MASTER

Developing an energy poverty risk index for social housing in the Netherlands

van Swam, Max

Award date:
2023

DEVELOPING AN ENERGY POVERTY RISK INDEX FOR SOCIAL HOUSING IN THE NETHERLANDS

Author: M.F.T. van Swam (1013005)

1st Supervisor: Stephan Maussen
2nd Supervisor: Theo Arentze
3rd Supervisor: Sander de Clerck

7Z45M0 Graduation project Urban Systems & Real Estate (45 ECTS)

Architecture, Building and Planning
Urban Systems and Real Estate
Technische Universiteit Eindhoven

26-6-2023

Figure on front page: (Metropoolregio Amsterdam, n.d.)

Summary

While energy poverty has been recognised as a societal problem in the United Kingdom since 1970, it has only recently gained attention from policy makers of the Dutch national government. Although increasing in the past few years, research about energy poverty in the Netherlands remains in its infancy. Furthermore, policies aimed at reducing energy poverty in the Netherlands have only been introduced recently while the amount of households with energy poverty have increased in the past years. A household that experiences energy poverty cannot meet their energy needs unless expenses are made to other necessary common consumables. Energy poverty remains a relevant societal problem and it can therefore be beneficial to increase the knowledge about energy poverty trough research. If energy poverty is analysed and mapped in a clearer and more detailed way, this can help to reduce the occurrence of energy poverty and help with the energy transition.

Currently, various indicators are used to assess if a household has energy poverty. Using these indicators, the determination if a household has energy poverty is often only based on two factors thereby ignoring many relevant factors for determining energy poverty. Additionally, the existing indicators do not present energy poverty as a continuous value and only determine if a household has energy poverty or does not have energy poverty. Because of this, the current energy poverty indicators cannot be used to determine how close households are to experience energy poverty or how close they are to solving their energy poverty problems. Furthermore, since the existing indicators use different factors to determine if a household has energy poverty, a household may experience energy poverty according to one indicator and not according to another one.

This research aims to improve the method of identifying energy poverty for households through the research question: Can a new model be created that predicts the risk of energy poverty? In order to answer this research question, research on energy poverty is reviewed and relevant factors are identified from this literature. In existing research, the energy quote (EQ), low income, high costs (LIHC), minimum income standard (MIS), housing costs overburden rate (HCOR), and the low income low energy quality (LILEQ) indicators are used to determine the presence of energy poverty. The literature review shows that multiple factors with an effect on energy poverty are already included in the existing indicators. However, at the same time, most existing energy poverty indicators do not include dwelling and personal characteristics that have a relation with energy poverty according to the literature review.

Based on the findings of the literature review, a conceptual model is constructed in which all relations between characteristics found in the literature review are included. All relevant variables are selected from the Woononderzoek Nederland (WoON 2021) dataset and the data is processed and prepared for multiple analyses. A descriptive analysis is conducted to visualise the dataset and filter outliers and value categories with few respondents from the dataset. After this analysis, multiple bivariate analyses are performed in order to determine the sample representativeness and to test for multicollinearity between independent variables. The results of these analyses show that the data sample is very representative for the social housing stock of the Netherlands and that there is almost no multicollinearity between the independent variables. Because of a high correlation between two variables, a variable is removed from the dataset in order to increase the reliability of the results of further analyses.

An exploratory factor analysis (EFA) is conducted to determine if all existing energy poverty indicators measure the same latent concept. All existing indicators included in the EFA show large factor loadings except for the MIS indicator. The results of the EFA show that all included existing energy poverty indicators measure the same common latent concept except for the MIS indicator. Because of this, the MIS indicator is removed from further analyses.

A structural equation model is made based on the conceptual model and structural equation modelling (SEM) is used to estimate the path coefficients of the model based on more than 8,500 subsamples. The results of the SEM show that most path coefficients are statistically significant except for a few value categories of some categorical variables. Furthermore, the results show that all path coefficients between the degree of urbanisation categories and income are non-significant. This finding indicates that there is no indirect relation between the degree of urbanisation and energy poverty through income. All statistically significant path coefficients are interpreted to determine the relations between all variables in the model and energy poverty.

An energy poverty risk prediction model is created based on the statistically significant path coefficients estimated by the SEM. Using all relevant factors identified in the literature review connected with the path coefficients of the SEM, the prediction model can determine the energy poverty risk index (EPRI) for all respondents of the WoON 2021 dataset. The results of the EPRI calculations for all respondents of the WoON 2021 dataset show that the EPRI has a relation with the other existing energy poverty indicators. Furthermore, a descriptive analysis of the respondents with a high EPRI is conducted in order to determine the characteristics of respondents that are overrepresented in the group with a high EPRI. The comparison between the descriptive analyses show that single-person and single-parent households with a low education living in neighbourhoods with a lower degree of urbanisation are overrepresented in the group with a high EPRI. Furthermore, the results show that there are relatively few households in the group with a high EPRI that live in dwellings constructed after 1992.

The EPRI prediction model is used to evaluate several policies and future scenarios. These predictions show that increasing energy price will increase the energy poverty risk for many respondents. Furthermore, the created EPRI model shows that the current government policy to reduce energy poverty may not be effective. Because the benefits of this policy are granted solely based on the income of a household, they are often granted to households that do not experience energy poverty instead of households that do experience energy poverty. Further results of the predictions show that improving the energy performance of dwellings reduce the EPRI more effectively.

The EPRI facilitates a simple and clear identification of risk groups and analysis of policy effects. It includes a large amount of factors to predict the energy poverty risk on a continuous scale, which enables a more accurate evaluation of the effect of policies compared to the existing energy indicators. The existing energy poverty indicators only predict “energy poverty or no energy poverty” and can therefore not predict how close a household is to experiencing energy poverty or how close they are to avoid energy poverty. The EPRI can be a useful tool to identify energy poverty risks for housing corporations and the government. Housing corporations can use the EPRI to more effectively identify dwellings in need of renovations to improve their energy efficiency. The government can use the EPRI to gain a clear overview where there are high and low energy poverty risks so they can more effectively target their allowances and other subsidies aimed at reducing energy poverty.

Preface

Before you lies the master thesis *Developing an energy poverty risk index for social housing in the Netherlands* written as graduation project for the Master study Urban Systems & Real Estate at the Technische Universiteit Eindhoven. This master thesis was partly executed during a graduation internship at Republiq in ‘s Hertogenbosch. This research is intended for housing corporations and government policymakers that demand a clear and detailed way to determine and analyse the energy poverty risk. For these stakeholders, gained insights on energy poverty risks will enable them to adjust their policies accordingly.

I would like to thank my supervisors from the university, Theo Arentze and Stephan Maussen, for their guidance during the process and their advises of research methods. Due to their different areas of expertise, coaching sessions with both supervisors complemented each other well during the process of my thesis. I would like to thank my supervisor from Republiq, Sander de Clerck, for his help and assistance during the process and for supplying data that could be used in the research. I would like to thank Peter van der Waerden for supplying the WoON 2021 data that was crucial for this research. Furthermore, I would like to thank the employees of Republiq for their help during the research. I am grateful that they taught me the KNIME program, which was very useful during this research. Additionally, it was nice that they showed me how consultants and housing corporations regard energy poverty. Finally, I would like to thank my father for giving me advise, explaining how to use Excel macro’s, and for checking my thesis.

I hope you enjoy your reading.

Max van Swam

Gassel, May 1, 2023

Table of Contents

1. Introduction

1.1 Background

At the beginning of May 2023, the Dutch minister for climate announced a new package of measures to reduce CO_2 emissions. This package focuses on “greening” residential, commercial, and public real estate. In order to achieve this, the government will spend nine billion euros on insulating buildings, installing heat pumps, and creating heat networks. The government will invest in increasing the sustainability of vulnerable neighbourhoods, installation of solar panels on rental dwellings, and renovation vouchers for homeowners. The government will furthermore increase the tax on natural gas while reducing the tax on electricity to increase the attractivity of renovating dwellings into gas-free dwellings. The measures in the climate package are partly aimed at people that do not have the financial capability to increase the sustainability of their dwellings but do suffer from a poorly insulated dwelling and high energy bills (Kraan, 2023). In addition to the energy prices themselves, food price have also increased in recent years due to these increased energy price and inflation. Because of the increasing costs for food, energy and other housing costs, the costs of living for many Dutch households are becoming untenable, resulting in more households ending up in energy poverty (Nu.nl, 2022). Between 2020 and 2022, the amount of people experiencing energy poverty in the Netherlands increased to over 600,000 (Basekin, 2023). While prices continue to rise and the cost of living for Dutch citizens increases, many wonder how the national government of the Netherlands will support households and if the suggested measures will prevent the increasing energy poverty (Nu.nl, 2022).

Energy poverty has been recognised as a societal problem by the United Kingdom since the 1970 (Delbeke, Verbeeck & Oosterlynck, 2013). However, until recently energy poverty research has been relatively limited within the European Union. For a long time, energy poverty as a problem and research into energy poverty gained little attention in the Netherlands. However, recently the national government of the Netherlands has increased its attention towards the problem of energy poverty. The COVID-19 pandemic has increased the awareness of reducing energy poverty and increasing the indoor climates of buildings. During this pandemic people spend more time at home than before, increasing their energy usage and therefore increasing their energy bills (Kruit, van Berkel & Dehens, 2021). As people were forced to spend a higher portion of their budget on energy, concerns about the adverse effects of energy poverty increased (Churchill & Smyth, 2021). Sped up by these relatively recent developments, energy poverty has gradually emerged on the agendas of EU policy makers (Bouzarovski, Thomson & Cornelis, 2021). However, since energy poverty has only recently been included in policies by policy makers, energy poverty research in the Netherlands remains in its infancy (Mulder, Dalla Longa & Straver, 2023). The recent increase in the amount of research about energy poverty in the Netherlands is also visible when looking at figure 1. This figure shows the amount of energy poverty research that was studied for the literature review of this research. Researches about energy poverty in the Netherlands have only been published since 2020 while international and European researches on energy poverty were already published prior to 2020.

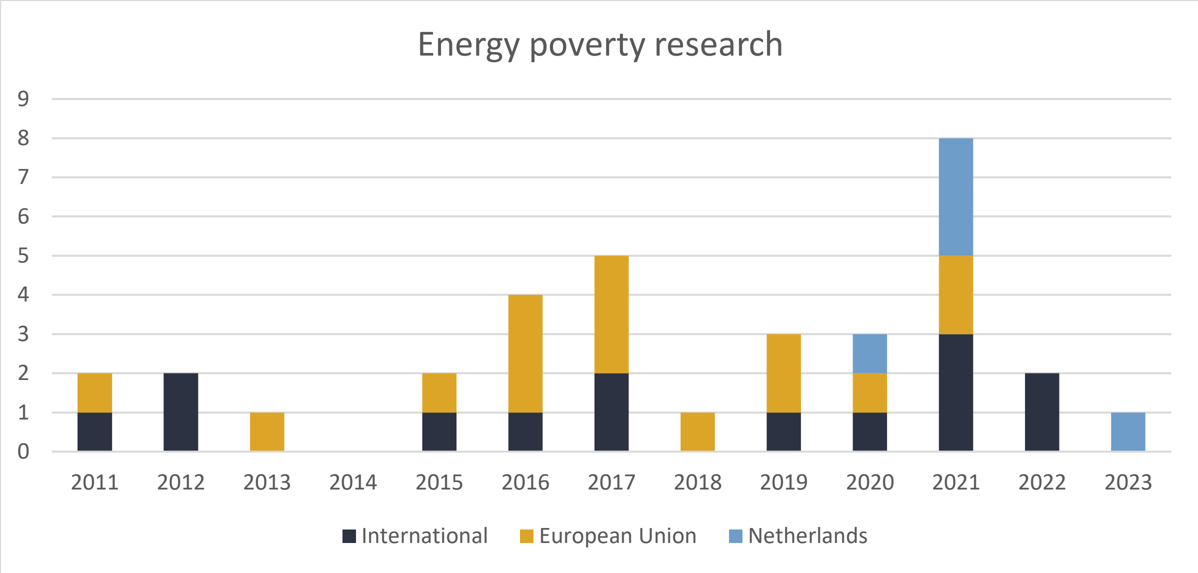

Figure 1. Publication dates of studied energy poverty literature

1.2 Definition

Currently, there is no common international definition for energy poverty. Households with energy poverty have difficulties with paying their energy bills and cannot invest in measures to reduce their energy usage. Energy poverty is connected to general poverty, however households that do not live in poverty may also experience energy poverty (Bonnard, Bruynoghe, Deprez & Kestemont, 2015). Huybrechs et al (2011) define energy poverty as a condition in which a household experiences difficulties with the supply of energy that is needed to satisfy their needs. The following definition of energy poverty will be used in this research:

"A household experiencing energy poverty cannot meet their energy needs unless this is at the expense of other necessary common consumables" (Bonnard et al, 2015).

This definition of energy poverty includes both regular energy poverty and hidden energy poverty. Regular energy poverty occurs when a household goes into debt because of using too much energy compared to their budget. Hidden energy poverty occurs when a household chooses not to meet their energy needs because they cannot afford this and do not want to get into debt (Delbeke et al, 2013).

1.3 Motivation

The aim of this research is to increase knowledge about energy poverty. Reducing energy poverty can lead to multiple socio-economic benefits and can speed up the energy transition (Mulder, Dalla Longa & Staver, 2021). Energy poverty is strongly connected to the energy transition where, according to the climate agreements, the aim is to stop using natural gas to heat buildings to reduce CO_2 emissions before 2050. This transition potentially reduces energy poverty by reducing the energy usage. However, this transition may also increase energy poverty through increasing energy prices or through the high investment that have to be made to improve the energetic efficiency of buildings. Because of this, energy poverty also influences the energy transition since the energy transition can only be a success when all households are included. Energy poverty must be considered in order to realise a successful energy transition that is affordable for every household. Because of this, it is important to know which households experience energy poverty and need additional support (Kruit et al, 2021).

1.4 Problem analysis

There are different measurement methods and indicators used in order to determine if a household has energy poverty. These indicators will be elaborated further during the literature review. Many of the existing indicators only use factors that are related to income and expenses. However, when predicting energy poverty housing and socio-demographic characteristics are often not included. Mulder et al (2021) develop an indicator that includes housing characteristics however the amount of housing characteristics that are included in this indicator are relatively limited. Other factors that are causally related to energy poverty according to the literature review are however not included in the existing indicators. Additionally, the various energy poverty indicators create significant differences in the measured number of households that experience energy poverty, even within researches. Next to this, many researches simply combine and add up the results of different energy poverty indicators, causing some households to be included twice or more in the total count of households experiencing energy poverty. The various characteristics of energy poverty currently create difficulties in order to estimate the energy poverty problem for all different sorts of households. The need to develop a new energy poverty indicator is present in existing literature. Bonnard et al (2015) suggest that a new method should be developed in order to improve the quantification of the scale of energy poverty.

Based on the problems of the multiple existing indicators, this research aims to add to the existing knowledge by creating a new energy poverty index. Using this new indicator, the energy poverty problem in the Netherlands can be analysed and the risk of energy poverty for a household can be determined. Existing energy poverty indicators will be analysed and the size of the current energy poverty problem in the Netherlands will be analysed. It will be determined which factors have a significant relation with energy poverty, who the people experiencing energy poverty are, and where these people live. When the new indicator is constructed, this can be used to predict the number of households experiencing energy poverty for multiple future scenarios. Based on the results, it will be determined what policy should be implemented to reduce energy poverty.

1.5 Research questions

Based on the background and problem definition, the following main research question has been created:

- Can a new model be created that predicts the risk of energy poverty?

In order to successfully answer this research question, several sub-questions were created. These sub-questions are:

  - Which indicators are currently used in the literature to measure energy poverty?
  - Which factors are significantly related to energy poverty?
  - What are the characteristics of households with a high energy poverty risk?
  - What advise can be given to reduce the energy poverty risk based on the created model?

1.6 Relevance

Academic

Energy poverty has been studied internationally for over a decade and only recently in the Dutch context. Despite this, the prevalence of energy poverty and the recently increased attention to this problem ensure that energy poverty remains a relevant topic of academic research. From a technical and built environment background, research into energy poverty remains important to provide policy makers, engineers, and architects with as much relevant knowledge as possible. These professions determine how the society handles energy poverty and how accessible and inclusive the society is for people that experience energy poverty. Energy poverty remains a relevant theme for research about energy and the transition towards a more sustainable society. When more knowledge about energy poverty and the people that experience energy poverty is gained, policies about energy poverty can focus more on people and energy usage instead of only on buildings. It remains relevant to increase knowledge about the most recent developments of a societal problem through academic research since this new knowledge can provide an important contribution to the society. Academic relevance can build upon suggestions for further research that are given in the existing literature to continue the research about energy poverty and increase the knowledge of the problem. For example, Bonnard et al (2015) suggest that a new method should be developed to improve the quantification of the scale of energy poverty.

Kruit et al (2021) suggest that more research is needed to identify the households that experience energy poverty. This can speed up the energy transition since when only the dwellings of the households experiencing energy poverty are improved, the largest effects can be achieved against the lowest costs. In order to do this in the most optimised way, it is vital to analyse this maximum potential and research which target groups will benefit most from this. These target groups can, for example, be categorised by housing type, energy label and income. Additional research can be performed about a financial support that is dependent on, for example, the housing type or income (Kruit et al, 2021). Indicators and data can form an effective tool for targeted measures and policies and can help to monitor the development of energy poverty and the effects of policies. This can be supported with location analyses with maps at the national, municipal and neighbourhood levels that can help to implement more targeted policies to reduce energy poverty (Mulder et al, 2021).

Societal

The societal relevance of this research is primarily aimed at housing corporations, municipalities and policymakers that require a knowledge of energy poverty to optimise policies and reduce energy poverty problems. Additional societal relevance is gained by testing and analysing the differences between existing indicators of energy poverty and by creating an indicator for energy poverty that includes more potential factors of energy poverty than the current indicators. The indicator that will be developed in this research can become a useful tool for housing corporations and municipalities by providing them with a more accurate representation of energy poverty. Additionally, these stakeholders will be able to use the model to predict what the effect of certain changes or policies on energy poverty will be. Based on this knowledge, housing corporations and municipalities can change their policies to optimally support people with an elevated risk of energy poverty. The model can furthermore create an insight for housing corporations and municipalities about which dwellings their portfolios should be renovated to reduce energy poverty. Policy makers can use the results of this research to develop optimised policies to reduce energy poverty. Finally, this research can provide an insight into which dwellings should be constructed to reduce the risk of experience energy poverty.

2. Literature review

This chapter will describe the literature review that is performed in order to analyse how energy poverty is currently researched and measured. The literature review will additionally review factors with a potential effect on energy poverty, the consequences of energy poverty, energy poverty mitigation and the future of energy poverty.

2.1 Energy poverty in existing research

In his research about the energy poverty in Groningen, Baardink (2020) used the WoonOnderzoek Nederland (WoON) 2018 dataset to analyse the potential energy poverty of 414 households using multiple indicators. Baardink concluded that 12-19% (14,909-23,606) of the households in Groningen experience in energy poverty. Additionally, Baardink concluded that a strong increase in energy poverty will occur if the energetic performance of a large number of dwellings is not or not timely improved. Compared to the results of his results of 2020, Baardink predicts that by 2030 the number of households experiencing energy poverty in Groningen will have doubled. The research of Baardink was however limited by a relatively small dataset and aged data that did not include recent prices and new municipalities. Additionally, the 10% indicator showed a lower prediction of energy poverty cases than other indicators in the research of Baardink, while this indicator yielded the most cases of energy poverty in other researches (Kruit et al, 2021; Mulder et al, 2021). At the national level, between 234,000 and 634,000 (3.3-8.8%) of the households in the Netherlands experience energy poverty according to Kruit et al (2021). In their research, Kruit et al (2021) used an indicator to include the hidden energy poverty that was relatively difficult to quantify and concluded that about 40,000 households (0.6%) in the Netherlands experience hidden energy poverty. The determined number of 550,000 (7%) households experiencing energy poverty in the Netherlands of Mulder et al (2021) is between the range described by Kruit et al (2021). However, Mulder et al (2021) found a larger number of households experiencing hidden energy poverty than Kruit et al (2021), namely 140,250 (1.8%). Mulder et al (2021) add to their research by concluding that more than 10% of the households experience energy poverty in five municipalities and 7% of all neighbourhoods in the Netherlands.

Internationally, multiple researches have been performed to analyse energy poverty, showing varying results and economic effects on energy poverty. Halkos & Gkampoura (2021) concluded that economic recessions have an impact on energy poverty, something that was also found by González-Equino (2015) who added that energy consumption decreases during economic recessions. Developing throughout the years, Halkos & Gkampoura (2021) found that energy poverty in Europe decreased between 2004 and 2008, increased between during 2008 and 2013 due to the energy crisis, and decreased again between 2013 and 2019. They also added that eastern and southern European countries have the most energy poverty while the Scandinavian countries have the least energy poverty. Coene & Meyer (2019) concluded that about 14% of all households in Belgium experienced energy poverty and Sokołowski, Lewandowski, Kiełczewska & Bouzarovski (2020) found that about 9.8% of all households in Poland experienced energy poverty in 2017. In their research about energy poverty in Greece, Papada & Kaliampakos (2016) concluded that 58% of all Greek households and more than 90% of all Greek households in poverty experience energy poverty. They add that Greek households spend on average about 14% of their income on energy and that 75% of all Greek households have reduced other essentials in favour of energy needs. Churchill & Smyth (2021) concluded that over 40% of the households in Bulgaria are experiencing energy poverty. Additionally, they added that over 20% of the households in the US experience energy poverty while Bednar & Reames (2020) concluded this amount to be 35.5% of all households in the US. Energy poverty is subject to seasonal differences and there is significantly more energy poverty compared to other seasons (Okushima, 2017).

2.2 Energy poverty measurement

Energy poverty in the context of developed countries differs from the energy poverty that is studied in developing countries. Where developed countries usually see energy poverty as the inability to pay for energy, developing countries see energy poverty as a lack of access to energy (Okushima, 2016; Bonatz, Guo, Wu & Liu, 2019). Because this research focuses on the Netherlands, energy poverty will be regarded from the developed countries perspective. It is difficult to construct a single indicator for energy poverty because of the different definitions in energy poverty between developing and developed countries and national and regional differences (Sokołowski et al, 2020). Because of this, a multidimensional index for energy poverty (MEPI) should be constructed that can be used as a single indicator for energy poverty in the Netherlands (Okushima, 2017).

In their research about effect of economic crisis on energy poverty in Europe, Halkos & Gkampoura (2021) describe two approaches to define and measure energy poverty: consensual approaches and expenditure approaches. Consensual approaches use various subjective indicators to identify households that have difficulties meeting basic energy needs. These approaches use surveys in order to gain subjective information about energy affordability, thermal comfort, and dwelling efficiency. The information used as indicators in the surveys may include: the inability to keep the home adequately warm, arrears on utility bills, and the presence of leaks, damp, and rot in the dwelling. Expenditure approaches use indicators that are evaluated against a critical threshold. These indicators may include a high share of energy costs, a low available income, or insufficient spending on energy (Halkos & Gkampoura, 2021).

Romero, Linares & López (2018) describe three expenditure approach indicators that are used to define and measure energy poverty: the energy quote (EQ), low income-high costs (LIHC), and minimum income standard (MIS). The most commonly used indicator for determining energy poverty is the EQ. According to the EQ, a household experiences energy poverty when:

- $10\%\ [Spendable\ income] < [Energy\ expenditure]$ (1)

The EQ is relatively simple to calculate, easy to communicate and relatively versatile from a pragmatic point of view. However, it is overly sensitive to energy prices, the 10% threshold is arbitrary selected, and the indicator has no reference to the household income (Romero et al, 2018).

The LIHC indicator defines energy poverty when the income is below a poverty threshold and when the energy costs are higher than an energy expenditure threshold. This indicator corrects the EQ by considering not only the expenditure on energy but also an income threshold (Romero et al, 2018). An additional benefit of the LIHC indicator is that it can easily distinguish between energy poverty and general poverty (Rademaekers et al, 2016). However, the LIHC is a complex and non-transparent indicator. Additionally, it is difficult to isolate causes and effects when analysing series and to find out households that can come out of energy poverty by reducing their energy costs (Romero et al, 2018). According to the LIHC indicator defined by Romero et al (2018) a household experiences energy poverty when:

- $[household\ expenditure\ on\ energy] > [median\ expenditure\ on\ energy]$ (2)

And:

- $[household\ income] - [household\ expenditure\ on\ energy] < 60\%\ [median\ household\ income - mean\ expenditure\ on\ energy]$ (3)

According to the MIS indicator a household is energy poor when they are not able to pay for their basic energy costs after covering housing and other needs. The MIS indicator is the most robust when measuring objective income-based energy poverty because it addresses the problem from its economic root. However, a con to this method is that it is difficult to determine the minimum income on an objective basis (Romero et al, 2018). Additionally, Rademaekers et al (2016) state that the MIS indicator predicts relatively low amounts of energy poverty for the lowest income groups and therefore question how accurate the MIS represents the reality. In order to determine if a household has energy poverty according to the MIS indicator, a MIS factor needs to be selected. This MIS factor represents all expenses on needs of a household, except for council tax, rent, mortgage and fuel costs. Baardink (2020) uses the Minimumvoorbeeldbudgetten provided by the Budgethandboek of Nibud for the MIS factor. Kruit et al (2021) define the MIS as the payment risk however they use the Referentiebudgetten voor levensonderhoud defined by the Sociaal en Cultureel Planbureau as the MIS factor. According to the MIS indicator defined by Romero et al (2018) a household experiences energy poverty when:

- $[energy\ costs] > [net\ household\ income] - [housing\ costs] - [MIS]$ (4)

An overview of the pros and cons of the energy poverty indicators is shown in table 1. Comparing all energy poverty indicators, Romero et al (2018) advise the usage of the MIS indicator for energy poverty. However, they add that an optimal MIS indicator should include the incomes and expenditures of households and all energy sources.

