totale woonlasten per maand	Transform to yearly costs for interpretation sake, as income is depicted as yearly income as well. Relabel to Housing expenses **per year and incorporate in Cost-to-income variable.**	Housing_expenses	Continious	
WOZ housing value	WOZwaarde	WOZ waarde huidige woning (peildatum 1 januari 2014) inclusief bijschatting		WOZ_housing_value	Continious	
Housing quality	Tonderho	(13.6) Woning is slecht onderhouden		Housing_quality	Ordinal	
Person-space ratio	CONSTRUCED		Number of persons in the household (AantalPP) / Living area (oppwon7)	Person_space_ratio	Continious	
Living area	OppWon7	Woon oppervlakte in 7 klassen		Living_area	Ordinal	
Neighbourhood characteristics (model 4)						
Neighbourhood satisfaction	TWoonOmg	(14.1) Tevredenheid met huidige woonomgeving		Neighbourhood_satisfaction	Ordinal	
Relationship with neighbours	ConBuur1	(14.25) Ik heb veel contact met mijn directe buren		Relationship_neighbours	Ordinal	
Safety	Brtveilig	(14.32) Bang in deze buurt om lastiggevallen of beroofd te		Safety	Ordinal	
Region (G4)	g4_2	G4 (ja/nee) (2)	Recode dummy into 0/1 and define other municipalites as base	Region_G4	Dummy	
Urbanity	stedgem	Stedelijkheid gemeente (5)		Urbanity	Ordinal	
Year controls						
Year	Year_2015		Construced dummy for each cycle (2012 / 2015 / 2018)	Year	Dummy	
				remaining cases (N):		**68.249**

university of groningen

Dataset: WoON 2018 (N = 67.523)					N= 69.339	
Variables	**Variable code WoON 2018**	**Label**	**Transformation**	**New variable name**	**Variable type**	**Missing values**
Dependent variable						
Satisfaction	twoning	(12.1) Tevredenheid met huidige woning	Duplicated and transformed into dummy variable (0 Not Satisfied 1 Satisfied)	Satisfaction_dummy	Ordinal	
Key independent variable (model 1)						
Energy label	energieklasse_vlp	Energielabel voorlopig (RVO 2018)	Recode energy labels into capital letters	Energy_label	Ordinal	
Personal socio-demographic characteristics households' (model 2)						
Age	leeftijd	Leeftijd respondent (7 klassen)		Age	Ordinal	
Ethnicity	etniop3	Etniciteit respondent naar herkomst (3 klassen) (Bron: BRP)		Ethnicity	Ordinal	
Education	nivbehop1	(23.3.1) Diploma gehaald respondent -	Recode ordinal variaible highest obtained education level into Low educated (up to VMBO/MAVO) secondary educated (up to MBO) and well educated (up to university). Drop if education is 'geen van deze'	Education	Ordinal	2.049
Disposable household income	vromhh_r	Besteedbaar inkomen huishouden (definitie VROM/BZK), voorlopig inkomen	Transformed into categories low (up to €35.000 yearly, middle (€35.001-70.000) and high (€70.000 and over).	Disposable_HH__inc	Continious	
Household type	samhh5	Samenstelling huishouden (5 klassen)	Redefine labels to match other WoON datasets in terms of spelling	Household_type	Ordinal	
Number of persons in the household	aantalpp5	Aantal personen in huishouden (5 klassen)		Number_of_persons_in_HH	Ordinal	
Number of years living at this adress	jrkomwon	(14.1) Sinds welk jaar woont u op dit adres	Variable displays year of arrival, whereas the intended variable displays the amount of years living at the current adress. Therefore replace Years_at_current_adress = 2018-Years_at_current_adress.	Years_at_current_adress	Continious	
Self-reported health	gezond	(25.1) Oordeel gezondheid		h	Ordinal	

Housing characteristics (model 3)						
Tenure	eigendom	Eigenaar, sociale of overige huurder		Tenure	Ordinal	
Construction year	bjaarbag	Bouwjaar van de huidige woning (Bron: BAG peildatum 1 januari 2017)	Variable displays construction year, though the intended variable displays the age of the building in years. Therefore: replace Construction_year = 2018-Construction_year	Construction_year	Continious	
Housing type	vorm	Een-/meergezins huidige woning (2 klassen)	Recode dummy into 0/1 and define 'eengezinswoning' as base	Housing_type	Dummy	
Cost-to-income	CONSTRUCED		Self constructed variables based on new variable Housing_expenses, to examine the housing quote. Housing expenses (totwl) / disposable household income (vromhh)	Cost_to_income	Continious	
Housing expenses	totwlw_r	Totale woonlasten per maand na revisie	Transform to yearly costs for interpretation sake, as income is depicted as yearly income as well. Relabel to Housing expenses **per year and incorporate in Cost-to-income variable.**	Housing_expenses	Continious	
WOZ housing value	wozwaarde	WOZ-waarde huidige woning (Bron: LV WOZ, peildatum WOZ-waarde: 1 januari 2016),		WOZ_housing_value	Continious	
Housing quality	tonderho	(12.8) Woning is slecht onderhouden		Housing_quality	Ordinal	
Person-space ratio	CONSTRUCED		Number of persons in the household (aantalpp5) / Living area (oppwon7)	Person_space_ratio	Continious	
Living area	oppwon7	Woon Oppervlakte in 7 klassen		Living_area	Ordinal	
Neighbourhood characteristics (model 4)						
Neighbourhood satisfaction	twoonomg	(13.1) Tevredenheid met huidige woonomgeving		Neighbourhood_satisfaction	Ordinal	
Relationship with neighbours	conbuur1	(13.9) Ik heb veel contact met mijn directe buren		Relationship_neighbours	Ordinal	
Safety	brtveilig	(13.16) Bang in deze buurt om lastiggevallen of beroofd te		Safety	Ordinal	
Region (G4)	g4_2	G4 (ja/nee) (2)	Recode dummy into 0/1 and define other municipalites as base	Region_G4	Dummy	
Urbanity	stedgem	Stedelijkheid gemeente (5)	Redefine labels to match other WoON datasets in terms of spelling	Urbanity	Ordinal	
Year controls						
Year	Year_2018		Construced dummy for each cycle (2012 / 2015 / 2018)	Year	Dummy	
				remaining cases (N):		**67.290**

APPENDIX E: Multicollinearity

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Satisfaction_dummy	1	10.000														
Energy label	2	-0,0775	10.000													
Year	3	0,0303	-0,0094	10.000												
Age	4	0,122	-0,0213	0,03	10.000											
Ethnicity	5	-0,0913	0,0153	-0,0358	-0,0576	10.000										
Education	6	0,0394	-0,0008	0,1339	-0,3124	0,0262	10.000									
Disposable HH income	7	0,1156	-0,0471	0,16	-0,1387	-0,0386	0,3096	10.000								
Household type	8	-0,0345	-0,0129	0,0321	-0,3457	0,0409	0,1168	0,3119	10.000							
Number of persons in the HH	9	0,0124	-0,0516	0,0653	-0,3475	0,0183	0,1517	0,436	0,7581	10.000						
Amount of years lived at this adress	10	0,0893	0,166	0,0718	0,5536	-0,0701	-0,2126	-0,0583	-0,1462	-0,1414	10.000					
Self-reported health status	11	-0,1288	0,0024	-0,0519	0,2707	0,0359	-0,239	-0,2068	-0,1405	-0,1747	0,1196	10.000				
Tenure	12	-0,2312	0,0613	-0,2183	-0,0187	0,0947	-0,1623	-0,3274	-0,1317	-0,248	-0,1376	0,138	10.000			
Building age	13	0,0457	-0,6197	-0,0093	0,0474	-0,0165	-0,0485	0,0076	0,0007	0,0274	-0,0985	0,0244	-0,0486	10.000		
Housing type	14	-0,1518	0,0199	-0,1346	-0,0264	0,1049	-0,0507	-0,2602	-0,2529	-0,3279	-0,2583	0,1015	0,4053	0,0049	10.000	
Cost-to-income ratio	15	-0,0677	0,0342	0,0077	-0,0146	0,0482	-0,1229	-0,4327	-0,1804	-0,2351	-0,1452	0,1057	0,2962	-0,0299	0,1821	10.000

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Housing expenses	16	0,081	-0,029	0,2431	-0,2284	-0,0095	0,2723	0,4307	0,2606	0,3596	-0,2325	-0,1847	-0,1602	-0,0125	-0,1778	0,3212
Housing value (WOZ tax value)	17	0,1766	-0,0295	0,1362	0,0842	-0,0619	0,2263	0,4803	0,1496	0,2305	0,0742	-0,145	-0,3188	-0,0328	-0,3154	-0,1152
Housing quality	18	0,4048	-0,1784	0,0876	0,1054	-0,0894	0,0857	0,2002	-0,0013	0,0688	0,0634	-0,1353	-0,3495	0,1148	-0,1793	-0,1407
Person-space ratio	19	-0,1549	0,0523	-0,1306	-0,3443	0,0877	0,0082	0,0937	0,4852	0,5838	-0,2056	-0,0608	0,0923	-0,0458	0,085	-0,096
Square metres of living area	20	0,2081	-0,087	0,2069	0,0882	-0,0835	0,1803	0,4078	0,2392	0,3295	0,1636	-0,1374	-0,4304	0,0448	-0,5372	-0,1753
Satisfaction with the neighbourhood	21	-0,3393	0,0189	-0,0778	-0,1034	0,0508	-0,0377	-0,1147	0,0104	-0,0304	-0,0525	0,1346	0,1484	-0,0015	0,1317	0,0439
Relationship with neighbours	22	-0,1546	0,0113	-0,0473	-0,1415	0,0278	0,0524	-0,0561	-0,0269	-0,0816	-0,1225	0,0389	0,1473	-0,0083	0,142	0,0566
Region (G4)	23	-0,0887	0,1151	-0,0378	-0,0932	0,1333	0,0771	-0,0275	-0,0032	-0,0466	-0,0896	0,007	0,1838	-0,1347	0,325	0,0662
Urbanity level	24	0,096	-0,0649	0,0621	0,0814	-0,1369	-0,094	0,037	0,0281	0,078	0,1128	-0,0138	-0,188	0,0697	-0,3945	-0,0728

