Essays on the Empirical Analysis of Residential Energy Demand

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln
2021

vorgelegt von
M.Sc. Sebastian Mertesacker
aus Duisburg

Referent: Prof. Van Anh Vuong, Ph.D.

Korreferent: Prof. Dr. Johannes Münster

Tag der Promotion: 23. April 2021

Acknowledgments

There are many people that supported me over the years and to which I want to express my highest gratitude.

First, I want to thank Van Anh Vuong for supervising this thesis, starting an exciting research project with me and sticking to it even at times when progress was slow. Her support has been extremely valuable at all stages of the projects and I will always appreciate it.

Officially, Johannes Münster is the second supervisor of this thesis, but he has truly been much more than that. He provided countless advice and inspiration for my research, my study programme and my general thinking about economics since the beginning of my master studies. This guidance has been extremely valuable and I will always be thankful for it.

Further, I will always be grateful to Felix Höffler for giving me the opportunity to work at his chair. It allowed me to teach two of the most fascinating and best prepared courses I know. It was a pleasure to work for him and the chair and to enjoy his sharp reasoning and subtle irony.

I have been lucky to work in a team with two kind, able and supportive colleagues, Ute Büttner and Razi Farukh, for most of my time at the chair. We overcame some obstacles and hurdles and it has always been good to know you on my side.

Many thanks also to Marc Oliver Bettzüge, who repeatedly provided flexible and pragmatic support in teaching the curriculum of the chair. He also gave valuable suggestions to my research on several occasions. I am very thankful for all the support I received.

I would also like to thank Paul Gorny, Anna Kerkhof, Jonas Löbbing and Alexander Matuschke for the joy they brought to the first year of doctoral courses. Anna Kerkhof and Jonas Löbbing have been the friends who shared successes and failings during the dissertation. They were always there for a chat, for advice or for helping out, even with a place to sleep when trains stopped their service. Thanks a lot for your company on this journey.

This thesis would be unthinkable without the endless support of my parents. Thank you so much for creating the environment for me to become the person who I am. Many thanks also to my brother, whose education to a building engineer has proved extremely useful. I also thank my wife for her patience and for always keeping me aware of what is really important in life.

Contents

Introduction **1**

Overview of the thesis 2

Contribution to chapter 3 6

1 Introducing engineering knowledge into the econometric analysis of households' heat demand **7**

1.1 Introduction 7

1.2 Related literature using engineering knowledge for empirical analyses of residential heat demand 10

1.3 An economist's look into engineering fuel requirement calculations 12

1.3.1 A generic representation of fuel requirement calculations 13

1.3.2 The functional form of the input demand equation 14

1.3.3 Validity of fuel requirement calculations 18

1.4 Empirical specification 20

1.4.1 Obtaining an empirical specification if the true efficiency state is observed 20

1.4.2 Allowing for measurement errors in the generated efficiency states 21

1.5 Data 25

1.5.1 The German Residential Energy Consumption Survey 25

1.5.2 Generation of efficiency states 27

1.6 Empirical analysis 30

1.7 Conclusion 37

2 A structural analysis of households' heat consumption **39**

2.1 Introduction 39

2.2 A household production model of residential heat demand 42

2.2.1 The input demand for thermal comfort 43

2.2.2 Households' utility from indoor temperature 44

2.2.3 Model Solution 47

2.3 Empirical specification 50

2.4 Data . . . 53
2.5 Empirical analysis . . . 55
2.5.1 Estimation of the structural model . . . 55
2.5.2 Predictions of mean temperature choices and elasticities . . . 61
2.5.3 Economic interpretation and heterogeneity analysis . . . 63
2.6 Conclusion . . . 66

3 Households' dynamic investment in domestic energy efficiency 69
3.1 Introduction . . . 69
3.2 Theoretical model . . . 72
3.2.1 Energy efficiency and thermal heat consumption . . . 72
3.2.2 Households' dynamic investment choice . . . 75
3.3 Data . . . 77
3.4 Estimation and empirical results . . . 79
3.4.1 Estimation of the period utility function . . . 79
3.4.2 Estimation of the evolution process . . . 82
3.4.3 Estimation of investment cost . . . 84
3.4.4 Simulation and policy analysis . . . 90
3.5 Conclusion . . . 92

Appendices 95

A Appendix to chapter 1 97
A.1 Robustness analyses and descriptives . . . 97
A.2 Theorems . . . 108
A.3 Details on the generation of efficiency states . . . 109
A.3.1 Detailed descriptions of the outlier corrections applied to the generated efficiency states . . . 109
A.3.2 Detailed analyses of the generated fuel requirements . . . 112

B Appendix to chapter 2 117
B.1 Robustness analyses . . . 117

C Appendix to chapter 3 125
C.1 Descriptives of actual monetary cost of retrofitting in Germany . . . 125

Bibliography 129

List of Tables

1.1 Coefficients of naive regressions . . . 30
1.2 Estimated coefficients of the full regression model . . . 32

2.1 Regression coefficients for different specifications of the adjustment term . . . 57
2.2 Model Predictions . . . 61
2.3 Heterogeneity of households in their preferences for and choices of temperature increases . . . 65

3.1 Estimates of utility function and adjustment term parameters . . . 81
3.2 Estimates of evolution process . . . 83
3.3 Estimates of modernisation costs . . . 86
3.4 Model predictions . . . 89
3.5 Simulation of counterfactual policy scenarios . . . 91

A.1 Coefficients of naive regression using the Poisson estimator . . . 97
A.2 Estimated coefficients of the full regression model using the Poisson estimator . . . 98
A.3 Robustness checks using alternative data restrictions . . . 101
A.4 Regression coefficients using an alternative dependent variable . . . 105
A.5 Descriptive statistics over different variable restrictions for the year 2008 . . . 107
A.6 Coefficients of regression of predicted fuel consumption on dwelling characteristics 113

B.1 Regression coefficients for different calibrations of $\overline{\tau}^{in}$ . . . 117
B.2 Model Predictions for different calibrations of $\overline{\tau}^{in}$ . . . 120
B.3 Model Predictions for regressions under different data restrictions . . . 121
B.4 Regression coefficients obtained under different data restrictions . . . 122

C.1 Descriptives of monetary modernisation costs of home owners based on SOEP data between 2010 and 2015 . . . 126

List of Figures

1.1 Distribution of generated fuel requirements 29

2.1 Distribution of predicted elasticities 64

3.1 Density of an exponential distribution with mean γ 86

A.1 Boxplot of outliers in predicted fuel requirements 109
A.2 Boxplot of implied indoor temperatures 110
A.3 Boxplot of fuel consumption and predicted fuel requirements 111
A.4 Boxplot of distances to the ideal temperature level 111
A.5 Generated fuel requirements of dwellings constructed before and after 1969 . . . 112

C.1 Distribution of monetary modernisation cost of home owners based on SOEP data between 2010 and 2015 128

Introduction

The ability of households to consume large amounts of energy at low cost is an essential reason for the high level of well being of individuals in modern societies. It allows people to enjoy thermal warmth when outdoor temperatures are low, to use artificial light after the sun has set and to operate myriads of devices that make life more interesting, comfortable and enjoyable.

However, the discovery of men made climate change has revealed a substantial cost that can be associated to the consumption of energy. The emittance of carbon dioxide into the atmosphere during the generation of energy from conventional sources, increases global temperature levels and threatens to make the planet a less desirable place to live (Pachauri et al., 2014). These cost make it a primary policy goal to set incentives for households to switch to carbon free energy sources and to control and – if possible – reduce the consumption of energy, despite its eminent effect on the welfare of individuals.

To mitigate global warming, policy makers around the world have set ambitious targets to reduce greenhouse gas emissions. Until 2030 the European Commission wants to reduce greenhouse gas emissions by at least 40 % over 1990 levels and has proposed to raise the reduction target further to 55 % in September 2020 (European Commission, 2020a). In its long-term strategy it furthermore defines the goal to achieve climate neutrality until 2050 (European Commission, 2020b).

Being responsible for 26.1 % of the final energy consumption in the European Union in 2018 (Eurostat, 2020), the residential energy sector plays an important role to meet these climate targets. The design of good policies to achieve emission reductions in the residential sector requires a detailed understanding of the mechanisms that determine households' behavior. Only a good understanding of how households make their decisions and how they react to incentives allows to assess government programs with respect to their effectiveness to achieve the desired goals and with respect to their welfare effects more generally.

The research presented in this thesis contributes to the understanding of residential energy consumption. The analyses focus on the demand for thermal heat, which allows to model and empirically estimate households' choices in greater detail compared to broader studies on domestic energy consumption. In 2018 space heating was responsible for 63.6 % of final domestic energy consumption in the European Union (Eurostat, 2020), which makes it the natural first choice of an in depth analysis.

Overview of the thesis

The three chapters of this thesis analyse two types of household choices relevant for the final consumption of heating energy. First, chapters 1 and 2 consider households' heat consumption in a static environment, in which the characteristics of the dwelling, such as the level of thermal insulation and the type of heating system used, are given. In this short term perspective households determine the final amount of energy used for heating only via their choice of the level of indoor temperature in the dwelling, the number of rooms and time of the day it should be heated and other factors affecting the level of thermal comfort experienced in the dwelling. Households' choices vary with the cost associated to the consumption of thermal comfort, their disposable income and individual preferences, which might differ across households depending on their socioeconomic characteristics. Understanding in detail how these choices are made is essential to understand how different households react to specific changes in their environment, which potential effects on their individual welfare might be associated to these changes, and how they contribute to the intended policy targets to reduce carbon emissions. The first two chapters contribute to this understanding, by empirically studying the factors and mechanisms that drive households' decisions.