Table 1: *Pros and cons of expenditure approach indicators (Based on: Rademaekers et al, 2016; Romero, Linares & López, 2018)*

Energy quote (EQ)	Low income, high costs (LIHC)	Minimum income standard (MIS)
+ Simple to calculate	+ Corrects the energy quote by considering not only the expenditure on energy but also an income threshold	+ Robust when measuring objective income-based energy poverty by addressing the problem form its economic root
+ Easy to communicate		
+ Relatively versatile from a pragmatic point of view	+ Easily distinguishes between energy poverty and general poverty	
- Excessive sensitivity to energy prices	- Overlay complex and not transparent indicator	- Predicts relatively low amounts of energy poverty for the lowest income groups
- Arbitrary selection of the threshold at 10%	- Difficult to find out those households that can come out of energy poverty by reducing their energy costs	
- Lack of any reference to household income	- Difficult to isolate causes and effects when analysing the series	- Difficult to determine the minimum income on an objective basis

Adding to the energy poverty indicators described by Romero et al (2018), Baardink (2020) adds the housing costs overburden rate (HCOR) that includes the net living costs, energy costs and the spendable income. According to this indicator, someone experiences energy poverty when:

- $[net\ living\ costs] - [energy\ costs] > 40\%\ [spendable\ income]$ (5)

Mulder et al (2021) describe an energy poverty indicator that includes the energy performance of a dwelling. This indicator is called the low income and low energy quality (LILEQ) indicator. According to this indicator a household experiences energy poverty when they have a relatively low income and live in a house with a relatively low energy quality. The low income is defined as having an income that is lower than 130% of the legal social minimum and having a financial capital that is part of the lowest 10% of the Netherlands. Mulder et al (2021) use two measures to define a dwelling with a low energy quality. The first is if a dwelling has an energy label D or lower. The second, more accurate, definition defines a dwelling with a low energy quality when the median energy usage of the dwelling class is higher than the median energy usage of all dwellings in the Netherlands. The different dwelling classes that were included in their research were determined based on the dwelling characteristics: construction year, dwelling type, and dwelling size. According to the analysis of Mulder et al (2021), about 50% of all dwellings with an energy label C and all dwellings with a lower energy label were defined as having a low energy quality according to their more accurate definition.

Next to the general LILEQ indicator, Mulder et al (2021) add the LILEQ- and LILEQ+ indicators. LILEQ- is a variant of LILEQ aimed to measure hidden energy poverty and the households that under consume energy because of financial difficulties. LILEQ- adds an underconsumption of energy factor to LILEQ and this factor includes if the energy costs of a household are part of the lowest 25% in their dwelling class. LILEQ+ is another variant of LILEQ, aimed to include households with a strikingly high energy consumption. LILEQ+ adds an energy overconsumption factor to LILEQ in order to include if the energy costs are part of the highest 75% of the corresponding dwelling class.

Both LILEQ and its variants add to the other measurements of energy poverty by including some dwelling characteristics instead of primarily looking at household incomes and expenditures. However, because of this, LILEQ and its variants are insensitive to the effect of energy factors on energy poverty, such as the energy prices and the energy consumption of households (Mulder et al, 2021). Another disadvantage of the LILEQ indicator is that measuring and comparing the energetic quality of a dwelling is not a simple task in practice. Reliable energy label data is often absent and hence the dwelling energy quality in the LILEQ measurement is operationalised in an indirect way (Mulder et al, 2023).

Selecting the most optimal energy poverty indicators is important to measure all elements of energy poverty. The measurements should be able to include households with an excessive energy burden or energy expenditure, households with a residual income below a monetary poverty line after their energy expenses have been deducted, and households with a low actual energy consumption that can be seen as hidden energy poverty (Herrero, 2017). Kruit et al (2021) include hidden energy poverty in their measurement by analysing the households that have energy poverty according to the 10% and MIS indicator for standard energy consumption but that do not have this when looking at the measured energy usage. Figure 2 shows the average energy poverty percentages in the European Union that were found using the different energy poverty indicators that were analysed. This figure shows that the amount of energy poverty varies greatly between different energy poverty indicators.

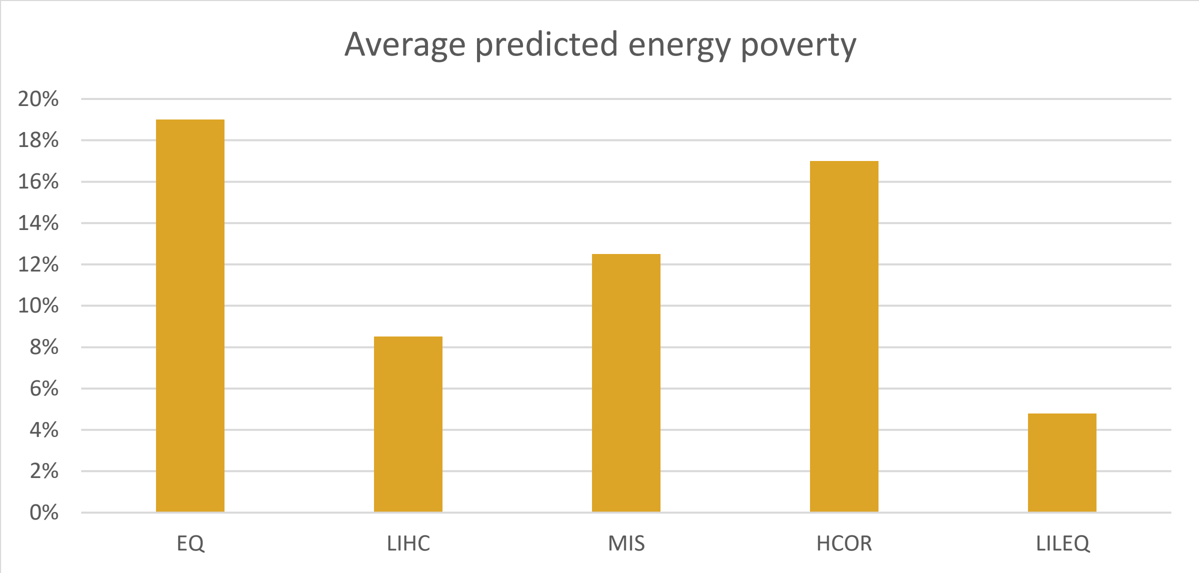

Figure 2. Average energy poverty of analysed indicators (Based on: Baardink, 2020; Coene & Meyer, 2019; Kruit et al, 2021; Mulder et al, 2021; Romero et al, 2018; Sokołowski et al, 2020)

2.3 Energy poverty factors

Energy

Multiple existing researches stress the importance of relation between energy prices and energy poverty. All these researches agree that when energy prices increase, so do the number of households experiencing energy poverty when no measures to improve energy performance are implemented. In their research about energy poverty in Europe, Halkos & Gkampoura (2021) identify energy prices as the main cause of energy poverty and Mulder et al (2023) add that a minor increase of the gas price may lead to a significant increase in energy poverty. While both the prices for electricity and gas have increased in the past years, there is a trend in developed countries where the electricity usage increases while the gas usage decreases (Coene & Meyer, 2019). The energy price is especially important when using the EQ, since this indicator is extremely sensitive to energy price changes (Romero et al, 2018).

Dwelling

A second factor determining energy poverty that most energy poverty researchers agree on is a poor energy performance of dwellings. Both Baarding (2020) and De La Paz et al (2022) conclude that energy poverty is characterised by dwellings with a poor energy performance. When a building has poor insulation and an aged heating system more energy is required to create a comfortable indoor climate. The energy label is often used as a simple indicator of the energy performance of a building and according to Kruit et al (2021) there is an increased risk of energy poverty when a building has an energy label C or worse.

Bonnard et al (2015) and Coene & Meyer (2019) concluded that rent prices have an effect on energy poverty. Especially households that live in a relatively cheap rental dwelling are confronted with a substantial increase in the rental prices, reducing their spendable income after deduction of housing costs (Coene & Meyer, 2019).

Regarding dwelling type and energy poverty, most researches agree that households living in detached houses have a larger risk of experiencing energy poverty, followed by semi-detached houses. This increased risk of experiencing energy poverty for household living in detached houses is caused by the fact that detached households often have a lower energy quality. Additionally, detached houses require more energy to heat the dwelling to a comfortable temperature compared to other housing types. Households living in apartments have a lower risk of experiencing energy poverty. (Delbeke et al, 2013; Papada & Kaliampakos, 2016; Romero et al, 2018; Sokołowski et al, 2020). Although Maxim et al (2016) could not determine a significant relation between energy poverty and the dwelling type they argue that the differences in the energy poverty between different housing types are related to the dwelling exposure to the environment. Detached and semi-detached houses require additional energy for heating and cooling and have therefore a higher risk for energy poverty compared to dwellings with a higher thermal efficiency. Papa & Kaliampakos (2016) did find a significant relation between energy poverty and the dwelling type. In their research, they discovered that energy poverty occurred twice as often in detached houses compared to apartments.

Delbeke et al (2013), Okushima (2017), Kruit et al (2021), and De La Paz et al (2022) agree that a larger dwelling size increases the risk of energy poverty since the size of the dwelling determines its energy performance through increasing energy usage with dwelling size.

Another factor related to the energy performance of dwellings is the age of the dwelling. Households living in older dwellings experience more energy poverty than households living in newer dwellings. Older dwellings are in general less insulated and therefore require more energy to create a comfortable indoor climate, increasing the energy bill for households living in these older dwellings (Kruit et al, 2021). Papada & Kaliampakos (2016) elaborate that the energy performance of 71% of the Greek residences built before 1979 have not been improved throughout the years. Sokołowski et al (2020) concluded that the risk of energy poverty increases with building age and the effect of building age on energy poverty was largest for buildings constructed before 1946. Adding to this, Kruit et al (2021) concluded that the number of households experiencing energy poverty living in dwellings constructed after 1992 is relatively minor compared to the older dwellings. Furthermore, they concluded that households living in dwellings constructed before 1964 experience the most hidden energy poverty while hidden energy poverty does not exist for households living in dwellings constructed after 2006.

The location of the dwelling has an influence on the risk of energy poverty. Most researchers conclude that people living in areas with a low degree of urbanisation experience more energy poverty than people living in cities (De La Paz et al, 2022). This difference is caused by both lower average incomes in rural areas and because of the urban heat island effect, reducing the amount of energy needed to heat a dwelling to a comfortable temperature in urban areas (Sokołowski et al, 2020; Halkos & Gkampoura, 2021). Bouzarkovski et al (2021) reach the same conclusion, except for Spain where energy poverty is less common in the rural areas. Focusing on the Dutch context, Kruit et al (2021) concluded that households in Groningen and Friesland experience the most energy poverty. Mulder et al (2021) concluded that energy poverty primarily occurs for households living in peripheral regions of the Netherlands and some densely populated urban districts outside of the Randstad. Mulder et al (2023) add that while poverty is generally found in urban areas, energy poverty is primarily located in the rural areas. According to their research it is not clear if there is a link between location and income. Additionally, they question whether the energy poverty differences between urban and rural areas are caused by a higher number of recent urban renovation projects resulting in dwellings with better energy qualities compared to dwellings in rural areas. Although they find differences in energy poverty figures between urban and rural areas, they conclude that these differences are not significant.

Socio-demographics

The income is another factor that is considered to be relevant in the occurrence of energy poverty by all existing literature about energy poverty. Baardink (2020) and De La Paz et al (2022) found that the chance of experiencing energy poverty increases with the decrease of income and Pachauri & Spreng (2011) also found a significant relation between monetarily poverty and energy poverty. Okushima (2017) add to these findings, stating that energy poverty is especially present among the lowest 30% of incomes. The relation between income and energy proves to be a source of social injustice. The share of a household income that is spend on energy is higher for the lower incomes while this group often uses less energy than households with higher incomes. Because of this, the amount of energy poverty indicated by the 10% energy quote indicator may be different for people, depending on if they have a high, average, or low income. Additionally, someone can experience energy poverty while not living in monetarily poverty (Bonnard et al, 2015). Energy poverty is an issue that is primarily targeted at low-come households that often have difficulties to eliminate the causes of energy poverty (Li et al, 2021). Because of this, most existing researches agree that unemployment is a significant predictor of energy poverty (Baardink, 2020).

Another factor that can be used to analyse energy poverty is the payment risk, occurring when a household has insufficient funds for life costs after paying for the energy bills and housing costs (Kruit et al, 2021). According to Kruit et al (2021) about 5.5% of the Dutch households have a payment risk and Kruit et al (2021) define energy poverty as an affordability problem through a combination of a high energy quote and payment risk.

The risk of experiencing energy poverty differs between different household types. According to Delbeke et al (2013) especially single-person, single-parent, and large households experience energy poverty. Except for the large households, most other researchers agree with the finding that single-person and single-parent households experience more energy poverty (Okushima, 2017; Baarding, 2020; Kruit et al, 2021; Rao, Tang, Chau, Iqbal & Abbas, 2022). According to Coene & Meyer (2019) more than a third of all single-parent households experience energy poverty while this group is only 7.2% of all households. More than half of all households that experience energy poverty are single-person households while this group is only a third of all households. The increased risk of experiencing energy poverty for single-parent households is related to an average lower income and increased poverty risk for these types of households. Coene & Meyer (2019) expect that with the current demographic developments, energy poverty will increase further, especially for single-parents.

Multiple conclusions about the effect of age on energy poverty are found in during the literature review. According to Romero et al (2018) members of a household aged under fourteen increase the chance of energy poverty while members of a household aged 65 or over reduce the probability of being energy poor. Romero et al (2018) add that the elderly do not seem to be a vulnerable group, a finding that is supported by Halkos & Gkampoura (2021) who conclude that elderly seem to be affected less by energy poverty. These findings are contrary to other researches that concluded that elderly have a higher probability of experiencing energy poverty (Okushima, 2017; Coene, & Meyer, 2019; Churchill & Smyth, 2021; Rao et al, 2022). Delbeke et al (2013) support these findings by explaining that elderly spend more time at home and prefer a higher indoor temperature which causes them to use more energy. According to De La Paz et al (2022), both old and young household members increase the chance to experience energy poverty.

The education level is another factor that has a relation with energy poverty, where a someone is more likely to be energy poor when they have a low level of education (Delbeke et al, 2013; Romero et al, 2018; Lamain, Elbert & Ottema, 2021). People with a low educational level have higher risks of becoming unemployed and apply less energy saving measures because they are less aware of these measures (Delbeke et al, 2013; Lamain et al, 2021).

The tenure status a considerable influence on the probability for a household to experience energy poverty. Households living in owned properties are less likely to experience energy poverty compared to households living in rented dwellings. Living in a rented dwelling doubles the probability of being energy poor compared to living in an owned dwelling according to Romero et al (2018). However, according to Bonnard et al (2015), households renting a dwelling experience four times as much energy poverty as households living in an owned dwelling. Baardink (2020) concluded that energy poverty in Groningen primarily occurs in the social and private rental sectors. However, multiple studies have contradicting conclusions to whether energy poverty is more prevailing in the social or private rental sector. According to Huybrechs et al (2011) energy poverty occurs most in the private rental sector, followed by the social rental sector. Coene & Meyer (2019) concluded that tenants in the private sector represented a third to half of all households experiencing energy poverty, while this group is only 22% of the entire population. Kruit et al (2021) and Mulder et al (2021) concluded that most households experiencing energy poverty live in a social rental dwelling. According to Mulder et al (2021) even 75% of all households experiencing energy poverty live in a dwelling owned by a social housing corporation. Households living in rented dwellings often have low incomes and therefore these households cannot invest in improving the energy-efficiency of their dwelling when the owner of the dwelling is not willing to invest in improving the energetic performance of the building (Bonnard et al, 2015). Next to this, tenants often have little to no interest to improve the energetic performance of their dwelling because they are not the owners of the dwelling. Adding to this, existing measures to improve the energy performance of dwellings are currently mainly aimed at owners and therefore owner-occupied dwellings are more often improved energetically compared to rental dwellings (Delbeke et al, 2013). An overview of the identified characteristics and their expected effect on energy poverty are shown in appendix 1.

2.4 Energy poverty consequences

Energy poverty has multiple consequences such as: impacts on health, the economy, and the environment (González-Eguino, 2015). Thomson, Snell & Bouzarovski (2017) concluded that there is a significant difference in health between the energy poor and the non-energy poor. Energy poverty coupled with the poor energetic performance of a dwelling can form a threat to the health of residents (Goedemé et al, 2017). Energy poverty can cause moist and mould problems to develop within dwellings, resulting in indoor pollution (Delbeke et al, 2013). This indoor pollution can lead to respiratory and cardiovascular diseases, lung cancer and pneumonia (González-Eguino, 2015). Energy poverty can deteriorate the mental wellbeing of people and can be a large cause for stress and depression (Bouzarovski et al, 2021; Kruit et al, 2021). Additionally, energy poverty can limit the activities that a person can perform, resulting in social isolation and exclusion (Huybrechs et al, 2011; Coene & Meyer, 2019; Baardink, 2020). Countries with relatively high amounts of energy poverty also reported increased mortality rates and access to energy can decrease mortality rates and increase life expectancy (Churchill & Smyth, 2011; Pan, Biru & Lettu, 2021). Pan et al (2021) concluded that countries with a higher standard of living could mitigate the negative effect of energy poverty on health.

Energy poverty impacts the economy of a country by reducing the gross domestic product and can decrease the social welfare of a country dramatically if the energy poverty is long-term consistent and not mitigated (Li et al, 2021). The environment is impacted by energy poverty through climate change, where energy poverty can lead to deforestation, land-use changes, and increased greenhouse gas emissions (Sovacool, 2012).

2.5 Energy poverty mitigation

There are multiple barriers to successfully mitigate or eliminate energy poverty, these include technical, financial, political, and social (Sovacool, 2012). In order to overcome these barriers energy poverty needs to be incorporated into the design of policies (González-Equino, 2015). Since inefficient policies can increase energy poverty (Li et al, 2021), the design of policies to reduce energy poverty requires detailed information about which subgroups of the population are most affected by energy poverty and the determinants of energy poverty (Pachauri & Spreng, 2011). In general, three tracks are used to mitigate energy poverty: increasing the awareness about energy poverty to reduce energy usage, financial measures to alleviate energy burdens, and sustainability measures to improve the energy-efficiency of dwellings (Kruit et al, 2021). Measures aimed at increasing the awareness about energy poverty are designed to tutor people about saving energy and energy policies (Churchill & Smyth, 2021). However, these measures should always be combined with other measures since awareness measures alone yield too little to create a structural energy transition (Kruit et al, 2021).

Financial measures to alleviate energy burdens such as social tariffs and income supports are specifically aimed at improving the affordability of energy for low-income and vulnerable households (Okushima, 2017). However, social tariffs are only effective when the vulnerable households in the target group are the actual recipients (Romero et al, 2018). In a financial program in Greece, residents could be supplied with a subsidy based on financial criteria. However, this financial program was not successful since the low-income and vulnerable households were not the recipients of these subsidies and they were often the first ones that were excluded from the subsidies (Papada & Kaliampakos 2016). Additionally, financial supports are only effective to temporarily lift the energy burden of low-income households and they do not reduce the high initial energy burdens and support the energy transition (Kruit, 2021). Opinions about financial incentives to mitigate energy poverty differ between people that approach energy poverty from an ecological view and those that approach energy poverty from a social view. People with an ecological view on energy poverty strive to make households reduce their energy usage while people with a social view strive to provide households experiencing energy poverty with more energy. Because of this, solely financial initiatives such as a reduction of taxes on energy are not effective from an ecological view since this will result in increased energy consumption and negative effects on the environment (Delbeke et al, 2013). Additionally, reducing the energy costs for households removes the incentive to invest into improving the energy-efficiency of dwellings and does therefore also not contribute to other environmental goals such as the reduction of CO_2 emissions (Ürge-Vorsatz & Tirado Herrero, 2012; Goedemé et al, 2017).

The final common track to mitigate energy poverty is to apply sustainability measures to improve the energy-efficiency of dwellings to reduce energy bills (Kruit et al, 2021). Improving the energy-efficiency of buildings has the potential to contribute to both the elimination of energy poverty and reduction of CO_2 emissions in the long term (Ürge-Vorsatz & Tirado Herrero, 2012). Additionally, stimulating improvements to the energy-efficiency of dwellings can be done through measures that are satisfactory from both an ecological and a social view (Delbeke et al, 2013). Improving the energetic performance of a dwelling can be done according to the three consecutive steps described by the Trias Energetica of Lysen (1996). The first step is a reduction in the need for energy through insulation and airtightness of the dwelling. Because of this step, residents of dwelling will need less energy to create a comfortable indoor climate. The second step is to supply the household energy needs as much as possible by sustainable energy generated by the dwelling through solar panels, solar collectors, and heat pumps or by solely purchasing sustainable energy. The third step is to use technical installations that are as efficient as possible to minimise heat and energy losses. The improvements of the Trias Energetica reduce the energy requirements of households, making them less sensitive to increasing energy prices and reducing their risks of experiencing energy poverty (Lysen, 1996).

More municipalities in the Netherlands aim to improve the sustainability of their housing stock. However, improving the energetic performance of dwellings is not always affordable for everyone. The benefits of improving the energetic performance of a dwelling are primarily long-term benefits while poorer households are often only interested into short-term benefits. Because of this, energetic improvements of dwellings of poorer households are often not performed (Baardink, 2020). The energy transition offers different opportunities for dwelling owners than for dwelling tenants. Tenants cannot improve the energetic performance of their dwelling themselves and are dependent on the willingness of the housing corporation to improve the energetic performance of the dwelling (Mulder et al, 2021). Housing corporations are often not interested in improving the performance of dwelling because the costs for improving the energetic performance of a building are generally not paid back through energy savings (Kruit et al, 2021). Additionally, so-called split incentives occur when a property owner invests in improving the energy efficiency of their building but cannot experience the direct benefits from these investments that are often only benefits for tenants such as reduced energy needs. Because of this, investing in sustainability improvements become less interesting for the property owner, unless the rent price can be increased. However, if the rent price is increased, this can be detrimental to the often financially vulnerable tenants that may still experience energy poverty. Social housing corporations have to pay for sustainability investments to their housing stock while the benefits of these investments regarding energy savings are mainly for tenants. However, because social rents are capped by law, the cost of the energy improvements cannot simply be passed on to the tenant, even if their energy costs are reduced (Goedemé et al, 2017).

Delbeke et al (2013) advise improve the energetic performance of dwellings on the private rental market. Households with limited financial resources should be provided with an opportunity to invest in sustainability and be able to perform energy-saving improvements. Additionally, the information about energy poverty and solutions should remain accessible and simple so everyone can understand the problem and what can be done to mitigate it (Delbeke et al, 2013). Serial renovations should be executed to make sure that many social housing dwellings can be renovated in a brief period, with minimal impacts to their tenants. This is however only possible when the dwellings in the series are relatively similar (Goedemé et al, 2017). National governments can stimulate the affordability of sustainability measures through sustainability subsidies and the payback period of sustainability investments is currently relatively favourable (Baardink, 2020). Policies that stimulate the improvement of the energetic performance of dwellings can reduce energy poverty and support the energy transition. Investment costs for improving the energetic performance of dwellings are relatively high however these investments create a structural long-term positive effect. The effect of improving the energy performance of dwellings on energy poverty is however relatively slow. Because of this, the largest effects to reduce energy poverty can be achieved when the dwellings of households experiencing energy poverty are given priority in renovations since these dwellings will offer the largest potential savings (Kruit et al, 2021).

Kyprianou et al (2019) also advise to target the most effective and vulnerable target group with policies and financial aids. The government can play a vital role in order to reduce energy poverty by investing in the housing sector and stimulating the renovations of dwellings to improve their energy performance (Bonnard et al, 2015). Government policies should reduce the effects of energy poverty and increase the motivation of residents to invest into energy poverty mitigation (Li et al, 2021). Policies should focus on energy saving and improving the energy performance of the building stock. Additionally, these policies should be targeted at the population experiencing energy poverty and not only the low-income population. Finally, a more decentralised approach can be beneficiary to the mitigation of energy poverty since regional policies to reduce energy poverty are more effective than those at the national level (Kyprianou et al, 2019; Bouzarovski et al, 2021).

2.6 Energy poverty future

Because of the unpredictable nature of energy poverty and some of its factors, it is difficult to predict how energy poverty will develop in the next decade. Much will depend on how the energy prices will develop and the speed in which the energy-efficiency of dwellings is improved (Huybrechs et al, 2011; Delbeke et al, 2013). According to Mulder et al (2021) many people that indicated to currently have no problem to afford their energy bills will experience energy poverty when the gas price increases further in the future. If the energy performance of the housing stock is not improved, the combination of the current energy transition policy and energy prices prognoses will cause an increase in the number of households that experience energy poverty of more than 30% by 2030. In order to avoid this increase, measures and investments will be needed that are targeting households with an elevated risk of experiencing energy poverty. These measures and investments should provide financial support and reduce the energy needs of the households (Kruit et al, 2021).

3. Research design

Multiple analyses will have to be performed in order to achieve the set research goals and answer the research questions. This chapter will explain the created conceptual model, the scope of the research, and the methods and programs that were used to analyse the data.

3.1 Conceptual model

Figure 3 shows the conceptual model that was created with the aim to predict energy poverty through multiple dwelling, socio-demographic and economic factors. The conceptual model is aimed at determining the indirect effects of variables on energy poverty and the effect of variables on energy poverty that had not been included in existing energy poverty research.