university of groningen

		16	17	18	19	20	21	22	23	24
Housing expenses	16	10.000								
Housing value (WOZ tax value)	17	0,4109	10.000							
Housing quality	18	0,1137	0,2325	10.000						
Person-space ratio	19	0,0646	-0,1481	-0,1301	10.000					
Square metres of living area	20	0,3263	0,5964	0,2633	-0,4066	10.000				
Satisfaction with the neighbourhood	21	-0,1013	-0,2096	-0,2325	0,0927	-0,1826	10.000			
Relationship with neighbours	22	-0,0294	-0,0926	-0,1336	0,0214	-0,1256	0,2698	10.000		
Region (G4)	23	0,011	-0,032	-0,1147	0,1339	-0,2254	0,0867	0,0566	10.000	
Urbanity level	24	-0,0052	0,0804	0,122	-0,1217	0,289	-0,1202	-0,0803	-0,4702	10.000

university of groningen

	Variable	VIF	1/VIF
Energy label	1	2.35	0.425554
	2	2.81	0.355407
	3	3.69	0.270639
	4	2.11	0.474659
	5	2.40	0.416304
	6	1.91	0.522804
	7	BASE	
Year	2012	BASE	
	2015	3.20	0.312211
	2018	3.32	0.301342
Age	1	BASE	
	2	6.21	0.161050
	3	7.23	0.138233
	4	8.71	0.114780
	5	9.01	0.110983
	6	8.53	0.117227
	7	6.88	0.145393
Ethnicity	1	BASE	
	2	1.19	0.841776
	3	1.02	0.978848
Education	1	BASE	
	2	1.54	0.650469
	3	1.82	0.550111
Disposable HH income		2.46	0.406168

	Variable	VIF	1/VIF
Household type	1	BASE	
	2	3.19	0.313354
	3	10.17	0.098328
	4	2.04	0.491348
	5	1.56	0.642016
Number of persons in the HH		11.08	0.090264
Amount of years lived at this address		1.82	0.548298
Self-reported health status	1	BASE	
	2	1.86	0.538284
	3	1.73	0.579312
	4	1.44	0.696033
	5	1.24	0.805507
Tenure	1	BASE	
	2	2.08	0.479731
	3	1.31	0.761247
Building age		1.71	0.584830
Housing type		2.15	0.464146
Cost-to-income ratio		2.48	0.403056
Housing expenses		2.80	0.357430
housing value (WOZ tax value)		2.21	0.453261
Housing quality	1	BASE	
	2	3.34	0.299397
	3	4.84	0.206517
	4	10.24	0.097645
	5	9.53	0.104979

	Variable	VIF	1/VIF
Person-space ratio		7.19	0.139072
Square metres of living area	1	BASE	
	2	5.82	0.171901
	3	11.65	0.085832
	4	21.05	0.047516
	5	18.35	0.054505
	6	14.64	0.068304
	7	10.38	0.096349
Satisfaction with the neighbourhood	1	BASE	
	2	1.42	0.703133
	3	1.35	0.738800
	4	1.19	0.840370
	5	1.09	0.916259
Relationship with neighbours	1	BASE	
	2	2.60	0.384523
	3	2.43	0.411044
	4	2.17	0.460966
	5	1.43	0.699353
Region (G4)		1.96	0.511371
Urbanity level	1	BASE	
	2	2.74	0.364678
	3	2.40	0.417474
	4	2.73	0.365903
	5	1.76	0.568522
	Mean VIF	**4.49**	

APPENDIX F: EXTENDED TABLE 5.1

	Model I	Model II	Model III	Model IV	Model V
Satisfaction dummy (0 = NS, 1 = S)					
A	2.100***	2.093***	2.416***	1.506***	1.734***
	(0.0983)	(0.0980)	(0.118)	(0.0870)	(0.105)
B	1.674***	1.720***	1.689***	1.245***	1.335***
	(0.0613)	(0.0631)	(0.0646)	(0.0582)	(0.0656)
C	1.133***	1.182***	1.249***	1.133**	1.219***
	(0.0335)	(0.0351)	(0.0389)	(0.0451)	(0.0511)
D	1.079*	1.267***	1.185***	0.987	1.023
	(0.0403)	(0.0489)	(0.0482)	(0.0474)	(0.0517)
E	0.496***	0.520***	0.696***	0.904**	0.988
	(0.0153)	(0.0161)	(0.0228)	(0.0355)	(0.0413)
F	1.099**	1.143***	1.107**	0.982	1.019
	(0.0399)	(0.0416)	(0.0420)	(0.0432)	(0.0470)
G	Base	Base	Base	Base	Base
Year=2012		Base	Base	Base	Base
Year=2015		1.689***	1.268***	0.831***	0.796***
		(0.0444)	(0.0356)	(0.0262)	(0.0268)
Year=2018		1.433***	0.962	0.624***	0.573***
		(0.0368)	(0.0268)	(0.0198)	(0.0194)
17-23 year			0.718***	0.918**	0.990
			(0.0184)	(0.0258)	(0.0295)
35-64 year			Base	Base	Base
65 years and older			1.962***	1.799***	1.517***
			(0.0526)	(0.0516)	(0.0460)
Native			Base	Base	Base
Non-Western			0.444***	0.651***	0.601***
			(0.0125)	(0.0204)	(0.0205)
Western			0.716***	0.806***	0.791***
			(0.0212)	(0.0256)	(0.0266)
Low educated			Base	Base	Base
Secondary educated			1.008	0.862***	0.916***
			(0.0230)	(0.0212)	(0.0238)
Well educated			1.112***	0.846***	0.941*
			(0.0283)	(0.0239)	(0.0280)
Low			0.541***	0.880***	0.883***
			(0.0131)	(0.0278)	(0.0292)
Middle			Base	Base	Base
High			1.868***	1.144*	1.193**
			(0.111)	(0.0742)	(0.0800)

Single person HH	Base	Base	Base
Couple	1.070* (0.0311)	1.010 (0.0327)	1.069 (0.0364)
Parent(s) with kids	0.699*** (0.0311)	0.744*** (0.0360)	0.791*** (0.0404)
Non family household	0.737*** (0.0466)	1.002 (0.0694)	1.154 (0.0850)
Number of persons in the hold	1.033* (0.0168)	1.031 (0.0211)	1.030 (0.0222)
Amount of years lived at this address	1.012*** (0.000880)	1.003** (0.000958)	1.005*** (0.00102)
Very well	Base	Base	Base
Good	0.758*** (0.0206)	0.755*** (0.0219)	0.829*** (0.0253)
Ok	0.436*** (0.0144)	0.482*** (0.0171)	0.565*** (0.0212)
Differs from time to time	0.351*** (0.0130)	0.428*** (0.0172)	0.519*** (0.0221)
Bad	0.276*** (0.0123)	0.363*** (0.0175)	0.465*** (0.0239)
Owner-occupier		Base	Base
Social rent		0.403*** (0.0112)	0.438*** (0.0128)
private rent		0.387*** (0.0137)	0.367*** (0.0136)
Age of the building		1.000 (0.000257)	1.000 (0.000267)
Single-family home		Base	Base
Multi-family home		0.939* (0.0247)	0.993 (0.0292)
Yearly housing expenses / Yearly Disposable HH income		1.523*** (0.150)	1.588*** (0.163)
Housing expenses		1.000 (0.0000040 5)	1.000* (0.0000041 6)
Housing value (WOZ tax value)		1.000*** (0.0000001 79)	1.000*** (0.0000001 77)
Agreed		Base	Base
Disagreed		5.275*** (0.126)	5.128*** (0.132)
Persons in HH / Living area		0.669***	0.637***

				(0.0262)	(0.0263)
up to 69m2				Base	Base
70 up to 119m2				1.042 (0.0356)	1.070 (0.0387)
over 120m2				1.188*** (0.0585)	1.250*** (0.0651)
Very satisfied					Base
Satisfied					0.548*** (0.0171)
Neither satisfied, nor dissatisfied					0.146*** (0.00516)
Dissatisfied					0.105*** (0.00452)
Very dissatisfied					0.0661*** (0.00442)
Not agreed					Base
Agreed					1.425*** (0.0308)
Other municipalities					Base
Big four (A'dam, R'dam, Utrecht, the Hague)					1.073* (0.0352)
Strong					Base
Moderately					0.921** (0.0278)
Little to none					0.929** (0.0256)
Constant	6.305*** (0.156)	4.155*** (0.140)	9.899*** (0.575)	3.438*** (0.310)	7.852*** (0.779)
Observations	116920	116920	116920	116920	116920
Log likelihood	-44575.003	-44382.441	-40984.716	-36153.697	-32671.842
df	6	8	25	36	44
Pseudo R^2	0.024	0.028	0.103	0.208	0.285

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Satisfaction dummy, independent variable Energy label with label G as base category. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively. Inclusion of control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics denoted with yes/no.