In contrast, in chapter 3 households' decisions are analysed in a dynamic environment, in which they are able to improve the energetic performance of their dwellings via modernisation investments. Incentivising households to modernise the building stock, is the most important area of activity for policies targeting emission reductions in the residential heating sector. It provides opportunities to save energy without households having to give up on thermal comfort and thus individual utility. However, the details of the decision process households face are involved. Households face high upfront cost in the period a modernisation decision is made that have to be traded off against lower energy cost and higher utility from the consumption of thermal comfort in future periods. Understanding the investment decisions households make, requires an understanding of the monetary and non-monetary cost households face, their expectations about the efficiency gains that can be realized and the ultimate gain in utility associated to it. To study all these factors a dynamic investment model is developed and estimated in chapter 3.

The aim of the analyses in all chapters is to empirically study households' choices that determine the final demand for thermal heat in detail. For this purpose, households' behavior is modelled carefully making use of reliable external information where possible. This imposes additional structure to the regression equation that helps to identify the fundamental parameters in households' decision process and thus provides clear interpretations of their meaning and the basis for counterfactual policy scenarios.

A challenge in the analysis of households' heat demand is, that it is a function of many input variables representing households' choices for thermal warmth as well as the thermal characteristics of the dwellings they inhabit, that potentially interact in nonlinear ways. This makes it difficult to acquire all the relevant data and to specify a regression equation that correctly captures the interdependencies of dwelling properties and household choices and allows to identify the parameters of interest.

The research conducted in chapter 1, **"Introducing engineering knowledge into the econometric analysis of households' heat demand"**, is based on the observation that engineering models contain substantial information on how dwelling characteristics interact with households' choices to determine the amount of fuel used for heating, that could be useful to address the aforementioned problems. These models commonly combine data on the characteristics of a dwelling with assumptions on household behavior to predict the amount of fuel the dwelling requires within a year. Different to the economic analysis of households' heat demand the models do not focus on finding explanations *how* households' choices for thermal comfort are made. In chapter 1, I develop an empirical model of households' fuel demand that employs the information contained in the engineering model efficiently, by using it to fix the effect of dwelling characteristics on fuel demand, while estimating households' choices from the observed fuel data. The approach makes use of the engineering knowledge, where it is expected to provide reliable guidance on households' individual fuel consumption, and uses the observed fuel consumption to replace assumptions on households' choices by empirical estimates that vary with households' individual cost in the consumption of thermal comfort, their income and socioeconomic characteristics.

The proposed model introduces engineering knowledge in a more comprehensive way into the empirical analysis than previous approaches. It explicitly considers the different types of information contained in the underlying model by Loga et al. (2005) and carefully evaluates how it can be used in the econometric analysis. Importantly, these analyses reveal that empirical specifications based on engineering models should control for potential systematic errors in the engineering prediction, to obtain reliable results. For this purpose, I suggest a simple approach to proxy for systematic errors in the engineering model.

I estimate the developed regression model using German survey data from the time period between 2006 and 2008 provided by RWI and forsa (2016). The price elasticity is found to be roughly -0.66. Once socioeconomic characteristics are included in the regression and controls for potential errors in the underlying engineering model are added, there is only a relatively weak positive effect of income on household behavior. This contrasts previous studies, which have found very strong income effects using models that relate observed and predicted fuel consumption to households' income in a less rigorous way. Households' age and size are found to have strong positive effects on fuel consumption. Ownership status does not explain consumption, once the model controls for systematic errors in the engineering prediction.

Chapter 2, **"A structural analysis of households' heat consumption"**, extends the analysis from chapter 1, by explicitly deriving households' optimal consumption of thermal comfort from a theoretical economic model, instead of specifying it as a reduced form of explanatory variables. For this purpose, the relationship between dwelling characteristics, household behavior and fuel consumption, derived in chapter 1, is used to define the input demand function in a household production model of heat demand. In terms of the model, this function defines the amount of fuel the household has to use to consume the desired level of thermal warmth and represents a constraint in his consumption behavior.

I explicitly model the utility households receive from the consumption of thermal warmth to have a satiation point, describing the choice of a household facing no monetary constraints. Deriving the regression equation based on the solution to the resulting decision problem, this introduces structure on the estimation, that is used to identify additional parameters that cannot be identified in the ad hoc specification of households' temperature choice used in chapter 1. Concretely, the empirical model of chapter 2 allows to separately identify the effect of one explanatory variable on both, households' preferences for thermal comfort and the size of the error in engineering predictions of a dwelling's fuel requirement. Importantly, this allows to calculate the level of households' temperature choices, instead of its relative changes with observed covariates only, which allows interpretations of the estimated coefficients in terms of the theoretical model of households' heat consumption.

The structural empirical model developed in chapter 2 shows that engineering knowledge can be introduced in an economic analysis of households' temperature choice in a theoretically consistent way. The explicit link of the regression equation to the theoretical model, implies that the estimated parameters have a clear economic interpretation in terms of households' utility from thermal comfort. Larger and more educated households are found to have stronger preferences for high indoor temperatures. In contrast, there is no evidence for a positive effect of income on temperature choices.

In addition, the analysis of chapter 2 provides a novel approach to the estimation of the elasticity of temperature choices with respect to price changes. Instead of estimating the elasticity through a separate coefficient on a measure of households' marginal heating cost, it is derived theoretically from the structural model of heat consumption. Given the estimates of households' preferences for thermal comfort, the elasticity can be predicted for every household in the sample. The analysis of predicted elasticities reveals that there is substantial heterogeneity across households. The median elasticity in the sample is -0.214, implying that there are many households that are fairly inelastic. Few substantially more elastic households result in a mean elasticity of -0.302, which confirms earlier results in the literature (Aydin et al., 2017; Sorrell et al., 2009). The theoretical model explicitly pins down the source of the heterogeneity in households' elasticity. Households' with higher preferences for thermal comfort are less elastic towards price changes, because they consume temperature levels closer to the satiation point before the price change.

In chapter 3 of the thesis, **"Households' dynamic investment in domestic energy efficiency"**, the focus of the analysis is switched from the study of households' decisions, given the characteristics of the dwelling they inhabit, to the decision how they invest into superior properties of the dwelling, which require less fuel for the consumption of the same amount of thermal comfort. To analyse this decision in detail, a dynamic model is developed, that explicitly differentiates between the (large) one time investment cost households face in the period the investment is conducted and the discounted long-run gains in period utility they enjoy in future periods. To define the period utility households' receive from the consumption of thermal warmth and other goods, the utility function developed in chapter 2 of the thesis is applied. This ensures, that the dynamic model explicitly accounts for the increase in households' period utility after an efficiency increase, that results from the monetary savings associated to the reduction in heating cost as well as the substitution towards the now relatively cheaper good thermal heat. That is, in the model the (positive) impact of households' rebound behavior on their period utility is endogenized in the investment decision.

We estimate the period utility function based on the empirical model applied in chapter 2 and predict the improvements in energy efficiency after modernisations using an engineering model by Loga et al. (2005). This allows to calculate the long-run utility gain associated to the alternatives to invest or not to invest for every household in the sample. Given the assertion that households only invest if the associated utility gains exceed the cost, the cost of investing are then estimated such they rationalize the modernisations observed in the sample.

A virtue of the dynamic framework is that it allows to disentangle benefits from the investment from the associated cost, which is not possible with static utility models, such as logit or probit. The clear association of estimated coefficients to underlying parameters of households' decision process is crucial to perform counterfactual policy scenarios and thus to assess potential policy programs with respect to their impact. Also, model quantities such as the temperature choice, the expected gain in lifetime utility associated to an investment and the resulting investment probability can be predicted and thus quantified for every household in the sample. This allows a much more detailed study of households' incentives to invest compared to static utility models.

Our results show that households' valuation for thermal comfort as well as the marginal cost associated to its consumption matter for households' temperature and investment choice. A household living in a less efficient dwelling consumes lower temperature levels and has, ceteris paribus, higher incentives to invest. A household with stronger preferences for high temperature levels, e.g. an older household, chooses a higher mean indoor temperature and has stronger incentives to retrofit the dwelling to further decrease the cost of temperature consumption. The results thus clarify the importance of understanding the sources of heterogeneity in households' static temperature choice to achieve a detailed understanding of the mechanisms that determine their investment choices.

We simulate three policy scenarios that could be introduced to increase households' retrofit investments and decrease overall fuel consumption in the economy. We find that direct subsidies to the purchasing cost increase investment rates, but that this does only little to decrease households' energy consumption. To the contrary, an increase in energy prices, for instance via a tax increase, creates high energy saving incentives for households. This leads to an increase in the investment rate by 22.1 % and a reduction of households' mean temperature choice by 4.8 %.

Contribution to chapter 3

While chapters 1 and 2 are based entirely on my own work, chapter 3 results from joint work with Van Anh Vuong. The initial research idea was developed by Van Anh Vuong. I contributed the preparation of the data, the development of the static model and the estimation of purely monetary investment cost from a second data source. Van Anh Vuong was leading on the development of the dynamic investment model and its implementation and estimation in MATLAB, which I supported with inspections of the model and code and suggestions for improvements. I wrote the first version of the chapter, which we both revised and edited multiple times afterwards.