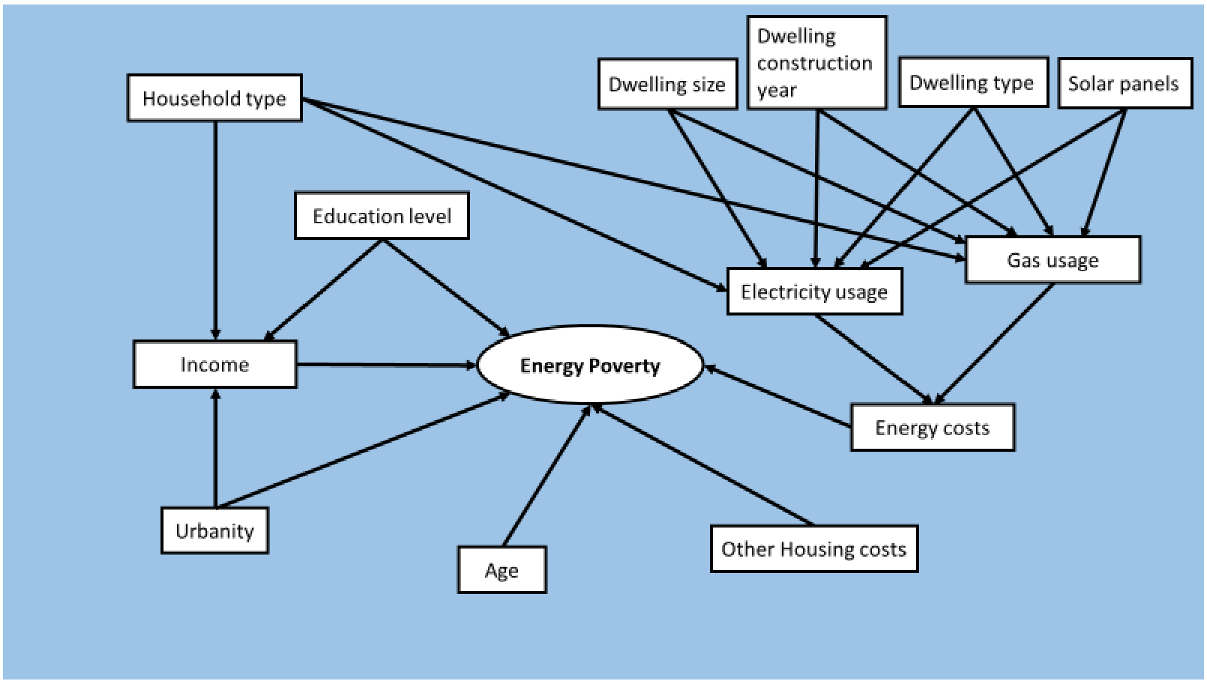

Figure 3. Conceptual model

Based on the literature review, it is expected that the household type, dwelling size, dwelling construction year, dwelling type and the presence of solar panels will not have a direct effect on energy poverty. It is however expected that these variables will have an indirect effect on energy poverty through energy usage and energy expenditure. It is expected that multiple variables will have an indirect on energy poverty through income. Based on the literature review, it is expected that the education level, income, urbanity, age, other housing costs, and the energy costs will have a direct effect on energy poverty. The expected relations between the variables and energy poverty are shown in appendix 1. The relation between age and energy poverty included in the conceptual model remains a research gap caused by the many contradictory findings about this relation in the literature review. The created conceptual model includes factors of energy poverty that are included in existing energy poverty indicators and additional factors that were identified to be related to energy poverty in the literature review and that have not yet been included in existing energy poverty indicators. These new factors are, for example, personal characterisitcs such as the household type, education level, and age. Other factors that have not yet been included in existing energy poverty indicators are dwelling characterisitcs such as the dwelling type, size, construction year, and presence of solar panels. Finally, the effect of the urbanity of a neighbourhood on energy poverty has not yet been included in existing energy povety indicators.

3.2 Scope

The scope of this research will not include the negative effects of energy poverty on health since this topic itself is relatively broad and has already been the topic of many existing researches. The research is focused on households living in social housing since this group is the most affected by energy poverty according to the literature review. Dwellings connected to heat networks will be filtered from the dataset since these dwellings are not comparable to other dwellings regarding energy usage and expenditure and will therefore cause skewed results. Energy prices will not be included in this research since these may vary greatly depending on time and are very dependent on the energy contracts that respondents have. Additionally, this research will not include hidden energy poverty. While hidden energy poverty is similar in name to energy poverty, it is determined in a completely different way and should therefore be analysed in another research. Finally, since this research focusses on energy poverty and energy usage in dwellings, transport poverty and its effects on energy poverty will not be included.

3.3 Methods

Descriptive/bivariate analyses

The first analysis that will be performed after the required data is prepared is a descriptive analysis. The descriptive analysis will show the frequencies of the value categories of each of the selected variables. Additionally, the descriptive analysis will show how the respondent cases are distributed for all original and recoded variables. The second analysis that will be performed is a bivariate analysis of the selected and recoded variables. The bivariate analysis will test how well the demographic and socio-economic variables in the data sample represent the larger population. Additionally, other bivariate analyses will test the relations between exogenous variables in the created conceptual model. Both the descriptive and bivariate analyses will be performed and analysed using the Statistical Package for the Social Sciences (SPSS). SPSS is a statistical software platform that was introduced in 1968. SPSS can perform statistical analyses and create visualisations with the aim to extract insights from big data.

Exploratory factor analysis (EFA)

Energy poverty is regarded as a latent variable in this research since the occurrence of energy poverty cannot be observed and has to be interpreted according to indicators. Latent variable modeling techniques are required to analyse models that include latent variables. Factor analysis (FA) will be used to analyse the relations between existing energy poverty indicators and the latent concept of energy poverty. FA was introduced in 1904 and allows for the reduction of multiple variables into a smaller set by establishing the underlying dimensions between the observed variables and a latent construct. Exploratory factor analysis (EFA) is a version of FA that is used when a researcher has no expectations or prior theory about the relations between observed variables included in the factors (Taherdoost, Sahibuddin & Jalaliyoon, 2022). EFA can be used to determine if the observed variables measure the same concept. The basic hypothesis of a EFA is that there is an underlying latent concept that relates multiple observed variables in the dataset. The goal of the EFA is to find the smallest number of factors that will account for the correlations with the latent factor (Young & Pearce, 2013). If the observed variables are unrelated, an EFA will not find a common underlying concept. Figure 4 shows an overview of all steps that will be taken during the EFA.

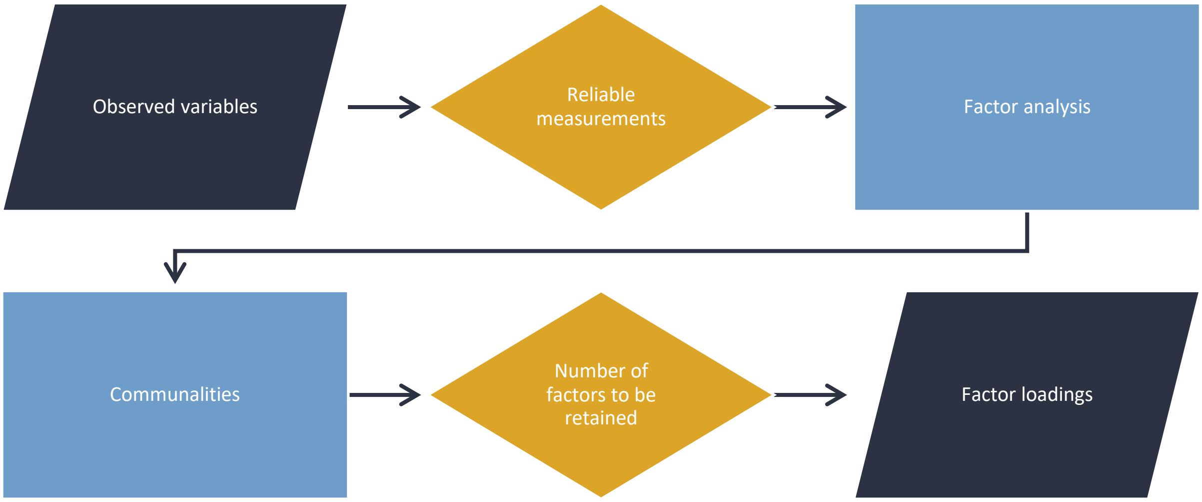

Figure 4. Overview of EFA steps (Based on: Kootstra: 2004)

When conducting an EFA, the first step is to create a list of observed variables that will be included in the EFA model (Latif, n.d.). All observed variables should be measured on a continuous scale and the sample size should be sufficient. There is no general consensus about a sufficient sample size for EFA and suggestions for the number of respondents vary between at least 10-15, at least 50, at least five times as much as the included variables (Kootstra, 2004); at least 300 (Young & Pearce, 2018); at least 150 and 10 per variable (Hooper, 2012).

The second step of conducting the EFA is to determine if the variables in the model are reliable measurements. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy can be tested and if the value of the KMO is greater than 0.5, the data sample is adequate for an EFA (Kootstra, 2004). The correlation matrix has to be analysed in order to analyse the correlation among the observed variables. In order to successfully conduct the EFA, no or limited correlation between the observed variables should be avoided (Latif, n.d.). To be appropriate for EFA the observed variables have to be correlated however they should not have too high correlations, as this can lead to multicollinearity and singularity. Bartlett's test of sphericity can be used to check the correlations between observed variables used in the EFA. The EFA can be successfully conducted if Bartlett's test of sphericity is significant. Multicollinearity can be checked by analysing the determinant of the correlation matrix. If the determinant is greater than 0.00001, it can be concluded that there is no multicollinearity (Kootstra, 2004).

In the third step of the EFA the FA is conducted and in the fourth step the communalities are estimated. If the goal of the FA is to summarize the data of multiple observed variables into a latent variable, the principal axis factoring technique is selected. The communalities describe the amount of variance that a variable shares with the other variables taken into account in the EFA (Latif, n.d.). In FA it is assumed that the variables do not account for 100% of the variance (Kootstra, 2004). When a variable included in the EFA has a relatively small communality value, the variable does not contribute much to the measurement of the underlying factor and should therefore be removed from the FA (Latif, n.d.).

In the fifth step of the EFA, the researcher has to determine the number of factors that have to be retained. Most commonly, this is done according to the Kaiser-Guttman rule which suggests that only factors with an eigenvalue larger than 1 should be retained (Kootstra, 2004). The eigenvalue represents the amount of the common variance of the observed variables that a factor explains. Factors with an eigenvalue larger than 1 explain more variance than a single observed variable while factors with eigenvalues less than 1 explain less variance than a single observed variable. Factor with an eigenvalue less than 1 are therefore not retained in the analysis (Qualtrics, n.d.). Other methods that are used to determine the number of factors that should be retained in the EFA are keeping the factors which account for about 70-80% of the variance or keeping all factors before the breaking point of a scree-plot (Kootstra, 2004). A scree plot is a graph that visualises the eigenvalues and factor numbers according to the order of extraction (Latif, n.d.) Since the goal of this research is to determine if the factors have one common underlying factor, only one factor will be retained.

The final step of the EFA concerns the interpretation of the results. The EFA will determine factor loadings, also known as factor scores, which describe how strongly one of the observed variables is related to a given factor (Kootstra, 2004). The factor loadings are included in the classical factor analysis model as described by Yong & Pearce (2013):

- $X_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m + e_j$ (6)

In this model, m represents the number of observed variables (F_1, F_2,…,F_m), X_j represents the latent factor and the factor loadings are represented as: a_{j1}, a_{j2},…, a_{jm}. The model assumes that there is an underlying factor consisting of a linear function of observed variables and a residual variate. Using this model, it can be determined how much of each observed variable contributed to the latent factor since a larger factor loadings means that the observed variable contributes more. Since factor loadings represent the strength of the correlation between an observed variable and the factor, they are similar to regression coefficients in multiple regression analysis (Yong & Pearce, 2013). In fact, factor loadings can be interpreted like standardised regression coefficients. In an EFA, factor loadings larger than 0.65 are considered to be strong associations (The Analysis Factor, n.d.). In this research all steps of the EFA will be conducted using SPSS.

Structural equation modeling (SEM)

A structural equation model (SEM) will be developed in order to predict energy poverty. SEM is a framework that integrates measurement theory, latent variable factor analysis, path analysis, regression, and simultaneous equations. By including both direct and indirect effects of variables on other variables in the model, SEM enables the analysis of a system of relationships rather than the effects of predictors on a single dependent variable. By looking at the relationships between variables, SEM enables the application of path analysis with latent concepts that are not directly observable such as energy poverty (Sturgis, 2016).

Multiple indicators with a high correlation with the latent factor (factor loadings) are needed to successfully identify components of the SEM. Additionally, a SEM must be over-identified, the SEM must include more observed than latent variables, and there must be enough information in the data to estimate the SEM model coefficients. SEM includes both exogeneous and endogenous variables and within a SEM a variable can be both a predictor and an outcome (Sturgis, 2016).

SEM is a popular technique used to model complex and multivariate systems. In a SEM, latent variables are constructed through a confirmatory factor analysis (CFA) (Jaya, Hermina & Sunengsih, 2019). CFA differs from EFA since the measurement model of the CFA is specified before identifying the data. CFA is used when researchers want to confirm existing hypotheses or theories rather than explore data (Frost, n.d.). Because of this, it should be predetermined which indicators measure which factors and which indicators are unrelated to which factors. Since latent variables have no measurements, a SEM allows a researcher to constrain one factor loading to the value of 1. By doing so, the connected variable will become a reference item and the latent variable will have the same scale as this reference item, resulting in a fully standardised solution (Sturgis, 2016). In a SEM, path analysis is used to construct the relationships between the latent and observed variables.

There are two common approaches of SEM: variance based-SEM (VB-SEM) and covariance based-SEM (CB-SEM). CB-SEM can evaluate complex SEM models though a parametric approach however strong assumptions must be satisfied to have a good result. Examples of these assumptions are that a large sample size is needed, and that the data must be normally distributed. CB-SEM is very sensitive to non-normality and a high variability in the data will violate the assumptions of CB-SEM. VB-SEM can be used when the assumptions of CB-SEM are violated. VB-SEM, also called partial least square SEM (PLS-SEM) is more flexible regarding the normality of the distribution of the data and sample size requirements. Additionally, where CB-SEM is used for confirmative purposes only, VB-SEM can also be used for predictive purposes. The aim of VB-SEM is to maximise the explained variance of the latent variable (Jaya, Hermina & Sunengsih, 2019). Table 2 shows the main differences between VB-SEM and CB-SEM. In this research, the VB-SEM approach will be used since the main objective of this research is prediction oriented. Additionally, VB-SEM will be used because of the non-normality of some of the observed variables that were identified in the literature review and that will be included in the SEM model.

Table 2: *Differences between VB-SEM and CB-SEM (Based on: Jaya, Hermina & Sunengsih, 2019; Hanafiah, 2020)*

	VB-SEM	CB-SEM
Objective	Prediction oriented	Parameter oriented
Approach	Variance	Covariance
Distribution assumption	Non-parametric	Parametric
Required sample size	Small (min. 30-100)	Large (min. 100-800)
Model complexity	High	Low to average (high complexity models become problematic)
Indicators per construct	No constraint	Minimum three to meet identification criteria
Measurement model	Reflective and formative	Reflective only
Goodness of fit measure	None	Many
Implication	Optimal for prediction	Optimal for parameter estimation
Software	SmartPLS, WarpPLS, PLS-Graph	Amos, Lisrel, MPlus

Three types of SEM models can be created to measure latent constructs: reflective, formative, and network models. In a reflective model the latent construct reflects on the observed variables and the model analyses how much the latent construct influences the observed variables. A requirement of a reflective model is that all observed variables should have a high collinearity with each other. In a formative model, the causality goes from indicators to the latent construct since observed variables are considered to form the latent construct (Sekar & Rai, 2018). Contrary to reflective models, the latent construct that is analysed is affected by the types of indicators that represent the construct and the number of constructs in the SEM model (Hanafiah, 2020). In formative models, all observed variables represent a unique part of the latent construct and therefore high correlations between the observed variables should be avoided. Reflective models are usually analysed with more subjective data while formative models are usually analysed with objective data. Network models are clusters with no defined directions between a latent construct and the observed variables (Sekar & Rai, 2018). For this research, a formative SEM model will be created and analysed since the latent construct energy poverty will be formed by the model and it is expected that all observed variables will have an unique effect on this latent construct.

There are relatively few evaluation criteria in order to determine the validity and reliability of formative measurement models (Jaya, Hermina & Sunengsih, 2019). Formative measurement models can be assessed by analysing the convergent validity, indicator collinearity, and significance of the indicator weights determined by the SEM model (Hair et al, 2021). However, analysing the convergent validity is only possible when the SEM model contains both formative and reflective constructs. VB-SEM software programs determine the significances of calculated indicator weights, and the indicator collinearity can be assessed through the variance inflation factor (VIF). Both the bivariate analysis and VIF will be analysed to determine if there are high correlations between the observed variables. These high correlations should be avoided since they have an impact on the estimations of the path coefficients and their statistical significances. If the VIF is larger than 3.3 there is a potential issue with the indicator collinearity (Diamantopoulos & Siguaw, 2006) and if the VIF is larger than 5.0 there is a certain indicator collinearity problem (Hair, Ringle & Sarstedt, 2011).

For this research, a VB-SEM will be created using the SmartPLS software. SmartPLS allows for the creations of VB-SEM models that include complex relationships. SmartPLS was released in 2005 and represents the SEM in the standard SEM notation used in academic research. Latent variables are represented by an ellipse, observed variables are represented by a rectangle and error terms are represented by a circle. Covariance paths are represented by a double headed arrow while regression paths are represented by a single headed arrow. The results of the VB-SEM model in SmartPLS will be path coefficients between the observed variables and the latent variable. These 'structural equations' explain how the variables are related to each other and the independent variable, both directly and indirectly. The VB-SEM that will be created in SmartPLS will include all relevant variables and will be able to predict energy poverty with the largest prediction power that is possible. From this SEM model it will be possible to draw several results that will be concluded and discussed in later sections of this research.

3.4. Data

This section describes the dataset and the variables of the dataset that were selected to perform this research. The variables selected from the dataset include all relevant variables from the literature review that were included in the conceptual model. All different data analyses in this research will be performed with these variables selected from the dataset.

The dataset that was selected for this research is the WoonOnderzoek Nederland (WoON) 2021 dataset. The WoON dataset is created through the WoON research, a successor of the Dutch housing preferences research that was performed since 1981. The WoON research is a large national research conducted every three years by the Centraal Bureau voor de Statistiek (CBS) since 2006. Since 2009, the WoON research is conducted by a collaboration of the CBS and the Ministerie van Binnenlandse Zaken en Koningrijkrelaties. The aim of the WoON research is to collect statistical data about the housing situation of the Dutch population and to gain insights into their wishes and demands regarding multiple topics. These topics include the housing situation of households, housing environments, satisfaction, housing preferences, relocation preferences, and housing costs (Ministerie van Binnenlandse Zaken en Koningsrelaties, 2022).

The WoON research is a sample survey aimed at inhabitants of the Netherlands over the age of seventeen. The data of the research is collected through personal interviews, telephone interviews and online surveys. Afterwards, this data is augmented by data from existing registers. The WoON survey has a sample size of at least 60,000 respondents in order to provide reliable information. A weighting factor is added to the results of the survey to correct for differences between the composition of the respondents in the sample and the total population of the Netherlands (Centraal Bureau voor de Statistiek, n.d.). The result of the WoON research is a dataset that is highly representative of the Dutch population. Because of this, the results of the WoON research function as an important base for the housing policy of the Dutch government (Rijksoverheid, n.d.a). The WoON 2021 dataset selected for this research is the most recent version of the WoON database. This version of the WoON dataset was presented in June 2022 and represents the period between 2018 and 2021 (Ministerie van Binnenlandse Zaken en Koningsrelaties, 2022).

3.5 Data preparation

KNIME: the Konstanz information miner tool was used to prepare the data for the analyses. KNIME is an open-source data analytics platform that was developed at the University of Konstanz in Germany. The first version of the tool was released in 2006 as an open-source platform with the aim to process, transform and analyse big data. KNIME has no limitations regarding the size of the data that needs to be processed (KNIME, n.d.). KNIME is a modular environment in which an interactive data pipeline (workflow) consisting of data processing nodes and flows can be assembled and executed. New data processing and model building algorithm nodes can be added to the model through a drag & drop function to easily expand an existing model. Nodes are selected from a large node repository consisting of data readers, data manipulation, data transformation, mining algorithms, machine learning, statistics, visualisation, and scripting nodes. Because of the graphical overview of the tool, the user is always able to view the results of the workflow between all data processing steps. This enables the user to easily explore and check the data throughout the workflow. All nodes in the model show a node status (configured, executed, failed). Because of this, a KNIME user can easily spot errors in the execution of the workflow and the user can then check these errors in the node dialog. KNIME has been designed as an open-source teaching, research, and collaboration platform. Because of this new algorithm, data manipulation or visualisation method nodes are still being added to increase the power of the tool (Berthold et al, 2006).

The KNIME workflow that was created consists of over 300 nodes to prepare the WoON 2021 dataset for the analyses. Nodes used in the created KNIME workflow include data readers, column and row filters, rule engines, math formulae, and data writers. First, all variables relevant for this research are selected from the large dataset. These variables and their measurement scales are shown in appendix 2. Second, the names or categories of some variables are recoded. Since variables in the WoON 2021 dataset often have abbreviated or coded names, some variables were recoded to have a different name that more clearly explained the variable that is represented. The value categories of some variables were recoded to reduce the amount of value categories that a categorical value could have, or to combine categories with relatively few responses. Next, all data was filtered for missing values. Since SEM cannot be performed with data that includes missing values, the respondents that had some data missing were removed from the dataset. After all data was filtered and the missing cases were removed, 8,907 respondents remained in the dataset for the analyses. For the remaining respondents in the dataset the EQ, LIHC, MIS, HCOR, and LILEQ energy poverty indicators were calculated to be analysed in the FA. Finally, the value categories of all categorical variables that are included in the SEM were recoded into dummy variables. This was needed since SEM models cannot include categorical variables unless they are recoded into dummy variables.

4. Results

In this chapter, the results of the analyses are conducted will be described. First the descriptive analysis shows the value categories of the included variables. Next, the bivariate analyses show the representativeness of the sample and the correlations between independent variables. The results of the EFA show if all existing energy poverty indicators measure the same latent concept. The coefficients of the SEM model and prediction model are shown. Finally, the analyses of the high risk group and effects of policies will be shown.

4.1 Descriptive analysis

The first analyses that were performed after all required data had been prepare were multiple descriptive analyses. The descriptive analyses of all categorical variables included in the research will be described in this section. All descriptive analyses will present both the frequencies and percentages of the value categories of the selected categorical variables. Additionally, the descriptive analyses will help to visualise how the respondent cases are distributed for all variables.

Categorical socio-demographic characteristics

The descriptive analyses of the socio-demographic variables included in the model were the first to be conducted. These variables consider the household or the respondent of the WoON 2021 questionnaire. The frequencies of the household type show that more than half of all respondents in the dataset (53.7%) are single-person households. The second largest value category of household type in the data is couple without children and the smallest value category is the other household composition category. Just over 20% of the respondents in the data live in a household with children. The frequencies and corresponding percentages of the household types of the respondents are shown in table 3. Because of the limited number of respondents of the other household composition category, respondents of this category were removed from the dataset. After the removal of these respondents, 8,658 respondents remained in the dataset.

Table 3: *Household type frequencies and percentages*

Household type		
	Frequency (count)	Percent
Single-person	4,780	53.7
Single-parent with children	917	10.3
Couple without children	2,054	23.0
Couple with children	907	10.2
Other household composition	249	2.8
Total	8,907	100.0

Most respondents in the dataset have a HAVO, VWO, or MBO education level, followed by the value category of the VMBO and MBO1 education levels. These education level value categories consist of over 60% of all respondents. 18.5% of the respondents in the dataset have a primary education level while 19.6% of the respondents have a bachelor or master’s degree. Table 4 shows the frequencies and percentages of the education levels of the respondents in the dataset.

Table 4: *Education level frequencies and percentages*

Education level		
	Frequency (count)	Percent
Primary education	1,599	18.5
VMBO, MBO1	2,318	26.8
HAVO, VWO, MBO	3,042	35.1
HBO/WO-bachelor	1,181	13.6
HBO/WO-master	518	6.0
Total	8,658	100.0

The final socio-demographic variable that was included in the descriptive analysis is the age of the respondents. The descriptive analysis shows that most respondents are in the value categories for the higher ages. Most respondents in the dataset are aged between 65 and 74, followed by those who are 75 years or older and those aged between 55 and 64. More than 40% of all respondents are aged 65 or olde. The frequencies and percentages of the variable representing the age of the respondents are shown in table 5.

Table 5: *Age frequencies and percentages*

Age		
	Frequency (count)	Percent
34 or younger	1,320	15.2
35-44	1,047	12.1
45-54	1,189	13.7
55-64	1,598	18.5
65-74	1,832	21.2
75 or older	1,672	19.3
Total	8,658	100.0

Categorical dwelling characteristics

After the descriptive analysis of the socio-demographic variables, the variables of the dwelling characteristics were analysed. Regarding the degree or urbanisation in which the households in the dataset live, most respondents indicated to live in a highly urbanised area. More than 60% of all respondents in the dataset live in an area with a high or very high urbanisation while about 20% of the respondents live in an area with a low or no urbanisation. The frequencies and percentages of the value classes for the degree of urbanisation of the locations of the households is shown in table 6.

Table 6: *Degree of urbanisation frequencies and percentages*

Degree of urbanisation		
	Frequency (count)	Percent
No	606	7.0
Low	1,133	13.1
Average	1,450	16.7
High	2,654	30.7
Very high	2,815	32.5
Total	8,658	100.0

When looking at the value categories of the dwelling sizes, the descriptive analysis shows a relatively normal distribution. Most households in the dataset live in a dwelling with a size between 75 and 99 square metres. Less than 1 percent of the respondents in the dataset live in a dwelling with a floor space of 150 square metres or more. Table 7 shows the frequencies and percentages of the dwelling sizes in the dataset. Because of the limited number of respondents of the 150m2 or more category, respondents of this category were removed from the dataset. After the removal of these respondents, 8,600 respondents remained in the dataset.

Table 7: *Dwelling size frequencies and percentages*

Dwelling Size		
	Frequency (count)	Percent
Less than 50m2	539	6.2
50-74m2	2,318	26.8
75-99m2	3,466	40.0
100-149m2	2,277	26.3
150m2 or more	58	0.7
Total	8,658	100.0

Most dwellings in the dataset were constructed between 1975 and 1991, followed by the dwellings constructed between 1965 and 1974. 12.6% of all dwellings in the dataset are in the value category of the newest dwellings while only 8.0% of all dwellings are in the oldest value category. The frequencies and percentages of the dwelling construction years are shown in table 8.

