Residential energy consumption in theory and practice—the effect of home ownership

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Management of Technology**

Faculty of Technology, Policy and Management

by

Kim Fernández Gómez

Student number: 4629507

To be defended in public on August 9, 2022

Graduation committee

Chairperson	:	Prof.dr. K. Blok, Energy and Industry
First supervisor	:	Dr. E. Schröder, Economics of Technology and Innovation
Second supervisor	:	Prof.dr. K. Blok, Energy and Industry

Acknowledgements

I would like to express my gratitude to the chair of my graduation committee and second supervisor, Prof. dr. Kornelis Blok, for his insightful feedback. I would also like to thank my first supervisor, Dr. Enno Schröder, for the guidance and encouragement he provided to me during the project.

Abstract

One of the main strategies for reducing energy demand—and thereby carbon emissions—is increasing energy efficiency. In the Netherlands, energy efficiency policy and energy reduction targets in the residential sector are based on the theoretical energy consumption of buildings, which often over- or underestimates the measured energy consumption. The difference between the actual energy consumption and the calculated energy consumption of a building is referred to as the energy performance gap (EPG). As a consequence of this discrepancy, the estimated energy savings are inaccurate and the energy-saving targets are unattainable. Narrowing the EPG is necessary to develop more accurate estimates of energy savings and realistic targets, and to improve the effectiveness of energy reduction policies and campaigns. Most studies on the EPG have focused predominantly on data from the social rental housing sector, failing to represent the national distribution of home ownership type. At the same time, homeowners and tenants have been shown to behave differently regarding energy consumption.

This thesis investigates the effect of home ownership on the actual natural gas consumption and the EPG by descriptive statistics, correlation analysis, and multiple linear regression on a representative sample of the Dutch housing stock. The multiple regression analysis controls for building and occupant characteristics that are expected to influence the actual gas consumption and the EPG, in order to measure the ceteris paribus effect of ownership type on the actual gas consumption and the EPG.

The results show that ownership type does not have a practically significant effect on actual gas consumption or the EPG, while controlling for building and household characteristics. However, without controlling for these factors, there is a moderate positive correlation between home ownership and actual gas consumption, and a weak positive correlation between home ownership and the magnitude of overpredictions. This suggests that observed differences in gas consumption or EPG between ownership groups may be explained by building and household characteristics, rather than by potential behavioral differences. Specifically, the positive correlation between home ownership and actual gas consumption can be explained by the larger floor area, type of buildings, higher income, and larger household size. The positive correlation between home ownership and the size of overpredictions is explained by type of building and larger floor area. Thus, there are no major differences in energy consumption behavior between homeowners and tenants that cause large differences in their actual gas consumption or EPG. Nevertheless, the distinction of ownership type may still be of practical use to policymakers. Targeting homeowners could be an efficient way to promote energy-saving measures in the largest and highest energy-consuming dwellings.

Contents

1 Introduction

1.1 Climate change and energy efficiency

Nearly 75% of green house gas emissions responsible for global warming have originated from the energy sector, making this sector key in the mitigation of climate change (International Energy Agency (IEA), 2021) Climate change has caused severe, irreversible damages to people and nature, and the impacts have recently appeared worse than previously estimated (Intergovernmental Panel on Climate Change (IPCC), 2022). Ocean acidification, sea level rise, and more frequent and intense extreme weather events like droughts, hot extremes and heavy precipitation are some of the consequences of human-induced climate change. These, in turn, have lead to numerous adverse impacts, including increased heat-related human mortality, jeopardized food and water security, and deterioration of ecosystems including species extinction. But, further damages caused by climate change are expected to be reduced substantially if global warming is limited to 1.5°C in the near term (by 2040) (Intergovernmental Panel on Climate Change (IPCC), 2022), and that requires reducing emissions. At the same time, growing economies and increasing world population contribute to a rising demand for energy services. Because reducing energy intensity is essential for decarbonising the energy supply, using energy more efficiently is one of four key measures in the aim to achieve the 1.5°C limit in the next ten years (International Energy Agency (IEA), 2021). In addition to climate change, other reasons to invest in energy efficiency to reduce energy demand are ensuring the security of energy supply, affordability, and reducing the dependence on foreign fossil fuel imports (European Commission, n.d.). The latter has received particular attention since the Russian invasion of Ukraine in February 2022 (International Energy Agency (IEA), 2022). In sum, focusing on increasing energy efficiency is vital.

1.2 Energy efficiency in the residential sector

An important area for improving energy efficiency is the buildings sector, one of Europe's largest energy consumers (Visscher et al., 2016). A large part of this consumption is attributable to residential buildings; the residential sector demands over one quarter of the total EU final energy use (Eurostat, 2022). In the Netherlands, measures and policies to promote energy efficiency and to decrease energy demand of buildings are based on their theoretical energy consumption, which is calculated using building characteristics and installations, and assumptions about its occupants. For example, the theoretical figure is used for calculating the cost-effectiveness of efficiency measures and for setting energy-saving targets. It is evident that the theoretical consumption must reflect the real consumption if these decisions are to be accurate and effective (Laurent et al., 2013; Majcen et al., 2013a; van den Brom et al., 2017; van den Brom, 2020). However, research has shown that the theoretical consumption deviates significantly from the real, measured consumption. This discrepancy is referred to as the energy performance gap (EPG). In some cases, the estimate is off by 100% from the actual consumption (Majcen et al., 2013a). According to projections, using these inaccurate estimates leads to unattainable energy savings targets, misleading predictions of energy demand and energy savings after a renovation, and affects the cost-effectiveness of energy efficiency measures and policies (Laurent et al., 2013; Majcen et al., 2013a). For more effective energy-saving policy and campaigns, studying actual energy consumption data is essential (Laurent et al., 2013; van den Brom et al., 2017). In addition, in order to narrow the gap between actual and theoretical consumption, it is necessary to understand what factors contribute to it and to what extent. Previous studies have revealed a number of factors influencing the EPG, including certain building characteristics, occupant behavior, and occupant characteristics (Majcen et al., 2013b; van den Brom et al., 2017).

However, these studies are either based on data from social housing associations or in which the social housing stock is over-represented compared to private rental or owner-occupied dwellings. Social housing is defined as rental homes with a rent below a certain limit. These houses are usually owned by housing associations, which must let the majority of their properties to people with incomes below a certain limit. In contrast, the private housing sector is more expensive; there is no limit on the rent that can be charged (Government of the Netherlands, n.d.b). While social housing represents a substantial portion of the Dutch housing stock (around 30%), it has some limitations to the generalizability. First, the average income of occupants in the social housing sector is below the Dutch population average. Second, the social housing data consists exclusively of rental dwellings (van den Brom et al., 2017). This has been pointed out as a limitation of the research, because several studies have discovered behavioral differences of tenants and homeowners in energy consumption (Madlener and Hauertmann, 2011; Aydin et al., 2017). Finding out the influence of the type of ownership on the EPG may add to the understanding of the relationship between occupant characteristics and energy consumption. This knowledge may contribute to modeling a more accurate theoretical energy consumption, allowing for more effective energy-saving policy and achieving the energy reduction targets.

1.3 The energy performance gap

In the buildings sector, the largest potential for energy savings lies in currently existing buildings, which are expected to comprise about 75% of the total building stock by 2050 (Visscher et al., 2016). To stimulate energy efficiency improvements in existing buildings, the Energy Performance Building Directive introduced in 2003 requires EU countries to issue an energy performance certificate for all buildings (ISSO, 2020). The method of calculation of the energy performance certificate is specific to each country. In the Netherlands, the energy performance certificate is known as the energy label. This document contains a rating of the energy performance of the building, and a theoretical energy consumption of the building, estimated based on building characteristics and assumptions about occupant behavior (Visscher et al., 2016; van den Brom, 2020).

While the initial aim of the directive was to create awareness on the energy efficiency of a building to be bought or rented, it is currently used for a variety of purposes related to energy savings, including formulating policy goals (Visscher et al., 2016; van den Brom, 2020). At the same time, actual household energy consumption data is collected by energy suppliers, and research shows a discrepancy between the theoretical and actual energy consumption—the energy performance gap (Visscher et al., 2016; Majcen et al., 2013a). Specifically, the research shows that actual energy consumption is higher than calculated in newer, more efficient buildings, whereas it is lower than expected in buildings with a poor energy efficiency (Majcen et al., 2013a). For the more efficient buildings, the explanation for the higher actual energy consumption is a combination of construction faults causing the building to underperform, and occupant behavior. The occupant behavior is partly due to the rebound effect; a higher energy efficiency increases energy service demand, thereby offsetting part of the potential energy savings from the efficiency improvement. For example, occupants in buildings with a more efficient heating system may be more inclined to increase the temperature setting. In contrast, there is evidence that the performance of the less efficient buildings may be underestimated. Several studies showed that solid walls turned out to allow less heat transmission than previously assumed, which explains the overestimation of the energy consumption. In addition, due to their poor insulation, their occupants may be more frugal in their heating behavior than accounted for in the theoretical models. This is referred to as the prebound effect (Visscher et al., 2016).

The theoretical calculation of a building's energy consumption as contained in the energy label is described in ISSO publication 82.1 and 82.3.[1] According to the calculation method introduced in 2011, the energy label, ranging from A++ as the most energy efficient through label G being the least energy efficient, is based on the magnitude of the energy index. The energy index is a function of the floor area, the heat loss area, and an estimate of the total energy consumption of the dwelling. The latter is based on the insulation, type of heating system, type of domestic hot water system, type of ventilation system, and air-tightness of the building, and on assumptions such as an average indoor temperature of 18°C and the number of occupants based on the floor area (van den Brom, 2020). As a result of the assumptions, the validity of the theoretical energy consumption is limited. Nevertheless, theoretical energy consumption is often used by in by policymakers, such as for developing energy-saving targets and policies, monitoring energy performance, and determining maximum rent and subsidies. In addition, it is used in practice for estimating the outcomes of energy efficiency renovations (van den Brom, 2020). Thus, for optimal effectiveness of these policies and renovation measures, it is crucial that the theoretical energy accurately represents the actual energy consumption of the dwelling, i.e., narrowing the EPG.

In particular, the EPG with respect to natural gas use is informative to study. This is because electricity consumption of household appliances, which in practice accounts for about one third of total electricity consumption, is not included in the calculation of theoretical electricity consumption. For natural gas consumption, the only end use excluded in the theoretical calculation is gas for cooking, which represents less than 3% of total gas use. Since the theoretical estimate for gas consumption accounts for essentially all types of actual gas uses, the EPG in gas use measures how much the predicted gas use actually deviates from the real gas use. In contrast, the EPG in electricity use is for a large part expected.

1.4 Research questions and objectives

This master thesis adds to the investigation of the factors influencing the EPG. Specifically, it formulates the following main research question:

How does home ownership influence the energy performance gap in residential gas consumption?

In order to answer the main question, the research is broken down into the following sub-questions:

- To what extent does the type of ownership—owner-occupied or rental—predict a) the actual natural gas consumption and b) the size of the EPG?

[1] An updated version of the calculation method of the energy label was introduced in 2021 (Government of the Netherlands, 2020; ISSO, 2020), but since the database of this master thesis contains energy label data as registered in 2018 (van Zoelen and Gopal, 2019), an earlier version is described.

- What empirical variables influence the relationship between ownership and the EPG?
- What is the mechanism of the relationship between ownership and the EPG?

The aim of this research is threefold: first, to deliver an analysis of the energy performance gap in a sample of households representative of the Dutch population; second, to determine to what extent ownership of the dwelling influences the size of the energy performance gap; third: to investigate the mechanism that could explain the influence of ownership on the energy performance gap. After realizing these objectives, the findings could be used for developing more accurate estimates of energy consumption and savings for policy and practice. In addition, they could serve for designing more effective energy-saving policy and campaigns, targeted toward the right household groups.

2 Literature review

2.1 Evidence of the EPG

This thesis aims to fill a knowledge gap regarding the factors that explain the EPG. In the past, research on the relationship between energy labels and actual energy consumption has been limited by the lack of accessible energy label data. Early studies are based on small samples, which can be problematic for the statistical significance of the results. For example, Guerra Santín and Itard (2012) compared the theoretical with the actual energy consumption of a few hundred Dutch households. When viewing the comparison by building performance category, it appeared that, from high to low-performance buildings, the theoretical consumption was increasingly higher than the real consumption. In fact, the differences between actual energy consumption over the different performance categories were found insignificant, although this was attributed to the small sample size.

A large sample of nearly 200,000 Dutch dwellings was first analyzed by Majcen et al. (2013a). Energy label data (including theoretical energy use) was acquired from Agentschap NL[2] and was merged with actual energy consumption data from Statistics Netherlands (in Dutch: Centraal Bureau voor de Statistiek, CBS) on an address level, which is collected by energy suppliers. Despite its size, the sample failed to reflect the national distribution of ownership types. For example, the percentage of owner-occupied dwellings in the sample was 20%, while the national figure was 55%. In contrast, social housing comprised 79% of the ownership types in the sample, while on a national level it was only 33%. As a result, owner-occupied dwellings were underrepresented, which might introduce bias. Over the average of the sample, the theoretical gas consumption was much higher than actual gas use, while the opposite was true for electricity consumption. In the latter case this is expected; electricity consumption of household appliances, which in practice accounts for about one third of total electricity consumption, is not included in the theoretical calculation. For natural gas consumption, the only end use excluded in the theoretical calculation is gas for cooking, which represents less than 3% of total gas use. The study compared mean theoretical and actual natural gas consumption (Figure 1) and mean theoretical and actual natural gas consumption per unit floor area of each energy label (Figure 2), showing that the mean actual consumption was higher than the theoretical for higher performance labels (A and B), while it was lower than the theoretical in lower performance labels (D-G).

[2] In 2014, Agentschap NL was merged with another government organization and is since known as Netherlands Enterprise Agency (in Dutch: Rijksdienst voor Ondernemend Nederland, RVO) (Government of the Netherlands, 2014).

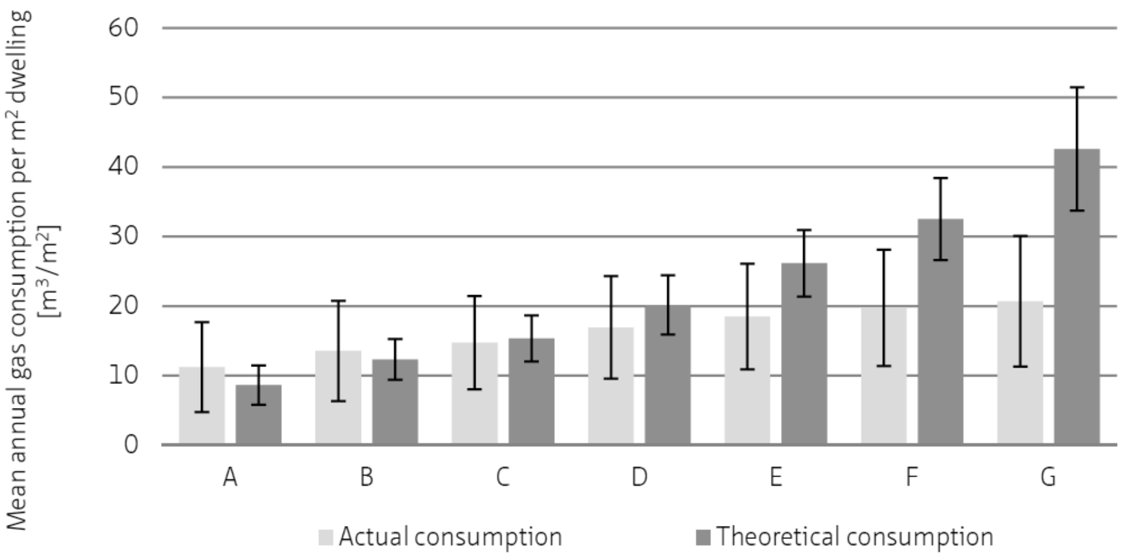


Figure 1: Comparison by energy label of the actual and theoretical natural gas consumption by Majcen (2016), revealing the EPG. The theoretical consumption increasingly exceeds the actual consumption toward lower-performance labels, while the opposite occurs toward higher-performance labels. Error bars represent ±1 standard deviation.

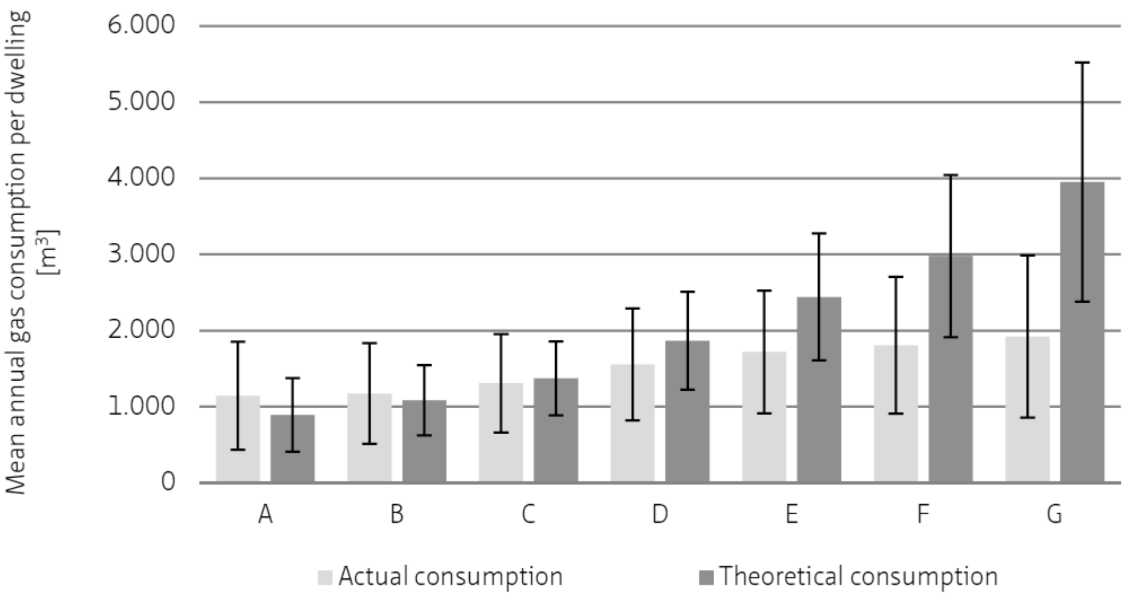


Figure 2: Comparison by energy label of the actual and theoretical natural gas consumption per m^2 floor area by Majcen (2016). Error bars represent ±1 standard deviation. Correcting for floor area does not change the pattern in the EPG as seen in Figure 1.

Similar findings were obtained in earlier studies of the Netherlands (Guerra Santín and Itard, 2012; Tigchelaar et al., 2011), and while other EU member states use different models for calculating theoretical energy consumption, they find similar results as well (Cayre et al., 2011; Hens et al., 2010). Later, very similar results were found in the analysis of a database of over 1.4 million dwellings from the Dutch social housing sector (van den Brom, 2020). If the theoretical gas consumption is to be used for policy targets and predicting the cost-effectiveness of energy efficiency measures, it must represent the actual consumption more accurately. And for that, the factors contributing to their discrepancy must be identified.

2.2 Explaining the EPG

2.2.1 The rebound effect

As mentioned earlier in subsection 1.3, one of the explanations for the EPG is the rebound effect (Visscher et al., 2016). Rebound effects cause an increase in energy demand as a result of higher energy efficiency, partly offsetting the energy savings (Berkhout et al., 2000). The increase in energy efficiency leads to a lower energy consumption of a particular amount of energy service use, such as heating or lighting. The resulting money savings can be used to increase the use of the energy service, e.g., by increasing the temperature setting after efficiency improvements in heating, such as better insulation. As a consequence, the energy savings originally predicted by engineers following an efficiency improvement may not be fully realized (Sorrell, 2015). Considering the engineers' calculations only and neglecting rebound effects is problematic when policymakers set targets for the reduction of energy consumption through energy efficiency improvements, as rebound effects can cause them to be less effective (Berkhout et al., 2000). Some studies have found correlations between certain household characteristics and the rebound effect. For example, income appears to influence the size of the rebound effect, as well as the type of ownership of the dwellings (rental or owner occupied) (Madlener and Hauertmann, 2011; Aydin et al., 2017).

2.2.2 Type of home ownership

Madlener and Hauertmann (2011) studied the rebound effect in German residential space heating, and discovered that its size differs significantly between tenants and homeowners; for homeowners the rebound effect was 12.2%, while for tenants it was 40%. When differentiating between low and high income groups within homeowner and tenant groups, the rebound effect for homeowners was similar in both groups. This suggests that there is no saturation effect, i.e., the energy service use of homeowners is already satiated and not limited by affordability. However, low-income tenants showed a rebound effect as high as 50%, implying that half of the potential energy savings are lost due to increased energy service use. For high income tenants, the rebound effect was 31%. The lower rebound effect for higher incomes than for lower is in accordance with the saturation effect, meaning that for lower income groups, the affordable use of energy services is often further away from the comfortable level. As a result, increases in efficiency induce an increased use of energy services, e.g., a higher room temperature. Higher-income groups, on the other hand, are expected to already consume energy services near the level of comfort, and therefore a lower rebound effect is expected. While the differences between income levels are explained and also in accordance with other studies, the explanation for the differences between ownership type remains unclear. Perhaps, it is in a way related to the absolute income of tenants and homeowners; the separation between high and low monthly disposable income groups was drawn at €2,710 for the homeowner group while at €1,920 for the tenant group, suggesting overall lower incomes in tenants. While the existence of the difference in rebound effects between ownership type is relevant for policy (Madlener and Hauertmann, 2011), the question of where this difference comes from remains unanswered.

Aydin et al. (2017) investigated the rebound effect in residential heating in the Netherlands, and also found a significant difference between homeowners and tenants: 26.7% and 41.3%, respectively. The authors indicate that tenants are more inclined to changing behavior than homeowners. They relate this difference to the expectation that higher-income households are less affected by changes in the affordability of heating. The results are similar to those found in German households by Madlener and Hauertmann (Madlener and Hauertmann, 2011). However, when analyzing different income levels within the homeowner group, they find a rebound effect of 40% for the lower quantile and 19% for the upper quantile. In fact, the rebound effect of the lower-income homeowners is approximately the same as the average rebound effect of tenants. While this difference is in line with the expected effect of income, it deviates from the German results where homeowners displayed similar rebound effects regardless of their income level (Madlener and Hauertmann, 2011). Within the tenant group, Aydin et al. found a similar influence of income on the rebound effect as in the homeowner group, which is in accordance with the German study (Madlener and Hauertmann, 2011). From both studies, it appears that income may be related to the influence of ownership on the rebound effect.

Besides the influence of ownership on the rebound effect, previous studies have also investigated the influence of ownership type on actual and theoretical gas consumption and the EPG. A multiple regression analysis of the Dutch Housing Survey 2012 (WoON 2012) by Majcen et al. (2015) found ownership type to be significant for both theoretical energy consumption and the EPG, but not for actual energy consumption (all expressed per unit floor area, and controlling for a number of building and occupant characteristics). However, in a sample of nearly 49,000 dwellings in the Amsterdam area, the type of ownership was found insignificant for all three dependent variables, most likely because it consists mainly of social housing and is therefore not representative for ownership type (Majcen et al., 2015). In another sample of about 40,000 dwellings, Majcen et al. (2013b) showed that owner-occupied dwellings correlate with a slightly higher theoretical gas consumption than social housing (rental) dwellings, although in this sample, social housing was overrepresented as well. Possibly, since the theoretical calculation does not consider ownership type, it makes sense that there is in fact not a causal relationship, but rather that it should be mediated by a third variable, e.g., the size of the house. Regarding actual energy consumption, the study showed that owner-occupied dwellings consume significantly less volume of natural gas than social housing dwellings. The author attributes this to possibly better insulation in owner-occupied dwellings or to different behavior (Majcen et al., 2013b). In order to understand how exactly ownership type matters for gas consumption and/or the EPG, it is important to identify the variables that may interfere with this relationship, such that they can be controlled for in multiple regression analyses.