APPENDIX G: LIKELIHOOD RATIO TEST MODELS

Testing the model fit by adding control variables

lrtest Model_I Model_II

Likelihood-ratio test — LR chi2(2) = 385.12

(Assumption: Model_I nested in Model_II) — Prob > chi2 = 0.0000

lrtest Model_II Model_III

Likelihood-ratio test — LR chi2(17) = 6795.45

(Assumption: Model_II nested in Model_III) — Prob > chi2 = 0.0000

lrtest Model_III Model_IV

Likelihood-ratio test — LR chi2(11) = 9662.04

(Assumption: Model_III nested in Model_IV) — Prob > chi2 = 0.0000

lrtest Model_IV Model_V

Likelihood-ratio test — LR chi2(8) = 6963.71

(Assumption: Model_IV nested in Model_V) — Prob > chi2 = 0.0000

Housing satisfaction

Testing the model fit of the segmented regression models per Year (robustness & heterogeneity)

lrtest (Model_IV2012 Model_IV2015 Model_IV2018) Model_IV

Likelihood-ratio test — LR chi2(68) = 238.73

Prob > chi2 = 0.0000

Assumption: (Model_IV) nested in (Model_IV2012, Model_IV2015, Model_IV2018)

Testing the model fit of the interaction model VI compared to model V

lrtest Model_IV Model_VI_HS

Likelihood-ratio test — LR chi2(3) = 48.51

(Assumption: Model_VI_HS nested in Model_IV) — Prob > chi2 = 0.0000

Testing the model fit of the segmented regression models per Tenure (robustness & heterogeneity)

lrtest (Model_IVTenure1 Model_IVTenure2 Model_IVTenure3) Model_IV

Likelihood-ratio test — LR chi2(68) = 402.85

Prob > chi2 = 0.0000

Assumption: (Model_IV) nested in (Model_IVTenure1, Model_IVTenure2, Model_IVTenure3)

lrtest Model_IV Model_VII_HS

Likelihood-ratio test LR chi2(3) = 17.73

(Assumption: Model_VII_HS nested in Model_IV) Prob > chi2 = 0.0005

Residential Satisfaction

Testing the model fit of the segmented regression models per Year (robustness & heterogeneity)

lrtest (Model_V2012 Model_V2015 Model_V2018) Model_V

Likelihood-ratio test LR chi2(84) = 241.04

Prob > chi2 = 0.0000

Assumption: (Model_V) nested in (Model_V2012, Model_V2015, Model_V2018)

Testing the model fit of the interaction model VI compared to model V

lrtest Model_V Model_VI

Likelihood-ratio test LR chi2(3) = 52.86

(Assumption: Model_VI nested in Model_V) Prob > chi2 = 0.0000

Testing the model fit of the segmented regression models per Tenure (robustness & heterogeneity)

lrtest (Model_VTenure1 Model_VTenure2 Model_VTenure3) Model_V

Likelihood-ratio test LR chi2(84) = 464.09

Prob > chi2 = 0.0000

Assumption: (Model_V) nested in (Model_VTenure1, Model_VTenure2, Model_VTenure3)

lrtest Model_V Model_VII

Likelihood-ratio test LR chi2(3) = 37.84

(Assumption: Model_VII nested in Model_V) Prob > chi2 = 0.0000

APPENDIX H: Alternative extended table 5.2 & 5.7

All energy labels become significant when changing the interpretation of the regression output and assessing the likelihood of being satisfied with lower energy labels than label A. Appendix H reports the regression results of model V, but with label A as base category. All variables regarding the energy label become statistically significant at 1%. Consequently, all odds ratios are less than 1 indicating that a lower energy label decreases the likelihood of a household being satisfied. These results are in line with earlier conclusions based on model V indicating that green energy labels result in a higher likelihood of being satisfied, though it also indicates that the opposite reasoning might be plausible. Lower energy labels are making it less likely that a household is satisfied as well, or in other words inefficient homes make it more likely to be less satisfied. These results can be related to the fact that energy labels also serve as a representation of energy costs and living comfort (Ebrahimigharehbaghi et al, 2019; Wong, 2020).

Housing satisfaction	Model IV Alternative base energy label A
A	BASE
B	0.827***
	(0.0427)
C	0.752***
	(0.0355)
D	0.656***
	(0.0368)
E	0.600***
	(0.0300)
F	0.652***
	(0.0365)
G	0.664***
	(0.0384)
Year controls	Yes
Socio-demographics	Yes
Housing characteristics	Yes
Neighbourhood characteristics	No
Constant	2.459
	(1.294)
Observations	116920
Log likelihood	-36153.697
Df	36
Pseudo R^2	0.208

Note: Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable Energy label with **label A** as base category. Inclusion of control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics denoted with yes/no. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

Similar effects are visible for residential satisfaction, though again the effects are stronger compared to housing satisfaction. After all, the odds ratios are smaller, more below one, in the residential satisfaction model than in the housing satisfaction model.

Residential satisfaction	Model V Alternative base energy label A
A	BASE
B	0.770***
	(0.0417)
C	0.703***
	(0.0347)
D	0.590***
	(0.0347)
E	0.570***
	(0.0299)
F	0.588***
	(0.0345)
G	0.577***
	(0.0349)
Year controls	Yes
Socio-demographics	Yes
Housing characteristics	Yes
Neighbourhood characteristics	Yes
Constant	13.61***
	(1.407)
Observations	116920
Log likelihood	-32671.842
Df	44
Pseudo R^2	0.285

Note: Depending variable: Satisfaction dummy: Not Satisfied (0) Satisfied (1), independent variable Energy label with **label A** as base category. Inclusion of control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics denoted with yes/no. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively.

APPENDIX I: STATA do-file

```
*** PREPERATION OF DATASET WoON 2012 BEFORE MERGING WITH OTHER DATASETS ***

**Import original WoON 2012 data sheets
    use "C:\Users\roelb\OneDrive\Real Estate Studies - RUG\Master\Master's Thesis Real Estate Studies\STATA\STATA_WoON2012_e_1.1.dta"

**Generate new variable to indicate which year the datacycle belongs to
    generate Year_2012 = 2012
    label variable Year_2012 "2012"

**Keep relevant variables, drop everything else to get a clean dataset to work with
    keep TWoning e_label Leeftijd etniop3 VOplOP vromhh SAMHH5 AantalPP JrKomWon Gezond eigendom BJaar vorm totwl wozwaarde
    Tonderho oppwon7 TWoonOmg ConBuur1 Brtveilg g4_2 stedgem Year_2012

** Recode variables
        ren TWoning Satisfaction
        ren e_label Energy_label
        ren Leeftijd Age
        ren etniop3 Ethnicity
        ren VOplOP Education
        ren vromhh Disposable_HH__income
        ren SAMHH5 Household_type
        ren AantalPP Number_of_persons_in_HH
        ren JrKomWon Years_at_current_adress
        ren Gezond Self_reported_health
        ren eigendom Tenure
        ren BJaar Construction_year
        ren vorm Housing_type
        ren totwl Housing_expenses
        ren wozwaarde WOZ_housing_value
        ren Tonderho Housing_quality
        ren oppwon7 Living_area
        ren TWoonOmg Neighbourhood_satisfaction
        ren ConBuur1 Relationship_neighbours
        ren Brtveilg Safety
        ren g4_2 Region_G4
        ren stedgem Urbanity
        ren Year_2012 Year

** Recode Energy_label to combine the classes A++ A+ A into A
recode Energy_label (1=1) (2=1) (3=1) (4=2) (5=3) (6=4) (7=5) (8=6) (9=7)
label define Energy_label 1 "A" 2 "B" 3 "C" 4 "D" 5 "E" 6 "F" 7 "G"
label values Energy_label

** Recode Education
    recode Education (1/5=1) (6/8=2) (9/10=3)
    drop if Education==11
    label variable Education "Education level"
    label define Education_label 1 "Low educated" 2 "Secondary educated" 3 "Well educated"
    label values Education Education_label

** Recode labelname Household_type to fit with datasets 2015&2018
label define Household_type_label 1 "eenpersoonshuishouden" 2 "paar" 3 "paar + kind(eren)" 4 "1-oudergezin" 5 "niet-gezinshuishouden"

label values Household_type Household_type_label

** Recode labelname Number_of_persons_in_HH to fit with datasets 2018
recode Number_of_persons_in_HH (0/1=1) (2=2) (3=3) (4=4) (5/max=5)
label define Number_of_persons_in_HH_label 1 "1 persoon" 2 "2 personen" 3 "3 personen" 4 "4 personen" 5 "5 of meer personen"
label values Number_of_persons_in_HH Number_of_persons_in_HH_label

** Transform variable “since when do you live at this address” into the amount of years you’ve lived at this address
replace Years_at_current_adress = 2012-Years_at_current_adress
label variable Years_at_current_adress "Years at current adress"
```

```
** Recode labelname self-reported health status to fit with datasets 2015&2018
label define Self_reported_health_label 1 "Zeer goed" 2 "Goed" 3 "Gaat wel" 4"Soms goed en soms slecht" 5 "Slecht"
label values Self_reported_health Self_reported_health_label

** Recode labelname Tenure to make it comparable with datasets 2015&2018 & drop unkown observations
label define Tenure_label 1 "Koopwoning" 2 "Sociale huur" 3 "Particuliere huur" 5 "Onbekend (OP geen lid van hh)"
label values Tenure Tenure_label

** Recode labelname Housing_quality to fit with datasets 2015&2018
label define Housing_quality_label 1 "Helemaal mee eens" 2 "Mee eens" 3 "Niet mee eens, maar ook niet mee oneens" 4 "Mee oneens" 5
"Helemaal mee oneens"
label value Housing_quality Housing_quality_label

** Recode labelname Neighbourhood_satisfaction to fit with datasets 2015&2018
label define Neighbourhood_satisfaction_label 1 "Zeer tevreden" 2 "Tevreden" 3 "Niet tevreden, maar ook niet ontevreden" 4 "Ontevreden" 5
"Zeer ontevreden"
label value Neighbourhood_satisfaction Neighbourhood_satisfaction_label

** Recode labelname Relationship_neighbours to fit with datasets 2015&2018
label define Relationship_neighbours_label 1 "Helemaal mee eens" 2 "Mee eens" 3 "Niet mee eens, maar ook niet mee oneens" 4 "Mee oneens"
5 "Helemaal mee oneens"
label value Relationship_neighbours Relationship_neighbours_label

** Replace Housing_expenses per month for Housing expenses per year, for clearer interpretation
replace Housing_expenses = Housing_expenses * 12
label variable Housing_expenses "Housing expenses per year"

** Generate variable cost to income by dividing yearly housing cost by yearly disposable income
generate Cost_to_income = Housing_expenses / Disposable_HH__income
label variable Cost_to_income "Yearly housing expenses / Yearly Disposable HH income"

** Recode and relabel dummy Region G4, other municipalities as base category.
recode Region_G4 (1=1) (2=0)
label define Region_G4_label 0 "overige gemeenten" 1 "vier grote steden"
label values Region_G4 Region_G4_label

** Recode and relabel dummy housing type, 'eengezinswoning' as base category
recode Housing_type (1=0) (2=1)
label define Housing_type_label 0 "eengezinswoning" 1 "meergezinswoning"
label values Housing_type Housing_type_label

** Relabel variables to match datasets
label variable Satisfaction "Satisfaction level"
label variable Energy_label "Energy label
label variable Age "Age
label variable Ethnicity "Ethnicity
label variable Education "Level of education
label variable Disposable_HH__income "Disposable household income
label variable Household_type "Household type
label variable Number_of_persons_in_HH "Number of persons in the hold
label variable Years_at_current_adress "Amount of years lived at this adress
label variable Self_reported_health "Self reported health status
label variable Tenure "Tenure
label variable Construction_year "Building period
label variable Housing_type "Housing type
label variable Housing_expenses "Housing expenses
label variable WOZ_housing_value "Housing value (WOZ tax value)
label variable Housing_quality "Housing quality
label variable Living_area "Square metres of living area
label variable Neighbourhood_satisfaction "Satisfaction with neighbourhood
label variable Relationship_neighbours "Relationship with neighbours
label variable Safety "Safety
label variable Region_G4 "Region
label variable Urbanity "Urbanity level
label variable Year "Year
```