Table 8: *Dwelling construction year frequencies and percentages*

Dwelling construction year		
	Frequency (count)	Percent
1945 or older	690	8.0
1946-1964	1,518	17.7
1965-1974	1,851	21.5
1975-1991	2,471	28.7
1992-2005	985	11.5
2006 or newer	1,085	12.6
Total	8,600	100.0

Regarding the dwelling type of the dwellings in the dataset, the descriptive analysis shows that most dwellings are apartments (53.8%). Terraced and semi-detached dwellings both have more than 1,000 respondents in the dataset however there are only thirteen detached dwellings in the dataset. Because of the limited number of respondents of the detached dwelling category, this category was removed from the dataset. After this category was removed, 8,587 respondents remained in the dataset. Table 9 shows the frequencies and percentages of the dwelling types in the dataset.

Table 9: *Dwelling type frequencies and percentages*

Dwelling type		
	Frequency (count)	Percent
Apartment	4,624	53.8
Terraced	2,597	30.2
Semi-detached	1,366	15.9
Detached	13	0.2
Total	8,600	100.0

The final categorical dwelling variable considered in this research is the presence of solar panels on the dwellings. The descriptive analysis shows that most dwellings do not have solar panels, while 14.8% of the dwellings do have solar panels. The frequencies and percentages of the value categories of the presence of solar panels are shown in Table 10.

Table 10: *Solar panels frequencies and percentages*

Solar panels		
	Frequency (count)	Percent
Yes	1,269	14.8
No	7,318	85.2
Total	8,587	100.0

An overview of the categories, frequencies, and percentages of the categorical variables remaining in the dataset after the descriptive analysis is shown in appendix 3.

Continuous characteristics

After the descriptives of the categorical variables were analysed, the same was done for all continuous variables in the dataset. A histogram of the annual spendable incomes of the respondents is shown in figure 5.

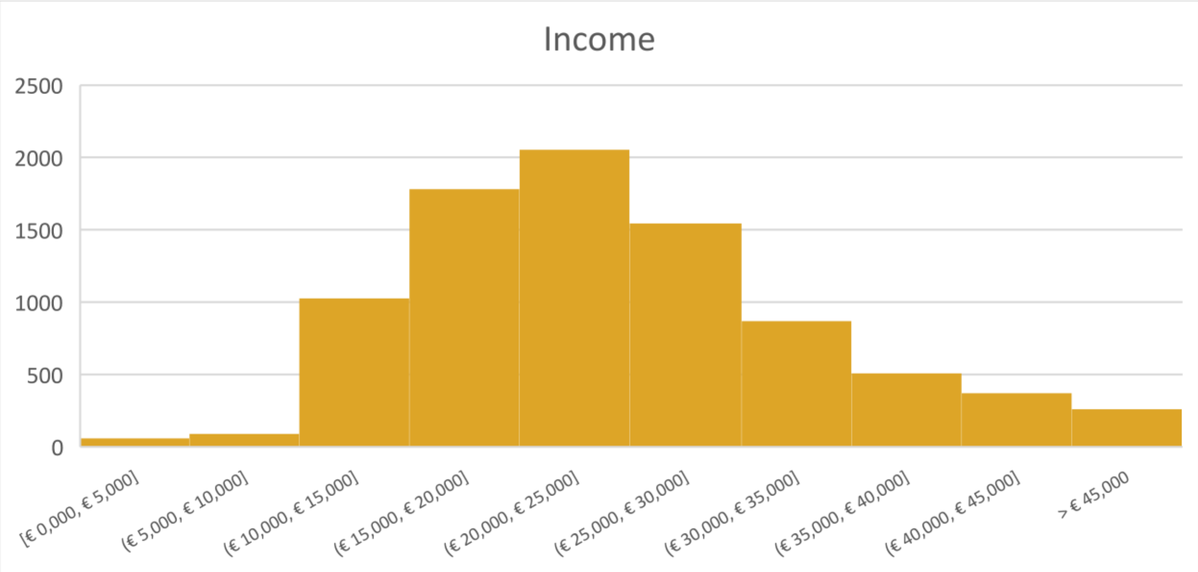

Figure 5. Histogram annual spendable income of the respondents.

Figure 5 shows that the annual spendable incomes of the respondents in the dataset have relatively normal distribution with a slight positive skew. Very few respondents have an annual spendable income below €10,000 and most respondents have an annual spendable income between €20,000 and €25,000. The average spendable income of all respondents in the dataset is €24,350.

After the descriptives of the continuous variable income were analysed, the same could be done for the annual energy expenditure of the households. Figure 6 shows a histogram of the annual energy expenditures of all households in the dataset.

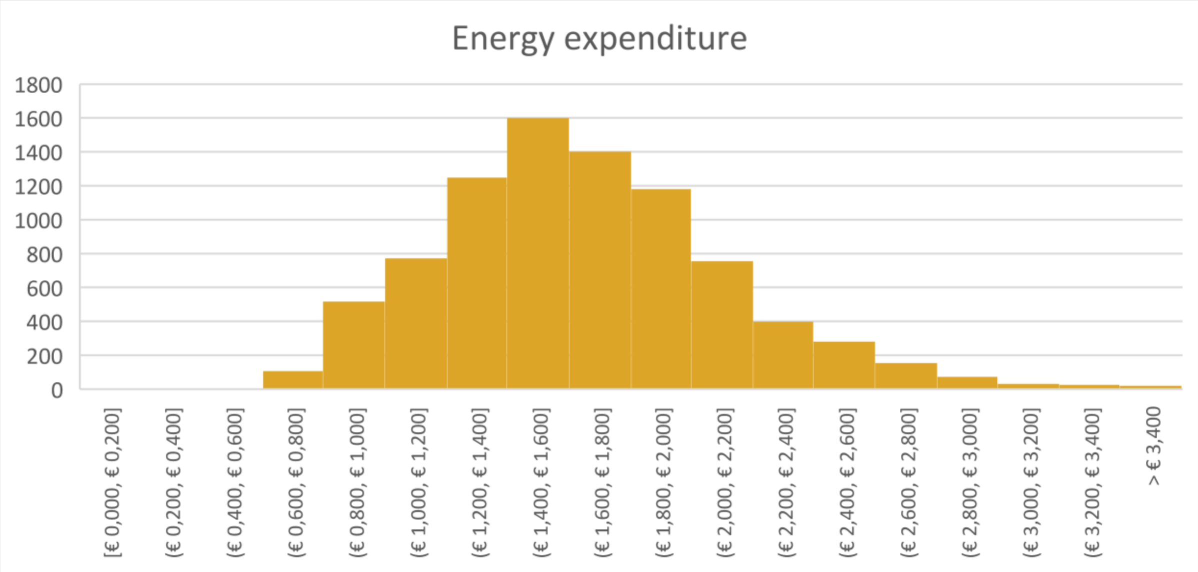

Figure 6. Histogram annual energy expenditures of the respondents.

The histogram in figure 6 shows that the annual energy expenditure of the households has a relatively normal distributions with a slight positive skew. Most respondents have an annual energy expenditure between €1,400 and €1,600 and the average annual energy expenditure of the households in the dataset is €1,650.

Figure 7 shows the histogram of the annual electricity usage of the households in the dataset.

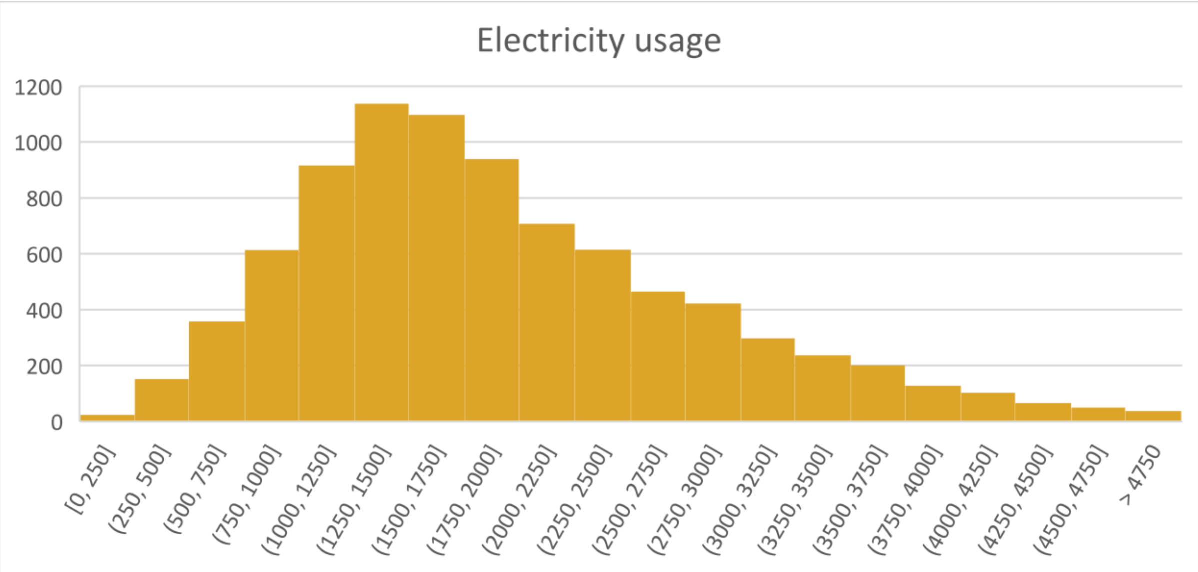

Figure 7. Histogram annual electricity usage [kWh] of the households.

Figure 7 shows that the annual electricity usage of the respondents is normally distributed with a positive skew. Most have an annual electricity usage between 1,250 and 1,500 kWh and the average annual electricity usage of the respondents in the dataset is 1,920 kWh.

The histogram of the annual gas usage of the households in the dataset is shown in figure 8.

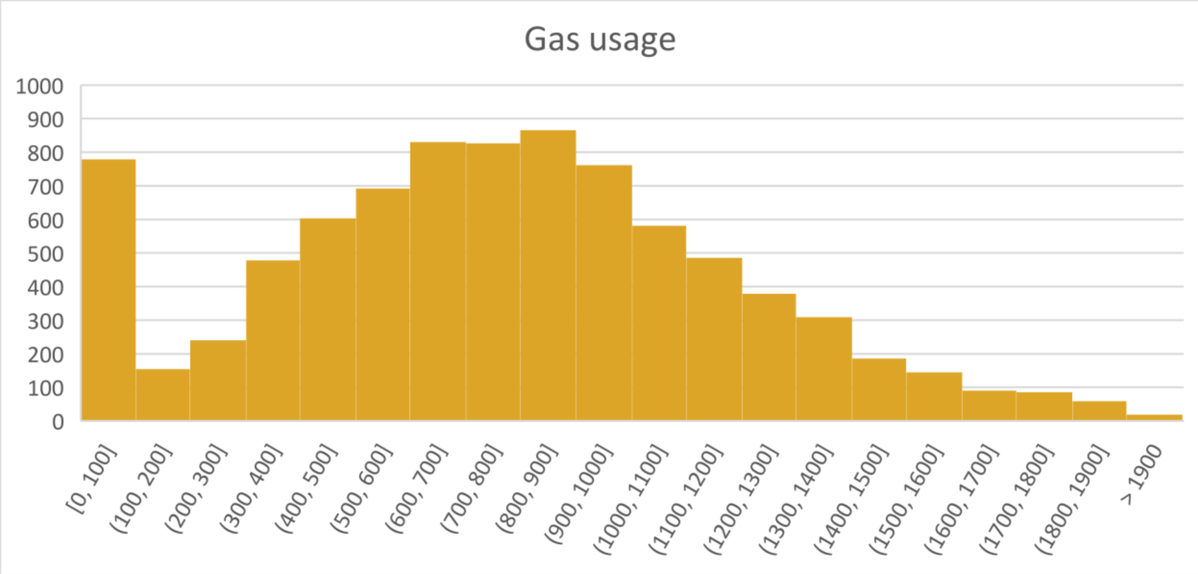

Figure 8. Histogram annual gas usage [m^3] of the households.

The histogram in figure 8 shows that there are relatively many households that have an annual gas usage in the category between 0 and 100 m^3. This category is relatively large since it includes all gas-free dwellings in the dataset. Other than the first category, the histogram shows a relatively normal distribution with a positive skew. Most households in the dataset have an annual gas usage between 800 and 900 m^3 and the average annual gas usage of the respondents in the dataset is 762 m^3.

The final categorical variable included in the conceptual model and dataset is the variable representing the annual other housing costs of the households. A histogram of the annual other housing costs is shown in figure 9.

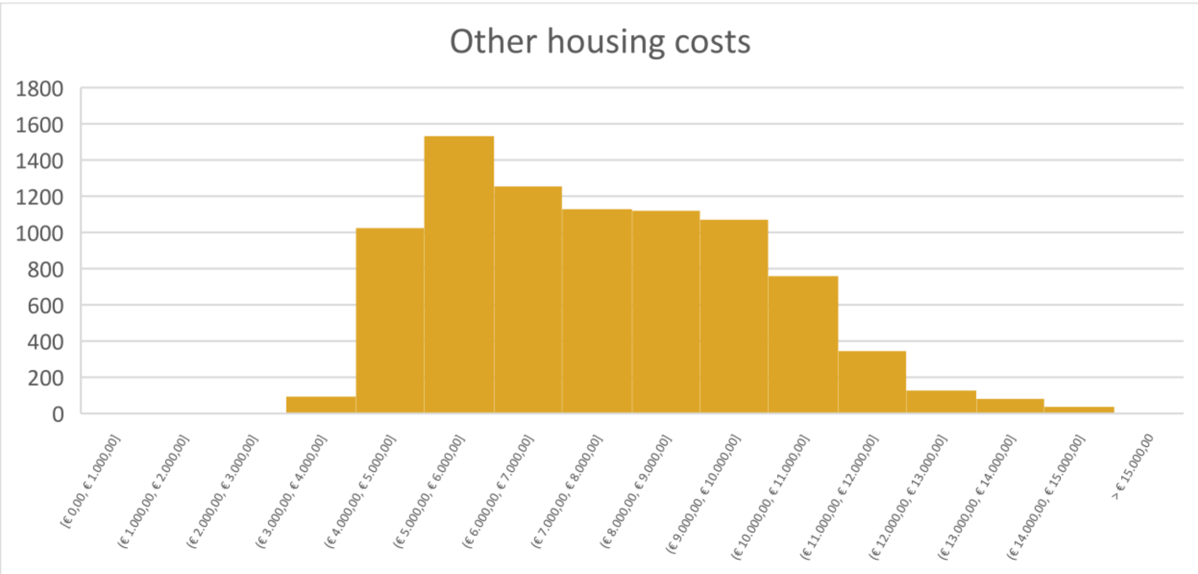

Figure 9. Histogram annual other housing costs of the respondents.

The results in figure 9 show that the distribution of the annual other housing costs is positively skewed. Most respondents have annual other housing costs between €5,000 and €6,000 and the average annual other housing costs of all respondents is €7,559.

4.2 Bivariate analysis

In this section the bivariate analyses that were performed will be described. First, the sample representativeness of the WoON 2021 dataset will be tested for multiple variables. These variables include both socio-demographic and dwelling characteristics. The results of the sample representativeness bivariate analyses will show how representative the WoON 2021 sample is for all social housing dwellings and the inhabitants of these dwellings. After the sample representativeness is tested, multiple other bivariate analyses will be conducted to determine the potential relations between exogeneous variables in the created model.

Sample representativeness

Multiple chi-square tests were conducted to determine if the selected WoON 2021 data is representative for all social housing dwellings and their inhabitant. In the chi-square tests, the observed percentages of the value classes of the WoON 2021 data were compared with the value categories of the social housing sector. If the social housing sector information was unavailable, the WoON 2021 data was compared to the distributions of these value classes of the entire population of the Netherlands. Some of the additional data that was gathered had to be recoded to match the value categories of the variables in the WoON 2021 data. The KNIME program was used to do this where needed. The results of the chi-square tests will be interpreted by looking at the p-values that result from the analyses. If the p-value of a chi-square test is less than 0.05, this indicates that there is a significant difference between the sample data and the larger datasets that were gathered.

The first variables included in the model that were tested for their representativeness were the education level and age of the residents. Since the data of these characteristics was not available in the CBS StatLine database for social housing only, the distributions of respondents were compared to data of all persons and dwellings in the Netherlands. The results of the chi-square tests of the education level and age of respondents both showed p-value below 0.05. These results, shown in table 11 and 12 respectively, mean that there is a significant difference between the sample and the total population. The sample cannot be considered as representative for the entire population of the Netherlands regarding education level and age. However, this is expected since it is known that there are differences between the people living in social housing and the total population of the Netherlands.

Table 11: *Sample representativeness education level.*

Education level	WoON 2021	CBS	Chi-square test p-value
Primary education	18.4	8.9	0.000
VMBO, MBO1	26.8	11.3	
HAVO, VWO, MBO	35.2	45.1	
HBO/WO-bachelor	13.6	21.8	
HBO/WO-master	6.0	12.9	

Table 12: *Sample representativeness age.*

Age	WoON 2021	CBS	Chi-square test p-value
34 or younger	15.2	27.9	0.002
35-44	12.1	14.5	
45-54	13.7	16.8	
55-64	18.5	16.7	
65-74	21.2	13.7	
75 or older	19.3	10.4	

Similarly to the previously mentioned variables, the data regarding the degree of urbanisation of the social housing dwellings in the Netherlands was not available in the database of CBS StatLine. However, contrary to the previously mentioned variables, the degree of urbanisation of social housing dwellings could be determined based on two datasets present in the CBS StatLine database. KNIME was used to combine the information of the degree of urbanisation of each municipality in the Netherlands with the amount of social housing dwellings in each municipality in the Netherlands. Using this data, the chi-square test of the degree of urbanisation resulted in a p-value of 0.636. This result means that the WoON 2021 sample is representative for the degree of urbanisation of all social housing dwellings in the Netherlands. The result of the chi-square test for the representativeness of the degree of urbanisation in the WoON 2021 dataset is shown in table 13.

Table 13: *Sample representativeness degree of urbanisation.*

Degree of urbanisation	WoON 2021	CBS	Chi-square test p-value
No	7.0	4.7	0.636
Low	13.0	16.0	
Average	16.6	13.4	
High	30.8	31.8	
Very high	32.6	34.1	

Since the data of the household types of households living in social housing dwellings was not available in the CBS StatLine database, this information was gathered from the Vereniging van Nederlandse Gemeenten (VNG). Testing the sample representativeness of the household types using a chi-square test, the results shown in table 14 show a p-value of 0.545. This result means that the WoON 2021 sample is representative for the entire social housing population of the Netherlands regarding household type.

Table 14: *Sample representativeness household type.*

Household type	WoON 2021	VNG	Chi-square test p-value
Single-person	55.3	51.8	0.545
Single-parent with children	10.3	13.8	
Couple without children	23.0	20.8	
Couple with children	10.2	13.6	

For testing the sample representativeness of the dwelling characteristics in the WoON 2021 data sample, Republiq provided a dataset that included the dwelling characteristics of all social housing dwellings in the Netherlands. This provided dataset combines both information from the Eigendomsdata Kadaster and the Basisadministratie gebouwen (BAG). Since this dataset was complete and very elaborate, the dataset provided ideal data to test the sample representativeness of the dwelling size, construction year and dwelling types included in the WoON 2021 data. The results of the chi-square tests of these variables are shown in tables 15, 16, and 17, respectively.

Table 15: *Sample representativeness dwelling size.*

Dwelling size	WoON 2021	Republiq	Chi-square test p-value
Less than 50m2	6.5	5.6	0.942
50-74m2	26.9	29.0	
75-99m2	40.1	40.4	
100m2 or more	26.5	25.2	

Table 16: *Sample representativeness dwelling construction year.*

Dwelling construction year	WoON 2021	Republiq	Chi-square test p-value
1945 or older	8.2	9.3	0.914
1946-1964	17.7	20.0	
1965-1974	21.5	21.1	
1975-1991	28.4	28.8	
1992-2005	11.5	11.3	
2006 or newer	12.7	9.5	

Table 17: *Sample representativeness dwelling type.*

Dwelling type	WoON 2021	Republiq	Chi-square test p-value
Apartment	53.8	52.5	0.968
Terraced	30.3	30.9	
Semi-detached	15.9	16.5	

The results of the chi-square tests of the social housing dwelling characteristics all have a value above 0.05, meaning that the WoON 2021 sample is representative for the analysed dwelling characteristics. Especially the division of dwelling types among the value categories was extremely representative for all dwellings with a p-value of almost 1.

The final variable in the WoON 2021 data sample that could be tested for representativeness for all social housing dwellings wat the presence of solar panels. The data including all social housing dwellings in the Netherlands used for the comparison was gathered from Aedes. The results of the chi-square test shown in table 18 show a p-value relatively close to 1. This result means that there is a no significant difference between the sample data and the larger dataset. Because of this, the WoON 2021 dataset can be considered to be representative for all social housing dwellings regarding the presence of solar panels.

Table 18: *Sample representativeness solar panels.*

Solar panels	WoON 2021	Aedes	Chi-square test p-value
Yes	14.8	16.1	0.998
No	85.2	83.9	

Correlations between exogeneous independent variables

After the representativeness of the data sample was tested, the relations between exogeneous independent variables in the model had to be tested for multicollinearity. Multicollinearity occurs when the independent variables are strongly correlated. If there is a high correlation between independent variables, the effect of one of these variables on energy poverty may be explained through the other variable and not by its direct effect on energy poverty. Multicollinearity between variables is not desired since this may decrease the statistical reliability of the results. If the results of the bivariate analysis show that there is multicollinearity between independent variables, the choice will be made to remove one of these variables from the model to improve its reliability. The bivariate analysis technique that is needed to analyse the relations between independent variables is based on the measurement scales of the variables. An overview of the independent variables, their measurement scales, and the needed bivariate analysis methods based on the measurement scales is shown in table 19.

Table 19: *Measurement scales of the exogenous independent variables and the needed bivariate analysis techniques*

<table>
<tr><td colspan="2" rowspan="7"></td><td rowspan="6"></td><td colspan="4">Independent variable B</td></tr>
<tr><td rowspan="5">Solar panels</td><td rowspan="3">Household type</td><td>Education level</td><td rowspan="5">Other housing costs</td></tr>
<tr><td>Age</td></tr>
<tr><td>Degree of urbanisation</td></tr>
<tr><td rowspan="2">Dwelling type</td><td>Dwelling size</td></tr>
<tr><td>Dwelling construction year</td></tr>
<tr><td>Measurement scale</td><td>Nominal (dichotomous)</td><td>Nominal</td><td>Ordinal</td><td>Ratio</td></tr>
<tr><td rowspan="9">Independent variable A</td><td>Solar panels</td><td>Nominal (dichotomous)</td><td colspan="2" rowspan="3">Chi-square test</td><td>Mann-Whitney U-test</td><td>Independent samples t-test</td></tr>
<tr><td>Household type</td><td rowspan="2">Nominal</td><td rowspan="2">Kruskal-Wallis H-test</td><td rowspan="2">Analysis of variance (ANOVA)</td></tr>
<tr><td>Dwelling type</td></tr>
<tr><td>Education level</td><td rowspan="5">Ordinal</td><td rowspan="5">Mann-Whitney U-test</td><td rowspan="5">Kruskal-Wallis H-test</td><td rowspan="5">Spearman correlation</td><td rowspan="5">Spearman correlation</td></tr>
<tr><td>Age</td></tr>
<tr><td>Degree of urbanisation</td></tr>
<tr><td>Dwelling size</td></tr>
<tr><td>Dwelling construction year</td></tr>
<tr><td>Other housing costs</td><td>Ratio</td><td>Independent samples t-test</td><td>Analysis of variance (ANOVA)</td><td>Spearman correlation</td><td>Pearson correlation</td></tr>
</table>

The variables in the model are measured on a dichotomous nominal scale, nominal scale, ordinal scale, or on a ratio scale. Because of this, the bivariate analysis techniques that are needed are the chi-square test, Mann-Whitney U-test, Kruskal-Wallis H-test, independent samples t-test, analysis of variance (ANOVA), and the Spearman correlation test. For each of these tests, the correlation coefficients or effect sizes between variables will be calculated. Both the se correlation coefficient and effect sizes can then be interpreted using Cohen's standard. According to Cohen's standard, coefficients less than 0.10 represent negligible correlations (Schober, Boer & Schwarte, 2018). Coefficient between 0.10 and 0.29 represent weak correlations, coefficients between 0.30 and 0.49 represent moderate correlations and coefficients between 0.50 and 0.9 represent strong correlations (Statistics Solutions, n.d.). Finally, coefficients stronger than 0.9 represent very strong correlations according to Cohen's standard (Schober et al, 2018). The presence of multicollinearity is often assumed with coefficients larger than 0.7 however the reliability of a model is better when strong correlations are avoided (M., 2019).

Nominal-nominal

The bivariate analyses between dichotomous nominal and nominal exogeneous independent were the first analysed performed to test the relations between the exogeneous independent variables. Because of the measurement scales of these variables, chi-square tests were needed to test for the independence of the variables. For all of the chi-square tests results, the Pearson chi-square could be interpreted. The first chi-square analysis performed analyses the relations between the only dichotomous nominal variable, solar panels, and the other nominal variables. The results of this analysis are shown in appendix 4. The results show relatively large Pearson chi-square values that are all significant at the 5% level, indicating statistically significant relations between the variables. In order to determine if these statistically significant relations are strong, moderate, weak, or negligible, the symmetric measure Cramer's V had to be calculated since one of the variables in each chi-square test has more than two value classes. The symmetric measures for the dichotomous nominal and nominal variables show statistically significant results for Cramer's V. These results show that the presence of solar panels has a negligible relation with the household type and a weak relation with the dwelling type.