1 Introducing engineering knowledge into the econometric analysis of households' heat demand

1.1 Introduction

Residential energy demand for space heating is a major contributor to the overall energy demand in the economy. In 2018 the residential energy sector was responsible for 26.1 % of total final energy consumption in Europe. Roughly 63.3 % of the energy was consumed to heat residential dwellings (Eurostat, 2020). Consequently, understanding the factors that determine households' heat demand is important for policy makers interested in reducing energy consumption of the economy to meet climate targets.

Economic theory analyses households' fuel demand in household production models. In these models the good fuel is considered an input factor used in the production of the commodity thermal comfort, which directly enters households' utility function. That is, the amount of fuel, $F_{i,t}$, demanded by a household, i, in period t, results from the input required for the production of the consumed indoor temperature, $\tau_{i,t}$, given the characteristics of the dwelling stored in a vector, $\boldsymbol{D_{i,t}}$:

$$F_{i,t} = I(\tau_{i,t}, \boldsymbol{D_{i,t}}). \tag{1.1}$$

Explaining households' fuel demand thus requires an understanding of their decision for thermal comfort, as well as knowledge of the input demand function $I(\tau_{i,t}, \boldsymbol{D_{i,t}})$.

The focus of the economic analysis is the study of households' behavior. However, the challenge for the econometric analyses of households' heat demand is that households' choices cannot easily be studied independently from the input demand equation, primarily for two reasons. First, households' temperature choices are rarely observed in empirical data, such that they cannot be studied directly. Instead, most empirical studies analyse households' heat demand from observed fuel consumption. Second, from the perspective of the decision making household, the input demand function $I(\tau_{i,t}, \boldsymbol{D_{i,t}})$ represents the technology available to produce thermal comfort. It determines the marginal cost of temperature consumption and thus constraints households' choice set together with their budget constraints. Consequently, it directly enters into the temperature choice of interest in the econometric study (see Pollak and Wachter, 1975).

Without knowledge of the functional form of $I(\tau_{i,t}, \boldsymbol{D_{i,t}})$ it is thus difficult to study households' demand for thermal heat. A common approach of the literature is to regress logarithmised fuel demand on a linear function of fuel prices, income, socioeconomic characteristics of the household and dwelling characteristics (see, e.g., Rehdanz, 2007). However, in this ad hoc approach it is unclear, whether the regression equation is correctly specified or whether the implicit assumptions on the input demand function are invalid. Also, this regression approach does not allow to introduce the true marginal cost of temperature consumption into the regression equation. Most fundamentally, since parameters of the input demand function $I(\tau_{i,t}, \boldsymbol{D_{i,t}})$ as well as households' preferences have to be estimated from the fuel data in the same regression and since they interact in unknown ways – directly via equation (1.1) and indirectly via their impact on the optimal temperature choice – they are likely to be mixed up in estimates from a standard linear regression. This poses a problem to the detailed understanding of the mechanisms determining residential energy consumption and in particular prohibits the calculation of counterfactual policy scenarios (Lucas, 1976).

The discussion clarifies that using engineering knowledge on the functional form of $I(\tau_{i,t}, \boldsymbol{D_{i,t}})$ can be very useful. It could avoid misspecification of the regression equation, increase precision of the estimation by utilising existing knowledge and provide clear interpretations for the estimated parameters providing the foundation for counterfactual policy scenarios.

Engineers have substantial knowledge on the functional form of interest. In fuel requirement calculations for dwellings, they commonly combine data on the characteristics of a dwelling with assumptions on household behavior to predict the amount of fuel the dwelling requires within a year. That is, they provide very detailed descriptions of the input demand function $I(\tau_{i,t}, \boldsymbol{D_{i,t}})$. Different to the economic analysis of households' heat demand the models do not focus on finding explanations *how* households' temperature choices are made.

The goal of this paper is to show how engineering knowledge can be used to improve specifications of energy demand equations in applied empirical work. For this purpose, I conceptually decompose fuel requirement calculation procedures into two components: the aggregation of dwelling characteristics, $\boldsymbol{D_{i,t}}$, to the highest possible level that does not require assumptions on households' temperature choice – providing an overall state variable, $s_{i,t}$, of the dwellings' efficiency level – and the function $I(\tau_{i,t}, s_{i,t})$, that maps $s_{i,t}$ and $\tau_{i,t}$ into a fuel demand. The decomposition allows to employ the engineering knowledge efficiently by using it to obtain a) state variables that make efficiency levels comparable across dwellings and b) a functional form for the empirical analysis, while allowing households' choices to be determined by the data. That is, it allows to make use of the engineering knowledge, where it is expected to provide reliable guidance on households' individual fuel consumption and to replace assumptions on households' choices by empirical estimates obtained from observed fuel data.

Concretely, I make three major contributions to the existing literature. First, I use a novel approach to generate efficiency states based on dwelling characteristics frequently observable in micro datasets, using a simplified fuel requirement calculator developed by Loga et al. (2005). Their calculator has originally been designed to facilitate the calculation of energy performance certificates for home owners. The idea to use the simplified procedure to calculate fuel requirements for households in a micro dataset used for empirical analyses, greatly simplifies the use engineering predictions in empirical studies, as it does not require these to be observed as part of the data. Second, I also derive the concrete functional form of the input demand function, $I(\tau_{i,t}, s_{i,t})$, from the engineering model by Loga et al. (2005). I show that it is linear in households' temperature choice and thus supports linearity assumptions implicitly made in much of the related literature. Finally, I show how, based on the input demand equation, an empirical specification can be obtained, that controls for potential flaws in the engineering model that are responsible for its tendency to systematically overpredict actual consumption.

I estimate fuel demand as a function of household characteristics, income and the marginal heating cost using data from "The German Residential Energy Consumption Survey" (RWI and forsa, 2016). The results indicate limitations of existing approaches to use engineering knowledge for the empirical analysis of households' heat demand, presented in section 1.2. It is important to control for household characteristics in the empirical analysis of households' temperature choices and to allow for potential errors in the engineering calculations, both of which is rarely done in the previous literature. The inclusion of household characteristic substantially reduces the impact of income on temperature choice. Once the model controls for potential systematic errors in the engineering model, income has no statistically significant effect at a 5 % confidence level and also the effect of other socioeconomic characteristics decreases.

The next section reviews prior literature, that has used engineering knowledge for the econometric analysis of residential heat demand and points out limitations the model presented in this paper aims to address. I then provide a careful discussion of fuel requirement calculations developed by engineers in section 1.3, that provides the basis for the subsequent specification of the regression model. The discussion includes a generic representation of the engineering model in section 1.3.1, the derivation of a functional form of the input demand equation in section 1.3.2 as well as an inspection of the source of systematic biases that are inherent to engineering predictions in section 1.3.3. Section 1.4 develops the empirical regression equation. In section 1.5, I present the data and give a detailed description how the efficiency states are generated, before the estimation results are reported and discussed in section 1.6. Section 1.7 concludes.

1.2 Related literature using engineering knowledge for empirical analyses of residential heat demand

Some previous work has tried to introduce engineering knowledge into the econometric analysis of heat demand. Hsueh and Gerner (1993) carefully derive an input demand equation for fuel based on heat loss equations used by engineers. They specify the demand for thermal warmth as a function of fuel prices, income and further household demographics and insert this as the behavioral component into the technical input demand equation to obtain an estimable regression equation. A problem of their approach is the lack of crucial dwelling information required for the calculation of precise heat losses. They use binary indicators, whether different surfaces have been insulated as proxies for the quality of the thermal shell and run separate regressions by fuel type to account for differences in heating system efficiency. This implies that the underlying physical properties of the dwelling are part of the estimated coefficients, as they note themselves. Furthermore, their input demand equation indicates that the linear functions describing household behavior and the technological state of the dwelling are multiplicatively related. The large number of interactions of explanatory variables that would result, forces the authors to simplify the functional form derived from engineering knowledge to avoid multicollinearity. In addition, their specification also does not explicitly consider the indirect effect of overall efficiency on fuel consumption via households' temperature choice. A virtue of the approach used in this paper is that a simplified fuel requirement calculation is used to condense the available dwelling information into aggregate variables of dwelling efficiency that interact with household behavior. The absence of the need to estimate technological states based on limited dwelling information, allows to actually specify the interactions of technology and behavior in a thorough and complete way.

Haas et al. (1998) and Haas and Biermayr (2000) are able to avoid some of the limitations of the analysis by Hsueh and Gerner (1993) by the use of better data. They also derive a functional form of the input demand from engineering principles, but observe thermostat settings and measures of insulation quality as well as heating system efficiency (the latter only in Haas and Biermayr (2000)) in their data. However, besides the strong data requirements, a limitation of their approach is that they do not model the effect of technology on household behavior.

An alternative approach of using engineering knowledge for the empirical analysis of residential heat consumption is to compare the consumption levels predicted by some engineering model to the actual measured consumption of households. The idea is that, given that the engineering predictions for all dwellings are based on a fixed set of assumptions about household behavior, deviations from the predicted reference values can be informative about households' choices of thermal comfort. In this sense, the ratio of actual and predicted consumption is often referred to as the "service factor", indicating the level of energy services consumed by the household (Haas et al., 1998).[1] A general finding of the literature is that predicted consumption exceeds actual consumption for the vast majority of households. The common interpretation is that economically constrained households consume lower levels of energy services than those assumed in engineering models. Various studies have related the service factor to households' income and the level of predicted fuel consumption, taking the latter as an indicator of the dwellings' relative cost in producing a given level of thermal comfort (Cayla et al., 2011; Aydin et al., 2017; Majcen et al., 2013; Laurent et al., 2013; Bakaloglou and Charlier, 2018). Typically, these studies simply plot the service factor against either of the two economic variables into a two-dimensional graph. They find that the service factor substantially increases with income and decreases with cost of temperature provision, consistent with a positive income and a negative price elasticity of energy service consumption, respectively.