*** PREPERATION OF DATASET WoON 2015 BEFORE MERGING WITH OTHER DATASETS ***

```
**Import original WoON 2015 data sheets
    use C:\Users\roelb\OneDrive\Real Estate Studies - RUG\Master\Master's Thesis Real Estate Studies\STATA\Log STATA_2015 prep.smcl

**Generate new variable to indicate which year the datacycle belongs to
    generate Year_2015 = 2015
    label variable Year_2015 "2015"

**Keep relevant variables, drop everything else to get a clean dataset to work with
    keep TWoning Energieklasse_vlp Leeftijd etniop3 NivBehOP vromhh SamHH5 AantalPP JrKomWon Gezond Eigendom bjaarbag vorm totwl
    WOZwaarde Tonderho OppWon7 TWoonOmg ConBuur1 Brtveilig g4_2 stedgem Year_2015

** Recode variables
        ren TWoning Satisfaction
        ren Energieklasse_vlp Energy_label
        ren Leeftijd Age
        ren etniop3 Ethnicity
        ren NivBehOP Education
        ren vromhh Disposable_HH__income
        ren SamHH5 Household_type
        ren AantalPP Number_of_persons_in_HH
        ren JrKomWon Years_at_current_adress
        ren Gezond Self_reported_health
        ren Eigendom Tenure
        ren bjaarbag Construction_year
        ren vorm Housing_type
        ren totwl Housing_expenses
        ren WOZwaarde WOZ_housing_value
        ren Tonderho Housing_quality
        ren OppWon7 Living_area
        ren TWoonOmg Neighbourhood_satisfaction
        ren ConBuur1 Relationship_neighbours
        ren Brtveilig Safety
        ren g4_2 Region_G4
        ren stedgem Urbanity
        ren Year_2015 Year

** Redefine labels Energy_label to match with other datasets
label define Energy_label 1 "A" 2 "B" 3 "C" 4 "D" 5 "E" 6 "F" 7 "G"
label values Energy_label

** Combine multiple categories regarding Education of respondent and merge into new categorical variable Education (low educated, secondary
educated, well educated)
    recode Education (1/5=1) (6/8=2) (9/10=3)
    drop if Education==11
    label variable Education "Education level"
    label define Education_label 1 "Low educated" 2 "Secondary educated" 3 "Well educated"
    label values Education Education_label

** Recode labelname Number_of_persons_in_HH to fit with datasets 2018
recode Number_of_persons_in_HH (0/1=1) (2=2) (3=3) (4=4) (5/max=5)
label define Number_of_persons_in_HH_label 1 "1 persoon" 2 "2 personen" 3 "3 personen" 4 "4 personen" 5 "5 of meer personen"
label values Number_of_persons_in_HH Number_of_persons_in_HH_label

** Transform variable “since when do you live at this address” into the amount of years you’ve lived at this address
replace Years_at_current_adress = 2015-Years_at_current_adress
label variable Years_at_current_adress "Years at current adress"

** Recode labelname Tenure to make it comparable with datasets 2012&2018
label define Tenure_label 1 "Koopwoning" 2 "Sociale huur" 3 "Particuliere huur" 5 "Onbekend (OP geen lid van hh)"
label values Tenure Tenure_label

** Replace Housing_expenses per month for Housing expenses per year, for clearer interpretation
replace Housing_expenses = Housing_expenses * 12
label variable Housing_expenses "Housing expenses per year"
```

```
** Generate variable cost to income by dividing yearly housing cost by yearly disposable income
generate Cost_to_income = Housing_expenses / Disposable_HH__income
label variable Cost_to_income "Yearly housing expenses / Yearly Disposable HH income"

** Recode and relabel dummy Region G4, other municipalities as base category.
recode Region_G4 (1=1) (2=0)
label define Region_G4_label 0 "overige gemeenten" 1 "vier grote steden"
label values Region_G4 Region_G4_label

** Recode and relabel dummy housing type, 'eengezinswoning' as base category
recode Housing_type (1=0) (2=1)
label define Housing_type_label 0 "eengezinswoning" 1 "meergezinswoning"
label values Housing_type Housing_type_label

** Relabel variables to match datasets
label variable Satisfaction "Satisfaction level"
label variable Energy_label "Energy label
label variable Age "Age
label variable Ethnicity "Ethnicity
label variable Education "Level of education
label variable Disposable_HH__income "Disposable household income
label variable Household_type "Household type
label variable Number_of_persons_in_HH "Number of persons in the hold
label variable Years_at_current_adress "Amount of years lived at this adress
label variable Self_reported_health "Self reported health status
label variable Tenure "Tenure
label variable Construction_year "Building period
label variable Housing_type "Housing type
label variable Housing_expenses "Housing expenses
label variable WOZ_housing_value "Housing value (WOZ tax value)
label variable Housing_quality "Housing quality
label variable Living_area "Square metres of living area
label variable Neighbourhood_satisfaction "Satisfaction with neighbourhood
label variable Relationship_neighbours "Relationship with neighbours
label variable Safety "Safety
label variable Region_G4 "Region
label variable Urbanity "Urbanity level
label variable Year "Year

*** PREPERATION OF DATASET WoON 2018 BEFORE MERGING WITH OTHER DATASETS ***

**Import original WoON 2018 data sheets
    use C:\Users\roelb\OneDrive\Real Estate Studies - RUG\Master\Master's Thesis Real Estate Studies\STATA\Log STATA_2018 prep.smcl

**Generate new variable to indicate which year the datacycle belongs to
    generate Year_2018 = 2018
    label variable Year_2018 "2018"

**Keep relevant variables, drop everything else to get a clean dataset to work with
    keep twoning energieklasse_vlp leeftijd etniop3 nivbehop1 vromhh_r samhh5 aantalpp5 jrkomwon gezond eigendom bjaarbag vorm totwlw_r wozwaarde tonderho oppwon7 twoonomg conbuur1 brtveilig g4_2 stedgem Year_2018 nivbehop2 nivbehop3 nivbehop4 nivbehop5 nivbehop6 nivbehop7 nivbehop8 nivbehop9 nivbehop10 nivbehop11 nivbehop12 nivbehop13 nivbehop14 nivbehop15 nivbehop16 nivbehop17

** Recode variables
        ren twoning Satisfaction
        ren energieklasse_vlp Energy_label
        ren leeftijd Age
        ren etniop3 Ethnicity
        ren nivbehop1 EDUC_DUMMY1
        ren vromhh_r Disposable_HH__income
        ren samhh5 Household_type
        ren aantalpp5 Number_of_persons_in_HH
        ren jrkomwon Years_at_current_adress
        ren gezond Self_reported_health
```

```
ren eigendom Tenure
ren bjaarbag Construction_year
ren vorm Housing_type
ren totwlw_r Housing_expenses
ren wozwaarde WOZ_housing_value
ren tonderho Housing_quality
ren oppwon7 Living_area
ren twoonomg Neighbourhood_satisfaction
ren conbuur1 Relationship_neighbours
ren brtveilig Safety
ren g4_2 Region_G4
ren stedgem Urbanity
ren Year_2018 Year
ren nivbehop2 EDUC_DUMMY2
ren nivbehop3 EDUC_DUMMY3
ren nivbehop4 EDUC_DUMMY4
ren nivbehop5 EDUC_DUMMY5
ren nivbehop6 EDUC_DUMMY6
ren nivbehop7 EDUC_DUMMY7
ren nivbehop8 EDUC_DUMMY8
ren nivbehop9 EDUC_DUMMY9
ren nivbehop10 EDUC_DUMMY10
ren nivbehop11 EDUC_DUMMY11
ren nivbehop12 EDUC_DUMMY12
ren nivbehop13 EDUC_DUMMY13
ren nivbehop14 EDUC_DUMMY14
ren nivbehop15 EDUC_DUMMY15
ren nivbehop16 EDUC_DUMMY16
ren nivbehop17 EDUC_DUMMY17

** Redefine labels Energy_label to match with other datasets
label define Energy_label 1 "A" 2 "B" 3 "C" 4 "D" 5 "E" 6 "F" 7 "G"
label values Energy_label

** Combine multiple dummies regarding Education of respondent and merge into new categorical variable Education (low educated, secondary educated, well educated)
generate Education=0
replace Education=1 if EDUC_DUMMY1==1
replace Education=2 if EDUC_DUMMY2==1
replace Education=3 if EDUC_DUMMY3==1
replace Education=4 if EDUC_DUMMY4==1
replace Education=5 if EDUC_DUMMY5==1
replace Education=6 if EDUC_DUMMY6==1
replace Education=7 if EDUC_DUMMY7==1
replace Education=8 if EDUC_DUMMY8==1
replace Education=9 if EDUC_DUMMY9==1
replace Education=10 if EDUC_DUMMY10==1
replace Education=11 if EDUC_DUMMY11==1
replace Education=12 if EDUC_DUMMY12==1
replace Education=13 if EDUC_DUMMY13==1
replace Education=14 if EDUC_DUMMY14==1
replace Education=15 if EDUC_DUMMY15==1
replace Education=16 if EDUC_DUMMY16==1
drop if EDUC_DUMMY17==1
recode Education (1/4=1) (5/10=2) (11/16=3)
label variable Education "Education level"
label define Education_label 1 "Low educated" 2 "Secondary educated" 3 "Well educated"
label values Education Education_label

** Transform variable “since when do you live at this address” into the amount of years you’ve lived at this address
replace Years_at_current_adress = 2018-Years_at_current_adress
label variable Years_at_current_adress "Years at current adress"

** Recode labelname Tenure to make it comparable with datasets 2012&2015
label define Tenure_label 1 "Koopwoning" 2 "Sociale huur" 3 "Particuliere huur" 6 "Onbekend (OP geen lid van hh)"
label values Tenure Tenure_label
```

```
** Recode labelname Urbanity to fit with datasets 2012&2015
label define Urbanity_label 1 "Zeer sterk stedelijk" 2 "Sterk stedelijk" 3 "Matig stedelijk" 4 "Weinig stedelijk" 5 "Niet stedelijk"
label value Urbanity Urbanity_label

** Replace Housing_expenses per month for Housing expenses per year, for clearer interpretation
replace Housing_expenses = Housing_expenses * 12
label variable Housing_expenses "Housing expenses per year"

** Generate variable cost to income by dividing yearly housing cost by yearly disposable income
generate Cost_to_income = Housing_expenses / Disposable_HH__income
label variable Cost_to_income "Yearly housing expenses / Yearly Disposable HH income"

** Recode and relabel dummy Region G4, other municipalities as base category.
recode Region_G4 (1=1) (2=0)
label define Region_G4_label 0 "overige gemeenten" 1 "vier grote steden"
label values Region_G4 Region_G4_label

** Recode and relabel dummy housing type, 'eengezinswoning' as base category
recode Housing_type (1=0) (2=1)
label define Housing_type_label 0 "eengezinswoning" 1 "meergezinswoning"
label values Housing_type Housing_type_label

** Relabel variables to match datasets
label variable Satisfaction "Satisfaction level"
label variable Energy_label "Energy label
label variable Age "Age
label variable Ethnicity "Ethnicity
label variable Education "Level of education
label variable Disposable_HH__income "Disposable household income
label variable Household_type "Household type
label variable Number_of_persons_in_HH "Number of persons in the hold
label variable Years_at_current_adress "Amount of years lived at this adress
label variable Self_reported_health "Self reported health status
label variable Tenure "Tenure
label variable Construction_year "Building period
label variable Housing_type "Housing type
label variable Housing_expenses "Housing expenses
label variable WOZ_housing_value "Housing value (WOZ tax value)
label variable Housing_quality "Housing quality
label variable Living_area "Square metres of living area
label variable Neighbourhood_satisfaction "Satisfaction with neighbourhood
label variable Relationship_neighbours "Relationship with neighbours
label variable Safety "Safety
label variable Region_G4 "Region
label variable Urbanity "Urbanity level
label variable Year "Year

*** MERGE DATASETS WoON 2012 & WoON 2015 & WoON 2018, SAFE AS POOLED DATASET ***
append using "C:\Users\roelb\Desktop\Final version 2015.dta" "C:\Users\roelb\Desktop\Final version 2012.dta"
(label Housing_type_label already defined)
(label Region_G4_label already defined)
(label Tenure_label already defined)
(label Education_label already defined)
(label TWONING already defined)
