Energy Efficiency and Energy Poverty in Dutch Households

How Housing-related Factors Determine the Residential Energy Consumption and Energy Affordability in the Netherlands

Mark Jan Uijl
MSc Complex System Engineering and Management

TU Delft

Energy Efficiency and Energy Poverty in Dutch Households

How Housing-related Factors Determine the Residential Energy Consumption and Energy Affordability in the Netherlands

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex System Engineering and Management**

Faculty of Technology, Policy and Management

by

Mark Jan Uijl

Student number: 4550803

To be defended in public on August 16th 2022

Graduation committee

Chairperson	: Dr.ir. M. Kroesen, Section Transport and Logistics
First Supervisor	: Dr. N. Goyal, Section Organisation & Governance
Second Supervisor	: Dr. E. Schröder, Section Economics of Technology and Innovation

Preface

Dear Reader,

Before you lies the result of my graduation project, the final project I worked on in fulfilment of the master Complex System Engineering and Management at TU Delft. I have enjoyed writing this thesis in the past six months, as I found it a great way to apply what I have learned over the years and see it as an incredible project to finish off my time as a student here in Delft. Thank you for taking the time to read my thesis.

For carrying out this project, I especially would like to thank Nihit Goyal, who always provided me with constructive feedback, and gave me new ideas when I wasn't sure about how to proceed. Our weekly meetings were often inspiring and have improved the quality of my thesis tremendously. I also want to thank Maarten Kroesen and Enno Schröder for their sharp comments and their interest in my thesis; It was a pleasure to have you in my graduation committee.

Furthermore, I would like to thank my supervisors at BearingPoint for their support during this project. Michiel Musterd, for his in-depth knowledge on analytics, data visualisation and academic writing, which was very helpful, both in the data analyses and the writing of this thesis. And also, Koen Graat, for his support in the data preparation. Without your help the data quality wouldn't have been where it is now; I really appreciate all the hours that you put into this, sometimes even in your free evenings.

All in all, I am happy to have finished my thesis, and I hope that it brings us a small step closer to reducing the energy efficiency and energy poverty problems in the Netherlands. I am passionate about these topics and this thesis reveals that there's plenty of room for policy improvements in this area. I am interested to see where it goes, especially given the energy transition and the rising energy prices, and hope that this thesis can be a contribution here. Enjoy reading my thesis!

Mark Jan Uijl

Abstract

The residential energy consumption and energy affordability are crucial in policy design, with regard to energy efficiency and energy poverty. This has become even more relevant now, as the energy transition is taking off and the energy prices are soaring. Given the inequality in the energy consumption and energy affordability, it is therefore critical to know which factors are impacting these phenomena, and how they are impacting these phenomena. The housing-related factors are the least understood factors impacting the energy consumption and energy affordability, even though these factors are key pillars of energy policy. Thus, this research assesses how housing-related factors impact the residential energy consumption and energy affordability in the Netherlands. For this assessment, multiple data sources are combined over the years 2016 to 2018, covering a total of more than 8 million Dutch households. A regression analysis is performed on a postal code 6 level using the housing-related factors as predictors, while controlling for socio-demographic and weather factors. This revealed that the housing-related factors impacting energy consumption and energy affordability are relatively similar. This study finds that the WOZ-value is a strong predictor for both phenomena as houses with a higher WOZ-value have a lower energy efficiency. Furthermore, it shows that rental households consume less energy but spend a similar percentage of their income on energy, and that the impact of the year of construction on both phenomena has increased significantly, when compared to earlier studies. Based on this, it is concluded that rental houses and houses with an older year of construction deserve more attention in Dutch energy policy. Furthermore, low-income households should be targeted in short-term energy policy, as this study shows that the current soaring energy prices have a severe impact on the energy affordability in the Netherlands. Insights are however broader, and can be pivotal in energy demand and energy affordability projections, both in the Netherlands and the rest of Western Europe.

Keywords: Residential energy consumption; Residential energy affordability; Housing-related factors; households; The Netherlands; Energy efficiency; Energy poverty

Executive Summary

Societal Relevance

The residential energy efficiency and energy poverty are critical in Dutch policy design as the Netherlands aims to reduce greenhouse gas emissions and energy inequality. Energy efficiency measures target the energy consumption, while energy poverty policy increases the energy affordability of households. Nevertheless, the inequality in the energy consumption and energy affordability is growing in the Netherlands, and the policies trying to reduce these problems are underperforming. This raises the question which factors determine the energy consumption and affordability. Furthermore, the energy transition increases the need for a lower energy consumption, and the current soaring energy prices reduce the energy affordability significantly, resulting in more energy poverty among Dutch households. The inequality in the Dutch residential energy consumption and energy affordability is for a large part dependent on housing-related factors, although the impact of these factors on the energy consumption and affordability is poorly understood. Thus, more insight in the relation between the housing-related factors and the energy consumption and affordability is needed to improve energy efficiency and energy poverty policy in the Netherlands.

Research Objective

Therefore, the objective of this research is to provide insight in the Dutch residential energy consumption and energy affordability, and their dependence on housing-related factors. The energy consumption includes the gas and electricity consumption, and the energy affordability is measured using the energy expenditure to income ratio. Housing-related factors include the dwelling characteristics, the urbanity and the ownership structure of a house. The research objective is summarized in the following research question: *How do housing-related factors impact the Dutch residential energy consumption, and the affordability of this energy consumption?*

Methodology

In order to answer this question, a literature review was performed first on the different housing-related factors impacting the energy consumption and energy affordability, and on the control variables needed for the analyses, which included socio-demographic and weather factors. Afterwards, the descriptive statistics of these factors, and of the energy consumption and energy affordability were analysed. Finally, multiple regression analyses were performed to assess the impact of the housing-related factors on the energy consumption and energy affordability, when controlling for other variables. For these analyses, multiple datasets from the CBS, the BAG, the RVO and the KNMI were combined over the years 2016, 2017 and 2018, covering yearly data on more than 8 million Dutch households. As most data was only available per postal code 6, this is the unit of analysis used in this study.

Distribution Energy Consumption and Energy Affordability

The descriptive statistics revealed that the average residential energy consumption per postal code over 2016-2018 is 17302 kWh per year (St.dev. 6275), while the average energy expenditure to income ratio is 6,16% (St.dev. 2,04). The distributions of the energy consumption and energy affordability are relatively similar, which was expected as the energy affordability is retrieved by dividing the income by the energy consumption. Geographically, the energy consumption and energy expenditure to income ratio are lower in the Randstad and higher in the eastern and north-eastern part of the Netherlands. An additional affordability analysis showed that based on April 2022 prices, the average energy expenditure to income ratio has increased to 11,32%, leaving more than half of the Dutch households under the energy poverty line of 10%, when controlling for a 30% decrease in energy consumption.

Housing-Related Factors Impacting the Energy Consumption and Energy Affordability

Next, the regression analyses revealed how housing-related factors are impacting the energy consumption and affordability. Both for the energy consumption and the energy expenditure to income ratio, the surface area, WOZ-value, house type and construction year showed the highest regression coefficients and thus the strongest relation. A larger surface area and higher WOZ-value resulted in higher values for the energy consumption and energy expenditure to income ratio, while a higher construction year resulted in lower values. When considering the house type, apartments had a lower energy consumption and energy expenditure to income ratio, while values were higher for detached houses. Differences between the energy consumption and energy affordability were mainly in the strength of the relation. The most relevant differences were that terraced houses are more similar in energy consumption to apartments, while they are in between detached houses and apartments in energy affordability, and that rental households consume less energy, although they don’t show a relevant difference in energy affordability.

New Insights

The study adds to the literature by revealing that the WOZ-value has a strong positive relationship with the energy consumption and energy expenditure to income ratio. This is due to the fact that it captures additional housing information that is not included in other variables, and because the energy efficiency of houses with a higher WOZ-value is worse. Furthermore, the impact of the year of construction has increased when compared to previous studies, showing that the energy efficiency of new houses is improving faster than before. Next to that, the urbanity has a limited impact on the energy consumption in the Netherlands, even though it is an important factor in other countries. Rental houses are consuming less energy, while they don’t spend significantly less on energy, which suggests that households living in rental houses have affordability issues. Moreover, the affordability analysis based on the energy prices in April 2022 revealed that the soaring energy prices have a large impact on the energy affordability, and that further research is needed here.

Policy Implications

All in all, this research has several policy implications. First the study highlights that policies for improving the energy efficiency in rental houses should be redesigned in order to create incentive for landlords to invest in energy efficiency. Second, additional energy efficiency policies should be implemented based on construction year, as older houses consume significantly more energy, which affects both the energy consumption and the energy affordability of households. Third, more policies should be designed to prevent low-income households from falling into energy poverty, as financial barriers are currently high for this group. The need for this has increased even more now that the energy prices are soaring. Additionally, the government should target the current energy prices, as the current prices will push more than 50% of the Dutch households into energy poverty, when no measures are taken. Finally, the insight in the factors impacting energy consumption and energy affordability can help improve energy demand and energy affordability projections, which are useful for the Dutch government, the network operators and the energy providers.

Table of Contents

Preface....iv

Abstract....vi

Executive Summary....vii

1. Introduction....1

2. Literature Review....8

  2.1 Factors influencing Energy Consumption....8

  2.2 Factors influencing Energy Affordability....14

  2.3 Summary of Core Concepts and Impacting Factors....19

3. Methodology & Data....22

  3.1 Research Approach & Methodology....22

  3.2 Overview Datasets and Data Preparation....26

  3.3 Included Variables and Variable Transformations....30

  3.4 Data Description....34

4. Descriptive Statistics: Energy Consumption, Energy Affordability and Housing-related Factors in the Netherlands....37

  4.1 Energy Consumption....37

  4.2 Energy Affordability....39

  4.3 Correlations Socio-Demographic and Weather Factors....44

  4.4 Correlations Housing-related Factors....45

5. Inferential Statistics: How Housing-Related Factors impact Energy Consumption and Energy Affordability....51

  5.1 How Housing-Related Factors impact the Energy Consumption....51

  5.2 How Housing-Related Factors impact the Energy Affordability....53

6. Discussion & Conclusion....57

  6.1 Summary of Findings and Comparison with Literature....57

  6.2 Conclusion....62

  6.3 Implications....64

  6.4 Limitations & Future Research....67

Bibliography....70

Appendix A: Data Preparation....77

Appendix B: Excluded House Types for Regression....81

1. Introduction

Relevance of Residential Energy Consumption

Residential energy consumption is responsible for a large part of the total energy consumption of a country, and thus for a significant amount of its greenhouse gas emissions. In Western Europe, households are responsible for 15-20% of the carbon dioxide emissions, which are contributing to global warming (Abrahamse & Steg, 2009). As countries aim to limit global warming, they are implementing different policies to mitigate these greenhouse gas emissions (Gillingham & Stock, 2018). Measures are aimed at energy efficiency optimization and adding renewables to the energy mix (Khezri, Mahmoudi, & Haque, 2020; Krarti & Aldubyan, 2021)). With regard to residential energy consumption and energy efficiency, this raises the question which factors determine the residential energy use, as knowing the determinants of the energy use can contribute to designing effective policies for reducing the energy use (Guo et al., 2018). This knowledge is needed, as current energy efficiency policies in the Netherlands are underperforming (Vega, Van Leeuwen, & Van Twillert, 2022). Moreover, previous studies have looked into some of these determinants, such as household behaviour, energy conservation and demand response, and revealed the need for policy change (Cheng & Steemers, 2011; Li, Yao, Yang, & Zhou, 2018; Guo et al., 2018).

Relevance of Residential Energy Affordability

A concept that is closely related to the residential energy consumption is the residential energy affordability, which is decreasing dramatically due to the rising energy prices and the COVID-19 pandemic. In Europe, the energy prices have soared since the summer of 2021. Especially the gas prices experienced a large price increase, growing with more than 200% between July and October 2021 (Tesio, Conti, & Cervigni, 2022). Currently, the war in Ukraine is expected to increase the energy prices even more (CPB, 2022). Next to that, the COVID-19 pandemic causes people to stay home more often, which leads to more residential energy consumption. Households use up to 20% more energy in COVID-19 times, depending on the government measures (Krarti & Aldubyan, 2021). Both the higher energy prices and the higher energy demand lead to higher energy costs, which decreases the affordability of energy, worsens the inequality in energy affordability and increases energy poverty among Dutch households (Lin & Wang, 2020). Households are considered energy poor when they can't afford or don't have access to the energy they need. Energy poverty is considered a problem in the EU, and an increase in the inequality in energy affordability is also undesirable from a policy perspective (Galvin & Sunikka-Blank, 2018). In the Netherlands however, the policies for reducing energy poverty are limited, and no adequate safety net is available for households experiencing energy poverty (Mashhoodi et al., 2018; Vega et al., 2022). Therefore, it is critical to study the factors that are responsible for the energy consumption and the energy affordability of households, especially given the current energy market.

Inequality in Energy Consumption and Affordability

The determinants of energy consumption and energy affordability are often used to explain the differences or inequalities in residential energy consumption and affordability. Cheng and Steemers (2011), and Guo et al. (2018) reveal how the energy consumption differs for households living in different types of dwellings, and how a higher income and socio-economic class leads to more energy consumption. Next to this, the energy efficiency differs per house depending on its characteristics, and not all households have access to the same energy sources, even in the Netherlands (Brounen et al., 2012; Lin & Wang, 2020). ŠTreimikienė (2014) and Dong et al. (2018) show how this leads to an unequal division of energy consumption and energy affordability, when comparing different regions and different groups of households. The unequal division in energy consumption is often referred to as energy inequality, which is characterized as the unfair distribution of energy use or access among different groups in a population (Nguyen, Hoang, Wilson, & Managi, 2019). The unequal division in energy affordability can be quantified with affordability ratios, such as the expenditure-to-income ratio (Haffner & Boumeester, 2015). Furthermore, the inequality in energy consumption and energy affordability in a certain population can be measured using the Atkinson index or the Gini coefficient and is often visualized with a Lorenz curve (Sun, Zhang, Peng, & Zhang, 2015; Nguyen et al., 2019). The economic inequality in Western Europe has increased over the past decades, which suggests that the energy affordability is decreasing, and energy inequality is rising (Christophers, 2019; Galvin & Sunikka-Blank, 2018). This is also interesting from a policy perspective, as the Netherlands aims to reduce energy inequality with policies targeting energy poverty and energy efficiency (Van Middelkoop, Van Polen, Holtkamp, & Bonnerman, 2018; Vega et al., 2022). This means that inclusive energy policies are promoted, which give households similar opportunities in terms of becoming more sustainable and energy efficient (Van Middelkoop et al., 2018). Therefore, insight in the inequalities in the residential energy consumption and energy affordability is crucial for future policy design. This also holds for other countries in western Europe, as they are relatively similar to the Netherlands in terms of demographics and energy use (Neagu & Teodoru, 2019).

Relevance of Housing-related Factors

The residential energy consumption and energy affordability are however dependent on many factors, which can be divided in different categories. The factors include demographic and behaviour factors (Huebner et al., 2016; Sanquist, Orr, Shui, & Bittner, 2012), but also housing-related factors, including dwelling characteristics, the location of the house and the owning structure of the household, which have a large impact on the energy consumption (Xie, Yan, Zhang, & Wei, 2020; Wei, Zhu, & Glomsrød, 2014; Yohanis, Mondol, Wright, & Norton, 2008). Moreover, the housing-related factors are important determinants for the inequality between different societal groups in terms of economics and welfare (Lin & Wang, 2020; Krarti & Aldubyan, 2021). The energy consumption is a core part of this inequality, on the one hand because certain dwellings do not have access to certain sources of energy (Miah, Foysal, Koike, & Kobayashi, 2011) and on the other hand because house characteristics such as a lack of insulation affect the energy affordability and can even lead to energy poverty (ŠTreimikienė, 2014). In high-income countries, such as the Netherlands, energy access is relatively good, and thus problems with regard to energy inequality go hand in hand with problems in energy affordability (Galvin & Sunikka-Blank, 2018). Therefore, the energy inequality in high-income countries is closely linked with the housing-related factors, as they influence both the energy consumption and the energy affordability. Thus, the housing-related factors can be used to reveal the differences in energy consumption and energy affordability between different societal groups.

Within the housing-related factors, the differences between rural and urban households and rented and owned homes is particularly interesting. Research on these factors is limited, but the literature suggests that they do have a large impact on the energy consumption and the energy affordability, (Iraganaboina & Eluru, 2021; Haffner & Boumeester, 2015). Rural households in the Netherlands are using more gas, which makes them less sustainable and more vulnerable to energy poverty (Mashhoodi, 2021). Households living in rented homes on the other hand have an economic disadvantage when compared to homeowners (Christophers, 2019), which means that they do experience energy poverty more often and have less opportunities in the energy transition (Chapman & Okushima, 2019). Haffner and Boumeester (2015) confirm this by showing that the affordability of housing and energy is a core aspect of inequality in the Netherlands, and that the home ownership structure is closely related to energy poverty, as rental homes are often less affordable for occupants. Next to that, the ownership structure of a house is often used as a parameter in policy design, both for energy efficiency and for energy poverty, increasing the relevance of this factor even more (Kontokosta et al., 2019; Vega et al., 2022; Riva et al., 2021).

Furthermore, the WOZ-value or property value and the year of construction are interesting factors to research. The WOZ-value wasn't included in previous studies found on energy consumption and energy affordability. Nevertheless, other literature reveals that it influences the energy consumption as houses with a higher WOZ-value have a worse energy label (Boesveld, 2021) and lag behind in terms of the implementation of energy efficiency policies (Van Middelkoop, Vringer, & Visser, 2017). The year of construction on the other hand is included in multiple energy consumption and energy affordability studies, although the size of its impact differs in the literature. Some studies find only a minimal relation with the energy consumption and energy affordability, while others find a strong negative relation, meaning that households living in older houses consume more and spend more on energy (Santin et al., 2009; Besagni & Borgarello, 2018; Riva et al., 2021; Das et al. ,2022). The energy efficiency of new houses is however increasing, as both Dutch and European policy requires new homes to have a minimal level of energy efficiency and strives towards energy-neutral houses (Tambach & Visscher, 2012; Visscher, Meijer, Majcen, & Itard, 2016). This suggests that the relation between the year of construction and the energy consumption and affordability is becoming stronger, that the disadvantage that households in older houses have is growing, and that policy intervention is needed.

Geographical Differences in Energy Consumption and Affordability

Most research on the determinants of energy consumption and affordability analyses households in the US, Europe or China. Where literature on European and US households focusses more on behaviour (Abrahamse & Steg, 2009; Huebner et al., 2016; Sanquist et al., 2012), literature on Chinese households focusses more often on demographics and environmental characteristics, such as dwelling type or the difference between rural and urban residents (Chen et al., 2013; Xie et al., 2020; Miah et al., 2011; Wei et al., 2014). Nevertheless, the year of construction is covered in most articles in the US, Europe and China, where it has a negative relation with the energy consumption and energy affordability (Brounen et al., 2012; Iraganaboina & Eluru, 2021; Xie et al., 2020). As mentioned before, the WOZ-value wasn't included in previous studies as a predictor of energy consumption or energy affordability.

As the difference between rural and urban energy consumption hasn't been researched in Europe, it is interesting to explore this difference in the Netherlands, because the demographics and living conditions there are different than in China or the US. Energy access is for example generally good in the Netherlands, while it is a core differentiator in the countries mentioned above. Next to this, the ownership status is also an interesting factor to include, as current research on this is limited. The only 2 studies on energy consumption found in Europe that included this variable were performed more than a decade ago, showed contradicting results, and lacked an explanation of the observed difference in energy consumption (Santin et al., 2009; Yohanis et al., 2008). For the energy affordability, the urbanity and the ownership status have been studied in Europe. Differences are however large between different countries, both in terms of the average affordability and in terms of the impact of urbanity and ownership (Karasek & Pojar, 2018; Mashhoodi et al., 2018; Papada & Kaliampakos, 2017).

Research Objective

Thus, the objective of this research is to identify how housing-related factors impact the Dutch residential energy consumption and energy affordability. Knowing the inequality in the Dutch residential energy consumption and energy affordability, and what the determinants of this inequality are, can help regulating stakeholders, including national and local governments, and energy providers, to steer the energy consumption and affordability towards a more sustainable and equal future. They could adapt or specify their policies based on the determining factors to improve the equality of the residential energy consumption, both in the Netherlands and in the rest of Western Europe.

Relevance to the Program

This research fits the objectives of a CoSEM thesis, as it considers a technical system, namely the energy demand, which is affected by social factors, including housing and demographics. The interaction between these aspects makes the analysis of this socio-technical system relevant to the program. Next to that, the study also considers the relation that different actor groups have with the energy consumption, and the impact that the energy consumption is making on society in terms of energy affordability and energy inequality, which fits the CoSEM Master well.

Knowledge Gap and Main Research Question

The study adds to the literature by comparing the energy consumption of rural and urban households, on which no previous research was found in Europe. Furthermore, the research also includes data related to the owning structure of the housing. Housing is a crucial factor in wealth inequality in European countries (Christophers, 2019), and the difference in energy consumption and energy affordability between rental houses and owner-occupied houses is therefore an interesting factor to consider. Next to that, this factor is underrepresented in previous research, which means that the impact of this factor is unclear, and that additional research is needed. A third addition to the literature lays in including the WOZ-value, as previous research on the impact of this factor is limited, and because houses with a high WOZ-value are expected to have a higher energy consumption and a lower energy affordability. The year of construction is also an intriguing factor in this study, as new built houses are becoming more energy efficient than ever before. Moreover, the analysis of the residential energy affordability is currently very interesting, given the rising energy prices, which decrease the affordability of energy for households. Thus, the energy affordability increases the impact of the inequalities discussed above, and therefore both the policy and academic relevance of this study. A final addition of this thesis is that it includes both the energy consumption and affordability, whereas previous studies tend to focus on only one of these concepts. This allows for the comparison of these phenomena and the factors impacting them.

All in all, this results in the following research question: *How do housing-related factors impact the Dutch residential energy consumption, and the affordability of this energy consumption?* In this question, housing-related factors include ownership, dwelling characteristics and location, which is used for determining the urbanity. The average residential energy consumption and energy affordability are analysed per postal code 6, based on socio-demographic, weather and housing data on the majority of the Dutch households. The energy consumption is analysed in kilowatt-hour (kWh) per year, while the yearly energy expenditure-to-income ratio is used as a measure for analysing the energy affordability.

Sub-Questions

The main research question can be divided in several sub-questions. The first sub-question is: *How do different factors influence the Dutch residential energy consumption and energy affordability?* This includes two literature reviews on the factors impacting the residential energy consumption and energy affordability. The qualitative data from the literature reviews is used to determine which factors have to be included in the statistical analyses, and to determine the expected impact of these factors on the energy consumption and energy affordability. Next to this, the core concepts of this thesis and their interconnection are summarized in a conceptual model.

The second sub-question is: *How is the energy consumption distributed among households in the Netherlands?* This question is answered by analyzing the data on the Dutch residential energy consumption and visualizing the differences between different groups in the population, based on the socio-demographical characteristics of households and weather data. The impact that the socio-demographic and weather factors have on the energy consumption is analysed descriptively using Tableau and SPSS. The energy inequality is also calculated in this question, resulting in a Gini coefficient and Lorenz curve.

The third sub-question is: *How is the energy affordability distributed among households in the Netherlands?* A similar approach as in sub-question 2 is used here to analyze and visualize the energy affordability in the form of the energy expenditure to income ratio, based on socio-demographic and weather factors. Next to the inequality, the energy poverty in Dutch households is also discussed when answering this question. The differences in affordability are visualised in Tableau and SPSS, and the inequality in affordability is calculated using the Gini coefficient and the Lorenz curve.

The fourth sub-question is: *How do the current energy prices influence the residential energy affordability in the Netherlands?* For answering this question, a speculative longitudinal analysis is used to see how the current energy prices impact the energy affordability. This analysis compares the energy expenditure to income ratio from 2018 with the estimated expenditure to income ratios from October 2021 and April 2022 using Tableau, based on the energy prices in these months and the price elasticity of residential energy consumption. As mentioned before, the affordability is a critical factor in the energy inequality and energy poverty in the Netherlands, making this analysis a key part of this research.

The fifth sub-question is: *How do housing-related factors correlate with the Dutch residential energy consumption and energy affordability?* Analyzing and visualizing the data is also core in this question, only now the analyses is focusing on the differences in energy consumption and energy affordability, based on the housing-related factors. This analysis covers the differences in ownership, dwelling characteristics and urbanity, using Tableau and SPSS. Thus, this sub-question reveals how housing-related factors are associated with the energy consumption and energy affordability.

The sixth sub-question is: *How do housing-related factors explain the residential energy consumption in the Netherlands?* This question combines the energy consumption and the housing-related factors in a statistical analysis, where the housing-related factors are used to analyze and predict the Dutch residential energy consumption. The analysis controls for the weather and socio-demographic factors and provides insight in the extent to which the housing-related factors can explain the energy consumption. This question is answered with the help of a multivariate regression model, built in SPSS. The factors that are included in the analysis, and the order in which they are included, are based on the literature review from sub-question 1 and the insights from the visualisations of sub-question 2 and 5. The deliverable from this sub-question is a regression model on energy consumption that reveals how the energy consumption differs in the Netherlands, based on the housing-related factors.

The seventh sub-question is: *How do housing-related factors explain the residential energy affordability in the Netherlands?* In this question, the housing-related factors are used to predict the energy affordability, while controlling for weather and socio-demographic factors. Again, a multivariate-regression model in SPSS is used, in this case however with the affordability as the dependent variable, which is included as the expenditure-to-income ratio. The data from sub-questions 1, 3 and 5 is combined to determine which factors to include in the regression-model and the order in which these factors are included. All in all, this question gives insight in the explaining power that the housing-related factors have with regard to energy affordability.

Research Flow Diagram

The main research question is answered by combining the deliverables from the different sub-questions, as is shown in the research flow diagram in figure 1. Sub-questions 2, 3 and 5 visualize how respectively the energy consumption, the energy affordability and the housing-related factors differ among households, based on the literature review performed in question 1. This information is used in sub-questions 6 and 7 to design the regression models, which reveal how the energy consumption and the energy affordability are dependent on socio-demographic, weather and housing-related factors. Next to that, sub-question 4 makes use of the energy affordability data gathered in sub-question 3 in order to determine the current energy affordability based on recent energy prices. All in all, the statistical models, together with the Gini coefficient and Lorenz curve, explain the inequality in the Dutch residential energy consumption and energy affordability, and the impact that housing-related factors have on these phenomena.

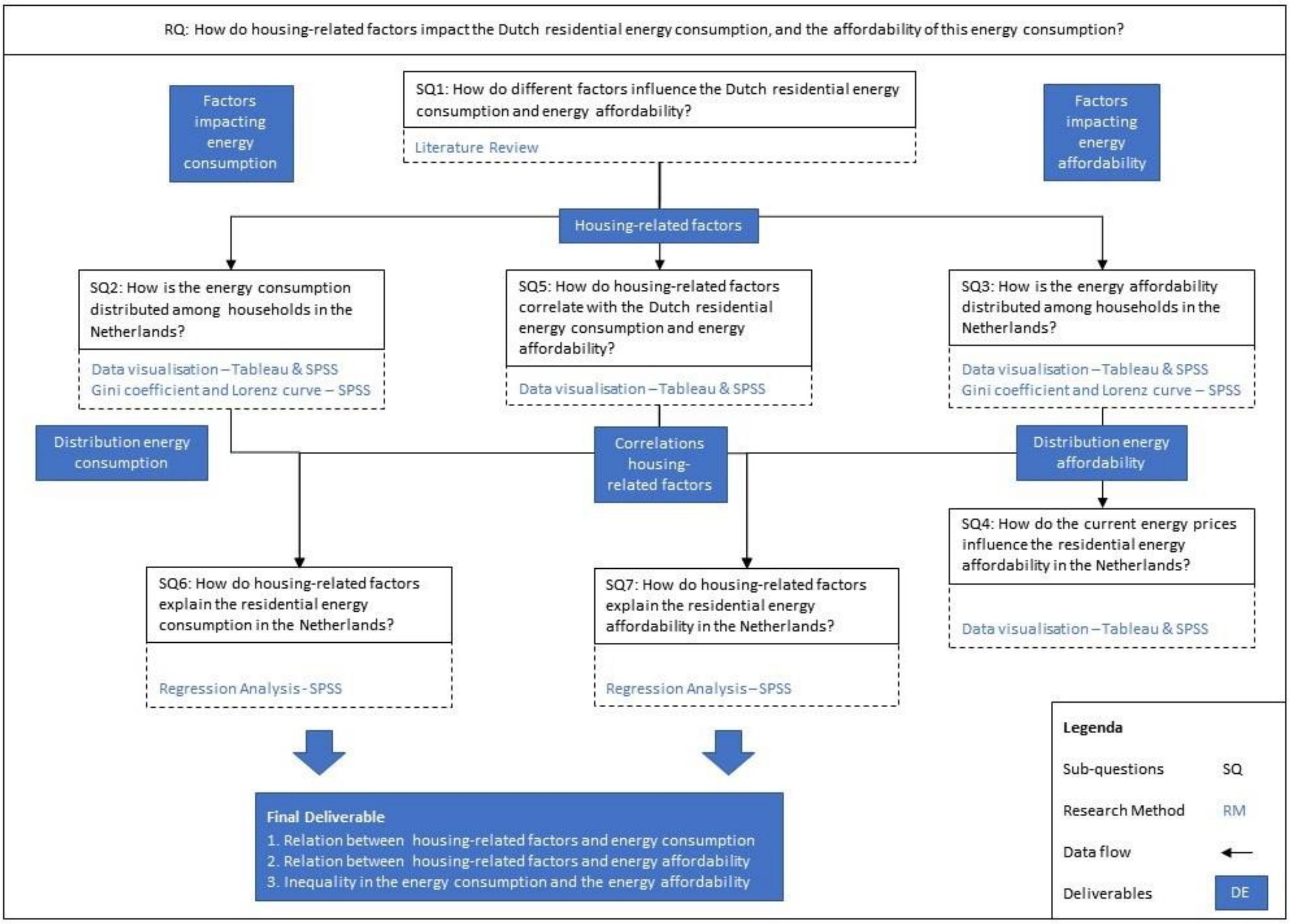

Figure 1: Research Flow Diagram

Structure of the Thesis

This thesis is structured as follows. First, a literature review is presented in chapter 2. This chapter identifies which factors are influencing the energy consumption and the energy affordability. Subsequently, chapter 3 describes the research approach and methodology, and gives an overview of the datasets used. Third, chapter 4 presents an overview of how the core concepts in this research, the energy consumption, energy affordability and housing-related factors, differ among households in the Netherlands. Fourth, chapter 5 elaborates on the correlation between the housing-related factors, and the energy consumption and affordability, based on the regression analyses performed. Finally, chapter 6 discusses the results from the analyses performed in the previous chapters based on the literature, concludes this thesis by answering the main research question, and identifies policy implications and areas of future research.

2. Literature Review

This chapter describes the insights from the literature with regard to the factors impacting the Dutch residential energy consumption and energy affordability. First, the residential energy consumption is defined, and the factors impacting the energy consumption are identified in paragraph 2.1. Second, the energy affordability is described after which the factors impacting this phenomenon are elaborated on in paragraph 2.2. Finally, paragraph 2.3 summarizes the relation between the core concepts found in the literature and reveals which variables are included in further analyses based on the literature. Thus, this chapter answers sub-question 1: How do different factors influence the Dutch residential energy consumption and energy affordability?

2.1 Factors influencing Energy Consumption

Literature Review Process

In order to get an overview of the research being done on the determinants of residential energy consumption, a literature review was performed. The final search term, which is shown in figure 2, was first used in the Scopus database, after which Web of Science was checked to see if articles were missing. To filter out only the most relevant articles, the papers had to satisfy certain criteria. Only papers that covered residential energy consumption solely, included multiple factors impacting this consumption and provided insight in the predicting power of these factors, were selected. The third criteria led to the exclusion of many machine learning studies, which focus on the prediction of the energy consumption with limited factors and without being able to explain the contribution of predicting variables. Two of the most relevant articles related to this were included in the review, to get an overview of this area of research (Szoplik, 2015; Wijaya, Vasirani, Humeau, & Aberer, 2015). The Scopus and Web of Science search resulted in a total of 20 articles. Forward and backward snowballing was applied to these articles, which added another 19 articles to the literature review. All in all, this made a total of 39 relevant articles. An overview of the relevant articles, and the factors covered in them is summarized in table 1. The literature review process is visualised in figure 2.