The second chi-square test that was performed analyses the relation between the exogeneous independent nominal variables of the model, household type and dwelling type. The results of this chi-quare test are show in appendix 5. Again, the Pearson chi-square value is relatively large and statistically significant, indicating a significant relation between the variables. Cramer's V was calculated to determine the effect size of the relation between household type and dwelling type. The results of this calculation show a statistically significant value of 0.203 for Cramer's V. Based on this result, it can be determined that there is a weak relation between household type and dwelling type.

Nominal (dichotomous) – ordinal

After the relations between the nominal variables were analyses, the relations between the dichotomous variable solar panels and all ordinal variables were tested. Based on the measurement scales of these variables, a Mann-Whitney U-test should be used for the bivariate analysis. The results of these Mann-Whitney U-tests are shown in appendix 6. The results of the Mann-Whitney U-tests all show large Mann-Whitney U-values that are all statistically significant except for the variable representing the education level. Because of this, it can be determined that there is no statistically significant relation between the presence of solar panels and the education level of the inhabitants. All other ordinal variables do show a significant relation with the presence of solar panels. Because of this, it should be determined how large these effects are so multicollinearity can be avoided. The relation between the presence of solar panels and the ordinal variables can be calculated using formula 7.

- $r^2 = \frac{Z^2}{n}$ (7)

In this formula *Z* represents the standardised value for the U-value, *n* represents the total numbers of observations on which *Z* is based, and r^2 represents the effect size index (Tomczak & Tomczak, 2014). The results of the r^2 calculations shown in appendix 6 all have a value of less than 0.1. Because of this it can be determined that the relations between the presence of solar panels and the ordinal variables included in this research are negligible. The education level forms an exception to this, since no statistically significant relations was found between the presence of solar panels and the education level.

Nominal (dichotomous) – ratio

The relation between the dichotomous nominal variable solar panels and the ratio variable other housing costs in the model can be analysed by using an independent samples t-test. This independent samples t-test was conducted, and the results of this test are shown in appendix 7. The results of this independent samples t-tests shows a significant value for Levene's tests for equality of variances. Because of this result, equal variances should not be assumed and the significance of the t value for this category should be interpreted. The t value when equal variances are not assumed is statistically significant, indicating significant difference between groups and a significant relation between the presence of solar panels and the other housing costs. Because of this, the effect sizes of the presence of solar panels on the other housing costs should be calculated to check for multicollinearity. Hedges' g coefficient is calculated to determine the effect size of the independent samples t-test. Hedges' g coefficient is calculated using formula 8.

- $g = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{(n_1-1){s_1}^2 + (n_2-1){s_2}^2}{n_1+n_2-2}}}$ (8)

In this formula x_1 and x_2 represent the means of the first and second sample. n_1 and n_2 represent the number of observations in groups (group 1, group 2) and s_1 and s_2 represent the standard deviation in groups (group 1, group 2) (Tomczak & Tomczak, 2014). The result of the calculation of Hedges' g is shown in appendix 7. Based on the result of this calculations, it can be determined that the presence of solar panels has a weak relation with the other housing costs.

Nominal – ordinal

Kruskal-Wallis H-tests were conducted in order to analyse the potential relations between the nominal variables household type and dwelling type and the ordinal variables included in the model. First, the Kruskal-Wallis H-tests for the household type and the ordinal variables were conducted. The results of these Kruskal-Walls H-tests are shown in appendix 8. The Kruskal-Wallis H-tests shown in appendix 8 all yield significant results, indicating statistically significant relations between household type and the ordinal variables. Because of these statistically significant results, it is important to test the strengths of the relations between the household type and the ordinal variables. In order to test this, the epsilon square coefficient can be calculated and interpreted. The epsilon squared coefficient can be calculated using formula 9.

- $E_R^2 = \frac{H}{(n^2-1)/(n+1)}$ (9)

In this formula E_R^2 represents a coefficient with a value between 0 (no relationship) and 1 (perfect relationship). Additionally, *H* represents the Kruskal-Wallis H-test statistic obtained in all Kruskal-

Wallis H-tests and *n* represents the total number of observations (Tomczak & Tomczak, 2014). The calculated epsilon squared coefficients shown in appendix 8 show relatively low values. Based on these coefficients it can be determined that the household type has a weak relation with age and the dwelling size. Furthermore, the results of the epsilon squared tests show that household type has a negligible relation with the education level, degree of urbanisation, and dwelling construction year.

After all Kruskal-Wallis H-tests for the household type and the ordinal variables were performed, the same had to be done for the nominal variable representing the dwelling types. The results of these Kruskal-Wallis H-tests are shown in appendix 9. The results of these Kruskal-Wallis H-tests are all significant, indicating statistically significant relations between the dwelling type and the ordinal variables. Again, the epsilon square coefficients were calculated to determine the strengths of the relations between the variables. The calculated epsilon square coefficients shown in appendix 9 show various strengths of relations between the dwelling type and the ordinal variables. By interpreting the epsilon square coefficients, it can be determined that the relations between the dwelling type and the education level, age and dwelling construction year are negligible. Additionally, the relation between dwelling type and degree of urbanisation is a weak relation while the relation between dwelling type and dwelling size is a moderate relation.

Nominal – ratio

The relations between the nominal variables and the ratio variable in the model are analysed using the bivariate technique ANalysis Of VAriance (ANOVA). The first ANOVA that was performed analysed the relations between the nominal variable household type and the other housing costs. The result of this ANOVA is shown in appendix 10. The results of the ANOVA show a significant value for the Levene statistic, indicating the presence of statistically significant relation between the variables. The results furthermore show a significant result of the F statistic and therefore significant differences between groups. Because of these results, it is important to determine the strength of the relation between the household type and the other housing costs. Formula 10 is used to calculate the omega squared value that can be interpreted as the effect size of an ANOVA.

- $\omega^2 = \frac{df_{ef}(MS_{ef} - MS_{er})}{SS_t + MS_{er}}$ (10)

In formula 11, df_{ef} represents the degrees of freedom of the effect between groups and SS_t represents the total sum of squares. Additionally, MS_{ef} represents the mean square of the effect, while MS_{er} represents the mean square error (Tomczak & Tomczak, 2014). The results of the omega squared calculations for the relations between the household type and the other housing costs are shown in appendix 10. From these omega squared calculations it can be determined that the relation between household type and other housing costs is a weak relation.

A second ANOVA had to be performed to analyse the relation between the dwelling type and the other housing costs. The result of this ANOVA is shown in appendix 11 and does not show a significant value for the Levene statistic. This result indicates that there is no significant relation between the dwelling type and other housing costs.

Ordinal – ordinal

The Spearman correlation was used to determine the relations between the ordinal variables in the model. The results of the Spearman correlation analyses performed in SPSS are shown in appendix 12. The results of the Spearman analyses between the ordinal variables all show a statistically significant result except for the relation between the dwelling construction year and the education level of inhabitants. All other ordinal variables have a statistically significant relation with various strengths.

The education level has a moderate relation with age, a weak relation with the degree of urbanisation, and a negligible relation with the dwelling size. Age has a weak relation with the degree of urbanisation and dwelling size and a negligible relation with dwelling construction year. Furthermore, the degree of urbanisation has weak relations with both the dwelling size and construction year. Finally, the relation between the size of a dwelling and the construction year of a dwelling is negligible.

Ordinal – ratio

The final bivariate analysis technique that was needed to analyse the relations between the exogenous independent variables is the Spearman correlation. The Spearman correlation used to determine the correlations between the ordinal variables could also be used to determine the relations between the ordinal and the ratio variable in the model. The results of the Spearman correlation analyses between the ordinal and ratio variables in SPSS are shown in appendix 13. The results show significant relations between the education level, degree of urbanisation, dwelling size and construction year and the other housing costs. The other housing costs do not have a statistically significant relation with age. When the results of the spearman correlation coefficients are interpreted, it can be determined that the education level and dwelling construction year have a weak relation with the other housing costs. Additionally, the relation between the other housing costs and the dwelling size is a moderate relation while the relation between the degree of urbanisation and the other housing costs is negligible.

The results of the bivariate analyses show that while most exogeneous independent variables have a statistically significant relation, most of these relations are weak or negligible. A relation of moderate strength was found between education level and age. Furthermore, the dwelling size has relations of moderate strength with the dwelling type and other housing costs. No relations stronger than moderate were found between the exogeneous independent variables. Because of this, no relations stronger than 0.7 were found and therefore multicollinearity between these variables can be ruled out. The absence of multicollinearity ensures that the reliability of the results of further analyses is not reduced.

Correlations between exogeneous independent and endogenous dependent variables

After the relations between exogeneous independent variables in the model were tested for multicollinearity, the same had to be done for some relations between exogeneous independent and endogenous dependent variables. Most exogeneous variables in the model are directly or indirectly related to the endogenous variables however two variables, age, and other housing costs, are not. For these variables, multicollinearity with the endogenous variables is not desired since this may reduce the reliability of the result. If multicollinearity is found between these variables, a choice will be made to remove one of the variables from the model to increase its reliability. Again, table 19 was used to determine the needed bivariate analysis techniques based on the measurement scales of the variables. The Spearman correlation was used to analyse the relations between age and the endogenous variables electricity usage, gas usage, income, and energy expenditure. The results of these Spearman correlations are shown in appendix 14. The results show no significant relation between age and the electricity usage. Furthermore, age has a weak relation with income and a negligible relation with gas usage and energy expenses. Because of the measurement scales of the variables, the Pearson correlation was used to analyse the correlations between the other housing costs and the endogenous variables. The results of these Pearson correlations are shown in appendix 15. All relations between the other housing costs and the endogenous variables are statistically significant. The results show that the other housing costs have a weak relation with energy expenses and the gas usage, a moderate relation with the electricity usage, and a strong relation with the income. From the results of the bivariate analyses it can be concluded that there was no multicollinearity between the exogenous and endogenous variables since there was no correlation coefficient of 0.7 or higher. However, because of the high correlation between income and the other housing costs, the choice was made to remove the other housing costs from further analyses to increase the reliability of the results.

4.3 Exploratory factor analysis

In this section the EFA will be described that was used to analyse if the existing energy poverty indicators represent and measure the same latent construct. If the existing energy poverty indicators represent the latent construct energy poverty, they will all have significant factor loading with this construct. The variables included in the EFA are the EQ, LIHC, MIS, HCOR, and LILEQ energy poverty indicators that are all measured on a continuous scale. Since the information of the existing energy poverty indicators was available for all respondents remaining in the used WoON dataset the sample size is sufficient with 8,587 respondents. Figure 10 shows the EFA model of the existing energy poverty indicators.

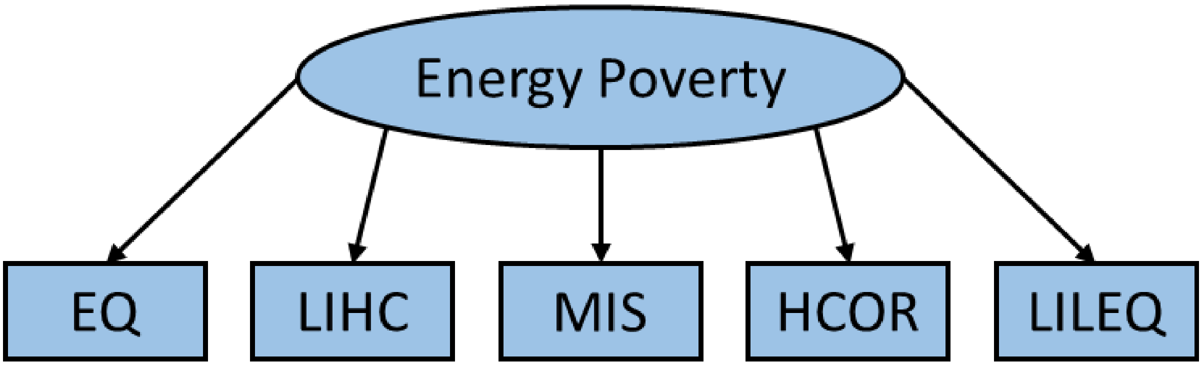

Figure 10. EFA model energy poverty indicators

After the variables that included in the EFA were identified, the next step of the FA process is to determine if the variables in the model are reliable measurements. The KMO measure of sampling adequacy and Barlett's test of sphericity were determined to analyse the reliability of the variables included in the EFA. The results of the KMO and Bartlett's tests are shown in appendix 16. The result of the KMO measure of sampling adequacy is larger than 0.5, indicating that the data sample is adequate for an EFA. The result of Bartlett's test of sphericity shows a high chi-square value that is statistically significant. This result means that there exist significant relations among the variables and the variables relate to each other enough to perform an EFA. The strengths of the relations between the existing indicators could be analysed using the correlation matrix shown in appendix 17. This correlation matrix shows that most correlations between the energy poverty indicators are statistically significant except for the relations between MIS and the EQ, LIHC, and LILEQ indicators. All correlations are less than 0.8 except for the relation between the EQ and HCOR indicators. Because of this result, it is crucial that the determinant of the correlation matrix is checked to avoid multicollinearity. Because the determinant of the correlation matrix is greater than 0.00001, it can be concluded that there is no multicollinearity between the observed variables.

Since it was determined that the observed variables in the EFA model are reliable measurements, the FA could be conducted. After the FA was conducted, the initial and extraction values of the communalities were estimated. The communalities shown in appendix 18 describe the amount of variance that a variable shared with the other variables considered in the FA. The communalities show high extraction values for the communalities of the EQ, LIHC, HCOR and LILEQ indicators however the MIS indicators has a communality extraction factor less than 0.4. This result indicates that the MIS indicators does not contribute much to the measurement of the underlying latent factor. In the fifth step of the EFA it is usually decided how many factors will be retained. However, since the goal of this EFA is to determine if all existing energy poverty indicators measure one common underlying latent factor, only one factor will be retained. When looking at the total variance explained in appendix 19, it can be seen that two factors would have been retained if the Kaiser-Guttman rule was used since two factors have an eigenvalue larger than 1. Since only one component is retained from the FA, no rotation is applied to the FA since this is only possible when there are two or more components. According to appendix 19, the component that is retained from the FA explains just over 50% of the total variance.

In the final step of the EFA, the results are interpreted by looking at the component matrix. In the component matrix shown in appendix 20 the factor loadings are shown. These factor loadings are Pearson correlations that represent how strong the observed indicators represent the common underlying latent factor. Since factor loadings greater than 0.65 are considered to be strong associations, it can be seen that most observed indicators have a strong association with the underlying latent concept. This confirms the findings of Mulder et al (2023) who found that the EQ, LIHC, and LILEQ energy poverty indicators were relatively strongly correlated. Most of the observed indicators seem to load as one construct and are therefore considered to be good component factor scores. The MIS is an exception to this, as the component matrix and factor loadings and all other EFA tests show that this observed variable does not measure the same common underlying latent concept. Because of this, the MIS indicator will be omitted from further analyses with the existing energy poverty indicators.

Because it was found that the MIS indicator did not measure a common underlying concept with the other energy poverty indicators, a second EFA was conducted without the MIS indicator. This second EFA resulted in the same values for almost all the steps and results of the first EFA except for the total variance explained. The total variance explained in the second EFA shown in appendix 21 shows that when the MIS indicator is removed from the EFA, the total variance explained of the first component increases from 50.180 to 61.212%.

4.4 Structural equation modeling (SEM)

After the EFA had been performed, the SEM model was created in SmartPLS. First, the processed data of the respondents remaining in the dataset was loaded into the SmartPLS programme. After the correct measurement scales of the variables were selected in SmartPLS, the loading of the data was completed. Second, the SEM model was drawn in SmartPLS by drawing rectangles for the observed variables and a circle for the latent variable. In the final step of the creation of the SEM model in SmartPLS, the relations between the variables were drawn in the SEM model. These relations are all based on the expected relations between the variables that were discovered in the literature review. The created SEM model in SmartPLS is shown in figure 11.

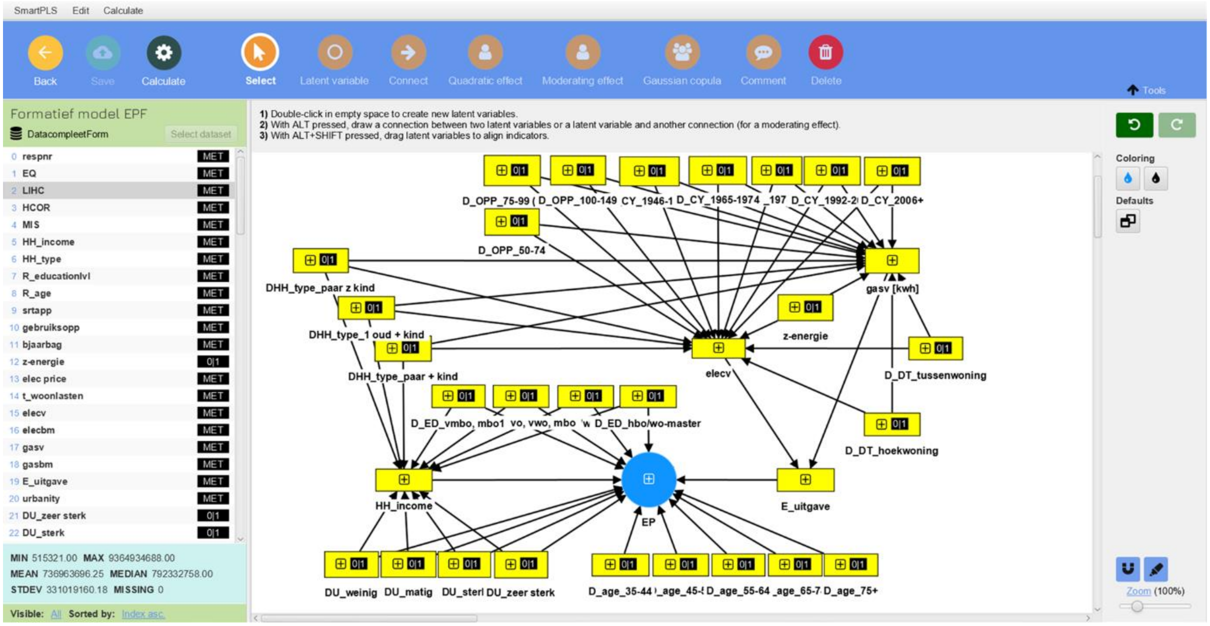

Figure 11. SEM model in SmartPLS

After the SEM model was created and the variables of the data were coupled to the model in SmartPLS, the path coefficients of the SEM model could be estimated. Based on more than 8,500 subsamples, the SEM model could estimate the standardised path coefficients between the variables included in the model. The results of the estimations of the path coefficients and P values are shown in table 20.

Table 20: *Path coefficient estimations and P values*

Variable 1		Variable 2	Path coefficient	P value
Household type	Single-person	Electricity usage	REF	.
	Couple without children	Electricity usage	0.587	0.000
	Single-parent	Electricity usage	0.599	0.000
	Couple with children	Electricity usage	1.228	0.000
Dwelling size	Less than 50	Electricity usage	REF	.
	50-74	Electricity usage	0.172	0.000
	77-99	Electricity usage	0.315	0.000
	100 or larger	Electricity usage	0.512	0.000
Construction year	1945 or older	Electricity usage	REF	.
	1946-1964	Electricity usage	0.018	0.658
	1965-1974	Electricity usage	0.092	0.022
	1975-1991	Electricity usage	0.096	0.009
	1992-2005	Electricity usage	0.103	0.012
	2006 or later	Electricity usage	0.192	0.000
Dwelling type	Apartment	Electricity usage	REF	.
	Terraced	Electricity usage	0.117	0.000
	Semi-detached	Electricity usage	0.208	0.000
Solar panels	No	Electricity usage	REF	.
	Yes	Electricity usage	-0.325	0.000
Household type	Single-person	Gas usage	REF	.
	Couple without children	Gas usage	0.129	0.000
	Single-parent	Gas usage	0.197	0.000
	Couple with children	Gas usage	0.224	0.000
Dwelling size	Less than 50	Gas usage	REF	.

	50-74	Gas usage	0.173	0.000
	77-99	Gas usage	0.243	0.000
	100 or larger	Gas usage	0.393	0.000
Construction year	1945 or older	Gas usage	REF	.
	1946-1964	Gas usage	0.048	0.290
	1965-1974	Gas usage	-0.082	0.074
	1975-1991	Gas usage	-0.168	0.000
	1992-2005	Gas usage	-0.406	0.000
	2006 or later	Gas usage	-0.677	0.000
Dwelling type	Apartment	Gas usage	REF	.
	Terraced	Gas usage	0.240	0.000
	Semi-detached	Gas usage	0.431	0.000
Solar panels	No	Gas usage	REF	.
	Yes	Gas usage	-0.075	0.008
Electricity usage		Energy expenses	0.359	0.000
Gas usage		Energy expenses	0.379	0.000
Education level	Primary school	Income	REF	.
	Vmbo, mbo1	Income	0.100	0.000
	Havo, vwo, mbo	Income	0.257	0.000
	Hbo, wo-bachelor	Income	0.458	0.000
	Hbo, wo-master	Income	0.589	0.000
Household type	Single-person	Income	REF	.
	Couple without children	Income	0.508	0.000
	Single-parent	Income	0.365	0.000
	Couple with children	Income	0.922	0.000
Degree of urbanisation	Not urbanised	Income	REF	.
	Hardly urbanised	Income	0.031	0.423
	Moderately urbanised	Income	-0.002	0.950
	Strongly urbanised	Income	-0.036	0.305
	Extremely urbanised	Income	-0.003	0.931
Degree of urbanisation	Not urbanised	Energy poverty	REF	.
	Hardly urbanised	Energy poverty	-0.074	0.080
	Moderately urbanised	Energy poverty	-0.037	0.425
	Strongly urbanised	Energy poverty	-0.122	0.003
	Extremely urbanised	Energy poverty	-0.118	0.008
Education level	Primary school	Energy poverty	REF	.
	Vmbo, mbo1	Energy poverty	-0.050	0.030
	Havo, vwo, mbo	Energy poverty	-0.051	0.046
	Hbo, wo-bachelor	Energy poverty	-0.032	0.392
	Hbo, wo-master	Energy poverty	-0.014	0.832
Age	34 or younger	Energy poverty	REF	.
	35-44	Energy poverty	-0.183	0.000
	45-54	Energy poverty	-0.133	0.002
	55-64	Energy poverty	-0.105	0.017
	65-74	Energy poverty	-0.221	0.000
	75 or older	Energy poverty	-0.274	0.000
Energy expenses		Energy poverty	0.238	0.000
Income		Energy poverty	-0.532	0.000

The results of the estimation of the SEM model show significant results for all path coefficients of variables connected to the electricity usage except for the path coefficient of dwellings constructed between 1946 and 1964. The results show that single-parent households and couples with and without children have a higher electricity usage than single-person households. The results furthermore show that electricity usage increases with the dwelling size of a building and that newer dwellings use more electricity than older dwellings. Compared to apartments, both terraced and semi-detached dwellings have a higher electricity usage and the presence of solar panels reduces the energy usage.

All variables related to the gas usage show similar significant path coefficients to those related to the electricity usage. Single-parent households and couples with and without children have a higher gas usage than single-person households. Larger dwellings have a higher gas usage than smaller dwellings and terraced and semi-detached dwellings have a higher gas usage than apartments. Where the path coefficients connected to electricity usage showed that newer dwellings use more energy, the opposite can be seen when looking at the path coefficients connected to the gas usage. These path coefficients show that there are no statistically significant differences in the amount of gas used for dwellings constructed before 1945, between 1946 and 1964 and between 1965 and 1974. However, from the dwellings constructed from 1975 and on, newer dwellings use increasingly less gas. Solar panels reduce the gas usage of a dwelling however this effect is smaller than the reduction in electricity usage when solar panels are present.

The results of the SEM model show that both the electricity and gas usage have a statistically significant relation with a household's energy expenses. The effects of the electricity and gas usage on the household energy expenses are relatively similar. However, the effect of electricity usage (0.359) is slightly less than the effect of gas usage (0.379) energy expenses of a household.

When looking at the path coefficients and their P values of the variables connected to the income of a household, the results show that there is no statistically significant relation between household income and the degree of urbanization. However, both other variables connected to the household income do show a statistically significant result. The results show that the income of a household increases with the education level. Furthermore, single-parent households and couples with and without children have higher household incomes than single-person households.

While most of the variables connected to the latent variable energy poverty showed a statistically significant relation in the SEM model, this was not the case for all variables or variable categories. Regarding the degree of urbanisation, the results show that hardy and moderately urbanised have no statistically significant difference in energy poverty compared to areas that are not urbanised. However, the model shows that people living in strongly and extremely urbanised areas experience less energy poverty. The results of the estimations show that energy poverty does not differ for those having a hbo or wo-bachelor or master's degree compared to those who have a primary education degree. However, according to the results, people with a vmbo, mbo1, havo, vwo or mbo degree experience slightly less energy poverty. The path coefficients between all age categories and energy poverty included in the SEM model are statistically significant. All categories included in the model have less energy poverty than the reference category of people aged 34 or younger. This effect decreases when age increases up to the category of people aged between 55 and 64. For people older than 64, the amount of energy poverty decreases again, and the least amount of energy poverty was estimated for those aged 75 years or older. Both the path coefficients between energy poverty and energy expenses and income are statistically significant. Finally. the results showed that energy poverty increases when energy expenses increase and decreases when income increases.

In order to assess the validity of the model, only the VIF values could be checked. The convergent validity of the model could not be estimated because the created model only includes a formative construct. Looking at the VIF, the indicator collinearity could be determined. The results of the VIF estimates show a value of 1.0 for all variables included in the SEM model. Because these values are all less than 3.3 there are no potential issues with the indicator collinearity in the created model.

4.5 Energy poverty risk prediction model

The path coefficients that were estimated in the SEM model can be interpreted as standardised regression coefficients. Using these standardised regression coefficients and the standardised values of the continuous variables, a new model could be made that can be used to predict the energy poverty of a household based on the variables included in the model. An energy poverty prediction model that includes all statistically significant variables and relations was created in Excel. This model includes both direct and indirect effects of the included variables to determine the risk of experience energy poverty that a household has. Using multiple formulas and references, the created energy poverty model can predict the energy poverty risk based on twelve input cells. Figure 12 shows an example response of the input cells that the energy poverty prediction model uses to predict the energy poverty risk of a respondent.