A more rigorous analysis is provided by Dubin et al. (1986). They use the ratio of observed and predicted consumption to conduct a multivariate regression on income, an engineering estimate of marginal cost and household characteristics. This allows them to derive explicit estimates of income and price elasticities, while controlling for additional factors affecting household behavior, which are potentially correlated to income and the marginal cost of heat consumption. However, similar to the approaches discussed before, they assume the engineering model to be correct and differences between actual and predicted consumption to solely originate from differences in household behavior. This might bias estimation results if technical errors correlate to characteristics of the dwelling, which might be related to income and other variables affecting household behavior. In contrast, I propose a framework that allows to control for such errors in the engineering model, while estimating households' temperature from observable fuel data.

Dubin et al. (1986) obtain predicted consumption values by applying the thermal load model developed by Dubin and McFadden (1983) to their micro dataset. The ability to generate fuel requirements based on dwelling information available in household survey data substantially reduces the data requirements for the empirical analysis. This paper shows how a more recent

[1] Alternative names used in the literature for the ratio of actual and predicted consumption include the "heating factor" or the "intensity of use" (Laurent et al., 2013).

simplified fuel requirement calculation procedure Loga et al. (2005) developed for Germany can be used to generate engineering predictions of households' fuel consumption based on dwelling information frequently observed in micro datasets.

Finally, Aydin et al. (2017) use the predicted fuel requirements they observe in their panel of Dutch households as a proxy for dwellings' energy efficiency, which they include into a standard energy demand equation. Engineering knowledge is thus not used to disentangle technology and household preferences in their approach. The model provides a simple way to estimate rebound effects from efficiency increases, if predicted fuel requirements are observed in the data. Aydin et al. (2017) also acknowledge the possibility that the engineering predictions might not be entirely correct, introducing measurement error in their independent variable. They address this problem by an instrumental variable approach. In this paper, the engineering model is related to observed fuel demand in a theoretically consistent way. Systematic differences between actual and predicted consumption are explained by prices, income and household characteristics affecting household behavior and imperfections in the engineering models' ability to correctly calculate the efficiency state of a dwelling, which is allowed to vary with dwelling characteristics.

1.3 An economist's look into engineering fuel requirement calculations

Fuel requirement calculations are complex engineering models that combine data on dwelling characteristics and assumptions on household behavior with physical laws to calculate the amount of fuel the given dwelling is likely to consume. A fundamental component of these procedures is the calculation of the total heat loss, which is obtained by summation of transmission and infiltration heat losses across all components of the thermal shell. The behavior of the household, the physical properties of the dwelling as well as outdoor temperatures play a crucial role in determining the total amount of heat that evaporates from the dwelling to the outside. The total demand for heating energy is obtained by subtracting heat gains (e.g. from solar energy) from the total heat loss. This is the amount of useful heat consumed in the dwelling. Accounting for inefficiencies in the production and distribution of this useful heat the final energy demand is obtained.

The details of the fuel requirement calculations are complicated. They are full of technical modelling elements that are crucial for a correct prediction of fuel demands, but make it hard to incorporate them into an economic framework. The next section therefore develops a simple characterisation of such models that emphasizes the role of households' choices. Model parts less interesting from an economic perspective are interpreted as providing aggregate measures of the dwellings' efficiency that interact with the household behavior to determine the fuel demand of the household. The decomposition facilitates the empirical estimation of households' choices, while making use of the information contained in engineering models where possible. Section 1.3.2 derives an input demand function for fuel based on the fuel requirement calculation by Loga et al. (2005). Finally, section 1.3.3 discusses the performance of these procedures and how this affects their inclusion into an economic model of heat demand.

1.3.1 A generic representation of fuel requirement calculations

Let a full engineering model be described by a function $I^e(\psi^e(\boldsymbol{D^e}), \boldsymbol{\tau^e})$. The superscript e indicates the engineering assumptions on input variables and functional forms used. The model maps n^{D^e} dwelling characteristics from the vector $\boldsymbol{D^e}$ and household behavior stored in the vector $\boldsymbol{\tau^e}$, into a prediction of the amount of fuel the dwelling consumes.[2] The relevant household behavior entering the input demand equation, includes the decision on the level of indoor temperature as well as the area of the dwelling that is heated, the amount of time the temperature level is consumed and potentially other factors such as the ventilation behavior.

The dwelling characteristics required as inputs in the calculation procedure are numerous. They include detailed descriptions of the thermal shell, the size and shape of the body as well as the heating system in use. To reduce the complexity of the representation, it is useful to summarize the n^{D^e} characteristics into a vector $\boldsymbol{s^e}$ of length $n^{s^e} \leq n^{D^e}$ of elements that directly interact with households' choices. That is, $\boldsymbol{s^e}$ is constructed such that it contains the highest level of aggregation of technical characteristics, that is possible without making assumptions on the behavior of the household. It represents the sufficient statistic for the efficiency of the technology available to the household. The nested function $\psi^e(\boldsymbol{D^e})$ reflects the structure of the fuel requirement calculation that maps observed dwelling characteristics into the vector $\boldsymbol{s^e}$. That is,

$$\boldsymbol{s^e} = \psi^e(\boldsymbol{D^e}) \tag{1.2}$$

and thus

$$I^e(\boldsymbol{s^e}, \boldsymbol{\tau^e}) = I^e(\psi^e(\boldsymbol{D^e}), \boldsymbol{\tau^e}). \tag{1.3}$$

The correct specification of the function $\psi^e(\boldsymbol{D^e})$ to transfer the available data into a good approx-

[2] I focus on fuel requirement calculations for heating. While complete models determining the fuel requirement of a dwelling often also include the input demand for hot water generation, this can be analysed separately and is generally a minor contributor to the overall fuel demand.

imation of the actual physical properties of the dwelling is a difficult as well as important step of any fuel requirement calculation.

Consider the amount of heat a dwelling looses through the surface of an outer wall as an example how the function $\psi^e(\cdot)$ aggregates dwelling properties. The exact heat loss is a function of its size, environment (e.g., whether it is a detached wall) and – most importantly – the different materials forming the wall and their respective thickness and U-Values. The U-Values are physical properties of the materials indicating the thermal resistance of the body that might be known more or less precisely, depending on the level of detail of the information available. Clearly, even the calculation of the heat lost through a single wall (i.e. its "thermal performance") thus requires the combination of many different pieces of information (about its physical properties) from the vector $\boldsymbol{D^e}$ and potentially some approximations. Furthermore, the function $\psi^e(\cdot)$ typically aggregates the thermal loss across multiple walls and other properties of the dwelling up to a statistic in $\boldsymbol{s^e}$ that captures all thermal characteristics of the dwelling that are independent of the behavior of the household.

The characterisation of the fuel requirement calculation as a composite function helps to clarify how the engineering model by Loga et al. (2005) is used to empirically analyse households' choice of thermal comfort. First, the functions $\psi^e(\boldsymbol{D^e})$ and $I^e(\boldsymbol{s^e}, \boldsymbol{\tau^e})$ are used in two distinct ways. The former is employed to generate data about the efficiency of the dwellings based on observable dwelling characteristics. For this purpose, the ability of the model by Loga et al. (2005), to map basic information on the house into elementary physical properties is employed. In contrast, the latter determines the functional form of the regression equation. By using the engineering knowledge to fix the role of technology and its interaction with household behavior in the determination of the final energy demand, the factors driving households' choice of thermal warmth can be identified in an empirical analysis. Second, while I allow for errors to occur in $\psi^e(\boldsymbol{D^e})$ and discuss how empirical models can be specified to avoid these to bias estimation results in section 1.4, I assume the functional form of $I(\cdot)$ to reflect the true technology available to the household given his efficiency states.

1.3.2 The functional form of the input demand equation

I use the engineering model by Loga et al. (2005) to derive a functional form for the input demand function, $I^e(\boldsymbol{s^e}, \boldsymbol{\tau^e})$, describing the interaction between a given household behavior and the current state of the dwelling's technology.[3] Households' fuel demand from heating depends on

[3] Note, that while Loga et al. (2005) call their model a "simplified" fuel requirement calculation procedure, the simplifications involved relate to the data collection process and the calculation of the fundamental physical characteristics of the dwelling, i.e. the function $\psi^e(\boldsymbol{D^e})$ in terms of the generic characterisation developed in the previous section. They use a common functional form to map the dwelling characteristics and assumptions on

the total heat demand as well as the efficiency of the heating system used to produce that heat. Following their notation, let Q_L refer to the total heat loss of the dwelling and let Q_A denote the difference of the sum of losses from transfer, distribution and storage of the produced heat and alternative heat gains (e.g. from solar and internal sources). A generic characterisation of the amount of fuel the household needs to heat the dwelling is

$$F^e = I^e(\cdot) = (Q_L + Q_A) \cdot \xi, \tag{1.4}$$

where ξ refers to the efficiency of the heating system used. While Q_A is independent of household behavior, the heat loss, Q_L, depends on the choice of thermal comfort in a way that is explored in detail in the remainder of this section.