2.2.3 Other variables explaining the EPG

The two studies that investigated the effect of ownership on the EPG (Majcen et al., 2013b, 2015) also studied and controlled for other factors that contribute to the EPG in the Dutch residential sector. In the research by van den Brom (2020) the possible causes of the EPG were investigated as well. Each study has done multiple regression analysis on different data sets and sample sizes. The most important variables investigated in these three studies are reviewed here and can be categorized into three types: occupant behavior, occupant characteristics, and building characteristics.

Majcen et al. (2013b) studied the influence of several variables on theoretical and actual gas and electricity consumption using the same large sample as in their previous study discussed in subsection 2.2 (Majcen et al., 2013a). After supplementing it with additional databases to add more variables, the total sample size remained at about 40,000 dwellings. The methods used were descriptive statistics and multiple regression analysis. The study computed the prediction power of multiple variables on theoretical and actual gas consumption.

In a second study, the same authors investigated the correlation of similar variables on the theoretical and actual gas consumption per m^2 floor area and on the EPG per m^2 floor area, this time using two different data sets (Majcen et al., 2015). First, a data set provided by Rekenkamer Amsterdam, which after cleaning resulted in a sample size of nearly 49,000 dwellings in the Amsterdam area. Second, the data set from the Dutch Housing Survey 2012 (In Dutch: WoonOnderzoek Nederland (WoON) 2012), a sample of about 4,800 representative Dutch dwellings. Both data sets were coupled with energy consumption data from CBS. Correlation and multiple regression analysis allowed to control for variables that explain the variation in the dependent variables.

The third study investigated the EPG in a database from 2014 of the social housing sector, supplemented with data from CBS, resulting in a sample size of over 1.4 million dwellings. Gas consumption was expressed per m^2 of floor area, and different occupant characteristics were grouped into 'household categories', which were then used as dummy variables in multiple linear regression. Another analysis looked for the occupant characteristics and building characteristics that correlated with the highest and lowest 10% groups of energy usage for a high-efficiency label (B) and a low-efficiency label (E), as to show more clearly which ones correlate with higher-than-expected or lower-than-expected energy consumption.

Occupant characteristics Salary per occupant and total salary of the household was an insignificant predictor of theoretical gas consumption in the Agentschap NL database (Majcen et al., 2013b). In the WoON 2012 and Amsterdam samples, a correlation was found between theoretical gas consumption per m^2 floor area, and the amount of spendable income. This was attributed to the likeliness of higher-income households occupying better-performing dwellings, i.e., with a lower theoretical gas consumption. In addition, higher-income households have lower actual gas use per m^2 and a smaller EPG in the WoON 2012 data set (Majcen et al., 2015). In contrast, the analysis of the Agentschap NL database indicated that per €10,000 increase of annual total income, absolute actual gas consumption increases by 8 m3 (Majcen et al., 2013b).

The number of occupants was significant for actual gas consumption, but not for theoretical consumption. In the theoretical calculation, the number of occupants is estimated based on floor area and thus it is not an independent factor in the calculation (Majcen et al., 2013b). However, the analysis of the WoON database showed a negative correlation of theoretical gas use per m^2 with increasing number of occupants. In the Amsterdam sample larger number of occupants correlated with higher actual gas use per m^2, but this was not observed in the WoON 2012 sample. In the social housing database, single occupants occur more frequently in low energy-consuming groups (per unit floor area), while households of two or more members are found more often in the highest 10% gas-consuming group.

Building characteristics In the Agentschap NL sample, floor area was found to be a good predictor of both theoretical and actual gas consumption, and larger floor areas corresponded to a larger EPG—the theoretical gas consumption is higher than the actual consumption. This may be explained by the assumption made in the theoretical calculation, that the entire area of the dwelling is heated, while in practice it is less likely that in large houses, all rooms are equally heated (Majcen et al., 2013b). Analysis of the Amsterdam sample indicated that even after correcting for floor area, i.e., expressing actual energy consumption per unit floor area, this variable is a good predictor: larger dwellings have a lower natural gas consumption per m^2. However, this correlation was not observed in the WoON 2012 sample, which is more representative for the Dutch housing stock (Majcen et al., 2015).

The age of the building was also found to be a significant predictor of both theoretical and actual gas consumption per m^2 floor area: older buildings correlate with higher gas consumption, and the strength of the correlation for the theoretical consumption is double that of the actual consumption (Majcen et al., 2013b, 2015). Older dwellings also correlate with and predict a larger EPG per m^2 floor area, while controlling for some other building characteristics: the type of building and the type and efficiency of the heating installation (Majcen et al., 2015). In addition, van den Brom et al. (2017) found that, within the high-efficiency label A, older buildings occur more frequently in the group of the 10% highest actual energy consumers than newer buildings. At the same time, it is recognized that often for older buildings, less documentation on building characteristics is available. Consequently, their theoretical energy consumption (and the energy label) is more often based on estimates and assumptions, limiting its accuracy (van den Brom, 2020). This could partly explain the observed effect of building age, despite correcting for energy label or other building characteristics.

In the analysis on the Amsterdam sample and the Agentschap NL sample, the correlation of the label with the theoretical consumption is stronger than with actual consumption, and the other way around in the WoON 2012 sample. In any case, the label was a significant predictor for both actual and theoretical consumption and for the EPG (Majcen et al., 2013b, 2015). A stronger correlation of the label with theoretical consumption makes sense since actual consumption is influenced by more factors besides the label compared to theoretical consumption, e.g., occupant behavior. Within actual consumption, as the efficiency of the label is increased, the strength of the predictor increases as well (Majcen et al., 2013b). The analysis by van den Brom et al. (2017) was done per label, in other words, the label was not investigated as a characteristic itself.

The type of dwelling predicts both theoretical and actual gas consumption with similar power, suggesting little influence on the EPG (Majcen et al., 2013b). The study of the smaller samples showed that gallery apartments have the lowest theoretical and actual gas consumption, and also the lowest EPG (absolutely and relatively) (Majcen et al., 2015). In the social housing study, the distribution of housing type among the lowest and highest energy consumers was found to be significantly different both in high- and low-efficiency label categories. Specifically, single-family houses are more frequent in the high-consuming groups, while apartments occur more frequently in the lower consumption group (van den Brom et al., 2017). This can be explained by the larger building envelope of single-family houses.

More efficient installation systems of a dwelling, e.g, high-efficiency boilers, correlate with lower theoretical and actual gas consumption, as expected (Majcen et al., 2015). However, the installation type is a worse predictor of actual gas consumption than of theoretical gas consumption. In the study using the Agentschap NL database, the accuracy of this result is limited by the lack of data on hot tap water systems (Majcen et al., 2013b). Lower-efficiency installations correlate with a larger EPG, which could mean that their efficiency is underestimated in the theoretical calculation (Majcen et al., 2013b, 2015). At the same time, some installations contribute to lower energy usage despite their low efficiency, such as a gas fire, which only heats individual rooms (Majcen et al., 2013b). The social housing analysis confirmed this finding as well (van den Brom et al., 2017).

Additional building characteristics that were found significant were the number of rooms, the value of the house, and the degree of insulation. Larger number of rooms leads to a larger EPG, according to WoON 2012 data, but this was not observed in the Amsterdam sample, possibly due to poorly represented dwellings with large numbers of rooms (Majcen et al., 2015). Dwellings with a value of over 100,000 (in 2009), are found to consume higher volumes of natural gas both theoretically and in reality (Majcen et al., 2013b). Finally, insulated buildings are found more frequently in the lower 10% gas consuming social dwellings, while poorly or non-insulated buildings occur most often in the 10% highest gas-consuming group of social homes, as expected (van den Brom et al., 2017).

Occupant behavior Behavior includes factors like the setpoint temperature, number of rooms that are heated, the occupancy time of the house, etc. It refers to lifestyle and habits (Majcen et al., 2015). Previous studies have shown that the effect of occupant behavior on the EPG is complex, in part because it is related to other factors, such as climate or building characteristics (Guerra Santín, 2010). For example, the type of thermostat influences the heating time. In addition, the rebound effect, which results from (changes in) behavior, is also believed to influence the EPG (Visscher et al., 2016); the energy efficiency condition of the building influences the demand for energy services, e.g., a higher heating efficiency leads occupants to use a higher temperature setting (Majcen, 2016). While difficult to quantify, it is suggested that occupant behavior accounts for a large part of the variability in actual gas demand and is key in the explanation of the EPG (Majcen, 2016).

In conclusion, knowing that ownership matters for the size of the rebound effect means that it may also be of influence on the EPG, since the rebound effect is partly responsible for the EPG (Visscher et al., 2016). Many other factors of influence to the EPG have been studied, but the influence of ownership has not yet been clearly researched as a consequence of using data of predominantly rental dwellings. Studying the EPG in a sample with a more representative distribution of owner-occupied and rental dwellings and thereby assessing the importance of ownership type on the size of the EPG would fill this gap in the literature.

2.3 Theoretical framework

In this section, the possible relationships between variables as revealed by the literature review in subsection 2.2 are derived and the resulting conceptual model is supported by theory. The theoretical framework is the basis of and guides the subsequent research. From this theoretical framework, hypotheses are formulated as tentative answers to the research questions, which this thesis will test, and subsequently accept or reject.

2.3.1 Conceptual model

Figure 3 shows the proposed conceptual model of the relationship between ownership type and actual and theoretical energy consumption, and the EPG as a result. The relationships are explained according to the numbering of the arrows in Figure 3.

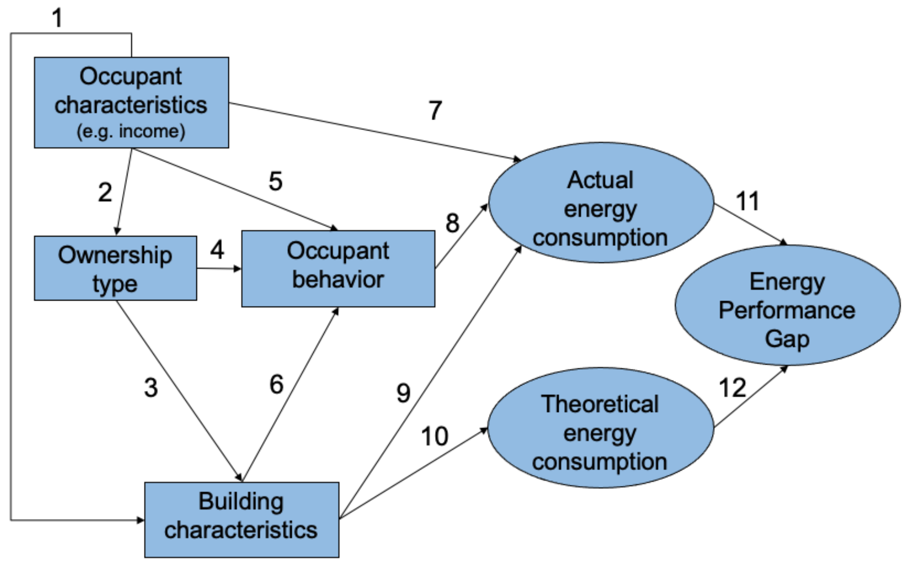


Figure 3: Diagram of the relationships between ownership type, occupant characteristics, building characteristics, and occupant behavior and the dependent variables actual energy consumption, theoretical energy consumption and energy performance gap. It is expected that ownership type affects actual energy consumption and the EPG through building characteristics and behavior.

1. Some occupant characteristics are expected to determine building characteristics. For example, the household's income may determine the size of the house, the value of the house, the type of building (such as apartment or detached house).

2. Occupant characteristics are also expected to determine ownership type. For example, income can determine whether an occupant qualifies for social housing. Another example is type of employment. For people that are employed it may be easier to get a mortgage than for freelancers, and therefore may be more likely to become homeowners. The same applies to household size: single-person households are less likely to be able to afford to buy a home compared to couples.

3. Ownership type is expected to determine building characteristics. It is expected that (social) rental houses are more often smaller in size, and more often apartments or terraced houses rather than (semi-)buildings.

4. Ownership type is also expected to influence occupant behavior. As shown in the literature review, tenants have been shown to display larger rebound effects than homeowners (Madlener and Hauertmann, 2011; Aydin et al., 2017). Since rebound effects are behavioral changes in energy consumption depending on energy efficiency

changes, it may be expected that homeowners and tenants have different energy consumption behavior. At least, the energy consumption behavior of tenants seems to be more sensitive to changes in energy efficiency than that of homeowners, hence the higher rebound effect in rental dwellings. Another way in which ownership type can influence occupant behavior is through the concept of "all-in" rent. In the Netherlands, all-in rent refers to the rent paid by tenants that includes the use of utilities and is independent of the actual consumption Government of the Netherlands (n.d.a). As a result, energy consumption is not limited or influenced by money, and tenants may consume energy more excessively without financial consequences. Evidence for this was found by Maruejols and Young (2011): Canadian households for which the landlord pays the energy bills have higher temperature settings during the day, and are less likely to turn down the thermostat when no one is home. These households were also found to have a higher energy consumption per unit floor area. Similar findings were reported by Levinson and Niemann (2004) for rental apartments in the United States.

5. Occupant behavior is also determined by other occupant characteristics. For example, income has been shown to determine behavior (Guerra Santín, 2010). Households with higher incomes tend to keep higher indoor temperatures and seem less concerned about saving energy. Also, the age of the occupant can influence the indoor temperature choice: elderly occupants tend to have higher indoor temperatures. A higher level of education of occupants was found to be related to fewer hours of heating at the highest chosen temperature (Guerra Santín, 2010).

6. Occupant behavior is determined by building characteristics. Occupants in detached houses choose lower indoor temperatures compared to apartments (Lindén et al., 2006). The type of thermostat in the house also influences the heating behavior (Guerra Santín and Itard, 2010; de Groot et al., 2008). Programmable thermostats lead to more hours of heating than manual thermostats. Moreover, the energy performance of the dwelling influences heating behavior: more insulation leads to higher indoor temperature demands. This could be interpreted as a rebound effect (Haas et al., 1998; Shipworth et al., 2010).

7. Besides through occupant behavior, occupant characteristics also influence actual energy consumption directly. For example, the number of people in the household is a household characteristic that directly influences energy use: the more occupants in the household, the higher the energy consumption. This can occur for example because each individual member uses hot water for showering. Also, more occupants may lead to more rooms needing to be heated.

8. Occupant behavior is believed to influence actual energy consumption and play a key role in the EPG (Majcen et al., 2015; Guerra Santín, 2011; Gill et al., 2010). Behavior includes factors like the setpoint temperature, number of rooms that are heated, the occupancy time of the house, etc. It refers to lifestyle and habits (Majcen et al., 2015).

9. Building characteristics also influence actual energy consumption directly. For example, the energy efficiency aspects such as insulation, as well as the building type (i.e., apartment, terraced, detached, etc.), size, type of heating installation, etc.

10. Building characteristics determine the theoretical energy consumption according to the calculation method of the energy label in ISSO publication 82.3 and described by Majcen (2016).

11. and 12. Together, theoretical energy consumption ($Q_{\text{theoretical}}$) and actual energy consumption (Q_{actual}) determine the EPG, according to Equation 1.

$$\text{EPG} = Q_{\text{theoretical}} - Q_{\text{actual}} \tag{1}$$

As can be seen in Figure 3, ownership type is not expected to have a direct (causal) effect on actual energy consumption or the EPG. Rather, it is expected to affect actual energy consumption and the EPG through building characteristics and behavior. While it is certain that there are differences in building characteristics between the ownership types, differences in behavior seem to be less known.

Differences in energy consumption behavior have been found in Canada and the United States for tenants with utility-included rent (Maruejols and Young, 2011; Levinson and Niemann, 2004). In the United States, over a quarter of rental apartments include utility costs in their rent (Levinson and Niemann, 2004). Although it exists in the Netherlands too, it is not clear how often "all-in" rent occurs (Government of the Netherlands, n.d.a). Nevertheless, all-in rent is discouraged in the Netherlands, because it prevents tenants from verifying whether the yearly rent increases are justified (Rent Tribunal, n.d.). This makes it more probable that all-in rent in the Netherlands is mostly uncommon. As a result, the effect of utility-included rent on energy consumption in the Netherlands could be limited. In any case, it is not clear to what extent—if at all—potential different behavior influences differences in actual energy use between ownership types.

Differences in the rebound effect of tenants and homeowners have been identified in Germany and the Netherlands, by Madlener and Hauertmann (2011) and Aydin et al. (2017), respectively. In the Netherlands, the rebound effect in tenants was 26.7% while in homeowners it was 41.3%. Since the rebound effect is believed to be part of the explanation for the EPG, it may be that different ownership types lead to different sizes of EPG, through differences in behavior resulting from different extents of rebound effects. Because of the (potentially) less conserving behavior of tenants and their larger rebound effects, it is expected that tenants have a higher energy consumption than homeowners and a larger EPG, when other influencing factors are corrected for.

2.3.2 Theory

Theory is needed in order to support the conceptual model and guide the research. Theory may help explain what determines consumer demand—in this case demand for natural gas and related energy services.

For this thesis, the neoclassical theory of consumer demand forms a good basis for the conceptual model. There are three important assumptions in this theory. First, the unit of analysis in this theory is the representative consumer, who makes their decisions independently from others. Second, the consumer is instrumentally rational, meaning that they seek the highest possible utility, limited by their income—the consumer is insatiable. Third, marginal utility diminishes, which means that for each additional unit of a commodity that is consumed, the added amount of utility gained is smaller. This theory is in accordance with the rebound effect: as explained in subsection 2.2, higher energy efficiency has a similar effect on the demand for an energy service as lowering its price, since less energy is needed to achieve the same energy service. The money saved from higher energy efficiency is then spent on more of that energy service or on other commodities, to continue maximizing utility and being only limited by income. At the same time, with regard to energy and energy services, it does not seem realistic that the consumer is insatiable, as described previously in subsection 2.2. There may be a maximum demand for energy services, after which utility decreases with increasing consumption. For example, a consumer may prefer a certain indoor temperature setting in their house, and any temperature higher or lower than that will decrease their utility, even if they could afford to increase the use of heating. In that case, maximum utility is not being limited by income. This is contrary to the second assumption of the neoclassical theory. Nevertheless, this theory helps to understand the relationship between income, consumer behavior, the rebound effect and energy demand, and is at the same time a reminder that in this case, other factors besides income and price likely play a role as well.

2.3.3 Hypotheses

From the theoretical framework hypotheses are developed that are tentative answers to the research sub-questions presented in subsection 1.4, and are tested and evaluated in the remainder of the report.

- *To what extent does the type of ownership—owner-occupied or rental—predict a) the actual natural gas consumption and b) the size of the EPG?*
  When building characteristics and occupant characteristics are controlled for, rental dwellings have a higher gas consumption and a higher EPG than owner-occupied dwellings.

- *What empirical variables influence the relationship between ownership and the EPG?*
  Building characteristics, such as building type and floor area, contribute to differences in ownership type as well as the EPG.

- *What is the mechanism of the relationship between ownership and the EPG*
  Because of the (potentially) less conserving behavior of tenants and their larger rebound effects, it is expected that tenants have a higher energy consumption than homeowners and a larger EPG, when other influencing factors are corrected for.

3 Data

The database used in this research is the Dutch Housing Survey 2018 (in Dutch: Woononderzoek Nederland, WoON, 2018). This database is owned by the Ministry of the Interior and Kingdom Relations (BZK) and Statistics Netherlands (CBS) (2019a,b), and is the result of a large housing market research project that is carried out every three years in the Netherlands. In this thesis, two modules of the database are used: the housing module and the energy module.

3.1 WoON 2018 housing module

The housing module is the main module of the WoON database and contains data of over 67,000 dwellings regarding occupant characteristics, occupant behaviour and preferences related to housing, and building characteristics. These data have been collected through a survey, a technical inspection of the dwelling, and from other databases. The variables used in this thesis include natural gas consumption in 2017[3], electricity consumption in 2017, floor area, definitive energy label (from RVO database), preliminary energy label (from RVO database), ownership structure, disposable income, highest level of education, type of building, number of household members, main income source, value of the house, age group of the building and the type of heating and domestic hot water (dhw) installations.

3.2 WoON 2018 energy module

The energy module consists of a sample from the housing module of 4,506 dwellings and contains additional energy-related technical and survey data on this sample. The energy module data set can be coupled to the housing module through the respondent numbers. The variables from the energy module used in this thesis are the energy index, the energy label (from technical inspection), the indoor temperature setting on weekdays and the occupancy time of the house on weekdays.

3.3 Cleaning data

The raw data sets were cleaned to prepare them for analysis, guided by previous work by Majcen (2016) and van den Brom (2020). From the housing module, cases with unrealistic floor areas were removed. For Dutch social housing cases, floor areas under 15 m^2 and above 300 m^2 were removed (van den Brom, 2020), and for non-social housing cases floor areas of over 1000 m^2 were deleted (Majcen, 2016). Also, dwellings with collective systems for heating and domestic hot water—district and block heating—were removed, as their reported energy use is considered unreliable (Majcen, 2016; van den Brom, 2020). In addition, dwellings with more than one heating installation system are deleted (Majcen, 2016), as well as dwellings lacking access to domestic hot water (only 19 cases). Also, cases with more than one household living on the same address were removed—about 3% of the total number of observations. This was done because the gas consumption is reported for the address, whereas many occupant characteristics are related only to the household responding to the survey. With a single household living at each address, the gas consumption of the address can be related to the corresponding household and its characteristics. Also,

[3] The natural gas consumption recorded in the data set is standardized, meaning that it takes into account the weather conditions of the year by being corrected for annual degree days in 2017. This allows for a fair comparison of the gas consumption between different years (van den Brom, 2020; Stuart-Fox et al., 2019).

cases with rare types of building, such as farms or dwellings with attached shop or office, were excluded for simplicity, since these represented only 3% of the total housing module. Furthermore, because many cases missed the value for the energy label, the missing value was replaced by the value reported in the variable representing the preliminary energy label. The preliminary energy label is not the definitive registered label of the building, but it is a preliminary estimate based on the technical inspection of the building performed in WoON 2018 (Stuart-Fox et al., 2019).

From the energy module, cases that indicate the inability to open windows and/or doors as a possible method to ventilate the dwelling were removed. This was done because in the documentation of the energy module it is emphasized that it is unusual that a dwelling lacks the possibility to open windows and/or doors (Cremers, 2017). Also, all cases are removed that indicate no possible method of ventilation, for the same reason.

Finally, in both modules, responses such as "unknown" or "refuses" were treated as missing data and cases with missing data were removed from the data sets. The decision to delete cases with missing data was justified by evaluating the representativeness of the resulting samples and thereby ensuring that deleting cases with missing data did not introduce bias to the samples. Specifically, the relative frequencies of different variables were compared to national data. The evaluation of the representativeness is described in subsection 3.5.
The cleaning steps resulted in a final estimation sample of the housing module containing 41,971 cases. The cleaned housing module estimation sample was coupled to the cleaned energy module data set based on the respondent number, and the resulting energy module estimation sample consisted of 2,010 observations.