(label TONDERHO already defined)
(label TWOONOMG already defined)
(label CONBUUR1 already defined)
(label BRTVEILI already defined)
(label JRKOMWON already defined)
(label GEZOND already defined)
(label ETNIOP3 already defined)
(label SAMHH5 already defined)
(label LEEFTIJD already defined)
(label OPPWON7 already defined)
(note: variable Urbanity was byte, now double to accommodate using data's values)
(note: variable Region_G4 was byte, now double to accommodate using data's values)
```

```
(note: variable Number_of_persons_in_HH was byte, now double to accommodate using data's values)
(note: variable Household_type was byte, now double to accommodate using data's values)
(note: variable Construction_year was int, now double to accommodate using data's values)
(note: variable Self_reported_health was byte, now double to accommodate using data's values)
(note: variable Satisfaction was byte, now double to accommodate using data's values)
(note: variable Housing_quality was byte, now double to accommodate using data's values)
(note: variable Neighbourhood_satisfaction was byte, now double to accommodate using data's values)
(note: variable Relationship_neighbours was byte, now double to accommodate using data's values)
(note: variable Safety was byte, now double to accommodate using data's values)
(note: variable Years_at_current_adress was int, now double to accommodate using data's values)
(note: variable Education was float, now double to accommodate using data's values)
(note: variable Age was byte, now double to accommodate using data's values)
(note: variable WOZ_housing_value was long, now double to accommodate using data's values)
(note: variable Tenure was byte, now double to accommodate using data's values)
(note: variable Housing_type was byte, now double to accommodate using data's values)
(note: variable Living_area was byte, now double to accommodate using data's values)
(note: variable Ethnicity was byte, now double to accommodate using data's values)
(note: variable Energy_label was byte, now double to accommodate using data's values)
(label Housing_type_label already defined)
(label Region_G4_label already defined)
(label Tenure_label already defined)
(label Number_of_persons_in_HH_label already defined)
(label Education_label already defined)

*** safe as Final version pooled dataset
save "C:\Users\roelb\Desktop\Final version pooled dataset.dta"
file C:\Users\roelb\Desktop\Final version pooled dataset.dta saved

***  Workfile: Final version pooled dataset                                  ***

*** Data cleaning
**Identify missing values using mdesc
mdesc
drop if Satisfaction==.
drop if Energy_label==.
mdesc
drop if Housing_type==.
drop if Living_area==.
drop if Housing_expenses==.
drop if Cost_to_income==.
mdesc

**
**Exclude tenure unkown from analysis
tabulate Tenure
tabulate Tenure, nol
drop if Tenure==5
drop if Tenure==6
tabulate Tenure

**Delete negative and odd values housing expense: negative or over 5.000 a month so 60.000 a year
sum
drop if Housing_expenses<0
drop if Housing_expenses>60000

** drop if disposable household income is lower than social assistance benefit 2012 (12*668,21=8018) or over a quarter of a million euro's
drop if Disposable_HH__income<8018

*** Data adjustments
**Recode Satisfaction into dummies
recode Satisfaction (1=1) (2=1) (3=0) (4=0) (5=0), gen(Satisfaction_dummy)
label define Satisfaction_dummy_label 1 "Satisfied" 0 "Not satisfied"
label values Satisfaction_dummy Satisfaction_dummy_label
label variable Satisfaction_dummy "Satisfaction dummy (0 = NS, 1 = S)

** Generate variable person-space ratio by dividing the number of persons in the household by the size of the living area
```

```
gen Person_space_ratio = Number_of_persons_in_HH / Living_area
label variable Person_space_ratio "Persons in HH / Living area"

**Drop dummy regarding educational level, needed for creating the Education variable as used in the dataset
keep Satisfaction Satisfaction_dummy Energy_label Year Age Ethnicity Education Disposable_HH__income Household_type Number_of_persons_in_HH Years_at_current_adress Self_reported_health Tenure Construction_year Housing_type Cost_to_income Housing_expenses WOZ_housing_value Housing_quality Person_space_ratio Living_area Neighbourhood_satisfaction Relationship_neighbours Region_G4 Urbanity

**Order the dataset according to the model specifications for the regression equation
order Satisfaction Satisfaction_dummy Energy_label Year Age Ethnicity Education Disposable_HH__income Household_type Number_of_persons_in_HH Years_at_current_adress Self_reported_health Tenure Construction_year Housing_type Cost_to_income Housing_expenses WOZ_housing_value Housing_quality Person_space_ratio Living_area Neighbourhood_satisfaction Relationship_neighbours Region_G4 Urbanity

***  Test assumptions                                                    ***
**Generate descriptive statistics table
sum

**Correlation matrix
corr Satisfaction_dummy Energy_label Year Age Ethnicity Education Disposable_HH__income Household_type Number_of_persons_in_HH Years_at_current_adress Self_reported_health Tenure Construction_year Housing_type Cost_to_income Housing_expenses WOZ_housing_value Housing_quality Person_space_ratio Living_area Neighbourhood_satisfaction Relationship_neighbours Region_G4 Urbanity

**Linear regression
reg Satisfaction_dummy ib7.Energy_label i.Year i.Age i.Ethnicity i.Education Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity

**vif
estat vif

**recoding variables to solve multicollinearity issues
tabulate Living_area
tabulate Living_area, nol
recode Living_area (1=1) (2=1) (3=2) (4=2) (5=3) (6=3) (7=3)
label define Living_area_label 1 "up to 69m2" 2 "70 up to 119m2" 3 "over 120m2"
label values Living_area Living_area_label

tabulate Housing_quality
tabulate Housing_quality, nol
recode Housing_quality (1=0) (2=0) (3=1) (4=1) (5=1)
label define Housing_quality_label 0 “Agreed” 1 “Disagreed”
label values Housing_quality Housing_quality_label

tabulate Household_type
tabulate Household_type, nol
recode Household_type (1=1) (2=2) (3=3) (4=3) (5=4)
label define Household_type 1 "Single person HH" 2 "Couple" 3 "Parent(s) with kids" 4 "Non family household"
label values Household_type

**Reducing amount of parameters in control variables
tabulate Relationship_neighbours
tabulate Relationship_neighbours, nol
recode Relationship_neighbours (1=1) (2=1) (3=0) (4=0) (5=0)
label define Relationship_neighbours_label 1 "Agreed" 0 "Not agreed"
label values Relationship_neighbours Relationship_neighbours_label

tabulate Age
tabulate Age, nol
recode Age (1=1) (2=1) (3=2) (4=2) (5=2) (6=3) (7=3)
label define Age_label 1 "17-23 year" 2 "35-64 year" 3 "65 years and older"
label values Age Age_label
```

```
tabulate Urbanity
tabulate Urbanity, nol
recode Urbanity (1=1) (2=1) (3=2) (4=3) (5=3)
label define Urbanity_namecat 1 "Strong" 2 "Moderately" 3 "Little to none"
label values Urbanity Urbanity_namecat

recode Disposable_HH__income (0/35000=1) (35001/75000=2) (75001/max=3)
label define Disposable_HH_income_label 1 “Low” 2 “Middle” 3 “High”
label values Disposable_HH__income Disposable_HH__income_label

reg Satisfaction_dummy ib7.Energy_label i.Year i.Age i.Ethnicity i.Education i.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity
estat vif

**Translate labels
label variable Housing_quality "Poor housing quality

tabulate Ethnicity
label define Ethnicity_label 1 "Native" 2 "Non-Western" 3 "Western"
label values Ethnicity Ethnicity_label

tabulate Self_reported_health
label define Self_reported_health_label 1 "Very well" 2 "Good" 3 "Ok" 4 "Differs from time to time" 5 "Bad"
label values Self_reported_health Self_reported_health_label

tabulate Tenure
label define Tenurelabel 1"Owner-occupier" 2 "Social rent" 3"private rent"
label values Tenure Tenurelabel

tabulate Neighbourhood_satisfaction
label define Neighbourhood_satisfaction_label 1 "Very satisfied" 2 "Satisfied" 3 "Neither satisfied, nor dissatisfied" 4 "Dissatisfied" 5 "Very dissatisfied"
label values Neighbourhood_satisfaction Neighbourhood_satisfaction_label

tabulate Region_G4
label define Region_G4Label 0"Other municipalities" 1 "Big four (A'dam, R'dam, Utrecht, the Hague)"
label values Region_G4 Region_G4Label

tabulate Housing_type
label define Housing_type_Label  0"Single-family home" 1 "Multi-family home"
label values Housing_type Housing_type_Label

**Generate adjusted descriptive statistics table
sum i.Satisfaction_dummy i.Energy_label i.Year i.Age i.Ethnicity i.Education i.Disposable_HH__income i.Household_type i.Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity

by Year, sort : summarize Satisfaction_dummy Energy_label Age Ethnicity Education Disposable_HH__income Household_type Number_of_persons_in_HH Years_at_current_adress Self_reported_health Tenure Construction_year Housing_type Cost_to_income Housing_expenses WOZ_housing_value Housing_quality Person_space_ratio Living_area Neighbourhood_satisfaction Relationship_neighbours Region_G4 Urbanity

***  Analysis                                                    ***
**Generate first analyse table, comparing the model fit
logistic Satisfaction_dummy ib7.Energy_label
estimate store Model_I

logistic Satisfaction_dummy ib7.Energy_label i.Year
estimate store Model_II

logistic Satisfaction_dummy ib7.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health
estimate store Model_III
```

```
logistic Satisfaction_dummy ib7.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area
estimate store Model_IV

logistic Satisfaction_dummy ib7.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity
estimate store Model_V

esttab Model_I Model_II Model_III Model_IV Model_V using NEWTable5.1.rtf, con b se pr2 eform label nonumbers mtitles("Model I" "Model II" "Model III" "Model IV" "Model V")

**LRtest models
lrtest Model_I Model_II
lrtest Model_II Model_III
lrtest Model_III Model_IV
lrtest Model_IV Model_V

**Run alternative model V
logistic Satisfaction_dummy ib1.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity

estimate store Model_V_baselabelA

esttab Model_V_baselabelA using NEWTable5.2_baselabelA.rtf, con b se pr2 eform label nonumber mtitles("Model V Alternative base energylabel A")

***HOUSING SATISFACTION

esttab Model_IV using NEWTableFull_HS.rtf, con b se pr2 eform label nonumbers mtitles("Model IV")

**Run alternative model
logistic Satisfaction_dummy ib1.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area

estimate store Model_IV_HS_baselabelA

esttab Model_IV_HS_baselabelA using NEWTable5.2_baselabelA_HS.rtf, con b se pr2 eform label nonumber mtitles("Model IV Alternative base energylabel A")

**Run robustness checks, segmenting the logistic regression into years HOUSING SATISFACTION
logistic Satisfaction_dummy ib7.Energy_label ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area if Year==2012

estimate store Model_IV2012

logistic Satisfaction_dummy ib7.Energy_label ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area if Year==2015

estimate store Model_IV2015

logistic Satisfaction_dummy ib7.Energy_label ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area if Year==2018

estimate store Model_IV2018

esttab Model_IV Model_IV2012 Model_IV2015 Model_IV2018 using NEWTableHS5.3.rtf, con b se pr2 eform label nonumbers mtitles("Pooled
```