Household type	Single-person
Age	34 or younger
Education level	Primary school
Income	27786,983
Degree of urbanisation	Not urbanised
Energy costs	1649,996
Electricity usage (kwh)	2023,814
Gas usage (m3)	784,681
Solar panels	No
Dwelling type	Apartment
Dwelling constrcution year	1945 or earlier
Dwelling size	Less than 50

Figure 12. Example response input cells energy poverty risk model

The prediction model is aimed at housing corporations or policy makers that, when using the prediction model, can note the exact values of the metric variables in the input cells. For all categorical variables in the model, a dropdown menu can be opened and the most appropriate category can be selected. When all cells have an input value, the model calculates an energy poverty risk index (EPRI). The EPRI can be used by housing corporations and policy makers to analyse and identify energy poverty risks within a housing stock and evaluate the effect of policies on the energy poverty risk. Using the macro function in Excel, a macro was written where all input cells were "answered" one-by-one by all respondents in the original WoON 2021. The results of the calculations of the EPRI for all respondents in the WoON 2021 dataset used in this research are shown in figure 13. This figure shows that the results of EPRI for all WoON 2021 respondents show a relatively normal distribution.

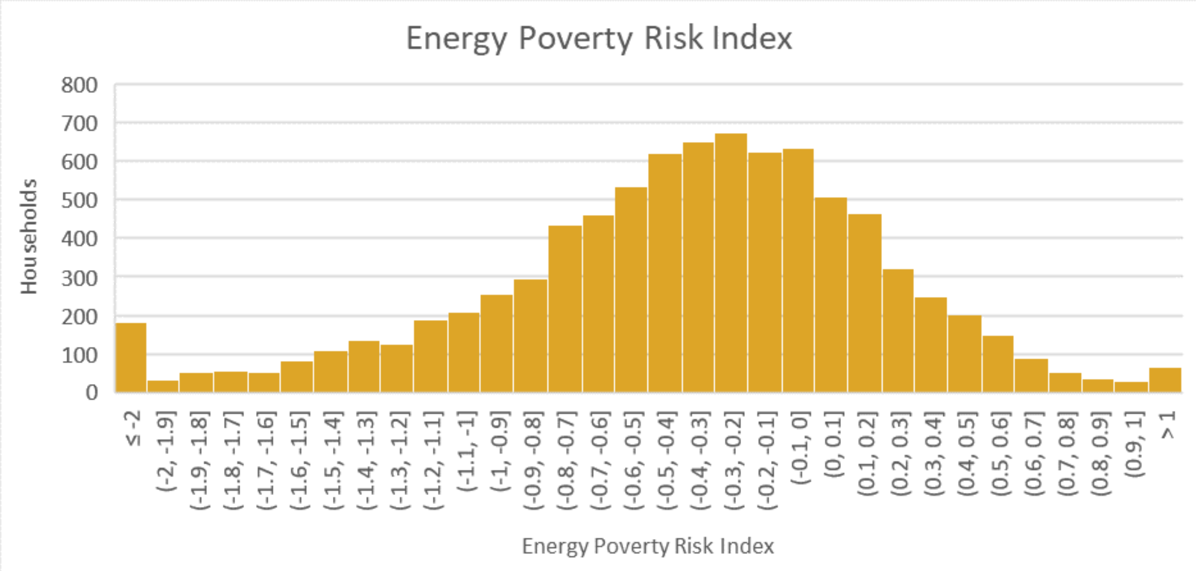

Figure 13. Results EPRI WoON 2021 respondents

In order to interpret the results of the EPRI, the results were compared to the four relevant existing energy poverty indicators described previously. The respondents of the dataset were given a number, ranging from 0 to 4, based on whether they have energy poverty based on the existing energy poverty indicators. If a respondent has energy poverty according to two of these indicators, they will score a 2. If the respondents have energy poverty according to all existing indicators, they will score a 4. When comparing how respondents score on the EPRI and the existing energy poverty indicators, a clear relation can be seen. This relation, visible in figure 14, shows that when the EPRI increases, the probability of having energy poverty according to the other energy poverty indicators increases too.

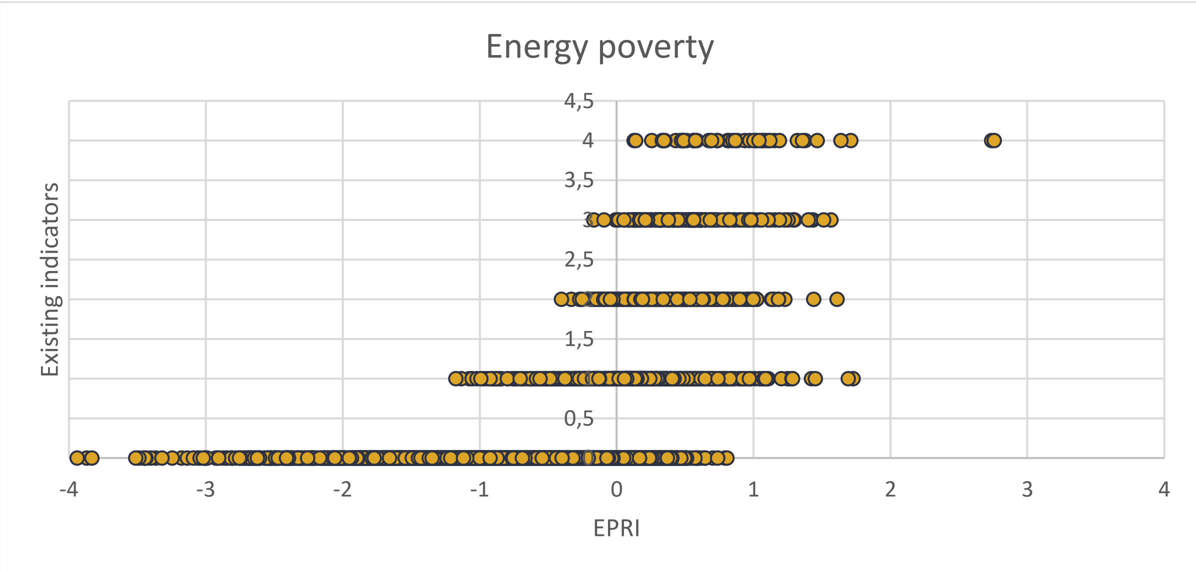

Figure 14. EPRI compared to existing energy poverty indicators.

Respondents that have energy poverty according to none of the existing energy poverty indicators score an average of -0.556 on the EPRI. Respondents that have energy poverty according to 1, 2, or 3 of the existing indicators on average score 0.100, 0.373, and 0.544 on the EPRI, respectively. Finally, respondents that have energy poverty according to all the existing energy poverty indicators score an average of 0.876 on the EPRI.

4.6 High energy poverty risk group

A descriptive analyse was conducted to determine the personal and dwelling characteristics of the respondents that have an increased risk of experiencing energy poverty. After the EPRI of all respondents in the dataset was calculated, all respondents with an EPRI over 0.373 were selected for the descriptive analysis. These respondents were selected since an EPRI value over 0.373 means that these respondents are likely to have energy poverty according to two or more of the existing indicators and can therefore be considered to have an increased energy poverty risk. The descriptive characteristics of these high EPRI respondents can be compared to the descriptive characteristics of all respondents to determine the personal and dwelling characteristics of the respondents with an increased energy poverty risk. The value category percentages of the personal characteristics of all respondents and those with a high EPRI are shown in appendix 22.

All figures in this section enable a comparison between the respondents with a high EPRI and all respondents in the WoON dataset. In these figures, both the WoON 2021 and high EPRI groups add up to 100%. By comparing the value categories of both these groups represented on the y-axis of the figures, it can be determined which value categories are overrepresented in the high EPRI group. The percentages of the value categories of the education level of the respondents are shown in figure 15.

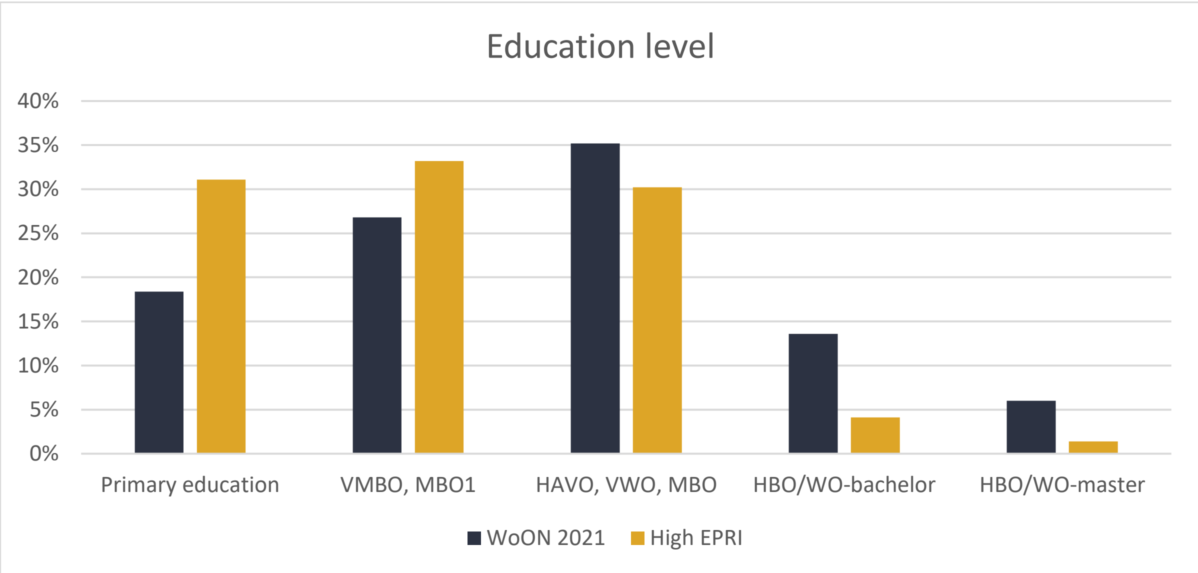

Figure 15. Education level value category percentages WoON 2021 respondents and respondents with high EPRI

When comparing the education levels of all respondents of the dataset to those with an increased EPRI, it can be seen that respondents with a primary, VMBO or MBO1 education are overrepresented in the group with a high EPRI. Additionally, the results in figure 15 show that there are relatively few HBO/WO-bachelor and master respondents that have a high EPRI.

The percentages of the value categories of the age of all respondents and those with a high EPRI are shown in figure 16.

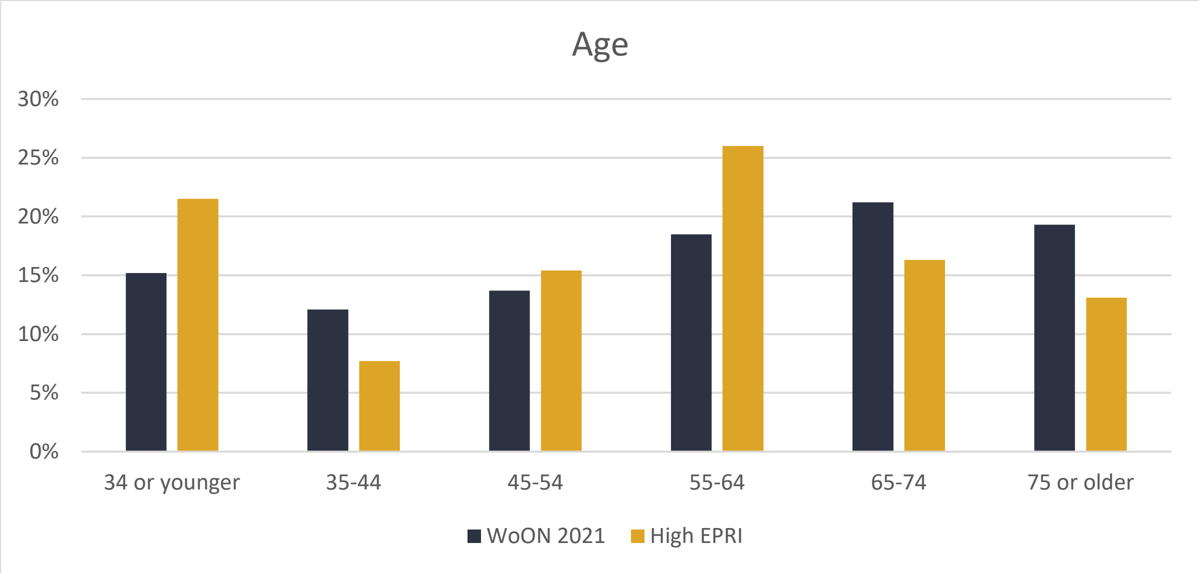

Figure 16. Age value category percentages WoON 2021 respondents and respondents with a high EPRI

When comparing these percentages of the value categories, it can be seen that respondents aged 34 or younger or between 55 and 64 are especially overrepresented in the group of respondents with a high EPRI Additionally, there are relatively few respondents with a high EPRI aged between 35 and 44 or 65 and older.

The final personal characteristic of which the value categories of respondents with a high EPRI were compared to respondents is the household type. Figure 17 shows the percentages of the value categories of the household type for all respondents and those with a high EPRI.

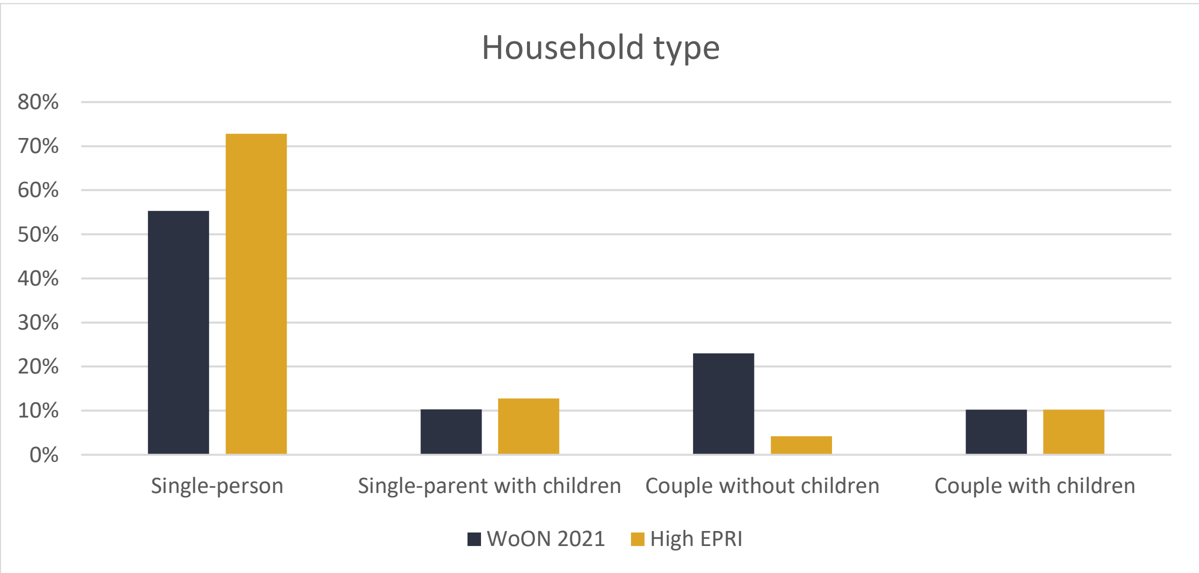

Figure 17. Household type value category percentages WoON 2021 respondents and respondents with a high EPRI

From the results in figure 17, it can be seen that single-person and single-parent households are overrepresented in the group of respondents with a high energy poverty risk factor. On the other hand, couples without children are extremely underrepresented in the high EPRI group while the percentage of couples with children is equal in both respondent groups.

After the personal characteristics of those with a high EPRI were compared to the entire WoON 2021 sample, the same could be done for the dwelling characteristics of these respondents. A descriptive analysis of the dwelling characteristics was conducted and the value category percentages of the dwelling characteristics of all respondents and those with a high EPRI are shown in appendix 23.

The percentages of value categories of the degree of urbanisation are shown in figure 18.

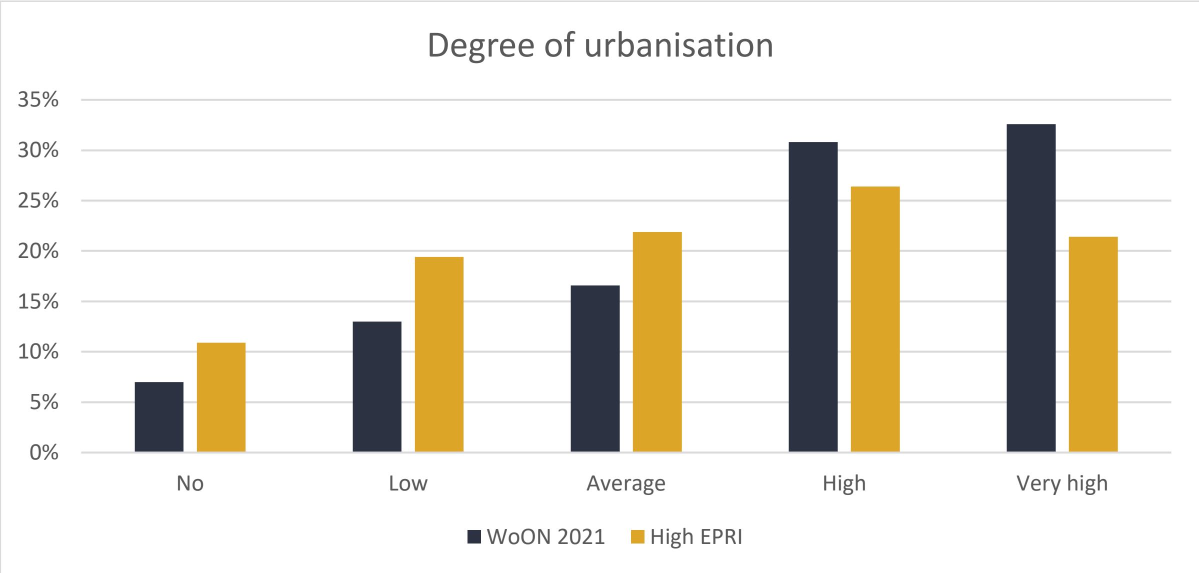

Figure 18. Degree of urbanisation value category percentages WoON 2021 respondents and respondents with high EPRI

When comparing the group with a high EPRI to the entire sample, it can be seen that all degrees of urbanisation of average and lower are overrepresented in the high energy poverty risk group. The results show that dwellings in areas with a high and especially a very high degree of urbanisation are relatively underrepresented in the high energy poverty risk group.

Figure 19 shows the value categories of the dwelling sizes of all respondents and those with a high EPRI

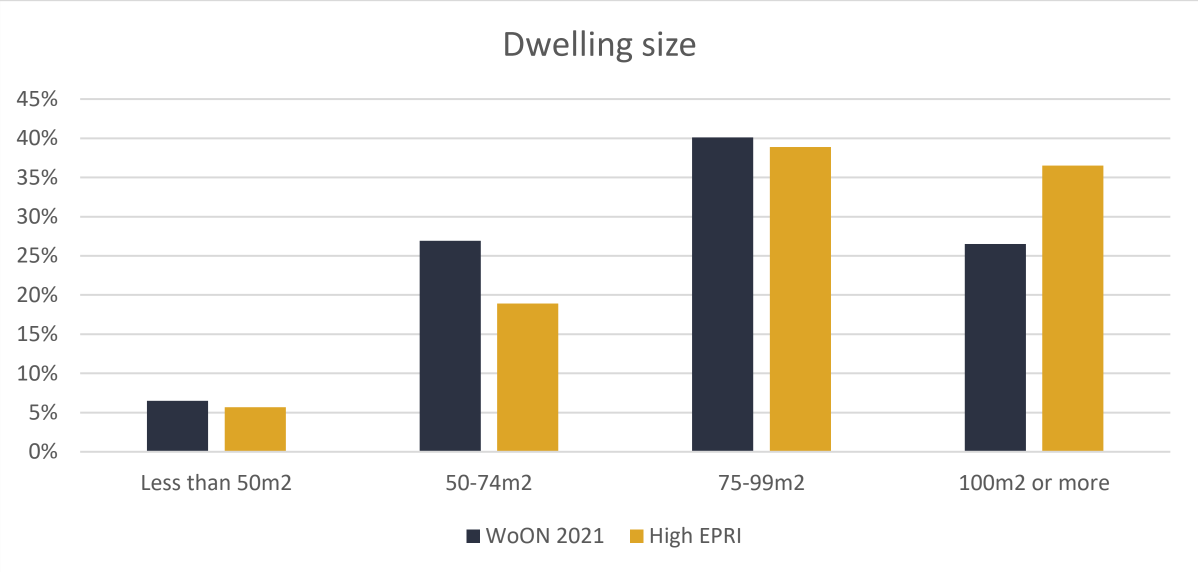

Figure 19. Dwelling size value category percentages WoON 2021 respondents and respondents with high EPRI

When comparing the respondents with a high EPRI to all respondents it can be seen that, except for the largest dwelling sizes, all dwelling size categories are underrepresented with the group with a high EPRI. Dwellings with a dwelling size of 100m2 or more are relatively overrepresented in the high EPRI respondents group.

The descriptive statistics of the dwelling construction year value categories of the respondents with a high EPRI and all respondents are shown in figure 20.

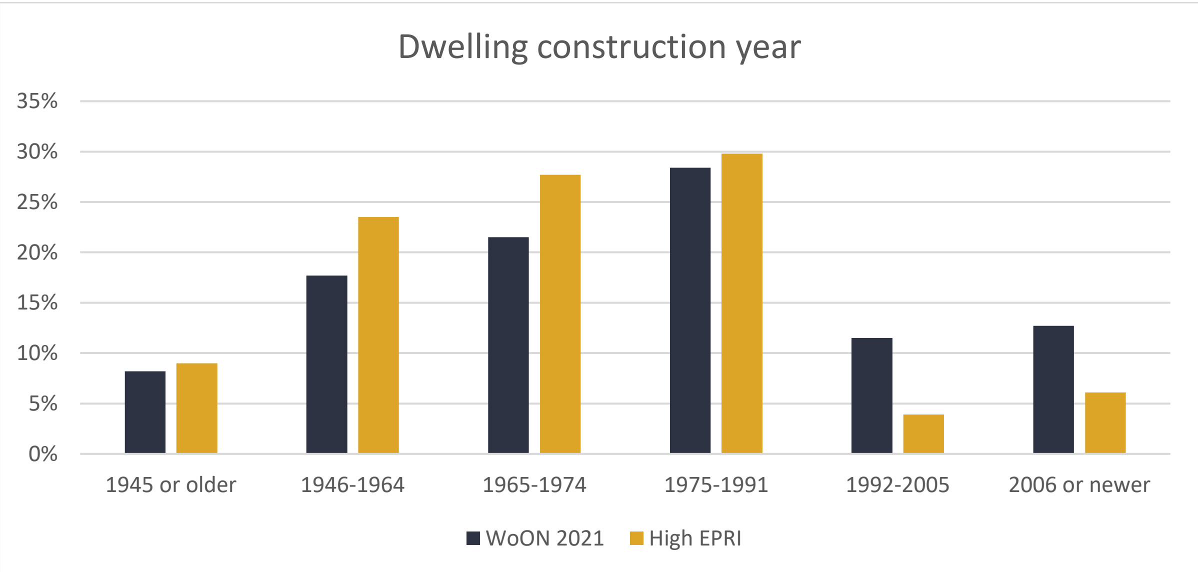

Figure 20. Dwelling construction year value category percentages WoON 2021 respondents and respondents with high EPRI

When analysing the results of the descriptive analysis, it can be seen that dwellings constructed before 1992 are overrepresented in the group with a high EPRI. Dwellings constructed between 1965 and 1974 are especially overrepresented while there are relatively few respondents with a high energy poverty risk factor that live in dwellings constructed after 1991.

Figure 21 shows the value categories of the dwelling types of all respondents and of those with a high EPRI.

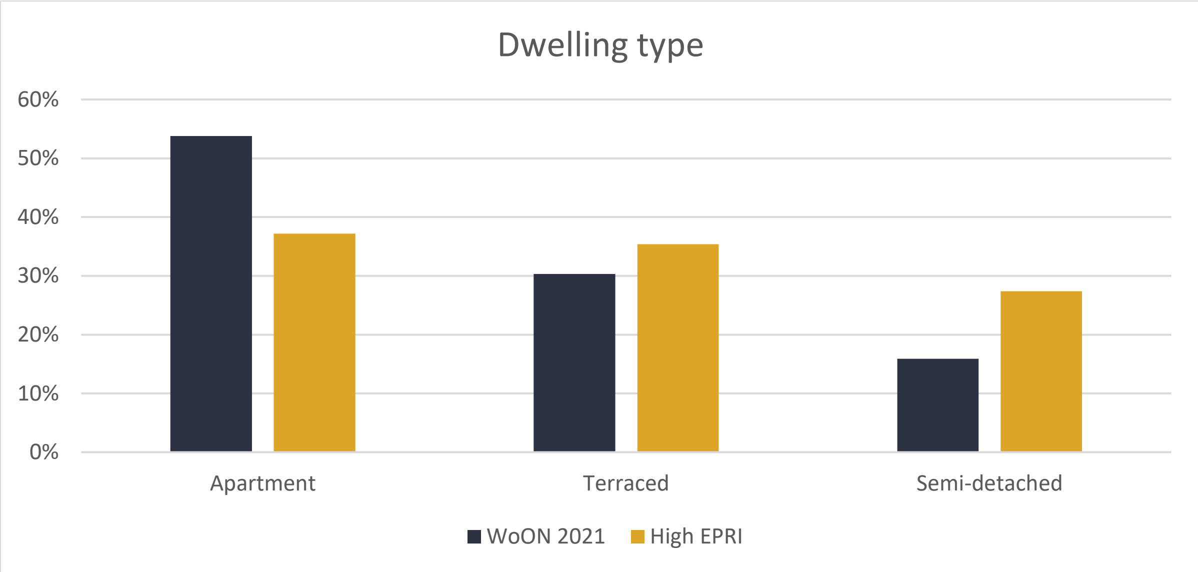

Figure 21. Dwelling type value category percentages WoON 2021 respondents and respondents with high EPRI

The results of the bivariate analysis show that relatively few respondents with a EPRI live in apartment dwellings when comparing to the entire WoON 2021 sample. On the other hand, there relatively more respondents living in terraced and especially in semi-detached dwellings in the high EPRI group.