The dwelling looses thermal heat through its transmission across the building elements and the infiltration of (unheated) air. The size of the transmission and infiltration heat losses per degree temperature difference to the outside, H_T and H_V, are properties of the dwelling. To obtain the total amount of heat that is required over the year, they can be multiplied with the total number of degree-hours temperature increases over the outside temperature, that have to be produced during a year to keep the dwelling at the desired indoor temperature. In the given model these heating degree days (HDD) are obtained as the product of the difference between the mean indoor and outdoor temperature during the heating period, τ^{in} and τ^{out}, the length of the heating period in days, t_{HP}, and the number of kilo hours per day:

$$HDD = (\tau^{in} - \tau^{out}) \cdot t_{HP} \cdot 0.024. \tag{1.5}$$

Because households' may not heat the entire dwelling to the desired temperature level during the entire heating period, the model allows the temperature increase to be adjusted by a factor for the intensity of use, f_I. It is the product of the fraction of rooms and time of the day that the dwelling is heated, f_a and f_d:[4]

$$f_I = f_a \cdot f_d.$$

households' choices into predictions of their fuel consumption. See for instance Aydin et al. (2017) and Majcen et al. (2013) for formalisation of a fuel requirement calculation similar to the one sketched below.

[4]For instance, if the household decides to heat his entire dwelling of size m to the desired indoor temperature, τ^{in}, but only for 16 hours a day, this implies $f_a = \frac{m}{m} = 1$ and $f_d = \frac{16}{24} = \frac{2}{3}$. The original model by Loga et al. (2005) also includes a general intensity factor, f_n, that allows to easily calibrate the assumed household behavior such that the predicted fits well to the observed consumption.

Overall, the heat loss is thus obtained as a function of household decision parameters:

$$Q_L = (H_T + H_V) \cdot (\tau^{in} - \tau^{out}) \cdot t_{HP} \cdot f_a \cdot f_d \cdot 0.024. \tag{1.6}$$

In terms of the general characterisation of the input demand equation discussed before, the vector $\boldsymbol{\tau}$, representing households' decisions determining the final fuel demand, consists of the elements τ^{in}, t_{HP}, f_a and f_d. The indoor temperature, τ^{in}, is the most important variable governing households' level of thermal comfort. The difference between the indoor temperature chosen and the outdoor temperature determines by how many degrees the dwelling has to be heated up on average. Multiplying this with the adjustment factors for partial use, one obtains the mean number of degrees the entire dwelling effectively has to be heated up over the entire heating period. Following the model by Loga et al. (2005), I also assume that the decision to turn the heating system on (and off), determining the length of the heating period, and which indoor temperature to set once the system is in use are two separate and independent choices households' make.[5] The vector $\boldsymbol{\tau}$ is thus reduced to a scalar variable, τ, that represents the households' single choice variable in the input demand function and can be interpreted as the mean temperature increase over the entire dwelling if the heating system is turned on:

$$\tau = (\tau^{in} - \tau^{out}) \cdot f_a \cdot f_d. \tag{1.7}$$

This reduction of the dimensionality of households' choice set implies that their fuel demand can be predicted based on one simple measure of household behavior that is straight forward to interpret. This greatly simplifies the analysis of their heat demand in an empirical or theoretical economic framework. In particular, it might allow the identification of household behavior in empirical studies based on observed fuel consumption data.[6]

[5]The assumption greatly facilitates the analysis of the households' decision problem, since the effect of the temperature choice on the number of days the dwelling has to be heated (i.e. the extensive margin), does not have to be modelled. The plausibility of the assumption depends on the heating technology typically used in a country. In Germany, where a central system that produces heat and requires some time to start-up is typically located outside the living area, it seems a reasonable assumption. In contrast, it might be less plausible in regions that use for instance electric heating that can be turned on and off flexibly from within the living area. A model that explicitly accounts for the interdependency of temperature choice and the length of the heating period is provided by Hausman (1979).

The model by Loga et al. (2005) determines the length of the heating period as the number of days with a mean outdoor temperature below a dwelling specific heating limit temperature. Implicitly, households are assumed to turn their heating system on/off once a year, as soon as outdoor temperatures get sufficiently cold/warm. Since better insulated dwellings allow the household to enjoy comfortable indoor temperatures without heating at lower outdoor temperatures, they have lower heating limit temperatures, and thus fewer heating degree days, than dwellings with inferior thermal insulation.

[6]Note that it would generally be very difficult to identify several behavioral parameters from fuel consumption data alone as any observed fuel demand could be the outcome of various combinations of the different choices households make.

Combining equations (1.7), (1.6) and (1.4), the fuel demand from heating is obtained as a linear function of the households' temperature choice. Since the difference of additional heat losses and heat gains, Q_A, is relatively small in practice and it is independent of behavior, and thus irrelevant for the calculation of the marginal cost of heating, I simplify the engineering model slightly by dropping the term $Q_A \cdot \xi$.[7] I thus obtain

$$F^{\tilde{e}} = (H_T + H_V) \cdot t_{HP} \cdot 0.024 \cdot \xi \cdot \tau, \tag{1.8}$$

where the tilde in the superscript indicates the slight deviation from the engineering model.

I then collect all technical parameters into one variable and scale it by the size of the dwelling in square metres, m. This yields a one-dimensional state variable, indicating the amount of fuel a household has to use per square metre to increase the mean indoor temperature over the entire dwelling by one degree for the entire heating period. The variable $s \equiv \frac{1}{m}((H_T + H_V) \cdot t_{HP} \cdot 0.024 \cdot \xi)$ allows to compare the relative efficiency of the technology used to produce thermal warmth across dwellings.[8] The input demand as a function of the efficiency state and households' temperature choice is thus obtained as

$$I^{\tilde{e}}(\tau, s) = s \cdot m \cdot \tau. \tag{1.9}$$

Equation (1.9) shows that an input demand function that is linear in an univariate variable representing household behavior can be in line with a fuel requirement calculation procedure developed by engineers. Linearity of the input demand function greatly facilitates the economic analysis of households' heat demand. Theoretically, it implies that the marginal cost of temperature production are independent of household behavior and thus the temperature choice is a standard consumer problem and easy to solve explicitly.[9] Econometrically, it ensures that regression equations resulting from specifications of households' temperature choices stay fairly simple and thus comparably easy to estimate and interpret.

[7] Dropping the term $Q_A \cdot \xi$ from the fuel requirement calculation, the predicted fuel consumption is reduced to 91.41 % of the value that would be obtained based on the complete engineering model on average. In absolute terms, the average reduction in fuel consumption is $24.65\ ^{kWh}/_{m^2}$. Table A.4 in appendix A.1 provides results from an alternative regression equation that does not drop the constant term $Q_A \cdot \xi$. The results are qualitatively the identical to the main results discussed in section 1.6.

[8] Energy efficiency is often defined as the amount of energy services the technology allows to produce per unit of energy input. Note that, according to this definition, s is the reciprocal of efficiency, often referred to as energy intensity (see Galvin (2014) for a very detailed discussion of alternative definitions of energy efficiency). Throughout the paper, I refer to the variable s as the "efficiency state", while acknowledging that increases in the efficiency level are represented by lower levels of this state variable.

[9] See Pollak and Wachter (1975) for a detailed discussion of the conceptual problems that arise if the price of the desired good depends on households tastes as well as the technology they use in its production.

Previous research has often used linear input demand functions more or less implicitly. For instance, the "service factor" developed by Haas et al. (1998) relates household behavior to actual fuel demand in a linear way. More fundamentally, approaches estimating the elasticity, $\varepsilon_{\tau,s}$, of household behavior with respect to changes in dwellings' efficiency using observed fuel data are based on the assertion that it relates to the efficiency elasticity of fuel consumption, $\varepsilon_{F,s}$, as indicated by the formula

$$\varepsilon_{\tau,s} = 1 + \varepsilon_{F,s}. \tag{1.10}$$

Equivalent relationships have for instance been developed by Aydin et al. (2017) and Sorrell and Dimitropoulos (2008). Yet, equation (1.10) holds if and only if the input demand function is linear in household behavior.[10]

Not all fuel requirement calculations are linear in households' temperature choice. For instance, Dubin and McFadden (1983) explicitly derive a nonlinear form of the input demand equation, resulting from different assumptions on households' ventilation behavior compared to the model by Loga et al. (2005). Also the length of the heating period, t_{HP}, or the efficiency of the heating system, ξ, could depend on the temperature choice, resulting in equation (1.8) being nonlinear in indoor temperature. In this work, I assume that the engineering model by Loga et al. (2005) used to derive equation (1.9) correctly describes reality. Potential errors in the model's prediction of actual fuel consumption are thus assumed to be due to wrong assumptions on household behavior or the use of erroneous efficiency states, and not due to how these variables interact to determine the actual fuel demand.

1.3.3 Validity of fuel requirement calculations

While engineering predictions of fuel consumption are generally found to perform well in providing comparable measures of dwellings' thermal performance under standardised conditions (see, e.g., Aydin et al. (2017) for a recent discussion), is a pervasive finding of the literature that predicted fuel requirements systematically exceed actual consumption, and often substantially (Sunikka-Blank and Galvin, 2012; Laurent et al., 2013). I assume that the fuel consumption observable in empirical data, F^d, precisely measures actual consumption, F°, and is thus not responsible for the gap.[11] A

[10] More precisely, I show in appendix A.2 that for input demand functions of type $I(s, h(\tau)) = smh(\tau)$ equation (1.10) only holds if and only if $h(\tau) = \tau$. For more general forms of the input demand function, equation (1.10) is in general unlikely to hold.