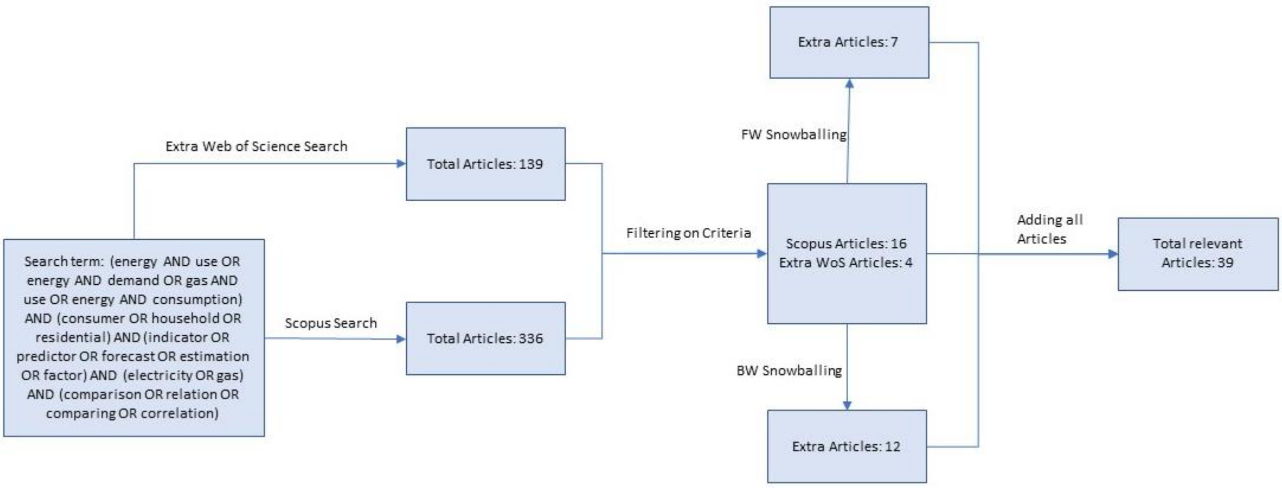

Figure 2: Literature Review Process Energy Consumption

Residential Energy Consumption in this Research

Households consume different sources of energy for various purposes. This research focuses on natural gas and electricity, as these are the main energy sources for residential energy consumption in the Netherlands (Santin, Itard, & Visscher, 2009; Brounen et al., 2012). Natural gas is often used for heating and cooking, whereas electricity is generally used for lighting, refrigeration and other electric appliances (Iraganaboina & Eluru, 2021). In the Netherlands, more than 70% of the residential energy consumption comes from natural gas, which is high compared to other European countries (Mashhoodi, 2021). Most other energy consumption comes from electricity, although district heating exists in some municipalities. The dwellings with district heating are however left outside the scope of this paper, as they are not representative for the Dutch households, and as data on this kind of energy use is limited. In the literature, residential energy consumption is most often measured in kWh per year, which is also used in this study (Cheng & Steemers, 2011; Abrahamse & Steg, 2009; Chalal et al., 2017). Incidentally, other units of energy, such as BTU or m^3 are used (Iraganaboina & Eluru, 2021; Brounen et al., 2012)

Sources of Energy
Residential Energy Consumption covers the use of various sources of energy by a household, for different purposes, such as heating, lighting and cooking (Iraganaboina & Eluru, 2021). Within this area of research, a significant number of articles are written solely on electricity use (Huebner et al., 2016; Ndiaye & Gabriel, 2011), which only covers a part of the energy consumption. Most articles do however cover the total energy consumption of a household. In Europe, this is often limited to gas and electricity (Chalal et al., 2017; Brounen et al., 2012), while users outside Europe generally use a more diverse mix of energy sources. In the US this mainly includes fuel oil next to gas and electricity (Tso & Guan, 2014; Iraganaboina & Eluru, 2021). In Asia, the total energy consumption covers an even more diverse mix, with energy sources such as coal and biomass (Zheng et al., 2014; Miah et al., 2011).

Factors impacting Energy consumption
Different categories of factors were found to impact the residential energy use. Most of the relevant articles focus on the socio-demographic variables of the households and dwelling characteristics (Boukarta & Berezowska-Azzag, 2018; Huebner et al., 2016; Fumo & Rafe Biswas, 2015; Chen, Wang, & Steemers, 2013), which are often powerful predictors of the energy consumption. There are also articles that take into account the behaviour and the attitudes of the users, which can have a significant impact too (Abrahamse & Steg, 2009; Guo, 2018). Kavousian, Rajagopal and Fischer (2013) show this by presenting how climate awareness and behaviour focussed on climate change can result in a lower energy consumption. Weather factors are often considered when comparing users in geographical areas with a different climate (Dong et al., 2018; Iraganaboina & Eluru, 2021), or when focussing on short-term energy consumption, as weather changes impact the energy demand throughout the day (Szoplik, 2015; Wijaya et al., 2015). In some studies, the use or ownership of appliances is included as a determinant (Khan, 2019; Huebner et al., 2016). Sanquist et al. (2012) reveal that for example TV and AC use have a large impact on the total energy consumption of a household.

Socio-Demographic and Weather Factors
Within the socio-demographics and weather variables, several factors have a clear relation with the energy consumption. First, the household income has a strong positive relationship with the energy consumption, meaning that higher income households consume more energy (Brounen et al., 2012; Abrahamse & Steg, 2009). Second, the household size is also clearly related with the energy consumption, as the energy consumption is higher for households with more residents (Iraganaboina, 2021; Huebner et al., 2016). Third, the variable age is often included, although different articles show conflicting results with regard to this factor. Abrahamse and Steg (2009) find for example that no relation exists with age, whereas Brounen et al. (2012) conclude that older residents consume more energy. Fourth, the energy consumption is impacted by the type of household. One person households and households with two parents and children consume less when controlling for other factors (Chalal et al., 2017; Brounen et al., 2012). Fifth, the ethnic background is an important factor, as residents with a migration background consume less energy (Tso & Guan, 2014). Sixth, the temperature or a measure

related to the temperature is often included. The literature reveals that a lower average temperature increases the energy consumption (Santin et al., 2009; Iraganaboina & Eluru, 2021). Factors that are often included as control variables, but don't show a consistent relation with the energy consumption are the education level, the gender and whether household members have social benefits (Abrahamse & Steg, 2009; Santin et al., 2009).

Housing-related Factors

The housing-related factors are also clearly correlated with the energy consumption. First, houses with a larger surface area consume more energy (Iraganaboina & Eluru, 2021; Huebner et al., 2016). Second, the type of house also shows a clear relationship, although the direction can differ per country. Overall, apartments consume less energy, whereas detached houses consume more (Brounen et al., 2012; Tso & Guan, 2014). Third, the relation of energy consumption with year of construction is negative, meaning that older houses consume more energy on average (Santin et al., 2009; Brounen et al., 2012). Fourth, the number of dwellings in a building also influences the energy consumption; a larger number of dwellings in a building results in a lower average energy consumption (Tso & Guan, 2014). Furthermore, the number of floors in a building is sometimes included in the analysis, although no clear relation exists (Huebner et al., 2016).

Urban and Rural

Another housing-related factor related to inequality, that was often included, is the difference in residential energy consumption between rural and urban consumers (Iraganaboina & Eluru, 2021; Miah et al., 2011; Xie et al., 2020). In the US, the main difference between these two groups was in the type of energy sources used. In China and Bangladesh however there's also a clear difference in total energy consumption. Rural households use less energy, partly because of demographic and house characteristics, but also because they often don't have access to certain energy sources (Xie et al., 2020; Miah et al., 2011). The difference in house characteristics and demographic factors such as income does have a large impact on the energy consumption and can be used to reveal the differences in quality of life between rural and urban households (Zheng et al., 2014; Miah et al., 2011). All in all, the literature reveals that energy inequality often exists between rural and urban households. Even in the US, rural households have access to fewer energy sources (Iraganaboina & Eluru, 2021), which raises the question whether this inequality also exists in Europe.

Ownership Structure

Furthermore, some articles included the ownership structure of the dwellings that the residential consumers live in as a key housing-related factor. A study in the US showed that rental homes consume more energy than owner-occupied homes, which could be due to the fact that utilities are often included in the rent (Ndiaye & Gabriel, 2011). In Taiwan, owner-occupied houses consume more energy because they own more appliances (Huang, 2015). Yohanis et al. (2008) support this claim with their research in Northern Ireland, which suggests that homeowners consume 25% more energy, as rental homes are often occupied by low-income families. This does however contradict with research from Santin et al. (2009), which suggests that households with private rent consume more energy due to lower dwelling quality. Overall, it can thus be stated that knowledge in this area is limited and often contradicting, which makes it an interesting area of research. With regard to the difference between social rent and private rent, no clear conclusions can be drawn as Haffner & Boumeester (2015) suggest that there is no difference in energy consumption between these groups, whereas Santin et al. (2009) suggest that private rent consumes more.

Table 1: Literature Overview Energy Consumption

Energy sources: E = Electricity, G = Gas, GE = Electricity & Gas, D = Diverse													
Authors	Year	Title	Energy sources	Analyzed Factors									Region
				Demo-graphics	Dwelling	Attitude	Appliances	Behaviour	Weather	Owning	Urban-Rural	Other	
Abrahamse & Steg	2009	How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings?	D	X		X		X					NL
Baker & Rylatt	2008	Improving the prediction of UK domestic energy-demand using annual consumption-data	GE		X			X					EU
Bedir et al.	2013	Determinants of electricity consumption in Dutch dwellings	E	X	X		X	X					NL
Besagni & Borgarello	2018	The determinants of residential energy expenditure in Italy	D	X	X		X						EU
Borozan	2018	Regional-level household energy consumption determinants: The european perspective	D	X					X			X	EU
Boukarta & Berezowska-Azzag	2018	Assessing Households' Gas and Electricity Consumption: A Case Study of Djelfa, Algeria	GE	X	X		X						AF
Brounen et al.	2012	Residential energy use and conservation: Economics and demographics	GE	X	X								NL
Chalal et al.	2017	The impact of the UK household life-cycle transitions on the electricity and gas usage patterns	GE	X	X			X		X			EU
Chen et al.	2013	A statistical analysis of a residential energy consumption survey study in Hangzhou, China	E	X	X			X				X	ASIA
Cheng & Steemers	2011	Modelling domestic energy consumption at district scale: A tool to support national and local energy policies	GE	X	X			X	X				EU
Dong et al.	2018	A Comparative Analysis of Residential Energy Consumption in Urban and Rural China: Determinants and Regional Disparities	D	X					X		X		ASIA
Frederiks et al.	2014	The Socio-Demographic and Psychological Predictors of Residential Energy Consumption: A Comprehensive Review	D	X	X	X							-
Fumo & Rafe Biswas	2015	Regression analysis for prediction of residential energy consumption	D	X	X		X		X				US

Guo	2018	Residential electricity consumption behavior: Influencing factors, related theories and intervention strategies	E	X	X	X		X						-
Huang	2015	The determinants of household electricity consumption in Taiwan: Evidence from quantile regression	E	X	X					X				ASIA
Huebner et al.	2016	Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes	E	X	X	X	X	X						EU
Iraganaboina & Eluru	2021	An examination of factors affecting residential energy consumption using a multiple discrete continuous approach	D	X	X		X		X		X			US
Jiang et al.	2019	Energy consumption by rural migrant workers and urban residents with a hukou in China: quality-of-life-related factors and built environment	D	X	X		X			X	X			ASIA
Jones & Lomas	2015	Determinants of high electrical energy demand in UK homes: Socio-economic and dwelling characteristics	E	X	X									EU
Jones et al.	2015	The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings	E	X	X		X							EU
Kavousian et al.	2013	Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior	E	X	X	X	X	X	X					US
Khan	2018	Household factors and electrical peak demand: a review for further assessment	E	X	X		X	X	X				X	OC
McLoughlin et al.	2012	Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study	E	X	X									EU
Miah et al.	2011	Domestic energy-use pattern by the households: A comparison between rural and semi-urban areas of Noakhali in Bangladesh	D	X	X						X			ASIA

Ndiaye & Gabriel	2011	Principal component analysis of the electricity consumption in residential dwellings	E	X	X		X	X		X			CA
Rhodes et al.	2014	Clustering analysis of residential electricity demand profiles	E	X			X					X	US
Sanquist et al.	2012	Lifestyle factors in U.S. residential electricity consumption	E	X	X		X	X	X				US
Santin et al.	2009	The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock	GE	X	X			X		X			NL
Steemers & Yun	2010	Household energy consumption: a study of the role of occupants	D	X	X			X	X				US
ŠTreimikienė	2014	Residential energy consumption trends, main drivers and policies in Lithuania	D	X	X		X						EU
Swan & Ugursal	2009	Modeling of end-use energy consumption in the residential sector: A review of modeling techniques	D	X	X		X	X	X				-
Szoplik	*2015*	*Forecasting of natural gas consumption with artificial neural networks*	*G*						*X*				*EU*
Tso & Guan	2014	A multilevel regression approach to understand effects of environment indicators and household features on residential energy consumption	D	X	X		X		X	X			US
Wang et al.	2018	Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns	E	X	X	X	X						EU
Wei et al.	2014	Energy spending and household characteristics of floating population: Evidence from Shanghai	D	X			X				X		ASIA
Wijaya et al.	*2015*	*Cluster-based aggregate forecasting for residential electricity demand using smart meter data*	*E*				*X*		*X*			*X*	*EU*
Xie et al.	2020	Does urbanization increase residential energy use? Evidence from the Chinese residential energy consumption survey 2012	D	X	X						X	X	ASIA
Yohanis et al.	2008	Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use	E	X	X		X	X		X			EU

Zheng et al.	2014	Characteristics of residential energy consumption in China: Findings from a household survey	D	X	X		X	X			X		ASIA

2.2 Factors influencing Energy Affordability

Literature Review Process

In order to get insight in the factors impacting the energy affordability, a literature review was performed, following a similar process as the literature review on the energy consumption. Again Scopus was used first to gather the relevant articles, after which Web of Science was checked for missing articles. Articles were selected on three criteria. First, the dependent variable of the research had to be related to residential energy affordability, meaning all research that covers the energy expenditure of households. Second, the articles needed to include multiple factors impacting this affordability. Finally, the research needed to contain a quantitative indication of the impact that these factors have. In the end, 18 relevant articles were selected, as is visualised in figure 3.

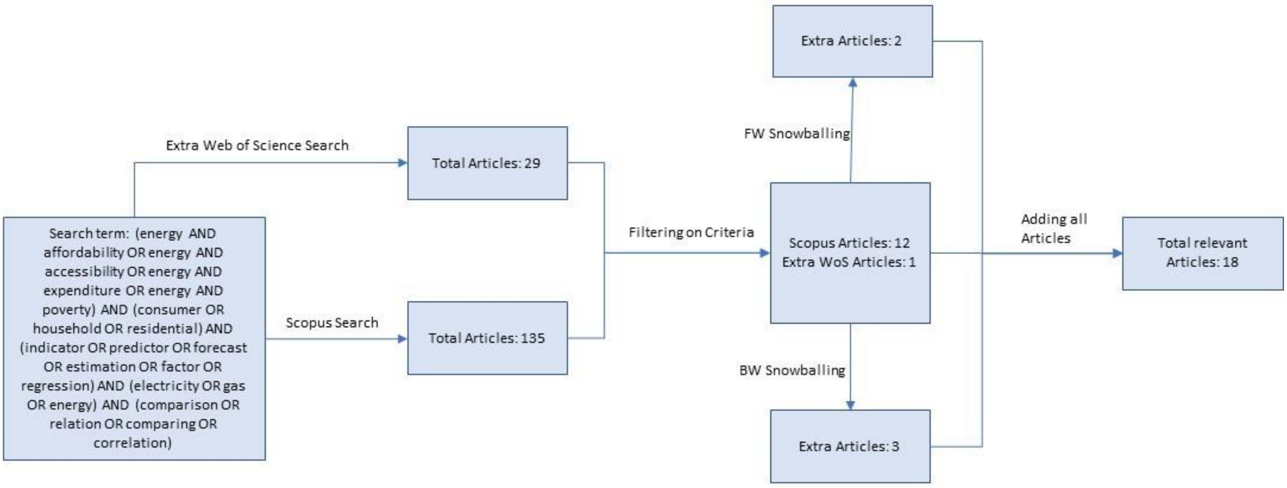

Figure 3: Literature Review Process Energy Affordability

Overall, it can be stated that less literature is available on energy affordability, when compared to energy consumption. Furthermore, the articles on energy affordability are also more diverse. Some of them relate to overall residential expenditures, where others are linked to 'energy poverty' or 'fuel poverty'. And even within these domains, the unit of analyses differs, ranging from energy expenditures only to the expenditure-to-income ratio and other more complex measures (Bardazzi, Bortolotti, & Pazienza, 2021). Energy affordability is therefore lacking a strong literary base, making it an interesting addition to this research. An overview of the relevant articles found and the factors used to predict the affordability in these articles is shown in table 2.

Residential Energy Affordability in this Research

The concept of energy affordability can be complex. When discussing energy affordability, the expenditure-to-income ratio is most commonly used as a measure, which is the household income divided by the energy costs. This ratio is mainly dependent on the the energy prices, the thermal quality of the house and the income of the households (Galvin & Sunikka-Blank, 2018; Haffner & Boumeester, 2015). As the income and dwelling characteristics are powerful determinants of energy use, the energy consumption and energy affordability are closely related. An important difference is however that the affordability shows what percentage of the household income is spend on energy. By doing this, the affordability reveals the economic situation of a household, and the impact of its energy consumption on its residual income (Haffner & Boumeester). This explains why the energy affordability is a key inequality indicator, as mentioned in the introduction.

Income is a crucial part of energy affordability, on the one hand as a key determinant of the energy consumption, and on the other hand as a levelling factor in the expenditure-to-income ratio. It is determining the energy consumption both in a direct and indirect way. Directly, as high-income households heat more rooms and use more appliances (Galvin & Sunikka-Blank, 2018; Cheng & Steemers, 2011), and indirectly as high-income households live in larger dwellings, which consume more energy. Research in England has shown that up to 85% of the variance in energy demand can be explained by the income and dwelling characteristics of a household (Cheng & Steemers, 2011). On the other hand however, when looking to income as a levelling factor, the literature reveals that low-income households do relatively spend more on energy. This is mainly because of the thermal quality of a dwelling, and because some consumption patterns such as lighting and heating cannot be changed easily (ŠTreimikienė, 2014; Galvin & Sunikka-Blank, 2018). These higher energy costs lead to a lower affordability, and sometimes even to energy poverty (Haffner & Boumeester, 2015; Lin & Wang, 2020).

Another important aspect of the energy affordability is the energy price, as an increasing price can cause low-income households to spend a disproportional amount of their income on energy. This leads to a low affordability, more energy poverty, and a higher energy inequality (Kontokosta, Reina, & Bonczak, 2019; Galvin & Sunikka-Blank, 2018). On the other hand, a low price decreases the energy costs and improves the energy affordability. As the energy price is an external factor that households can't influence, price changes often result in an unexpected change in their energy affordability (Kontokosta et al., 2019).

All in all, the affordability provides insight in the impact of the energy inequality on people's lives, as the percentage of income spend on energy reveals to what extend households experience poverty due to their high energy costs. A household is considered energy poor when more than 10% of their income is spend on energy (Lin & Wang, 2020; Haffner & Boumeester, 2015). Such households can't keep their home adequately warm (Galvin & Sunikka-Blank, 2018). High energy costs can push people into poverty, especially as other large costs such as rent are often fixed. The current high energy prices are thus likely to decrease the affordability of energy and increase the energy poverty and energy inequality in the Netherlands.

Sources of Energy

Similar to the literature on energy consumption, the literature on energy affordability covers different kinds of energy sources, ranging from gas and electricity to a more diverse energy mix (Charlier & Kahouli, 2019; Ismail & Khembo, 2015). However, when compared with the articles on energy consumption, fewer articles are written on electricity consumption only. This is due to the fact that in order to give an indication of the energy affordability of a household, the full energy mix used needs to be considered. As affordability is a relative measure, dependent on the income, excluding a major energy source would result in a lower calculated affordability, that can differ substantially from the actual energy affordability. Thus, the majority of articles analyse the affordability based on a diverse energy mix (Riva et al., 2021; Karásek & Pojar, 2018; Belaïd, 2018). This does however mean that even within articles that cover a diverse energy mix, differences in the energy sources used remain significant. Most European and North American countries, such as Italy or Canada, have a high share of modern energy sources, including electricity and natural gas (Das et al., 2022; Besagni & Borgarello, 2019b). Other more traditional fuels, such as firewood and coal are more limited in these countries, while they are still a main source of energy in developing countries in Africa and Asia (Nguyen et al., 2019; Khundi-Mkomba, Saha, & Wali, 2021).

Factors Impacting Energy Affordability

The factors used to predict the energy affordability, are relatively similar to those for the prediction of the energy consumption. Socio-demographic variables and dwelling characteristics have the highest impact in most articles (Longhi, 2015; Papada & Kaliampakos, 2017; Kontokosta et al., 2019). Furthermore, the weather or climate is included in case geographical differences are significant (Bardazzi et al., 2021; Papada & Kaliampakos, 2017). Contrary to energy consumption, behaviour and attitudes aren't important predictors for energy affordability. This could be because energy affordability is a more high-level measure that often covers a diverse mix of fuels and thus behaviours. A measure related to behaviour, that is included in some cases, is the ownership of specific appliances, usually when these appliances are responsible for a large part of the energy consumption. Tabata and Tsai (2020) reveal for example that air conditioning modules have a significant impact on the energy affordability in Japan during the summer months. The energy price was only considered a crucial factor in one of the papers (Charlier & Kahouli, 2019), opposed to expectations, as energy prices are increasing and do have a large impact on the affordability (Karásek & Pojar, 2018; Bardazzi et al., 2021). In most cases however, the energy price was not included because it is viewed as a given factor that households can't impact, or is considered a stable factor that is similar for different households (De Arce & Mahía, 2019; Riva et al., 2021).

Socio-Demographic and Weather Factors

There are multiple socio-demographic and weather variables that have a strong relationship with the energy affordability. First, the ethnicity of residents influences the affordability, as people with a foreign background have a lower affordability (Belaïd, 2018; Kontokosta et al., 2019). Second, the income has a clear negative relation with the expenditure to income ratio (Mashhoodi et al., 2018; De Arce & Mahia, 2019), as mentioned because the income is a crucial part of the affordability. Third, the type of household impacts the affordability; one person households spend more on energy, whereas married couples spend less on energy (Das et al., 2022; Riva et al., 2022). Fourth, age is often included, but not always significant. When age is significant, the elderly have a lower energy affordability (Besagni & Borgarello, 2018; Riva et al., 2022). Fifth, the temperature is correlated with the energy affordability; a lower average temperature results in a higher expenditure to income ratio (Mashhoodi et al., 2018). Furthermore, the household size, education level, gender, social benefits and employment status are often included in analyses on the energy affordability. Nevertheless, their impact is limited in most cases, meaning that no significant relation exists with the energy affordability (Belaïd, 2018; Mashhoodi et al., 2018; De Arce & Mahia, 2019).

Housing-related Factors

When considering the housing-related factors, there are also strong relations with the energy affordability, represented by the expenditure to income ratio. First, the type of house is an important factor, as apartments have lower energy expenditures and detached houses have higher energy expenditures, when controlling for other variables (Besagni & Borgarello, 2018; Riva et al., 2021). Second the surface area has a strong positive relationship with the energy expenditure to income ratio, meaning that houses with a larger surface area have a higher expenditure to income ratio (Belaïd, 2018; Besagni & Borgarello, 2018). Third, the year of construction has a strong influence on the energy affordability in the literature. However, in contrast to the previous factor it has a negative relationship with the expenditure to income ratio, meaning that households living in older houses spend a larger percentage of their income on energy (De Arce & Mahia, 2019; Das et al., 2022).

Urban and Rural

A large number of articles also discusses the difference between urban and rural energy affordability. In developing countries, such as Vietnam and Rwanda, rural households spend a larger part of their income on energy, mainly because they are poorer (Nguyen et al., 2019; Khundi-Mkomba et al., 2021). However, even in North America, rural households have a significantly lower affordability, when compared to urban households (Das et al., 2022; Riva et al., 2021). In Europe, the research on the rural and urban difference in energy affordability is limited. The only studies found in France and the UK show no large differences (Belaïd, 2018; Longhi, 2015), which might be the reason that this area has received less attention so far. Nevertheless, uncertainty remains about the difference between rural and urban energy affordability, as no clear conclusions are drawn. Thus, this study explores whether this difference exists in Europe, and more specifically in the Netherlands.

Ownership Structure

The ownership structure is in some of the cases also included as a housing-related predictor of energy affordability. A distinction is made between owned and rental homes, and occasionally the difference between private and social rent is also considered (Mashhoodi, Stead, & Van Timmeren, 2018). In Canada, households living in rental homes spend a larger proportion of their income on energy. Riva et al. (2021) state that this is because of the principal-agent problem, which causes landlords to refrain from investing in energy efficient rental homes, leaving the households in these homes with high energy costs. Research in France and the Netherlands confirms this, by revealing that households in privately rented homes experience energy poverty more often, meaning that they struggle with their energy affordability (Belaïd, 2018; Mashhoodi et al., 2018). In Spain however, households in rental homes have a slightly higher energy affordability, which is contradicting to the other European research (De Arce & Mahía, 2019). Nevertheless, the article from the Netherlands is most applicable to this research, suggesting that households living in privately rented homes experience a lower energy affordability (Mashhoodi et al., 2018). No clear conclusion can however be drawn from this article only, as this article only covers private rent, and makes no distinction between homeowners and social rent. Haffner and Boumeester (2015) do however state that all tenants have a lower affordability when compared to homeowners, and Hoekstra (2017) adds to this that social rent households have a significantly lower income, which is likely to lower their energy affordability. Rental households are thus expected to have a higher expenditure to income ratio. With regard to social and private rent there is no clear expectation, as Hoekstra (2017) and Mashhoodi et al. (2018) show contradicting findings.

Table 2: Literature Overview Energy Affordability

Energy sources: E = Electricity, G = Gas, GE = Electricity & Gas, D = Diverse													
Authors	Year	Title	Energy sources	Analyzed Factors								Expense/income-ratio	Region
				Demographics	Dwelling	Energy Price	Appliances	Weather	Owning	Urban-Rural	Other		
Bardazzi et al.	2021	To eat and not to heat? Energy poverty and income inequality in Italian regions	GE	X				X			X	X	EU

Belaïd	2018	Exposure and risk to fuel poverty in France: Examining the extent of the fuel precariousness and its salient determinants	D	X	X			X	X	X			EU
Besagni & Borgarello	2019	Measuring Fuel Poverty in Italy: A Comparison between Different Indicators	D	X	X			X			X		EU
Besagni & Borgarello	2019	The socio-demographic and geographical dimensions of fuel poverty in Italy	G/D	X	X			X					EU
Charlier & Kahouli	2019	From Residential Energy Demand to Fuel Poverty: Income-induced Non-linearities in the Reactions of Households to Energy Price Fluctuations	GE	X	X	X		X				X	EU
Das et al.	2022	Quantifying the prevalence of energy poverty across Canada: Estimating domestic energy burden using an expenditures approach	D	X	X		X		X	X	X	X	NA
De Arce & Mahía	2019	Drivers of Electricity Poverty in Spanish Dwellings: A Quantile Regression Approach	E	X	X			X	X		X	X	EU
Hartono et al.	2020	Modern energy consumption in Indonesia: Assessment for accessibility and affordability	D	X	X					X		X	ASIA
Ismail & Khembo	2015	Determinants of energy poverty in South Africa	D	X	X					X	X	X	AF
Karásek & Pojar	2018	Programme to reduce energy poverty in the Czech Republic	D	X	X							X	EU
Khundi-Mkomba et al.	2021	Examining the state of energy poverty in Rwanda: An inter-indicator analysis	D	X			X			X			AF
Kontokosta et al.	2019	Energy Cost Burdens for Low-Income and Minority Households	GE	X	X							X	NA
Longhi	2015	Residential energy expenditures and the relevance of changes in household circumstances	D	X	X				X	X	X		EU
Mashhoodi et al.	2018	Spatial homogeneity and heterogeneity of energy poverty: a neglected dimension	GE	X	X			X	X		X	X	NL
Nguyen et al.	2019	Energy transition, poverty and inequality in Vietnam	D	X						X			ASIA

Papada & Kaliam-pakos	2017	Energy poverty in Greek mountainous areas: a comparative study	D	X	X			X				X	EU
Riva et al.	2021	Energy poverty in Canada: Prevalence, social and spatial distribution, and implications for research and policy	D	X	X				X	X	X	X	NA
Tabata & Tsai	2020	Fuel poverty in Summer: An empirical analysis using microdata for Japan	D	X	X		X					X	ASIA

2.3 Summary of Core Concepts and Impacting Factors

This paragraph elaborates on the connections between the core concepts in this research and summarizes these connections in a conceptual model. It describes how the energy consumption and the energy affordability are related, and how the different factors impact the energy consumption and energy affordability. Thus, the goal of this paragraph is to put the literature review into context, show how the different concepts discussed link together, and describe how these concepts are analysed in the remainder of this thesis.

How Different Factors Impact the Energy Consumption and Energy Affordability

The literature review revealed that the residential energy consumption and energy affordability are influenced by different factors. The socio-demographic factors, weather factors and the housing-related factors are however responsible for most of the variance in the energy consumption and energy affordability (Guo et al., 2018; Brounen et al., 2012; Cheng & Steemers, 2011; ŠTreimikienė, 2014). Other variables related to behaviour generally have less predicting power and are thus not included in this study. That doesn’t mean that behaviour is not an important determinant of energy consumption and affordability. Guo et al. (2018) state that socio-demographic factors influence the energy consumption through behaviour. An example of this is that elderly people often prefer a warmer home, which results in them turning on the heating more often. This behaviour results in a higher energy consumption and lower energy affordability, and is based on the socio-demographic variable ‘age’. Thus, socio-demographics are the core determinants of residential energy consumption (Abrahamse & Steg, 2009).

Housing-related factors do have a high predicting power but are causally dependent on socio-demographic and weather factors such as income, household size and temperature. Households choose in what kind of dwelling they live based on their socio-demographics, and often also within the limitations that socio-demographic factors such as income provide (ŠTreimikienė, 2014; Jones & Lomas, 2015). Furthermore, the weather influences factors such as insulation, and thus also housing-related factors linked to this, for example the house type and the year of construction (Das, Martiskainen, & Li, 2022). Nevertheless, housing-related factors are important predictors of the energy consumption and affordability, as these factors do differ among houses in the same region and among people with a similar socio-demographic background. This is partly explained by the fact that housing isn’t a flexible factor, meaning that households don’t necessarily move to another house when their socio-demographic situation changes, as moving barriers are significant (Haffner & Boumeester, 2015).

Connection of Core Concepts

The relation between the socio-demographic factors, the housing-related factors, the energy consumption and the energy affordability is visualised in figure 4. The figure shows how socio-demographic factors impact the energy consumption directly and via the housing-related factors. The energy expenditure to income ratio, which represents the energy affordability, is the income divided by the total energy costs, which depend on the energy consumption and the energy price. The blue concepts in the figure, including the socio-demographics, housing-related factors and energy price, are viewed as independent variables in this research, while the orange concepts, are the dependent variables. The energy consumption and energy affordability are the core units of analysis in this thesis, while the total energy costs is only used to compute the energy affordability.

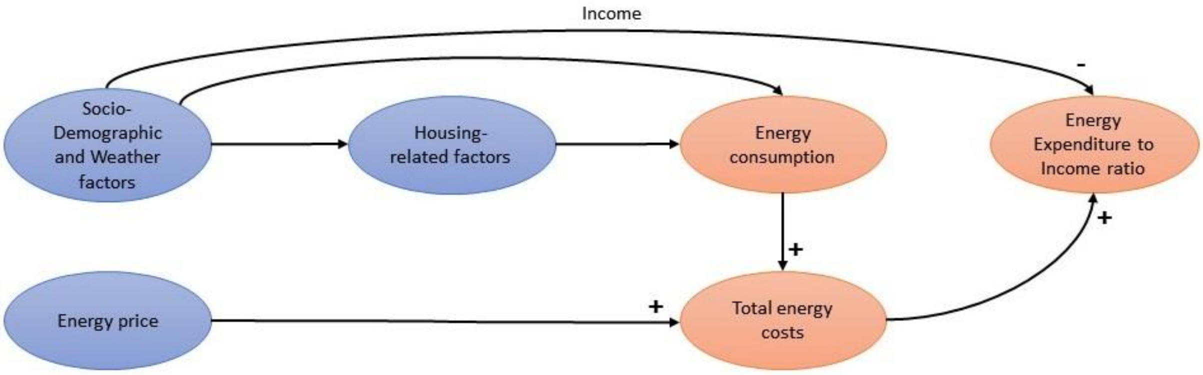

Figure 4: Conceptual Model of Core Concepts

Energy Consumption Factors Included in the Analyses

The factors that are included in the correlation and regression analyses of energy consumption are presented in figure 5, where they are added to the conceptual model. Factors that are expected to have a positive relationship with energy consumption based on the literature are listed with a plus-sign, whereas factors with a negative expected relationship are listed with a minus-sign. Variables that have no clear expected relation with energy consumption are listed with a tilde. On top of the factors discussed in the literature review in 2.1, the WOZ-value of houses is added as a variable, as it is expected to have a positive relationship with the energy consumption due to the fact that it captures additional dwelling information that is not represented by other housing-related variables, and because houses with a high WOZ-value have a lower energy efficiency (Boesveld, 2021; Abidoye & Chan, 2016). The percentage of district heating is also added for data quality reasons, which is further elaborated on in paragraph 3.2. The education of residents is left out of the analysis, as no data was available on this. This is however not expected to have a significant impact on the analysis, as the education is not a crucial control variable in previous literature (Iraganaboina & Eluru, 2021).

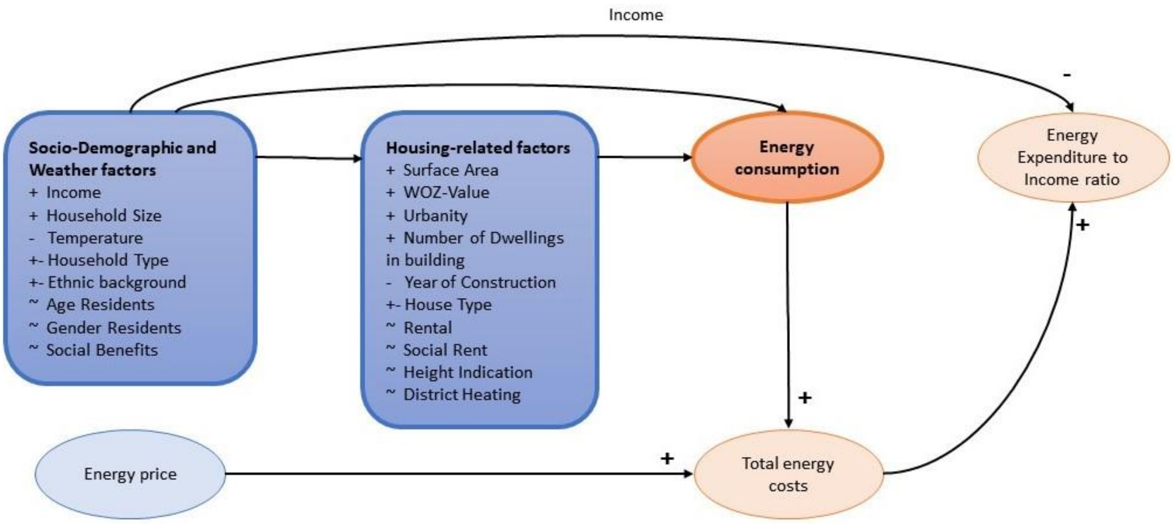

Figure 5: Conceptual Model with factors impacting the Energy Consumption

Energy Affordability Factors Included in the Analysis

The factors that are used in the statistical affordability analyses based on the literature are presented in figure 6, where they are included in the conceptual model. As with the analyses on energy consumption, the WOZ-value and district heating are also included in the analyses. This is again because the WOZ-value explains additional housing information and because houses with a high WOZ-value have a lower energy efficiency (Boesveld, 2021; Abidoye & Chan, 2016); including district heating is needed for data quality control. The education level and the employment status are excluded from the analysis as no data on these variables was found. This is justifiable as they are no key factors in the literature (Mashhoodi et al., 2018; Kontokosta et al., 2019). The number of dwellings per building, and the height indication weren't found to be relevant in energy affordability literature, but are used in the regression analysis, as these factors are expected to impact the energy consumption. Thus, including these variables is desirable, on the one hand because the energy consumption is related to the energy affordability, and on the other hand to be able to compare the regression analyses on energy consumption and energy affordability. All in all, when comparing figure 5 and 6, the socio-demographic, weather and housing-related factors included in the analyses are similar for energy consumption and energy affordability. The relationship that these factors have with these phenomena does however differ, confirming the need for analysing the energy consumption and the energy affordability separately in the statistical analyses.