3.4 Accuracy of the energy label

As indicated in Sections 3.1 and 3.2, the energy label contained in the energy module has been obtained from a technical inspection of the dwellings carried out for the purpose of the Dutch Housing Survey, and are therefore up to date and reflecting the actual energy performance of the dwelling. This is however not necessarily the case for the energy labels recorded in the housing module, because they originate from the Netherlands Enterprise Agency (in Dutch: Rijksdienst voor Ondernemend Nederland, RVO) energy label database. The RVO database does not necessarily contain up-to-date labels, because these labels are only required to be updated when a house is sold or rented. When comparing the labels from the housing module and the energy module of the dwellings in the coupled energy module sample (N = 2,010), it appears that merely 41% of the observations have matching energy labels, i.e. the label recorded in the RVO database is equal to that obtained from the technical inspection by WoON 2018. For 68% of cases that do not match, the label recorded in the RVO database is of poorer performance than the label obtained by the technical inspection. This may possibly indicate that often a house with a low-performance label in the RVO database would qualify for a higher performance label at the time of WoON 2018, for example because it has been retrofitted while not being sold or rented to a new tenant, and therefore not having to be updated in the RVO database.

In addition, as described in subsection 3.3, missing data of the energy label in the housing module, i.e., the energy labels provided by RVO, were replaced by the preliminary energy label. The estimate of the preliminary label is simplified compared to the actual energy label definition and it often deviates from the definitive label (Stuart-Fox et al., 2019), thus introducing inaccurate labels for part of sample. At the same time, using only the observations with the definitive label would introduce bias as well, because the resulting sample is not representative for the national housing stock in other aspects (Stuart-Fox et al., 2019). In addition, it would reduce the sample size.

In sum, it may be important to keep in mind that the energy labels in the housing module are often not up to date, and/or not accurately estimated.

3.5 Representativeness

Overall, the WoON database is a representative sample of the whole Dutch housing stock (Janssen-Jansen, 2019). This subsection analyses the representativeness of the cleaned estimation samples used in the analyses of this thesis: the housing module sample (N = 41,971), and the energy module sample (N = 2,010). Their representitaveness is evaluated based on a number of occupant and building characteristics.

This thesis focuses on the the relationship between ownership structure and energy consumption data and therefore the representation of the Dutch distribution of ownership structure is especially important to draw conclusions on the national level based on the analysis. The distribution of ownership structure in the housing module sample is 65% owner occupied, 27% social rental, and 9% private rental[4]. In the energy module sample the distribution is similar: 65% owner occupied, 26% social housing, 9% private rental. This distribution is not far from the distribution in the whole Dutch household population in 2018: approximately 58%, 30%, and 12%, respectively, according to Statistics Netherlands (CBS) (2018). The comparison is visualized in Figure 4.

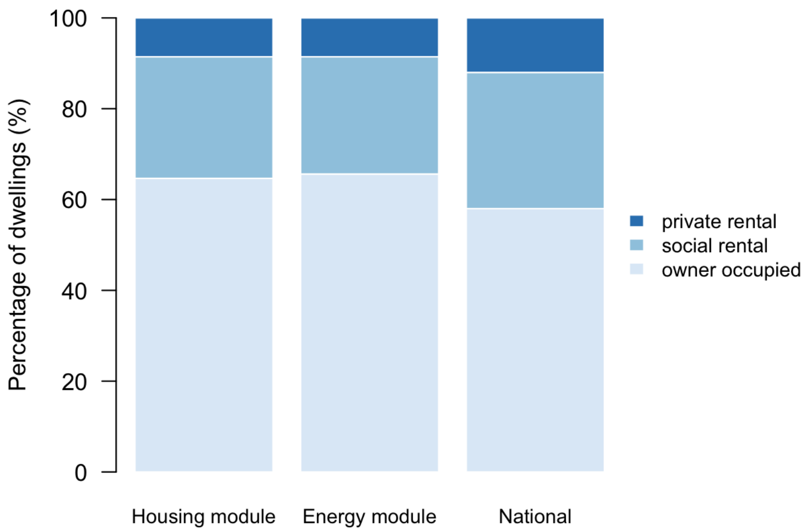


Figure 4: Comparison of the shares of dwellings with different ownership types in the housing module sample, the energy module sample, and the total Dutch housing stock (Statistics Netherlands (CBS), 2018). Both samples are representative for the national distribution of ownership structure.

[4]Percentages do not add up to 100 due to rounding.

The representativeness of the sample was also assessed based on the energy label (Figure 5). For that, cases with energy label A+ or A in the energy module were grouped together as energy label A, as done by Majcen et al. (2013a). This was done because only a few cases had a label A+ (1.4% of the energy module). The housing module did not contain any A+ labels, and neither module contained cases of A++ labels. The samples were compared to energy label data from 2018 by Environmental Data Compendium (2020). This data consists of about 640,000 houses, but the source of the energy label data is the RVO, meaning that the labels may not be up to date, as explained in subsection 3.4. As can be seen in Figure 5, the labels where the sample data deviates the most from the national data are the highest efficiency labels, A and A+. The most efficient energy labels comprise around 11% and 16% in the housing and energy sample, respectively, while in the national they represent almost 30%: double to triple the frequency as in the samples. The opposite occurs for the lowest efficiency labels, F and G, in the case of the housing module sample: the relative frequency in the housing module is approximately double that of the national data. In the energy module sample the relative frequencies of the lower efficiency labels are more similar to the national data, especially label F. The different distributions in the two samples may be explained by the different sources of the energy label in both data sets, as explained in subsection 3.4. The majority of dwellings in both samples, about 32%, have an energy label C. In the national data, about a quarter of all dwellings have this label. The remaining labels B, D and E have representative relative frequencies in both samples.

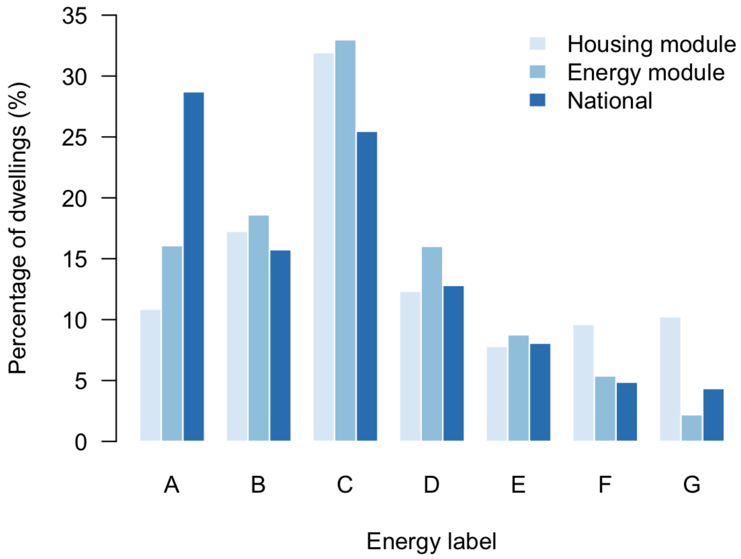
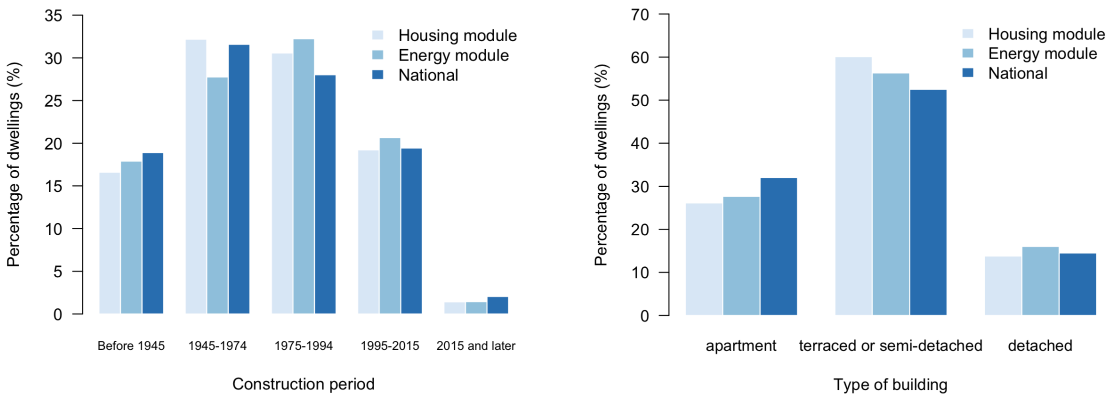


Figure 5: Comparison of the shares of energy labels in the housing module sample, the energy module sample and in the total Dutch housing stock in 2018 (Environmental Data Compendium, 2020). In both samples, high energy performance dwellings with labels A or A+ are underrepresented. Poor energy performance dwellings with labels F or G are overrepresented in the housing module, while well represented in the energy module.

Construction period (Figure 6a) and type of building (Figure 6b) were also compared to national data to assess the representativeness of the sample. Overall, both building characteristics in both samples are very similar to the national situation. As seen in Figure 6a, houses built between 1945 and 1974 are slightly more uncommon in the energy module sample than in the housing module sample and the national data, whereas houses built between 1975 and 1994 are most common in the energy module. In Figure 6b, terraced and semi-detached houses are not distinguished. This is because in the samples, mid-terraced and end-terraced houses are indistinguishable in the category "terraced", while semi-detached houses are in a separate category "semi-detached". At the same time, in the national data of CBS from 2009 presented by Majcen et al. (2013a), end-terraced houses are separated from mid-terraced houses, while semi-detached houses are indistinguishable from end-terraced houses. In any case, detached houses are well-represented in both samples. Apartments are slightly underrepresented in the samples, while terraced or semi-detached houses are slightly overrepresented.



(a) (b)

Figure 6: Comparison of the shares of households in the housing module sample, energy module sample, and total Dutch housing stock according to the period of construction (Statistics Netherlands (CBS), 2022b) (a) and the type of building (Majcen et al., 2013a) (b). The samples reflect the national distribution quite well for both characteristics.

Besides building characteristics, the representativeness of the samples based on some occupant characteristics was analysed as well. Table 1 shows the distribution of disposable income of occupants. The distributions in the two samples is similar to each other. The higher percentiles are well-represented by the samples, whereas lower incomes are underrepresented. The average disposable income in 2017 in both the housing module sample and the energy module sample are similar to each other (about €42,000 and €43,000, respectively) but higher than the national average (about €38,000).

The average number of household members in the housing module sample completely reflects the national average in 2018—2.2 (van Duin et al., 2018). However, it must be noted that in the sample, when the number of household members of 5 or higher, the reported response is 5. As a result, the average may be biased towards a lower value than in reality, as 6 or more household members are counted in the computation of the mean as 5 members as well. In any case, the response of "5 or more household members" accounts for less

Table 1: Comparison of the percentiles of disposable income (in 1000 €) in the sample and in the Netherlands in 2017 (Statistics Netherlands (CBS), 2021).

Percentile	10%	20%	30%	40%	50%	60%	70%	80%	90%
Housing module	18.0	22.7	27.1	31.8	37.0	42.8	49.6	57.6	70.5
Energy module	18.2	23.9	28.2	33.4	38.1	43.8	50.4	59.4	72.4
National	15.1	19.8	24.3	28.8	34.1	40.8	48.4	57.6	72.4

than 5% of the total responses in the sample. In the energy module sample, the average number of household members is slightly lower: 2.0. Dwellings occupied by people aged 65 or older (single or couple) occur more frequently in both the housing module sample and the energy module sample than in the national population (31%, 45% and 15%, respectively). This may be due to people aged 65+ showing higher response rates than most younger groups in the WoON 2018 research (Janssen-Jansen, 2019). The overrepresentation of occupants aged 65+ occured in previously studied databases as well van den Brom (2020).

Some of the differences between the samples and the national data could possibly be corrected for by using survey weights, which are provided in the WoON datasets. Nevertheless, it has been shown that both samples are in many aspects a good representation of the population.

4 Method

Descriptive statistics, correlation and simple and multiple linear regression were used to analyze the data sets. The statistical software and language used for this thesis were RStudio (version 2022.02.0) and R.

4.1 Correlation

To establish whether ownership type and actual gas consumption, and ownership type and the EPG are correlated, the point-biserial correlation coefficients were computed. This type of correlation coefficient is applied when one of the variables is dichotomous and the other one is measured on the interval or ratio scale. The correlation coefficients can take on a value between -1.00 and $+1.00$. A value of $\pm.00$ and $\pm.29$ is referred to as none (.00) to weak correlation, a value between $\pm.30$ and $\pm.69$ indicates a moderate correlation, and a value between $\pm.70$ and ±1.00 means a strong or perfect (1.00) correlation (Jackson, 2013). The point-biserial correlation coefficients were computed in R by the `cor.test` function.

4.2 Simple linear regression

Simple linear regression is used to model the relationship between home ownership and the dependent variables—the actual gas consumption and the EPG—in the population, i.e., the Dutch housing stock (Equation 2). Like correlation analysis, the simple regression can indicate whether ownership type has a statistically significant effect on the dependent variables. In addition, the simple regression model quantifies by how much the dependent variables are affected by the ownership types.

$$y = \beta_0 + \beta_1 owner + u \tag{2}$$

In Equation 2, the independent variable *owner* indicates whether or not the dwelling is owner occupied. The variable y represents the dependent variables actual gas consumption or the EPG. The parameter β_0 is the intercept, i.e., the value of y when the value of *owner* is equal to 0, i.e., when the dwelling is rental. The population parameter β_1 measures the change in y with respect to *owner*. The error term u accounts for the remaining variation in y. The population parameters β_0 and β_1 are estimated by the estimation method of ordinary least squares using the `lm` function in R (Heiss, 2020). The `stargazer` package is used for generating output tables (Hlavac, 2022). .

4.3 Multiple linear regression

The method for the analysis of the ceteris paribus effect of ownership type on gas consumption and the EPG is multiple linear regression, since there are multiple variables affecting the dependent variables that may correlate with the ownership type as well. Another goal of this analysis is to identify the variables that show predictive power of the actual gas consumption and the EPG. The extent to which the independent variables in the model explain the variation in the dependent variable is referred to as the goodness-of fit of the model, and it is measured by the coefficient of determination, R^2. The value of R^2 represents the fraction of the variation in y that is explained by the regressors included in the model.

4.3.1 Model and estimation method

The relationship in the population can be modeled in the form of Equation 3. The model represents the relationship between the dependent variable and the independent variables of a population, for example, the Dutch housing stock. The unit of analysis is a dwelling with a certain unique address. This means that the variables should represent properties of the dwelling, such as characteristics of the building or characteristics of the occupants of the dwelling.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_{\mathrm{k}} x_{\mathrm{k}} + u \tag{3}$$

In Equation 3, y is the explained variable and x_1 through x_{k} are the explanatory variables. β_0 is the intercept, i.e., the value of y when all explanatory variables take on a value of 0. The population parameters β_1 through β_{k} measure the change in y with respect to x_1 through x_{k}, respectively, holding other factors constant—ceteris paribus. Finally, the error term u accounts for the remaining variation in y.

This thesis is concerned with the role of ownership type in explaining the variation in the dependent variables. Thus, the model equation can be specified as in Equation 4.

$$y = \beta_0 + \beta_1 owner + \beta_2 private + \boldsymbol{x\delta} + u \tag{4}$$

In Equation 4, the variables *owner* and *private* represent the ownership type of owner occupied and private rental, whereas the category for social rental dwellings is excluded to avoid perfect collinearity—it is the reference category. The notation $\boldsymbol{x\delta}$ is the vector representing other explanatory variables besides *owner* and *private*, such as building characteristics and other occupant characteristics. The exact variables are described in Table A1 in Appendix A. The dependent variables, y, are the actual gas consumption and the EPG. Gas consumption is studied in absolute terms (m^3 natural gas per year) and relative to the floor area of the dwelling, i.e. m^3 natural gas per m^2 per year. The latter is referred to as the specific gas use, analogous to the concept of specific energy use. Specific energy use refers to the amount of energy required to achieve a unit of energy service, e.g. the number of megajoule needed to heat a m^2 of a home per year (Blok and Nieuwlaar, 2021). The advantage of this quantity is that it reflects a kind of efficiency: the lower the amount of natural gas needed for a certain amount of energy service, the higher the efficiency in terms of gas use. In addition, it corrects for the variation in floor area (van den Brom, 2020). The exact dependent variables are described in Table A1 in Appendix A as well.

The population parameters are estimated by the method of ordinary least squares (OLS) using the `lm` function in R (Heiss, 2020). The `stargazer` package is used for generating output tables (Hlavac, 2022).

4.3.2 Dummy variables

To prepare the variables for the regression analyses, the qualitative variables, i.e., nominal or ordinal variables, were first transformed into binary (dummy) variables (Wooldridge, 2012; Heiss, 2020). This was done by creating a separate, binary variable for each category of the qualitative variable. When a certain category applied to an observation, its corresponding binary variable was assigned the value 1, whereas it was assigned the value 0 when it did not apply.

For simplicity and to limit the number of variables in the analysis, some categories were not converted to a dummy variable each, but grouped together into one dummy variable.

For example, for the ordinal variable from WoON 2018 representing the construction year class of the building, the categories "1945-1959" and "1960-1969" were grouped into one dummy variable *45to69*. Table 2 shows how multiple categories were grouped into one dummy variable. In the regression analysis, one of the dummy variables for each original variable is left out to avoid perfect collinearity. This dummy variable is referred to as the reference category and is chosen to enable a certain interpretation of the regression coefficients.

Table 2: Formation of the dummy variables from the categorical variables for cases where categories were grouped together into a single dummy variable. See Table A1 in Appendix A for the variable descriptions.

	Dummy variables and reference category (ref.)	Categories in WoON 2018	Variable in WoON 2018
Housing module	*noUniversity* (ref.)	Primary school Lower secondary school (in Dutch: vmbo, havo-, vwo-onderbouw, mbo 1) Higher secondary school (in Dutch: havo, vwo, mbo 2-4) Bachelor's degree (in Dutch: hbo-, wo-bachelor)	Highest level of education of respondent or their partner
	university	Master's or doctoral degree (in Dutch: hbo-, wo-master, doctor)	
	before45 (ref.)	Before 1945	Construction period of the dwelling
	45to69	Between 1945 and 1959 Between 1960 and 1969	
	70to89	Between 1970 and 1979 Between 1980 and 1989	
	90to09	Between 1990 and 1999 Between 2000 and 2009	
	after09	In 2010 or later	
	boiler (ref.)	Gas-fueled boiler (in Dutch: CV-ketel)	Type of heating installation
	wood	Wood-fueled heating installation (in Dutch: houtgestookte CV-ketel, houtkachel, houthaard, inzethaard) Pellet-fueled heating installation (in Dutch: pellet CV-ketel, pelletkachel)	
	gasHeater	Gas furnace or gas fireplace (in Dutch: gaskachel, gashaard)	
	heatPump	Heat pump	
	otherHeat	Other type of installation	
Energy module	*away** (ref.)	Almost never at home between 9:00 and 12:00 Almost never at home between 12:00 and 15:00 Almost never at home between 15:00 and 18:00	Occupancy of the dwelling on weekdays by at least one occupant
	*home**	Almost always at home between 9:00 and 12:00 Almost always at home between 12:00 and 15:00 Almost always at home between 15:00 and 18:00	
	*depends**	Not almost never or almost always at home between 9:00 and 12:00 Not almost never or almost always at home between 12:00 and 15:00 Not almost never or almost always at home between 15:00 and 18:00	

Note: * All three categories must be TRUE simultaneously.

Besides the dummy variables, also the binary variables that already existed in the data sets were converted to zeros and ones. Assigning the values 0 and 1 is in principle arbitrary, but it allows for a more intuitive interpretation of the regression parameters.

4.4 Estimating the theoretical energy consumption

As opposed to WoON 2012 (van den Brom, 2020), WoON 2018 does not include the theoretical energy consumption of the dwellings as a variable. Therefore, estimates of the theoretical gas consumption were used in this thesis. According to the calculation method of the energy label in ISSO publication 82.3 and described by Majcen (2016), theoretical gas use, $Q_{\text{gas,t}}$, and theoretical electricity use, $Q_{\text{electricity,t}}$, are determined by long calculations including numerous variables related to building characteristics. However, in this thesis $Q_{\text{gas,t}}$ was instead approximated based on the energy index contained in the energy module.

As described by Majcen (2016), according to ISSO publication 82.3, the energy index (EI) is calculated by Equation 5.

$$\text{EI} = \frac{Q_{\text{total,t}}}{155 \cdot A_{\text{floor}} + 106 \cdot A_{\text{loss}} + 9560} \tag{5}$$

In Equation 5, A_{floor} is the floor area of the dwelling in m^2 and A_{loss} is the heat loss area in m^2 (van den Brom, 2020). $Q_{\text{total,t}}$ is the total theoretical primary energy consumption in MJ, which is the sum of the energy consumption by gas and the energy consumption by electricity (Equation 6).

$$Q_{\text{total,t}}[\text{MJ}] = Q_{\text{gas,t}}[m^3] \cdot 35.17 \left[\frac{\text{MJ}}{\text{m}^3}\right] + \frac{Q_{\text{electricity,t}}[\text{kWh}] \cdot 3.6 \left[\frac{\text{MJ}}{\text{kWh}}\right]}{0.39} \tag{6}$$

In Equation 6, the theoretical energy consumption by natural gas is calculated as the theoretical consumption of gas in m^3, $Q_{\text{gas,t}}$, multiplied by the assumed higher heating value of natural gas: 35.17 MJ/m^3. The theoretical energy consumption by electricity is calculated as the theoretical consumption of electricity in kWh, $Q_{\text{electricity,t}}$, converted to MJ, divided by the assumed efficiency of the electricity network, 0.39. To avoid long and complex calculations to determine $Q_{\text{gas,t}}$ according to the ISSO publication 82.3 (Majcen, 2016), $Q_{\text{gas,t}}$ was estimated using Equation 6 and by estimating $Q_{\text{total,t}}$, and $Q_{\text{electricity,t}}$. By combining and rearranging Equations (5) and (6), the expression for the theoretical gas consumption, $Q_{\text{gas,t}}$ becomes Equation 7.

$$Q_{\text{gas,t}} = \frac{1}{35.17} \cdot \left(\text{EI} \cdot (155 \cdot A_{\text{floor}} + 106 \cdot A_{\text{loss}} + 9560) - Q_{\text{electricity,t}} \cdot \left(\frac{3.6}{0.39}\right)\right) \tag{7}$$

For the variables in Equation 7 that were not readily available in the WoON 2018 data sets, an estimate was used. Specifically, A_{loss}, was estimated as the sum of the total surface areas of (straw) roof, ground floor, glass, facade, walls, panels and doors of each house. These surface areas are included in the energy module and are obtained from the technical inspection. Together they approximately form the heat loss area of a building, according to the description of the heat loss area by Netherlands Enterprise Agency (RVO) (2021). The mean of the resulting estimates of A_{loss} for each building type is shown in Table 3.

The other estimated variable is $Q_{\text{electricity,t}}$, which was estimated using the actual electricity consumption contained in the housing module of WoON 2018, and assuming the same ratio of theoretical to actual electricity consumption for each energy label as found by Majcen (2016). The ratios for each energy label used to estimate $Q_{\text{electricity,t}}$ are shown in Table 4, as well as the mean of the resulting estimated $Q_{\text{electricity,t}}$ per energy label. The

Table 3: Mean estimated heat loss area for each building type.

Building type	Mean A_{loss} (m^2)
Apartment	109
Terraced	203
Semi-detached	290
Detached	431

exact numbers used for calculating the ratios based on Majcen (2016) can be found in Table C1 in Appendix C.

Table 4: The ratio between actual and theoretical electricity consumption by Majcen (2016), and the resulting estimated theoretical electricity consumption, $Q_{\text{electricity,t}}$, per energy label.