[11] Even if measurement errors exist, they are unlikely to be systematic in a way, that the observed underconsumption relative to the engineering prediction could be explained (see Laurent et al., 2013).

formal representation of the gap is then

$$F^d = F^\circ = I(\boldsymbol{\tau}^\circ, \boldsymbol{s}^\circ) < I(\boldsymbol{\tau}^e, \boldsymbol{s}^e), \tag{1.11}$$

where nodes indicate unobservable true values of the variables and e superscripts engineering assumptions, respectively. Equation (1.11) emphasizes, that the systematic overprediction of actual consumption either results from engineering models assuming levels of thermal comfort above actual levels, $\tau^e > \tau^\circ$, or because they operate with efficiency states that are too large, $s^e > s^\circ$.[12]

In line with most discussions of the problem, Laurent et al. (2013) identify wrong behavioral assumptions as the primary source of the gap. A reason is, that engineering models typically calibrate $\boldsymbol{\tau}^e$ to normed levels of thermal comfort, which reflect desirable consumption levels rather than actual choices. As a consequence households making decisions under (monetary) constraints are likely to underconsume relative to the engineering specification. The strong correlations between the service factor and households' income and dwellings' efficiency, commonly found in the literature (as discussed in section 1.2), is consistent with this view on the gap.

Yet, overpredictions of the efficiency states used in engineering models can equivalently explain the gap. Equation (1.2) indicates two potential sources for erroneous efficiency states. They may be due to limited data availability (i.e. $\boldsymbol{D}^e \neq \boldsymbol{D}^\circ$, because the relevant data is not available or imperfectly measured) or difficulties in aggregating the available data, because the true aggregation function, $\psi^\circ(\cdot)$, is imperfectly known or not applicable to the observable dwelling characteristics, $\boldsymbol{D}^e$. In fact, evaluations of potential sources of the gap acknowledge that errors in the efficiency states are also likely to contribute to the gap (Sunikka-Blank and Galvin, 2012; Laurent et al., 2013; Majcen et al., 2013). In particular, the difficulties in raising complete and correct data on the fundamental physical properties of the dwelling are often emphasized. Majcen et al. (2013) argue that the inaccuracies in the predicted consumption levels might be larger for older dwellings, because "inspecting older dwellings is often difficult and instead of measuring U-values, a guess is made as to whether the cavity walls were insulated at the time of construction and what the quality of that insulation may be after many years" (Majcen et al., 2013, p. 461). Knissel and Loga (2006) find in their study of German households that the magnitude of the overprediction varies in dwelling characteristics such as the number of apartments per dwelling, its age and the type of heating system used. The variation can be substantial. While the overprediction in a dwelling with $1-2$ apartments is 43.8 %, predicted consumption is only 26.5 % above actual consumption

[12] Systematic errors in the functional relationship that describes how household behavior and dwelling efficiency are mapped into a fuel demand are not considered a relevant factor behind the gap in this paper. This is consistent with Laurent et al. (2013) finding that the "core modelling" behind the engineering models represents rather good approximations of reality.

in dwellings with 8 or more apartments. Even though these observations do not reveal the precise source of the gap, they indicate that its size is correlated to technical covariates in a systematic way that makes household behavior unlikely to be its only cause.

For the use of the engineering model in an economic study, errors in the prediction of the efficiency state are of primary concern. In fact, the goal and outcome of the economic analysis based on engineering knowledge should be to improve the understanding – and eventually also the modelling – of households' temperature choices. In contrast, economists have much less expertise in assessing and possibly correcting flaws in the fuel requirement calculation unrelated to households' choices. Errors in the efficiency state provided by an engineering model, enter as a measurement error into the econometric analysis. In particular, the apparent correlation of the gap between predicted and actual consumption to dwelling characteristics, discussed above, strongly indicates that such measurement errors are likely correlated to household characteristics and other independent variables of interest in an empirical specification.

The following section provides a detailed discussion how an empirical regression equation based on engineering knowledge can be obtained, that treats both potential sources of the error in engineering predictions in a systematic way.

1.4 Empirical specification

I derive the empirical specification in two steps, based on the input demand function in equation (1.9). First, I show that a regression approach, closely related to the use of the service factor in earlier research, can be obtained by assuming the true efficiency state $s^{\circ}_{i,t}$ to be observed. I then show how the regression framework can be extended to relax this assumption.

1.4.1 Obtaining an empirical specification if the true efficiency state is observed

The functional form of the input demand equation derived in section 1.3.2 and stated in equation (1.9), depicts a simple and systematic way to relate households' temperature choice to observable fuel consumption. Assuming the true efficiency state, $s^{\circ}_{i,t}$, to be observed for every household i and time period, t, and that individual temperature choices are the result of some general optimal behavior that can be described by a function $\tau^*(\boldsymbol{x^\tau_{i,t}}, \boldsymbol{\beta^\tau}, \varepsilon^\tau_{i,t})$, a regression equation can be defined that allows to estimate the effects of household characteristics and economic variables, stored in a vector $\boldsymbol{x^\tau_{i,t}}$, on the chosen level of thermal comfort:

$$F^d_{i,t} = s^{\circ}_{i,t} \cdot m_{i,t} \cdot \tau^*(\boldsymbol{x^\tau_{i,t}}, \boldsymbol{\beta^\tau}, \varepsilon^\tau_{i,t}). \tag{1.12}$$

The error term $\varepsilon^{\tau}_{i,t}$ associates variation in fuel demand that cannot be explained by the set of covariates included in $\boldsymbol{x^{\tau}_{i,t}}$ to idiosyncracies in households' behavior. Since the functional form of the input demand function, as well as the efficiency state, are assumed to be known, no parameter has to be estimated on the term $s^{\circ}_{i,t} \cdot m_{i,t}$ and variation in the dependent variable can directly be attributed to household behavior. An algebraically equivalent regression equation can be obtained by dividing both sides of the equation by $s^{\circ}_{i,t} m_{i,t}$:

$$\frac{F^d_{i,t}}{s^{\circ}_{i,t} m_{i,t}} \equiv \tau^d_{i,t} = \tau^*(\boldsymbol{x^{\tau}_{i,t}}, \boldsymbol{\beta^{\tau}}, \varepsilon^{\tau}_{i,t}). \tag{1.13}$$

The dependent variable is thus transformed into the implied temperature choice, $\tau^d_{i,t}$, that would have lead to the fuel demand $F^d_{i,t}$ according to the input demand function (1.9). Equation (1.13) emphasizes that the parameters, $\boldsymbol{\beta^{\tau}}$, indeed identify the effect of the variables contained in $\boldsymbol{x^{\tau}_{i,t}}$ on the economic decision variable of interest.

Equation (1.13) is directly related to previous approaches relying on the service factor, i.e. the ratio of actual and predicted consumption, $F^d_{i,t}$ and $F^e_{i,t}$, to study households' consumption of thermal warmth. Dividing both sides of the equation by the mean indoor temperature assumed in the engineering model, $\tau^e_{i,t}$, one obtains

$$\frac{F^d_{i,t}}{s^{\circ}_{i,t} m_{i,t} \tau^e_{i,t}} \equiv \frac{F^d_{i,t}}{F^e_{i,t}} = \frac{\tau^*(\boldsymbol{x^{\tau}_{i,t}}, \boldsymbol{\beta^{\tau}}, \varepsilon^{\tau}_{i,t})}{\tau^e_{i,t}}. \tag{1.14}$$

These approaches thus identify the size of actual temperature choices relative to those assumed in the engineering model. The advantage of the framework developed in this paper is that the regression equation is derived from an explicit functional form of the input demand equation, emphasizing the role of the linearity assumption. Moreover, it also allows interpretations with respect to the actual temperature choice directly. Finally, while the existing literature using the service factor has assumed that the robust gap between actual and predicted consumption is entirely driven by erroneous assumptions on household behavior in the engineering model, the developed framework allows a straight forward extension that is sketched in the next section.

1.4.2 Allowing for measurement errors in the generated efficiency states

The discussion in section 1.3.3 has pointed out that calculations of the efficiency state, $s_{i,t}$, that are based on engineering models are likely to suffer from substantial upward biases that should not be ignored in an empirical analysis. I consider the case that the size of the overprediction by the engineering model is proportional to the size of the efficiency state and can be captured by a scalar variable, $\lambda_{i,t} \in \mathbb{R}^+$, such that

$$s^{\circ}_{i,t} = \lambda_{i,t} \cdot s^e_{i,t}. \tag{1.15}$$

Combining equations (1.15) and (1.12) and solving again for $\tau^d_{i,t}$ as the dependent variable, the empirical model becomes

$$\frac{F_{i,t}}{s^e_{i,t} m_{i,t}} \equiv \tau^d_{i,t} = \tau^*(\boldsymbol{x^{\tau}_{i,t}}, \boldsymbol{\beta^{\tau}}, \varepsilon^{\tau}_{i,t}) \cdot \lambda_{i,t}. \tag{1.16}$$

The bias in the predicted efficiency state, $s^e_{i,t}$, results in a systematic measurement error in the dependent variable, $\tau^d_{i,t}$, that is represented by an additional error term, $\lambda_{i,t}$, on the right hand side of equation (1.16). Two problems might result from the error. First, the engineering models' tendency to systematically overpredict actual consumption implies that $\lambda_{i,t}$ is expected to have mean smaller than one, which biases the intercept or the estimated coefficients in case the intercept is omitted. Second and more problematic in most applications, does the literature find $\lambda_{i,t}$ to be correlated to dwelling characteristics (see the discussion in section 1.3.3 for details). This is likely to cause correlation between the error term and variables in the vector $\boldsymbol{x^{\tau}_{i,t}}$, making them endogenous.