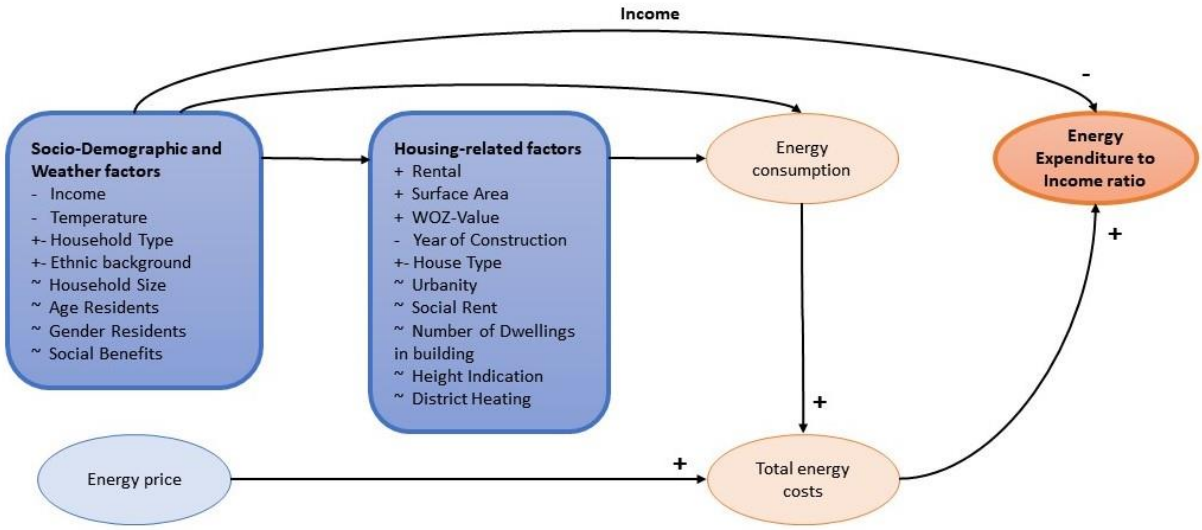

Figure 6: Conceptual Model with factors impacting the Energy Affordability

3. Methodology & Data

This chapter elaborates on the methodology used to answer the research questions, and the data that is analysed in this process. Paragraph 3.1 provides an overview of the research approach and methodology, motivates the regression analyses, and introduces the data analysed. Afterwards, paragraph 3.2 gives a more extensive overview of the datasets used and the data preparation operations performed to integrate the different datasets and make the data analysable. Next, paragraph 3.3 provides an overview of the variables included in the final dataset and the variable transformations required for the statistical analyses. Finally, paragraph 3.4 presents the descriptive statistics of these variables, and discusses remarkable observations.

3.1 Research Approach & Methodology

Description Research Approach

In order to evaluate the relationship between the housing-related factors and the residential energy consumption and affordability, a quantitative study is performed. A quantitative approach fits this research well, because a relation between different factors is analyzed, and as quantitative empirical data is available on the topic (Sukamolson, 2007). This research leverages energy consumption data, weather data, socio-demographic data and housing-related data to answer the main research question.

The study combines a correlational and a causal-comparative approach, as it is non-experimental and analyzes both the relations between factors and the differences between separate groups (Johnson, 2001). On the one hand, the correlational approach results in analyzing the relationship between the quantitative housing-related factors, such as the surface area and the year of construction, and the energy consumption and affordability. The causal-comparative approach on the other hand, is put into practice by analyzing the differences in the energy consumption and affordability between groups, e.g. between urban and rural households and rental and owner-occupied households. The research is cross-sectional, meaning that it aims to give insight in the current relationship between the housing-related factors and the energy consumption and energy affordability. Longitudinal data over multiple years are pooled to strengthen the power of the analyses, but not to analyze how the energy consumption is developing over time. Furthermore, the study is not only descriptive, but also explanatory (Johnson, 2001) as it aims to reveal how different factors impact the residential energy consumption and affordability.

Strengths & Limitations Research Approach

Although a quantitative approach as described above is generally well-structured, allows the exploration of large sets of information and provides clear results, limitations do also exist. Quantitative studies often have a smaller scope and lack context of the problem. Not all information can be captured in numbers, and correlation and statistical significance can't always explain the underlying mechanism of a phenomena. Thus, it is important to be cautious while interpreting the results of the analyses performed, and to focus on understanding the environment surrounding the problem (Queirós, Faria, & Almeida, 2017). This study aims to achieve this with an extensive literature study on the factors impacting energy consumption, providing the quantitative analyses with context. In case the literature suggests that a certain relationship exists, quantitative analyses are used to test this claim and determine the strength of the relationship. Another limitation is that the data quality is sometimes not good enough for the required analyses, e.g. because of missing information. In this study data preprocessing methods, such as imputation, are used to be able to analyze the data.

Review of Methods Used in Literature

The articles on the factors determining residential energy consumption and energy affordability use different methods, ranging from the application of theories through qualitative research (Abrahamse & Steg, 2009) to machine learning and AI (Szoplik, 2015). Most research does however make use of statistical methods, among which linear regression is the most common method used (Fumo & Rafe Biswas, 2015). Regression analyses are used to determine the relevance of different factors to the energy consumption and affordability, measure the correlation between these factors and predict the residential energy consumption and affordability based on these factors (Borozan, 2018; Chen et al., 2013; Chalal et al., 2017). Machine learning and AI are often used in combination with historical smart meter data and weather data to predict short term, e.g. hourly, energy consumption, without considering demographic or housing factors. For long-term prediction, which is focussed on the predictors of the energy consumption, these methods are however often inferior to conventional statistical analyses. Although the prediction accuracy of machine learning algorithms may be good, the explainability of these algorithms is usually poor, which makes most of them inadequate for analysing the relation between specific factors and the energy consumption or energy affordability (Szoplik, 2015; Wijaya et al., 2015).

Methodology Used: Multivariate Regression Analysis

Therefore, this study performs a statistical analysis of the residential energy consumption, making use of regression. A regression model combines data on multiple factors in a model, where the dependent variable, in this case the residential energy consumption and energy affordability, is predicted based on multiple coefficients or parameters, which match the relevant factors. The model is optimized based on its fit with the data (Swan & Ugursal, 2009). Multivariate regression is used, as both the predictor and the response variables are expected to correlate with each other (Fumo & Rafe Biswas, 2015). The main response variables in this research are the energy consumption and energy affordability, which do correlate with each other (Cheng & Steemers; Wei et al., 2014). In case there is a strong relationship between different variables, multicollinearity may occur, meaning that the model becomes less stable and reliable due to the correlation between the predictor variables. As many factors are expected to be correlated, multicollinearity is checked for and reduced when needed, by applying a principal component analysis (Fumo & Rafe Biswas, 2015). Next to the regression analysis, the descriptive statistics of the variables included are also analysed, using Tableau and SPSS.

Strengths & Limitations Multivariate Regression Analysis

A strength of multivariate regression models is that they clearly show the relationship between different variables, which means that they can be used to understand causal relations between certain factors (Jeon, 2015). Correlation does however not have to mean that a causal relation exists, and therefore it is important to combine regression analysis with theory before using its results for prediction or explanation. Another risk with regression analyses is that they are built on certain assumptions, including a normal distribution of the errors and a linear relationship between the variables. Therefore, it is crucial to check whether the assumptions hold for the data analysed. Furthermore, the quality of the analyses is off course heavily dependent on the quality of the data. Pre-processing is needed before being able to use the data in a regression analysis (Nunes et al., 2015). Moreover, it is critical to not only use the statistical significance and the sum of squares of variation to determine the strength of the model. Jeon (2015) states that it is important to interpret the slope estimate and look to the actual distribution of the data to see if the regression model makes sense. Overall however, multivariate regression models are very suitable for this research, given that these models are able to identify the relation between different factors and a certain phenomenon, which is exactly what this research is focussing on.

Regression Model Design

This study uses SPSS to perform the regression analyses. The enter regression method is chosen for building the models as it is the most common method for this type of analysis, and because it includes all variables. This is desirable because the dataset includes the data of the entire population, meaning that it is unnecessary to exclude variables because they don't have a significant regression coefficient, as the model doesn't have to be extrapolated to the population, but is build on population data. The regression model is divided in two blocks. The first block contains the control variables, the socio-demographic data and the weather data, while the second block contains the housing-related factors. This two-block distinction improves the interpretability of the model. First, it reveals the extra variance that is added by the housing-related factors, which gives an indication of the impact that these factors have on the energy consumption and energy affordability. Second, it allows for the interpretation of the socio-demographic and weather variables without the bias of the housing-related factors. As the housing-related factors don't influence the demographics or the weather, it would be unrealistic to build a model in which the housing-related factors explain a certain part of the variation in the socio-demographics and weather variables. The two blocks are therefore based on the causal relationship; the control variables are added first because they influence the housing-related factors, and not the other way around. Furthermore, categorical variables such as ownership are entered as percentual ratio-variables due to the fact that a postal code contains multiple households. A principal component analysis is used to check for multicollinearity.

Datasets Used

Seven datasets are used for the statistical analyses described above. First, the 'CBS-postcode-6' database is included (CBS, 2016–2018), which contains socio-demographic and dwelling-related variables for households per postal code, as well as their electricity and gas consumption. The age, gender, migration background, household composition, number of households, household size and number of people with social benefits, electricity consumption and gas consumption are recorded. On top of that some dwelling-related variables are included: Tenure and WOZ value. The income class of households is in there as well, which is used for calculating the energy affordability. This research includes datasets from 2016 up until 2018, which is the most recent year with income data available.

Second, the 'BAG-adressen-woning' dataset is used (NLExtract, 2021–2022), which is mainly based on Kadaster data. This dataset contains data for each dwelling in the Netherlands, including the type of dwelling, square meters, indication of height, number of dwellings per building, surface area, and year of construction. The dataset contains the data of the dwellings of 2021, and of March 2022, which is the most up-to-date data at the time of writing. As the timeframe of this data differs slightly from the timeframe of the first dataset, the validity of the data was checked, and dwellings from later years were removed. Similar variables in the CBS-postcode-6 data and the BAG-adressen-woning data were compared, and pre-processing was applied to the dwelling data as described in paragraph 3.2.

Third, the 'CBS-inkomen-van-huishoudens' dataset (CBS, 2016–2020), which includes the average income for different income groups, is combined with the CBS-postcode-6 data to approximate the average household income per postal code. Fourth, the 'CBS-aardgas-en-elektriciteit-prijzen' dataset is used to calculate the energy affordability (CBS, 2016–2021). It contains the delivery price, network price and transaction price for natural gas and electricity for households in the Netherlands, from 2016 to 2021. Fifth, the 'CBS-energietarieven' dataset is analysed (CBS, 2018–2022), which also contains the different cost components of the electricity and gas prices for households in the Netherlands; the data is aggregated per month from 2018 to April 2022. The difference with the CBS-aardgas-en-elektriciteit-prijzen data is that this dataset includes prices per month, and contains the most recent prices, which are needed for the additional affordability analysis. Thus, the energy prices in this dataset are used for calculating the affordability of the energy consumption for 2021 and 2022.

Sixth, the ‘RVO-warmtenetten’ dataset is merged with the CBS-postcode-6 data to be able to exclude the households with district heating (RVO, 2017–2021), as the district heating energy consumption is not included in the energy data. This dataset includes the municipality, neighbourhood code and percentage of dwellings with district heating. The data covers district heating from 2017 to 2021, and thus all current relevant district heating areas. Seventh, the ‘KNMI-dagwaarnemingen’ dataset is used to control for the weather differences within the Netherlands (KNMI, 2016–2018). It contains the average temperature, the minimum temperature and the sunshine duration, from 2016 until 2018. In the end, all these datasets were merged on a postal code 6 level, covering the years 2016, 2017 and 2018. An overview of the relevant datasets can be found in table 3.

Table 3: Relevant Datasets

Dataset	**Relevant Variables**	**Temporal Granularity**	**Timeframe**	**Aggregation level**
CBS – Kerncijfers postcode-6	Postal code, age, municipality, gender, migration background, household composition, number of households, household size, percentage of people with social benefits, ownership structure, WOZ-value, income, gas consumption, electricity consumption	Yearly data	2016-2018	Postal code
BAG-adressen-woning	Postal code, type of dwelling, number of dwellings per building, indication of height, square meters, year of construction	Cross-section (one per year)	2021, 2022	Building or address
CBS – Inkomen van huishoudens	Average income for each 10% group	Yearly data	2016-2020	Country
CBS – Aardgas en Elektriciteit prijzen	Transaction price gas, delivery price gas, network price gas, transaction price electricity, delivery price electricity, network price electricity	Quarterly & yearly data	2016-2021	Country
CBS – energie-tarieven	Transport costs, variable costs, sustainable energy storage tax, energy tax (all variables are for gas and electricity)	Monthly & yearly data	Jan 2018 – April 2022	Country
CBS – Gebieden in Nederland	Municipality, urbanity level, addresses per squared kilometre	Yearly data	2016-2018	Municipality
RVO – Warmte-netten	Neighbourhood code, municipality, percentage of dwellings with district heating	Yearly data	2017-2021	Neighbour-hood
KNMI – dag-waarnemingen	Average temperature, minimum temperature, sunshine duration	Daily data	2016-2018	Weather station

Research Overview

All in all, the proposed research aims to explore how housing-related factors impact the residential energy consumption and energy affordability in the Netherlands. For this, data on the electricity and gas consumption of Dutch households, is combined with data from the CBS and the Kadaster on respectively socio-demographics and building characteristics. The data analysed is aggregated per 6-digit postal code, to prevent privacy issues, and covers the years 2016, 2017 and 2018. The research focusses on housing-related factors, including dwelling characteristics, ownership status and location, which can be used to compare urban and rural areas. Socio-demographic variables, such as age, household composition and household size are controlled for, as these heavily correlate with energy use. The data analysis makes use of multivariate regression to see which variables are the best predictors of residential energy use and energy affordability, and how these variables relate to each other. The results of these analyses are used to assess the equality of the residential energy consumption in the Netherlands. This is done by comparing different groups of households using the regression models, and by applying inequality metrics, including the Gini-coefficient and the Lorenz curve, to quantify the inequality in the population.

3.2 Overview Datasets and Data Preparation

Data preparation was needed before the data could be analysed. This included removing irrelevant variables, checking and improving the data quality, transforming the data to analysable variables, and merging the data on a postal code 6 level. For each of the datasets used, the required operations in terms of data preparation are described below.

CBS Kerncijfers Postcode-6

First, the irrelevant data was removed from the dataset. After this, the following variables remained in the dataset: Postal code, age, gender, number of residents, migration background, household composition, number of households, household size, number of dwellings, rental/owned, dwellings in social rent, WOZ-value, gas consumption, electricity consumption, income, and the number of residents with social benefits.

Second, when considering the quality of the data, it should be noted that the data quality is low for a significant part of this dataset, as much privacy sensitive data is included. This means that cells which contain information on less than 5 residents, or less than 5 households are registered as missing values, depending on the aggregation level of the variable. Thus, all postal codes including 5 or less households were removed from the dataset, as the majority of variables in the dataset is missing for these cases. The removed cases account for 21% of the postal codes in the Netherlands, which means that this operation significantly reduced the total number of cases in the dataset. The operation is however justified, as the number of households in a postal code has no relation with the energy consumption and energy affordability, or with one of the predictors of these variables.

Without the removed postal codes, the data quality is relatively good, meaning that for most variables less than 5% of the values is missing. The missing values from these variables are not imputed, as this would lower the data quality. Furthermore, imputation is not needed, as only a small percentage of the cases is missing, and as these cases are missing at random. There is however one exception, which is the percentage of rental and owned homes, for which 24% of the cases contain missing values. These postcode-6 cases are imputed with the percentages of rental and owned homes in their Postcode-5 areas, for which the CBS also has data available. Each postcode-5 area contains 14 postcode-6 areas on average, which are part of the same neighborhood and often lie within the same street. This means that both dwellings and residents are relatively similar within a postcode-5 area, and that taking the percentage of rental and owned homes from the postcode-5 area as an approximation of the percentage in the postcode-6 area can be well justified.

Third, several variables in the dataset were transformed in order to make them suitable for the analyses to be performed. The variables age, gender, household composition, dwellings in social rent and number of residents with social benefits are all listed as an absolute number. However, to make these variables analyzable, the absolute number was transformed to a percentage, based on the total number of residents or dwellings within a postal code.

The variable income is given as a range within which the different incomes of the households with the same postal code fall. The range is marked by percentiles, and covers 20% in most cases (e.g. when the income is between the 40^{th} and 60^{th} percentile). As the income is expected to have a linear relationship with the energy consumption and with the energy affordability, it is however desirable to include the income as a numerical ratio variable. Furthermore, this increases the power of the income measure, and allows for the numerical calculation of affordability. Therefore, the percentiles from this dataset are combined with the average disposable income of the different percentiles from the CBS Inkomen van huishoudens dataset. This allows the calculation of the average income of the percentile range, which can be used as an approximation of the average income of the households in a certain postal code. It should however be noted that this approximation is not the same as the actual average, as the fact that the income of different households falls within a certain range does not mean that the average of the different cases is the same as the average of the range. However, as the dataset contains more than 80% of the households in the Netherlands, it can be stated that although the approximation of the average may differ from the actual average in individual cases, it can be expected to be an accurate measure of the average when analyzing the complete dataset.

BAG-adressen-woning

The BAG-adressen-woning dataset includes data on all addressable objects in the Netherlands, making it a large dataset of 4GB. Thus, to make the dataset more analyzable, all irrelevant variables were removed first. The variables left in the dataset were: status of addressable object, neighboorhood code, municipality, postal code, province, type of dwelling, city, indication of height, number of addresses per building, square meters addressable object, and year of construction. The addressable objects with 'living function = 0' and 'year of construction > 2018' were removed from the dataset, as they don't include current households.

Second, the data quality was checked, which proved to be high, as for all variables less than 1% of the cases is missing. Nevertheless, the numerical missing values were imputed with the average value of their postal code, as this is the most accurate approximation of this value and allows for a more complete analysis without missing cases. The most important reason for the imputation is however that all of this data was later aggregated to postal code data, after which the average is the only measure left of these cases.

Third, the relevant variables were transformed to analysable variables, mainly by aggregating them from an addressable object level to a postal code 6 level. For the numerical variables, including the year of construction, number of addresses per building, square meters addressable object and the indication of height, the average of the values was taken as the postal code value. The status of the addressable object and the status of the building are transformed into the percentage of objects and buildings in use in a certain postal code. For the type of addressable object, and the type of dwelling, multiple variables were created, each including the percentage of a category within a certain postal code (e.g. the % of dwellings and the % of detached houses). For all other variables, including the neighbourhood code, the municipality, the postal code, the province and the city, the values of the addresses within the same postal code are similar. Thus, the most occurring value within a postal code was chosen for these variables when aggregating to a postal code level. After this, the data was merged with the CBS Kerncijfers Postcode-6 dataset.

CBS – Inkomen van huishoudens
This data contains the average disposable income of each 10% income decile within the population. The deciles are equally sized groups, ranked on income, and can be used to quantify the income, as described in the CBS Kerncijfers Postcode-6 paragraph. The data is yearly from 2016 to 2020, which matches the energy consumption data. Thus, the average income of the deciles is merged with the income categories in the CBS Kerncijfers Postcode-6 dataset to compute an approximation of the average income per postal code.

CBS - Aardgas en Electriciteit prijzen
Both for gas and electricity, the transaction prices for households were taken from the CBS – Aardgas en Electriciteit prijzen dataset. Average yearly prices were used for 2016, 2017 and 2018, in order to match the energy consumption data. The price data was combined with the income and the energy consumption to compute the average expenditure-to-income ratio within a postal code.

CBS – Energietarieven
The different price components of the residential gas and electricity price in the Netherlands were taken from the CBS – Energietarieven dataset to calculate the energy expenditure to income ratio for October 2021 and April 2022. These ratios are needed for the additional speculative affordability analysis. The dataset includes the fixed transport costs, the variable transaction costs, and the taxes, both for electricity and gas. It matches the data from the CBS – Aardgas en Electriciteit prijzen dataset to a large extent, although there are some key differences. First, the CBS – Energietarieven dataset makes use of monthly data, which means in this case that it includes price data on the most recent months in 2022. This data is needed, as the energy prices soared during this period, which is interesting from an affordability perspective. Second, this dataset includes the prices on the market, whereas the CBS – Aardgas en Electriciteit prijzen covers the actual prices paid by the households. There is a delay between the prices on the market and the consumer price paid, as households often have yearly energy contracts, based on previous prices. Thus, the speculative affordability calculated with these prices doesn't match the exact timeframe of the prices in this dataset. The data contains prices from January 2018 to April 2022, although only the 2021 and 2022 data is used in the analysis.

In order to calculate the energy expenditure for October 2021 and April 2022, the energy prices were combined with the income and the energy consumption. For this, several assumptions were made, as not all data was available for these timeframes. First, the baseline energy consumption is assumed to be similar to the energy consumption between 2016 and 2018. Second, the CBS – Inkomen van huishoudens dataset was used to determine the income of the households in 2021, which is used in the speculative affordability analysis. As income data was only available until 2020, the 2021 incomes were estimated by extrapolating the yearly income growth from 2016 to 2020. Based on the average growth rate in these years, incomes were increased with 3,56% for 2021. Third, there is the impact of the energy price on the energy demand, which is quantified by the price elasticity. Higher prices decrease the demand on energy, although the exact impact of the recent soaring energy prices remains unclear, because of the magnitude of the price increase, which hasn't been this large after the oil crisis in 1973 (Tesio et al., 2022; Labandeira, Labeaga, & López-Otero, 2017). Thus, 4 different scenarios were analyzed, 2 for the October 2021 energy prices and 2 for the April 2022 energy prices. For both energy prices, one scenario assumes no change in demand, whereas the other does assume a change in demand, based on the price elasticity. The price elasticity used for these analyses is -0.215, which is the average short term price elasticity of residential energy (Labandeira et al., 2017).

CBS – Gebieden in Nederland
The CBS – Gebieden in Nederland dataset includes the urbanity of all municipalities in the Netherlands. It contains 3 variables related to this: the average number of addresses per squared kilometre, 5 categories of urbanity based on the number of addresses per squared kilometer, and urbanity codes from 1 to 5, linked to the previous categories. The data is yearly and was merged with the other datasets on a municipality level.

RVO - Warmtenetten
The RVO – Warmtenetten data is used to remove households with district heating from the dataset. This research only uses gas and energy consumption data to calculate the total energy consumption and the total energy expenditures. Thus, households that also use district heating need to be excluded from the analyses, as the gas and electricity consumption of these households combined doesn't equal their total energy consumption. The data includes data on each neighbourhood that makes use of district heating. For each of these neighbourhoods, the following variables are listed: The municipality, the number of dwellings, the percentage of dwellings with district heating, and the neighbourhood code.

The neighbourhood code is used to merge the data with the CBS – Kerncijfers Postcode-6 dataset. After merging the datasets, the postal codes in neighbourhoods in which more than 25% of the households have district heating were deleted from the final dataset. Postal codes in neighbourhoods with 25% or less households using district heating were kept in the dataset, to prevent the removal of large amounts of postal codes that don't make use of district heating. A more elaborate motivation for this choice is given in appendix A. After removing most households with district heating from the dataset, a dummy variable containing whether a postal code lies in a neighbourhood with district heating was added to the dataset, to be able to distinguish postal codes in a neighbourhood with district heating.

KNMI - Dagwaarnemingen
Data related to weather differences within the Netherlands was retrieved from the KNMI – Dagwaarnemingen dataset. The dataset includes daily data for the different weather stations in the Netherlands, including the average temperature, the minimum temperature and the sunshine duration. Data from 2016 until 2018 was included for all Dutch weather stations active during this period. In order to be able to merge this data, the daily data was aggregated to yearly data. Yearly averages were created to match the yearly energy consumption. Furthermore, the weather stations were linked to municipalities, after which the weather data could be merged with the other datasets based on municipality.

Merging the datasets
All in all, the separate datasets were transformed to yearly data and merged on a postal code 6 level, using python. For some of the datasets with a different data granularity, such as the 'BAG-adressen-woning' dataset, this meant that different cases had to be aggregated first. The two largest datasets, the CBS Kerncijfers postcode-6 and the BAG-adressen-woningen were merged first, after which the variables or computations from the other datasets were added.

3.3 Included Variables and Variable Transformations

After merging the datasets, all variables were combined in a final dataset, which includes a total of 1.196.531 different cases, divided over 2016, 2017 and 2018. Each variable is aggregated on a postal code 6 level, based on yearly data. An overview of all the relevant variables in the final datasets is given in table 4.

Table 4: Overview Included Variables

Variable Name	**Description**	**Original Dataset**	**Original Aggregation level**	**Unit**
GAS_CONSUMPTION_ kWh	The average natural gas consumption of a household within a postal code in kWh	CBS – Kerncijfers Postcode-6	Postal Code 6	(Avg kWh / Year) / Postal Code
ELECTRICITY_ CONSUMPTION	The average electricity consumption of a household within a postal code in kWh	CBS – Kerncijfers Postcode-6	Postal Code 6	(Avg kWh / Year) / Postal Code
TOTAL_ENERGY_ CONSUMPTION	The average total energy consumption, including gas and electricity, of a household within a postal code in kWh	CBS – Kerncijfers Postcode-6	Postal Code 6	(Avg kWh / Year) / Postal Code
ENERGY_EXPENDITURE_ TO_INCOME	The average percentage of household income that is spend on energy consumption within a postal code. The formula for this is: (Energy price*Energy consumption) / income	CBS – Kerncijfers Postcode – 6, Inkomens van huishoudens, Aardgas en Elektriciteit prijzen	Postal Code 6 / Country	(Avg € / Year) / Postal Code
RENT_PR	The percentage of rental homes within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Dwellings / Postal Code
OWNED_PR	The percentage of owner-occupied homes within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Dwellings / Postal Code
SOCIAL_RENT_PR	The percentage of social rent homes within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Dwellings / Postal Code
URBAN_RURAL_SCALE	The urbanity of a certain postal code, displayed on a scale from 1 to 5, with 1 being extremely urban and 5 being rural. The scale is based on the number of residents per squared kilometre within a municipality	CBS – Gebieden in Nederland	Municipality	(# Dwellings / km2) / Municipality 1 = > 500 2 = 500-1000 3 = 1000-1500 4 = 1500-2500 5 = < 2500
YEAR_OF_ CONSTRUCTION	The average year of construction within a postal code	BAG – Adressen-woning	Address	Avg Year / Postal Code
WOZ_VALUE	The average WOZ-value of a dwelling within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	Avg € (x1000) / Postal Code
DISTRICT_HEATING_PR	The percentage of households with district heating within the neighbourhood of a certain postal code	RVO – Warmte-netten	Neighbour-hood	% households / Neighbour-hood

DISTRICT_HEATING	A boolean variable that has the value 1 when a postal code lies within a neighbourhood with district heating	RVO – Warmte-netten	Neighbour-hood	Neighbour-hood
HEIGHT_INDICATION_ BUILDING	An indication of the average height of a building within a postal code. Technically, this is the ratio between the surface area of all dwellings in the building, and the surface area of the building itself.	BAG – Adressen-woning	Address	(Avg m2 / m2) / Postal Code
DWELLINGS_IN_BUILDING	The average number of dwellings in a building within a postal code	BAG – Adressen-woning	Address	(Avg # Dwellings / Building) / Postal Code
SQUAREMETERS_ DWELLING_m2	The average surface area of a dwelling within a postal code	BAG – Adressen-woning	Address	Avg m2 / Postal Code
APARTMENT_PR	The percentage of apartment homes within a postal code	BAG – Adressen-woning	Address	% Dwellings / Postal Code
CORNER_HOME_PR	The percentage of corner homes within a postal code	BAG – Adressen-woning	Address	% Dwellings / Postal Code
TERRACED_HOUSE_PR	The percentage of terraced houses within a postal code	BAG – Adressen-woning	Address	% Dwellings / Postal Code
SEMI_DETACHED_ HOUSE_PR	The percentage of semi-detached houses within a postal code	BAG – Adressen-woning	Address	% Dwellings / Postal Code
DETACHED_HOUSE_PR	The percentage of detached houses within a postal code	BAG – Adressen-woning	Address	% Dwellings / Postal Code
AVG_TEMP	The average daily temperature of the weather station closest to a postal code	KNMI – Dagwaar-nemingen	Weather Station	Avg °C / Postal Code
AVG_MIN_TEMP	The average minimum daily temperature of the weather station closest to a postal code	KNMI – Dagwaar-nemingen	Weather Station	Avg °C / Postal Code
AVG_SUNSHINE _DURATION	The average daily sunshine duration in hours of the weather station closest to a postal code	KNMI – Dagwaar-nemingen	Weather Station	Avg # Hours / Postal Code
MALE_PR	The percentage of male residents within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Residents / Postal Code
FEMALE_PR	The percentage of female residents within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Residents / Postal Code
RESIDENTS_AGE_0014_PR	The percentage of residents within a postal code in the 0-14 age group	CBS – Kerncijfers Postcode-6	Postal Code 6	% Residents / Postal Code
RESIDENTS_AGE_1425_PR	The percentage of residents within a postal code in the 15-24 age group	CBS – Kerncijfers Postcode-6	Postal Code 6	% Residents / Postal Code

RESIDENTS_AGE_2544_PR	The percentage of residents within a postal code in the 25-44 age group	CBS – Kerncijfers Postcode-6	Postal Code 6	% Residents / Postal Code
RESIDENTS_AGE_4564_PR	The percentage of residents within a postal code in the 45-64 age group	CBS – Kerncijfers Postcode-6	Postal Code 6	% Residents / Postal Code
RESIDENTS_AGE_65PL_PR	The percentage of residents within a postal code in the 65+ age group	CBS – Kerncijfers Postcode-6	Postal Code 6	% Residents / Postal Code
DUTCH_BACKGROUND _PR	The percentage of residents with a Dutch ethnic background within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Residents / Postal Code
WESTERN_MIGRATION _PR	The percentage of residents with a western migration background within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Residents / Postal Code
NONWESTERN_ MIGRATION_PR	The percentage of residents with a non-western migration background within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Residents / Postal Code
ONEPERSON_PR	The percentage of one person households within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Households / Postal Code
MULTIPLE_PERSON_ WITHOUT_CHILDREN_PR	The percentage of 'multiple person without children' households within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Households / Postal Code
ONEPARENT_PR	The percentage of one parent households within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Households / Postal Code
TWOPARENTS_PR	The percentage of two parent households within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Households / Postal Code
HOUSEHOLD_SIZE	The average household size within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	# Avg Number / Postal Code
INCOME	An approximation of the average income of a household within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	(Avg € / Year) / Postal Code
SOCIAL_BENEFITS_PR	The percentage of households with social benefits within a postal code	CBS – Kerncijfers Postcode-6	Postal Code 6	% Households / Postal Code
MUNICIPALITY	The municipality connected to a certain postal code	BAG – Adressen-woning	Address	Municipality
YEAR	The year over which the energy consumption, the income and the other variables where measured.	CBS – Kerncijfers Postcode-6	Postal Code 6	Year

Types of Variables in the Dataset

The different variables can be divided in 4 different categories. First, there are the dependent variables: the yearly gas and electricity consumption, which are combined in the total energy consumption, and the energy expenditure to income ratio. Second, the housing-related factors are listed, including the percentage of rented and owned houses and the urbanity level of a postal code. These variables describe characteristics of the dwelling, or aspects related to the location of the dwelling. Third, weather variables are listed, covering the average daily temperature, the average minimum daily temperature, and the average daily hours of sunshine. These variables are used as control variables in the regression. Fourth, there are the socio-demographic variables, which are also used as control variables. Socio-demographic variables include characteristics of households, such as the average household size and income, and characteristics of residents, such as age and ethnic background.

Variables Transformed for the Statistical Analyses

Some variables were transformed to make them more suitable for analysis, because they have a large number of outliers, or because they only have a relationship with energy consumption and energy affordability for a certain range of values. First, for the year of construction, all values below 1900 were altered to 1900, as the relationship between energy consumption and year of construction is based on the thermal quality of a house, which doesn't decrease further for houses build before 1900. Second, for the number of dwellings, all values above 50 were transformed to 50, because the relationship between the number of dwellings in a building and the energy consumption doesn't increase significantly for values above 50. Third, for the height indication of the building, all values above 10 were transformed to 10. On the one hand this was done because the energy consumption doesn't differ between buildings with a height indication of 10 and buildings with a higher height indication. On the other hand however, there were also some outliers with a large impact on the correlations with energy consumption and energy affordability, which made this transformation necessary. Fourth, for the surface area in square meters, all values above 400 were altered to 400. Above this value, a higher number of square meters no longer leads to a higher energy consumption. Next to that, outliers do have a high impact as well, which justifies this operation. A more elaborate explanation and justification of the transformations described above is given in Appendix A.

Furthermore, the year variable was transformed into 2 dummy variables, to be able to check if there are significant differences between the included years, even when controlling for other factors such as temperature. Dummies were created for 2017 and 2018, meaning that 2016 serves as the reference category. Also, the surface area in square meters was transformed into a logarithmic variable for the regression analyses, as both figure 15 and the literature reveal that its relationship with the energy consumption is logarithmic (Brounen et al., 2012; Santin et al., 2009; Iraganaboina & Eluru, 2021).