Label	A	B	C	D	E	F	G
Ratio	0.45	0.40	0.36	0.34	0.34	0.56	0.40
$Q_{\text{electicity,t}}$ (kWh/year)	1255	1186	1051	1003	1019	1586	1419

With the estimated theoretical gas consumption, $Q_{\text{gas,t}}$, by Equation 7, and the actual gas consumption in the data set provided by the energy network operators, $Q_{\text{gas,a}}$, the EPG of the dwellings in the energy module was calculated by Equation 8.

$$\text{EPG} = Q_{\text{gas,t}} - Q_{\text{gas,a}} \tag{8}$$

Equation 8 implies that the EPG is positive when the theoretical gas consumption is larger than the actual gas consumption. This phenomenon is referred to as overprediction in the literature. The opposite, underprediction, occurs when the theoretical gas consumption is smaller than the actual gas consumption, and the EPG is negative (Majcen et al., 2015) (Table 5). Because over- and underprediction seem to have different causes, it may be important to distinguish them in the analyses (Majcen et al., 2015).

Table 5: Definitions of over- and underprediction.

Overprediction	$Q_{\text{gas,t}} > Q_{\text{gas,a}}$	EPG > 0
Underprediction	$Q_{\text{gas,t}} < Q_{\text{gas,a}}$	EPG < 0

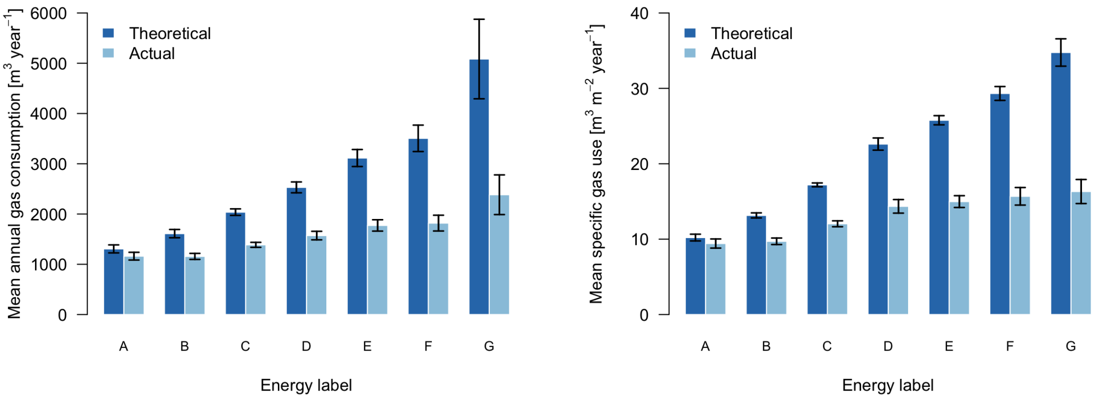
5 Results and Discussion

5.1 Descriptive statistics

Descriptive statistics are used in 5.1.1 to identify the EPG in the energy module sample and evaluate the quality of the estimated theoretical gas consumption, and in 5.1.2 to compare the different ownership types in terms of the other variables, and identify similarities and differences. Descriptive statistics on the whole housing module and energy module samples can be found in Appendix B.

5.1.1 The energy performance gap in WoON 2018 energy module

Figure 7 demonstrates the presence of energy performance gaps in the energy module sample, by comparing the mean theoretical and actual (specific) gas consumption per energy label. The theoretical gas consumption was approximated as described in subsection 4.4.



(a) (b)

Figure 7: Comparison of the mean theoretical and actual gas consumption (a) and of the mean theoretical and actual specific gas consumption (b). The theoretical gas consumption is approximated as described in subsection 4.4. Error bars represent the 95% confidence interval.

In accordance with the results by Majcen et al. (2013a), standardizing the gas consumption to the consumption per square meter floor area does not narrow the relative gaps, as can be seen when comparing Figures 7a and 7b; the patterns in the two graphs are nearly identical. This means that the sizes of the energy performance gap cannot be attributed to potentially different average dwelling sizes per energy label category, as was initially expected by Majcen et al. (2013a). A comparison of mean floor area per label category can be found in Appendix B.

As previously found by Majcen et al. (2013a) as well, the variation in actual gas consumption over the different energy labels is much smaller than the variation of the theoretical gas consumption. Furthermore, a similar pattern with regard to the EPG as reported by Majcen et al. (2013a) and shown in Figures 1 and 2 can be recognized: at low-performance energy labels, the EPG is the largest and most positive and increasingly narrows towards higher-performance labels. However, Majcen et al. (2013a) found that around energy label C, the pattern reverses and the average EPG becomes negative, i.e., the mean actual gas consumption exceeds the theoretical gas consumption. This is not the case in Figure 7, where the mean actual gas consumption is not exceeded by theoretical gas consumption in any label. This may possibly indicate that the theoretical gas consumption is overestimated by the approach in subsection 4.4 through the use of the approximations. For example, the estimated heat loss area could have been systematically overestimated if the way it was defined is incorrect. Nevertheless, it may also be the consequence of studying different populations, since the sample used by Majcen et al. (2013a) consisted primarily of social housing. This explanation is supported by the findings of Tigchelaar et al. (2011). Tigchelaar et al. (2011) analyzed the WoON 2006 database (approximately 4,700 households) and showed that the mean ratio of actual to theoretical specific gas use was 0.88 for the highest efficiency label, A, and gradually decreased per energy label towards 0.53 for the lowest efficiency label, G. The mean ratios of actual to theoretical specific gas use found by Tigchelaar et al. (2011) are in fact very similar to what is found in Figure 7. Since WoON 2006 is considered a representative data set for the Dutch housing stock (although for the year 2006), like WoON 2018, it is possible that the difference between Figures 7a and 7b and Figures 1 and 2 are caused by the different composition of the samples, and that the estimates of the theoretical gas consumption according to subsection 4.4 are in fact close to the real theoretical values.

5.1.2 Comparison of the ownership types

The comparison of owner-occupied dwellings, private rental dwellings and social rental dwellings with respect to each variable is useful to identify differences or similarities in the gas use and the EPG and to understand where these potential differences may come from.

Continuous variables Tables 6 and 7 contain the mean and standard deviation of the continuous variables per ownership type, for the housing and energy module samples, respectively. In both samples, the average actual yearly gas use is largest in owner-occupied dwellings, lower in private rental dwellings, and the lowest in social rental dwellings. This is expected when comparing the average dwelling size, which is the largest for owner-occupied dwellings and smallest for social rental dwellings. Whereas actual gas use increases with floor area, specific actual gas use, i.e. gas use per unit floor area, decreases with increasing floor area. This is believed to occur because a smaller percentage of the floor area is heated in large homes compared to smaller homes, for example by heating less rooms (Majcen et al., 2013b).

As expected, average disposable income and value of the house are highest for owner-occupied dwellings and smallest for social housing.

Table 7 corresponding to the energy module sample contains additional variables related to energy. The average energy index is similar for all ownership types, while the average estimated heat loss area of owner-occupied dwellings is substantially larger than rental dwellings. Like actual gas use, the average theoretical gas use is also largest for owners-occupied dwellings and smallest in social housing, although in all three cases it is substantially higher than the actual gas use. Consequently, the mean EPG for all ownership types is positive, meaning that on average gas use is overpredicted, as could also be seen in Figure 7a. However, the size of the EPG does differ per ownership type. It is the most overpredicted in owner-occupied dwellings and least overpredicted in social rental dwellings, both absolutely and relative to average actual gas use.

The average temperature settings during the day on weekdays between 9:00 and 18:00 and during the evening and night on weekdays between 18:00 and 9:00 are similar for all ownership types. On average, temperature setting behavior seems therefore similar in homeowners and tenants, although the standard deviation is larger in rental dwellings. This means that the temperature settings in rental dwellings are spread over a wider range around the mean than in owner-occupied dwellings, where the values are closer to the mean, i.e., among rental dwellings the temperature setting varies more.

Table 6: Summary statistics of the continuous variables within each ownership type category (housing module sample). See Table A1 for variable definitions and units.

	Owner occupied		Private rental		Social rental	
Variable	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
gasA	1566.9	800.2	1247.3	770.7	1043.2	491.9
gasSpecA	12.2	6.0	14.5	10.5	12.4	6.3
floorArea	135.4	60.6	97.5	53.6	87.5	25.0
dispInc	5.1	3.2	3.2	2.9	2.6	1.2
value	27.2	14.2	20.2	12.4	15.7	5.7

Number of observations = 41,971

Table 7: Summary statistics of the continuous variables within each ownership type category (energy module sample). See Table A1 for variable definitions and units.

	Owner occupied		Private rental		Social rental	
Variable	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
gasA	1590.0	791.2	1233.1	691.0	1035.8	497.8
gasSpecA	11.9	6.0	13.0	7.5	12.4	6.1
floorArea	140.2	54.8	101.7	44.4	87.0	25.6
dispInc	5.2	2.5	3.5	3.5	2.5	1.1
value	29.7	13.8	20.7	10.4	16.3	5.9
EI	1.6	0.5	1.7	0.5	1.6	0.4
ALoss	270.7	142.0	156.5	90.5	136.8	60.6
gasT	2460.3	1333.7	1802.5	1027.4	1513.1	638.9
gasSpecT	18.0	7.8	18.3	7.6	17.8	6.7
EPG	870.2	1006.7	569.4	922.3	477.3	586.3
tempDay	19.0	2.0	19.1	2.8	19.1	2.7
tempNight	18.3	1.5	18.0	2.9	18.1	2.5

Number of observations = 2,010

Categorical variables Tables 8 and 9 show how the categories of the categorical variables—i.e. the dummy variables—are distributed within each ownership type, in the housing and energy module, respectively.

In terms of household size, single-person households occur relatively more frequently in social housing (51.5% of social housing in the housing module consists of single-person households) than in private rental dwellings (44.2% of private rental dwellings in the housing module), and least relatively frequently in owner-occupied dwellings (21.9% of owner-occupied dwellings in the housing module). Four-person households are more common in owner-occupied dwellings than in rental dwellings.

Regarding energy labels in the energy module sample, poorest energy performance labels E through G are relatively the most frequent in private rental houses and the least relatively frequent in social housing. The relative frequency of highest energy performance labels A and B is similar in owner-occupied dwellings and social housing, and lower in private rental dwellings. An explanation for this could be that homeowners and housing associations have more incentive to increase the energy performance of their properties than private owners of rental properties. In owner-occupied dwellings, the homeowners benefit from the improvements themselves, in addition to adding value to the house. On the other hand, housing associations are pushed by the government to improve the energy performance of social housing, and need to comply with agreements regarding energy efficiency (House of Representatives of The Netherlands, 2008). In contrast, private owners of rental properties may have less interest in improving the label of their properties since they do not benefit from the improvements directly.

The distributions with respect to apartments and (semi-)detached houses are very different in owner-occupied houses and rental dwellings. Whereas more than 40% of owner-occupied dwellings in the housing module are either semi-detached or detached houses and only 13% are apartments, nearly 60% of private rental dwellings and 74% of social housing are apartments and only a few percent of rental dwellings are (semi-)detached houses.

Also construction periods are differently distributed within the different ownership types. Nearly 70% of social housing was built between 1945 and 1989. Private rental properties were most frequently built before 1945. At the same time, private rental dwellings were also relatively more often built after 2010 than owner-occupied dwellings or social housing.

The distribution of occupants with or without university education is similar in owner-occupied and private rental dwellings, and different to social housing: over 20% of owner-occupied and private rental dwellings are occupied by at least one university-educated occupant. This is only the case for less than 6% of social rental dwellings. This may be explained by the expectation that people without university education may have a lower income and therefore be more likely to occupy a social rental dwelling.

The relative frequencies of the main sources of income of the household also differ depending on the ownership type. The main source of income of the majority of households in owner-occupied dwellings, nearly 66% of the owner-occupied dwellings in the housing sample, is employment. Only about 4% of this group receives state benefits as the main source of income. In contrast, more than 20% of social housing occupants receive state benefits an less than 40% receives income from employment.

Regarding heating and dhw installations, owner-occupied dwellings and social housing are very similar: in the housing module, more than 95% have a boiler as a heating installation and more than 92% have a boiler as a dhw installation. Private rental dwellings more often have a gas furnace or gas fireplace as a heating installation and a geyser or electric boiler for dhw.

As can be seen in Table 9, less than 5% of respondents in social rental dwellings indicate that there is usually no one at home on weekdays between approximately 9:00 and 18:00, whereas in owner-occupied dwellings and private dwellings this answer is relatively more common: about 8%. This could perhaps be related to the higher relative frequencies of employment in owner-occupied and rental dwellings, compared to social housing.

Table 8: Distribution of the dummy variables within each ownership type (housing module sample). See Table A1 for variable definitions.

Category	Owner occupied (%)	Private rental (%)	Social rental (%)
n1	21.9	44.2	51.5
n2	41.9	35.3	31.6
n3	13.6	10.7	9.0
n4	16.7	6.3	5.3
n5more	5.9	3.5	2.6
A	10.8	12.0	10.7
B	17.5	11.5	18.4
C	31.4	21.9	36.3
D	10.6	9.9	17.3
E	6.6	10.9	9.8
F	11.8	8.9	4.6
G	11.3	24.9	3.0
apartment	13.0	58.9	47.3
terraced	46.1	29.0	49.6
semiDetached	20.5	5.3	2.9
detached	20.4	6.7	0.2
before1945	18.0	34.2	7.6
45to69	19.5	17.4	31.8
70to89	33.0	24.4	37.7
90to09	25.3	16.7	18.1
after10	4.1	7.3	4.9
noUniversity	78.7	79.6	94.3
university	21.3	20.4	5.7
noIncome	0.4	1.8	0.7
employed	65.8	57.8	38.6
benefits	3.9	13.6	21.2
retired	29.9	26.8	39.5
boiler	96.3	90.7	95.7
wood	0.5	0.3	0.1
gasHeater	1.6	6.3	2.4
heatPump	1.0	1.6	1.0
otherHeat	0.7	1.1	0.8
dhwGasBoiler	92.7	86.2	92.2
dhwElecBoiler	1.7	5.0	2.0
dhwGeyser	2.6	6.3	2.9
dhwSolar	1.8	0.9	1.6
dhwHeatPump	0.9	1.1	0.9
dhwOther	0.3	0.4	0.4

Number of observations = 41,971

Table 9: Distribution of the dummy variables within each ownership type (energy module sample). See Table A1 for variable definitions.

Category	Owner occupied (%)	Private rental (%)	Social rental (%)
n1	20.8	42.4	55.9
n2	53.8	41.9	31.2
n3	10.6	9.9	6.9
n4	11.4	2.3	2.7
n5more	3.3	3.5	3.3
A	16.8	15.1	14.6
B	18.1	17.4	20.2
C	32.6	27.3	35.8
D	16.1	16.3	15.8
E	8.5	12.2	8.3
F	5.5	8.1	4.0
G	2.4	3.5	1.2
apartment	15.8	57.0	48.2
terraced	40.0	34.3	48.9
semiDetached	20.3	5.8	2.7
detached	24.0	2.9	0.2
before1945	19.4	27.9	10.8
45to69	16.5	14.0	27.2
70to89	32.4	29.1	37.6
90to09	27.4	21.5	19.1
after10	4.2	7.6	5.4
noUniversity	68.2	74.4	91.5
university	31.8	25.6	8.5
noIncome	0.2	2.9	0.6
employed	52.5	41.3	26.8
benefits	3.6	7.6	24.1
retired	43.8	48.3	48.6
boiler	97.9	95.3	97.9
wood	0.2	0.0	0.4
gasHeater	0.4	1.7	0.4
heatPump	0.9	2.9	1.2
otherHeat	0.7	0.0	0.2
dhwGasBoiler	94.0	90.7	95.4
dhwElecBoiler	0.8	1.2	1.2
dhwGeyser	1.4	2.9	0.4
dhwSolar	2.5	2.3	1.9
dhwHeatPump	0.8	2.9	1.0
dhwOther	0.5	0.0	0.2
away	8.3	8.1	4.8
home	42.3	30.2	37.6
depends	49.4	61.6	57.6

Number of observations = 2,010

5.2 Correlation of actual gas use and the EPG with home ownership

The previous section showed that owner-occupied dwellings consume on average more gas than rental dwellings, and that the average EPG is also more positive in owner-occupied dwellings than in rental dwellings, both absolutely and relatively (Tables 6 and 7). To further quantify the potential correlation between gas use and EPG and ownership, correlation analysis and simple linear regression were used.

Point-biserial correlation coefficients were computed to assess the linear relationship between home ownership (versus tenancy) and actual gas consumption and between home ownership and the EPG, using the dummy variable *owner* and the continuous variables *gasA* and *EPG*. As mentioned in subsection 4.4, the phenomena of overprediction (EPG $>$ 0) and underprediction (EPG $<$ 0) can be distinguished and should be analyzed separately. For that, the energy module was divided into cases of overprediction and underprediction, leading to two separate subsamples.

There is a moderate positive correlation between owner-occupied dwellings and actual gas consumption in the energy module sample, $r(2{,}008) = .32$, $p < .001$ and a similar result was obtained in the housing module sample, $r(41{,}969) = .30$, $p < .001$. This result means that—without controlling for any other factors—owner-occupied dwellings have a larger actual gas consumption than rental dwellings. A weak positive correlation was found between owner-occupied dwellings and the EPG in case of overprediction, $r(1{,}707) = .20$, $p < .001$, meaning that the overprediction of gas consumption in owner-occupied dwellings is larger than in rental dwellings. In contrast, no significant correlation was found between owner-occupied dwellings and the EPG in case of underprediction, $r(299) = -.05$, $p = .43$. In this case, the EPG is neither significantly smaller or larger for owner-occupied dwellings than for rental dwellings. The lack of a correlation with the EPG in case of underprediction, in contrast to a positive correlation in case of overprediction, may be explained by previous findings by Majcen et al. (2015). They show that overprediction and underprediction are two different phenomena with very different explanatory factors. Specifically, the size of the EPG in case of overprediction is largely explained by building characteristics, whereas in case of underprediction building characteristics are insignificant for predicting the size of the EPG, and behavioral factors play the main role (Majcen et al., 2015). Therefore, the positive correlation of ownership with the EPG in case of overprediction and no correlation in case of underprediction may suggest that building characteristics, such as building type or floor area (Tables 6 and 7), rather than behavioral aspects are most important in determining the differences in actual gas use and EPG observed between homeowners and tenants.

A simple linear regression analysis was done on the dependent variables *gasA* and *EPG*, with the dummy variable *owner* as the single regressor. The simple regression shows similar information to the correlation analysis, but in addition it shows how much more gas is consumed by owner-occupied dwellings compared to rental dwellings, and how much larger the EPG is in owner-occupied dwellings in case of overprediction compared to rental dwellings. As shown in Table 10, owner-occupied dwellings consume significantly more gas than rental dwellings (social or private). Specifically, owner-occupied dwellings consume 474 m^3 per year more than rental dwellings in the housing module sample and 505 m^3 more in the energy module sample. Owner-occupied dwellings have a significantly larger EPG in case of overprediction than rental dwellings, by 350 m^3 per year. These differences are quite substantial, given that the average actual gas consumption in both samples was about 1,400 m^3 per year (Tables B1 and B2), and the average EPG in cases of overprediction was about 940 m^3 per year (Table B3). As shown by the correlation analysis as well, there is no significant difference in EPG in case of underprediction between owner-occupied and rental dwellings.

In sum, the correlation and simple regression analyses showed a moderate correlation between ownership and actual gas use and a weak correlation between ownership and the EPG in case of overprediction; owner-occupied dwellings consume substantially more gas than rental dwellings and also exhibit substantially larger overpredictions. However, these analyses are not determining causation. In fact, ownership type is not expected to directly cause a difference in actual gas consumption or the EPG. Yet, these results could suggest that the correlation between ownership type and actual gas use and between ownership type and the EPG in case of overprediction is mainly caused by differences in building characteristics between the types of ownership, rather than by potential differences in behavior in homeowners and tenants. This is because underpredictions, for which no correlation is found, are less determined by building characteristics, and mostly by occupant behavior (Majcen et al., 2015). In contrast, overpredictions, for which a weak correlation is found, are mostly determined by building characteristics (Majcen et al., 2015). The potential importance of building characteristics is also indicated by the results of subsection 5.1. Descriptive statistics showed that, for example, owner-occupied dwellings are much more often semi-detached or detached than rental dwellings, and owner-occupied dwellings are on average larger than rental dwellings.

Table 10: Simple regression coefficients of the actual gas consumption and the EPG with home ownership. The standard errors are given in parentheses. See Appendix A for variable definitions and units.

	Dependent variable:			
	gasA	*gasA*	*EPG* Overpredictions	*EPG* Underpredictions
owner	474.318***	505.123***	350.125***	−52.346
	(7.450)	(33.797)	(42.041)	(66.687)
Constant	1,092.619***	1,084.902***	704.378***	−348.500***
	(5.990)	(27.378)	(34.517)	(49.673)
Observations	41,971	2,010	1,709	301
R^2	0.088	0.100	0.039	0.002
Adjusted R^2	0.088	0.100	0.038	−0.001
Note:				***p<0.01

5.3 Explaining the correlation with multiple regression

In subsection 5.2 it was shown that there is a correlation between ownership and gas use and between ownership and EPG in case of overprediction. Tables 6 to 9 in subsection 5.1 revealed differences between the ownership types that are likely to explain at least part of this correlation. This subsection aims to answer the question: do the building and occupant characteristics fully explain the correlation between ownership and gas use, and ownership and EPG, or could differences in behavior between ownership types play an important explanatory role? In order to answer this question, multiple linear regression was applied to control for building characteristics, which were expected to cause (part of) the correlation between ownership type and actual gas use or the EPG. In addition, other variables that were expected to influence the actual gas consumption directly were controlled for as well, namely occupant characteristics such as disposable income, the number of occupants, whether occupants are usually away or at home during the day, etc. This was previously depicted in the conceptual model (Figure 3).

In 5.3.1 and 5.3.2, the multiple regressions on actual (specific) gas use using the housing module sample and the energy module sample, respectively, are discussed. In 5.3.3, the EPG is used as the dependent variable for the multiple regression analysis. As described in section 3, the energy module sample (N = 2,010) is a subset of the housing module sample (N = 41,971) containing several additional energy-related variables, such as the estimated theoretical gas consumption and the EPG. Besides the sample size and the number of variables, another important difference between the two samples is the source and accuracy of the energy labels, as explained in subsection 3.4. Because the variable of actual gas consumption is available in the housing module sample, it is also in the energy module sample and therefore, the multiple regression analysis on actual gas use was done using both samples.

5.3.1 Actual energy consumption (housing module sample)

The interpretation of the regression coefficients shown in Table 11 can be understood when looking at the corresponding regression equation of *gasA* (Equation 9). For the categorical variables, the interpretation of a regression coefficient is the amount of additional gas use of a dwelling in that category compared to a dwelling in the reference category, ceteris paribus. For example, according to Table 11, a dwelling with energy label G per year consumes about 442 m^3 of gas more than a dwelling with label A (reference category), in the case that all the other variables are fixed, i.e., equal floor area, the same period of construction, etc. For the continuous variables, the coefficient is the amount of additional yearly gas use per unit of the continuous variable, e.g., a dwelling consumes per year about 2.6 m^3 more gas for each additional square meter of floor area, ceteris paribus.

$$gasA = 330.460 + 24.442 \cdot owner + 78.169 \cdot private + 45.749 \cdot n2 + ... + 169.369 \cdot dhwOther \quad (9)$$

Ownership type In terms of actual gas consumption in m^3, owner-occupied dwellings consume significantly more gas than social rental dwellings, even with building characteristics and other occupant characteristics held fixed. This may suggest that the correlation found in subsection 5.2 is being caused by more than the building and occupant characteristics controlled for in the regression. However, as indicated previously, large sample sizes can lead even very small differences to be statistically significant that may not be meaningful in practice. Owner-occupied dwellings consume only about 24 m^3 more gas per year than social housing. In terms of practical significance, this is a relatively small difference given the average natural gas use in 2017 of approximately 1,400 m^3 (Table B1). This result is in disagreement with previous multiple linear regression results by Majcen

Table 11: Estimates of the regression coefficients by OLS on the actual gas use in the housing module sample. The standard errors are given in parentheses. See Table A1 in Appendix A for variable definitions, units and reference categories.