To avoid endogenity of my explanatory variables of interest in the empirical analysis, I assume that $\lambda_{i,t}$ can be represented by a function, $\lambda(\boldsymbol{x^{\lambda}_{i,t}}, \boldsymbol{\beta^{\lambda}}, \varepsilon^{\lambda}_{i,t})$, of observed covariates stored in a vector $\boldsymbol{x^{\lambda}_{i,t}}$ and an additional error $\varepsilon^{\lambda}_{i,t}$, that is orthogonal to variables in the vector $\boldsymbol{x^{\tau}_{i,t}}$ affecting household behavior. Estimating $\lambda(\boldsymbol{x^{\lambda}_{i,t}}, \boldsymbol{\beta^{\lambda}}, \varepsilon^{\lambda}_{i,t})$ as part of equation (1.16), then allows to obtain unbiased estimates of the parameters of interest.[13]

Equation (1.16) indicates a multiplicative connection of the quantities of interest, that results from a fundamental relationship between households' temperature choice and the physical characteristics of the dwelling. This imposes restrictions on the functional forms that can be used to model $\tau^*(\boldsymbol{x^{\tau}_{i,t}}, \boldsymbol{\beta^{\tau}}, \varepsilon^{\tau}_{i,t})$ and $\lambda(\boldsymbol{x^{\lambda}_{i,t}}, \boldsymbol{\beta^{\lambda}}, \varepsilon^{\lambda}_{i,t})$. In particular, a linear specification of both functions would lead to a large number of interactions that do not identify the parameters. To obtain an estimable regression equation, I therefore make the technical assumption that $\tau^*(\boldsymbol{x^{\tau}_{i,t}}, \boldsymbol{\beta^{\tau}}, \varepsilon^{\tau}_{i,t})$ and $\lambda(\boldsymbol{x^{\lambda}_{i,t}}, \boldsymbol{\beta^{\lambda}}, \varepsilon^{\lambda}_{i,t})$ can be represented by exponential functions. That is,

$$\tau^*(\boldsymbol{x^{\tau}_{i,t}}, \boldsymbol{\beta^{\tau}}, \varepsilon^{\tau}_{i,t}) = e^{\boldsymbol{x^{\tau}_{i,t}} \boldsymbol{\beta^{\tau}} + \varepsilon^{\tau}_{i,t}} \tag{1.17}$$

[13] In the empirical analysis I sometimes refer to the term $\lambda(\boldsymbol{x^{\lambda}_{i,t}}, \boldsymbol{\beta^{\lambda}}, \varepsilon^{\lambda}_{i,t})$ as the "adjustment factor" since, according to equation (1.15), it can be interpreted to indicate how much the state variables, $s_{i,t}$, generated in the engineering model have to be scaled downward to obtain reliable values of the dwellings' efficiency level.

and

$$\lambda(\boldsymbol{x^\lambda_{i,t}}, \boldsymbol{\beta^\lambda}, \varepsilon^\lambda_{i,t}) = e^{\boldsymbol{x^\lambda_{i,t}}\boldsymbol{\beta^\lambda} + \varepsilon^\lambda_{i,t}}, \tag{1.18}$$

where the error terms $\varepsilon^\tau_{i,t}$ and $\varepsilon^\lambda_{i,t}$ are assumed to be normally distributed. Making the intercepts, β^τ_0 and β^λ_0 explicit,[14] regression equation (1.16) then becomes

$$\tau^d_{i,t} = e^{\beta^\tau_0 + \beta^\lambda_0 + \boldsymbol{x^\tau_{i,t}}\boldsymbol{\beta^\tau} + \boldsymbol{x^\lambda_{i,t}}\boldsymbol{\beta^\lambda} + \varepsilon^\tau_{i,t} + \varepsilon^\lambda_{i,t}}, \tag{1.19}$$

which can be estimated using a Poisson regression model with robust errors. Alternatively, logarithmising equation (1.19) yields a linear specification with a log dependent variable:

$$\ln \tau^d_{i,t} = \beta^\tau_0 + \beta^\lambda_0 + \boldsymbol{x^\tau_{i,t}}\boldsymbol{\beta^\tau} + \boldsymbol{x^\lambda_{i,t}}\boldsymbol{\beta^\lambda} + \varepsilon^\tau_{i,t} + \varepsilon^\lambda_{i,t}. \tag{1.20}$$

Consistency of the estimates of $\boldsymbol{\beta^\tau}$ requires the remaining measurement error, ε^λ, as well as ε^τ to be orthogonal to the regressors. Equation (1.20) also illustrates that if $\mathbf{E}[\lambda] < 1$, this is captured by the intercept $\beta_0 \equiv \beta^\tau_0 + \beta^\lambda_0$ and thus does not bias the coefficients of primary interest. Since it implies $\beta^\lambda_0 < 0$ the estimated constant underestimates the true intercept in households' decision process, β^τ_0. The lack of identification of the individual parameters β^τ_0 and β^λ_0 implies that level predictions of households' temperature choice, $\tau^*_{i,t}$, and the error term, $\lambda_{i,t}$, are not possible. Also, the partial effects of these quantities with respect to the observed covariates depend on all parameters in the respective exponents, including the intercepts, and are thus not identified. The interpretation of the regression results is therefore limited to the relative changes the regressors induce on $\tau^*_{i,t}$ and $\lambda_{i,t}$, respectively.[15]

Moreover, the relative effect of an explanatory variable $x_{i,t,k}$ on the temperature choice $\tau^*_{i,t}$ is only identified if $x_{i,t,k}$ is a component of the vector $\boldsymbol{x^\tau_{i,t}}$ but not of $\boldsymbol{x^\lambda_{i,t}}$. While household characteristics can often savely be assumed not to affect the size of the measurement error, $\lambda_{i,t}$, only the joint effect $\beta_k = \beta^\tau_k + \beta^\lambda_k$ can be identified in cases that this assumption cannot be maintained.[16]

Unfortunately, households' temperature choice depends on the marginal cost of temperature consumption – and thus the dwellings' efficiency level – which further complicates the consistent estimation of the parameters of interest. Knowing the functional form of the input demand function,

[14]With small abuse of notation the vectors $\boldsymbol{x^\tau_{i,t}}$ and $\boldsymbol{x^\lambda_{i,t}}$ as well as $\boldsymbol{\beta^\tau}$ and $\boldsymbol{\beta^\lambda}$ keep the same names even though their dimensionality is reduced by a constant equal to one and a parameter, respectively.

[15]Of course, it is possible to predict the level of the implied temperature choice, $\tau^d_{i,t}$, as well as the respective partial effects based on the regression coefficients. Since $\tau^d_{i,t}$ suffers from measurement error, this is not very useful.

[16]The intercept is a trivial example for which only the joint effect can be identified leading to the problems in the identification of level effects discussed above.

stated in equation (1.9), the marginal cost can be calculated. Making use of equation (1.15) and denoting the marginal cost as $c_{i,t}$, their dependence on the measurement error, $\lambda_{i,t}$, becomes explicit:

$$c_{i,t}^{\circ} \equiv p_{i,t}^{F} \cdot \frac{\partial I(\tau_{i,t}, s_{i,t}^{\circ})}{\partial \tau_{i,t}} = p_{i,t}^{F} \cdot s_{i,t}^{\circ} \cdot m_{i,t} = p_{i,t}^{F} \cdot \lambda_{i,t} \cdot s_{i,t}^{e} \cdot m_{i,t} = \lambda_{i,t} \cdot c_{i,t}^{e}. \tag{1.21}$$

The overpredictions in the engineering models thus also introduce a measurement error in one of the explanatory variables. In addition to the endogenity concerns addressed before, this always results in an attenuation bias in all estimated coefficients, unless $\lambda_{i,t}$ is completely independent of $s_{i,t}^{e}$ (and thus by definition entirely driven by the true efficiency state $s_{i,t}^{\circ}$).[17] In the empirical analysis, I include many dwelling characteristics in the empirical specification of equation (1.18), such that the remaining error, $\varepsilon_{i,t}^{\lambda}$, and associated biases are arguably neglectable.

Finally, the fact that the efficiency state, $s_{i,t}$, enters the input demand function twice, implies that so does $\lambda(\boldsymbol{x_{i,t}^{\lambda}}, \boldsymbol{\beta^{\lambda}}, \varepsilon_{i,t}^{\lambda})$ used to model the measurement errors associated to it. Including $\ln(\lambda(\boldsymbol{x_{i,t}^{\lambda}}, \boldsymbol{\beta^{\lambda}}, \varepsilon_{i,t}^{\lambda}) \cdot c_{i,t}^{e})$ in the vector of explanatory variables affecting the temperature choice, $\boldsymbol{x_{i,t}^{\tau}}$, and simplifying equation (1.20) becomes:[18]

$$\ln \tau_{i,t}^{d} = \beta_0 + \boldsymbol{x_{i,t}^{\tau}} \boldsymbol{\beta^{\tau}} + \beta^{c} \ln c_{i,t} + \boldsymbol{x_{i,t}^{\lambda}} \boldsymbol{\beta^{\lambda}} \cdot (1 + \beta^{c}) + \varepsilon_{i,t}^{\tau} + (1 + \beta^{c}) \varepsilon_{i,t}^{\lambda}, \tag{1.22}$$

where β^{c} denotes the elasticity of temperature consumption with respect to the marginal cost of heating. The regression thus identifies a function of two parameters, $\boldsymbol{\gamma} \equiv \boldsymbol{\beta^{\lambda}} \cdot (1 + \beta^{c})$, which underestimates the true adjustment of the efficiency state unless the level of indoor temperature is a Giffen good. Intuitively, if the true measurement error, $\lambda_{i,t}$, scales the efficiency state down, this has two opposing effects on the dependent variable of which only the net effect can be estimated by the inclusion of $\lambda(\boldsymbol{x_{i,t}^{\lambda}}, \boldsymbol{\beta^{\lambda}}, \varepsilon_{i,t}^{\lambda})$ in the regression model. First, it directly reduces the implied indoor temperature, $\tau_{i,t}^{d}$, since a higher efficiency of the dwelling implies a lower fuel consumption for a given temperature choice, $\tau_{i,t}^{*}$. Second, the decrease in the marginal heating cost associated to a downscaled efficiency level triggers the consumption of a higher level of indoor temperature, $\tau_{i,t}^{*}$. Consistent estimates of β^{c} allow to calculate the true impact of technical covariates on the adjustment factor, by multiplying the respective parameters with $1/1+\beta^{c}$.