model IV" “2012” “2015” “2018”)

**LR test to estimate the robustness of the model by segmenting the data into years HOUSING SATISFACTION
lrtest (Model_IV2012 Model_IV2015 Model_IV2018) Model_IV

**Create new variable Energy efficient green label vs Energy inefficient
recode Energy_label (1=1) (2=1) (3=1) (4=2) (5=2) (6=2) (7=2), gen(Energy_efficient)
label define Energy_efficient_label 1 "Energy efficient green label" 2 "Energy inefficient"
label values Energy_efficient Energy_efficient_label

**Run new logistic regression model including the interaction variable of interest: Year*energy efficiency HOUSING SATISFACTION
logistic Satisfaction_dummy i.Year##ib2.Energy_efficient ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area
estimate store Model_VI_HS

lrtest Model_IV Model_VI_HS

esttab Model_VI_HS using NEWTableHS5.4.rtf, con b se pr2 eform label nonumber mtitles("Pooled model with interaction effects")

**Generate regression analysis with pooled model V and segregated models based on Tenure HOUSING SATISFACTION
logistic Satisfaction_dummy ib7.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area if Tenure==1

estimate store Model_IVTenure1

logistic Satisfaction_dummy ib7.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area if Tenure==2

estimate store Model_IVTenure2

logistic Satisfaction_dummy ib7.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area if Tenure==3

estimate store Model_IVTenure3

esttab Model_IV Model_IVTenure1 Model_IVTenure2 Model_IVTenure3 using NEWTableHS5.5.rtf, con b se pr2 eform label nonumbers mtitles("Pooled model IV" “Owner-occupier” “Social rent” “Private rent”)

**LRtest segmented models based on Tenure
lrtest (Model_IVTenure1 Model_IVTenure2 Model_IVTenure3) Model_IV

**Run new logistic regression model including the interaction variable of interest: Tenure*energy efficiency HOUSING SATISFACTION
logistic Satisfaction_dummy i.Tenure##ib2.Energy_efficient i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area
estimate store Model_VII_HS

lrtest Model_IV Model_VII_HS

esttab Model_VII_HS using NEWTableHS5.6.rtf, con b se pr2 eform label nonumber mtitles("Pooled model with interaction effects")

***RESIDENTIAL SATISFACTION
**Run robustness checks, segmenting the logistic regression into years RESIDENTIAL SATISFACTION
logistic Satisfaction_dummy ib7.Energy_label ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity if Year==2012

estimate store Model_V2012

logistic Satisfaction_dummy ib7.Energy_label ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity if Year==2015

estimate store Model_V2015

logistic Satisfaction_dummy ib7.Energy_label ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity if Year==2018

estimate store Model_V2018

esttab Model_V Model_V2012 Model_V2015 Model_V2018 using NEWTable5.3.rtf, con b se pr2 eform label nonumbers mtitles("Pooled model V" “2012” “2015” “2018”)

**LR test to estimate the robustness of the model by segmenting the data into years RESIDENTIAL SATISFACTION
lrtest (Model_V2012 Model_V2015 Model_V2018) Model_V

~~**Create new variable Energy efficient green label vs Energy inefficient~~
~~recode Energy_label (1=1) (2=1) (3=1) (4=2) (5=2) (6=2) (7=2), gen(Energy_efficient)~~
~~label define Energy_efficient_label 1 "Energy efficient green label" 2 "Energy inefficient"~~
~~label values Energy_efficient Energy_efficient_label~~

**Run new logistic regression model including the interaction variable of interest: Year*energy efficiency RESIDENTIAL SATISFACTION
logistic Satisfaction_dummy i.Year##ib2.Energy_efficient ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity
estimate store Model_VI

lrtest Model_V Model_VI

esttab Model_VI using NEWTable5.4.rtf, con b se pr2 eform label nonumber mtitles("Pooled model with interaction effects")

**Generate regression analysis with pooled model V and segregated models based on Tenure RESIDENTIAL SATISFACTION
logistic Satisfaction_dummy ib7.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity if Tenure==1

estimate store Model_VTenure1

logistic Satisfaction_dummy ib7.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity if Tenure==2

estimate store Model_VTenure2

logistic Satisfaction_dummy ib7.Energy_label i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity if Tenure==3

estimate store Model_VTenure3

esttab Model_V Model_VTenure1 Model_VTenure2 Model_VTenure3 using NEWTable5.5.rtf, con b se pr2 eform label nonumbers mtitles("Pooled model V" “Owner-occupier” “Social rent” “Private rent”)

**LRtest segmented models based on Tenure
lrtest (Model_VTenure1 Model_VTenure2 Model_VTenure3) Model_V

**Run new logistic regression model including the interaction variable of interest: Tenure*energy efficiency RESIDENTIAL SATISFACTION

```
logistic Satisfaction_dummy i.Tenure##ib2.Energy_efficient i.Year ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity
estimate store Model_VII

lrtest Model_V Model_VII

esttab Model_VII using NEWTable5.6.rtf, con b se pr2 eform label nonumber mtitles("Pooled model with interaction effects")
```