The final dwelling characteristic that could be analysed for differences between the entire sample and the high energy poverty risk group is the presence of solar panels. The value category percentages of the presence of solar panels for the entire sample and the high energy poverty risk group is shown in figure 22.

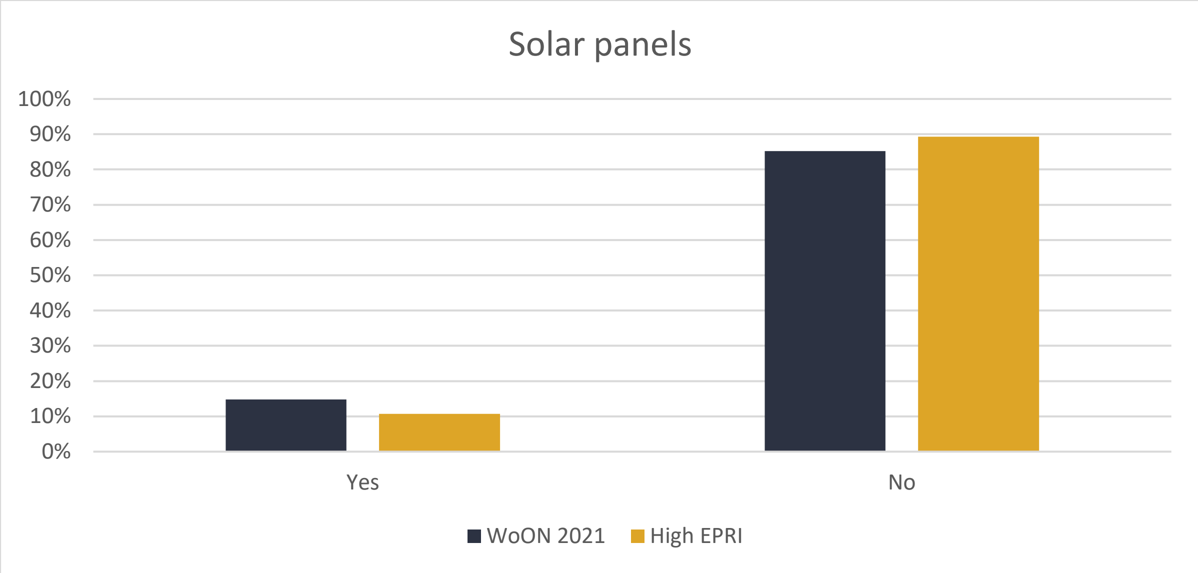

Figure 22. Solar panels category percentages WoON 2021 respondents and respondents with high EPRI

When comparing the value percentages of the presence of solar panels for all respondents and those with a high EPRI, no major differences can be seen. However, respondents with a high EPRI live in relatively less dwellings with solar panels compared to the entire WoON 2021 sample.

Concluding this section, the results of the descriptive analysis and comparison between the WoON 2021 sample and the respondents with a high EPRI give several insights. People with a relatively low education aged under 35 or between 45 and 64 are relatively overrepresented in the high EPRI group. Single-person and single-parent households are additionally relatively overrepresented when comparing to the entire WoON 2021 sample. The respondents with a high EPRI live in relatively less urbanised areas and large dwellings compared to the entire sample. Additionally, respondents living in dwellings constructed before 1992 show higher percentages in the high EPRI group while only a few respondents in this group live in dwellings constructed after 1991. Relatively few respondents with a high EPRI live in apartment dwellings while relatively many of these respondents live in semi-detached dwellings. Finally, there are relatively few high EPRI respondents that live in dwellings with solar panels compared to the entire WoON 2021 sample.

4.7 Policy effect

In this section the effect of policies developed to reduce energy poverty will be evaluated based on the developed EPRI. The effects of the policies that will be analysed, will result in changes from the already determined base scenario in the developed energy risk prediction model. By analysing these differences, it can be identified how effective the policies are in reducing the energy poverty risk. The current Dutch energy poverty policy is primarily aimed to reduce or counteract increasing energy expenditure of households. Using the EPRI, the effect of increasing energy expenditures on the energy poverty risk of households could be determined. Figure 23 shows the number of households in different energy poverty risk classes in the base scenario and the scenarios in which the energy expenses increase with 50 and 100%.

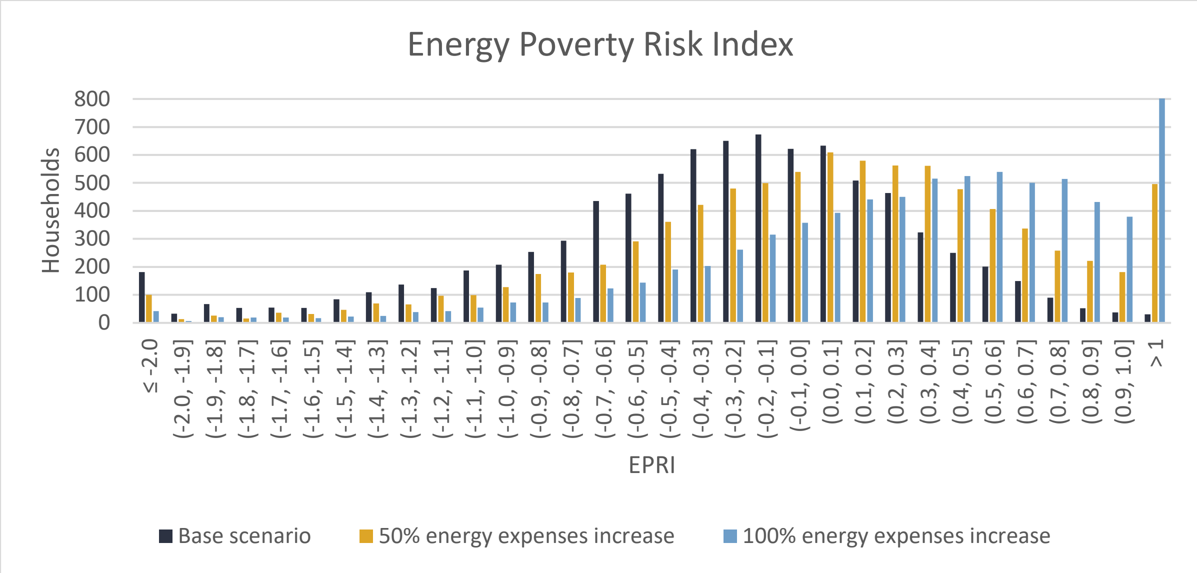

Figure 23. EPRI classes for households, base scenario and scenarios with 50 and 100% energy expenditure increases

In order to mitigate energy poverty, the Dutch government has introduced an energy allowance. This allowance is financial measure of 1,300 euro aimed to compensate for the increased energy costs and alleviate energy burdens for vulnerable households with a low income (Rijksoverheid, n.d.b). To receive this allowance, an applicant must have an income below a certain threshold. This threshold is different for various household compositions and age groups. To determine which respondents in the dataset are granted this allowance, KNIME was used to split the respondents based on their household type and age. Afterwards, KNIME was used to assess for all groups if they were above or below the set thresholds of their groups and therefore if they were granted the energy allowance. If the allowance was granted for a household, this was modelled in the data by adding 1,300 euro to the income of this household in the prediction model. Based on the incomes of the households in the WoON 2021 dataset, 1,221 households are granted an energy allowance when the current rules for this allowance are followed. This means that 14.2% of the households in the dataset will be given the 1,300-euro energy allowance. After these households were given the 1,300-euro additional income, the EPRI prediction model was calculated again. The result of the calculated scenario with the energy allowances is shown in figure 24.

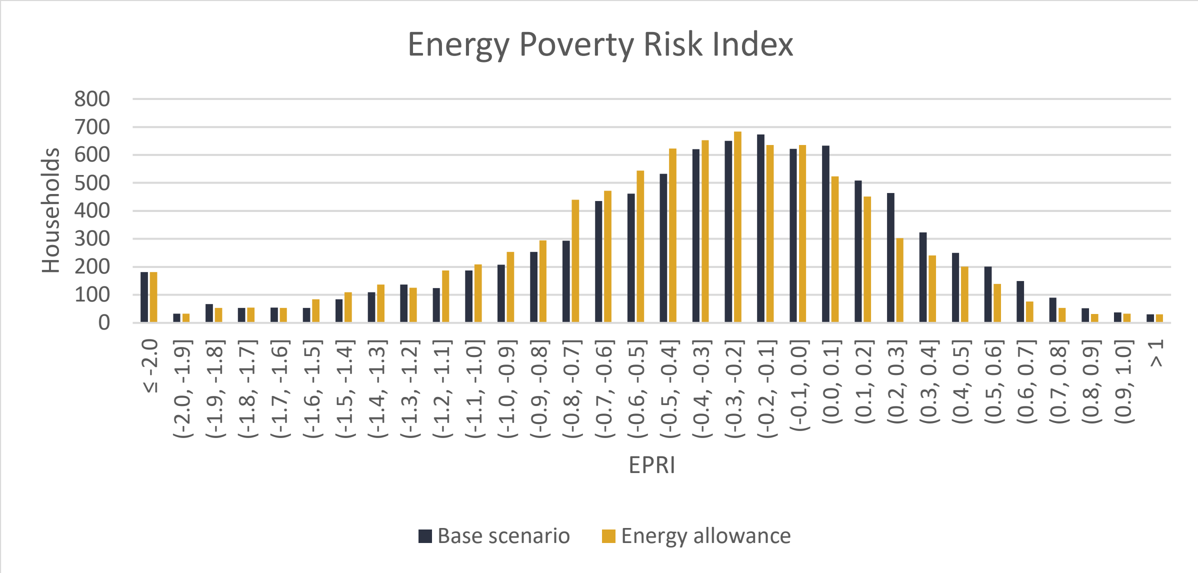

Figure 24. EPRI classes for households, base and energy allowance scenario

The results in figure 24 show that the energy allowance causes a slight reduction in the energy poverty risk. However, the effect of the energy allowance is relatively small since it is only granted to a small proportion of all households. When analysing the results in more detail, it can be seen that some households with no or a low energy poverty risk are still granted the energy allowance because they have a low income. Furthermore, the energy allowance is not granted to all households with a high energy poverty risk since some have a monthly income above 1,300 euros. The effect of the energy allowance policy does not seem optimal because the targeted vulnerable households are not always granted the allowance. A second disadvantage of this policy is that the allowance is only granted once. Because of this, the energy poverty risk is only temporarily reduced and will return when the allowance is spent.

Contrary to a financial measure, the effect of improving the energy performance of the dwelling stock on energy poverty is more aimed at the long term. Improving the energy performance of the dwelling stock was proposed in the new package of measures by the Dutch minister for climate at the beginning of May 2023. Where the energy allowance to reduce energy poverty is arranged by the government, improving the energy performance of the housing corporation stock lies with the housing corporations themselves. Social housing tenants are therefore dependent on the willingness of the housing corporations to improve the energy performance of their housing stock. Next to reducing energy poverty, other goals of improving the dwelling stock are contributing to the energy transition and reducing the CO_2 emissions in the long term (Ürge-Vorsatz & Tirado Herrero, 2012). Often, measures to improve the energy performance of dwellings include insulation, making the dwellings gas-free, and installing solar panels so the dwellings can generate their own energy. A scenario was created to analyse the effect of improving the energy performance of dwellings. For this scenario, KNIME was used to determine how much additional electricity all respondents in the dataset would use to compensate for the reduction in gas usage. For this, it was assumed that all households would keep the same energy usage. However, in the new scenario all energy was provided through electricity while the gas usage went to zero. Afterwards, the effect of the installation of solar panels was modelled in the dataset by adding solar panels to all dwellings in the data that did not already have solar panels. Of all respondents in the dataset, 571 respondents already lived in a gas-free dwelling and 1,266 already had solar panels installed. This scenario will not affect 60 respondents in the dataset since these respondents already lived in gas-free dwellings with solar panels. After the changes to the dataset were made in KNIME, the energy poverty risk prediction macro was run again for the energy improvement scenario. The result of this calculation is shown in figure 25.

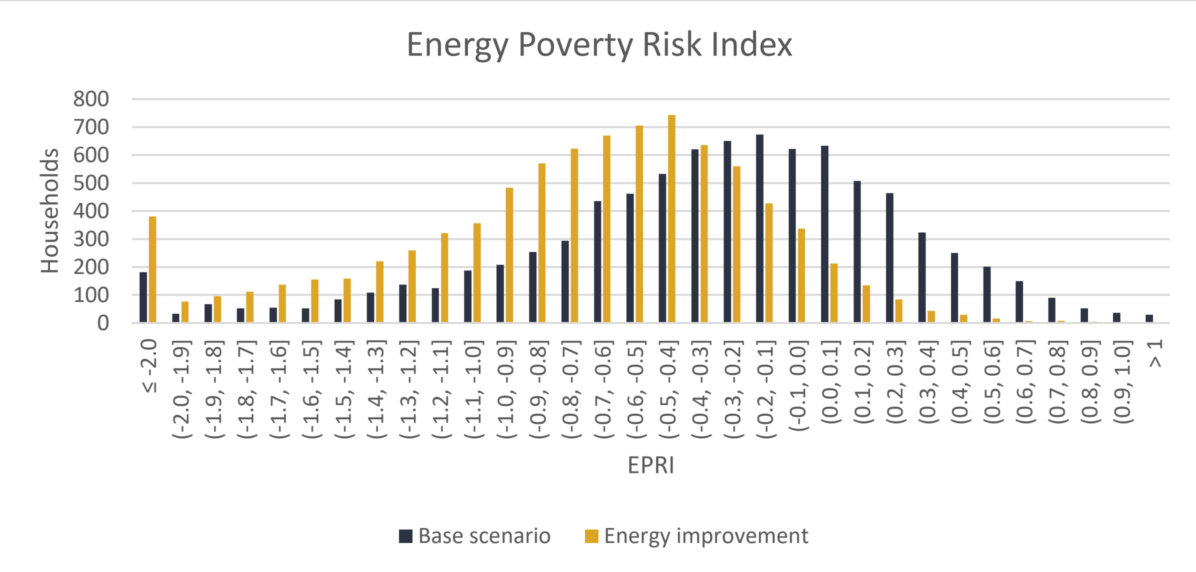

Figure 25. EPRI classes for households, base and energy improvement scenario

The results of the prediction show that when the energy performance of the dwellings is improved through the described measures, the energy poverty risk will decrease. While some respondents remain with an energy poverty risk, this risk is much smaller than in the base scenario. The energy poverty risk has especially been reduced in the higher risk categories of the base scenario. While the energy poverty risks are reduced significantly more in this scenario compared to the energy allowance scenario, the costs for improving the energy performance of the dwellings are much higher than a one-off energy allowance for households. Because housing corporations will not yield a return on their energy improvement investments through energy savings, it is crucial to convince housing corporations of the benefits of the of improving the energy performance of their housing stock. Improving the energy performance and sustainability of the housing stock is crucial for the energy transition, CO_2 emission reduction and energy poverty risk reduction. Contrary to a financial measure, the effect of the improvements of the housing stock will yield remaining long-term effects. It is crucial that the dwellings of households with a high energy poverty risk are given priority when energy improvements are made. By doing this, the largest potential energy poverty risk reduction can be achieved, when there are few monetarily resources available for renovations. In order to do this effectively, it is crucial to clearly identify households and dwellings with a relatively high energy poverty risk. Because this risk is caused and influenced by both personal and dwelling characteristics, the in this research developed EPRI can be particularly suitable for this.

5. Conclusion & Discussion

5.1 Conclusion

The aim of this research was to increase knowledge about energy poverty and improve the method to identify energy poverty. Many different indicators are used to measure energy poverty, however most of these indicators are only based on income and expenses, while other personal and dwelling characteristics are often not included. A literature study was performed to analyse the existing energy poverty indicators and determine the variables that have a potential relation with energy poverty. After the literature study, the data was collected and prepared for further analyses. A descriptive analysis was performed to determine the frequencies and percentages of all value categories and remove value categories with too few respondents. Next, multiple bivariate analyses were performed to determine how representative the data sample was for the entire population and to analyse relations between the predictor variables. Using an EFA it was determined if all existing energy poverty indicators measure a common latent construct and the amount of variance that these indicators cause in their common latent construct. A SEM was created to calculate path coefficients between all variables in the conceptual model to determine both direct and indirect effects on energy poverty. Using the path coefficients determined by the SEM, the EPRI prediction model was made in Excel. Using Excel macros, the EPRI could be determined for all respondents in the dataset and the effects of certain policies on the energy poverty risk could be analysed. The results gained in this research will all help to answer the created research questions.

- Which indicators are currently used in the literature to measure energy poverty?

During the literature study, several indicators were identified that are currently used to measure energy poverty. In general, these existing indicators can be grouped as consensual approaches or expenditure approaches. Consensual approaches are subjective indicators and expenditure approaches are objective indicators that are measured against a critical threshold to define if someone has energy poverty. Five expenditure approaches to measure energy poverty were identified from the literature: the EQ, LIHC, MIS, HCOR and LILEQ indicators. The EQ is the most commonly used indicator to measure energy poverty in existing research. The EQ is easy to calculate however it only includes the income and energy expenditure. The LIHC, MIS and HCOR energy poverty indicators were all developed to represent energy poverty in a more detailed way and be an improvement indicator for energy poverty over the EQ. The LIHC, MIS and HCOR all include more factors than the EQ however, like the EQ, they all do not take the energy performance of dwellings into account. The LILEQ indicator was developed to take the energy performance of dwellings into account however this indicator does not include the actual energy expenses of a household. The energy poverty indicators identified in the literature study all have their own pros and cons however none of the existing energy poverty indicators include personal characteristics. While all existing energy poverty indicators are developed to measure the same concept, energy poverty, the average amount of energy poverty measured by these indicators varies greatly. An EFA was performed to determine if all existing energy poverty indicators measured the same latent concept (energy poverty). The results of the EFA showed that all indicators had a factor loading larger than 0.8 except for the MIS indicator. Because this indicator has an EFA factor loading less than 0.4, it can be concluded that all existing energy poverty indicators measure the same latent concept except for the MIS indicator. After this indicator was removed from the EFA, all remaining energy poverty indicators predict 61.212% of the variance of the common latent construct (energy poverty).

- Which factors are significantly related to energy poverty?

Based the literature review, the expected relations between several factors and energy poverty were hypothesised. For all factors identified in the literature review that were present in the dataset, these hypothesised relations were confirmed or rejected in the CFA of the SEM. It was expected that income would have a negative effect on energy poverty and this was confirmed by the SEM that determined a negative path coefficient between income and energy poverty. Furthermore, the energy expenditure and dwelling size were hypothesised to have a positive effect on energy poverty and this was again confirmed by the SEM. It was expected that residents of detached and semi-detached dwellings would have a higher energy poverty risk. While this could not be tested for detached dwellings since these were not included in the final data, this hypothesis could be confirmed for semi-detached dwellings. Additionally, based on the SEM path coefficients, it can be concluded that residents of terraced dwellings have a higher risk of experiencing energy poverty compared to residents of apartments. Based on the literature review, it was hypothesised that the dwelling construction year would have a negative relation with energy poverty. The results of the SEM confirm this hypothesis; however, no differences were found between dwellings constructed between 1946 and 1965 and dwellings constructed before 1946. It was hypothesised that the degree of urbanisation would have a negative relation with energy poverty. This hypothesis was partly confirmed, since a negative relation was found. However, differences were only measured for the highest two degrees of urbanisation and not for the other value categories. Additionally, the results showed that the degree of urbanisation does not have an indirect effect on energy poverty though income, since the relations between the degrees of urbanisation and income were all non-significant. The SEM confirmed the hypothesis that single-person and single-parent households would have a higher energy poverty risk. It was hypothesised that the education level would have a negative relation with the energy poverty risk. Based on the results of the SEM this hypothesis can be rejected. Although the energy poverty risk factor decreases slightly when education level increases from primary education to vmbo, havo, vwo or mbo, this relation does not significantly continue for the higher education levels. Finally, the relation between energy poverty and age was considered a research gap due to many contradictory findings in existing studies. Based on the results of the SEM, it can be concluded that there is no linear relation between age and energy poverty risk. While the energy poverty risk is higher for people aged 34 or younger or between 45 and 65, the energy poverty risk was lower for people aged between 35 and 44 and all people aged 65 or older.

- What are the characteristics of households with a high energy poverty risk?

Based on the analysis of the respondents with a high EPRI, it can be determined what the common characteristics of this group are. Based on the results of this analysis, it can be concluded that the energy poverty risk is highest for single-person and single-parent households with a relatively low education level. Regarding age, there is no single age group with a highest energy poverty risk. However, the energy poverty risk is relatively low for respondents aged over 65. Dwelling characteristics of households with a relatively high energy poverty risk are semi-detached dwellings that larger than 99m2 and older than 1992. The dwellings of households with a relatively high energy poverty risk commonly have no solar panels and are located in areas with average, low or no urbanisation.

- What advise can be given to reduce the energy poverty risk based on the created model?

The results of the analysis of the current and alternative policies show that incorrect choices may be currently made to reduce energy poverty. The Dutch government currently invests heavily in energy allowances for the relatively poor and vulnerable households. However, because only the income of households is considered, this energy allowance does not always reach the target group and may therefore also sometimes be granted to households with a relatively low energy poverty risk. Partly caused by this, the effect of an energy allowance to reduce energy poverty remains relatively limited. Furthermore, this financial measure does not solve the problem of energy poverty in the long term since households will have energy poverty again when the energy allowance expires.

In the long term, the results show that improving the energy performance of the housing stock can contribute more to reducing energy poverty. Even though these improvements cost a relatively large amount of money in the short term, the effect of this policy, unlike a financial measure, is more effective in reducing energy poverty in the long term. Improving the entire housing stock to reduce energy poverty will take a long time, which is why it is important to prioritise the energy improvement of certain homes. By focusing more on project planning, the dwellings that score worst on the EPRI can be prioritised for energy improvements. Improving the relatively low-scoring dwellings first will yield the largest reduction in energy poverty risk in the short term.

- Can a new model be created that predicts the risk of energy poverty?

In this research a new model was created that predicts the risk of energy poverty, the EPRI. The development of this model started from the identification of relevant variables in the literature study. Because energy poverty is a latent construct, SEM provided the model with relations and path coefficients of both direct and indirect effects. Using these path coefficients, a calculation model could be created in Excel that could determine the energy poverty risk factor for a single household, or for the entire dataset using Excel macros. Contrary to the existing energy poverty indicators, the EPRI includes both multiple personal and dwelling characteristics. Furthermore, while the existing energy poverty indicators are measured on a dichotomous scale, the EPRI predicts the risk of energy poverty on a continues scale. This enables the user of the indicator to interpret in more detail what the effect of policies on energy poverty risk may be.

5.2 Discussion

The results of this research show that it is possible to develop an elaborate energy poverty risk indicator. The developed indicator facilitates a simple and clear identification of risk groups and analysis of policy effects. Contrary to the existing energy poverty indicators, the EPRI includes more factors to predict energy poverty. Because the EPRI predicts energy poverty risk as a continuous value, it enables a more accurate evaluation of the effect of policies compared to the existing energy indicators. The existing energy poverty indicators only predict "energy poverty or no energy poverty" and can therefore not predict how close a household is to experiencing energy poverty or how close they are to avoid energy poverty.

Important stakeholders for whom the results of this research can be relevant are: housing corporations, tenants of housing corporations and local and national governments. Housing corporations can use the EPRI to identify energy poverty risks among their housing stock and tenants. This knowledge can then be used to prioritise renovations to improve the energetic performance of the dwellings with a high EPRI. Because of this, the highest reduction in energy poverty risk can be achieved with the available budget for renovations. This can be beneficial for tenants that have a high energy poverty risk and are therefore in urgent need to have their dwelling renovated. The correct identification of the energy poverty risk may improve their housing situation, especially for tenants that have an elevated risk of suffering from energy poverty.

Similarly to housing corporations, the government can use the EPRI to identify the energy poverty risks of dwellings in certain regions. If the government has a clear overview of where there are high or low energy poverty risks, they can more effectively target their allowances or other subsidies. Because of this, especially vulnerable people will be helped as effectively as possible with the resources that are made available to reduce energy poverty. The government can also use subsidies to encourage homeowners or housing corporations to improve the energy performance of their dwellings.

The EPRI can be used to improve rental policies. Housing corporations can prioritise households with an increased EPRI when assigning dwellings to tenants. These high EPRI scoring tenants can then be assigned to low EPRI scoring dwellings in order avoid high total EPRI scores. Furthermore, the EPRI can be used to elaborate the rent points system used to determine the maximal rent price for a dwelling. By elaborating the rent points system with the EPRI score, more rent can be asked for a dwelling with a low EPRI score and less rent can be asked for a dwelling with a high EPRI score.

Housing corporations can additionally use the EPRI to determine which households could benefit from the advice of an energy coach. An energy coach can then advise the households with a high EPRI score how they can reduce their energy use by changing their energy use habits. Additionally, the energy coaches can advise housing corporations on improving energy performance of their dwellings if necessary.

Some problems were encountered while conducting this research. During the literature review, it quickly became clear that there are many definitions of energy poverty for different global contexts. Furthermore, the number of studies conducted on energy poverty in the Netherlands is still relatively limited, although more articles on energy poverty have been published in the Netherlands in recent years. A further limitation of this research is that not all desired variables were present in the available data. If this data would have been available, it would have been interesting to include the effect of the insulation and heating systems of dwellings, and energy prices on the energy poverty risk. Since the locations of the dwellings at the address level are not present in the WoON 2021 dataset, it was not possible to analyse where the households with a high EPRI live and determine if there is a relation between location and an increased energy poverty risk.