	Dependent variable:	
	gasA	*gasSpecA*
owner	24.442^{***} (7.848)	-0.024 (0.080)
private	78.169^{***} (11.380)	1.640^{***} (0.117)
n2	45.749^{***} (7.165)	0.159^{**} (0.073)
n3	125.434^{***} (10.328)	0.841^{***} (0.106)
n4	142.888^{***} (10.589)	0.967^{***} (0.109)
n5more	224.531^{***} (14.594)	1.544^{***} (0.150)
university	13.306^{*} (7.912)	-0.004 (0.081)
employed	-99.600^{***} (37.300)	-0.896^{**} (0.382)
benefits	-31.513 (38.081)	0.154 (0.390)
retired	18.363 (37.400)	-0.343 (0.383)
dispInc	16.380^{***} (1.171)	0.074^{***} (0.012)
floorArea	2.619^{***} (0.068)	-0.045^{***} (0.001)
value	7.994^{***} (0.294)	0.029^{***} (0.003)
B	102.805^{***} (12.634)	0.993^{***} (0.130)
C	164.833^{***} (13.858)	1.346^{***} (0.142)
D	316.390^{***} (15.895)	2.670^{***} (0.163)
E	315.063^{***} (17.805)	2.896^{***} (0.183)
F	383.920^{***} (17.855)	3.191^{***} (0.183)
G	441.898^{***} (19.831)	3.796^{***} (0.203)
terraced	213.025^{***} (7.849)	0.912^{***} (0.080)
semiDetached	427.895^{***} (10.719)	2.403^{***} (0.110)
detached	783.695^{***} (12.136)	4.912^{***} (0.124)
45to69	71.026^{***} (12.100)	0.065 (0.124)
70to89	41.844^{***} (13.403)	-0.747^{***} (0.137)
90to09	-176.256^{***} (15.475)	-2.523^{***} (0.159)
after10	-152.806^{***} (21.758)	-2.445^{***} (0.223)
wood	-342.538^{***} (46.155)	-1.934^{***} (0.473)
gasHeater	-203.775^{***} (25.009)	-1.495^{***} (0.256)
heatPump	-26.062 (52.079)	0.759 (0.534)
otherHeat	-39.290 (35.736)	-0.365 (0.366)
dhwElecBoiler	92.077^{***} (20.984)	1.369^{***} (0.215)
dhwGeyser	158.770^{***} (21.121)	1.540^{***} (0.216)
dhwSolar	-49.331^{**} (22.026)	-0.102 (0.226)
dhwHeatPump	35.962 (55.841)	1.088^{*} (0.572)
dhwOther	169.369^{***} (50.514)	1.398^{***} (0.518)
Constant	330.460^{***} (41.640)	14.550^{***} (0.427)
Observations	41,971	41,971
R^2	0.447	0.219
Adjusted R^2	0.447	0.219
Residual Std. Error (df = 41935)	568.331	5.826
F Statistic (df = 35; 41935)	969.063^{***}	336.361^{***}
Note:		*p<0.1; **p<0.05; ***p<0.01

et al. (2013b). They found that owner-occupied dwellings consume significantly—though relatively little—less gas than social rental dwellings by about 49 m^3 per year. Despite controlling for the energy label in the multiple regression, the authors suggest that this result may be explained by "better insulation in owner-occupied dwellings" (Majcen et al., 2013b, p. 466). Alternatively, "different behavior" is suggested as a possible explanation (Majcen et al., 2013b, p. 466). The discrepancy between the results by Majcen et al. (2013b) and Table 11 regarding ownership could be due to the overrepresentation of social housing in the sample studied by Majcen et al. (2013b).

Private rental dwellings consume significantly more gas than owner-occupied and social rental dwellings, while holding the other variables fixed. The difference of about 78 m^3 gas per year between private rental dwellings and social housing is somewhat more substantial, but still relatively small. The same applies to the difference of about 54 m^3 of gas consumption per year between private rental and owner-occupied dwellings.

A possible explanation for private rental dwellings having a higher gas consumption ceteris paribus than owner-occupied dwellings is that sometimes the rent paid by tenants includes the use of utilities and is independent of the actual consumption. This is referred to in the Netherlands as "all-in" rent (Government of the Netherlands, n.d.a). As a result, energy consumption is not limited or influenced by money. This might cause differences in energy consumption behavior between tenants of private rental properties and homeowners, leading to a higher energy use by tenants. This explanation is supported by previous studies. Maruejols and Young (2011) found that Canadian households for which the landlord pays the energy bills have a higher specific energy use than those who pay their own energy bills. In addition, they found differences in occupant behavior. Households that do not directly pay for heat have higher temperature settings during the day, and are less likely to turn down the thermostat when no one is home. Another study compared the energy consumption of utility-included rental apartments and metered rental apartments in the United States, and found that the energy consumption in heat-included apartments is higher, ceteris paribus (Levinson and Niemann, 2004). In the United States, over a quarter of rental apartments include utility costs in their rent (Levinson and Niemann, 2004). It is unclear how common the all-in rent is in the Netherlands. Given the weak effect, it is possible that utility-included rent in the Netherlands is not as common, thereby limiting its effect on the results in this sample. This is supported by the fact that all-in rent is discouraged in the Netherlands, because it prevents tenants from verifying whether the yearly rent increases are justified (Rent Tribunal, n.d.). Yet, this explanation would also support the results by Majcen et al. (2013b). They found that private rental dwellings were insignificant compared to social housing, i.e., there is no difference in gas use between social or private rental dwellings, and both types of rental dwellings consume more gas than owner-occupied dwellings.

However, utility-included rent may not explain why private rental dwellings consume more gas than social rental dwellings. Still, there could be certain behavior that differs between the social housing occupants and the private rental occupants. The underlying reason for the possible difference in behavior could be related to income. As seen in Table 6, the average disposable income in private rental dwellings is more than 20% higher than in social rental dwellings. The higher average income in private rental dwellings is expected: private rental dwellings are often more expensive than social housing, because they lack the rent price limit that social rent has. Perhaps, the lower income of social rental occupants causes more frugal behavior in energy use. Alternatively, there may be certain building characteristics private rental dwellings have in common, that lead to a higher gas use, that are not captured in the energy label or by the other building characteristics included in the regression model. These could perhaps be related to the difference in ownership of social and private rental houses: social housing is mostly owned by housing associations, whereas private rental dwellings are owned by private owners. Private owners may feel less responsibility for the energy efficiency of their properties than housing associations, which have been pushed by the government to improve the energy performance of their properties (House of Representatives of The Netherlands, 2008).

In terms of actual specific gas use, owner-occupied dwellings and social housing are not significantly different. This is in accordance with the results by Majcen et al. (2015), who showed that ownership type is insignificant for actual gas consumption per m^2. However, private rental specific gas use is significantly higher than in social housing, i.e. private rental dwellings consume more gas per m^2 than social dwellings and owner-occupied dwellings, while all other factors are constant. Nevertheless, the difference is relatively modest: 1.64 m^3/m^2 relative to the average 12.4 m^3/m^2 (Table B1). The larger specific gas use of private rental dwellings may have similar explanations as to why the absolute gas use in m^3 is also larger in private rental dwellings, namely behavioral differences caused by the payment of utility-included rent, behavioral differences caused by differences in income, or differences in building characteristics that have not been captured in the control variables.

Household size With increasing number of occupants and all other factors constant, both absolute and specific gas use increase. This is expected as larger households may use more dhw and/or heat more rooms. However, the size of the differences is relatively modest. According to Milieu Centraal (n.d.a), the difference in yearly gas use between a single-person household and a household of two or more members ranges between 140 m^3 and 200 m^3, depending on the type of building (apartment or mid-terraced), age category (old or new) and size category (small, medium or large) of the building. Possibly, the smaller effects can be explained by the ceteris paribus interpretation of the regression coefficients. The coefficients represent the unique contribution of the different household sizes to the gas consumption. In contrast, the expected figures could contain the contribution of other, related factors, e.g., the actual floor area. This could explain why the expected difference in gas use depending on the number of occupants is larger than the corresponding regression coefficients.

Employment, income and education Households for which the main source of income is employment consume significantly less gas (nearly 100 m^3 per year or 0.9 m^3/m^2 per year) than people with no source of income (i.e., people with income exclusively from properties or allowances). An explanation for this could be that employed people are more likely to spend less time at home than unemployed people, and therefore need less heating. Other main sources of income, state benefits or pension, are insignificant. The household's disposable income is significant for the (specific) gas use of the house, with all other factors equal. Per 10,000 EUR/year of additional disposable income, the gas use increases by approximately 16 m^3, and the specific gas use by 0.071 m^3/m^2. Both these coefficients are relatively very small compared to the average (specific) gas use in the sample (Table B1) and considering the units of the disposable income. While statistically significant, this variable could be considered practically insignificant. Whether the respondent or their partner has a university education or not is significant at a significance level of 0.10. Considering the large sample size, this is considered insignificant.

Building size and value Both floor area and the value of the house are statistically significant. Each additional square meter of floor area adds 2.6 m^3 of yearly gas consumption, with all other factors constant. Each additional 10,000 EUR of value of the house increases gas use with approximately 8 m^3 per year. Both these effects are also quite small.

Energy label As expected, absolute and specific gas use increase considerably with lower-performance energy labels. A dwelling labeled B—second-best energy performance—uses approximately 103 m^3 more gas per year than a dwelling labeled A, and a dwelling labeled G consumes nearly 442 m^3 more than a dwelling labeled A, while other factors are constant. An exception in the increasing pattern can be observed for label E, whose coefficient for the absolute gas use is similar to that of label D. However, in the case of specific gas use, the coefficient of label E does lie between those of labels D and F, as expected. Therefore, the inconsistency in the case of actual gas use may be related to the fact that the average floor area of label E dwellings in the sample is surprisingly small, as can be seen in Figure B1a. This may also explain why the average actual gas use in label E dwellings is lower compared to more energy-efficient dwellings (Table B5).

Type of building and construction year For the type of building, the results are also according to the expectations. Apartments (reference category) have the lowest gas use, likely because they are surrounded by other dwellings, limiting their heat loss. Terraced houses are only partially attached to other, similar dwellings, and have a higher gas consumption than apartments. Semi-detached houses have an even higher gas consumption, and detached houses have a yearly gas consumption of 784 m^3 higher compared to apartments. This is a considerable amount given than the average yearly gas consumption of the sample is approximately 1,400 m^3 (Table B1). The same can be said about the specific gas consumption, which is 4.9 m^3/m^2 higher per year for detached homes compared to apartments, while the average specific gas use in the sample is 12.4 m^3/m^2 per year (Table B1).

The period of construction is significant for the absolute gas consumption. Buildings built after 1989 have a substantially lower gas use than older buildings. The lower gas consumption of newer buildings must be explained by other factors than differences in energy label or any other control variable used in the analysis, because of the ceteris paribus interpretation of the multiple regression model. However, according to van den Brom (2020), older buildings are often less well documented regarding building characteristics than newer buildings. As a result, their energy labels rely on estimates and assumptions and are therefore more likely to be less accurate than those of newer buildings. As a consequence of the possible inaccuracy in the labels of older buildings, the regression coefficients may be influenced as well. Alternatively, the significance of the building year may perhaps be caused by building characteristics related to the construction year that are not taken into account in the energy label. These characteristics are then not controlled for in the regression analysis, such that they may influence the gas consumption. However, buildings built between 1945 and 1989 have a slightly higher gas consumption than older houses built before 1945. An explanation for this is that the energy labels in the housing module sample may be outdated, since they are only updated when a house is sold or rented (subsection 3.4). This means that in some cases, older, poor-performance buildings may have been retrofitted without their energy label being improved. Consequently, the old buildings from before 1945 could in reality have a better label than the regression analysis controls for. As a result, the regression analysis finds that buildings built before 1945 consume less gas than buildings built after 1945 and until 1989, ceteris paribus.

Heating and dhw installation Only some space heating and dhw installations are significant compared to a gas-fueled boiler. As expected, wood- and pellet-fueled heating installations result in a considerably lower gas consumption. This is probably because space heating is the most important use for natural gas in Dutch households.

Surprisingly, heat pumps as heating and/or dhw installation are not significant for (specific) gas use. This could be caused by the relatively small number of subjects in these categories, 433 and 374, respectively (Table B5). This can lead to less precise estimates (Wooldridge, 2012). An alternative explanation is that perhaps, dwellings with heat pumps often had the heat pump installed during 2017 or even in 2018, while the actual gas consumption in the sample is that of the year 2017. This is possible because the type of heating and dhw installation of the respondents is based on the survey which was answered partly in 2018 (Janssen-Jansen, 2019). As a result, the gas consumption data would not or only partly reflect the effect of a heat pump on gas use, which shows as insignificant in the regression analysis. In fact, according to Statistics Netherlands (CBS) (2022a), at the end of 2017 there were about 52,000 more heat pumps in use in the Dutch residential sector than at the beginning of the year, representing an almost 30% increase. During 2018, the number of heat pumps in use increased with about the same rate. This could also help explain why the mean actual gas consumption of all dwellings with a heat pump in the housing module sample (Table B5) is not much lower than the overall mean actual gas consumption of the sample (Table B1).

Dwellings with a solar boiler combined with a regular gas-fueled boiler as a dhw installation consume significantly less gas compared to a dwellings with just a regular gas-fueled boiler. The difference, about 50 m^3 less, is quite small compared to the average gas use, but considerable compared to the average gas use for dhw, 270 m^3 (Milieu Centraal, n.d.b). However, electric boilers and geysers would be expected decrease gas consumption, but have significant positive correlation coefficients. While geysers use gas to heat water, they do not maintain a storage tank of water at a hot temperature constantly like boilers, but rather heat it directly when used. The unexpected result may be caused by the relatively small number of observations in these categories (Table B5).

Practical significance and goodness-of-fit As expected, most regressors are highly significant ($p < 0.001$), since they were selected based on previous results (Majcen, 2016). Also, the high significance of the variables can be explained by the large sample size. Large sample sizes can lead even weak correlations to meet significance levels that could not be detected in smaller samples. Thus, using larger samples reduces the chance of accepting the null hypothesis hypothesis when in reality it should be rejected, i.e., reducing the probability of committing a Type II error (Sekaran and Bougie, 2016). At the same time, it is important to consider practical significance. Coefficients may be statistically significant, but when interpreting their size they may not be useful in practice (Sekaran and Bougie, 2016).

The model explains 44.7% of the variation in gas consumption and 21.9% of the variation in specific gas consumption, which is very close to the R^2 values of previous work by Majcen et al. (2013b, 2015). They attribute the limited explanatory power of the regression model to behavior and preferences of the occupants, which are believed to affect actual gas consumption (Majcen et al., 2013b). Also for actual specific gas use, occupant behavior appears to explain the most variance (Majcen et al., 2015). Aside from these missing explanatory variables, the low goodness-of-fit may also result from the assumption that the variables are linearly related, and could perhaps be improved by using alternative functional forms for the independent variables (Wooldridge, 2012). While indeed the model is far from a perfect linear fit, with microeconomic data it is common to achieve low values of R^2. Although a low R^2 means that there are more factors that affect the dependent variable that have not been accounted for, it does not influence the reliability of the estimated effects of the independent variables (Wooldridge, 2012).

5.3.2 Actual gas consumption (energy module sample)

Table 12 shows the results of the multiple regression analysis on the energy module sample (N = 2,010) using most of the same variables as used in the regression on the housing model sample (5.3.1). A difference in the energy module regression is the use of the energy label obtained through the technical inspection of WoON 2018, rather than the energy labels stored in the RVO database. In addition, some variables that are unique to the energy module were included as well: the dummy variables representing whether someone is usually home during weekdays between approximately 9:00 and 18:00 (*home*) or whether it varies (*depends*); *tempDay*, the average set temperature on weekdays between approximately 9:00 and 18:00, and *tempNight*, the average set temperature on weekdays between approximately 18:00 and 9:00. Even though *tempDay* and *tempNight* represent a behavioral aspect, it is included in the regression model because, as previously seen in subsection 5.1, there were no substantial differences in the average temperature setting between the different ownership types. In addition, point-biserial correlation coefficients show that *tempDay* is not correlated with *owner*, r(2,008) = -.03, p = .126, and *tempNight* only extremely weakly: r(2,008) = .07, p = .002. As a result, these variables were expected not to influence the difference in actual gas use or EPG between the different ownership types, and they could therefore be used as another control variable that is expected to partly determine actual gas use and the EPG.

Ownership type As opposed to the model of the housing module sample in 5.3.1, Table 12 shows that neither ownership type is significant for gas consumption in m^3 in the energy module sample. For specific gas use, only private rental dwellings are significant when the significance level is 0.1. Given the sample size, this result is considered insignificant. The insignificance implies that the differences in gas use between the ownership types identified in subsection 5.2 are accounted for by the control variables. Consequently, the presence of significant behavioral differences between ownership types is unlikely, according to this model.

The disagreement of the two analyses may be explained by sample size. While statistically significant, the effects in the housing module sample were relatively weak. It may therefore be that the weak effects are too small to be statistically significant for a smaller sample size, like the energy module sample. While the energy module sample itself is not particularly small, it contains many (dummy) variables. As a result, it may occur that in some categories there are actually very few observations. For example, as shown in Table B6, there are only 10 dwellings with no income, 4 dwellings with a wood- or pellet-fueled heating installation, and 18 dwellings with a electric boiler for dhw. These numbers are generally too small based on the rule of thumb that requires a minimum of 30 subjects in each category (Sekaran and Bougie, 2016). Thus, given that a larger sample size approaches the real population more than a smaller sample size, it can be argued that there is in fact a statistically significant difference in actual gas use between ownership types in the population, although with limited practical significance.

Control variables As expected, the energy label is also significant in the energy module and the coefficients display the expected pattern. An exception is label B, which is insignificant compared to label A. This suggests that whether a dwelling has label A or B does not matter for the gas consumption. It could therefore be that the the difference in energy label in these dwellings comes from the better performance in terms of electricity rather than in terms of natural gas, e.g., through photovoltaics. The type of building is also significant and shows the expected pattern: lowest for apartments and highest for detached houses. Both building type and energy label have a substantial practical significance as well.

The average temperature setting during the day on weekdays between 9:00 and 18:00, and during the evening and night between 18:00 and 9:00 are both significant for gas use. Specifically, for each degree Celsius added to the average temperature during the day or night, the yearly gas consumption is increased by 26.7 m^3 or 27.4 m^3, respectively. These effects are quite small, relative to the average annual gas use and the expected impact of temperature settings. The similarity of the coefficients for day and night may be explained by the longer period defined as “night” compared to the period defined as “day”. The longer period could compensate for the lower expected temperature setting during the night.

Regarding type of heating and dhw installation, the results differ from the housing module sample. In the energy module, dwellings with a heat pump consume a considerable amount of gas less than dwellings with a boiler, while controlling for the other factors. Although this is the expected result, the number of subjects in this category is small, 23. This could explain the low precision of this coefficient when comparing the housing module and energy module regressions in Tables 11 and 12, respectively.

Dwellings with a gas-fueled furnace or fireplace (*gasHeater*) also have a significantly lower gas use. This could be because there is a heater in one or some rooms only, instead of a central heating system with a boiler that can heat the entire dwelling. The installations corresponding to *wood, otherHeat, dhwElecBoiler and dhwSolar* are insignificant for gas use compared to a the reference categories *boiler* and *dhwBoiler*, although solar and electric installations would be expected to lower the gas consumption. The unexpected result is attributed to the smaller sample size, which causes there to be very few subjects in these categories (Table B6). A geyser as a dhw installation gives a significantly higher gas consumption, like was also the case in the housing module sample (Table 11). Surprisingly, dwellings with a heat pump for dhw consume significantly more gas compared to dwellings with a boiler for dhw. However, the average gas use of dwellings with a heat pump for dhw is lower than for any other dhw installation (Table B6). Therefore, it may be that the regression coefficient is not accurate. This may be the case because it is likely that *heatPump* and *dhwHeatPump* are correlated, i.e., dwellings with a heat pump as heating installation are likely to also have a heatpump as a dhw installation, and vice versa. As a result, some of the effect of *dhwHeatPump* may actually be accounted for by *heatPump*, which has a very large negative coefficient.

Also, the coefficients for disposable income, value of the house and floor area are significant and positive, in accordance with the regression on the housing module sample. Like in the housing module, the effects of these variables are relatively modest, with limited practical significance.

The number of occupants is much less significant than in the housing module sample. This may also be attributed to the smaller sample size and the fewer subjects per category. For example, in the housing module there are are 2,021 households of five or more members (Table B5), whereas in the energy module sample there are only 67 (Table B6).

Regarding the year of construction, it can be seen that houses built after 1989 significantly consume less gas than older buildings, which was also the case in the housing module sample (Table 11). One reason why building age may have an effect despite controlling for energy label is the potential inaccuracy of the energy labels of older buildings. Older buildings often lack documentation on building characteristics, resulting in more estimates and assumptions when calculating the energy label. Also, perhaps the energy label does not account for certain building characteristics that do influence efficiency in terms of natural gas and that are common in older buildings.

The dummy variables *university*, *employed*, *benefits* and *retired* are insignificant, similar to the housing module sample. Also, the occupancy of the dwellings during the day does not have a significant effect.

Practical significance and goodness-of-fit Overall, there are less statistically significant variables than in the housing module regression, which is explained by the relatively small effects of these variables combined with a smaller sample size in the energy module. Still, some statistically significant variables have limited practical significance.

The R^2 of 0.529 and 0.279 of the energy module regression models for actual gas use and actual specific gas use, respectively, are higher in the energy module analysis than in the housing module. This could be due to the addition of the variables *tempDay* and *tempNight*—occupant behavior is expected to explain a large part of the variance in actual gas use (Majcen et al., 2015). Also, the more accurate energy labels in this sample could contribute to a better fit.

Table 12: Estimates of the regression coefficients by OLS on the actual gas use in the energy module sample. The standard errors are given in parentheses. See Table A1 in Appendix A for variable definitions, units and reference categories.