APPENDIX J: LIFE SATISFACTION

STATA COMMANDS

```
***  Appendix life satisfaction                                                    ***
**Import original WoON 2015 data sheets
    use C:\Users\roelb\OneDrive\Real Estate Studies - RUG\Master\Master's Thesis Real Estate Studies\STATA\Log STATA_2015 prep.smcl

**Keep relevant variables, drop everything else to get a clean dataset to work with
    keep TWoning Energieklasse_vlp Leeftijd etniop3 NivBehOP vromhh SamHH5 AantalPP JrKomWon Gezond Eigendom bjaarbag vorm totwl
    WOZwaarde Tonderho OppWon7 TWoonOmg ConBuur1 Brtveilig g4_2 stedgem Leven

** Recode variables
        ren TWoning Residential_Satisfaction
        ren Energieklasse_vlp Energy_label
        ren Leeftijd Age
        ren etniop3 Ethnicity
        ren NivBehOP Education
        ren vromhh Disposable_HH__income
        ren SamHH5 Household_type
        ren AantalPP Number_of_persons_in_HH
        ren JrKomWon Years_at_current_adress
        ren Gezond Self_reported_health
        ren Eigendom Tenure
        ren bjaarbag Construction_year
        ren vorm Housing_type
        ren totwl Housing_expenses
        ren WOZwaarde WOZ_housing_value
        ren Tonderho Housing_quality
        ren OppWon7 Living_area
        ren TWoonOmg Neighbourhood_satisfaction
        ren ConBuur1 Relationship_neighbours
        ren Brtveilig Safety
        ren g4_2 Region_G4
        ren stedgem Urbanity
        ren Leven Life_Satisfaction

*** Data adjustments
**Drop if Tenure is unkown
Tabulate Tenure
Tabulate Tenure, nol
Drop if Tenure==5

**Recode Residential_Satisfaction into four categories
recode Residential_Satisfaction (1=1) (2=1) (3=0) (4=0) (5=0), gen(Residential_Satisfaction_d)
label define Residential_Satisfaction_d_l 1 "Satisfied" 0 "Not satisfied"
label values Residential_Satisfaction_d Residential_Satisfaction_d_l
label variable Residential_Satisfaction_d "Residential_Satisfaction dummy (0 = NS, 1 = S)

** Redefine labels Energy_label to match with other datasets
label define Energy_label 1 "A" 2 "B" 3 "C" 4 "D" 5 "E" 6 "F" 7 "G"
label values Energy_label

** Combine multiple categories regarding Education of respondent and merge into new categorical variable Education (low educated, secondary
educated, well educated)
    recode Education (1/5=1) (6/8=2) (9/10=3)
    drop if Education==11
    label variable Education "Education level"
    label define Education_label 1 "Low educated" 2 "Secondary educated" 3 "Well educated"
    label values Education Education_label

** Recode labelname Number_of_persons_in_HH to fit with datasets 2018
recode Number_of_persons_in_HH (0/1=1) (2=2) (3=3) (4=4) (5/max=5)
label define Number_of_persons_in_HH_label 1 "1 persoon" 2 "2 personen" 3 "3 personen" 4 "4 personen" 5 "5 of meer personen"
label values Number_of_persons_in_HH Number_of_persons_in_HH_label

** Transform variable “since when do you live at this address” into the amount of years you’ve lived at this address
```

```
replace Years_at_current_adress = 2015-Years_at_current_adress
label variable Years_at_current_adress "Years at current adress"

** Recode labelname Tenure to make it comparable with datasets 2012&2018
label define Tenure_label 1 "Koopwoning" 2 "Sociale huur" 3 "Particuliere huur" 5 "Onbekend (OP geen lid van hh)"
label values Tenure Tenure_label

** Replace Housing_expenses per month for Housing expenses per year, for clearer interpretation
replace Housing_expenses = Housing_expenses * 12
label variable Housing_expenses "Housing expenses per year"

** Generate variable cost to income by dividing yearly housing cost by yearly disposable income
generate Cost_to_income = Housing_expenses / Disposable_HH__income
label variable Cost_to_income "Yearly housing expenses / Yearly Disposable HH income"

** Recode and relabel dummy Region G4, other municipalities as base category.
recode Region_G4 (1=1) (2=0)
label define Region_G4_label 0 "overige gemeenten" 1 "vier grote steden"
label values Region_G4 Region_G4_label

** Recode and relabel dummy housing type, 'eengezinswoning' as base category
recode Housing_type (1=0) (2=1)
label define Housing_type_label 0 "eengezinswoning" 1 "meergezinswoning"
label values Housing_type Housing_type_label

label variable Housing_quality "Poor housing quality

tabulate Ethnicity
label define Ethnicity_label 1 "Native" 2 "Non-Western" 3 "Western"
label values Ethnicity Ethnicity_label

tabulate Self_reported_health
label define Self_reported_health_label 1 "Very well" 2 "Good" 3 "Ok" 4 "Differs from time to time" 5 "Bad"
label values Self_reported_health Self_reported_health_label

tabulate Tenure
label define Tenurelabel 1"Owner-occupier" 2 "Social rent" 3"private rent"
label values Tenure Tenurelabel

tabulate Neighbourhood_satisfaction
label define Neighbourhood_satisfaction_label 1 "Very satisfied" 2 "Satisfied" 3 "Neither satisfied, nor dissatisfied" 4 "Dissatisfied" 5 "Very dissatisfied"
label values Neighbourhood_satisfaction Neighbourhood_satisfaction_label

tabulate Region_G4
label define Region_G4Label 0"Other municipalities" 1 "Big four (A'dam, R'dam, Utrecht, the Hague)"
label values Region_G4 Region_G4Label

tabulate Housing_type
label define Housing_type_Label  0"Single-family home" 1 "Multi-family home"
label values Housing_type Housing_type_Label

** Recode and relabel Life_Satisfaction, from a continuous variable into binary variable Satisfied with life
recode Life_Satisfaction (1=0) (2=0) (3=0) (4=0) (5=0) (6=1) (7=1) (8=1) (9=1) (10=1)
label define Life_Satisfaction_label 0 "Dissatisfied with life" 1 "Satisfied with life"
label values Life_Satisfaction Life_Satisfaction_label

** Relabel variables to match datasets
label variable Residential_Satisfaction "Residential Satisfaction"
label variable Energy_label "Energy label
label variable Age "Age
label variable Ethnicity "Ethnicity
label variable Education "Level of education
label variable Disposable_HH__income "Disposable household income
label variable Household_type "Household type
label variable Number_of_persons_in_HH "Number of persons in the hold
```

```
label variable Years_at_current_adress "Amount of years lived at this adress
label variable Self_reported_health "Self reported health status
label variable Tenure "Tenure
label variable Construction_year "Building period
label variable Housing_type "Housing type
label variable Housing_expenses "Housing expenses
label variable WOZ_housing_value "Housing value (WOZ tax value)
label variable Housing_quality "Housing quality
label variable Living_area "Square metres of living area
label variable Neighbourhood_satisfaction "Satisfaction with neighbourhood
label variable Relationship_neighbours "Relationship with neighbours
label variable Safety "Safety
label variable Region_G4 "Region
label variable Urbanity "Urbanity level
label variable Life_Satisfaction "Satisfied with life

*** Data cleaning
**Identify missing values using mdesc
mdesc
drop if Residential_Satisfaction==.
drop if Life_Satisfaction==.
drop if Energy_label==.
mdesc
drop if Housing_type==.
drop if Living_area==.
drop if Housing_expenses==.
drop if Cost_to_income==.
mdesc

**Delete negative and odd values housing expense: negative or over 5.000 a month so 60.000 a year
sum
drop if Housing_expenses<0
drop if Housing_expenses>60000

** drop if disposable household income is lower than social assistance benefit 2012 (12*668,21=8018) or over a quarter of a million euro's
drop if Disposable_HH__income<8018

*** Data adjustments
**Recode Satisfaction into four categories
recode Residential_Satisfaction (1=1) (2=1) (3=0) (4=0) (5=0), gen(Residential_Satisfaction_dum)
label define Residential_Satisfaction_dum_lbl 1 "Satisfied" 0 "Not satisfied"
label values Residential_Satisfaction_dum Residential_Satisfaction_dum_lbl
label variable Residential_Satisfaction_dum "Satisfaction dummy (0 = NS, 1 = S)

** Generate variable person-space ratio by dividing the number of persons in the household by the size of the living area
gen Person_space_ratio = Number_of_persons_in_HH / Living_area
label variable Person_space_ratio "Persons in HH / Living area"

**recoding variables to solve multicollinearity issues
tabulate Living_area
tabulate Living_area, nol
recode Living_area (1=1) (2=1) (3=2) (4=2) (5=3) (6=3) (7=3)
label define Living_area_label 1 "up to 69m2" 2 "70 up to 119m2" 3 "over 120m2"
label values Living_area Living_area_label

tabulate Housing_quality
tabulate Housing_quality, nol
recode Housing_quality (1=0) (2=0) (3=1) (4=1) (5=1)
label define Housing_quality_label 0 "Agreed" 1 "Disagreed"
label values Housing_quality Housing_quality_label

tabulate Household_type
tabulate Household_type, nol
recode Household_type (1=1) (2=2) (3=3) (4=3) (5=4)
label define Household_type 1 "Single person HH" 2 "Couple" 3 "Parent(s) with kids" 4 "Non family household"
label values Household_type
```

```
**Reducing amount of parameters in control variables
tabulate Relationship_neighbours
tabulate Relationship_neighbours, nol
recode Relationship_neighbours (1=1) (2=1) (3=0) (4=0) (5=0)
label define Relationship_neighbours_label 1 "Agreed" 0 "Not agreed"
label values Relationship_neighbours Relationship_neighbours_label

tabulate Age
tabulate Age, nol
recode Age (1=1) (2=1) (3=2) (4=2) (5=2) (6=3) (7=3)
label define Age_label 1 "17-23 year" 2 "35-64 year" 3 "65 years and older"
label values Age Age_label

tabulate Urbanity
tabulate Urbanity, nol
recode Urbanity (1=1) (2=1) (3=2) (4=3) (5=3)
label define Urbanity_namecat 1 "Strong" 2 "Moderately" 3 "Little to none"
label values Urbanity Urbanity_namecat

recode Disposable_HH__income (0/35000=1) (35001/75000=2) (75001/max=3)
label define Disposable_HH_income_label 1 “Low” 2 “Middle” 3 “High”
label values Disposable_HH__income Disposable_HH__income_label

**Order variables
order Life_Satisfaction Residential_Satisfaction Residential_Satisfaction_dum Energy_label Year Age Ethnicity Education Disposable_HH__income Household_type Number_of_persons_in_HH Years_at_current_adress Self_reported_health Tenure Construction_year Housing_type Cost_to_income Housing_expenses WOZ_housing_value Housing_quality Person_space_ratio Living_area Neighbourhood_satisfaction Relationship_neighbours Region_G4 Urbanity

*** Test assumptions ***
**Generate descriptive statistics table
asdoc sum Life_Satisfaction i.Energy_label i.Age i.Ethnicity i.Education i.Disposable_HH__income i.Household_type i.Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure i.Residential_Satisfaction_dum Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity

**Correlation matrix
corr Life_Satisfaction Energy_label Age Ethnicity Education Disposable_HH__income Household_type Number_of_persons_in_HH Years_at_current_adress Self_reported_health Tenure Residential_Satisfaction_dum Construction_year Housing_type Cost_to_income Housing_expenses WOZ_housing_value Housing_quality Person_space_ratio Living_area Neighbourhood_satisfaction Relationship_neighbours Region_G4 Urbanity

reg Life_Satisfaction ib7.Energy_label i.Year i.Age i.Ethnicity i.Education i.Disposable_HH__income i.Household_type i.Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure i.Residential_Satisfaction_dum Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity

estat vif

*** Analysis ***
**Generate first analyse table, comparing the model fit
logistic Life_Satisfaction ib7.Energy_label
estimate store Model_I

logistic Life_Satisfaction ib7.Energy_label ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type i.Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health
estimate store Model_II

logistic Life_Satisfaction ib7.Energy_label ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type i.Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure i.Residential_Satisfaction_dum Construction_year i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area
estimate store Model_III

logistic Life_Satisfaction ib7.Energy_label ib2.Age i.Ethnicity i.Education ib2.Disposable_HH__income i.Household_type i.Number_of_persons_in_HH Years_at_current_adress i.Self_reported_health i.Tenure i.Residential_Satisfaction_dum Construction_year
```

```
i.Housing_type Cost_to_income Housing_expenses WOZ_housing_value i.Housing_quality Person_space_ratio i.Living_area i.Neighbourhood_satisfaction i.Relationship_neighbours i.Region_G4 i.Urbanity
estimate store Model_IV

esttab Model_I Model_II Model_III Model_IV using LifeTable.rtf, con b se pr2 eform label nonumbers mtitles("Model I" "Model II" "Model III" "Model IV")

**LRtest models
lrtest Model_I Model_II
lrtest Model_II Model_III
lrtest Model_III Model_IV
```