Variables Excluded from the Regression

For the regression, several variables were excluded because of multicollinearity. First, the percentage of owned houses was excluded because it contained the same information as the percentage of rental houses. Second, a principal component analysis was performed on the different housing types, as multicollinearity was high for these variables, meaning that removing one of the house types from the regression was not enough. The goal of the principal component analysis was not to replace the original variables with a component score, but to determine which of the variables explain the most variance and should be left in the model. The analysis, which is discussed more elaborately in appendix B, showed that the percentage of terraced houses, the percentage of apartments and the percentage of semi-detached houses explain the most variance. In the end however, the percentage of semi-detached houses was replaced by the percentage of detached houses, as this improved the interpretability of the model, and did not reduce the explained variance. The interpretability improved because the percentage of detached houses is the highest energy consuming house type, because it has the strongest correlation with the energy consumption and the energy expenditure to income ratio when compared to other house types, and because there are significantly more detached houses than semi-detached houses. Thus, the percentage of semi-detached houses, and the percentage of corner homes were excluded from the regression analyses.

Third, when considering the control variables, the percentage of residents from 0 to 14 years, the percentage of male residents, the percentage of one parent households and the percentage of residents with a Dutch background are left out of the analyses. Together with other percentage groups, these variables form 100%, causing multicollinearity as all the other percentage groups can be used to explain the same information, which is why they are excluded from the regression. These variables can however still be interpreted, as they serve as a reference category for the other variables in the same category, e.g. when the percentage of western and non-western immigrants is known and multiplied with the regression coefficients, the percentage of residents with a Dutch background can be filled in. None of the other variables in the dataset showed unacceptable multicollinearity scores, meaning that all other variables were included in the model.

Missing Values

Some variables, including the total energy consumption and the energy expenditure to income ratio, have a significant number of missing values, as shown in table 5. This is due to data unavailability, or because of privacy reasons. The variable income for example is only given in case the income of at least 10 households is available. The regression analyses make use of pairwise inclusion, meaning that missing values are kept out of the analysis, but that the cases containing these missing values are kept in the analysis, as values are present on other variables. In such a case where missing values are present, only a subset of the predictor variables is used for estimating the relation with the dependent variable. The regression coefficient is thus based on all the correlations available for each pair of variables. The advantage of pairwise inclusion is that all the available data is used, which increases the number of cases included, and therefore improves the validity of the model. Furthermore, removing cases with missing values, which is the case for listwise inclusion, can create a bias in the data, as the missing values may not be missing completely at random (Peugh & Enders, 2004; IBM, 2020). Thus, pairwise inclusion is the preferred method for dealing with missing values in the regression analyses. The dependent variables, the energy consumption and the energy expenditure, can however not be missing. Thus, as the number of valid cases for the energy expenditure to income ratio is lower than the number of cases for the total energy consumption, a different sample is used for these two regression analyses. This can be justified by the fact that the energy consumption of the missing energy expenditure values follows a random distribution, which is visualised in Appendix A.

3.4 Data Description

Table 5 displays the basic descriptive statistics of the variables included in the analyses. When combining the 2016, 2017 and 2018 data, a total of 1.196.531 postal codes is included in the analyses, which corresponds with 24.154.970 households. This means that on average 8.051.657 households are included per year.

Socio-demographics, Energy Consumption and Energy Affordability

The socio-demographic variables of the dataset match the socio-demographics of Dutch population, which confirms the quality of the data. Variables such as the percentage of men, the percentage of people with a Dutch background, and the percentage of one person households have a similar mean as the Dutch population (CBS, 2022). When looking at the energy consumption data, it can be stated that the total energy consumption is for a large part dependent on the gas consumption, as this is on average more than 4 times higher than the electricity consumption. The average total energy consumption of Dutch households is 17.302 kWh per year, when combining gas and electricity. The standard deviation is 6275 kWh, meaning that about 68% of the values fall between 11.000 and 23.600 kWh per year. The median, minimum and maximum of the total energy consumption suggest that it follows a relatively normal distribution, with a tail towards high energy consumption values. The energy expenditure to income ratio follows a similar distribution, and has a mean of 6,16, meaning that Dutch households spend on average 6,16% of their income on energy. The standard deviation of this variable is 2,04, which means that 68% of the values fall between 4,12% and 8,20% per year.

Table 5: Descriptive Statistics Included Variables

Descriptive Statistics

	N					
	Valid	Mean	Median	Std. Deviation	Minimum	Maximum
GAS_CONSUMPTION_kWh	1151716	14510,57	13677,00	5435,98	,00	74831,00
ELECTRICITY_CONSUMPTION	1162548	3014,03	2940,00	937,32	,00	9820,00
TOTAL_ENERGY_CONSUMPTION	1168453	17301,52	16547,00	6275,21	98,00	83091,00
ENERGY_EXPENDITURE_TO_INCOME	1082969	6,16	5,96	2,04	,04	38,99
RENT_PR	1195020	33,41	20,00	30,02	,00	100,00
OWNED_PR	1195020	66,39	80,00	30,16	,00	100,00
SOCIAL_RENT_PR	1186796	14,21	,00	30,15	,00	100,00
URBAN_RURAL_SCALE	1186872	2,89	3,00	1,34	1,00	5,00
WOZ_VALUE	1110855	224,18	196,00	124,07	11,00	4392,00
YEAR_OF_CONSTRUCTION	1195219	1965,53	1969,53	25,71	1900,00	2018,00
HEIGHT_INDICATION_BUILDING	1195219	2,08	1,82	1,05	,09	10,00
DWELLINGS_IN_BUILDING	1195219	5,57	1,00	11,82	1,00	50,00
SQUAREMETERS_DWELLING_m2	1196531	129,66	117,95	55,67	3,94	400,00
APARTMENT_PR	1196531	24,84	,00	39,13	,00	100,00
CORNER_HOME_PR	1196531	14,55	11,10	15,26	,00	100,00
TERRACED_HOUSE_PR	1196531	30,71	20,80	31,18	,00	100,00
SEMI_DETACHED_HOUSE_PR	1196531	11,65	,00	21,34	,00	100,00
DETACHED_HOUSE_PR	1196531	17,98	,00	30,57	,00	100,00
DISTRICT_HEATING_PR	46945	13,54	13,00	5,82	5,00	25,00
DISTRICT_HEATING	1196531	,04	,00	,19	,00	1,00
AVG_TEMP	1186872	10,96	10,99	,45	9,97	11,87
AVG_MIN_TEMP	1186872	6,52	6,46	,61	5,51	9,24
AVG_SUNSHINE_DURATION	1165401	5,17	5,03	,43	4,34	5,91
INCOME	1101461	27,47	28,68	9,38	11,70	55,25
SOCIAL_BENEFITS_PR	1194947	5,72	,00	10,43	,00	100,00
MALE_PR	1194947	49,55	50,00	9,62	,00	100,00
FEMALE_PR	1194947	50,41	50,00	9,54	,00	100,00
RESIDENTS_AGE_0014_PR	1194947	11,95	13,00	11,69	,00	77,00
RESIDENTS_AGE_1524_PR	1194947	7,83	,00	10,61	,00	100,00
RESIDENTS_AGE_2544_PR	1194947	21,67	21,00	16,11	,00	100,00
RESIDENTS_AGE_4564_PR	1194947	27,58	29,00	14,49	,00	100,00
RESIDENTS_AGE_65PL_PR	1194947	17,93	14,00	20,33	,00	100,00
DUTCH_BACKGROUND_PR	1196531	80,08	90,00	20,21	,00	100,00
WESTERN_MIGRATION_PR	1196531	6,83	,00	10,24	,00	100,00
NONWESTERN_MIGRATION_PR	1196531	8,83	,00	16,58	,00	100,00
ONEPERSON_PR	1196531	25,56	20,00	28,42	,00	100,00
MULTIPLE_PERSON_WITHOUT_CHILDREN_PR	1196531	21,40	20,00	21,85	,00	100,00
ONEPARENT_PR	1196531	1,38	,00	6,14	,00	100,00
TWOPARENTS_PR	1196531	19,79	,00	23,39	,00	100,00
HOUSEHOLD_SIZE	1196531	2,25	2,30	,58	1,00	6,30

Housing-related Factors

The housing-related factors have more diverse distributions. The percentage of rented houses for example has a mean of 33,41 and a median of 20, suggesting that most postal codes have a low percentage of rental homes, with more than half of the values between 0 and 20%. The mean of social rent is 14,21, meaning that of the 33,41% of rental homes, 14,21 percentage point is social rent. The urbanity scale has a less skewed distribution, with a mean of 2,89 and a median of 3,00. This reveals that the average postal code lies in a semi-urban environment with 1000-1500 dwellings per squared kilometre.

Two typical distributions are occurring relatively often among housing-related factors. First, there is the skewed distribution with a long tail in a certain direction, which can be seen in variables such as the WOZ-value, the height indication and the number of squared meters. These variables typically have a mean slightly higher than the median, and high maximum values that lie relatively far from the mean in terms of standard deviation. Second, there are the percentual distributions, such as the percentage of apartments or the percentage of dwellings with social rent. These distributions have values between 0 and 100, and have a low mean between 10% and 35%, except for the percentage of owned houses. The median is lower than the mean, and often 0, meaning that the different categories of houses aren't present in all postal codes, but are grouped in postal codes with a high percentage of houses from a certain category.

Remarkable Observations

Furthermore, some of the statistics need a more elaborate description. First, the income has a maximum value of 55,25. This may seem low, but is due to the fact that the after-tax income is measured per decile. Thus, this means that postal codes with households in the highest income decile have an average after-tax income of €55.250 per year. Therefore, there are postal codes with a higher average income, but the income given in the data is the average of the households in the highest income group. Another interesting observation is the fact that the minimum temperature has a higher standard deviation than the average temperature, showing the difference in temperature variation between coastal and more inland postal codes. As the differences in the minimum temperature are larger, this variable is expected to have a higher correlation with the energy consumption, when compared to the average temperature. The district heating has an average value of 0,04, meaning that 4% of the postal codes in the dataset have a percentage of houses in their neighbourhood with district heating. The DISTRICT_HEATING_PR variable shows that 46945 postal codes fall in this category, and that on average 13.54% of the dwellings in these postal codes have access to district heating. These households are not completely representative for the analysis, as district heating is not included in the total energy consumption, which is accounted for in the regression.

4. Descriptive Statistics: Energy Consumption, Energy Affordability and Housing-related Factors in the Netherlands

4.1 Energy Consumption

This paragraph presents an overview of the residential energy consumption in the Netherlands. Before diving into the impact that the housing-related factors have on energy consumption, it is important to have a good understanding of the phenomena itself, and how it is distributed in the Dutch society. Thus, this paragraph provides answers to sub-question 2: How is the energy consumption distributed among households in the Netherlands?

Energy Consumption Distribution

Figure 7 reveals that the energy consumption follows a relatively normal distribution, which starts at 0 kWh/year, and has a tail towards high values. The histogram shows that the spread in energy consumption is high, with more than 100.000 households per 1.000 kWh bar in the range from 7.000 kWh/ Year to 26.000 kWh/year. This means that a large number of postal codes in the population consumes more than 3 times as much energy as the low-consuming postal codes in the Netherlands.

The number of households consuming less than 6.000 kWh per year is relatively small, and the energy consumption of this group differs from the rest of the population. The percentages of gas and electricity in the total energy consumption in figure 7 disclose that these households consume mainly electricity and have a very low gas consumption. This suggests that these households make use of district heating, or use alternative fuels for heating, such as firewood. Furthermore, the figure reveals that also for households that consume more than 6.000 kWh per year, the percentage of gas consumption in their total energy consumption keeps increasing with a higher total energy consumption. As gas is mainly used for heating, this could be due to the fact that households with a higher energy consumption have larger houses.

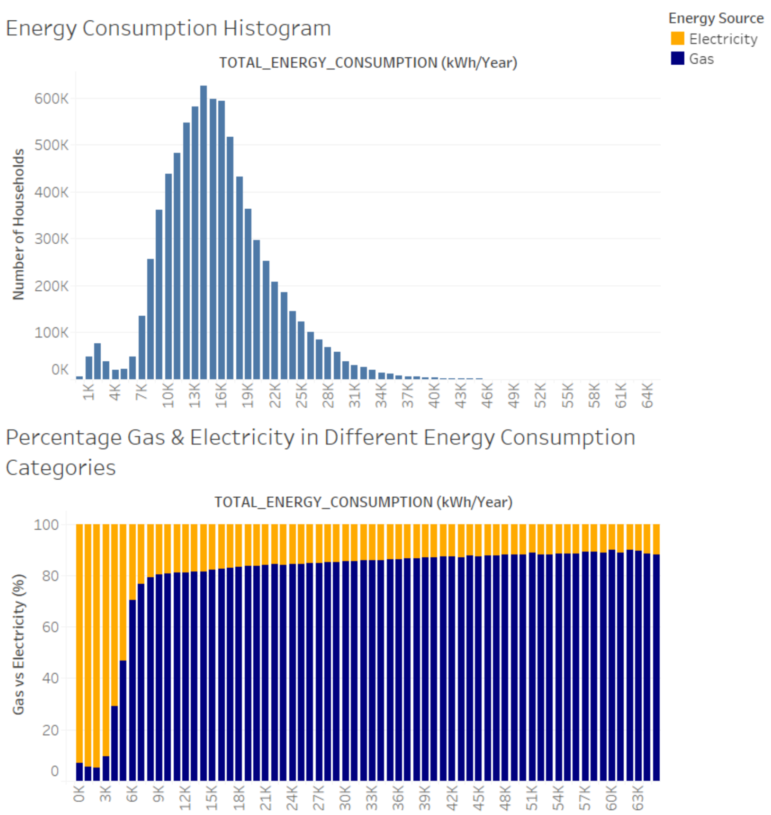

Figure 7: Energy Consumption Histogram

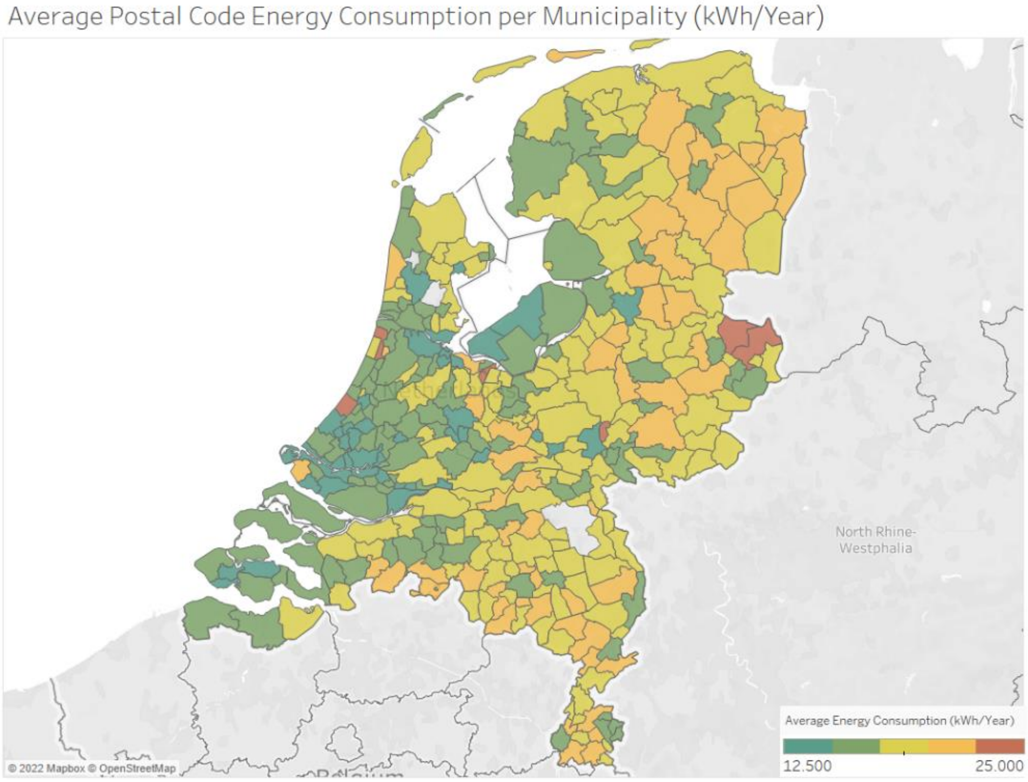

Figure 8: Energy Consumption Geographical Distribution

Geographical Differences in Energy Consumption

The geographical distribution of the average energy consumption per postal code is visualised in figure 8. It shows that differences in energy consumption are significant within the Netherlands, with households in some municipalities using twice as much as households in other municipalities. In general, it can be stated that the eastern part of the Netherlands consumes more energy, and that cities have a low average household energy consumption per postal code. Amsterdam, Rotterdam, the Hague and Utrecht all have an average energy consumption close to 12.500 kWh per year, whereas the eastern and more rural part of the Netherlands has a large number of municipalities with an average energy consumption above 20.000 kWh per year. It is unclear what the reasons are for these differences in energy consumption, whether it is related to the percentage of rental houses, the temperature, the type of dwelling or something else. The regression analyses in the nest chapter provide insight in these relationships.

It is interesting to note that there are some red coloured municipalities in the west of the country, that have a very high average energy consumption per postal code, while being surrounded by municipalities with a lower energy consumption. These municipalities, such as Wassenaar and Bloemendaal, are known as high income municipalities in the Netherlands which could explain this high energy consumption.

Inequality in Energy Consumption

Figure 9 shows the Lorenz curve and the Gini coefficient of the average energy consumption per postal code in the Netherlands. In general, it can be stated that a Gini coefficient of 0,200 is relatively low, meaning that the inequality in the energy consumption is limited. It should however be mentioned here that this Gini coefficient considers the postal code, and not the household level. As the energy consumption also differs within a postal code, the Gini coefficient on a household level would be considerably higher. The Lorenz curve is almost symmetrical, meaning that the differences in energy consumption cannot be linked to a specific group, and that the inequality in energy consumption follows a relatively normal distribution

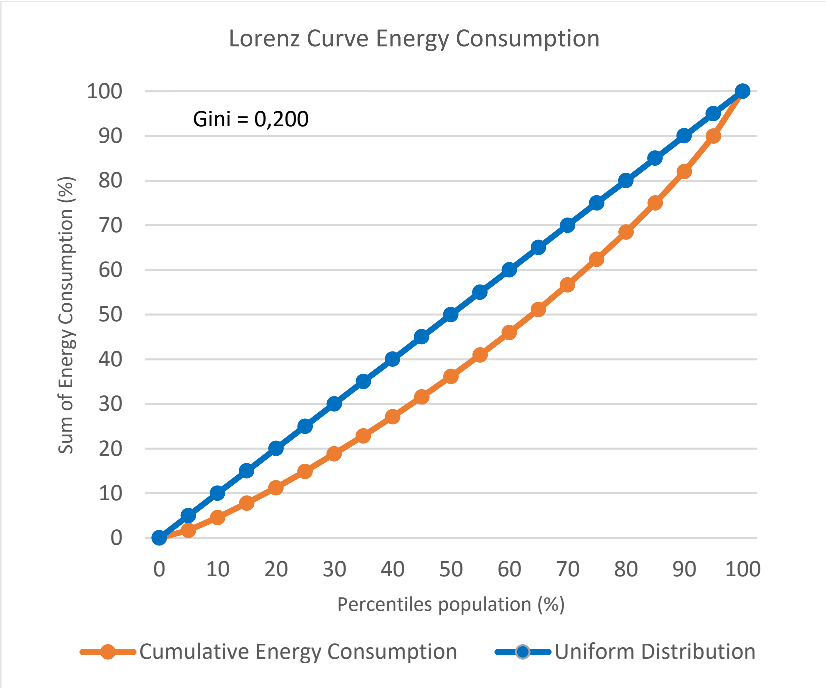

Figure 9: Lorenz Curve Energy Consumption

4.2 Energy Affordability

This paragraph provides insight in the energy affordability in the Netherlands, by showing how the energy expenditure to income ratio is distributed. This gives context to the relationship between the housing-related factors and the energy affordability, which is one of the focal points of this study. Thus, this paragraph presents answers to sub-question 3: How is the energy affordability distributed among households in the Netherlands? Furthermore, the impact of the current energy prices on the energy affordability distribution is also shown, thereby answering sub-question 4: How do the current energy prices influence the residential energy affordability in the Netherlands?

Energy Affordability Distribution

Figure 10 reveals the distribution of the energy expenditure to income ratio among households in the Netherlands. The distribution is relatively normal, with a small tail towards high ratios. The differences are large within the population; each bar, which is responsible for a 0,5% range, between 3% and 9% includes more than 200.000 households. Only a small number of households belong to a postal code that has an average energy expenditure to income ratio above 10%, which is considered the energy poverty line (Lin & Wang, 2020; Haffner & Boumeester, 2015). It should however be noted here, that the histogram shows the average energy expenditures per postal code, meaning that many more households are expected to fall above the energy poverty line. Postal codes with an average energy expenditure of 8% or 9% are for example likely to include a large percentage of households that do spend more than 10% of their income on energy.

The percentages of gas and electricity costs in energy expenditures in figure 10 follow a distribution that is relatively similar to that of the energy consumption in figure 7. This means that on average, households with a higher energy consumption, also have a higher energy expenditure to income ratio. Next to this, the comparison with energy consumption also reveals that the percentage of electricity expenditures is higher than the percentage of electricity consumption, which is due to the fact that electricity is more expensive than gas per kWh.

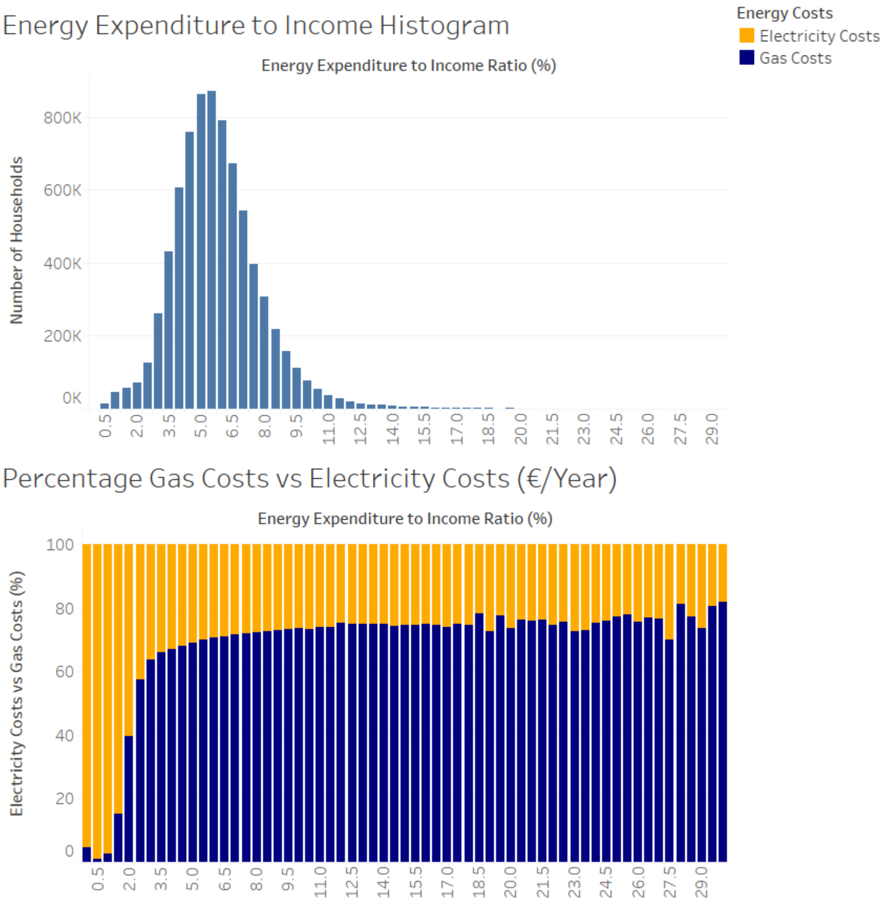

Figure 10: Energy Expenditures Histogram

Geographical Difference in the Energy Affordability

The geographical distribution of the energy expenditure to income ratio is visualised in figure 11, revealing that also on a municipality level, differences remain significant with averages between 4 and 9%. The geographical differences in energy expenditure are even more clear than those in energy consumption. The eastern and especially northern part of the Netherlands has a relatively high average energy expenditure, up to 9%. Municipalities in the Randstad on the other hand, the western part of the country surrounding the large cities, have a much lower average energy expenditure, with values between 4% and 6%. Therefore, it can be concluded that the average income in the Randstad is significantly higher than the average income in the north-eastern part of the country. When combined with a higher energy consumption, this results in a high energy expenditure to income ratio in the north-east of the Netherlands.

Furthermore, cities have a lower average energy expenditure than more rural municipalities, although this difference is smaller than the difference in energy consumption. Also, the municipalities with a high total energy consumption in the west of the country do not have a high energy expenditure to income ratio. This confirms that these municipalities have a high average income, which explains their higher energy consumption.

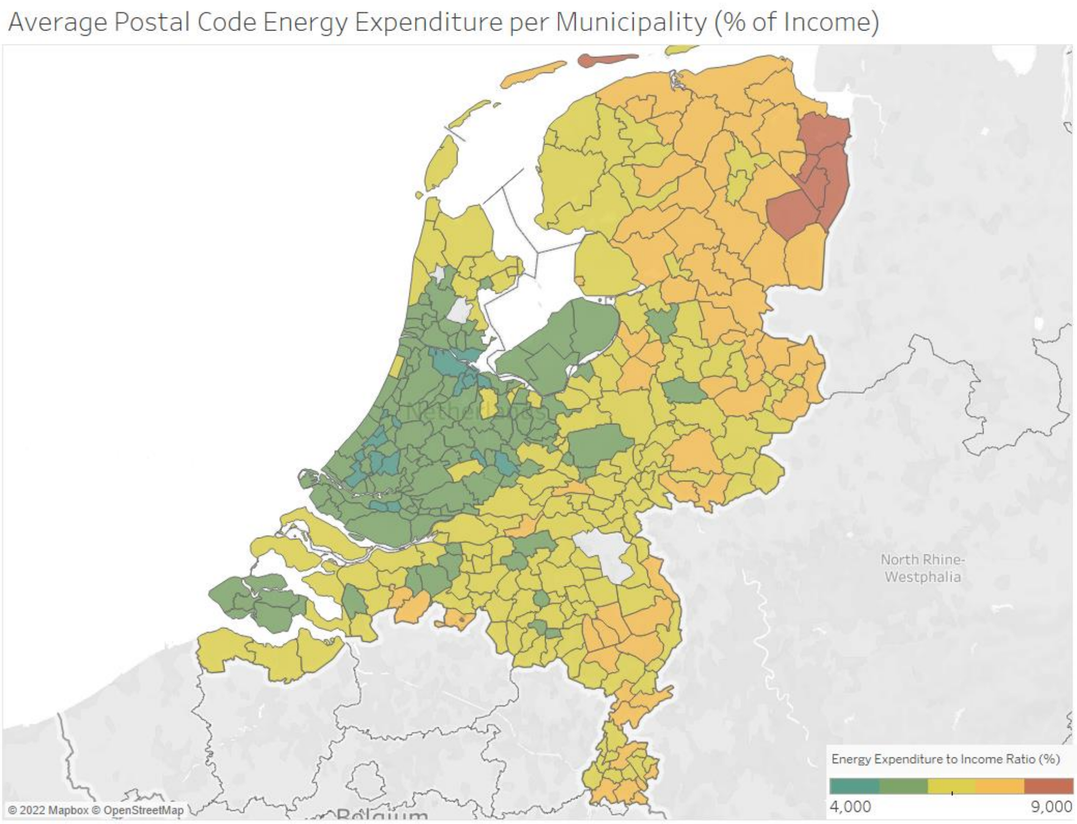

Figure 11: Energy Expenditure Geographical Distribution

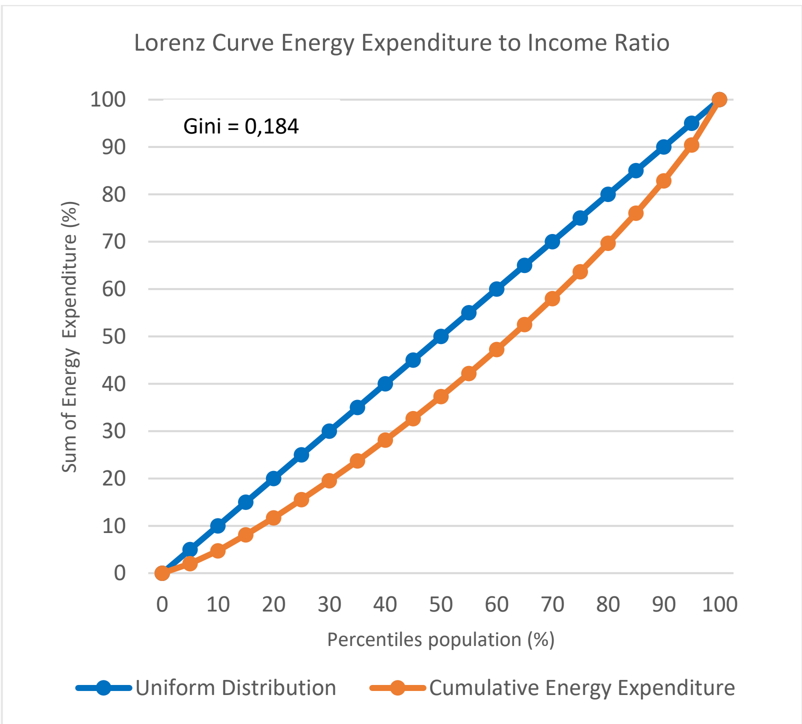

Figure 12: Lorenz Curve Expenditure to Income Ratio

Inequality in Energy Affordability

Figure 12 presents the inequality in the average energy expenditure to income ratio in the Netherlands by showing the Gini coefficient and the Lorenz curve. The Gini coefficient is 0,184, which suggests that the inequality in the average expenditure to income ratio per postal code is low in the Netherlands. The inequality is also lower than the inequality in energy consumption, which suggests that the energy consumption has a positive correlation with the income. The Gini coefficient is however based on the average expenditure to income ratio per postal code, which means that the inequality on a household level is larger. On top of that, the income is given per decile, and also averaged over the postal code, which decreases the inequality even more. The Lorenz curve is relatively symmetrical, meaning that the inequality in the energy expenditure to income ratio follows a relatively normal distribution.

Energy Affordability 2021-2022

Finally, the expenditure to income ratio was analysed for October 2021 and April 2022, to reveal how the rising energy prices impact the energy affordability. The 4 scenarios described in paragraph 3.2 are visualised in table 6 and figure 13. They confirm that the recent soaring energy prices increase the expenditure to income ratio. The expenditure to income ratio has the highest average in April 2022, followed by October 2021, and finally by the original energy expenditure to income ratio based on the 2016-2018 data. As expected, the scenarios that take into account the price elasticity, and thus a decrease in energy consumption, have a lower expenditure to income ratio than the scenarios that don't assume a decrease in energy consumption. The used price elasticity is -0,215, as mentioned in paragraph 3.2. It is interesting to note that the minimum values of the energy expenditure are negative in most scenarios. This is due to a fixed electricity subsidy for necessary minimal electricity consumption, which results in some households getting money from the Dutch government, as their energy consumption is below a certain value.

Table 6: Energy expenditure to Income Ratios October 2021 & April 2022 (PE = Price Elasticity Included)

Descriptive Statistics Additional Affordability Analysis based on Current Prices

	Mean	Median	Std. Deviation	Minimum	Maximum
Original Energy Expenditure to Income Ratio 2016-2018	6,16	5,96	2,04	,04	38,99
Energy expenditure to Income Ratio October 2021	7,69	7,46	2,50	-,57	46,04
Energy expenditure to Income Ratio October 2021 with PE	7,31	7,09	2,38	-1,53	44,12
Energy expenditure to Income Ratio April 2022	14,63	14,16	4,95	-1,39	92,57
Energy expenditure to Income Ratio April 2022 with PE	11,32	10,96	3,79	-2,69	70,93

The prices from October 2021 increase both the average and the standard deviation of the expenditure to income ratio. This means that on average people spend a larger percentage of their income on energy, but also that the inequality between different households increases. There is not much difference here between the scenario with the price elasticity and the scenario without the price elasticity, as the prices are on average only 20% higher than those in the 2016-2018 analysis. Nevertheless, the October 2021 histograms in figure 13 reveal that the number of orange and red coloured households in a postal code with an average energy expenditure above 10%, which is considered the energy poverty line, has increased tremendously, when compared to the original energy expenditure histogram in figure 10.

When considering the energy expenditure based on the prices of April 2022, a distinction has to be made between the scenario that corrects for fewer demand using the price elasticity, and the scenario that doesn't correct for fewer demand. The April 2022 with price elasticity scenario shows that the average energy expenditure almost doubles in this scenario, even when assuming that the demand decreases with 30% because of the 138% price increase between 2018 and April 2022. The standard deviation also increases to 3,79%, meaning that the inequality in society has increased even more in this scenario.