	Dependent variable:	
	gasA	gasSpecA
owner	-21.223 (34.488)	-0.070 (0.346)
private	52.327 (48.180)	0.810^{*} (0.484)
n2	52.384^{*} (31.725)	0.378 (0.318)
n3	113.910^{**} (49.542)	1.298^{***} (0.497)
n4	98.149^{*} (54.116)	0.706 (0.543)
n5more	44.713 (73.008)	0.554 (0.733)
university	15.474 (30.192)	0.176 (0.303)
employed	71.657 (170.560)	0.616 (1.712)
benefits	206.256 (174.244)	2.420 (1.749)
retired	96.350 (170.622)	0.781 (1.712)
dispInc	30.414^{***} (6.075)	0.128^{**} (0.061)
floorArea	2.957^{***} (0.341)	-0.055^{***} (0.003)
value	7.211^{***} (1.291)	0.033^{**} (0.013)
B	42.282 (43.142)	-0.276 (0.433)
C	238.403^{***} (43.739)	1.440^{***} (0.439)
D	426.135^{***} (52.691)	3.243^{***} (0.529)
E	577.759^{***} (60.221)	3.924^{***} (0.604)
F	615.177^{***} (69.619)	4.181^{***} (0.699)
G	937.401^{***} (94.344)	5.547^{***} (0.947)
terraced	197.338^{***} (32.648)	1.054^{***} (0.328)
semiDetached	428.501^{***} (44.581)	3.303^{***} (0.447)
detached	677.604^{***} (49.640)	4.901^{***} (0.498)
45to69	-60.156 (40.468)	-1.251^{***} (0.406)
70to89	37.544 (40.437)	-0.806^{**} (0.406)
90to09	-130.317^{***} (46.997)	-2.221^{***} (0.472)
after10	-212.540^{***} (73.261)	-3.379^{***} (0.735)
wood	115.145 (268.714)	2.578 (2.697)
gasHeater	-577.770^{***} (179.782)	-3.994^{**} (1.804)
heatPump	$-1{,}496.727^{***}$ (389.083)	-8.420^{**} (3.905)
otherHeat	-162.776 (174.933)	-0.677 (1.756)
dhwElecBoiler	-83.927 (127.435)	2.435^{*} (1.279)
dhwGeyser	437.600^{***} (112.624)	3.616^{***} (1.130)
dhwSolar	-3.286 (79.180)	0.192 (0.795)
dhwHeatPump	$1{,}579.825^{***}$ (406.001)	10.068^{**} (4.075)
dhwOther	631.629^{***} (212.506)	2.565 (2.133)
home	28.341 (56.294)	0.791 (0.565)
depends	13.732 (51.489)	0.614 (0.517)
tempDay	26.688^{***} (7.209)	0.199^{***} (0.072)
tempNight	27.382^{***} (7.851)	0.221^{***} (0.079)
Constant	-897.380^{***} (223.486)	5.476^{**} (2.243)
Observations	2,010	2,010
R^2	0.529	0.279
Adjusted R^2	0.520	0.265
Residual Std. Error (df = 1970)	525.594	5.275
F Statistic (df = 39; 1970)	56.758^{***}	19.594^{***}
Note:		*p<0.1; **p<0.05; ***p<0.01

5.3.3 The energy performance gap

After studying actual energy consumption, also the EPG was investigated. The EPG in terms of gas use is the difference between the actual gas consumption and the theoretical gas consumption of a dwelling, as in Equation 8. The literature review showed that the EPG may be caused partly by the (p)rebound effect, in addition to other causes such an the over- or underestimation of building performance (Visscher et al., 2016). At the same time, it also appears that the rebound effect varies in size depending on different occupant characteristics, such as income and also ownership type (Madlener and Hauertmann, 2011; Aydin et al., 2017). Therefore, it was hypothesized that ownership type may also have an effect on the EPG, although indirectly through differences in behavior.

As indicated in Table 5, a positive EPG is referred to as overprediction, and a negative EPG is referred to as underprediction. Because over- and underprediction have appeared to be two distinct phenomena with different explanatory factors (Majcen et al., 2015), a regression was run on the cases of overprediction and a separate regression was run on the cases of underprediction[5], both with the EPG as the dependent variable (Table 13).

Ownership type Like in the regression of actual gas consumption in the energy module sample, the ownership types are insignificant for the EPG, both in cases of underprediction and overprediction. This result is in disagreement with the hypothesis that the EPG is larger for tenants than for homeowners, even though previous studies found that tenants show higher rebound effects than homeowners. This suggests that the influence of rebound effects on the EPG is only limited. Nevertheless, van den Brom (2020) found that the rebound and prebound effects were responsible for up to 30% of the cases of lower-than-expected savings after an energy efficiency renovation. Thus, it may also be that the small sample sizes of the over- and underprediction subsets are responsible for the insignificance, similar to the regression of actual gas use in the energy module sample (5.3.2). In any case, even the statistically significant differences found in the housing module sample are very small and therefore have a low practical significance (5.3.1). This means that most likely there are not any major differences in energy consumption behavior between homeowners and tenants that cause large differences in their actual gas consumption or EPG.

This result, together with the findings in 5.3.1 and 5.3.2, are in accordance with the suggestion made in subsection 5.2 based on the correlation analysis: the correlations between ownership type and actual gas use and between ownership type and the EPG in case of overprediction are mainly caused by differences in building characteristics between the types of ownership, rather than by potential differences in behavior in homeowners and tenants. Specifically, subsection 5.1 showed that, for example, owner-occupied dwellings are much more often semi-detached or detached than rental dwellings, and owner-occupied dwellings are on average larger than rental dwellings. Both detached building type and floor area were also shown to be significant predictors of actual gas use and overpredictions. In the same way, disposable income and the number of occupants could additionally explain the correlation between ownership and actual energy use.

[5]For the regression on the cases of underprediction, the dummy variables *gasHeater*, *dhwOther* and *G* were omitted, because there were no observations in the underprediction subset with these types of installations and label. Also, *dhwHeatPump* was omitted because it correlated perfectly with *heatPump*, i.e., all the observations with a heat pump as a heating installation also had a heat pump as a dhw installation and vice versa.

Control variables The number of occupants is significant for overpredictions: overpredictions are significantly, but moderately smaller for dwellings with two or more occupants, compared to one occupant. The coefficients are between approximately 100 m^3 and 200 m^3, which is somewhat considerable compared to the mean overprediction in the sample of 940 m^3 (Table B3).

University education and income source are insignificant for both overpredictions and underpredictions.

With increasing floor area, the overprediction significantly increases. This could occur because an assumption in the theoretical gas consumption calculation is that the entire floor area of a dwelling is heated (Majcen et al., 2015), whereas for larger dwellings it is more likely that this is not the case, e.g., occupants of larger dwellings may not heat all of their rooms.

In accordance with previous studies, overprediction is increasingly larger for decreasingly energy-efficient labels (Majcen et al., 2013a, 2015; Tigchelaar et al., 2011). The difference in overprediction ranges from 222 m^3 for enery label B compared to A, to 2,051 m^3 for energy label G compared to A. This was also reflected in Figure 7.

For types of building, detached dwellings show significantly larger overpredictions than the other types (apartments, terraced, and semi-detached dwellings), by about 268 m^3. While occupants in detached houses have been shown to choose lower indoor temperatures compared to apartments (Lindén et al., 2006), average setpoint temperatures during the day and night are controlled for in the regression. Therefore, lower setpoint temperatures of detached houses than assumed in the theoretical calculation could not be an explanation their larger overpredictions. However, it may still be the case that for example, similar to choosing lower temperatures, occupants of detached dwellings choose to heat fewer spaces. At the same time, detached dwellings show significant and large underpredictions as well, compared to other types of buildings. This may suggest that the gas consumption of detached dwellings is very difficult to predict accurately by the theoretical energy consumption.

The construction year is insignificant for both overpredictions and underpredictions, i.e., the EPG is not significantly different for different construction periods, ceteris paribus.

Disposable income is insignificant for the overpredictions, but significant in case of underprediction: the underprediction becomes larger (i.e. the EPG more negative) with increasing disposable income. In other words, in cases where actual gas consumption is higher than theoretical gas consumption, higher income means a larger difference between actual and theoretical gas use. Possibly, this is the case because as seen in the regression models of actual gas consumption (Tables 11 and 12), disposable income positively correlates with actual gas consumption. It may be that households with higher income use gas less frugally than people with lower income, while the other factors are equal. Since income is not taken into account in the theoretical gas use, this could lead to higher gas use in higher income households than theoretically predicted. This result is in accordance with the neoclassical theory of consumer behavior described in subsection 2.3, in which the consumer maximizes its consumption as long as it is allowed by its income. Nevertheless, the effect of income on the size of the underprediction is relatively small, -49 m^3 per 10,000 EUR of yearly additional disposable income, compared to the mean underprediction of -378 m^3 (Table B4).

Dwellings with a geyser as a dhw installation show very large underpredictions compared to other types of installations. This result may be unreliable since the underprediction subset contains only 3 dwellings with this type of dhw installation. The same applies to the significance of *dhwOther* in overpredictions, of which there were only 7 subjects.

The temperature setting during the day between approximately 9:00 and 18:00 is significant for overpredictions: each additional degree Celsius reduces the EPG in overpredictions by 28 m^3. The temperature during the evening and night as a similar effect, although smaller, and with a lower level of significance ($p < 0.1$). Majcen et al. (2015) also found a negative effect of temperature with the size of overpredictions (in m^3/m^2).

Goodness-of-fit The models explain 60% of the variance in the EPG in the case of overprediction and only 18.5% in case of underprediction. This is similar to what was found by Majcen et al. (2015) with the EPG per m^2 as the dependent variable: underpredictions were explained much less well by the regression model than overpredictions. This may be attributed to behavioral aspects playing a larger role in underpredictions.

Appendix C contains additional regression tables for different dependent variables: the theoretical (specific) gas use, similarly to what was done by Majcen et al. (2015) (Table C3), the EPG per m^2 floor area (Table C4), and the EPG and EPG per m^2 for the whole energy module sample, i.e. over- and underpredictions combined (Table C5).

5.4 Limitations

There are several limitations to this thesis, relating to the data as well as to the analysis. Regarding the data, as described in subsection 3.3, missing data of the energy label in the housing module, i.e., the energy labels provided by RVO, were replaced by the preliminary energy label. The estimate of the preliminary label is simplified compared to the actual energy label definition and it often deviates from the definitive label (Stuart-Fox et al., 2019), thus introducing inaccurate labels for part of sample. At the same time, using only the observations with the definitive label would introduce bias as well, because the resulting sample is not representative for the national housing stock in other aspects (Stuart-Fox et al., 2019). In addition, it would reduce the sample size.

The definitive label in the housing module also has a limitation: it may not be up to date. These energy labels originate from the Netherlands Enterprise Agency (in Dutch: Rijksdienst voor Ondernemend Nederland, RVO) energy label database. The RVO database does not necessarily contain up-to-date labels, because these labels are only required to be updated when a house is sold or rented. It may therefore be possible that a house with a low-performance label in the RVO database has been retrofitted by the same owner and would qualify for a higher performance label at the time of WoON 2018. Nevertheless, the energy label (and the energy index) contained in the energy module have been obtained from a technical inspection of the dwellings carried out for the purpose of the Dutch Housing Survey, and are therefore up to date and reflecting the actual energy performance of the dwelling.

Another limitation of the data is related to how the survey categories were defined in WoON 2018. For example, the categories for the survey question regarding the type of building in the housing module do not differentiate between a mid-terraced house and an end-terraced house, or between an apartment on a middle floor and an apartment at the top floor. These differentiations could be relevant when studying gas use, since they are expected to matter for the heat loss area of the buildings. Nevertheless, the way the categories of control variables are defined should not affect the regression coefficients of the variable of interest, as long as they are not correlated with it. To illustrate, differentiating mid-terraced from end-terraced houses will most likely not influence the regression coeffi-

Table 13: Estimates of the regression coefficients by OLS of the EPG in case of overprediction or underprediction. The standard errors are given in parentheses. See Table A1 in Appendix A for variable definitions, units and reference categories.

	Dependent variable:	
	EPG	
	Overpredictions	Underpredictions
owner	−60.222 (38.241)	60.738 (94.361)
private	−56.393 (54.923)	−88.565 (118.576)
n2	-109.054^{***} (34.545)	150.726 (96.748)
n3	-203.717^{***} (55.311)	51.568 (132.535)
n4	-150.117^{**} (61.273)	118.853 (137.753)
n5more	-164.719^{**} (80.013)	−324.902 (208.546)
university	−16.284 (32.829)	70.851 (91.341)
employed	−28.107 (193.284)	452.376 (420.546)
benefits	−136.842 (197.055)	426.513 (432.699)
retired	−13.404 (192.993)	581.946 (422.511)
dispInc	−2.472 (6.720)	-48.528^{***} (17.010)
floorArea	8.229^{***} (0.363)	2.332^{*} (1.324)
value	−1.850 (1.413)	−2.230 (4.007)
B	222.192^{***} (50.819)	111.718 (100.721)
C	426.151^{***} (52.144)	98.245 (99.149)
D	734.467^{***} (60.905)	15.774 (170.263)
E	996.599^{***} (67.570)	−141.847 (267.099)
F	$1{,}369.525^{***}$ (76.981)	492.551 (318.904)
G	$2{,}051.172^{***}$ (99.871)	
terraced	−3.278 (36.112)	−20.132 (98.346)
semiDetached	−12.348 (48.430)	−61.228 (142.898)
detached	267.857^{***} (53.715)	-448.599^{***} (164.428)
45to69	15.314 (42.953)	55.466 (154.714)
70to89	−24.857 (43.693)	−15.278 (139.057)
90to09	−36.087 (51.963)	−49.945 (150.812)
after10	136.319 (89.751)	9.012 (172.978)
wood	−425.848 (314.835)	217.265 (590.054)
gasHeater	215.406 (183.696)	
heatPump	930.234^{**} (394.555)	−218.792 (163.995)
otherHeat	−164.108 (186.884)	−496.033 (746.359)
dhwElecBoiler	198.214 (140.024)	235.139 (433.437)
dhwGeyser	-204.095^{*} (123.071)	-786.627^{**} (333.303)
dhwSolar	−13.442 (88.289)	−39.081 (215.753)
dhwHeatPump	−518.522 (449.709)	
dhwOther	-617.889^{***} (215.763)	
home	−38.017 (61.770)	−173.658 (162.093)
depends	−7.058 (56.398)	−52.573 (148.181)
tempDay	-27.960^{***} (7.845)	−16.501 (22.130)
tempNight	-16.456^{*} (8.821)	1.763 (21.988)
Constant	403.060 (253.966)	−543.707 (549.330)
Observations	1,709	301
R^2	0.600	0.185
Adjusted R^2	0.590	0.078
Residual Std. Error	531.762 (df = 1669)	551.886 (df = 265)
F Statistic	64.097^{***} (df = 39; 1669)	1.721^{***} (df = 35; 265)
Note:		*p<0.1; **p<0.05; ***p<0.01

cient of ownership types, because it is not expected that the type of terraced house—mid or end—is related to whether the terraced house is owner occupied or rental.

Limitations of the analyses in this thesis should be considered when interpreting the results. First of all, regarding the representativeness of the samples, some of the differences in distributions between the samples and the national housing stock could possibly be corrected for by using survey weights, which are provided in the WoON datasets. Nevertheless, it has been shown that both samples are in many aspects a good representation of the population.

Second, the theoretical energy use was estimated, rather than obtained from existing data. Although the estimated theoretical gas use was similar to the theoretical gas use shown in previous studies, the estimation of the theoretical energy consumption—and therefore of the EPG—in this thesis was based on assumptions and approximations, and these values should therefore not be seen as the exact theoretical energy use or EPG. At the same time, estimating the theoretical gas use has allowed to analyze the EPG in the WoON 2018 dataset, which seems has not been done before in the previous analysis of this relatively recent WoON dataset (Stuart-Fox et al., 2019).

Third, the functional form of the regression models is questionable, since a linear model has been assumed. A linear model may not be the most appropriate form for these relationships, because in some cases, other types of relationships, e.g., logarithmic or exponential could seem more plausible. For example, the relationship between actual gas use and floor area could possibly be rather logarithmic than linear, since it makes sense that larger dwellings heat less rooms than smaller dwellings, i.e. they heat a smaller percentage of the dwelling. As a result, the relationship is not linear and gas use grows less fast as floor area increases. Improving the functional form of the models may also improve the goodness-of-fit (R^2 values) of the models predicting actual gas consumption and the EPG, since they are far from a perfect linear fit. Also, using more of the available and relevant variables in the WoON database in the model could improve the goodness-of-fit, particularly the variables related to behavior in energy use. Nevertheless, a low R^2 does not influence the reliability of the estimated effects of the independent variables (Wooldridge, 2012).

Finally, more insights into the predictive power of the different variables can be gained by using standardized coefficients. Standardized means that the variables are entered relative to the mean and standard deviation. Standardized coefficients are more suitable to compare to each other, i.e., the larger coefficients are the ones with most predictive power, since the effects are being measured in terms of standard deviations instead of the original units of the dependent variable, e.g. m^3 gas (Wooldridge, 2012).

Nevertheless, the limitations regarding the analyses in this thesis are at the same time opportunities for improvement of this study. In addition, the following considerations are recommended for further or similar research projects. First, defining over- or underpredictions as the theoretical gas consumption being simply unequal to the actual gas consumption may be too strict. It is very unlikely that the yearly actual gas consumption will exactly match the theoretical gas consumption. In practice, i.e., when predicting energy savings or for policy making, this is probably also not necessary. The prediction could be "good enough" if, for example, it only deviates 5% from the actual energy consumption. That way, the interpretation of over- and underprediction and their explanatory factors may be more meaningful.

Second, the WoON 2018 data base and especially the energy module contain numerous interesting variables regarding occupant characteristics, and behavior, motivations, and preferences regarding energy. This thesis only included a few of them, and therefore it is recommended to further explore these interesting data.

While these limitations must be kept in mind when interpreting the results and drawing conclusions, this thesis offered several additions to the existing literature. First of all, the large sample size of data used in this thesis and its representativeness of the Dutch housing stock are an advantage compared to most of the previously analysed data sets (Majcen et al., 2013a,b; van den Brom, 2020). In particular, the distribution of the ownership type is well represented in this thesis, while previous studies have mainly used data exclusively from the social housing sector (van den Brom, 2020) or in which social housing is much overrepresented (Majcen et al., 2013b, 2015). In addition to the large sample size and its representativeness, the data is relatively recent, compared to other, similar data sets previously used, e.g., WoON 2012 (Majcen et al., 2015). An additional advantage of using energy data from the year 2017 is that 2017 can be considered a relatively normal year and could therefore be representative for a year without disruptions, such as the COVID-19 pandemic in 2020-2021 or the Russian invasion of Ukraine in 2022, that may impact the economy and the use of energy. Also, the large number and the variety of variables used in the analyses of this thesis and especially the presence of behavioral data is valuable, since it has been recognized that studies investigating energy-related behavior are mostly found only for small sample sizes (van den Brom, 2020).

6 Conclusion

The conclusion of this thesis answers the research questions by summarizing the main findings in subsection 6.1 and reflects on the meaning and implications of the findings in a broader context in subsection 6.2.

6.1 Answering the research questions

The objectives of this thesis were to deliver an analysis of the actual gas use and the energy performance gap in a sample of households representative of the Dutch population; to determine to what extent ownership of the dwelling influences the size of the energy performance gap; and to investigate the mechanism that could explain the influence of ownership on the energy performance gap.

A large, representative, and relatively recent data set of the Dutch housing stock, WoON 2018, was analyzed by descriptive statistics and multiple linear regression to explain the variance in actual gas consumption of households and in the difference between actual gas consumption and theoretical gas consumption—the energy performance gap. The results of these analyses provide the answers to the sub-questions of this research.

To what extent does the type of ownership—owner-occupied or rental—predict the actual natural gas consumption and the energy performance gap?

A moderate positive correlation was identified between home ownership and actual gas consumption, This means that, without taking any other influencing factors into account, owner-occupied dwellings have a higher yearly gas consumption than rental dwellings.

Multiple regression analysis allowed to control for influencing factors, such as building characteristics and other occupant characteristics, that influence the gas consumption of a household. Within rental dwellings, social housing and private rental dwellings were distinguished. In the larger sample (N = 41,971), the type of ownership was significant at a 1% significance level for the yearly natural gas use of a household. Specifically, owner-occupied dwellings consumed more gas than social rental dwellings, and private rental dwellings consumed more gas than owner-occupied dwellings, ceteris paribus. A suggested explanation for the higher gas use of private rental dwellings was the so-called "all-in" rent, which eliminates tenants' financial incentive to conserve energy, leading to higher energy consumption (Government of the Netherlands, n.d.a; Levinson and Niemann, 2004; Maruejols and Young, 2011). However, despite the statistical significance of the ownership categories, their effects were relatively weak compared to the average actual gas consumption in the sample of 1,400 m^3/year. In fact, a similar analysis on a smaller sample (N = 2,010) indicated that neither ownership type was statistically significant for actual gas consumption, ceteris paribus. This was attributed to the smaller sample size, in combination with the weak effects identified in the larger sample.

Also, a weak positive correlation was found between home ownership and the energy performance gap in case of overprediction. This means that for cases where the theoretical gas consumption exceeds the actual gas consumption, the gap is larger in owner-occupied dwellings ccompared to rental dwellings. Multiple regression analysis of the energy performance gap showed, however, that ownership type was insignificant, ceteris paribus, both for overpredictions and underpredictions.

In sum, the multiple regression analyses showed that when occupant characteristics and building characteristics are controlled for, owner-occupied dwellings do not correlate with a higher gas consumption anymore. Furthermore, private rental dwellings do (practically) not have an effect on actual gas consumption and the EPG either, ceteris paribus.

What empirical variables influence the relationship between ownership and the energy performance gap?

While there is a correlation between homeowners and higher gas use and larger overpredictions, there is no (practical) effect of ownership type when building and occupant characteristics are controlled for. This means that there are most likely no major behavioral differences that could cause an effect of ownership type on actual gas use or the EPG. Rather, the correlation is probably influenced by differences in building characteristics and other household characteristics. Differences in building and occupant characteristic between the types of ownership were investigated using descriptive statistics. By in addition considering their explanatory effect on actual gas use and/or the energy performance gap, potential factors are identified that could explain the correlation. For actual gas consumption these are the household size (number of occupants), building type (apartment, terraced, semi-detached or detached), floor area, and disposable income. For the energy performance gap the potential factors are building type and floor area.

What is the mechanism of the relationship between ownership and the energy performance gap?

When building and occupant characteristics are controlled for, there is no (practical) effect of ownership type on actual gas consumption or on the EPG. This means that behavioral differences probably do not cause an effect of ownership type on actual gas use or the EPG. A number of building characteristics and other household characteristics were identified that could potentially explain the correlations through the following mechanisms.

A higher number of occupants increases the gas use of a dwelling. Larger households are likely to consume more hot water, e.g., for showering. Also, they might need to heat more rooms, since more rooms are likely to be occupied simultaneously. At the same time, owner-occupied dwellings more often contain larger households. Nearly eighty percent of owner-occupied dwellings contain households of two people or more, whereas in rental dwellings only about half contain two or more occupants. Hence, larger household sizes contribute to the higher gas consumption of owner-occupied dwellings.

Detached and semi-detached dwellings have the largest heat loss areas compared to other building types like terraced or apartments, because they are not surrounded by (similar) dwellings. Due to the higher thermal losses, these types of building need to consume more gas for heating. At the same time, previous research found that occupants of detached dwellings choose lower indoor temperatures (Lindén et al., 2006). This may indicate more conserving behavior of occupants of detached dwellings. This in turn may cause them to deviate more from the assumptions in the theoretical calculation, thereby leading to larger overpredictions. Almost half of all owner-occupied dwellings are either semi-detached or detached houses, whereas only a few percent of rental dwellings are in these categories. The higher gas requirement for heating of (semi-)detached buildings therefore contributes to a higher gas consumption in owner-occupied dwellings. At the same time, the potentially more conserving behavior of occupants in detached dwellings than assumed in the theoretical calculations may lead to larger overpredictions of owner-occupied dwellings.

A larger floor area increases the volume of the house, i.e., the space to be heated. As a consequence, larger houses need to consume more gas to heat their spaces. Also, larger floor areas increase the size of overpredictions. This is because an assumption in the theoretical calculation is that the entire floor area is heated, whereas in reality homes with larger floor areas are less likely to heat all their rooms and spaces. This causes the gas consumption to be more overestimated as floor area increases. Owner-occupied houses are on average larger than rental dwellings, thereby requiring a higher gas consumption for heating the dwelling, and increasing overpredictions.