A further limitation of this research is the relatively limited way to check the internal validity of the SEM model. The validity of formative SEM models can be assessed through the convergent validity, indicator collinearity and the significance of the indicator weights determined by the SEM model. However, for this research it was not possible to analyse the convergent validity because the created SEM model does not contain both formative and reflective constructs. The significances of the indicator weights could easily be checked and the indicator weights that were not statistically significant were not included in the prediction model. The indicator collinearity could be assessed by using the VIF and the VIF results showed that there were no problems with the indicator collinearity in the model. Based on the tested validity indicators, it can be determined that the validity of the model was good. However, the interpretation of the internal validity remains difficult because of the relatively few ways for it to be assessed.

Looking at the external validity of this research, it can be determined that the results of this research can be replicated for other housing situations and respondents such as those living in the private rental sector or those living in owner-occupied dwellings. However, when predicting the EPRI values for these other housing situations, it can be recommended to use the methods used in this research to calculate new coefficients for the EPRI prediction. The used methods can additionally be used to determine coefficients for EPRI prediction models for other developed countries since incomes and expenditures can vary widely between developed countries.

Several recommendations for future research can be made. Comparable research can be executed with a focus on dwellings in the private rental sector or owner-occupied dwellings. If these studies are performed, it will additionally be possible to analyse which of these tenure sectors has the highest average energy poverty risk. It would be interesting to perform this research again if a more elaborate dataset becomes available.. This would allow for a comparison with this research in order to determine if comparable results would have been obtained. Furthermore, the prediction model can be expanded with insulation types, heating systems and energy prices. If other relevant factors, such as personal or dwelling characteristics, for energy poverty are discovered in future literature review or studies, it would be interesting to expand the model with these factors. They will further detail the results and will make the energy poverty risk prediction value more reliable. When more knowledge about energy poverty becomes available, the energy poverty risk model can be improved further. With this knowledge, ever better advise can be given to stakeholders, strengthening the reduction of energy poverty.

6. References

Baardink, T. (2020). *Energiearmoede in Groningen* [Master's thesis]. Rijksuniversiteit Groningen.

Basekin, E. (2023). Ruim 600.000 huishoudens in problemen vanwege hoge energiekosten. Retrieved from: https://www.nu.nl/economie/6248820/ruim-600000-huishoudens-in-problemen-vanwege-hoge-energiekosten.html?_ga=2.164788282.467329069.1675151811-240414413.1661450869

Bednar, D. J., & Reames, T. G. (2020). Recognition of and response to energy poverty in the United States. *Nature Energy, 5*(6), 432-439.

Berthold, M. R., Cebron, N., Dill, F., Di Fatta, G., Gabriel, T. R., Georg, F., Meinl, T., Ohl, P., Sieb, C. & Wiswedel, B. (2006) KNIME: the Konstanz Information Miner. *Workshop on Multi-Agent Systems and Simulation (MAS&S), 4th Annual Industrial Simulation Conference (ISC)*, 05-07 June 2006, Palermo, Italy, pp. 58-61.

Bonatz, N., Guo, R., Wu, W., & Liu, L. (2019). A comparative study of the interlinkages between energy poverty and low carbon development in China and Germany by developing an energy poverty index. *Energy and Buildings*, *183*, 817-831.

Bonnard, S., Bruynoghe, C., Deprez, M., & Kestemont, B. (2015). Energieprijzen en energiearmoede. *Federale Overheidsdienst Economie, K.M.O., Middenstand en Energie*.

Bouzarovski, S., Thomson, H., & Cornelis, M. (2021). Confronting energy poverty in Europe: A research and policy agenda. *Energies, 14*(4), 858.

Centraal Bureau voor de Statistiek. (n.d.). WoonOnderzoek Nederland (WoON). Retrieved from: https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/woononderzoek-nederland--woon--

Churchill, S. A., & Smyth, R. (2021). Locus of control and energy poverty. *Energy Economics, 104*, 105648.

Coene, J., & Meyer, S. (2019). Barometer energiearmoede, 2009-2017. Koning Boudewijnstichting

De La Paz, P. T., Juarez Tarrega, F., Su, Z., & Monllor, P. (2022). Sources of Energy Poverty: A Factor Analysis Approach for Spain. *Frontiers in energy research, 10*.

Delbeke, B., Verbeeck, G., & Oosterlynck, S. (2013). Aanpak van energiearmoede via energie-efficiëntie: mogelijkheden en beperkingen. *De sociale staat van Vlaanderen 2013*, 177.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British journal of management, 17*(4), 263-282.

Frost, J. (n.d.). Factor Analysis Guide with an Example. Retrieved from: https://statisticsbyjim.com/basics/factor-analysis/

Goedemé, T., Coene, J., Hubeau, B., & Van Damme, R. (2017). Armoede, energie en wonen: creatieve ideeën voor een toekomst zonder energiearmoede (Onderzoeksdag Universitaire Stichting Armoedebestrijding, 18 oktober 2017).

González-Eguino, M. (2015). Energy poverty: An overview. *Renewable and sustainable energy reviews, 47*, 377-385.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *Partial least squares structural equation modeling (PLS-SEM) using R: A workbook* (p. 197). Springer Nature.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice, 19*(2), 139-152.

Halkos, G. E., & Gkampoura, E. C. (2021). Evaluating the effect of economic crisis on energy poverty in Europe. *Renewable and Sustainable Energy Reviews, 144*, 110981.

Hanafiah, M. H. (2020). Formative vs. reflective measurement model: Guidelines for structural equation modeling research. *International Journal of Analysis and Applications, 18*(5), 876-889.

Herrero, S. T. (2017). Energy poverty indicators: A critical review of methods. *Indoor and Built Environment, 26*(7), 1018-1031.

Hooper, D. (2012). Exploratory factor analysis. In H, Chen (Ed). Approaches to Quantitative Research – Theory and its Practical Application: A Guide to Dissertation Students. Cork, Ireland: Oak Tree Press

Huybrechs, F., Meyer, S., Vranken, J., Campaert, G., Moureau, H., & Storms, E. (2011, December). Energiearmoede in België. In *Rapport final, Convention Universiteit Antwerpen OASeS–ULB CEESE–Electrabel, à paraître*.

Jaya, I., Hermina, N., & Sunengsih, N. (2019). CB-SEM and VB-SEM: Evaluating measurement model of business strategy of internet industry in Indonesia. *International Journal of Scientific & Engineering Research*, *10*(10), 878-883.

KNIME. (n.d.). KNIME Open Source Story. Retrieved from: https://www.knime.com/knime-open-source-story

Kootstra, G. J. (2004). Exploratory factor analysis: Theory and application. *University of Groningen*.

Kraan, J. (2023) Slechts 3 procent 'rechtvaardig' klimaatpakket Jetten voor mensen in armoede. Retrieved from: https://www.nu.nl/klimaat/6262053/slechts-3-procent-rechtvaardig-klimaatpakket-jetten-voor-mensen-in-armoede.html

Kruit, K., van Berkel, P., & Dehens, J. (2021). Energiearmoede in de warmtetransitie-Onderzoek naar beleidsinstrumenten. *CE Delft*.

Kyprianou, I., Serghides, D. K., Varo, A., Gouveia, J. P., Kopeva, D., & Murauskaite, L. (2019). Energy poverty policies and measures in 5 EU countries: A comparative study. *Energy and Buildings, 196*, 46-60.

Lamain, W., Elbert, S., & Ottema, J. (2021). CO_2 uitstoot omlaag, inclusiviteit omhoog: een reactie op het recent verschenen TNO-rapport energiearmoede.

Latif, K. F. (n.d.). How to Perform Exploratory Factor Analysis (EFA) using SPSS. Retrieved from: https://researchwithfawad.com/index.php/lp-courses/data-analysis-using-spss/how-to-perform-exploratory-factor-analysis-using-spss/

Li, W., Chien, F., Hsu, C. C., Zhang, Y., Nawaz, M. A., Iqbal, S., & Mohsin, M. (2021). Nexus between energy poverty and energy efficiency: estimating the long-run dynamics. *Resources Policy, 72*, 102063.

Lysen, E. (1996). The Trias Energica. Solar Energy Strategies for Developing Countries. Eurosun Conference, Freiburg, 16-19 September 1996.

M., R. (2019). Correlation and Collinearity — How they can make or break a model. Retrieved from: https://blog.clairvoyantsoft.com/correlation-and-collinearity-how-they-can-make-or-break-a-model-9135fbe6936a#:~:text=Multicollinearity%20is%20a%20situation%20where,indicates%20the%20presence%20of%20multicollinearity.

Maxim, A., Mihai, C., Apostoaie, C. M., Popescu, C., Istrate, C., & Bostan, I. (2016). Implications and measurement of energy poverty across the European Union. *Sustainability, 8*(5), 483.

Metropoolregio Amsterdam. (n.d.). Onderzoek energiearmoede in de MRA: samen kijken naar kansen. Retrieved from: https://mraduurzaam.nl/energie/energiearmoede/

Ministerie van Binnenlandse Zaken en Koningsrelaties. (2022, June 9). WoonOnderzoek Nederland 2021. Retrieved from: https://www.rijksoverheid.nl/documenten/publicaties/2022/06/09/woononderzoek-nederland-2021

Mulder, P., Dalla Longa, F., & Straver, K. (2021). De feiten over energiearmoede in Nederland Inzicht op nationaal en lokaal niveau. *TNO Innovation for Life.*

Mulder, P., Dalla Longa, F., & Straver, K. (2023). Energy poverty in the Netherlands at the national and local level: A multi-dimensional spatial analysis. *Energy Research & Social Science, 96*, 102892.

Nu.nl. (2022, August 31). Leven wordt nog steeds duurder: inflatie stijgt naar nieuw record. Retrieved from: https://www.nu.nl/economie/6220915/leven-wordt-nog-steeds-duurder-inflatie-stijgt-naar-nieuw-record.html

Okushima, S. (2016). Measuring energy poverty in Japan, 2004–2013. *Energy policy*, *98*, 557-564.

Okushima, S. (2017). Gauging energy poverty: A multidimensional approach. *Energy, 137*, 1159-1166.

Pachauri, S., & Spreng, D. (2011). Measuring and monitoring energy poverty. *Energy policy, 39*(12), 7497-7504.

Pan, L., Biru, A., & Lettu, S. (2021). Energy poverty and public health: Global evidence. *Energy economics, 101*, 105423.

Papada, L., & Kaliampakos, D. (2016). Measuring energy poverty in Greece. *Energy Policy, 94*, 157-165.

Qualtrics. (n.d.). What is factor analysis and how does it simplify research findings?. Retrieved from: https://www.qualtrics.com/experience-management/research/factor-analysis/

Rademaekers, K., Yearwood, J., Ferreira, A., Pye, S. T., Hamilton, I., Agnolucci, P., Grover, D., Karásek, J. & Anisimova, N. (2016). Selecting indicators to measure energy poverty.

Rao, F., Tang, Y. M., Chau, K. Y., Iqbal, W., & Abbas, M. (2022). Assessment of energy poverty and key influencing factors in N11 countries. *Sustainable Production and Consumption, 30*, 1-15.

Rijksoverheid. (n.d.a). Woononderzoek Nederland. Retrieved from: https://www.woononderzoek.nl/

Rijksoverheid. (n.d.b). Hoe vraag ik de energietoeslag aan? 2023. Retrieved from: https://www.rijksoverheid.nl/onderwerpen/koopkracht/hoe-vraag-ik-de-energietoeslag-aan#anker-4-energietoeslag-in-2023

Romero, J. C., Linares, P., & López, X. (2018). The policy implications of energy poverty indicators. *Energy policy, 115*, 98-108.

Schober, P., Boer, C. & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*. *126*(5). 1763-1768.

Sekar, A., & Rai, V. (2018). Predicting Latent Constructs from Passive Datasets–Significance and Opportunities. *Available at SSRN 3285417*.

Sokołowski, J., Lewandowski, P., Kiełczewska, A., & Bouzarovski, S. (2020). A multidimensional index to measure energy poverty: the Polish case. *Energy Sources, Part B: Economics, Planning, and Policy, 15*(2), 92-112.

Sovacool, B. K. (2012). The political economy of energy poverty: A review of key challenges. *Energy for Sustainable Development, 16*(3), 272-282.

Statistics Solutions. (n.d.). correlation (Pearson. Kendall. Spearman). Retrieved from: https://www.statisticssolutions.com/correlation-pearson-kendall-spearman/

Sturgis, P. (2016, April 13) Structural Equation Modelling (SEM): What it is and what it isn't. National Centre for Research Methods, University of Southampton, online learning resource. Retrieved from: https://www.ncrm.ac.uk/resources/online/all/?main&id=10416

Taherdoost, H., Sahibuddin, S., & Jalaliyoon, N. (2022). Exploratory factor analysis; concepts and theory. *Advances in applied and pure mathematics, 27*, 375-382.

The Analysis Factor (n.d.) Factor analysis: a short introduction, part 1. Retrieved from: https://www.theanalysisfactor.com/factor-analysis-1-introduction/

Thomson, H., Snell, C., & Bouzarovski, S. (2017). Health, well-being and energy poverty in Europe: A comparative study of 32 European countries. *International Journal of Environmental Research and Public Health, 14*(6), 584.

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in sport sciences, 21*(1).

Ürge-Vorsatz, D. & Tirado Herrero, S. (2012). Building synergies between climate change mitigation and energy poverty alleviation. In: *Energy Policy, 49*, 83-90.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology, 9*(2), 79-94.

Appendices

Appendix 1: Characteristics identified in the literature review and their expected effect on energy poverty

Characteristic/feature	Expected effect on energy poverty
Economic characteristics	
Energy prices	Positive
Rent prices	Positive
Income	Negative
Housing costs	Positive
Dwelling characteristics	
Energy performance	Negative
Dwelling type	Energy poverty occurs more often for households living in detached and semi-detached dwellings
Dwelling size	Positive
Dwelling construction year	Negative
Degree of urbanisation	Negative
Socio-demographic characteristics	
Household type	Energy poverty occurs more often for single-person and single-parent households
Age	Research gap (caused by the many contradictory findings in the existing literature)
Education level	Negative
Tenure	Households living in rented dwellings are more likely to experience energy poverty

Appendix 2: Measurement scales of the selected variables

Variables	Measurement scale
Income	Ratio
Household type	Nominal
Education level	Ordinal
Age	Ordinal
Degree of urbanisation	Ordinal
Dwelling size	Ordinal
Dwelling construction year	Ordinal
Dwelling type	Nominal
Solar panels	Nominal (dichotomous)
Dwelling energy use	Ratio
Energy expenditure	Ratio
Other housing costs	Ratio

Appendix 3: Categorical variables descriptives

Household type		
	Frequency (count)	Percent
Single-person	4,751	55.3
Single-parent with children	909	10.6
Couple without children	2,032	23.7
Couple with children	895	10.4
Total	8,587	100.0
Education level		
	Frequency (count)	Percent
Primary education	1,584	18.4
VMBO, MBO1	2,302	26.8
HAVO, VWO, MBO	3,020	35.2
HBO/WO-bachelor	1,169	13.6
HBO/WO-master	512	6.0
Total	8,587	100.0
Age		
	Frequency (count)	Percent
34 or younger	1,310	15.3
35-44	1,040	12.1
45-54	1,175	13.7
55-64	1,587	18.5
65-74	1,814	21.1
75 or older	1,661	19.3
Total	8,587	100.0
Degree of urbanisation		
	Frequency (count)	Percent
No	603	7.0
Low	1,115	13.0
Average	1,430	16.7
High	2,643	30.8
Very high	2,796	32.6
Total	8,587	100.0
Dwelling Size		
	Frequency (count)	Percent
Less than 50m2	539	6.3
50-74m2	2,316	27.0
75-99m2	3,463	40.3
100-149m2	2,269	26.4
Total	8,587	100.0
Dwelling construction year		
	Frequency (count)	Percent
1945 or older	689	8.0
1946-1964	1,516	17.7
1965-1974	1,848	21.5
1975-1991	2,466	28.7
1992-2005	984	11.5
2006 or newer	1,084	12.6
Total	8,587	100.0

Dwelling type		
	Frequency (count)	Percent
Apartment	4,624	53.8
Terraced	2,597	30.2
Semi-detached	1,366	16.0
Total	8,587	100.0
Solar panels		
	Frequency (count)	Percent
Yes	1,269	14.8
No	7,318	85.2
Total	8,587	100.0

Appendix 4: Chi-square test and symmetric measures solar panels – nominal variables

Solar panels							
	Chi-square tests				Symmetric measures		
		Value	df	Asymptotic significance (2-sided)		Value	Approximate significance
Household type	Pearson chi-square	89.295	3	0.000	Cramer's V	0.097	0.000
Dwelling type	Pearson chi-square	512.435	2	0.000	Cramer's V	0.244	0.000

Appendix 5: Chi-square test and symmetric measures household type – dwelling type

Household type							
	Chi-square tests				Symmetric measures		
		Value	df	Asymptotic significance (2-sided)		Value	Approximate significance
Dwelling type	Pearson chi-square	706.072	6	0.000	Cramer's V	0.203	0.000

Appendix 6: Mann-Whitney U-tests: solar panels - ordinal variables

Mann-Whitney U-test: solar panels				
	Mann-Whitney U	Z	Asymp. Sig. (2-tailed)	r^2
Education level	4,497,254.000	-1.859	0.063	0.000
Age	4,395,575.000	-3.086	0.002	0.001
Degree of urbanisation	3,457,047.500	-15.095	0.000	0.027
Dwelling size	3,544,239.000	-14.241	0.000	0.024
Dwelling construction year	4,457,761.500	-2.326	0.020	0.001

Appendix 7: Independent samples t-test solar panels – other housing costs

Independent samples t-test: solar panels				
Group statistics				
	Solar panels	N	Mean	Std. Deviation
Other housing costs	Yes	1,269	7,851.640	2,151.966
	No	7,318	7,548.712	2,345.057

Independent samples t-test: solar panels							
		Levene's test for equality of variances		t-test for equality of means			Effect size
		F	Sig.	t	df	Sig.	g
Other housing costs	Equal variances assumed	7.730	0.005	4.298	8,585.000	0.000	
	Equal variances not assumed			4.566	1,830.583	0.000	0.140

Appendix 8: Kruskal-Wallis H-tests: household type - ordinal variables

Kruskal-Wallis H-test: household type				
	Kruskal-Wallis H	df	Asymp. Sig.	E^2
Education level	89.902	3	0.000	0.010
Age	1,009.079	3	0.000	0.117
Degree of urbanisation	89.881	3	0.000	0.010
Dwelling size	948.868	3	0.000	0.110
Dwelling construction year	64.578	3	0.000	0.007

Appendix 9: Kruskal-Wallis H-tests dwelling type-ordinal variables

Kruskal-Wallis H-test: dwelling type				
	Kruskal-Wallis H	df	Asymp. Sig.	E^2
Education level	44.783	2	0.000	0.005
Age	43.704	2	0.000	0.005
Degree of urbanisation	1,363.033	2	0.000	0.153
Dwelling size	2,792.531	2	0.000	0.313
Dwelling construction year	265.941	2	0.000	0.029

Appendix 10: ANOVA household type – other housing costs

Tests of Homogeneity of Variances				
		Levene Statistic	df1	Sig.
Other housing costs	Based on Mean	3.678	3	0.012

ANOVA							
		Sum of Squares	df	Mean Square	F	Sig.	ω^2
Other housing costs	Between groups	7,432,670,691,309.000	3	2,477,556,897,103.000	548.390	0.000	0.161
	Within Groups	38,776,941,622,242.000	8,583	4,517,877,388.000			
	Total	46,209,612,313,552.000	8,586				

Appendix 11: ANOVA dwelling type – other housing costs

Tests of Homogeneity of Variances				
		Levene Statistic	df1	Sig.
Other housing costs	Based on Mean	2.189	2	0.112

ANOVA							
		Sum of Squares	df	Mean Square	F	Sig.	ω^2
Other housing costs	Between groups	1,610,959,252,590.000	2	8,054,796,266.250	155.032	0.000	0.035
	Within Groups	44,598,653,060,962.000	8,584	5,195,556.042			
	Total	46,209,612,313,552.000	8,586				

Appendix 12: Spearman correlations ordinal variables

Correlations						
Spearman's rho		Education level	Age	Degree of urbanisation	Dwelling size	Dwelling construction year
Education level	Correlation coefficient	1.000				
	Sig. (2-tailed)	.				
Age	Correlation coefficient	-0.392	1.000			
	Sig. (2-tailed)	0.000	.			
Degree of urbanisation	Correlation coefficient	0.113	-0.110	1.000		
	Sig. (2-tailed)	0.000	0.000	.		
Dwelling size	Correlation coefficient	-0.072	0.143	-0.299	1.000	
	Sig. (2-tailed)	0.000	0.000	0.000	.	
Dwelling construction year	Correlation coefficient	0.009	0.062	-0.146	0.055	1.000
	Sig. (2-tailed)	0.382	0.000	0.000	0.000	.

Appendix 13: Spearman correlations ordinal - ratio variables

Correlations		
Spearman's rho		Other housing costs
Education level	Correlation coefficient	0.143
	Sig. (2-tailed)	0.000
Age	Correlation coefficient	-0.003
	Sig. (2-tailed)	0.780
Degree of urbanisation	Correlation coefficient	-0.067
	Sig. (2-tailed)	0.000
Dwelling size	Correlation coefficient	0.387
	Sig. (2-tailed)	0.000
Dwelling construction year	Correlation coefficient	0.164
	Sig. (2-tailed)	0.000

Appendix 14: Spearman correlations age - endogenous variables

Correlations					
		Electricity usage	Gas usage	Income	Energy expenses
Age	Correlation coefficient	-0.004	0.086	-0.109	0.058
	Sig. (2-tailed)	0.689	0.000	0.000	0.000

Appendix 15: Pearson correlations other housing costs - endogenous variables

Correlations					
		Electricity usage	Gas usage	Income	Energy expenses
Other housing costs	Correlation coefficient	0.328	0.174	0.584	0.294
	Sig. (2-tailed)	0.000	0.000	0.000	0.000

Appendix 16: KMO and Bartlett's test EFA

KMO and Bartlett's test			
Kaiser-Meyer-Olkin measure of sampling adequacy	Bartlett's test of sphericity		
	Approx. Chi-square	df	Sig.
0.575	22,984.982	10	0.000

Appendix 17: Correlation matrix EFA

Correlation matrix						
		EQ	LIHC	MIS	HCOR	LILEQ
EQ	Correlation	1.000				
	Sig. (1-tailed)	.				
LIHC	Correlation	0.456	1.000			
	Sig. (1-tailed)	0.000	.			
MIS	Correlation	-0.002	0.001	1.000		
	Sig. (1-tailed)	0.416	0.445	.		
HCOR	Correlation	0.883	0.244	-0.018	1.000	
	Sig. (1-tailed)	0.000	0.000	0.041	.	
LILEQ	Correlation	0.452	0.680	0.012	0.274	1.000
	Sig. (1-tailed)	0.000	0.000	0.134	0.000	.
Determinant = 0.076						

Appendix 18: Communalities EFA

Communalities		
	Initial	Extraction
EQ	1.000	0.945
LIHC	1.000	0.831
MIS	1.000	0.045
HCOR	1.000	0.950
LILEQ	1.000	0.825
Extraction method: principal component analysis		

Appendix 19: Total variance explained EFA

Total variance explained						
Initial eigenvalues				Extraction sums of squared loadings		
Component	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	2.509	50.180	50.180	2.509	50.180	50.180
2	1.087	21.745	71.926	1.087	21.745	71.926
3	0.996	19.928	91.853			
4	0.321	6.412	98.265			
5	0.087	1.735	100.000			
Extraction method: principal component analysis						

Appendix 20: Component matrix EFA

Component matrix					
	EQ	LIHC	MIS	HCOR	LILEQ
Component 1	0.901	0.733	-0.004	0.781	0.742
Extraction method: principal component analysis					

Appendix 21: Total variance explained second EFA

Total variance explained						
Initial eigenvalues				Extraction sums of squared loadings		
Component	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	2.448	61.212	61.212	2.448	61.212	61.212
2	1.010	25.245	86.456	1.010	25.245	86.456
3	0.454	11.350	97.807			
4	0.088	2.193	100.000			
Extraction method: principal component analysis						

Appendix 22: Personal characteristics value category percentages WoON 2021 respondents and respondents with a high EPRI

Education level		
	WoON 2021	High EPRI
Primary education	18.4	31.1
VMBO, MBO1	26.8	33.2
HAVO, VWO, MBO	35.2	30.2
HBO/WO-bachelor	13.6	4.1
HBO/WO-master	6.0	1.4
Age		
	WoON 2021	High EPRI
34 or younger	15.2	21.5
35-44	12.1	7.7
45-54	13.7	15.4
55-64	18.5	26.0
65-74	21.2	16.3
75 or older	19.3	13.1
Household type		
	WoON 2021	High EPRI
Single-person	55.3	72.8
Single-parent with children	10.3	12.8
Couple without children	23.0	4.2
Couple with children	10.2	10.2

Appendix 23: Dwelling characteristics value category percentages WoON 2021 respondents and respondents with a high EPRI

Degree of urbanisation		
	WoON 2021	High EPRI
No	7.0	10.9
Low	13.0	19.4
Average	16.6	21.9
High	30.8	26.4
Very high	32.6	21.4
Dwelling size		
	WoON 2021	High EPRI
Less than 50m2	6.5	5.7
50-74m2	26.9	18.9
75-99m2	40.1	38.9
100m2 or more	26.5	36.5
Dwelling construction year		
	WoON 2021	High EPRI
1945 or older	8.2	9.0
1946-1964	17.7	23.5
1965-1974	21.5	27.7
1975-1991	28.4	29.8
1992-2005	11.5	3.9
2006 or newer	12.7	6.1
Dwelling type		
	WoON 2021	High EPRI

Apartment	53.8	37.2
Terraced	30.3	35.4
Semi-detached	15.9	27.4
Solar panels		
Solar panels	WoON 2021	High EPRI
Yes	14.8	10.7
No	85.2	89.3