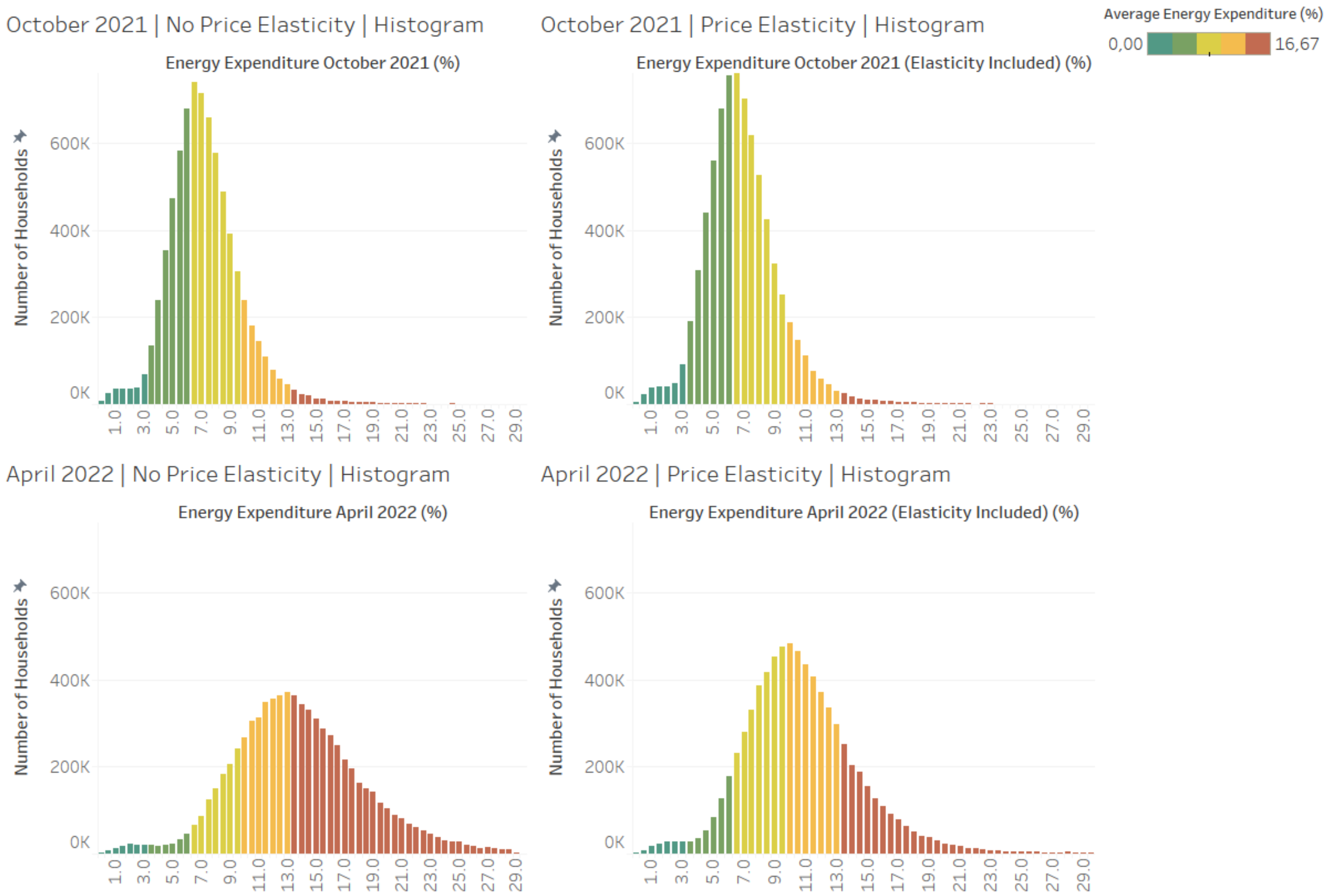

Figure 13: Histograms Energy Expenditure October 2021 & April 2022

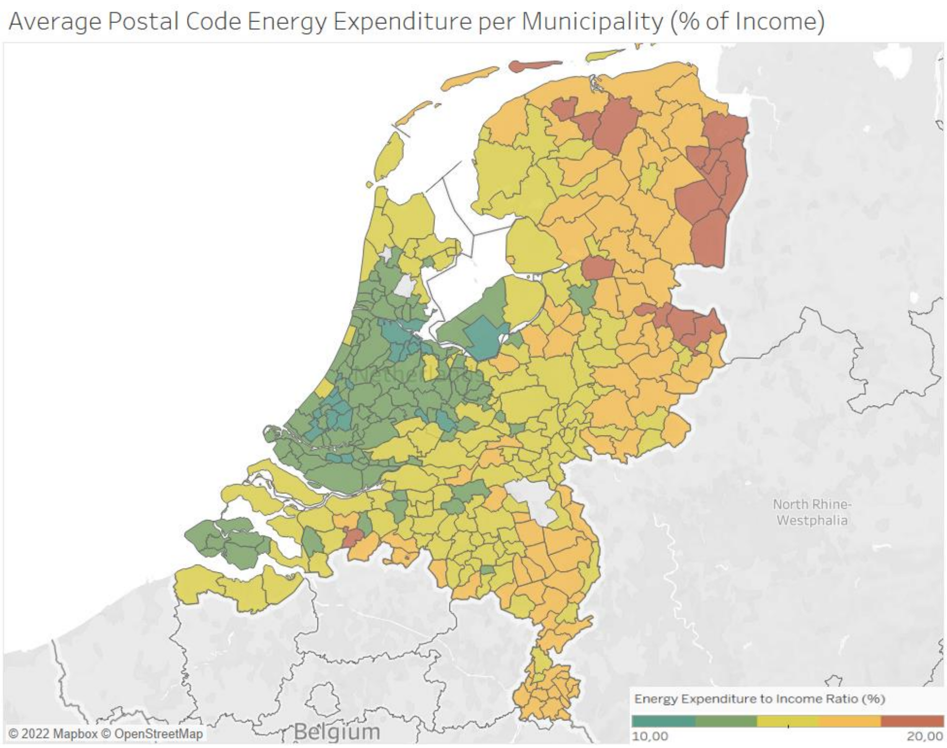

Figure 14: Energy Expenditure April 2022 Geographical Distribution (No Price Elasticity)

However, when households don't consume less because of this price increase, the expenditure to income ratio is even higher, as shown in the April 2022 without price elasticity scenario. The average energy expenditure to income ratio increases to 14,63%, and the histogram reveals approximately 75% of the households in the Netherlands spends more than 10% of their income on energy in this scenario. Thus, the majority of Dutch households would be considered energy poor in this scenario (Lin & Wang, 2020; Haffner & Boumeester, 2015), which is not unrealistic given the energy prices of April 2022.

Figure 14 presents the geographical distribution of the energy expenditure to income ratio in this scenario, based on April 2022 prices. It reveals that the differences in the energy expenditure to income ratio are larger between municipalities, ranging from 10% to 20%. All municipalities have an average ratio above 10% in this scenario, but municipalities in the northeast of the country have a ratio up to 20%, which suggests that energy is unaffordable for a large percentage of these households. Thus, the recent energy expenditure to income ratios reveal that affordability is expected to be a problem in the Netherlands in 2022, which increases the relevance of its relation with the housing-related factors.

4.3 Correlations Socio-Demographic and Weather Factors

In order to get a more in-depth overview of the energy consumption and the energy affordability, the correlations of the average total energy consumption and the average expenditure to income ratio per postal code with the control variables were analysed. Table 7 shows the correlations of the socio-demographic and weather variables with the energy consumption and energy expenditure, thereby answering sub-questions 2 and 3.

Correlations with Energy Consumption

The energy consumption has a strong correlation with the household size and the income of a household. Households with a higher income and a higher household size have a higher consumption, which was expected based on the literature (Wei et al., 2014; Huebner et al., 2016). The ethnic background and the percentage of one person households have a high correlation too. Households which a Dutch background consume more energy, while households with a non-western migration background and one-person households consume less energy.

Households with a large percentage of 25-44 year-old residents consume less energy on average, while households with a large percentage of 45-65 year-old residents consume more energy. The correlation of the percentage of residents with social benefits is also significant; postal codes with a large percentage of residents with social benefits have a lower average energy consumption. Next to this, it is interesting to observe that the minimum temperature has a much stronger correlation with energy consumption than the average temperature, suggesting that cold nights have a relatively large impact on the energy consumption.

Correlations with Energy Affordability

When considering energy affordability, it can be stated that the correlations with the socio-demographic and weather factors are significantly lower on average. This could be expected, because the income has a positive correlation with the energy consumption, and is also known to be correlated with other socio-demographic factors (Riva et al., 2021; Ismael & Khembo, 2015).

Table 7 reveals that the ethnic background, the percentage of one person households and the household size have a low correlation with the energy expenditure to income ratio, while they did have a high correlation with the energy consumption, suggesting a significant correlation with income. The percentage of people with social benefits even switches to a negative correlation, although it had a strong positive relation with the energy consumption, meaning that the lower energy consumption of this group can be explained by their low income. The income itself switches from a strong positive to a strong negative correlation, which can be explained by the fact that the income is divided by the energy consumption in order to calculate the energy expenditure to income ratio. Besides income, the energy expenditure has the highest correlation with the minimum temperature, which is interesting as this correlation is stronger than the correlation of the energy consumption with the minimum temperature. The correlation with the minimum temperature is negative, meaning that a higher minimum temperature results in a lower energy expenditure to income ratio and a higher affordability.

Table 7: Correlations Demographics & Weather Variables

Correlations

	TOTAL_ENERGY_CONSUMPTION	ENERGY_EXPENDITURE_TO_INCOME
MALE_PR	,106	,020
FEMALE_PR	-,084	-,015
RESIDENTS_AGE_0014_PR	,065	,016
RESIDENTS_AGE_1524_PR	,064	,150
RESIDENTS_AGE_2544_PR	-,323	-,151
RESIDENTS_AGE_4564_PR	,241	,005
RESIDENTS_AGE_65PL_PR	,006	,009
DUTCH_BACKGROUND_PR	,354	-,009
WESTERN_MIGRATION_PR	-,160	-,050
NONWESTERN_MIGRATION_PR	-,319	,044
ONEPERSON_PR	-,472	-,036
MULTIPLE_PERSON_WITHOUT_CHILDREN_PR	,183	-,054
ONEPARENT_PR	-,141	,075
TWOPARENTS_PR	,223	-,072
HOUSEHOLD_SIZE	,418	,036
INCOME	,501	-,423
SOCIAL_BENEFITS_PR	-,317	,151
AVG_TEMP	-,049	-,117
AVG_MIN_TEMP	-,169	-,219
AVG_SUNSHINE_DURATION	,067	,057
TOTAL_ENERGY_CONSUMPTION	1	,498
ENERGY_EXPENDITURE_TO_INCOME	,498	1

4.4 Correlations Housing-related Factors

This paragraph discusses the relation between the housing-related factors and the energy consumption and energy affordability. The correlations of the housing-related factors are discussed, and the factors with a strong correlation with the energy consumption or the energy expenditure to income ratio are elaborated on. Thus, this paragraph answers sub-question 5: How do housing-related factors correlate with the Dutch residential energy consumption and energy affordability?

Table 8 shows the correlations of the housing-related factors with the total energy consumption and the energy expenditure to income ratio. Overall, it can be stated that the correlations with the housing-related factors are higher than the correlations with the socio-demographics and the weather variables. This is interesting because the socio-demographics causally determine the housing-related factors and are less flexible (ŠTreimikienė, 2014; Jones & Lomas, 2015). On average the energy expenditure to income ratio has again a lower correlation when compared to the energy consumption, as the income correlates with the energy consumption, and with the housing related factors.

Correlations with Energy Consumption

The energy consumption has a high correlation with the percentage of rental and owner-occupied dwellings. Owner-occupied houses consume more energy than rental houses. The urban-rural scale has a lower correlation with the energy consumption, but still shows a significant correlation of 0,298. This means that rural households consume more energy than urban households in the Netherlands. The surface area of the dwelling (SQUAREMETERS_DWELLING_m2) and the percentage of detached houses in a postal code do however have the highest correlation with the energy consumption. The average energy consumption of a postal code is much higher if it has a large average surface area and if it has a large percentage of detached houses. The average WOZ-value of a house also has a high correlation. The percentage of apartments has a strong negative correlation, meaning that postal codes with a large percentage of apartments consume less energy. The number of dwellings in a building and the height indication of a building also have a significant negative correlation and thus a negative relation with the energy consumption.

Correlations with Energy Affordability

The energy expenditure to income ratio has a much lower correlation with the percentage of owned and rental houses. The correlations of these variables have even switched sign when compared to the energy consumption, suggesting that income is explaining a large part the difference between rental and owned homes. The urban-rural scale still has a significant correlation, meaning that rural households have a higher energy expenditure to income ratio than urban households. The strongest correlations of the energy expenditure to income ratio are with the percentage of detached houses and the average year of construction per postal code. Although the correlation is smaller than with energy consumption, households in detached houses do spend a higher percentage of their income on energy when compared to other households. Postal codes with a higher average year of construction have a lower expenditure to income ratio, which could be due to the fact that insulation is worse for older houses. It is interesting to note that the correlation with year of construction is stronger for the energy expenditure than for the energy consumption, which suggests that the income has a negative correlation with year of construction. The WOZ-value and the surface area have a much lower correlation with the energy expenditure than with the energy consumption, which could be because these factors have a positive correlation with the household income.

Table 8: Correlations Housing-related Variables

Correlations

	TOTAL_ENERGY_CONSUMPTION	ENERGY_EXPENDITURE_TO_INCOME
WOZ_VALUE	,571	-,071
YEAR_OF_CONSTRUCTION	-,255	-,280
HEIGHT_INDICATION_BUILDING	-,394	-,221
DWELLINGS_IN_BUILDING	-,404	-,219
SQUAREMETERS_DWELLING_m2	,724	,217
APARTMENT_PR	-,492	-,215
CORNER_HOME_PR	-,068	,032
TERRACED_HOUSE_PR	-,157	-,061
SEMI_DETACHED_HOUSE_PR	,280	,128
DETACHED_HOUSE_PR	,630	,247
OWNED_PR	,483	-,114
RENT_PR	-,484	,112
SOCIAL_RENT_PR	-,206	,053
URBAN_RURAL_SCALE	,298	,230
DISTRICT_HEATING	-,129	-,073
TOTAL_ENERGY_CONSUMPTION	1	,498
ENERGY_EXPENDITURE_TO_INCOME	,498	1

More in-depth visualisations were created for the highest correlating housing-related factors, the percentage of rental houses and the urban-rural scale, in order to get more insight in the relation that these variables have with the energy consumption and the energy expenditure to income ratio. The highest correlating housing-related factors are the surface area, the percentage of detached houses and the year of construction, which have the two strongest correlations with the energy consumption and the energy expenditure to income ratio.

Distribution Surface Area

Figure 15 describes the surface area distribution and its relation with the energy consumption and the energy affordability. The histogram reveals that the distribution is right-skewed. The relationship with the energy consumption is strong, as the correlation showed before, and follows a logarithmic curve. While an average surface area of 40 m2 results in an average energy consumption below 10.000 kWh per year, an average surface area of 160 m2 has an average energy consumption above 20.000 kWh per year. The relationship between the surface area and the energy consumption flattens out with values higher than 220 m2, although the number of cases with a higher surface area is also limited. The relationship between the surface area and the energy expenditure to income ratio is positive, but relatively weak.

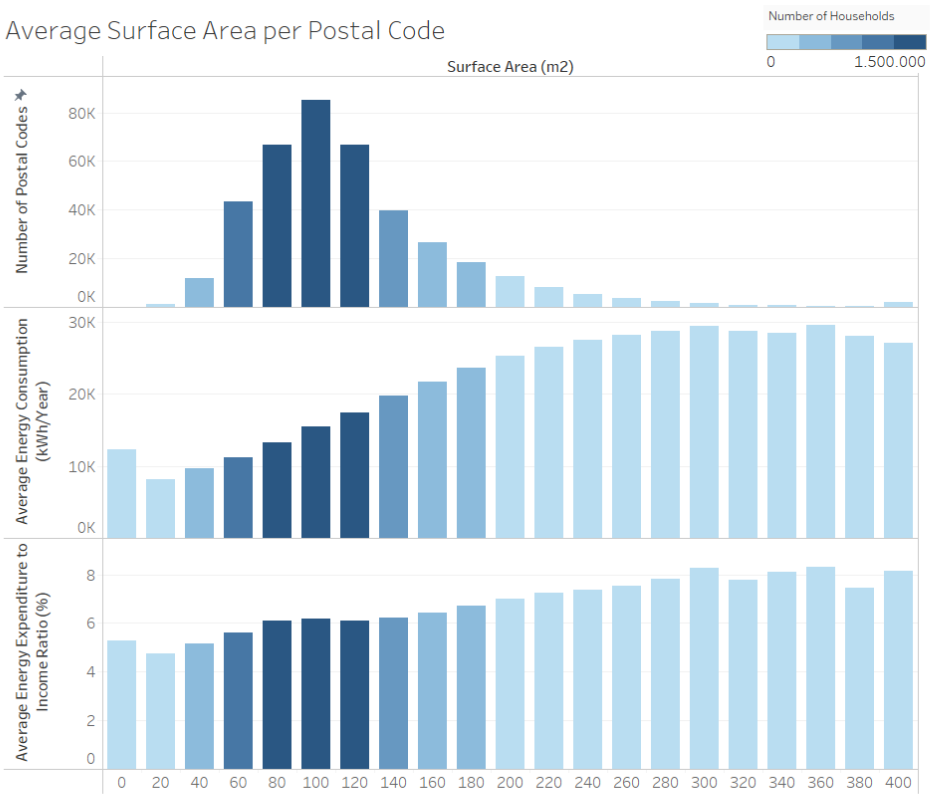

Figure 15: Surface Area Correlations

Distribution Percentage of Detached Houses

Figure 16 visualizes the distribution of the percentage of detached houses per postal code, and its relationship with the dependent variables of this research. The histogram reveals that the distribution is far from normal, and that a very large percentage of the postal codes have 0% of detached houses. Therefore, this 0% value determines the correlation for a large part. This is however not a problem when analysing this variable, as the relationship between the energy consumption and the percentage of detached houses is relatively linear. Postal codes with 0% detached houses have an average energy consumption around 15.000 kWh per year, while those with 100% of detached houses consume close to 26.000 kWh per year on average. The energy expenditure to income ratio has a less clear relationship with the percentage of detached houses, as the 90% and 100% bars don't have the highest ratios. Nevertheless, the trend is still positive on average, with a higher percentage of detached houses resulting in a higher energy expenditure to income ratio.

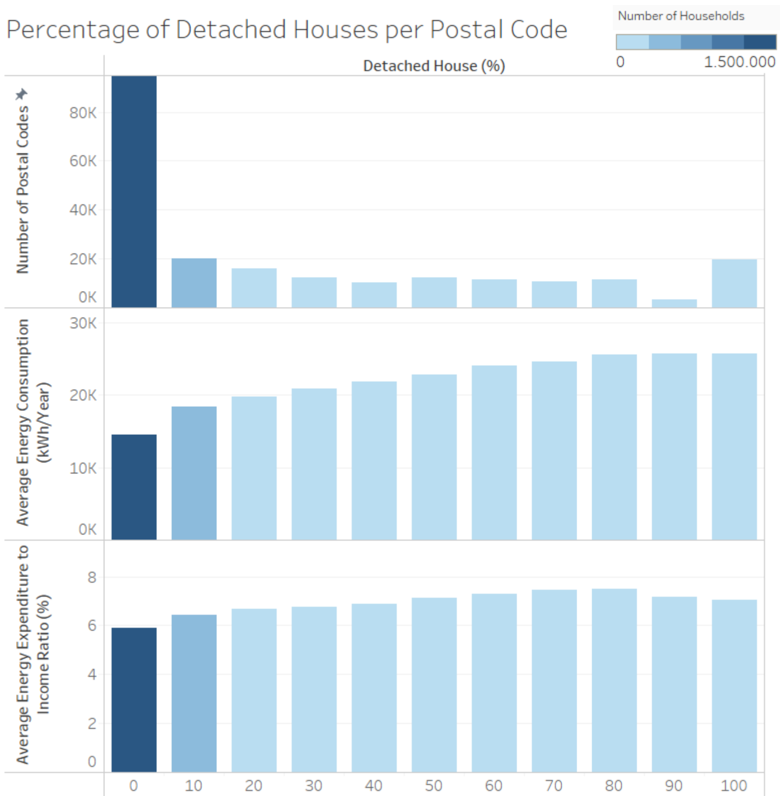

Figure 16: Detached Houses Correlations

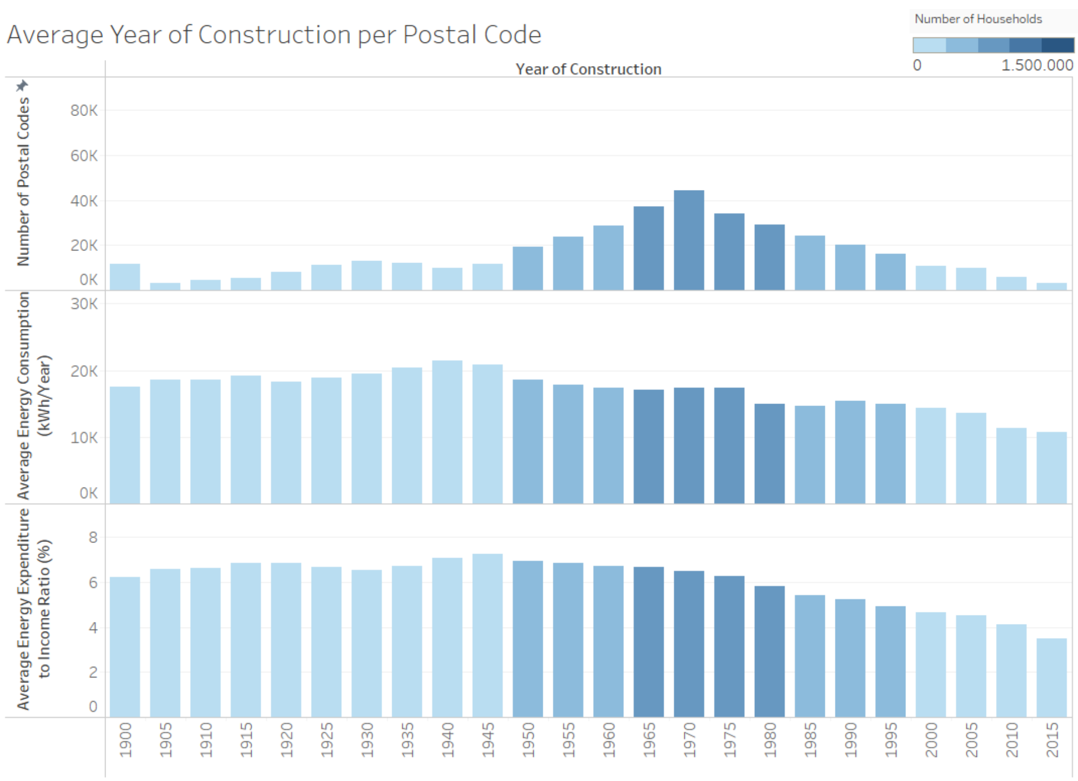

Figure 17: Year of Construction Correlations

Distribution Year of Construction

Figure 17 presents the distribution of the year of construction, and its correlation with the energy consumption and the energy expenditure to income ratio. The distribution of the construction year is quite normal, with a longer tail towards the left, the older houses. Both the relationships with energy consumption and energy expenditure are quite strong until 1945; older houses do however not show a significant correlation with either of these variables. Newer households consume less energy, probably due to better insulation. The energy expenditure to income ratio shows an even stronger negative relation than the energy consumption, with households in newer houses spending considerably less on energy. Postal codes with an average construction year of 1950 have an average expenditure to income ratio of 7%, while postal codes with an average of 2010 only have an average expenditure to income ratio of 4%.

Distribution Urbanity

Figure 18 shows the histograms of the urban-rural scale and the percentage of rental households, and their relationship with the energy consumption and the energy affordability. The urban-rural scale doesn't follow a typical distribution, but all values are represented well, with each of the 5 values representing more than 900.000 Dutch households. The relationship of the urban-rural scale with the energy consumption and the energy expenditure to income ratio is positive, meaning that more rural households consume more energy and spend a larger percentage of their income on energy. Both relationships are relatively linear, which confirms the quality of the ordinal distribution from the CBS and justifies its use in the regression. The average energy consumption per postal code differs from 14.000 kWh per year in very urban areas, represented by code 1, to 19.000 kWh per year in rural areas, represented by code 5. The average energy expenditure to income ratio differs from 5.6% in very urban areas to 6.9% in rural areas.

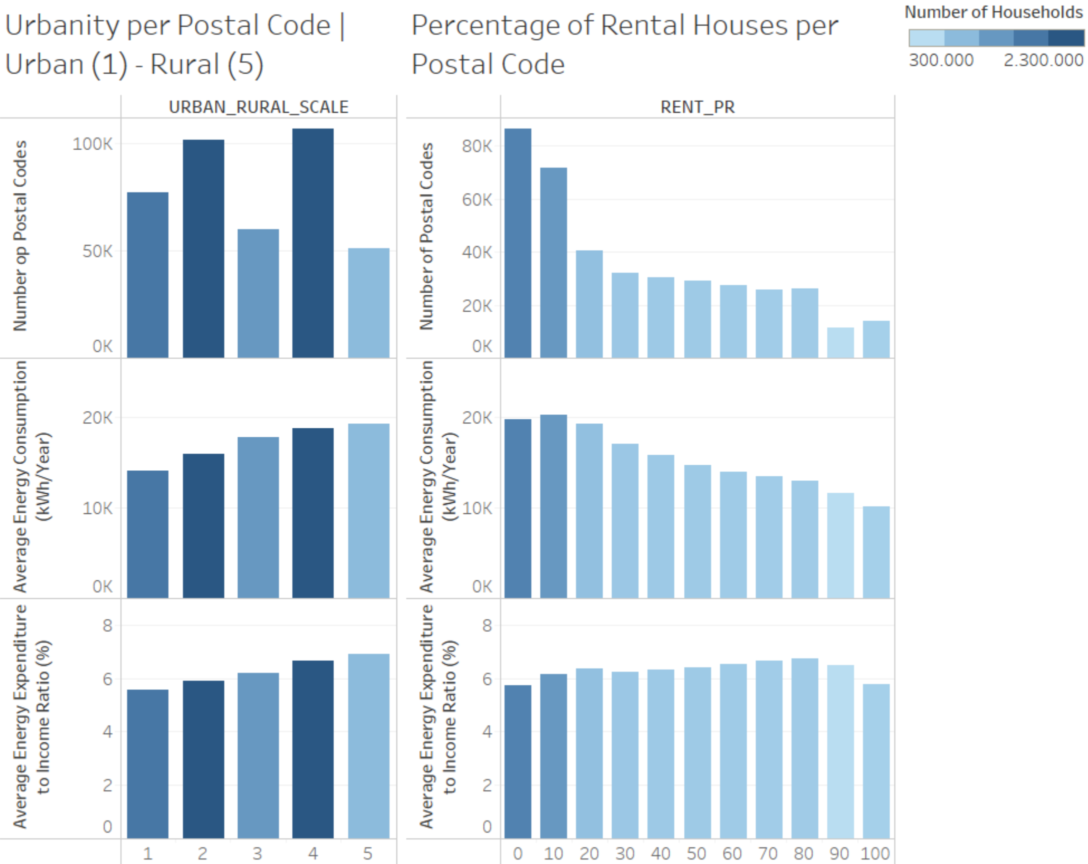

Figure 18: Rental houses and Urban - Rural Correlations

Distribution Rental Houses

The percentage of rental houses per postal code follows an exponential distribution, with many 0% and 10% cases, after which the number of cases per percentage decreases exponentially until it reaches 100%. This distribution is explainable as most houses in the Netherlands are owner occupied. Thus, the number of postal codes with a high percentage of rental houses is limited. The percentage of rental houses has a linear relationship with the average energy consumption within a postal code, which ranges from 20.000 kWh per year at 0% to 10.000 kWh per year at 100%. A higher percentage of rental houses does therefore result in a lower energy consumption. The relationship with the energy expenditure to income ratio is less clear, which explains the low correlation.

All in all, the housing-related factors show strong correlations with the energy consumption and the energy affordability. However, the question remains which variables are able to explain these two phenomena, as a high correlation doesn't have to mean causation. The correlations of the housing-related factors could be explained by the demographic factors or the weather variables, and it is also likely that different housing-related factors capture the same information. Therefore, a regression is needed to see what the individual impact of each variable is on the energy consumption and the energy affordability, which is what the next chapter elaborates on.

5. Inferential Statistics: How Housing-Related Factors impact Energy Consumption and Energy Affordability

5.1 How Housing-Related Factors impact the Energy Consumption

This paragraph presents the results from the regression model on energy consumption, thereby providing insight in the relation that the housing-related factors have with the energy consumption, when controlling for other factors. Thus, the results discussed answer sub-question 6: How do housing-related factors explain the residential energy consumption in the Netherlands?

Interpretation Regression Model

Table 9 and 10 summarize the regression model for the energy consumption. Table 9 reveals the R Square of the two models, which represents the percentage of the variance in the data that is explained by the regression models. Table 10 presents the regression coefficients of the variables included in the regression, and the multicollinearity statistics. The regression coefficients B can be used to estimate the average energy consumption for a specific postal code, by multiplying the coefficients with the relevant values of the postal code. The standardized regression coefficient shows the relative impact of a variable on the dependent variable. Not all variables have the same range, and thus this measure is needed to compare them to each other. The collinearity statistics show that the collinearity is acceptable for all of the variables included, as none of the variables have a VIF higher than 10. This is due to the fact that variables causing multicollinearity were left out of the analysis, as described in the paragraph 3.3.

Before assessing the impact of the housing-related factors it is important to get an overview of the distribution of the energy consumption based on the socio-demographics and the weather variables. These variables were added to the regression model in a separate block, apart from the housing-related factors, as they causally impact the housing-related factors, and aren't influenced by the housing-related factors. Including the variables in two separate blocks is therefore needed to prevent the housing-related factors from capturing the same information as the control variables. Thus, model 1, which only includes the control variables, is used to interpret the socio-demographics and the weather variables.

Impact Socio-Demographic and Weather Factors

Model 1 reveals that the household size and the income are the control variables with the most influence on the energy consumption, with standardized regression coefficients of respectively 0,288 and 0,329. For the household size, this means for example that when comparing a postal code with an average household size of 1,5 with a postal code with an average household size of 3,5, a postal code with an average household size of 3,5 consumes 6212 kWh per year more on average, when controlling for other variables. Households with a higher income also consume more energy, on average 2197 kWh per year more for every €10.000 of extra income.

Table 9:Explained Variance Regression Energy Consumption

Model Summary Energy Consumption

Model	R	R Square	Std. Error of the Estimate	Change Statistics		
				R Square Change	F Change	Sig. F Change
1	,650	,422	4770,163	,422	41843,639	,000
2	,838	,703	3422,683	,280	80959,510	,000

Table 10: Regression Model Energy Consumption

Regression Coefficients Energy Consumption

	Model 1			Model 2					
	Unstandardized Coefficients		Standardized Coefficients	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics
	B	Std. Error	Beta	B	Std. Error	Beta	t	Sig.	VIF
(Constant)	26561.502	203.450		26561.502	203.450		358.13	0.000	
FEMALE_PR	-43.230	0.517	-0.066	-7.204	0.375	-0.011	-19.20	0.000	1.128
RESIDENTS_AGE_1524_PR	48.761	0.510	0.082	32.869	0.369	0.056	89.07	0.000	1.348
RESIDENTS_AGE_2544_PR	-64.002	0.487	-0.164	-14.572	0.356	-0.037	-40.99	0.000	2.886
RESIDENTS_AGE_4564_PR	6.263	0.460	0.014	-8.893	0.333	-0.021	-26.69	0.000	2.051
RESIDENTS_AGE_65PL_PR	8.968	0.422	0.029	20.115	0.311	0.065	64.74	0.000	3.512
WESTERN_MIGRATION_PR	7.188	0.492	0.012	11.747	0.361	0.019	32.50	0.000	1.206
NONWESTERN_MIGRATION_PR	-42.512	0.345	-0.112	12.779	0.273	0.034	46.82	0.000	1.802
ONEPERSON_PR	-22.498	0.327	-0.102	-7.918	0.242	-0.036	-32.77	0.000	4.150
MULTIPLE_PERSON_WITHOUT_CHILDREN_PR	-6.879	0.272	-0.024	-3.075	0.197	-0.011	-15.59	0.000	1.635
TWOPARENTS_PR	-50.528	0.342	-0.188	-8.260	0.251	-0.031	-32.85	0.000	3.044
HOUSEHOLD_SIZE	3105.915	20.382	0.288	559.225	15.259	0.052	36.65	0.000	6.937
INCOME	219.694	0.673	0.329	52.150	0.637	0.078	81.86	0.000	3.145
SOCIAL_BENEFITS_PR	-13.776	0.564	-0.023	19.666	0.417	0.033	47.21	0.000	1.660
AVG_TEMP	-224.877	18.405	-0.016	-394.107	13.389	-0.028	-29.43	0.000	3.144
AVG_MIN_TEMP	-1176.744	11.080	-0.114	-467.770	8.209	-0.045	-56.98	0.000	2.191
AVG_SUNSHINE_DURATION	-1348.091	33.645	-0.093	-875.805	24.313	-0.061	-36.02	0.000	9.779
Year2017	-951.473	16.801	-0.071	-917.665	12.233	-0.069	-75.02	0.000	2.928
Year2018	402.583	24.657	0.030	-60.330	17.993	-0.005	-3.35	0.001	6.330
WOZ_VALUE				11.188	0.044	0.221	254.70	0.000	2.614
SOCIAL_RENT_PR				-2.877	0.136	-0.014	-21.17	0.000	1.478
RENT_PR				-8.801	0.178	-0.042	-49.36	0.000	2.522
URBAN_RURAL_SCALE				77.599	3.414	0.017	22.73	0.000	1.850
YEAR_OF_CONSTRUCTION				-57.216	0.151	-0.234	-379.65	0.000	1.321
HEIGHT_INDICATION_BUILDING				-368.469	4.805	-0.062	-76.68	0.000	2.242
DWELLINGS_IN_BUILDING				9.149	0.483	0.017	18.92	0.000	2.871
LOG_SQUAREMETERS_DWELLING_m2				9971.734	38.399	0.270	259.69	0.000	3.738
APARTMENT_PR				-37.617	0.212	-0.235	-177.13	0.000	6.076
TERRACED_HOUSE_PR				-28.708	0.206	-0.143	-139.37	0.000	3.630
DETACHED_HOUSE_PR				17.710	0.201	0.086	87.95	0.000	3.335
DISTRICT_HEATING				-958.307	17.901	-0.030	-53.53	0.000	1.063

Furthermore, the age, the household type, the non-western migration background and the minimum temperature also have a notable impact on the energy consumption. Households with a high percentage of 25-44 year-olds consume less energy. The same holds for one person households and two parent households, when controlling for the household size. It is an interesting observation that two parent households have a negative relation with the energy consumption in the regression model, as table 6 reveals that its correlation with the energy consumption is positive. This is probably because the regression controls for household size. Moreover, households with a high percentage of residents with a non-western migration background also have a lower energy consumption. Households experiencing a lower average minimum temperature do however have a higher energy consumption, as they need more energy for heating. The differences between the different years included in the analyses are relatively low, although the average energy consumption was lower in 2017 and higher in 2018, when compared to 2016, and controlling for other factors. Table 9 reveals that the control variables all together explain 42% of the variance in energy consumption, meaning that they have a high explaining power.

Impact Housing-Related Factors

The model with the housing-related factors on the other hand, explains 70% of the total variance, which is 28% more than the model with control variables. Thus, the housing-related factors add a large percentage to the variance, suggesting that they have a high impact on the energy consumption.