Higher disposable incomes lead to slightly higher gas consumption, because households with higher income may use gas less frugally than people with lower incomes. For example, they may choose higher indoor temperatures and heat more rooms. Average disposable income is higher in owner-occupied dwellings than in rental dwellings, which allows home-owners to afford a higher energy use.

How does home ownership influence the energy performance gap in residential gas consumption?
The main research question can be answered based on the answers to the sub-questions. To summarize, ownership does not have a practically significant effect neither on actual gas consumption or the EPG, ceteris paribus. However, there is a positive correlation between home ownership and actual gas consumption, and between home ownership and the size of overpredictions. These correlations have been found to be caused most likely by a number of building characteristics and household characteristics, rather than through behavioral differences, as initially hypothesized. Specifically, the correlation between home ownership and actual gas consumption can be explained by the larger floor area, type of buildings, higher income, and larger household size. The correlation between home ownership and the size of overpredictions is explained by type of building and larger floor area.

6.2 Reflection

Despite its limitations, this thesis provides an analysis of actual gas consumption and the EPG on a large, representative and relatively recent sample of Dutch households. It has shown the effect on actual gas consumption and the energy performance gap of multiple occupant characteristics, building characteristics and some behavioral aspects, using a large, representative sample of the Dutch housing stock. The findings in this report add to the understanding of the factors that determine the variation in the consumption of natural gas in the residential sector and in the energy performance gap. Understanding the energy performance gap and analysing actual residential energy consumption data is necessary to improve theoretical consumption calculations. Policymakers set energy saving targets, develop energy-saving plans and monitor the energy-saving progress based on theoretical energy consumption (van den Brom, 2020), but the discrepancy between actual and theoretical energy consumption has shown that these energy reduction targets cannot be met based on actual energy consumption (Majcen et al., 2013a). Combining actual energy consumption data with the theoretical models can narrow the energy performance gap, providing more reliable outcomes (van den Brom, 2020). Furthermore, knowing which type of household characteristics contribute to higher or lower energy consumption can also improve the effectiveness of energy-saving policies and campaigns, if the right audience is targeted. Besides policy-making, theoretical energy consumption is also used in practice by contractors and consultants, for instance to predict the effectiveness of renovation measures, to estimate their energy savings and to calculate their payback times. Understanding the energy performance gap and the factors that may affect it, can help practitioners take into account the possible inaccuracy of the predicted effects of renovation measures. More certainty on the payback time may increase the willingness of homeowners and housing associations to take renovation measures (van den Brom, 2020).

Occupants have been shown to explain a large part of the variation in actual energy consumption, but previous research has mostly been done on social rental housing data van den Brom (2020); Majcen (2016). At the same time, homeowners and tenants have been shown to behave differently with respect to energy use in some cases (Levinson and Niemann, 2004; Aydin et al., 2017; Madlener and Hauertmann, 2011; Maruejols and Young, 2011). This thesis has shown that there is no practically significant effect of ownership type itself on actual gas consumption or on the energy performance gap, when other occupant and building characteristics are controlled for. Thus, it shows that distinguishing owners and (social or private) tenants in policy is not necessary to improve the effectiveness of energy saving policies and campaigns. However, it has also shown that there is a correlation between home ownership and higher natural gas consumption, which can be explained by factors like the larger average size, higher frequency of (semi-)detached homes, and larger household sizes of home owners. In addition, home ownership is the most frequently occurring ownership type in the Netherlands. Thus, the distinction of ownership type may still be of practical use to, for example, municipalities. Targeting homeowners could be an efficient way to promote energy-saving measures in the largest, most common, and highest energy-consuming dwellings.

References

Erdal Aydin, Nils Kok, and Dirk Brounen. Energy efficiency and household behavior: the rebound effect in the residential sector. *RAND Journal of Economics*, 48:749–782, 9 2017. ISSN 17562171. doi: 10.1111/1756-2171.12190.

Peter H.G. Berkhout, Jos C. Muskens, and Jan W. Velthuijsen. Defining the rebound effect. *Energy Policy*, 28:425–432, 2000. ISSN 03014215. doi: 10.1016/S0301-4215(00)00022-7.

Kornelis Blok and Evert Nieuwlaar. *Introduction to Energy Analysis*. Routledge, third edition, 2021.

Emmanuelle Cayre, Benoit Allibe, Marie-Hélène Laurent, and Dominique Osso. There are people in the house! How the results of purely technical analysis of residential energy consumption are misleading for energy policies. *Proceedings of the European Council for an Energy Efficient Economy (ECEEE) Summer School*, 2011.

M. Cremers. Energieonderzoek - 2018: Vragenlijst, 2017. Centraal Bureau voor de Statistiek (CBS).

Ellie de Groot, Marleen Spiekman, and Ivo Opstelten. Dutch Research into User Behaviour in Relation to Energy Use of Residences, 2008. PLEA 2008 – 25$^{\text{th}}$ Conference on Passive and Low Energy Architecture, Dublin, 22$^{\text{nd}}$ to 24$^{\text{th}}$ October 2008.

Environmental Data Compendium. Energielabels van woningen, 2010 - 2019. `https://www.clo.nl/indicatoren/nl0556-energielabels-woningen`, 2020. Accessed 2022-07-10.

European Commission. Energy efficiency first principle. `https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficiency-targets-directive-and-rules/energy-efficiency-first-principle_en`, n.d. Accessed: 2022-03-30.

Eurostat. Energy statistics - an overview. `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_statistics_-_an_overview#Final_energy_consumption`, 2 2022. Accessed: 2022-03-27.

Zachary M. Gill, Michael J. Tierney, Ian M. Pegg, and Neil Allan. Low-energy dwellings: the contribution of behaviours to actual performance. *https://doi.org/10.1080/09613218.2010.505371*, 38:491–508, 9 2010. ISSN 09613218. doi: 10.1080/09613218.2010.505371. URL `https://www.tandfonline.com/doi/abs/10.1080/09613218.2010.505371`.

Government of the Netherlands. Rijksdienst voor Ondernemend Nederland van start. `https://web.archive.org/web/20150402133624/http://www.rijksoverheid.nl/nieuws/2014/01/06/rijksdienst-van-ondernemend-nederland-van-start.html`, 1 2014. Accessed: 2022-03-31.

Government of the Netherlands. Nieuw energielabel voor gebouwen per 1 januari 2021. `https://www.rijksoverheid.nl/actueel/nieuws/2020/12/18/nieuw-energielabel-voor-gebouwen-per-1-januari-2021`, 2020. Accessed 2022-03-28.

Government of the Netherlands. Waarom betaal ik servicekosten en kosten voor gas, water en elektriciteit? https://www.rijksoverheid.nl/onderwerpen/woning-huren/vraag-en-antwoord/waarom-betaal-ik-servicekosten-en-kosten-voor-gas-water-en-elektriciteit, n.d.a. Accessed: 2022-07-31.

Government of the Netherlands. Rented housing. https://www.government.nl/topics/housing/rented-housing, n.d.b. Accessed: 2022-08-04.

Olivia Guerra Santín. *Actual energy consumption in dwellings: The effect of energy performance regulations and occupant behaviour.* PhD thesis, Delft University of Technology, 2010.

Olivia Guerra Santín. Behavioural Patterns and User Profiles related to energy consumption for heating. *Energy and Buildings*, 43:2662–2672, 2011.

Olivia Guerra Santín and Laure Itard. Occupants' behaviour: determinants and effects on residential heating consumption. *Building Research & Information*, 38:318–338, 5 2010.

Olivia Guerra Santín and Laure Itard. The effect of energy performance regulations on energy consumption. *Energy Efficiency*, 5:269–282, 8 2012. ISSN 1570646X. doi: 10.1007/S12053-012-9147-9/TABLES/6. URL https://link.springer.com/article/10.1007/s12053-012-9147-9.

Reinhard Haas, Hans Auer, and Peter Biermayr. The impact of consumer behavior on residential energy demand for space heating. *Energy and Buildings*, 27:195–205, 1998.

Florian Heiss. *Using R for Introductory Econometrics*. Self-published, Düsseldorf, second edition, 2020. http://www.urfie.net/read/index.html.

Hugo Hens, Wout Parijs, and Mieke Deurinck. Energy consumption for heating and rebound effects. *Energy and Buildings*, 42:105–110, 1 2010. ISSN 0378-7788. doi: 10.1016/J.ENBUILD.2009.07.017.

Marek Hlavac. stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.3. https://CRAN.R-project.org/package=stargazer, 2022. Accessed: 2022-06-26.

House of Representatives of The Netherlands. Convenant Energiebesparing corporatiesector. https://zoek.officielebekendmakingen.nl/kst-31700-xviii-4-b1, 2008. Accessed: 2022-07-14.

Intergovernmental Panel on Climate Change (IPCC). Climate Change 2022 Impacts, Adaptation and Vulnerability. https://www.ipcc.ch/report/sixth-assessment-report-working-group-ii/, 2 2022.

International Energy Agency (IEA). World Energy Outlook 2021. https://www.iea.org/reports/world-energy-outlook-2021, 2 2021.

International Energy Agency (IEA). A 10-Point Plan to Reduce the European Union's Reliance on Russian Natural Gas. https://www.iea.org/reports/a-10-point-plan-to-reduce-the-european-unions-reliance-on-russian-natural-gas, 3 2022.

Stichting ISSO. ISSO-publicatie 82.1 Energieprestatie woningen en woongebouwen (methode 2020 - 3e druk). https://open.isso.nl/publicatie/isso-publicatie-82-1-energieprestatie-woningen-en-woongebouwen/2021, 2020.

Sherri L. Jackson. *Statistics Plain and Simple.* Wadsworth, Belmont, third edition, 2013.

Saskia Janssen-Jansen. WoON 2018 Onderzoeksdocumentatie en Kwaliteitsanalyse (conceptversie 0.1), 2019. Centraal Bureau voor de Statistiek (CBS).

Marie-Hélène Laurent, Benoît Allibe, Tadj Oreszczyn, Ian Hamilton, Ray Galvin, and Casper Tigchelaar. Back to reality: How domestic energy efficiency policies in four european countries can be improved by using empirical data instead of normative calculation. https://discovery.ucl.ac.uk/id/eprint/1403583/, 2013. European Council for an Energy Efficient Economy (ECEEE).

Arik Levinson and Scott Niemann. Energy use by apartment tenants when landlords pay for utilities. *Resource and Energy Economics*, 26:51–75, 2004.

A. Lindén, A. Carlsson-Kanyama, and B. Eriksson. Efficient and inefficient aspects of residential energy behaviour: what are the policy instruments for change? *Energy Policy*, 34:1918–1927, 2006.

Reinhard Madlener and Maximilian Hauertmann. Rebound Effects in German Residential Heating: Do Ownership and Income Matter? *SSRN Electronic Journal*, 2 2011. ISSN 1556-5068. doi: 10.2139/SSRN.1887030. URL https://papers.ssrn.com/abstract=1887030.

Daša Majcen. *Predicting energy consumption and savings in the housing stock: A performance gap analysis in the Netherlands.* PhD thesis, Delft University of Technology, 2016.

Daša Majcen, Laure Itard, and Henk Visscher. Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications. *Energy Policy*, 54:125–136, 3 2013a. ISSN 0301-4215. doi: 10.1016/J.ENPOL.2012.11.008.

Daša Majcen, Laure Itard, and Henk Visscher. Actual and theoretical gas consumption in Dutch dwellings: What causes the differences? *Energy Policy*, 61:460–471, 10 2013b. ISSN 0301-4215. doi: 10.1016/J.ENPOL.2013.06.018.

Daša Majcen, Laure Itard, and Henk Visscher. Statistical model of the heating prediction gap in Dutch dwellings: Relative importance of building, household and behavioural characteristics. *Energy and Buildings*, 105:43–59, 10 2015. ISSN 0378-7788. doi: 10.1016/J.ENBUILD.2015.07.009.

Lucie Maruejols and Denise Young. Split incentives and energy efficiency in Canadian multi-family dwellings. *Energy Policy*, 39:3655–3668, 2011.

Milieu Centraal. Gemiddeld energieverbruik. https://www.milieucentraal.nl/energie-besparen/inzicht-in-je-energierekening/gemiddeld-energieverbruik/, n.d.a. Accessed: 2022-07-31.

Milieu Centraal. Bespaartips warm water. https://www.milieucentraal.nl/energie-besparen/duurzaam-warm-water/bespaartips-warm-water/, n.d.b. Accessed: 2022-07-31.

Ministry of the Interior and Kingdom Relations (BZK) and Statistics Netherlands (CBS). WoON2018: release 1.0 - Woononderzoek Nederland 2018. https://doi.org/10.17026/dans-z6v-chq9, 2019a. DANS.

Ministry of the Interior and Kingdom Relations (BZK) and Statistics Netherlands (CBS). WoON2018energie: release 1.0 – Energiemodule WoON2018. https://doi.org/10.17026/dans-xt4-jxzw, 2019b. DANS.

Netherlands Enterprise Agency (RVO). Standaard en streefwaarden voor woningisolatie. https://www.rvo.nl/onderwerpen/wetten-regels/standaard-streefwaarden-woningisolatie#de-standaard-op-het-energielabel, 07 2021. Accessed: 2022-07-07.

Rent Tribunal. Splitsen all-in huurprijs. https://www.huurcommissie.nl/onderwerpen/huurprijs-en-punten/splitsen-all-in-huurprijs, n.d. Accessed: 2022-08-03.

Uma Sekaran and Roger Bougie. *Research Methods for Business.* Wiley, Chichester, seventh edition, 2016.

Michelle Shipworth, Steven K. Firth, Michael I. Gentry, Andrew J. Wright, David T. Shipworth, and Kevin J. Lomas. Central heating thermostat settings and timing: building demographics. *Building Research & Information*, 38:50–69, 2010.

Steve Sorrell. Reducing energy demand: A review of issues, challenges and approaches. *Renewable and Sustainable Energy Reviews*, 47:74–82, 2015. ISSN 18790690. doi: 10.1016/j.rser.2015.03.002.

Statistics Netherlands (CBS). Voorraad woningen; eigendom, type verhuurder, bewoning, regio. https://opendata.cbs.nl/#/CBS/nl/dataset/82900NED/table, 11 2018. Accessed: 2022-05-16.

Statistics Netherlands (CBS). Welvaart; grenzen van 10%-groepen inkomen & vermogen. https://opendata.cbs.nl/#/CBS/nl/dataset/83934NED/table, 10 2021. Accessed: 2022-06-02.

Statistics Netherlands (CBS). Warmtepompen; aantallen, thermisch vermogen en energiestromen. https://opendata.cbs.nl/#/CBS/nl/dataset/82380NED/table, 6 2022a. Accessed: 2022-08-02.

Statistics Netherlands (CBS). Voorraad woningen; gemiddeld oppervlak; woningtype, bouwjaarklasse, regio. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82550NED/table?fromstatweb, 4 2022b. Accessed: 2022-06-02.

Michael Stuart-Fox, Tom Kleinepier, and Kenneth Gopal. Energie besparen in de woningvoorraad: inzichten uit de Energiemodule WoON 2018. https://www.eumonitor.eu/9353000/1/j9vvik7m1c3gyxp/vl1thzqe5nyx, 2019. ABF Research.

Casper Tigchelaar, Bert Daniëls, and Marijke Menkveld. Obligations in the existing housing stock: Who pays the bill? https://repository.tudelft.nl/islandora/object/u uid%3A4d66e962-438f-409d-9623-25b5e0bebe73, 2011. Energieonderzoek Centrum Nederland (ECN).

Paula van den Brom. *Energy in Dwellings: A comparison between Theory and Practice*. PhD thesis, A+BE | Architecture and the Built Environment, 2020.

Paula van den Brom, Arjen Meijer, and Henk Visscher. Performance gaps in energy consumption: household groups and building characteristics. *Builing Research & Information*, 46:54–70, 1 2017. ISSN 14664321. doi: 10.1080/09613218.2017.1312897. URL https://www.tandfonline.com/doi/abs/10.1080/09613218.2017.1312897.

Coen van Duin, Saskia te Riele, and Lenny Stoeldraijer. Huishoudensprognose 2018-2060: opmars eenpersoonshuishoudens zet door, 2018. Centraal Bureau voor de Statistiek (CBS).

Sandra van Zoelen and Kenneth Gopal. Dataprocessing WoON Energie 2018. https://disco.datawonen.nl/jaargang?bronid=WOON&jaargangid=2018_EM, 7 2019. ABF Research.

Henk Visscher, Frits Meijer, Daša Majcen, and Laure Itard. Improved governance for energy efficiency in housing. *Building Research & Information*, 44:552–561, 8 2016. ISSN 14664321. doi: 10.1080/09613218.2016.1180808. URL https://www.tandfonline.com/doi/abs/10.1080/09613218.2016.1180808.

Jeffrey M. Wooldridge. *Introductory Econometrics*. South-Western College Pub, Mason, fifth edition, 2012. http://www.urfie.net/read/index.html.

A Variable definitions

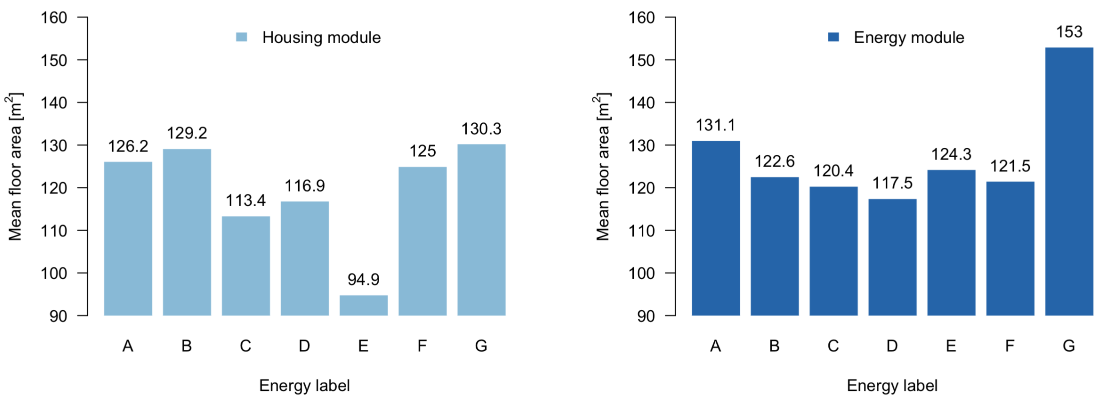
Table A1: Description of the variables used in the regression analysis.

	Name	Description [units]
Dependent variables	*gasSpecA*	Actual specific gas use, i.e. actual gas consumption of a dwelling in 2017 per unit floor area [$m^3 \cdot m^{-2} \cdot year^{-1}$]
	gasSpecT	Estimated theoretical specific gas use, i.e. estimated theoretical gas consumption per unit floor area [$m^3 \cdot m^{-2} \cdot year^{-1}$]
	gasA	Actual gas consumption of a dwelling in 2017 [m^3]
	gasT	Estimated theoretical gas consumption of a dwelling [m^3]
	EPG	Difference between theoretical and actual gas consumption (Equation 8) [m^3]
	EPGm2	Difference between theoretical and actual specific gas consumption [$m^3 \cdot m^{-2} \cdot year^{-1}$]
Independent variables	*social* (ref.)	= 1 if social rental dwelling, 0 otherwise
	owner	= 1 if owner-occupied dwelling, 0 otherwise
	private	= 1 if private rental dwelling, 0 otherwise
	floorArea	Floor area of the dwelling [m^2]
	n1 (ref.)	= 1 if household consists of a single person, 0 otherwise
	n2	= 1 if household consists of two people, 0 otherwise
	n3	= 1 if household consists of three people, 0 otherwise
	n4	= 1 if household consists of four people, 0 otherwise
	n5m	= 1 if household consists of two people, 0 otherwise
	A (ref.)	= 1 if energy label A, 0 otherwise
	B	= 1 if energy label B, 0 otherwise
	C	= 1 if energy label C, 0 otherwise
	D	= 1 if energy label D, 0 otherwise
	E	= 1 if energy label E, 0 otherwise
	F	= 1 if energy label F, 0 otherwise
	G	= 1 if energy label G, 0 otherwise
	apartment (ref.)	= 1 if flat or apartment, 0 otherwise
	terraced	= 1 if mid-terraced or end-terraced house, i.e. a single-family house attached to other similar houses on both sides (mid-terraced) or one side (end-terraced), 0 otherwise
	semiDetached	= 1 if semi-detached dwelling, i.e. a single-family house attached to a different type of building, 0 otherwise
	detached	= 1 if single-family house detached from other buildings, 0 otherwise
	before1945 (ref.)	= 1 if building constructed before 1945, 0 otherwise
	45to69	= 1 if building constructed between 1945 and 1969, 0 otherwise
	70to89	= 1 if building constructed between 1970 and 1989, 0 otherwise
	90to09	= 1 if building constructed between 1990 and 2009, 0 otherwise
	after10	= 1 if building constructed in 2010 or after, 0 otherwise
	noUniversity (ref.)	= 1 if neither respondent or their partner has university education (master's degree or doctoral degree) school, 0 otherwise
	university	= 1 if respondent and/or their partner has university education (master's degree or doctoral degree), 0 otherwise
	noIncome (ref.)	= 1 if household has no source of income (i.e., has income exclusively from properties or allowances), 0 otherwise
	employed	= 1 if main source of income is employment, 0 otherwise
	benefits	= 1 if main source of income are state benefits, 0 otherwise
	retired	= 1 if main source of income is pension, 0 otherwise
	dispInc	Disposable income [10,000 EUR / year]
	value	Value of the house in 2017 [10,000 EUR]
	boiler (ref.)	= 1 if heating installation is a gas-fueled boiler, 0 otherwise
	wood	= 1 if heating installation fueled by wood or pellets, 0 otherwise
	gasHeater	= 1 if heating installation is a gas furnace or gas fireplace, 0 otherwise
	heatPump	= 1 if heating installation is a heat pump, 0 otherwise
	otherHeat	= 1 if heating installation is neither a boiler, wood-fueled, heater, or heat pump, 0 otherwise
	dhwGasBoiler (ref.)	= 1 if dhw installation is a gas-fueled boiler, 0 otherwise
	dhwSolar	= 1 if dhw installation is a gas-fueled boiler combined with a solar boiler, 0 otherwise
	dhwGeyser	= 1 if dhw installation is a gas-fueled geyser, 0 otherwise
	dhwElecBoiler	= 1 if dhw installation is an electric boiler, 0 otherwise
	dhwHeatPump	= 1 if dhw installation is a heat pump, 0 otherwise
	dhwOther	= 1 if dhw installation is neither a (solar or electric) boiler, geyser or heat pump, 0 otherwise
	away (ref.)	= 1 if occupant is almost never at home on weekdays between approximately 9:00 and 18:00, 0 otherwise
	home	= 1 if occupant almost always at home on weekdays between approximately 9:00 and 18:00, 0 otherwise
	depends	= 1 if occupant neither almost always or almost never at home on weekdays between approximately 9:00 and 18:00, 0 otherwise
	tempDay	Average indoor temperature setting on weekdays between approximately 9:00 and 18:00 [°C]
	tempNight	Average indoor temperature setting on weekdays between approximately 18:00 and 9:00 [°C]

B Descriptive statistics on whole samples

B.1 Dwelling size

Because average dwelling size (floor area) has been reported to vary among different energy labels (Majcen et al., 2013a), they are visualized for the housing module sample and the energy module sample in Figures B1a and B1b, respectively. Specifically, in the study by Majcen et al. (2013a), label A dwellings are on average considerably larger than dwellings with any of the other labels. This is not the case in the housing or energy module sample Figure B1. In the housing module sample (Figure B1a), label E consists of considerably smaller dwellings compared to other labels. In the energy module sample (Figure B1b), dwellings with a label G are considerably larger on average than dwellings with higher efficiency labels. Moreover, the average floor areas found in the WoON 2018 samples are generally larger than those found in Majcen et al. (2013a). In Figure B1a, the mean areas vary between 94.9 m^2 and 130.3 m^2, and in Figure B1b, between 117.5 m^2 and 153 m^2, whereas those found by Majcen et al. (2013a) vary between 90.2 m^2 and 105.1 m^2. This, and the different distribution, may be explained by the large overrepresentation of social housing in the sample used by Majcen et al. (2013a) (almost 80% of the sample), since social housing is more likely to be smaller in size than private rental or owner-occupied dwellings. The disagreement between Figure B1a and Figure B1b may be attributable to the possible inaccuracy of the energy labels in the housing module sample and their discrepancies between the two samples (subsection 3.4).