*** VIF ***

Variable	VIF	1/VIF
Energy_label		
1	2.39	0.418120
2	2.72	0.368067
3	3.46	0.288927
4	1.67	0.597090
5	2.36	0.424317
6	1.83	0.546443
Age		
2	2.53	0.395035
3	3.30	0.303224
Ethnicity		
2	1.17	0.857312
3	1.02	0.978303
Education		
2	1.54	0.650016
3	1.79	0.559729
Disposable~e		
2	2.26	0.442874
3	1.84	0.543390
Household_~e		
2	2.56	0.390448
3	7.17	0.139548
4	1.46	0.686838
Number_of_~H	7.79	0.128381
Years_at_c~s	1.74	0.575609
Self_repor~h		
2	1.86	0.538998
3	1.70	0.586917
4	1.43	0.696933
5	1.24	0.803958

Variable	VIF	1/VIF
Tenure		
2	1.79	0.557131
3	1.34	0.748719
1.Resident~m	1.34	0.746223
Constructi~r	1.73	0.577041
1.Housing_~e	2.12	0.471583
Cost_to_in~e	2.61	0.382643
Housing_ex~s	2.50	0.400445
WOZ_housin~e	1.82	0.548422
1.Housing_~y	1.20	0.830396
Person_spa~o	4.42	0.226050
Living_area		
2	5.13	0.194882
3	8.38	0.119278
Neighbourh~n		
2	1.39	0.721240
3	1.37	0.729256
4	1.21	0.828010
5	1.09	0.916331
1.Relation~s	1.08	0.925544
1.Region_G4	1.37	0.730949
Urbanity		
2	1.27	0.788908
3	1.41	0.710106
	Mean VIF	**2.36**

*** Regression results ***

	Model I	Model II	Model III	Model IV
Satisfied with life				
A	1.154	1.340*	0.935	0.961
	(0.144)	(0.183)	(0.149)	(0.154)
B	1.186	1.214	0.978	0.959
	(0.130)	(0.145)	(0.133)	(0.132)
C	0.933	1.138	0.985	0.981
	(0.0858)	(0.113)	(0.117)	(0.118)
D	1.657**	1.329	1.141	1.084
	(0.267)	(0.226)	(0.210)	(0.201)
E	0.549***	1.110	1.205	1.215
	(0.0524)	(0.116)	(0.141)	(0.145)
F	1.186	1.155	1.038	1.032
	(0.143)	(0.149)	(0.146)	(0.146)
G	Base	Base	Base	Base
17-23 year		1.148	1.266*	1.301**
		(0.108)	(0.122)	(0.126)
35-64 year		Base	Base	Base
65 years and older		2.162***	1.904***	1.789***
		(0.171)	(0.155)	(0.147)
Native		Base	Base	Base
Non-Western		0.417***	0.533***	0.539***
		(0.0341)	(0.0458)	(0.0483)
Western		0.681***	0.726***	0.722***
		(0.0621)	(0.0675)	(0.0674)
Low educated		Base	Base	Base
Secondary educated		1.092	1.026	1.053
		(0.0768)	(0.0737)	(0.0762)
Well educated		1.226*	1.073	1.126
		(0.100)	(0.0915)	(0.0967)
Low		0.562***	0.819	0.817
		(0.0502)	(0.0843)	(0.0843)
Middle		Base	Base	Base
High		2.151*	1.512	1.525
		(0.648)	(0.461)	(0.466)
Single person HH		Base	Base	Base
Couple		1.890***	1.646***	1.666***
		(0.190)	(0.173)	(0.176)
Parent(s) with kid(s)		1.139	1.093	1.133
		(0.169)	(0.166)	(0.172)
Non family household		0.667*	0.654*	0.710
		(0.127)	(0.126)	(0.138)
Number of persons in the hold		1.156*	1.135	1.120
		(0.0673)	(0.0811)	(0.0804)
Amount of years lived at this adress		1.014***	1.010**	1.010***
		(0.00284)	(0.00300)	(0.00303)
Very well		Base	Base	Base
Good		0.522***	0.546***	0.549***
		(0.0763)	(0.0800)	(0.0807)
Ok		0.177***	0.205***	0.212***
		(0.0269)	(0.0314)	(0.0325)

Differs from time to time	0.0851***	0.106***	0.110***
	(0.0129)	(0.0162)	(0.0169)
Bad	0.0224***	0.0287***	0.0298***
	(0.00335)	(0.00433)	(0.00452)
Owner-occupier		Base	Base
Social rent		0.897	0.907
		(0.0752)	(0.0761)
private rent		1.016	1.011
		(0.124)	(0.124)
Not satisfied		Base	Base
Satisfied		2.549***	2.029***
		(0.179)	(0.152)
Building period		1.001	1.001
		(0.000672)	(0.000665)
Single-family home		Base	Base
Multi-family home		0.931	1.000
		(0.0761)	(0.0856)
Yearly housing expenses / Yearly Disposable HH income		0.367***	0.361***
		(0.0858)	(0.0833)
Housing expenses		1.000**	1.000**
		(0.0000133)	(0.0000132)
Housing value (WOZ tax value)		1.000***	1.000**
		(0.00000054)	(0.00000053)
Agreed		Base	Base
Disagreed		1.202*	1.173*
		(0.0973)	(0.0955)
Persons in HH / Living area		0.979	0.945
		(0.151)	(0.148)
up to 69m2		Base	Base
70 up to 119m2		0.903	0.900
		(0.0936)	(0.0940)
over 120m2		0.811	0.801
		(0.127)	(0.126)
Very satisfied			Base
Satisfied			0.993
			(0.0835)
Neither satisfied, nor dissatisfied			0.710***
			(0.0730)
Dissatisfied			0.528***
			(0.0620)
Very dissatisfied			0.474***
			(0.0760)
Not agreed			Base
Agreed			1.477***
			(0.0943)

Other municipalities				Base
Big four (A'dam, R'dam, Utrecht, the Hague)				0.978 (0.0876)
Strong				Base
Moderately				1.135 (0.102)
Little to none				1.120 (0.0939)
Constant	37.58***	86.09***	3.400	3.679
	(2.912)	(17.87)	(4.453)	(4.782)
Observations	50165	50165	50165	50165
Log likelihood	-6241.1827	-4871.2837	-4728.9737	-4671.1131
Df	6	23	35	43
Pseudo R^2	0.010	0.227	0.250	0.259

Note: Variable coefficients denoted are the odds ratio's. Depending variable: Life Satisfaction dummy, independent variable Energy label with label G as base category. Standard errors in parentheses and significance depicted with *** p<0.01, ** p<0.05, * p<0.1 respectively. Inclusion of control variables for the years, socio-demographics, housing characteristics and neighbourhood characteristics denoted with yes/no.

lrtest Model_I Model_II

Likelihood-ratio test LR chi2(17) = 2739.80

(Assumption: Model_I nested in Model_II) Prob > chi2 = 0.0000

lrtest Model_II Model_III

Likelihood-ratio test LR chi2(12) = 284.62

(Assumption: Model_II nested in Model_III) Prob > chi2 = 0.0000

lrtest Model_III Model_IV

Likelihood-ratio test LR chi2(8) = 115.72

(Assumption: Model_III nested in Model_IV) Prob > chi2 = 0.0000