Although the correlations of the percentage of rental households and the urbanity were strong, the regression coefficients show a relatively weaker relation with the energy consumption, when controlling for other factors. The urban-rural scale has a standardized coefficient of 0,017, and the percentage of rental households has a standardized coefficient of -0,042. This means, that the average household in a postal code with 100% rental houses consumes 880 kWh per year less than a household in a postal code with 0% rental houses. A very urban household consumes on average 310 kWh per year less than a rural household. As the average energy consumption in the Netherlands is 17.302 kWh per year, these variables don't have a high explaining power. Much of their correlation with the energy consumption is thus explained by other variables, such as income, surface area and house type.

The WOZ-value and the surface area in square meters do have a high impact on the energy consumption, as the correlations suggested. A higher WOZ-value and a higher number of square meters leads to a higher energy consumption. The energy consumption increases with 1.119 kWh per year on average for every €100.000 of extra WOZ-value. The logarithmic scale of the surface area is harder to interpret as the impact of an extra square meter depends on the size of the house. When comparing an average house of 100 m2 to an average house of 250 m2, the average house of 100 m2 consumes 3968 kWh less per year.

The impact of the percentage of detached houses on the energy consumption is much weaker than its correlation, with a standardized regression coefficient of 0,086. Nevertheless, when comparing it to other housing types, detached houses do consume much more energy. A postal code with 100% of detached houses consumes on average 4.642 kWh more than a postal code with 100% of terraced houses, and 5.533 kWh more than a postal code with 100% apartments. This is mainly captured in the high negative regression coefficients of the percentage of terraced houses and the percentage of apartments, which respectively have a standardized regression coefficient of -0,143 and -0,235.

The year of construction has a strong negative regression coefficient, meaning that older houses have a higher energy consumption. A postal code with an average year of construction of 1950 for example consumes 2.861 kWh per year more than a postal code with an average year of construction of 2000, which confirms that the year of construction is one of the key factors impacting energy consumption.

Other factors have a much lower regression coefficient. The standardized regression coefficient of district heating is 0,030, meaning that the postal codes with district heating that are left in the analyses do indeed only have a small percentage of houses with district heating, as the regression coefficient is low. The percentage of social rental houses had no relevant impact on the energy consumption. Similarly, the number of dwellings in a building has a minimal positive relation with energy consumption. The height indication of the building has a somewhat stronger standardized regression coefficient of -0,062, pointing out that households in higher buildings have a lower average energy consumption.

5.2 How Housing-Related Factors impact the Energy Affordability

This paragraph provides insight in the relation between the housing-related factors and the energy affordability. It builds on a regression model which quantifies the relation between the housing-related factors and the energy expenditure-to-income ratio, which is used to measure the affordability. Thus, it presents answers to sub-question 7: How do housing-related factors explain the residential energy affordability in the Netherlands?

Interpretation Regression Model

Table 11 and 12 summarize the regression models on the energy expenditure to income ratio. Table 11 reveals the R Square of the models, while table 12 shows the regression coefficients and the collinearity statistics. When considering the regression coefficients, the first interesting observation is that they are significantly smaller than the regression coefficients of the energy consumption. This is due to the fact that the range and the standard deviation of the energy expenditure to income ratio are much smaller in absolute numbers, which makes them incomparable to the energy consumption. While the range of the energy consumption is more than 80.000 kWh, the range of the expenditure to income ratio is only 39%. The values of the standardized regression coefficients are however comparable, which are therefore also used for interpreting the regression.

Impact Socio-Demographic and Weather Factors

The impact of the socio-demographic and weather variables on the energy expenditure to income ratio is discussed first, as this provides context to the impact that the housing-related factors have. The average household income per postal code has by far the strongest standardized regression coefficient, with a coefficient of -0,596. This was expected, as the income is divided by the energy costs to create the expenditure to income ratio. Per €10.000 of extra yearly income, the expenditure to income ratio decreases on average with 1,29%.

Nevertheless, other variables have a strong regression coefficient as well. The household size has a relatively high standardized regression coefficient of 0,273, meaning that postal codes with a higher average household size spend a higher percentage of their income on energy. The age groups of 15-24 year-olds and 25-44 year-olds also have a strong regression coefficient. 15-24 year-olds have a higher energy expenditure to income ratio, while 25-44 year-olds have a significantly lower expenditure to income ratio. Postal codes with a high percentage of one person or two parent households also spend a lower percentage of their income on energy. The average minimum temperature has a strong negative relationship with the energy expenditure to income ratio. For example, if the average minimum temperature is 2°C lower, the energy expenditure to income ratio is on average 0,77% higher. This suggests that households in the northern and eastern part of the country spend a significantly higher percentage of their income on energy, as the minimum temperature is lower there. When considering the year dummies, it can be stated that households have spent a larger percentage of their income on energy in 2018 and a lower percentage of their income on energy in 2017, when compared to 2016. The observed differences are interesting, as they are significantly larger than the differences observed in energy consumption, and as the energy price fluctuations between these years are minimal.

All in all, 36% of the variance in the energy expenditure to income ratio can be explained by the control variables. This is a large part of the variance, but considerably less than the variance explained for the energy consumption, pointing out that the variance in the energy expenditure to income ratio is more diverse.

Table 11: Explained Variance Regression Energy Expenditure to Income Ratio

Model Summary Energy Expenditure to Income Ratio

				Change Statistics		
Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	,602	,362	1,628275	,362	32271,000	,000
2	,780	,608	1,276709	,246	53387,152	,000

Table 12: Regression Model Expenditure to Income Ratio

Regression Coefficients Energy Expenditure to Income Ratio

	Model 1			Model 2					
	Unstandardized Coefficients		Standardized Coefficients	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics
	B	Std. Error	Beta	B	Std. Error	Beta	t	Sig.	VIF
(Constant)	14.780	0.070		43.675	0.124		351.35	0.000	
FEMALE_PR	-0.013	0.000	-0.062	-0.003	0.000	-0.013	-20.53	0.000	1.128
RESIDENTS_AGE_1524_PR	0.022	0.000	0.117	0.018	0.000	0.093	129.83	0.000	1.348
RESIDENTS_AGE_2544_PR	-0.024	0.000	-0.188	-0.008	0.000	-0.065	-61.97	0.000	2.886
RESIDENTS_AGE_4564_PR	0.000	0.000	-0.003	-0.005	0.000	-0.035	-39.17	0.000	2.051
RESIDENTS_AGE_65PL_PR	-0.002	0.000	-0.024	0.003	0.000	0.027	23.21	0.000	3.512
WESTERN_MIGRATION_PR	0.006	0.000	0.029	0.006	0.000	0.030	43.66	0.000	1.206
NONWESTERN_MIGRATION_PR	-0.011	0.000	-0.091	0.006	0.000	0.051	60.93	0.000	1.802
ONEPERSON_PR	-0.007	0.000	-0.098	-0.001	0.000	-0.016	-12.30	0.000	4.150
MULTIPLE_PERSON_WITHOUT_CHILDREN_PR	0.002	0.000	0.017	0.003	0.000	0.029	36.35	0.000	1.635
TWOPARENTS_PR	-0.011	0.000	-0.125	0.002	0.000	0.023	21.30	0.000	3.044
HOUSEHOLD_SIZE	0.957	0.007	0.273	0.139	0.006	0.040	24.27	0.000	6.937
INCOME	-0.130	0.000	-0.596	-0.175	0.000	-0.805	-732.94	0.000	3.145
SOCIAL_BENEFITS_PR	0.005	0.000	0.027	0.015	0.000	0.076	95.30	0.000	1.660
AVG_TEMP	-0.099	0.006	-0.022	-0.179	0.005	-0.039	-35.76	0.000	3.144
AVG_MIN_TEMP	-0.387	0.004	-0.115	-0.178	0.003	-0.053	-57.77	0.000	2.191
AVG_SUNSHINE_DURATION	-0.419	0.012	-0.089	-0.252	0.009	-0.054	-27.64	0.000	9.779
Year2017	-0.384	0.006	-0.089	-0.367	0.005	-0.085	-80.11	0.000	2.928
Year2018	0.435	0.008	0.101	0.292	0.007	0.067	43.29	0.000	6.330
WOZ_VALUE				0.003	0.000	0.201	200.44	0.000	2.614
SOCIAL_RENT_PR				-0.001	0.000	-0.016	-21.68	0.000	1.478
RENT_PR				0.000	0.000	-0.004	-4.25	0.000	2.522
URBAN_RURAL_SCALE				0.040	0.001	0.026	31.17	0.000	1.850
YEAR_OF_CONSTRUCTION				-0.018	0.000	-0.229	-321.85	0.000	1.321
HEIGHT_INDICATION_BUILDING				-0.100	0.002	-0.052	-55.71	0.000	2.242
DWELLINGS_IN_BUILDING				-0.004	0.000	-0.022	-21.26	0.000	2.871
LOG_SQUAREMETERS_DWELLING_m2				3.312	0.014	0.276	230.29	0.000	3.738
APARTMENT_PR				-0.011	0.000	-0.211	-138.53	0.000	6.076
TERRACED_HOUSE_PR				-0.004	0.000	-0.063	-53.78	0.000	3.630
DETACHED_HOUSE_PR				0.005	0.000	0.076	66.78	0.000	3.335
DISTRICT_HEATING				-0.270	0.007	-0.026	-40.30	0.000	1.063

Impact Housing-Related Factors

The model with the housing-related factors explains 25% of extra variance, when compared to the model with only the control variables. This means that the housing-related factors play a significant role in the energy expenditures of households. However, also for the housing-related factors, the variance explained is lower than for the energy consumption, which confirms that the energy expenditure to income ratio has more random variance due to variables not in the analysis.

The standardized regression coefficients of the percentage of rental houses and the urbanity are again low. The percentage of rental households has a standardized regression coefficient of -0,004. On average, a postal code with 100% of rental households thus spends 0,03% of their income more on energy. The urban-rural scale has a standardized regression coefficient of 0,026, meaning a very urban household spends on average 0,16% of their income less on energy than a rural household. Especially for the urban-rural scale this low value is interesting, because the correlation with the energy expenditure to income ratio is notable. However, as the expenditure to income ratio has an average value of 6,16% in the Netherlands and a standard deviation of 2,04%, the influence of these variables is not very high.

The year of construction has a strong negative standardized regression coefficient of -0,229, which is in line with the negative correlation discussed in chapter 4. Households living in older houses spend a higher percentage of their income on their energy consumption. The model points for example out that postal codes with an average construction year of 1960 have an average energy expenditure to income ratio that is 0,91% higher than that of postal codes with an average construction year of 2010.

The variation in the energy expenditure to income ratio is also large for the different house types. The percentage of detached houses has a standardised regression coefficient of 0,076, the percentage of terraced houses has a coefficient of -0,063, and the percentage of apartments has a coefficient of -0,211. This means that the energy expenditure to income ratio is on average 0,92% higher for a postal code with 100% of detached houses, when compared to a postal code with 100% of terraced houses. However, when compared to a postal code with 100% of apartments, the difference is even larger, with an energy expenditure to income ratio that is 1,61% higher. This makes the house type the housing-related factor that has the most influence on the energy expenditure to income ratio.

The surface area in square meters has a positive standardised regression coefficient of 0,276. Therefore, households with a larger surface area spend a higher percentage of their income on energy. On average, a house with a surface area of 100 m2 spends 1,32% of its household income less on energy than a house with a surface area of 250 m2. The WOZ-value also has a regression coefficient of 0,201, meaning that postal codes with a higher WOZ-value have a higher expenditure to income ratio. This is interesting, as the correlation of the WOZ-value with the energy expenditure to income ratio was low and negative. This points out that when controlling for other factors, the WOZ-value does actually have a positive relationship with the energy expenditure to income ratio, in contrast to what the correlation was suggesting.

The regression coefficients of the other housing-related factors are relatively low. This is interesting, as the number of dwellings in a building and the height indication did have notable correlations with the energy expenditure to income ratio. When controlling for other variables however, the relationship that these factors have with the energy expenditure to income ratio is much weaker with standardised regression coefficients of respectively -0,022 and -0,052. Nevertheless, postal codes with a large average number of dwellings per building and a high average height indication do have a lower energy expenditure to income ratio. The percentage of social rent had again no relevant impact.

6. Discussion & Conclusion

This chapter discusses the results found, answers the research question based on these results and identifies the implications and limitations of this study. First, paragraph 6.1 combines the insights from the results found in chapter 4 and 5, and discusses them based on the literature. Second, paragraph 6.2 concludes this research by answering the main research question and summarizing the core findings. Third, the implications of this thesis are elaborated on in paragraph 6.3, both from an academic and a policy perspective. Finally, paragraph 6.4 points out the limitations of this study, and identifies interesting areas for future research.

6.1 Summary of Findings and Comparison with Literature

The aim of this research is to provide insight in the relation between the housing-related factors and the energy consumption and energy affordability. In order to do that, the distribution of energy consumption and energy affordability among households in the Netherlands was assessed first.

Distribution Energy Consumption

With an average electricity consumption of 3014 kWh per year, an average gas consumption of 14511 kWh per year, and a total energy consumption of 17302 kWh per year, the data analysed is representative for all Dutch households. Abrahamse and Steg (2009), Santin et al. (2009) and Brounen et al. (2012) reveal similar averages in their research on the energy consumption of the Dutch residential sector, confirming the validity of the data. The standard deviation is lower when compared to other research, e.g. 937 kWh for the electricity consumption instead of 1557 kWh in the data used by Abrahamse & Steg (2009). This can be explained by the fact that the unit of analyses in this research is postal code, which means that data from multiple households is aggregated, thus reducing the variance of the data.

The geographical distribution of the energy consumption revealed that the north-eastern and eastern part of the Netherlands consume significantly more energy, especially in rural areas. Where most municipalities in the Randstad have an average energy consumption between 12.500 and 17.500 kWh per year, in the north-eastern part of the country most municipalities have an average energy consumption that ranges between 17.500 and 22.500 kWh per year. Brounen et al. (2012) shows similar results on a province level, and research in other western countries also reveals that geographical differences in energy consumption can be large (Tso & Guan, 2014; Iraganaboina & Eluru, 2021).

Distribution Energy Affordability

When considering the energy affordability, it is more difficult to assess how this study compares to other research. First, the energy expenditure to income ratio is not always the measure used for calculating the energy affordability. Other methods, such as the Low Income High Costs (LHIC) method, and the energy poverty line of 10% are often used (Belaïd, 2018; Das et al., 2022). When considering other studies that do use the energy expenditure to income ratio, the differences are large. European studies reveal average energy expenditure to income ratios between 4% in Italy to 14% in Greece (Bardazzi et al., 2021; Karasek & Pojar, 2018; Mashhoodi et al., 2018; Papada & Kaliampakos, 2017). With an average of 6.16%, the mean energy expenditure to income ratio of this study is therefore relatively average when compared to the rest of Europe, although the more economically developed western Europe does have a lower expenditure to income ratio than other European countries. Another conclusion that can be drawn from this is however that the research done on energy affordability is relatively limited, and that the different measures and calculations of energy affordability impede a proper understanding of this concept. One of the reasons for this is that a large number of studies only uses a binary measurement for energy affordability, e.g. by assessing which households fall below the 10% energy poverty line, which makes these kind of studies very threshold-sensitive (Mashhoodi et al., 2018).

The geographical distribution of the energy affordability is similar to that of the energy consumption, as the energy expenditure to income ratio is higher in the eastern and especially north-eastern part of the country. The differences are however even larger than the differences in energy consumption, with the municipalities in the Randstad having an average energy expenditure to income ratio of 4% to 6%, while the north-eastern municipalities have an average ratio of 7% to 9%. This means that the average income is also lower in the eastern and north-eastern part of the Netherlands, as this is the denominator in the expenditure to income ratio. A comparable geographical distribution is also presented by Mashhoodi et al. (2018), confirming that the geographical inequality in energy affordability is relatively high in the Netherlands.

Additional Analyses Energy Affordability 2021-2022

An additional analysis was done on the energy affordability, based on the recent energy prices in October 2021 and April 2022. Even when correcting for the decreasing energy demand, this analysis reveals that the average energy expenditure to income ratio is increasing in the Netherlands, with an average ratio of 7,31% based on the prices in October 2021 and an average ratio of 11,32% based on the prices in April 2022. This is alarming, as the last scenario suggests that more than half of the Dutch population spends more than 10% of their income on energy, which is considered the energy poverty line (Lin & Wang, 2020; Haffner & Boumeester, 2015). In the north-eastern part of the country, the energy expenditure to income ratio is even higher, averaging over 15% in some municipalities. Therefore, this study confirms that the increasing energy prices are disrupting the energy market (Tesio, et al., 2022; CPB, 2022), and thus increasing the inequality among households in the Netherlands while pushing households into energy poverty.

Comparing Methodology to Similar Studies

The reasons for the inequality in the energy consumption and the energy affordability were assessed in the regression analyses, where the housing-related factors were used as predictors, next to the socio-demographic and weather variables, which were included as control variables. This study used multivariate regression analyses to assess the relation between the dependent and the independent variables, which is the most common method in this type of studies (Besagni & Borgarello, 2018; De Arce & Mahia, 2019; Tso & Guan, 2014; Santin et al., 2009; Brounen et al., 2012; Bardazzi et al., 2021). Other types of regression analysis, such as logistic regression are also used often, especially when dummy variables are used or when the dependent variable is binary, e.g. energy poverty or no energy poverty (Chalal et al., 2017; Riva et al., 2021; Khundi-Mkomba et al., 2021). As this study includes only one dummy variable and has no binary dependent variables, multivariate regression analyses produces reliable and accurate results and is thus the preferred method.

Impact of Control Variables on Energy Consumption

When applying the regression analysis to the energy consumption, the control variables were added first. The household size and the income had the strongest regression coefficients and are thus the most important socio-demographic predictors of energy consumption. Both variables have a positive relation with the energy consumption, which is similar to the relations found in other literature (Huebner et al., 2016; Abrahamse & Steg, 2009; Santin et al., 2009). The regression also revealed that the household type has a notable influence on the energy consumption, with one-person households and two-parent households consuming less energy. These types of households also consume less in previous studies, when controlled for the household size (Brounen et al., 2012; Chalal et al., 2017). The temperature has a negative relation with the energy consumption, meaning that energy consumption is higher when the average temperatures are lower, which is in line with previous research (Iraganaboina & Eluru, 2021; Santin et al., 2009).

It is interesting to note that postal codes with a high percentage of 25-44 year-olds consume significantly less energy; on average a postal code with 0% of 25-44 year-olds consumes 3200 kWh more per year than a postal code with an average of 50% 25-44 year olds. This is contrasting with previous research in the Netherlands, which suggests that there is no relation between energy consumption and age (Abrahamse & Steg, 2009), or that the elderly consume more energy (Brounen et al., 2012). This could either be explained by a change in the energy consumption per age group over the past 10 years, or by the fact that the research samples in previous research were not completely representative for the Dutch population, as this research considers most Dutch households.

Impact of Housing-Related Factors on Energy Consumption

Afterwards the impact of the housing-related factors on the energy consumption was assessed. Although the correlation of the percentage of rental houses with the energy consumption is strong, the standardized regression coefficient of -0,042 is relatively weak. This means that rental houses do consume significantly less energy, but that this can for a large part be explained by other housing-related factors such as the surface area and the house type. Nevertheless, the relation in the regression is still negative, revealing that even when corrected for other factors, rental houses do consume slightly less energy. This provides more clarity on the relationship between the percentage of rental houses and the energy consumption, as previous research on this is contradicting, presenting both positive and negative relationships (Santin et al., 2009; Huang, 2015). The reason for the negative relationship that the percentage of rental houses has with the energy consumption in the Netherlands could be explained by the lower socio-economic status of rental households and by the lack of energy-intensive appliances when compared to owner-occupied households (Yohanis et al., 2008; Huang, 2015).

The Urban-Rural scale has an even weaker regression coefficient, although it also has a significant correlation. Households in rural areas consume more energy, but this can be explained by the control variables, and the other housing-related factors. When controlling for these variables, no relevant relation remains. This is interesting, because research in other parts of the world reveals that rural households consume less energy (Iraganboina & Eluru, 2021; Zheng et al., 2014; Xie et al., 2020). This research reveals in contrast to the US and Asia, that there is no clear difference in energy consumption between urban and rural households in Western Europe, and more specifically in the Netherlands.

The housing-related factors that have the strongest relation with the energy consumption are the surface area in square meters, and the percentages of different house types. Apartments consume less energy, and detached houses consume more energy, which is in line with previous literature (Tso & Guan, 2014; Brounen et al., 2012). The same holds for the relation between the surface area and the energy consumption, as houses with a larger surface area consume more energy (Huebner et al., 2016; Iraganaboina & Eluru, 2021; Santin et al., 2009).

The construction year has a strong negative relation with the energy consumption, meaning that older houses consume more energy. The reason for this is that the thermal quality and insulation is worse for older houses (ŠTreimikienė, 2014). Similar research on this relation also notes a significant negative relation between the energy consumption and the year of construction, however this relation is not as strong as in this research. Santin et al. (2009) calculate a standardized regression coefficient of -0,082 for the year of construction in the Netherlands, which is weaker than the -0,234 found in this study. This suggests that the energy consumption has decreased even stronger in recent years, and that newer houses consume significantly less energy. This is confirmed by figure 17, which presents the energy consumption distribution for the different years of construction, and reveals that houses build after 2010 show a much lower energy consumption. Therefore, it can be concluded that the impact of the year of construction on the energy consumption has increased in the past decade.

In contrast to the year of construction, the WOZ-value has a positive relationship with the energy consumption. When the WOZ-value or property value increases, the energy consumption is higher on average. Although no articles were found in which the property value was included in a quantitative analysis on the energy consumption, it is a strong predictor of the energy consumption. This is partly due to the fact that the property value captures the value of housing-related factors that are not included in the regression, but do influence the energy consumption (Abidoye & Chan, 2016). Next to that, houses with a high WOZ-value have a lower energy label and thus a lower energy efficiency (Boesveld, 2021). Moreover, these houses lag behind in the implementation of energy efficiency policies (Van Middelkoop et al., 2017). An explanation for this could be that for houses with a high WOZ-value, taxes based on the property value increase significantly when households invest in energy efficiency (Boesveld, 2021; Dekker, 2014).

Impact of Control Variables on Energy Affordability

Next to the energy consumption, the energy affordability was also analysed using regression. The regression analysis on the energy expenditure to income ratio revealed that the income and the household size are again the strongest predictors among the control variables. However, the relation that these variables have with the energy affordability is different than with the energy consumption. The household size has a positive relationship with the energy expenditure to income ratio, but the income has a negative relationship with this variable, which shows that a higher income leads to a lower expenditure to income ratio. Both these relationships were expected, as these factors were also crucial in previous research (Mashhoodi et al., 2018; Das et al., 2022; De Arce & Mahia, 2019). The minimum temperature has a negative relationship with the energy expenditure to income ratio too, which is confirmed by Das et al. (2022) and Mashhoodi et al. (2018).

The household type is also an important predictor of the energy expenditure to income ratio. The regression showed that one-person households and two-parent households spend a smaller percentage of their income on energy. Two-parent households have a similar relation with the energy expenditure to income ratio in other western countries, confirming the result from the regression (Das et al., 2022; Riva et al., 2021). One-person households on the other hand have a significantly higher expenditure to income ratio in other western countries (Belaïd, 2018; Riva et al., 2021; Das et al., 2022), which suggests that one-person households in the Netherlands are different when compared to one-person households in North America and the rest of Western Europe.

The age of the residents also influences the energy affordability in a different way, when compared to similar countries. The literature notes either no relation between age and the energy expenditure to income ratio (Besagni & Borgarello, 2018), or a higher expenditure to income ratio for elderly residents (Riva et al., 2021). The regression in this study revealed however that in the Netherlands, the 15-24 year-old residents spend a higher percentage of their income on energy, while the 25-44 year-old residents spend a lower percentage of their income on this. The relation for 25-44 year olds can be seen in the energy consumption regression as well, but the lower energy affordability of 15-24 year olds is for a large part due to the fact that the average income of this age-group is lower. Furthermore, the energy expenditure to income ratio differed significantly over the years included in the analyses. This is however in line with previous literature, and can be explained by the yearly variations in energy needs, household needs and the changes in household composition (Riva et al., 2021).

Impact of Housing-Related Factors on Energy Affordability

The relation between the housing-related factors and the energy expenditure to income ratio was analysed next. Both the Urban-Rural scale, and the percentage of rental houses showed no relevant relationship with the energy expenditure to income ratio. The urbanity does have a strong relationship with the expenditure to income ratio in North America (Das et al., 2022; Riva et al., 2021), but was not found to be relevant in previous European research (Belaïd, 2018; Longhi, 2015). This research confirms this by showing that the urbanity also has a negligible influence on the energy affordability in the Netherlands. Although research on this is contradicting (Mashhoodi et al., 2018), the percentage of rental houses on the contrary was expected to show a positive regression coefficient. This is on the one hand due to the fact that households living in rental houses spend a higher percentage of their income on energy in France and Canada (Riva et al., 2021; Belaïd, 2018), and on the other hand because the average income of Dutch households living in rental households is lower (Hoekstra, 2017; Haffner & Boumeester, 2015). Rental households do however not have a lower energy affordability, as the percentage of rental houses has no positive relationship with the energy expenditure to income ratio in this study. This suggests that the lower energy consumption of rental households compensates their low income, thereby diminishing the relation between rental houses and energy affordability. The energy consumption regression confirms this, as it showed that rental houses consume less energy on average.

The surface area and the house type have a strong relationship with the energy affordability. Houses with a larger surface area spend a higher percentage of their disposable income on energy, which is confirmed by Kontokosta et al. (2019) and Besagni and Borgarello (2018). When considering the house type, detached houses spend the largest part of their income on energy, followed by terraced houses and finally apartments, which spend the smallest percentage of their income on energy. This is also in line with previous literature (Riva et al., 2021; Besagni & Borgarello, 2018), which confirms the regression results. The relation with the WOZ-value was again positive, with higher values leading to higher energy expenditure to income ratios. Although no previous research was found on the relation of this variable with the energy affordability, Van Middelkoop et al. (2017) show that houses with a high WOZ-value do not leverage energy efficiency policies as often as other houses. Furthermore, Boesveld (2021) reveals that houses with a high WOZ-value have a lower energy label, and thus a lower energy efficiency. This explains why the WOZ-value is a strong predictor of energy affordability.

The year of construction again has a strong negative regression coefficient, stating that the expenditure to income ratio is higher for older houses. Previous literature however doesn't agree on the influence of this variable, with Besagni and Borgarello (2018) finding no relation, Das et al. (2022) finding only a weak relation, and Riva et al. (2021) finding a strong relation with the energy affordability. In the Netherlands however, the relation is strong, which is again for a large part due to the energy efficient houses built in recent years, as revealed by figure 17. This energy efficiency increase is a result of the Dutch energy saving policies, which stimulated the construction of energy efficient housing (Vega et al., 2022). This study reveals the impact of these energy efficiency increases on the affordability, which is large, given that this variable has a standardized regression coefficient of -0,229.

Comparing Energy Consumption and Energy Affordability

Overall, it can be concluded that the housing-related predictors for energy consumption and energy affordability found in this study are relatively similar. The year of construction, house type, surface area and WOZ-value have a strong relation with both the energy consumption and the energy affordability, although the magnitude of this relation differs. The standardized regression coefficient of -0,143, for the percentage of terraced houses is for example relatively close to that of apartments in the energy consumption analysis, whereas in the energy affordability analysis it has a coefficient of -0,063, which is in between the percentage of detached houses and the percentage of apartments. Another important difference is that the percentage of rental houses has a notable relationship with the energy consumption, but is almost unrelated to the energy affordability. Furthermore however, the relations are often comparable, which was expected as the energy consumption is a crucial part of the energy expenditure to income ratio.

6.2 Conclusion

All in all, this thesis provides insight in the relationship between the housing-related factors and the energy consumption and energy affordability, by analysing postal code 6 data on 8 million Dutch households. The objective of this study is summarized in the following research question: *How do housing-related factors impact the Dutch residential energy consumption, and the affordability of this energy consumption?*

Factors impacting the Energy Consumption and Energy Affordability

This question was answered using multiple sub-questions. First, a literature review was performed, which identified the housing-related factors impacting the energy consumption and energy affordability, and the control variables that are expected to have a relation with these phenomena. The housing-related factors included in the analyses based on this literature review are: the surface area, the year of construction, the WOZ-value, the house type, the number of dwellings in a building, the height indication of the building, the urbanity, the ownership status and whether houses are in social rent. The impactful housing-related factors are similar for energy consumption and energy affordability, confirming the strong relation between these concepts.

Distribution of Energy Consumption and Energy Affordability

Afterwards, the distribution of the energy consumption and energy affordability among Dutch households was analysed. The analyses reveal that both the energy consumption and the energy affordability follow a relatively normal distribution in the Dutch population. The average household energy consumption per postal code is 17.302 kWh per year over the years 2016 to 2018, including both the gas and electricity consumption. The average energy expenditure to income ratio, which is used to measure the energy affordability, is 6,16%. Visualising the energy consumption and affordability geographically did however reveal that differences are large within the Netherlands. The Randstad and the large cities in the country have a lower energy consumption and a higher energy affordability, while the eastern and particularly north-eastern part of the country have a higher energy consumption and a lower affordability. Municipalities in the Randstad have an average household energy consumption between 12.500 and 17.500 kWh per year, while households in municipalities in the north-eastern part of the Netherlands consume 20.000 kWh on average. For the energy affordability, this ranges from 4% to 6% in the Randstad to 7% to 9% in the north-eastern part of the country. This revealed that the inequality is high within the Netherlands when comparing different parts of the country, both in terms of energy consumption and energy affordability.

Impact of Current Energy Prices on the Energy Affordability

An additional affordability analysis showed that the energy affordability is decreasing dramatically due to the current soaring energy prices. Based on April 2022 energy prices, the average energy expenditure to income ratio is increasing to 11,32%, even when taking into account that households consume 30% less due to these high prices. In this situation, more than half of the Dutch households spend more than 10% of their income on energy, which is considered the energy poverty line. Thus, the current energy prices have a large impact on the energy affordability, which makes it even more important to know which factors are influencing this affordability.

Correlations with Housing-Related Factors

The correlations with the housing-related factors revealed that the average surface area, the percentage of detached houses and the average WOZ-value have the highest correlation with the average household energy consumption per postal code. For different values of these factors, the variation in the average energy consumption is more than 10.000 kWh. For the energy affordability, the correlations are generally lower, suggesting that the income has a strong correlation with the housing-related factors. The average year of construction, the percentage of detached houses, and the urbanity show the largest differences in the energy expenditure to income ratio. However, different values of these variables only result in a spread of 2% in the energy expenditure to income ratio due to the lower correlations.

Impact of Housing-Related Factors on Energy Consumption and Affordability

Furthermore, the impact of the housing-related factors on the energy consumption and energy affordability was analysed using regression analyses, while controlling for socio-demographic and weather factors. The results showed that the impacting factors are similar for the energy consumption and the energy affordability, only the strength of the relation differs in some cases. The control variables that have the strongest relation with the energy consumption and affordability are the income and the household size. A higher income results in a higher energy consumption and energy affordability, while a larger household results in a higher energy consumption and a lower energy affordability.

When considering the housing-related factors, the year of construction, the surface area, the house type and the WOZ-value show the strongest relation, with standardized regression coefficients above 0,200. All of these variables have a similar relationship for the energy consumption and the energy affordability. Older houses consume more energy and spend more on energy and houses with a larger surface area or WOZ-value have a higher energy consumption and energy expenditure to income ratio. With regard to the house type, apartments have the lowest energy consumption and expenditure to income ratio, while detached houses consume the most energy and spend a higher percentage of their income on energy.

It is interesting to note that there is no relevant difference in urban and rural energy consumption in the Netherlands, while this is the case in other European and North American countries. Next to that rental households consume less energy, but don’t spend less on energy, which suggests that these households have problems with their energy affordability, even though they don’t have a higher expenditure to income ratio. This is confirmed by previous research, which found that rental households in similar countries do spend more on energy.

Furthermore, the relationship between the year of construction and the energy consumption and affordability has increased significantly when compared to previous research, due to the fact that newer houses built after 2010 are much more energy efficient. Another interesting factor is the WOZ-value, as it wasn’t considered in previous research, but has a strong relation with both the energy consumption and affordability. Houses with a higher WOZ-value implement less energy policies, and thus have a lower energy efficiency, which is interesting as energy efficiency measures such as insulation increase the WOZ-value of a house. Nevertheless, houses with a high WOZ-value consume more and spend a higher percentage of their income on energy.

In conclusion, it can be stated that housing-related factors explain a large part of the energy consumption and energy affordability. The energy efficiency of a house depends for a large part on the housing-related factors, and the socio-demographics also differ based on housing-related factors. The surface area, WOZ-value, year of construction and house type have the most impact here and reveal, together with other housing-related factors, a large part of the inequality in the Dutch residential energy consumption and energy affordability.

6.3 Implications

Societal Relevance Energy Consumption and Affordability

This study reveals how the housing-related factors influence the residential energy consumption and energy affordability in the Netherlands. It is critical to know the predictors of the energy consumption in the light of the energy transition, which is why energy efficiency is high on the policy agenda of the Dutch government (Vega et al., 2022). The same holds for the predictors of the energy affordability, as the inequality in the energy affordability is increasing (Christophers, 2019; Galvin & Sunikka-Blank, 2018). Next to that the energy consumption is rising due to COVID-19 (Krarti & Aldubyan, 2021), and the energy prices are soaring (Tesio et al., 2022; CPB, 2022), which puts the energy affordability even higher on the policy agenda.

Academic Implications Rental Houses and Urbanity

When comparing the results of this study with previous literature, several new insights arise. First, more clarity is provided on the relation of rental houses with the energy consumption and energy affordability. Although previous research was contradicting with regard to energy consumption (Santin et al., 2009; Huang, 2015), the regression clearly showed that rental houses are consuming less energy in the Netherlands. However, no relevant relation was found with energy affordability, even though the literature suggested that rental households have a higher energy expenditure to income ratio (Hoekstra, 2017; Haffner & Boumeester, 2015; Belaïd, 2018). The fact that rental households have a lower energy consumption and a similar energy affordability when compared to owner-occupied households suggests that rental households have a lower energy consumption because of their lower income and wealth, which is confirmed by Riva et al. (2021). Second, the regression revealed that contrary to other countries, the urbanity has no significant impact on the energy consumption and energy affordability in the Netherlands.