(a) (b)

Figure B1: Mean floor area per label in the housing module sample (a) and in the energy module sample (b).

B.2 Continuous variables

Tables B1 and B2 show the summary statistics of the continuous variables of the housing module sample and the energy module sample, respectively.

As mentioned in subsection 4.4, the phenomena of overprediction (EPG $>$ 0) and underprediction (EPG $<$ 0) can be distinguished and should be analyzed separately. For that, the energy module was divided into cases of overprediction and underprediction, leading to two

Table B1: Summary statistics of the ratio scaled variables of the housing module sample. See Table A1 for variable descriptions and units.

Variable	N	Mean	St. Dev.	Min	Max
gasA	41,971	1,399.3	764.0	1	9,196
gasSpecA	41,971	12.4	6.6	0.01	132.4
floorArea	41,971	119.3	57.2	14	1,000
dispInc	41,971	4.2	3.0	−101.3	126.4
value	41,971	23.5	13.4	1.7	272.2

Table B2: Summary statistics of the ratio scaled variables of the energy module sample. See Table A1 for variable descriptions and units.

Variable	N	Mean	St. Dev.	Min	Max
gasA	2,010	1,416.4	758.5	25	7,438
gasSpecA	2,010	12.1	6.2	0.2	113.5
floorArea	2,010	123.2	53.6	15	617
dispInc	2,010	4.3	2.6	−0.3	42.9
value	2,010	25.5	13.4	4.3	131.8
EI	2,010	1.6	0.5	−1.4	3.5
ALoss	2,010	226.4	136.6	15.5	2,362.6
gasT	2,010	2,159.4	1,241.1	−4,269.4	14,134.6
gasSpecT	2,010	18.0	7.5	−23.3	132.6
EPG	2,010	743.0	925.9	−6,326.4	9,858.2
tempDay	2,010	19.0	2.3	0.0	26.0
tempNight	2,010	18.2	2.0	0.0	26.0

separate samples. The summary statistics of the continuous variables in the overprediction and underprediction samples are shown in Tables B3 and B4, respectively.

Tables B5 and B6 show the mean and standard deviation of the actual (specific) gas consumption per category in the housing module sample and energy module sample, respectively, as well as the frequency (count) of the categories. For example, there are 27,136 owner-occupied dwellings in the housing module sample, and their average actual gas use is 1567 m^3 per year (Table B5).

Table B3: Summary statistics of the ratio scaled variables of the overpredictions (EPG > 0). See Table A1 for variable descriptions and units.

Variable	N	Mean	St. Dev.	Min	Max
gasA	1,709	1,356.0	712.5	25	7,438
gasSpecA	1,709	11.2	5.3	0.2	113.5
floorArea	1,709	125.4	54.3	15	617
dispInc	1,709	4.3	2.6	−0.3	42.9
value	1,709	25.8	13.3	5.8	125.9
EI	1,709	1.7	0.5	0.6	3.5
ALoss	1,709	231.8	136.8	15.5	2,362.6
gasT	1,709	2,296.4	1,250.5	556.9	14,134.6
gasSpecT	1,709	18.8	7.4	4.6	132.6
EPG	1,709	940.4	830.8	0.4	9,858.2
tempDay	1,709	18.9	2.2	0.0	24.0
tempNight	1,709	18.1	1.9	0.0	24.0

Table B4: Summary statistics of the ratio scaled variables of the underpredictions (EPG < 0). See Table A1 for variable descriptions and units.

Variable	N	Mean	St. Dev.	Min	Max
gasA	301	1,759.3	907.2	586	7,314
gasSpecA	301	17.1	7.9	4.9	58.7
floorArea	301	110.1	47.0	15	383
dispInc	301	4.4	2.8	0.3	24.1
value	301	23.5	13.3	4.3	131.8
EI	301	1.3	0.4	−1.4	2.7
ALoss	301	195.8	131.4	15.6	814.2
gasT	301	1,381.8	835.7	−4,269.4	5,114.2
gasSpecT	301	13.2	6.3	−23.3	45.9
EPG	301	−377.5	574.6	−6,326.4	−0.8
tempDay	301	19.6	2.4	0.0	26.0
tempNight	301	18.8	2.4	0.0	26.0

B.3 Categorical variables

Table B5: Summary statistics of the actual (specific) gas consumption per category in the housing module sample (N = 41,971). See Table A1 for variable descriptions.

		Actual gas use ($m^3 \cdot year^{-1}$)		Actual specific gas use ($m^3 \cdot m^{-2} \cdot year^{-1}$)	
Category	Count	Mean	St. Dev.	Mean	St. Dev
owner	27136	1566.94	800.22	12.18	5.96
social	11243	1043.20	491.89	12.43	6.28
private	3592	1247.31	770.68	14.54	10.54
n1	13318	1146.50	694.49	12.59	7.64
n2	16189	1472.28	788.07	12.27	6.12
n3	5087	1506.43	715.88	12.67	6.21
n4	5356	1581.98	693.59	12.32	5.49
n5more	2021	1726.52	847.24	12.77	6.33
A	4553	1049.74	589.75	8.79	4.78
B	7241	1276.94	688.77	10.38	5.28
C	13399	1305.75	640.45	11.86	5.41
D	5176	1563.94	819.77	14.01	6.83
E	3274	1330.02	652.26	14.58	6.56
F	4033	1742.44	838.29	14.99	6.88
G	4295	1800.06	986.70	15.74	9.08
apartment	10961	917.88	515.94	12.35	8.16
terraced	19136	1320.68	533.37	12.09	5.36
semiDetached	6085	1676.84	630.77	12.54	5.47
detached	5789	2278.86	1027.65	13.73	7.77
before1945	6971	1657.25	910.96	15.13	8.40
45to69	9508	1480.94	763.16	14.43	6.68
70to89	14071	1409.27	719.58	12.22	5.58
90to09	9508	1198.29	654.18	9.50	4.66
after10	1913	979.00	548.24	9.14	6.11
noUniversity	34829	1360.91	727.99	12.50	6.59
university	7142	1586.44	896.62	12.21	6.58
noIncome	239	1376.88	873.67	13.85	8.23
employed	24277	1409.31	736.62	12.25	6.42
benefits	3930	1166.02	609.88	13.64	8.20
retired	13525	1449.47	835.22	12.43	6.29
boiler	40153	1400.63	757.59	12.40	6.47
wood	155	1496.50	1112.41	12.26	11.60
gasHeater	920	1334.16	729.35	15.09	7.84
heatPump	433	1332.10	957.46	11.31	8.93
otherHeat	310	1463.57	1096.01	13.04	8.93
dhwGasBoiler	38627	1392.54	748.72	12.34	6.34
dhwElecBoiler	872	1530.18	1059.92	14.62	10.60
dhwGeyser	1265	1535.38	826.20	15.34	7.76
dhwSolar	686	1372.63	838.39	11.04	7.58
dhwHeatPump	374	1304.40	899.81	10.97	8.85
dhwOther	147	1589.67	1089.22	13.40	7.46

Table B6: Summary statistics of the actual (specific) gas consumption per category in the energy module sample (N = 2,010). See Table A1 for variable descriptions.

		Actual gas use ($m^3 \cdot year^{-1}$)		Actual specific gas use ($m^3 \cdot m^{-2} \cdot year^{-1}$)	
Category	Count	Mean	St. Dev.	Mean	St. Dev.
owner	1319	1590.03	791.16	11.88	5.98
social	519	1035.80	497.79	12.37	6.08
private	172	1233.08	690.99	13.00	7.47
n1	637	1093.74	579.11	11.97	6.47
n2	944	1553.56	807.67	12.05	6.12
n3	193	1542.45	672.73	12.72	6.03
n4	169	1663.75	825.98	12.29	5.51
n5more	67	1563.72	637.42	11.94	5.49
A	323	1161.69	704.38	9.42	5.55
B	374	1157.05	591.05	9.72	4.26
C	663	1386.83	654.49	12.04	5.16
D	322	1571.48	776.39	14.35	8.24
E	176	1773.20	764.00	14.98	5.30
F	108	1819.43	832.22	15.68	6.17
G	44	2383.77	1336.10	16.32	5.40
apartment	556	934.60	502.66	11.66	7.00
terraced	840	1324.47	527.58	11.99	4.93
semiDetached	292	1734.71	680.26	12.86	7.91
detached	322	2199.34	940.85	12.48	5.57
before1945	360	1637.62	893.16	14.57	8.14
45to69	383	1472.20	711.47	13.82	5.78
70to89	673	1460.49	713.37	12.25	5.44
90to09	497	1248.18	710.84	9.47	4.31
after10	97	930.47	523.09	8.69	5.15
noUniversity	1503	1345.33	697.91	12.13	6.37
university	507	1626.98	882.45	12.04	5.48
noIncome	10	1042.90	504.43	9.97	5.22
employed	902	1431.21	763.89	11.81	5.60
benefits	185	1183.12	574.75	14.00	7.54
retired	913	1453.07	779.38	12.03	6.32
boiler	1963	1420.10	759.69	12.12	6.14
wood	4	1267.75	540.77	13.64	7.13
gasHeater	10	1224.00	740.34	12.75	7.65
heatPump	23	1167.26	578.04	9.87	4.98
otherHeat	10	1509.20	955.03	12.39	9.53
dhwGasBoiler	1891	1406.04	744.81	12.05	5.98
dhwElecBoiler	18	1317.56	764.49	15.59	13.63
dhwGeyser	26	1990.00	1122.84	15.70	8.09
dhwSolar	47	1517.43	887.07	11.56	7.18
dhwHeatPump	21	1236.43	554.38	10.39	4.81
dhwOther	7	2192.00	1233.63	13.18	5.85
away	149	1142.92	652.44	10.71	5.12
home	805	1555.77	783.74	12.44	6.55
depends	1056	1348.69	733.25	12.05	5.95

C Theoretical gas use

Table C1 displays the values of the theoretical and actual electricty that were read from Figure 10 of Chapter 2 of Majcen (2016). The ratio was used to estimate theoretical electricity consumption, $Q_{\text{electricity}}$. Table C2 shows the mean and standard deviation of the theoretical (specific) gas consumption per category in the energy module sample, as well as the frequency (count) of the categories. For example, there are 1,319 owner-occupied dwellings in the energy module sample, and their average theoretical gas use is 2,460 m^3 per year. Tables C3 to C5 contain additional regression tables for different dependent variables: the theoretical (specific) gas use, similarly to what was done by Majcen et al. (2015) (Table C3), the EPG per m^2 floor area (Table C4), and the EPG and EPG per m^2 for the whole energy module sample, i.e. over- and underpredictions combined (Table C5).

Table C1: Actual and theoretical electricity consumption read from Figure 10 of Chapter 2 of Majcen (2016). The ratio is used to estimate theoretical electricity consumption, $Q_{\text{electricity}}$.

Label	A	B	C	D	E	F	G
Theoretical electricity use (kWh/year)	1300	1050	1000	1000	1000	1000	1100
Actual electricity use (kWh/year)	2900	2650	2750	2900	2900	1800	2750
Ratio	0.45	0.40	0.36	0.34	0.34	0.56	0.40

Table C2: Summary statistics of the theoretical (specific) gas consumption per category (N = 2,010). See Table A1 for variable descriptions.

		Theoretical gas use ($m^3 \cdot year^{-1}$)		Theoretical specific gas use ($m^3 \cdot m^{-2} \cdot year^{-1}$)	
Category	Count	Mean	St. Dev.	Mean	St. Dev
owner	1319	2460.26	1333.70	18.01	7.81
social	519	1513.06	638.90	17.80	6.74
private	172	1802.49	1027.44	18.25	7.59
n1	637	1756.69	906.06	18.50	7.20
n2	944	2379.55	1388.17	18.00	8.01
n3	193	2149.08	996.26	17.26	6.71
n4	169	2419.83	1250.69	17.20	6.30
n5more	67	2259.16	1459.64	16.82	8.40
A	323	1306.44	729.18	10.21	4.09
B	374	1610.29	814.07	13.15	3.31
C	663	2037.55	852.42	17.20	3.33
D	322	2530.41	989.17	22.61	7.38
E	176	3114.34	1146.29	25.77	4.11
F	108	3506.15	1394.00	29.33	4.87
G	44	5083.87	2678.38	34.76	6.14
apartment	556	1346.01	599.59	16.39	7.00
terraced	840	2005.50	785.68	18.02	6.54
semiDetached	292	2580.40	999.19	18.79	9.54
detached	322	3583.58	1759.60	19.86	8.20
before1945	360	2775.23	1650.92	23.52	9.07
45to69	383	2507.53	1229.56	22.98	6.62
70to89	673	2122.49	997.64	17.22	4.79
90to09	497	1709.23	949.37	12.75	4.38
after10	97	1061.92	469.96	9.73	3.85
noUniversity	1503	2041.28	1123.58	17.91	7.69
university	507	2509.58	1484.53	18.18	7.06
noIncome	10	1605.88	629.15	15.32	7.92
employed	902	2142.66	1150.34	17.48	6.90
benefits	185	1665.11	827.85	19.14	7.57
retired	913	2282.16	1369.41	18.26	8.07
boiler	1963	2167.58	1235.35	18.05	7.47
wood	4	2327.32	679.88	24.18	9.49
gasHeater	10	2568.41	1071.79	24.98	9.48
heatPump	23	1283.60	1596.14	8.50	4.45
otherHeat	10	2091.39	1198.09	15.71	5.60

Table C3: Estimates of the regression coefficients by OLS on the theoretical gas consumption [m^3] and theoretical specific gas use [$m^3 \cdot m^{-2}$]. The standard errors are given in parentheses. See Table A1 for variable definitions, units and reference categories.

	Dependent variable:	
	gasT	*gasSpecT*
owner	−31.565 (32.706)	0.074 (0.263)
private	−68.850 (45.691)	−0.235 (0.368)
n2	−21.128 (30.086)	−0.289 (0.242)
n3	−109.478** (46.983)	−0.733* (0.378)
n4	−80.995 (51.321)	−1.303*** (0.413)
n5more	−202.076*** (69.237)	−1.536*** (0.558)
university	18.669 (28.633)	0.272 (0.231)
employed	115.853 (161.749)	−0.157 (1.302)
benefits	92.457 (165.243)	0.009 (1.331)
retired	204.378 (161.808)	0.276 (1.303)
dispInc	7.660 (5.761)	−0.018 (0.046)
floorArea	11.663*** (0.324)	−0.047*** (0.003)
value	4.309*** (1.224)	0.025** (0.010)
B	413.304*** (40.914)	2.382*** (0.329)
C	804.300*** (41.480)	5.834*** (0.334)
D	1,338.400*** (49.969)	10.863*** (0.402)
E	1,763.502*** (57.110)	13.558*** (0.460)
F	2,174.142*** (66.023)	16.773*** (0.532)
G	3,258.829*** (89.470)	22.617*** (0.720)
terraced	242.002*** (30.962)	2.211*** (0.249)
semiDetached	416.979*** (42.278)	4.236*** (0.340)
detached	835.413*** (47.076)	6.871*** (0.379)
45to69	−40.312 (38.377)	−0.660** (0.309)
70to89	−33.239 (38.348)	−1.282*** (0.309)
90to09	−150.982*** (44.569)	−1.618*** (0.359)
after10	−85.349 (69.477)	−2.437*** (0.559)
wood	−343.156 (254.833)	−0.367 (2.052)
gasHeater	−243.267 (170.495)	−0.608 (1.373)
heatPump	−349.466 (368.984)	−0.870 (2.971)
otherHeat	−280.777* (165.896)	−2.068 (1.336)
dhwElecBoiler	20.667 (120.852)	1.268 (0.973)
dhwGeyser	70.850 (106.806)	1.012 (0.860)
dhwSolar	−82.640 (75.089)	−0.218 (0.605)
dhwHeatPump	136.355 (385.028)	−0.999 (3.100)
dhwOther	92.121 (201.529)	−0.675 (1.623)
home	−12.107 (53.386)	0.247 (0.430)
depends	24.560 (48.830)	0.282 (0.393)
tempDay	−9.277 (6.836)	−0.111** (0.055)
tempNight	2.554 (7.446)	−0.001 (0.060)
Constant	−523.621** (211.941)	17.114*** (1.707)
Observations	2,010	2,010
R^2	0.842	0.721
Adjusted R^2	0.839	0.716
Residual Std. Error (df = 1970)	498.443	4.014
F Statistic (df = 39; 1970)	268.854***	130.816***
Note:		*p<0.1; **p<0.05; ***p<0.01

Table C4: Estimates of the regression coefficients by OLS of the EPG per m^2 in case of overprediction or underprediction. The standard errors are given in parentheses. See Table A1 for variable definitions, units and reference categories.

	Dependent variable:	
	EPGm2	
	Overpredictions	Underpredictions
owner	-0.122 (0.279)	-0.029 (0.788)
private	-0.244 (0.401)	-0.775 (0.991)
n2	-1.106^{***} (0.252)	1.477^{*} (0.808)
n3	-1.967^{***} (0.404)	0.346 (1.107)
n4	-1.762^{***} (0.447)	1.942^{*} (1.151)
n5more	-1.883^{***} (0.584)	-0.963 (1.742)
university	-0.004 (0.240)	0.138 (0.763)
employed	-0.377 (1.411)	4.063 (3.513)
benefits	-1.442 (1.439)	3.982 (3.615)
retired	-0.717 (1.409)	5.374 (3.530)
dispInc	-0.047 (0.049)	-0.287^{**} (0.142)
floorArea	-0.005^{*} (0.003)	0.042^{***} (0.011)
value	-0.007 (0.010)	-0.007 (0.033)
B	1.215^{***} (0.371)	1.417^{*} (0.841)
C	2.883^{***} (0.381)	1.199 (0.828)
D	5.584^{***} (0.445)	0.419 (1.422)
E	7.445^{***} (0.493)	0.233 (2.231)
F	10.518^{***} (0.562)	5.610^{**} (2.664)
G	14.556^{***} (0.729)	
terraced	0.054 (0.264)	0.762 (0.822)
semiDetached	0.300 (0.354)	0.338 (1.194)
detached	2.176^{***} (0.392)	-2.734^{**} (1.374)
45to69	0.485 (0.314)	1.349 (1.292)
70to89	-0.117 (0.319)	0.591 (1.162)
90to09	0.152 (0.379)	1.071 (1.260)
after10	0.875 (0.655)	1.890 (1.445)
wood	-1.921 (2.299)	1.215 (4.929)
gasHeater	2.016 (1.341)	
heatPump	5.846^{**} (2.881)	-2.405^{*} (1.370)
otherHeat	-0.475 (1.365)	-6.465 (6.235)
dhwElecBoiler	1.461 (1.022)	-3.048 (3.621)
dhwGeyser	-1.642^{*} (0.899)	-5.972^{**} (2.784)
dhwSolar	-0.028 (0.645)	-0.615 (1.802)
dhwHeatPump	-5.375 (3.284)	
dhwOther	-3.869^{**} (1.575)	
home	0.007 (0.451)	-2.695^{**} (1.354)
depends	-0.024 (0.412)	-2.033 (1.238)
tempDay	-0.256^{***} (0.057)	0.029 (0.185)
tempNight	-0.166^{***} (0.064)	-0.150 (0.184)
Constant	13.567^{***} (1.854)	-9.404^{**} (4.589)
Observations	1,709	301
R^2	0.472	0.203
Adjusted R^2	0.460	0.098
Residual Std. Error	3.883 (df = 1669)	4.610 (df = 265)
F Statistic	38.273^{***} (df = 39; 1669)	1.931^{***} (df = 35; 265)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table C5: Estimates of the regression coefficients by OLS on the EPG and EPG per m^2 in the whole energy module sample. The standard errors are given in parentheses. See Table A1 for variable definitions, units and reference categories.

	Dependent variable:	
	EPG	*EPGm2*
owner	−10.343 (42.096)	−0.004 (0.272)
private	-121.177^{**} (58.809)	−0.118 (0.381)
n2	-73.511^{*} (38.724)	-1.253^{***} (0.251)
n3	-223.388^{***} (60.471)	-1.879^{***} (0.391)
n4	-179.144^{***} (66.055)	-2.131^{***} (0.427)
n5more	-246.789^{***} (89.115)	-1.738^{***} (0.577)
university	3.195 (36.853)	−0.108 (0.238)
employed	44.197 (208.187)	−1.072 (1.347)
benefits	−113.799 (212.684)	−2.083 (1.376)
retired	108.028 (208.263)	−1.493 (1.348)
dispInc	-22.754^{***} (7.416)	−0.009 (0.048)
floorArea	8.706^{***} (0.416)	-0.006^{**} (0.003)
B	371.022^{***} (52.660)	0.626^{*} (0.341)
C	565.897^{***} (53.389)	2.182^{***} (0.345)
D	912.265^{***} (64.315)	4.903^{***} (0.416)
E	$1,185.744^{***}$ (73.506)	6.863^{***} (0.476)
F	$1,558.965^{***}$ (84.978)	9.597^{***} (0.550)
G	$2,321.427^{***}$ (115.157)	14.168^{***} (0.745)
terraced	44.663 (39.851)	0.168 (0.258)
semiDetached	−11.522 (54.416)	0.284 (0.352)
detached	157.810^{***} (60.591)	2.242^{***} (0.392)
45to69	19.844 (49.395)	0.433 (0.320)
70to89	−70.784 (49.358)	−0.178 (0.319)
90to09	−20.665 (57.365)	0.082 (0.371)
after10	127.191 (89.423)	0.246 (0.579)
wood	−458.301 (327.995)	−2.374 (2.122)
gasHeater	334.504 (219.444)	1.953 (1.420)
heatPump	$1,147.260^{**}$ (474.919)	6.027^{**} (3.073)
otherHeat	−118.001 (213.525)	−0.991 (1.382)
dhwElecBoiler	104.594 (155.549)	1.389 (1.007)
dhwGeyser	-366.749^{***} (137.470)	−0.663 (0.890)
dhwSolar	−79.354 (96.648)	0.001 (0.625)
dhwHeatPump	$-1,443.469^{***}$ (495.569)	−4.517 (3.207)
dhwOther	-539.507^{**} (259.388)	-3.468^{**} (1.678)
value	-2.903^{*} (1.576)	−0.006 (0.010)
home	−40.448 (68.713)	0.319 (0.445)
depends	10.828 (62.848)	0.256 (0.407)
tempDay	-35.965^{***} (8.799)	-0.242^{***} (0.057)
tempNight	-24.827^{***} (9.584)	-0.131^{**} (0.062)
Constant	373.759 (272.789)	13.538^{***} (1.765)
Observations	2,010	2,010
R^2	0.529	0.418
Adjusted R^2	0.520	0.406
Residual Std. Error (df = 1970)	641.546	4.151
F Statistic (df = 39; 1970)	56.791^{***}	36.229^{***}
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$