Academic Implications WOZ-Value and Construction Year

Third, this research adds value to the existing body of literature by revealing a clear positive relation between WOZ-value or property value and the energy consumption and energy affordability. The literature on this factor was limited, but the regressions showed that a higher WOZ-value results in more energy consumption and a higher expenditure to income ratio. Fourth, the negative relation found between both the energy consumption and the energy affordability, and the year of construction is considerably stronger, when compared to Santin et al. (2009) and Besagni and Borgarello (2018). This means that the impact of the year of construction on the energy consumption and energy affordability has increased over the past years. The results show that this is due to the fact that recent houses build after 2010 have a significantly lower energy consumption and energy affordability. Thus, this study reveals that the differences in energy consumption and energy affordability based on the year of construction of a dwelling are growing within the Netherlands.

Unique Dataset

This thesis also adds to the existing literature because it builds on energy consumption data on the vast majority of Dutch households. This increases the validity of the claims made in this research when compared to previous research, as previous research in the Netherlands only included samples of households (Abrahamse & Steg, 2009; Brounen et al., 2012; Santin et al., 2009). Therefore, this study also validates the results from previous research.

Academic Implications Decreasing Affordability

Finally, this research confirms that the energy affordability is decreasing, which was already suggested by Tesio et al. (2022) and Christophers (2019). The additional affordability analysis reveals that the energy expenditure to income ratio has increased from 6,16% between 2016 and 2018, to 7,31% based on the October 2021 prices and 11,32% based on the April 2022 prices. Thus, this thesis shows that the energy affordability is decreasing in the Netherlands, mainly due to the high energy prices, and that the amount of energy poor households, which are categorized as households that spend more than 10% of their income on energy, is growing. This increases the relevance of this study even more as it highlights the impact of the predictors of the energy consumption and energy affordability.

Current Policy Limitations

When considering the policy relevance, it should be noted that there are a large amount of residential energy saving policies in the Netherlands that impact the energy consumption and energy affordability. Among others, households can make use of the Sustainable Energy Investment Subsidy (ISDE), the Homeowner Energy Saving Subsidy (SEEH) and the National Energy Saving Fund (NEF) to improve their energy efficiency. The number of households that makes use of these policies is however relatively low, as the transaction costs for implementing the policies are high due to the many arrangements that need to be made for such a transition (Vega et al., 2022). Next to this, low-income households can't make use of these policies, because they require an initial investment, which increases the inequality in energy efficiency (Riva et al., 2021; Mashhoodi et al., 2021). There are also policies for landlords to improve the energy efficiency of rental houses, such as the Energy Saving Fund for the Rental Sector (FEH) and the Energy Performance Incentive Scheme for the Rental Sector (STEP). The incentive for landlords to make use of these policies is however still low, which results in rental houses having a worse energy efficiency (Riva et al., 2021; Vega et al., 2022).

Policy Implications based on Year of Construction

This research confirms that the energy efficiency policies aren't resulting in a lower energy consumption, as the average energy consumption in this study is relatively similar to that observed by Abrahamse & Steg (2009) and Brounen et al. (2012). Furthermore, the differences in energy consumption and energy affordability are increasing based on the year of construction, which means that older houses are not improving their energy efficiency. Vega et al. (2022) argue that the financial barriers for the Dutch energy efficiency policies should be lowered in order to make the energy efficiency policies effective and reduce the energy consumption and energy expenditures. Based on this study it can be stated that the year of construction should be a policy criterion here, as older houses have a lower energy efficiency. Figure 17 showed that houses built before 1980 have a significantly higher energy consumption when compared to newer houses, suggesting that these houses should be targeted first. Mashhoodi (2021) even reveals that the year of construction and the lower building quality of houses are the main reasons that low-income households experience energy poverty. This means that when targeting the year of construction, policies will also improve the energy efficiency of low-income households, which increases the policy effectiveness.

Policy Implications for Rental Houses
Furthermore, this research substantiates that rental houses should deserve extra attention in Dutch energy saving policies. The fact that rental households spend relatively more on energy than homeowners, given their lower energy consumption and similar energy expenditure to income ratio, means that they are likely to struggle with the energy affordability (Riva et al., 2021), even though the regression on affordability only doesn't show this. When combining this knowledge with the lower building quality of rental houses and the low number of landlords that make use of the energy efficiency policies for rental houses (Vega et al., 2022; Riva et al., 2021) it can be concluded that policies need to be altered in order to improve the energy efficiency of rental houses. Incentive should be created for landlords to make use of the energy efficiency policies, and increase the energy affordability of tenants.

Policy Implications for Other Housing-Related Factors
Although energy policies in other countries are often based on the urbanity and location of households, this is not the case in the Netherlands (Vega et al., 2022). This study supports this decision, as the urbanity isn't an important predictor for energy consumption and energy affordability in the Netherlands. The house type, surface area and WOZ-value do have a strong relation with the energy consumption and energy affordability, but are generally not used for policy design because of their correlations with other variables such as income (Besagni & Borgarello, 2018; Riva et al., 2021; Vega et al., 2022). Nevertheless, the WOZ-value could be an interesting measure to consider in policy design, as both this research and external literature suggests that the energy efficiency of houses with a higher WOZ-value is significantly worse (Van Middelkoop et al., 2017; Boesveld, 2021). Further research on the nature of this relation and the correlations of the WOZ-value with other variables is however needed here before clear conclusions can be drawn.

Policy Implications for Low-Income Households
Next to the housing-related factors, it is critical to note that the income also has a strong relation with the energy consumption and especially the energy affordability. Thus, the income should be an important factor in policy design too. Riva et al. (2021) and Mashhoodi (2021) support this by showing that low-income households aren't able to leverage the energy efficiency policies and invest in energy savings, which decreases their energy affordability. Vega et al. (2022) even state that the current Dutch energy policies are increasing the energy poverty among low-income households, as they do pay energy taxes but can't invest in energy efficiency. New energy efficiency policies should thus aim to remove these financial barriers for low-income households. Moreover, this study reveals that the soaring energy prices are increasing the percentage of income that households spend on energy, which makes it even more clear that additional energy efficiency and energy affordability policies are needed for low-income households.

Implications for Resource Planning and Energy Price Policy
On a more general level, this thesis also provides an overview of the factors impacting the energy consumption and energy affordability. Knowing these factors and their impact can help to project the future energy demand, which can improve the resource planning of the Dutch network operators and energy providers (Brounen et al., 2012). Besides, the knowledge of the energy affordability predictors can help to project the future energy affordability, which is useful for setting policy targets in this field.

Furthermore, the additional affordability analysis based on the energy prices of October 2021 and April 2022 revealed that the energy prices currently also have a large influence on the energy affordability. Reducing the soaring energy prices should therefore also be a core policy objective. This can be achieved through improving the security of the gas supply or through financial compensation and tax adjustments (Tesio et al., 2022). This study showed that current energy prices can push a large percentage of the Dutch households into energy poverty, which makes this a crucial short-term policy target.

6.4 Limitations & Future Research

Although this study provides extensive insights in the factors impacting residential energy consumption and energy affordability in the Netherlands, it also has its limitations, which leave room for future research.

Data Limitations

First, there are the limitations related to the data. Even though using postal code 6 data allows the inclusion of the majority of households in the Netherlands in the research, it also means that the predictors could not be assessed on a household level. This resulted in predictors that capture less information or cover a wider range of cases. An example of this are the percentual variables such as the percentage of detached houses, as a certain percentage of detached houses could be either combined with a high percentage of apartments or a high percentage of terraced houses. This information is however not captured in the percentage of detached houses, which increases the uncertainty within this variable. Next to that, aggregating the data on a postal code level reduces the variance in energy consumption and energy affordability, as the variety within a postal code is averaged out (Abrahamse & Steg, 2009). This information loss could be reduced by including a sample of household-level data in the analysis to check how the energy consumption and energy affordability are distributed within a postal code, which is therefore recommended for future research.

Another limitation is in the income data, which was only available on the Dutch national income decile level per postal code, due to privacy sensitivity. The income for each postal code was calculated based on the national average disposable income per decile, which resulted in a lower data quality, as differences can be significant within an income group. A third limitation in this regard was the lack of data on the income and energy consumption for October 2021 and April 2022, which created the need for extrapolation in the additional affordability analysis. As extrapolation builds on assumptions which have its limitations as well, this reduced the reliability of the results (Armstrong & Collopy, 1993). Nevertheless, this is still considered the best available alternative, as the data on the recent energy price increase and its impact on the energy consumption is currently limited.

Methodological Limitations

Second, the methodology also has its limitations. The multivariate regression used builds on different assumptions, and one of these is that it assumes a linear relationship between the independent variables and the dependent variable (Jeon, 2015). This relationship is however not always linear, as figure 16 and 17 reveal. Sometimes the relation exists only for a certain range of the values or is logarithmic or quadratic. In this study, linear relations have been assumed, although linearity was also checked for the most relevant housing-related factors by assessing the literature and the histograms of these variables. Only the surface area in square meters was included as a non-linear variable, as previous research found this relation to be logarithmic (Brounen et al., 2012; Iraganboina & Eluru, 2021). Nevertheless, it could be that the relationships of some other variables with the energy consumption or energy affordability are logarithmic or quadratic, which is for example the case in the analysis of Besagni & Borgarello (2018) and Sanquist et al. (2012). Thus, it is interesting to deepen this research by assessing the nature of the relation between the independent and the dependent variables, in order to improve the validity of the regression analysis.

Next to this, the regression analyses also assumes that the error terms of the different cases included are completely independent from each other (Jeon, 2015). This is not the case in this research, as the same postal codes are included for 3 different years. Because the cases with the same postal code 6 correlate heavily with each other, the error terms are not independent, which results in the standard errors being lower. The impact of this limitation is however low, as this is not expected to impact the validity of the regression coefficients and because the errors are not considered a critical metric in this research.

Another limitation of the regression analysis is that a large number of variables included in the analysis capture similar value, as they correlate with each other. An example of this is the strong relationship between rental houses and apartments. Feijten and Mulder (2002) show that most rental houses in the Netherlands are apartments, and that the majority of the apartments are rental houses. In the regression analysis however, this shared explaining power can only be captured in one of the variables. Kavousian et al. (2013) confirm this by showing that when analysing the residential energy consumption, the regression coefficients change when excluding a variable, as the geographical location also captures information on the building type and as the surface area has a strong correlation with the income. Future research could control for this limitation by assessing the correlations between the predictor variables, and including interaction-variables that capture the shared explaining power of variables. Additional qualitative research could also add value here, by providing insight in the mechanisms through which factors impact the energy consumption and energy affordability.

Moreover, the methodology for assessing the energy affordability also has its limitations. The energy expenditure to income ratio is one of the most complete indicators of energy affordability (Kontokosta et al., 2019; Papada & Kaliampakos; Mashhoodi et al., 2018), but doesn't capture all the dimensions of this concept. One of the main drawbacks of this method is that it doesn't consider non-energy costs that can be restricting for a households' energy consumption pattern, such as housing costs (Charlier & Kahouli, 2019). Other measures, such as the Low-Income/High-Costs indicator and the After-Fuel-Costs-Poverty (Charlier & Kahouli, 2019), do take this into account, and could be combined with the energy expenditure to income ratio to get a better overview of the energy affordability in the Netherlands in future studies.

Scoping Limitations

Third, there are the limitations related to the scope of the study, which leave room for further research. This study is performed in the Netherlands, which has a lot of similarities to other Western European countries in terms of socio-demographics and energy use (Neagu & Teodoru, 2019). In some of these countries, like the UK, extensive research has been done on the energy consumption and energy affordability, but in other countries, such as Germany and France, research on this is relatively limited (Baker & Rylatt, 2008; Chalal et al., 2017; Jones & Lomas, 2015). It would thus be interesting to check whether the relations found in this study also hold for the other Western European countries, as insight in the factors impacting the energy consumption and energy affordability could be critical for future policies in these countries, especially in the current energy transition.

Furthermore, the change in energy consumption and energy affordability over time would be an intriguing topic for future research, given the increasing energy prices (Tesio et al., 2022; CPB, 2022) and the energy transition (Gillingham & Stock, 2018). This research has revealed that the average residential energy expenditure to income ratio is increasing, which raises multiple questions. How does this price increase impact the predictors of the energy consumption, how does it influence the average residential energy consumption and energy affordability, and how will the residential energy consumption change once the prices have stabilized, taking into account the energy transition? The answers to these questions are for a large part dependent on the price elasticity, which is unpredictable during these abnormal price variations (Labandeira et al., 2017). Therefore, it would also be interesting to include the price elasticity in the research, when answering these questions.

There are also some limitations regarding the dependent and independent variables used in the regressions in this study. The dependent variables only include the gas and electricity consumption in the Netherlands. Houses with district heating are left out of the analysis, and other energy sources such as firewood are not considered. Studies in other countries do include these type of energy sources and reveal that they can have a significant impact on the energy consumption for some households (Iraganaboina & Eluru, 2021; Besagni & Borgarello, 2018), making this a promising area of future research. The independent variables in this study range from socio-demographics and weather variables to housing-related factors. Although this are the factors with the highest expected impact on the energy consumption and energy affordability, other factors could still be added to improve the regression model. The use and ownership of appliances, such as a dishwasher or AC, has shown to be an important predictor for the electricity consumption (Khan, 2019; Sanquist et al., 2012; Huebner et al., 2016). Next to that, more specific housing-related factors such as the insulation, whether there is double glazing and the number of rooms were found to be helpful in explaining the energy consumption used for heating (Iraganaboina & Eluru, 2021; Santin et al., 2009). Adding these variables could thus improve the accuracy of the regression model and increase the insight in the determinants of the residential energy consumption and energy affordability.

Moreover, additional research on effective policy design is needed, with regard to the factors that should be included in energy efficiency and energy poverty policy according to this study. This thesis revealed that policy redesign is needed for rental houses and houses with an older year of construction, but didn't touch on the actual design of these policies. Given that current Dutch energy policies are underperforming, it is critical to find out how the effectiveness of future policies can be increased, and how the rental houses and the year of construction should be included in these policies (Vega et al., 2022). As mentioned, the WOZ-value could also be an interesting factor in policy design. More research is however needed on the nature of its relation with the energy consumption and affordability, on its correlations with other factors such as income, and on its potential effectiveness in energy policy, before determining whether this factor should be used in designing future energy policies.

A final limitation in the energy affordability analyses is that this study only considers a financial indicator for the energy affordability in the form of the expenditure to income approach. This means that the specific energy needs of households and the energy reduction of households who can't afford the energy they need is not taken into account (Riva et al., 2021). In order to get a more balanced overview of the residential energy affordability in the Netherlands, future research should include subjective affordability measures based on personal opinions and judgement, which can for example be gathered using surveys (Charlier & Kahouli, 2019). When combined with the expenditure to income approach, these subjective measures can provide a more in-depth overview of the residential energy affordability.

Bibliography

Abidoye, R. B., & Chan, A. P. (2016). Critical determinants of residential property value: professionals' perspective. *Journal of Facilities Management*, 14(3), 283–300. https://doi.org/10.1108/jfm-02-2016-0003

Abrahamse, W., & Steg, L. (2009). How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? *Journal of Economic Psychology*, 30(5), 711–720. https://doi.org/10.1016/j.joep.2009.05.006

Armstrong, J. S., & Collopy, F. (1993). Causal forces: Structuring knowledge for time-series extrapolation. *Journal of Forecasting*, 12(2), 103–115. https://doi.org/10.1002/for.3980120205

Baker, K. J., & Rylatt, R. M. (2008). Improving the prediction of UK domestic energy-demand using annual consumption-data. *Applied Energy*, 85(6), 475–482. https://doi.org/10.1016/j.apenergy.2007.09.004

Bardazzi, R., Bortolotti, L., & Pazienza, M. G. (2021). To eat and not to heat? Energy poverty and income inequality in Italian regions. *Energy Research & Social Science*, 73, 101946. https://doi.org/10.1016/j.erss.2021.101946

Bedir, M., Hasselaar, E., & Itard, L. (2013). Determinants of electricity consumption in Dutch dwellings. *Energy and Buildings*, 58, 194–207. https://doi.org/10.1016/j.enbuild.2012.10.016

Belaïd, F. (2018). Exposure and risk to fuel poverty in France: Examining the extent of the fuel precariousness and its salient determinants. *Energy Policy*, 114, 189–200. https://doi.org/10.1016/j.enpol.2017.12.005

Besagni, G., & Borgarello, M. (2018). The determinants of residential energy expenditure in Italy. *Energy*, 165, 369–386. https://doi.org/10.1016/j.energy.2018.09.108

Besagni, G., & Borgarello, M. (2019a). Measuring Fuel Poverty in Italy: A Comparison between Different Indicators. *Sustainability*, 11(10), 2732. https://doi.org/10.3390/su11102732

Besagni, G., & Borgarello, M. (2019b). The socio-demographic and geographical dimensions of fuel poverty in Italy. *Energy Research & Social Science*, 49, 192–203. https://doi.org/10.1016/j.erss.2018.11.007

Boesveld, M. (2021). *7/8 van de gemeenten met hoogste WOZ-waarden hebben slecht energielabel*. Vastelastenbond. Retrieved 15 July 2022, from https://www.vastelastenbond.nl/blog/78-van-de-gemeenten-met-hoogste-woz-waarden-hebben-slecht-energielabel/

Borozan, D. (2018). Regional-level household energy consumption determinants: The european perspective. *Renewable and Sustainable Energy Reviews*, 90, 347–355. https://doi.org/10.1016/j.rser.2018.03.038

Boukarta, S., & Berezowska-Azzag, E. (2018). Assessing Households' Gas and Electricity Consumption: A Case Study of Djelfa, Algeria. *Quaestiones Geographicae*, 37(4), 111–129. https://doi.org/10.2478/quageo-2018-0034

Brounen, D., Kok, N., & Quigley, J. M. (2012). Residential energy use and conservation: Economics and demographics. *European Economic Review*, 56(5), 931–945. https://doi.org/10.1016/j.euroecorev.2012.02.007

CBS. (2016–2018). *Kerncijfers per postcode-6* [Dataset]. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode

CBS. (2016–2020). *Inkomen van huishoudens; inkomensklassen, huishoudenskenmerken* [Dataset]. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/cijfers/detail/83932NED

CBS. (2016–2021). *Aardgas en elektriciteit, gemiddelde prijzen van eindverbruikers* [Dataset]. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/cijfers/detail/81309NED

CBS. (2018–2022). *Gemiddelde energietarieven voor consumenten* [Dataset]. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/cijfers/detail/84672NED

CBS. (2022, April 19). *Population dashboard*. Retrieved 20 June 2022, from https://www.cbs.nl/en-gb/visualisations/dashboard-population

Chalal, M. L., Benachir, M., White, M., Shahtahmassebi, G., Cumberbatch, M., & Shrahily, R. (2017). The impact of the UK household life-cycle transitions on the electricity and gas usage patterns. *Renewable and Sustainable Energy Reviews*, 80, 505–518. https://doi.org/10.1016/j.rser.2017.05.222

Chapman, A., & Okushima, S. (2019). Engendering an inclusive low-carbon energy transition in Japan: Considering the perspectives and awareness of the energy poor. *Energy Policy*, 135, 111017. https://doi.org/10.1016/j.enpol.2019.111017

Charlier, D., & Kahouli, S. (2019). From Residential Energy Demand to Fuel Poverty: Income-induced Non-linearities in the Reactions of Households to Energy Price Fluctuations. *The Energy Journal*, 40(2). https://doi.org/10.5547/01956574.40.2.dcha

Chen, J., Wang, X., & Steemers, K. (2013). A statistical analysis of a residential energy consumption survey study in Hangzhou, China. *Energy and Buildings*, 66, 193–202. https://doi.org/10.1016/j.enbuild.2013.07.045

Cheng, V., & Steemers, K. (2011). Modelling domestic energy consumption at district scale: A tool to support national and local energy policies. *Environmental Modelling & Software*, 26(10), 1186–1198. https://doi.org/10.1016/j.envsoft.2011.04.005

Christophers, B. (2019). A tale of two inequalities: Housing-wealth inequality and tenure inequality. *Environment and Planning A: Economy and Space*, 53(3), 573–594. https://doi.org/10.1177/0308518x19876946

CPB. (2022, march). *Centraal Economisch Plan 2022*. Centraal Planbureau. https://www.cpb.nl/sites/default/files/omnidownload/Centraal-Economisch-Plan-CEP-2022.pdf

Das, R. R., Martiskainen, M., & Li, G. (2022). Quantifying the prevalence of energy poverty across Canada: Estimating domestic energy burden using an expenditures approach. *The Canadian Geographer / Le Géographe canadien*. https://doi.org/10.1111/cag.12750

De Arce, R., & Mahía, R. (2019). Drivers of Electricity Poverty in Spanish Dwellings: A Quantile Regression Approach. *Energies*, 12(11), 2089. https://doi.org/10.3390/en12112089

Dekker, T. J. (2014). *Invloed van de energetische prestatie op de woningwaarde*. Delft University of Technology - Faculty of Architecture. https://repository.tudelft.nl

Dong, F., Yu, B., Hua, Y., Zhang, S., & Wang, Y. (2018). A Comparative Analysis of Residential Energy Consumption in Urban and Rural China: Determinants and Regional Disparities. *International*

Journal of Environmental Research and Public Health, 15(11), 2507. https://doi.org/10.3390/ijerph15112507

Feijten, P., & Mulder, C. H. (2002). The Timing of Household Events and Housing Events in the Netherlands: A Longitudinal Perspective. *Housing Studies*, 17(5), 773–792. https://doi.org/10.1080/0267303022000009808

Frederiks, E., Stenner, K., & Hobman, E. (2015). The Socio-Demographic and Psychological Predictors of Residential Energy Consumption: A Comprehensive Review. *Energies*, 8(1), 573–609. https://doi.org/10.3390/en8010573

Fumo, N., & Rafe Biswas, M. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews*, 47, 332–343. https://doi.org/10.1016/j.rser.2015.03.035

Galvin, R., & Sunikka-Blank, M. (2018). Economic Inequality and Household Energy Consumption in High-income Countries: A Challenge for Social Science Based Energy Research. *Ecological Economics*, 153, 78–88. https://doi.org/10.1016/j.ecolecon.2018.07.003

Gillingham, K., & Stock, J. H. (2018). The Cost of Reducing Greenhouse Gas Emissions. *Journal of Economic Perspectives*, 32(4), 53–72. https://doi.org/10.1257/jep.32.4.53

Guo, Z., Zhou, K., Zhang, C., Lu, X., Chen, W., & Yang, S. (2018). Residential electricity consumption behavior: Influencing factors, related theories and intervention strategies. *Renewable and Sustainable Energy Reviews*, 81, 399–412. https://doi.org/10.1016/j.rser.2017.07.046

Haffner, M., & Boumeester, H. (2015). Housing affordability in the Netherlands: the impact of rent and energy costs. *Journal of Housing and the Built Environment*, 30(2), 293–312. https://doi.org/10.1007/s10901-014-9409-2

Hartono, D., Hastuti, S. H., Balya, A. A., & Pramono, W. (2020). Modern energy consumption in Indonesia: Assessment for accessibility and affordability. *Energy for Sustainable Development*, 57, 57–68. https://doi.org/10.1016/j.esd.2020.05.002

Hoekstra, J. (2017). Reregulation and Residualization in Dutch Social Housing: A Critical Evaluation of New Policies. *Critical Housing Analysis*, 4(1), 31–39. https://doi.org/10.13060/23362839.2017.4.1.322

Huang, W. H. (2015). The determinants of household electricity consumption in Taiwan: Evidence from quantile regression. *Energy*, 87, 120–133. https://doi.org/10.1016/j.energy.2015.04.101

Huebner, G., Shipworth, D., Hamilton, I., Chalabi, Z., & Oreszczyn, T. (2016). Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes. *Applied Energy*, 177, 692–702. https://doi.org/10.1016/j.apenergy.2016.04.075

IBM. (2020). *Pairwise vs. Listwise deletion: What are they and when should I use them?* Retrieved 23 July 2022, from https://www.ibm.com/support/pages/pairwise-vs-listwise-deletion-what-are-they-and-when-should-i-use-them

Iraganaboina, N. C., & Eluru, N. (2021). An examination of factors affecting residential energy consumption using a multiple discrete continuous approach. *Energy and Buildings*, 240, 110934. https://doi.org/10.1016/j.enbuild.2021.110934

Ismail, Z., & Khembo, P. (2015). Determinants of energy poverty in South Africa. *Journal of Energy in Southern Africa*, 26(3), 66–78. https://doi.org/10.17159/2413-3051/2015/v26i3a2130

Jeon, J. (2015). The Strengths and Limitations of the Statistical Modeling of Complex Social Phenomenon: Focusing on SEM, Path Analysis, or Multiple Regression Models. *International Journal of Economics and Management Engineering*, 9(5), 1634–1642.

Jiang, Y., Zhang, L., & Zhang, J. (2019). Energy consumption by rural migrant workers and urban residents with a hukou in China: quality-of-life-related factors and built environment. *Natural Hazards*, 99(3), 1431–1453. https://doi.org/10.1007/s11069-019-03802-1

Johnson, B. (2001). Toward a New Classification of Nonexperimental Quantitative Research. *Educational Researcher*, 30(2), 3–13.

Jones, R. V., Fuertes, A., & Lomas, K. J. (2015). The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. *Renewable and Sustainable Energy Reviews*, 43, 901–917. https://doi.org/10.1016/j.rser.2014.11.084

Jones, R. V., & Lomas, K. J. (2015). Determinants of high electrical energy demand in UK homes: Socio-economic and dwelling characteristics. *Energy and Buildings*, 101, 24–34. https://doi.org/10.1016/j.enbuild.2015.04.052

Karásek, J., & Pojar, J. (2018). Programme to reduce energy poverty in the Czech Republic. *Energy Policy*, 115, 131–137. https://doi.org/10.1016/j.enpol.2017.12.045

Kavousian, A., Rajagopal, R., & Fischer, M. (2013). Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants’ behavior. *Energy*, 55, 184–194. https://doi.org/10.1016/j.energy.2013.03.086

Khan, I. (2019). Household factors and electrical peak demand: a review for further assessment. *Advances in Building Energy Research*, 15(4), 409–441. https://doi.org/10.1080/17512549.2019.1575770

Khezri, R., Mahmoudi, A., & Haque, M. H. (2020). Optimal Capacity of Solar PV and Battery Storage for Australian Grid-Connected Households. *IEEE Transactions on Industry Applications*, 56(5), 5319–5329. https://doi.org/10.1109/tia.2020.2998668

Khundi-Mkomba, F., Kumar Saha, A., & Wali, U. G. (2021). Examining the state of energy poverty in Rwanda: An inter-indicator analysis. *Heliyon*, 7(11), 08441. https://doi.org/10.1016/j.heliyon.2021.e08441

KNMI. (2016–2018). *KNMI - Daggegevens van het weer in Nederland* [Dataset]. Koninklijk Nederlands Meteorologisch Instituut. https://www.knmi.nl/nederland-nu/klimatologie/daggegevens

Kontokosta, C. E., Reina, V. J., & Bonczak, B. (2019). Energy Cost Burdens for Low-Income and Minority Households. *Journal of the American Planning Association*, 86(1), 89–105. https://doi.org/10.1080/01944363.2019.1647446

Krarti, M., & Aldubyan, M. (2021). Review analysis of COVID-19 impact on electricity demand for residential buildings. *Renewable and Sustainable Energy Reviews*, 143. https://doi.org/10.1016/j.rser.2021.110888

Labandeira, X., Labeaga, J. M., & López-Otero, X. (2017). A meta-analysis on the price elasticity of energy demand. *Energy Policy*, 102, 549–568. https://doi.org/10.1016/j.enpol.2017.01.002

Li, L., Yao, Y., Yang, R., & Zhou, K. (2018). Is it more effective to bring time-of-use pricing into increasing block tariffs? Evidence from evaluation of residential electricity price policy in Anhui province. *Journal of Cleaner Production*, 181, 703–716. https://doi.org/10.1016/j.jclepro.2018.01.209

Lin, B., & Wang, Y. (2020). Does energy poverty really exist in China? From the perspective of residential electricity consumption. *Energy Policy*, 143. https://doi.org/10.1016/j.enpol.2020.111557

Longhi, S. (2015). Residential energy expenditures and the relevance of changes in household circumstances. *Energy Economics*, 49, 440–450. https://doi.org/10.1016/j.eneco.2015.03.018

Lund, T. (2021). Research Problems and Hypotheses in Empirical Research. *Scandinavian Journal of Educational Research*, 1–11. https://doi.org/10.1080/00313831.2021.1982765

Mashhoodi, B. (2021). Who is more dependent on gas consumption? Income, gender, age, and urbanity impacts. *Applied Geography*, 137, 102602. https://doi.org/10.1016/j.apgeog.2021.102602

Mashhoodi, B., Stead, D., & Van Timmeren, A. (2018). Spatial homogeneity and heterogeneity of energy poverty: a neglected dimension. *Annals of GIS*, 25(1), 19–31. https://doi.org/10.1080/19475683.2018.1557253

McLoughlin, F., Duffy, A., & Conlon, M. (2012). Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. *Energy and Buildings*, 48, 240–248. https://doi.org/10.1016/j.enbuild.2012.01.037

Miah, M., Foysal, M. A., Koike, M., & Kobayashi, H. (2011). Domestic energy-use pattern by the households: A comparison between rural and semi-urban areas of Noakhali in Bangladesh. *Energy Policy*, 39(6), 3757–3765. https://doi.org/10.1016/j.enpol.2011.04.004

Ndiaye, D., & Gabriel, K. (2011). Principal component analysis of the electricity consumption in residential dwellings. *Energy and Buildings*, 43(2–3), 446–453. https://doi.org/10.1016/j.enbuild.2010.10.008

Neagu, O., & Teodoru, M. (2019). The Relationship between Economic Complexity, Energy Consumption Structure and Greenhouse Gas Emission: Heterogeneous Panel Evidence from the EU Countries. *Sustainability*, 11(2), 497. https://doi.org/10.3390/su11020497

Nguyen, T. T., Hoang, V. N., Wilson, C., & Managi, S. (2019). Energy transition, poverty and inequality in Vietnam. *Energy Policy*, 132, 536–548. https://doi.org/10.1016/j.enpol.2019.06.001

NLExtract. (2021–2022). *BAG-CBS - Adressen - Woning* [Dataset]. GeoCatalogus. https://geocatalogus.nl/tl/dataset/bag-cbs-adressen-woning

Nunes, C. A., Alvarenga, V. O., De Souza Sant'Ana, A., Santos, J. S., & Granato, D. (2015). The use of statistical software in food science and technology: Advantages, limitations and misuses. *Food Research International*, 75, 270–280. https://doi.org/10.1016/j.foodres.2015.06.011

Papada, L., & Kaliampakos, D. (2017). Energy poverty in Greek mountainous areas: a comparative study. *Journal of Mountain Science*, 14(6), 1229–1240. https://doi.org/10.1007/s11629-016-4095-z

Peugh, J. L., & Enders, C. K. (2004). Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement. *Review of Educational Research*, 74(4), 525–556. https://doi.org/10.3102/00346543074004525

Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and Limitations of Qualitative and Quantitative Research Methods. *European Journal of Education Studies*, 3(9), 369–387. https://doi.org/10.5281/zenodo.887089

Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F., & Webber, M. E. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, 135, 461–471. https://doi.org/10.1016/j.apenergy.2014.08.111

Riva, M., Kingunza Makasi, S., Dufresne, P., O'Sullivan, K., & Toth, M. (2021). Energy poverty in Canada: Prevalence, social and spatial distribution, and implications for research and policy. *Energy Research & Social Science*, 81, 102237. https://doi.org/10.1016/j.erss.2021.102237

RVO. (2017–2021). *Warmtenetten* [Aantal huishoudens die aangesloten zijn op een warmtenet per buurt]. Rijksdienst voor Ondernemend Nederland. https://nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/75f196bd-6fb2-4939-99a5-cd40d753183d?tab=general

Sanquist, T. F., Orr, H., Shui, B., & Bittner, A. C. (2012). Lifestyle factors in U.S. residential electricity consumption. *Energy Policy*, 42, 354–364. https://doi.org/10.1016/j.enpol.2011.11.092

Santin, O. G., Itard, L., & Visscher, H. (2009). The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. *Energy and Buildings*, 41(11), 1223–1232. https://doi.org/10.1016/j.enbuild.2009.07.002

Steemers, K., & Yun, G. Y. (2009). Household energy consumption: a study of the role of occupants. *Building Research & Information*, 37(5–6), 625–637. https://doi.org/10.1080/09613210903186661

ŠTreimikienė, D. (2014). Residential energy consumption trends, main drivers and policies in Lithuania. *Renewable and Sustainable Energy Reviews*, 35, 285–293. https://doi.org/10.1016/j.rser.2014.04.012

Sukamolson, S. (2007). Fundamentals of quantitative research. Language Institute Chulalongkorn University, 1(3), 1–20.

Sun, C., Zhang, Y., Peng, S., & Zhang, W. (2015). The inequalities of public utility products in China: From the perspective of the Atkinson index. *Renewable and Sustainable Energy Reviews*, 51, 751–760. https://doi.org/10.1016/j.rser.2015.06.055

Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8), 1819–1835. https://doi.org/10.1016/j.rser.2008.09.033

Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85, 208–220. https://doi.org/10.1016/j.energy.2015.03.084

Tabata, T., & Tsai, P. (2020). Fuel poverty in Summer: An empirical analysis using microdata for Japan. *Science of The Total Environment*, 703, 135038. https://doi.org/10.1016/j.scitotenv.2019.135038

Tambach, M., & Visscher, H. (2012). Towards Energy-neutral New Housing Developments. Municipal Climate Governance in The Netherlands. *European Planning Studies*, 20(1), 111–130. https://doi.org/10.1080/09654313.2011.638492

Tesio, E., Conti, I., & Cervigni, G. (2022). High gas prices in Europe: a matter for policy intervention? *Policy Brief*, 6, 1–7. https://doi.org/10.2870/260985

Tso, G. K., & Guan, J. (2014). A multilevel regression approach to understand effects of environment indicators and household features on residential energy consumption. *Energy*, 66, 722–731. https://doi.org/10.1016/j.energy.2014.01.056

Van Middelkoop, M., Vringer, K., & Visser, H. (2017). Are Dutch residents ready for a more stringent policy to enhance the energy performance of their homes? *Energy Policy*, 105, 269–282. https://doi.org/10.1016/j.enpol.2017.02.050

Van Middelkoop, M., Van Polen, S., Holtkamp, R., & Bonnerman, F. (2018). *Meten Met Twee Maten: Een studie naar de betaalbaarheid van de energierekening van huishoudens*. Planbureau voor de Leefomgeving. https://www.pbl.nl/sites/default/files/downloads/pbl-2018-de-betaalbaarheid-van-de-energierekening-van-huishoudens-3124.pdf

Vega, S. H., van Leeuwen, E., & van Twillert, N. (2022). Uptake of residential energy efficiency measures and renewable energy: Do spatial factors matter? *Energy Policy*, 160, 112659. https://doi.org/10.1016/j.enpol.2021.112659

Visscher, H., Meijer, F., Majcen, D., & Itard, L. (2016). Improved governance for energy efficiency in housing. *Building Research & Information*, 44(5–6), 552–561. https://doi.org/10.1080/09613218.2016.1180808

Wang, F., Li, K., Duić, N., Mi, Z., Hodge, B. M., Shafie-khah, M., & Catalão, J. P. (2018). Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns. *Energy Conversion and Management*, 171, 839–854. https://doi.org/10.1016/j.enconman.2018.06.017

Wei, T., Zhu, Q., & Glomsrød, S. (2014). Energy spending and household characteristics of floating population: Evidence from Shanghai. *Energy for Sustainable Development*, 23, 141–149. https://doi.org/10.1016/j.esd.2014.07.001

Wijaya, T. K., Vasirani, M., Humeau, S., & Aberer, K. (2015). Cluster-based aggregate forecasting for residential electricity demand using smart meter data. *2015 IEEE International Conference on Big Data (Big Data)*. https://doi.org/10.1109/bigdata.2015.7363836

Xie, L., Yan, H., Zhang, S., & Wei, C. (2020). Does urbanization increase residential energy use? Evidence from the Chinese residential energy consumption survey 2012. *China Economic Review*, 59, 101374. https://doi.org/10.1016/j.chieco.2019.101374

Yohanis, Y. G., Mondol, J. D., Wright, A., & Norton, B. (2008). Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use. *Energy and Buildings*, 40(6), 1053–1059. https://doi.org/10.1016/j.enbuild.2007.09.001

Zheng, X., Wei, C., Qin, P., Guo, J., Yu, Y., Song, F., & Chen, Z. (2014). Characteristics of residential energy consumption in China: Findings from a household survey. *Energy Policy*, 75, 126–135. https://doi.org/10.1016/j.enpol.2014.07.016

Appendix A: Data Preparation

Including District Heating Values Below 25%

Most postal codes with district heating are excluded from the analyses in this research, as this research focusses on the total energy consumption of Dutch households and as energy consumption data is only available for the gas and electricity consumption. Including households with district heating would thus create a bias, as the measured total energy consumption would not equal the actual total energy consumption. However, postal codes in neighbourhoods in which less than 25% of the households are connected to district heating are included in the analyses. The first reason for this is that a relatively large number of households belong to a postal code in which less than 25% of the households has district heating. The district heating histogram in figure 19 shows that more than 300.000 households fall in this category. The second reason for including this data is that a large majority of the postal codes in a neighbourhood with less than 25% district heating have an average energy consumption above 12.000 kWh per year, which corresponds to an orange or red colour in figure 19. As energy consumption levels above 12.000 kWh per year are common among Dutch households, also if they have 0% of district heating, this confirms that a large majority of these households have a normal energy consumption, comparable with other Dutch households. Thus, the postal codes with less than 25% of houses with district heating are present in the final dataset. Nevertheless, a dummy is used in the analyses performed to check whether these postal codes differ significantly from the rest of the Dutch population.

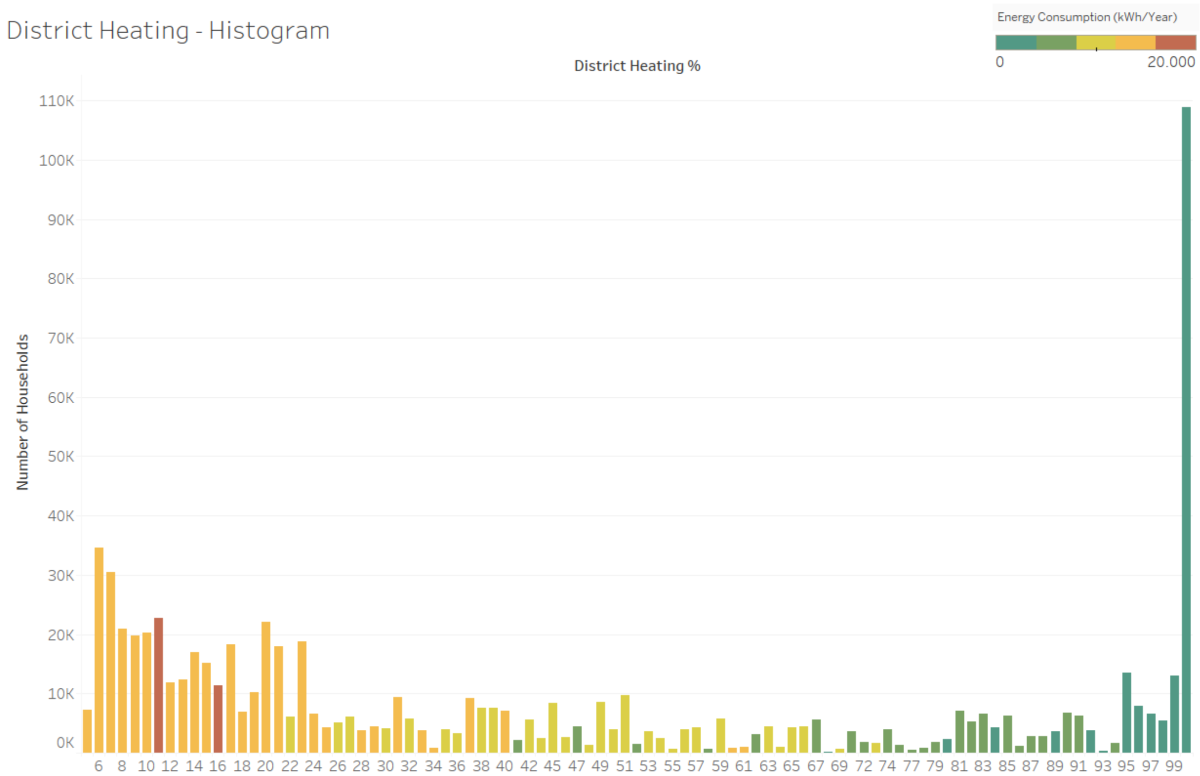

Figure 19: Histogram Percentage of District Heating

Transformed Variables

The year of construction, number of dwellings in building, height indication and surface area were transformed before performing the statistical analyses. All construction years before 1900 were transformed to 1900, all number of dwellings above 50 were transformed to 50, all height indications above 10 were transformed to 10, and all surface areas above 400 were transformed to 400. This paragraph explains and justifies these transformations using figures 20 to 23.

Figure 20 reveals the histogram of the original data on the year of construction, and its relation with the energy consumption and energy affordability. Outliers going back to the year 1000 were removed from this visualisation for readability. The relation between the year of construction and the energy consumption and energy affordability is approximately linear from 2015 until 1940, after which it flattens. However, as there are a large number of cases between 1940 and 1900, only cases before 1900 were adjusted, for the purpose of keeping the large majority of the year of construction data intact.

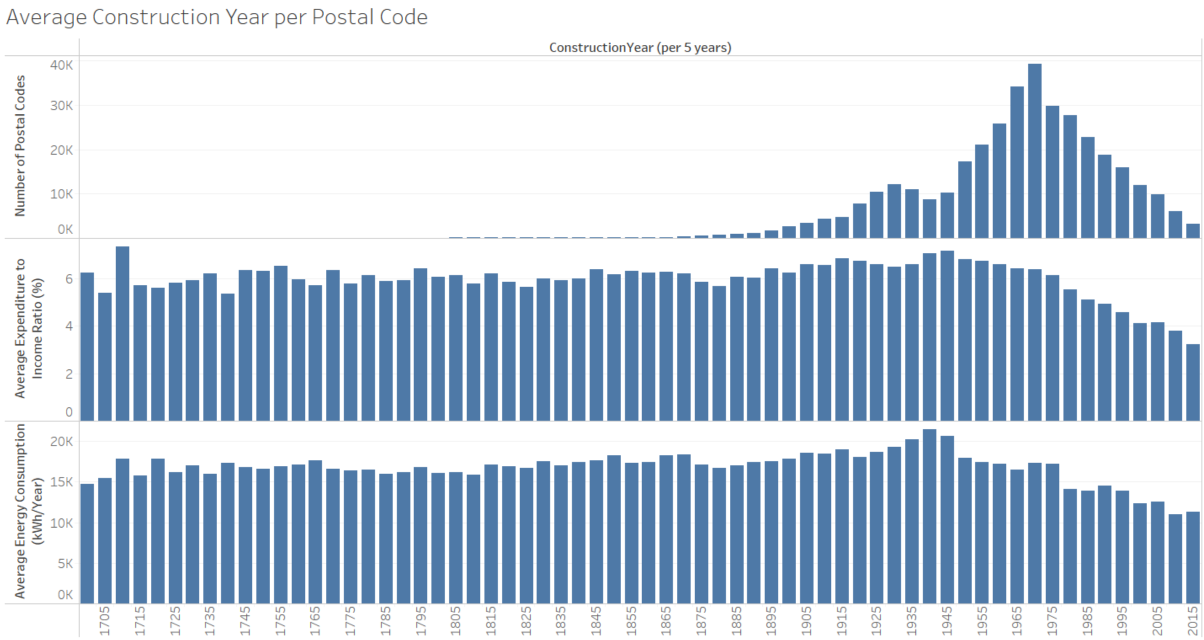

Figure 20: Original Distribution Construction Year

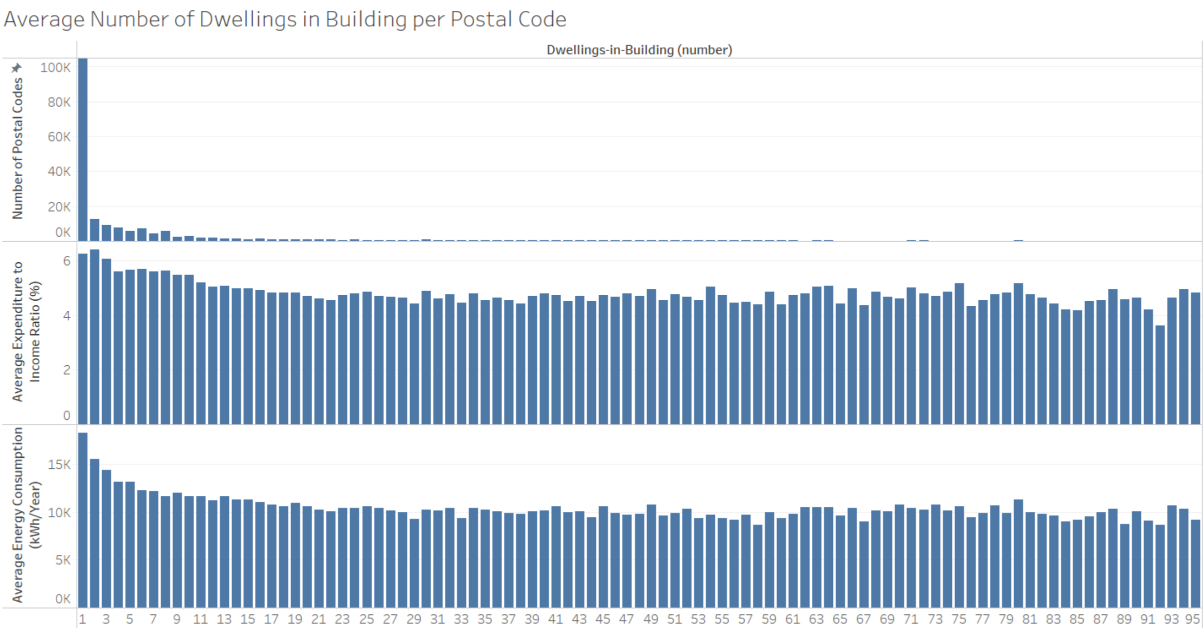

Figure 21: Original Distribution Number of Dwellings in Building

Figure 21 shows the distribution of the number of dwellings in a building, and its relation with the energy consumption and energy affordability. The relationship of the number of dwellings in a building with the energy consumption and energy affordability is relatively linear from 0 up until 30 dwellings;

afterwards the relation is no longer significant. Only cases with more than 50 dwellings are transformed, as the number of cases is notable until 50 dwellings, as a safe margin is needed from the 30 dwellings until which a clear relation is present, and as this research aims to minimize unnecessary transformations to keep the data interpretable.

Figure 22 presents an overview of the height indication and its relationship with the energy consumption and energy affordability. Outliers up until 169 are removed from the visualisations to keep the figure readable. A clear linear relationship with the energy consumption and energy affordability is present from 0 to 8, after which no relevant differences are observed. Thus, cases with a value higher than 10 are transformed, also because the number of cases with higher values is limited. Furthermore, the variation in the energy consumption and energy affordability is large above this value, although no longer relevant due to the limited number of cases.

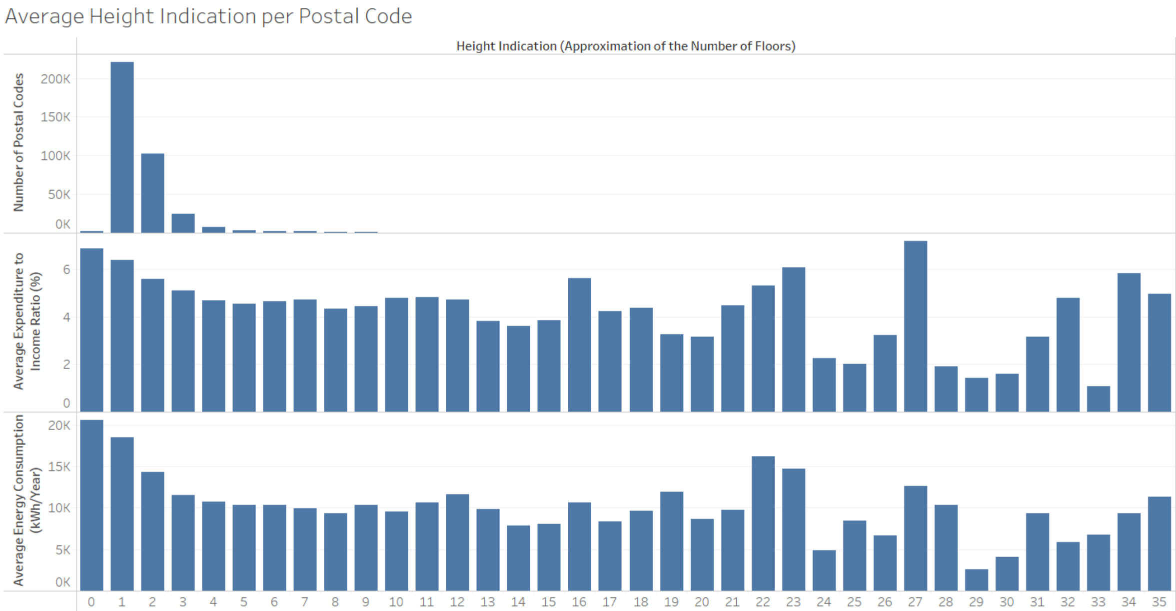

Figure 22: Original Distribution Height Indication

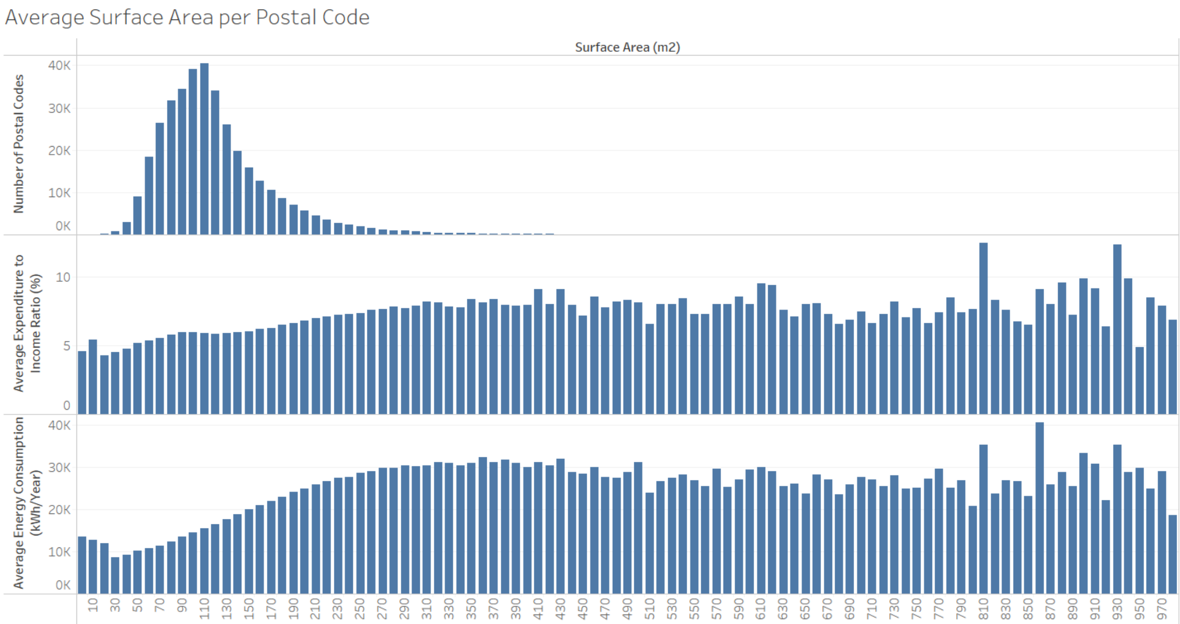

Figure 23: Original Distribution Surface Area

Figure 23 reveals the distribution of the surface area, and its relation with the energy consumption and energy affordability. The surface area included many outliers with values up to 20.900. However, in order to keep the visualisation readable, only values up to 1.000 square meters are included. The figure shows that from 0 to 350 square meters the relation between the surface area and the energy consumption and affordability is clear, and that it is relatively insignificant afterwards. Thus, the values above 400 are transformed, also since the number of cases with higher values is low.

Distribution of Energy Consumption in Missing Energy Affordability Cases

The energy consumption analyses in this study make use of a different data sample than the energy affordability analyses due to the fact that the energy expenditure to income ratio contains a large number of missing values, as described in paragraph 3.3. Thus, the distribution of the energy consumption was assessed within the missing energy expenditure to income cases, to check if the excluded energy expenditure to income values have a bias or if they are randomly distributed. Figure 24 shows that the energy consumption is randomly distributed within the missing energy expenditure to income cases, as the distribution presented is similar to the energy consumption distribution of the complete dataset in figure 7. This justifies that different data samples are used for the energy consumption analyses and energy affordability analyses.

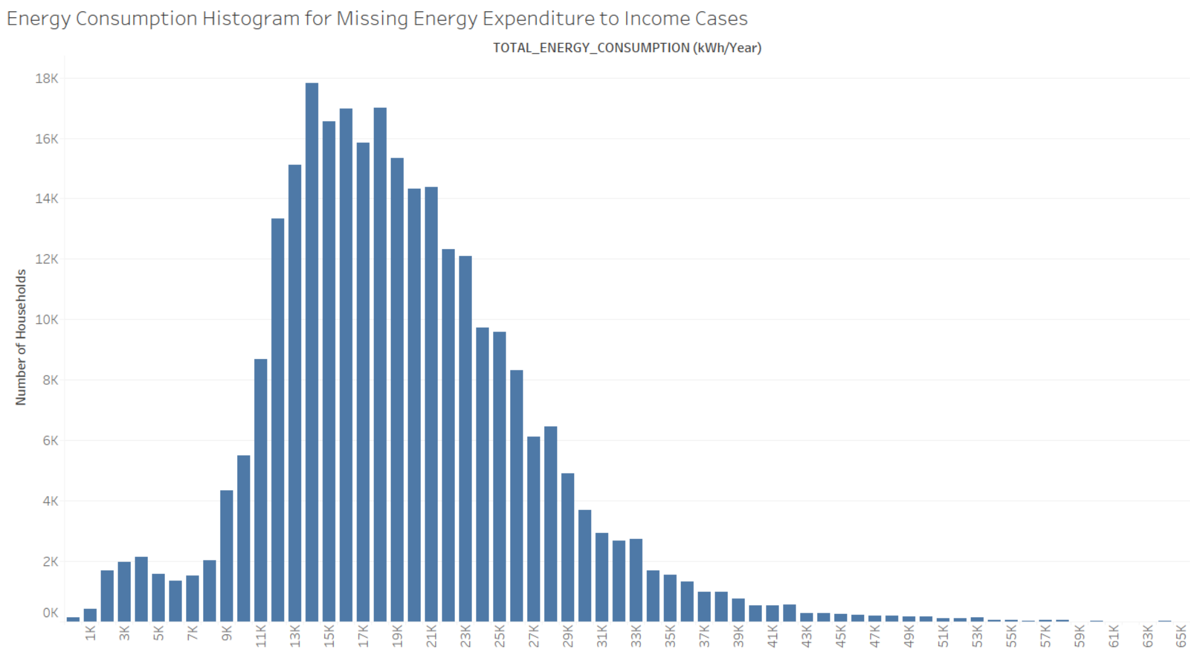

Figure 24: Energy Consumption Histogram for Missing Energy Expenditure to Income Cases

Appendix B: Excluded House Types for Regression

Some of the house types from the final dataset were excluded from the regression analysis because of multicollinearity. This appendix motivates their removal with collinearity statistics and a principal component analysis.

Table 13: Regression Model with All Variables Included

Regression Coefficients Energy Consumption with Semi-Detached & Detached House

	Model 1			Model 2					
	Unstandardized Coefficients		Standardized Coefficients	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics
	B	Std. Error	Beta	B	Std. Error	Beta	t	Sig.	VIF
(Constant)	26561.502	203.450		119107.763	334.889		355.66	0.000	
FEMALE_PR	-43.230	0.517	-0.066	-7.190	0.375	-0.011	-19.16	0.000	1.128
RESIDENTS_AGE_1524_PR	48.761	0.510	0.082	32.920	0.369	0.056	89.19	0.000	1.349
RESIDENTS_AGE_2544_PR	-64.002	0.487	-0.164	-14.567	0.356	-0.037	-40.97	0.000	2.886
RESIDENTS_AGE_4564_PR	6.263	0.460	0.014	-8.890	0.333	-0.021	-26.68	0.000	2.051
RESIDENTS_AGE_65PL_PR	8.968	0.422	0.029	20.129	0.311	0.065	64.78	0.000	3.512
WESTERN_MIGRATION_PR	7.188	0.492	0.012	11.720	0.361	0.019	32.43	0.000	1.206
NONWESTERN_MIGRATION_PR	-42.512	0.345	-0.112	12.839	0.273	0.034	47.00	0.000	1.804
ONEPERSON_PR	-22.498	0.327	-0.102	-7.900	0.242	-0.036	-32.69	0.000	4.151
MULTIPLE_PERSON_WITHOUT_CHILDREN_PR	-6.879	0.272	-0.024	-3.065	0.197	-0.011	-15.54	0.000	1.635
TWOPARENTS_PR	-50.528	0.342	-0.188	-8.242	0.252	-0.031	-32.77	0.000	3.045
HOUSEHOLD_SIZE	3105.915	20.382	0.288	559.278	15.259	0.052	36.65	0.000	6.937
INCOME	219.694	0.673	0.329	52.532	0.641	0.079	81.99	0.000	3.181
SOCIAL_BENEFITS_PR	-13.776	0.564	-0.023	19.665	0.417	0.033	47.21	0.000	1.660
AVG_TEMP	-224.877	18.405	-0.016	-394.544	13.389	-0.028	-29.47	0.000	3.145
AVG_MIN_TEMP	-1176.744	11.080	-0.114	-468.567	8.210	-0.045	-57.07	0.000	2.192
AVG_SUNSHINE_DURATION	-1348.091	33.645	-0.093	-875.915	24.313	-0.061	-36.03	0.000	9.779
Year2017	-951.473	16.801	-0.071	-917.939	12.233	-0.069	-75.04	0.000	2.928
Year2018	402.583	24.657	0.030	-60.330	17.993	-0.005	-3.35	0.001	6.330
WOZ_VALUE				11.186	0.044	0.221	254.66	0.000	2.614
SOCIAL_RENT_PR				-2.891	0.136	-0.014	-21.27	0.000	1.478
RENT_PR				-8.820	0.178	-0.042	-49.46	0.000	2.523
URBAN_RURAL_SCALE				77.112	3.415	0.017	22.58	0.000	1.851
YEAR_OF_CONSTRUCTION				-57.267	0.151	-0.235	-379.29	0.000	1.326
HEIGHT_INDICATION_BUILDING				-369.004	4.806	-0.062	-76.78	0.000	2.243
DWELLINGS_IN_BUILDING				9.182	0.483	0.017	18.99	0.000	2.872
LOG_SQUAREMETERS_DWELLING_m2				9967.497	38.406	0.270	259.53	0.000	3.740
APARTMENT_PR				-39.001	0.326	-0.243	-119.65	0.000	14.315
TERRACED_HOUSE_PR				-30.637	0.401	-0.152	-76.32	0.000	13.788
DETACHED_HOUSE_PR				16.350	0.316	0.080	51.82	0.000	8.190
SEMI_DETACHED_HOUSE_PR				-1.886	0.337	-0.006	-5.60	0.000	4.551
DISTRICT_HEATING				-956.771	17.903	-0.030	-53.44	0.000	1.063

Purpose of the Principal Component Analysis

Table 13 presents the regression model with all the house types except the percentage of corner homes included. Removing only the percentage of corner homes does however not decrease the interpretability of the regression model significantly, as the percentage of corner homes can be estimated once the other house type percentages are known since they together add up to 100%. The collinearity statistics of the model do however reveal that the percentage of apartments and the percentage of terraced houses have a VIF above 10, pointing out that collinearity is too high, which is reducing the validity of the regression model. Thus, at least one other house type needs to be excluded from the regression analyses, leading to an information loss, as this means that the percentages of the excluded variables can no longer be estimated by filling in the other house types.

In order to minimize this information loss and keep the predictive power of the model as high as possible, a principal component analysis was performed with all the different house types. The components created were not added to the regression, but used to determine which of the variables capture the highest variance. These variables were kept in the analyses, while variables with a lower explaining power were left out to deal with the multicollinearity.

Total Variance Explained

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2,288	45,768	45,768
2	1,489	29,790	75,558
3	,886	17,719	93,277
4	,332	6,635	99,912
5	,004	,088	100,000

Communalities

	Initial	Extraction
APARTMENT_PR	1,000	,980
TERRACED_HOUSE_PR	1,000	,869
SEMI_DETACHED_HOUSE_PR	1,000	,476
DETACHED_HOUSE_PR	1,000	,650
CORNER_HOME_PR	1,000	,803

Component Matrix

	Component				
	1	2	3	4	5
TERRACED_HOUSE_PR	,932	-,020	-,064	-,355	,032
CORNER_HOME_PR	,890	,101	,032	,443	,016
APARTMENT_PR	-,517	-,844	,106	,083	,040
DETACHED_HOUSE_PR	-,518	,618	-,589	,052	,032
SEMI_DETACHED_HOUSE_PR	-,303	,620	,723	-,021	,022

Extraction Method: Principal Component Analysis.

Figure 25: Results Principal Component Analysis

Results Principal Component Analysis

The results of the principal component analysis are shown in figure 25. The total variance table reveals how much variance each component explains, while the component matrix visualizes the correlations of the different house types with the components created. The communality table shows how much of a certain variable's variance can be explained by the other house types. In order to prevent VIF statistics higher than 10, two of the housing types need to be left out of the regression analysis. Thus, the variables that have the highest correlation with the three components explaining the most variance are left in the regression. Figure 25 reveals that TERRACED_HOUSE_PR has the highest correlation with component 1, APARTMENT_PR has the highest correlation with component 2, and SEMI_DETACHED_HOUSE_PR has the highest correlation with component 3. Therefore, a regression model was built in SPSS, excluding the percentage of detached houses and the percentage of corner homes.

Justification for Including the Percentage of Detached Houses

The new regression model without multicollinearity is summarized in table 14 and table 15. The regression model has an R Square of 0,702, and no longer has any VIF values above 10. Therefore, the variables included in this model could have been used for the final regression models. There is however a disadvantage to this, namely that the percentage of detached houses is excluded from the analysis. This is undesirable as including the percentage of detached houses would improve the interpretability of the regression model. First, the detached houses are the house type with the highest energy consumption, meaning that the variable can be used in the analysis to get insight in the variety in energy consumption and energy affordability based on house type, e.g. when comparing the percentage of detached houses to the percentage of apartments. Second, the percentage of detached houses has the second highest correlation with energy consumption and energy affordability, and thus it is interesting to see whether this correlation also results in a strong regression coefficient. Third, the percentage of detached houses in the dataset is significantly higher than the percentage of semi-detached houses, which is the variable that adds the least extra variance of the three variables included. 18,0% of the included houses is detached, while only 11,7% is semi-detached. Therefore, including detached instead of semi-detached would make the model better interpretable for a higher percentage of the dwellings included in the data.

Table 14: Regression Model with Semi-Detached Houses

Regression Coefficients Energy Consumption with Semi-Detached House

	Model 1			Model 2					
	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics VIF
(Constant)	26561.502	203.450		121122.556	333.057		363.67	0.000	
FEMALE_PR	-43.230	0.517	-0.066	-7.336	0.376	-0.011	-19.53	0.000	1.128
RESIDENTS_AGE_1524_PR	48.761	0.510	0.082	33.620	0.369	0.057	91.03	0.000	1.347
RESIDENTS_AGE_2544_PR	-64.002	0.487	-0.164	-14.360	0.356	-0.037	-40.34	0.000	2.886
RESIDENTS_AGE_4564_PR	6.263	0.460	0.014	-8.467	0.334	-0.020	-25.39	0.000	2.050
RESIDENTS_AGE_65PL_PR	8.968	0.422	0.029	20.349	0.311	0.066	65.41	0.000	3.512
WESTERN_MIGRATION_PR	7.188	0.492	0.012	11.114	0.362	0.018	30.73	0.000	1.205
NONWESTERN_MIGRATION_PR	-42.512	0.345	-0.112	13.952	0.273	0.037	51.17	0.000	1.793
ONEPERSON_PR	-22.498	0.327	-0.102	-7.638	0.242	-0.035	-31.57	0.000	4.149
MULTIPLE_PERSON_WITHOUT_CHILDREN_PR	-6.879	0.272	-0.024	-3.047	0.198	-0.011	-15.43	0.000	1.635
TWOPARENTS_PR	-50.528	0.342	-0.188	-8.404	0.252	-0.031	-33.38	0.000	3.044
HOUSEHOLD_SIZE	3105.915	20.382	0.288	561.407	15.279	0.052	36.74	0.000	6.937
INCOME	219.694	0.673	0.329	55.560	0.639	0.083	86.97	0.000	3.155
SOCIAL_BENEFITS_PR	-13.776	0.564	-0.023	19.836	0.417	0.033	47.55	0.000	1.660
AVG_TEMP	-224.877	18.405	-0.016	-430.640	13.389	-0.031	-32.16	0.000	3.136
AVG_MIN_TEMP	-1176.744	11.080	-0.114	-465.542	8.220	-0.045	-56.63	0.000	2.192
AVG_SUNSHINE_DURATION	-1348.091	33.645	-0.093	-840.767	24.335	-0.058	-34.55	0.000	9.771
Year2017	-951.473	16.801	-0.071	-905.826	12.246	-0.068	-73.97	0.000	2.927
Year2018	402.583	24.657	0.030	-68.219	18.016	-0.005	-3.79	0.000	6.329
WOZ_VALUE				11.277	0.044	0.223	256.61	0.000	2.610
SOCIAL_RENT_PR				-3.109	0.136	-0.015	-22.86	0.000	1.477
RENT_PR				-8.990	0.179	-0.043	-50.35	0.000	2.522
URBAN_RURAL_SCALE				79.051	3.419	0.017	23.12	0.000	1.851
YEAR_OF_CONSTRUCTION				-57.979	0.151	-0.238	-385.10	0.000	1.315
HEIGHT_INDICATION_BUILDING				-393.058	4.790	-0.066	-82.06	0.000	2.222
DWELLINGS_IN_BUILDING				10.782	0.483	0.020	22.32	0.000	2.860
LOG_SQUAREMETERS_DWELLING_m2				10341.839	37.769	0.280	273.82	0.000	3.607
APARTMENT_PR				-51.018	0.229	-0.318	-222.45	0.000	7.068
TERRACED_HOUSE_PR				-48.407	0.209	-0.241	-231.68	0.000	3.725
SEMI_DETACHED_HOUSE_PR				-15.330	0.215	-0.052	-71.20	0.000	1.853
DISTRICT_HEATING				-934.324	17.921	-0.029	-52.14	0.000	1.063

Table 15: Explained Variance Regression with Semi-Detached Houses

Model Summary Energy Consumption with Semi-Detached Houses

				Change Statistics		
Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	,650	,422	4770,163	,422	41843,639	,000
2	,838	,702	3427,085	,280	80531,081	,000

Conclusion on Excluded Variables

Thus, a regression model was built, including the percentage of detached houses, and excluding the percentage of semi-detached houses, to see whether this model performs significantly worse than the model with the percentage of semi-detached houses. If this is not the case, a model with the percentage of detached houses is desirable. This model is presented in table 9 and table 10 in the main text. The figures show that the R Square of the regression model with detached houses is 0,703, meaning that it doesn't lose predicting power when compared to the previous model with semi-detached houses. This is not surprising, as the percentage of detached houses also has a relatively high correlation with component 3 from the principal component analysis, as presented in figure 25, and thus captures much of the same variance. All in all, this resulted in the removal of the percentage of semi-detached houses, and in the addition of the percentage of detached houses. The final house types left out of the regression analysis are therefore the percentage of semi-detached houses and the percentage of corner houses.