[17] See for instance Wooldridge (2009, p. 320 ff) for a discussion of biases resulting from measurement error in an explanatory variable.

[18] In equation (1.22) the marginal cost of a temperature increase, $c_{i,t}$, and the associated parameter, β^{c}, are made explicit. With small abuse of notation the vectors $\boldsymbol{x_{i,t}^{\tau}}$ and $\boldsymbol{\beta^{\tau}}$ keep the same names even though this reduces their dimensionality by one variable and one parameter, respectively.

1.5 Data

The estimation of the model developed in the previous sections requires data on households' fuel consumption, household and dwelling characteristics as well as the marginal cost of heating. The cost of increasing the level of indoor temperature by one degree are determined by the fuel price, $p^F_{i,t}$ and the thermal efficiency level of the dwelling, $s_{i,t}$. While fuel type specific price data is easily obtained as yearly mean values from the German ministry of economics (Bundesministerium für Wirtschaft und Energie, 2018), I have to use a simplified fuel requirement calculation by Loga et al. (2005) to generate efficiency states. In this section, I first depict the primary data set provided by RWI and forsa (2016), before I give a detailed description of how the efficiency states are generated.

1.5.1 The German Residential Energy Consumption Survey

The primary data source is "The German Residential Energy Consumption Survey" (RWI and forsa, 2016). The survey is based on a random sample of $6,715$ German households, that have been interviewed in 2010. The dataset includes information about household and dwelling characteristics as well as the energy consumption between 2006 and 2008.

The energy consumption is directly obtained from households' billing data and transferred to calendar years based on household specific heating degree days observed in the data. The data cleaning mostly follows the descriptions of RWI and forsa (2011) and includes an outlier correction procedure to eliminate implausible reports.

The final analysis is restricted to households using natural gas, long distance heating or oil in their primary heating system. Since the consumption of natural gas and long distance heating is metered, the observed fuel consumption is very likely to reflect actual consumption within a year quite precisely for these fuel types. For the storable fuel input oil, the precise measurement is much more complex. Typically, only the amount of oil purchased at a specific point in time rather than the actual use is observed, such that the households' fuel consumption can only be inferred indirectly. I obtain the fuel consumption for the year 2007 by first taking the sum over the amount of oil purchased in the three consecutive years from 2006 to 2008. I then assign a fraction of the total energy purchased to the year of interest, based on its proportion of heating degree days over the same time period. The fuel consumption for the remaining years, 2006 and 2008, is calculated analogously.[19] Households that have lost their oil bills are allowed to estimate their consumption in the questionnaire. To ensure a high precision of the consumption data, I drop estimated consumption values and restrict the analysis to fuel consumption reported based on actual energy bills. All consumption values are converted into kilowatt hours based on a

[19]See RWI and forsa (2011) for more details on the problem of dealing with storable fuel types.

conversion table provided by RWI and forsa (2011).

There are two main selection concerns with respect to the observed consumption data. First, might households that use natural gas, long-distance heat or oil for heating be different from those using other fuel types. Second, might households that possess an energy bill differ from those that are not able to respond on their energy consumption in the previous years. Table A.5 in appendix A.1 reports mean actual and predicted fuel consumption values together with household and dwelling characteristics for various restrictions of the data set for the year 2008. The descriptive data indicates that the major selection concern results from the fact that households that own the dwelling they live in are more likely to report consumption data than tenants. Consistent with this, the households analysed in this paper are older, more educated, have higher income and live in larger dwellings than the average population. While, overall the differences to the representative population are not extreme, the researcher should be aware of the existing selection when interpreting the results.[20]

The dataset contains substantial information on household and dwelling characteristics that are useful for the empirical analysis. Unfortunately, most variables of interest are only observed in 2010, at the time the interview was conducted. Besides the fuel consumption, only the total number of persons living in the household as well as the heating degree days are available for the years 2006 to 2008 as well. For some variables, I impute the yearly data based on the information available if this is possible and reasonable. Another limitation of the survey is that individual specific information is only available for the survey respondent.[21] I use this information to proxy the characteristics of the entire household in the empirical analysis.

A number of household characteristics have to be modified or generated before their inclusion in the regression equation. I calculate the age of the household based on the year of birth of the survey respondent. Similarly, I track the number of children living in the household based on the reported dates of their births. The number of adult persons in the household is then obtained as the difference between the total number of household members reported for a given year and the calculated number of children. I top-code the number of adults and children at four and two people respectively. Based on categorical income information in the data, I obtain indicators for households being in a low, middle or high income group. Households are assigned to a low income

[20]Table A.3 in appendix A.1 replicates the preferred specification discussed in section 1.6 focussing on tenants and owners respectively. The effects are broadly similar for both types of households. The effect of income and the size of the household seem to be more pronounced for tenants than for owners. A rigorous way to deal with the selection concern would be to apply a selection model (Heckman, 1979), which is beyond the scope of this study.

[21]The exception is date of birth, which is also available for all children living in the households.

group if they have less than $1,500$ euros, to a middle income group if they earn between $1,500$ and $3,500$ euros and a high income group if they earn more than $3,500$ euros per month.[22] I also generate a variable whether a household has a job based on occupation information available in the data. Binary indicator variables whether the household has a high-school degree (the German "Abitur") and possesses the dwelling he lives in, in the year the survey is conducted, can be obtained from the survey directly. I assume that these variables have stayed the same between 2006 and 2010.

Only minor changes have to be made to prepare dwelling characteristics for the final analysis. To obtain an indicator of the relative size of the dwelling, I first calculate for each dwelling the relative deviation from the mean dwelling size (in square metres) over the entire sample. I then generate an indicator whether the dwelling belongs to the group of small, medium or large dwellings based on the terciles of the distribution.

1.5.2 Generation of efficiency states

To generate the efficiency states, $s_{i,t}$, I make use of the simplified fuel requirement calculation developed by Loga et al. (2005). Their calculation procedure has been designed to provide good approximations of the amount of fuel that will be consumed in a given dwelling based on limited input data. Instead of gathering as much and detailed information on the dwellings' characteristics – such as the thermal properties and thickness of all components of the thermal shell – as possible, their approach is based on data that is $a)$ easy to observe in practice and $b)$ informative on probable essential physical properties of the dwelling.

They develop three procedures that are capable of mapping this limited input data into the essential dwelling properties needed for the calculation of fuel requirements. These include the approximation of the areas of the components of the thermal shell, the definition of generalised U-values, indicating the quality of the thermal shell, and the definition of efficiency indicators of the installation engineering, that determine the efficiency of the production and distribution of heat respectively. While the approximation of areas of the components of the thermal shell is based on an empirical analysis of more than $4,000$ existing dwellings, do the authors refer to existing literature and standards for the definition of the remaining two procedures. The generalised values are defined for different construction years and types. In comparisons of fuel requirements calculated using the simplified procedures with those resulting from a full engineering model, the authors find their approximations to perform well (Loga et al., 2005). The median percentage

[22]The reduction of the number of income groups is convenient to a) ensure a sufficiently large number of observations in each income bin and b) reduce the number of variables included in the final regression. Unreported results confirm the robustness of the main empirical findings towards alternative definitions of income bins.

deviation between the simplified and the standard model is 1 % and the standard deviation at most 15 %.

The simplified calculation procedure developed by Loga et al. (2005), provides an opportunity to generate fuel requirements based on the dwelling characteristics observed in the primary dataset. The generation process implies that substantial information is added to the crude data. This information is drawn from standard engineering knowledge on the calculation of fuel requirements as well as the specific approximations provided by the authors. Ultimately, the efficiency state, $s_{i,t}$, a quantity predicting the amount of kilowatt hours the dwelling consumes in a year, per square metre and degree temperature increase, can be obtained. Different to the original input data, this can be used to relate household behavior and observed fuel demand in a theoretically consistent way.

To generate the fuel requirements, I modify the original program only at one point. I disable a technical adjustment of the fraction of time and area that are actually being heated, which is based on the quality of the dwellings' thermal insulation. This ensures that the generated efficiency states are mere indicators of the thermal performance of the dwelling. All behavioral reactions to variations in the efficiency level of the dwellings, are exclusively captured by the utility parameters to be estimated in the empirical analyses.[23]

Unfortunately, the dwelling characteristics, observed in the primary dataset, miss multiple inputs required in the simplified fuel requirement procedure. To run the program, I impute the missing data, referring to plausible values obtained from the authors' analyses and other external sources where possible. Importantly, all imputations are based on identical procedures to preserve the property of the resulting fuel requirements to indicate the relative level of efficiency of the dwelling.

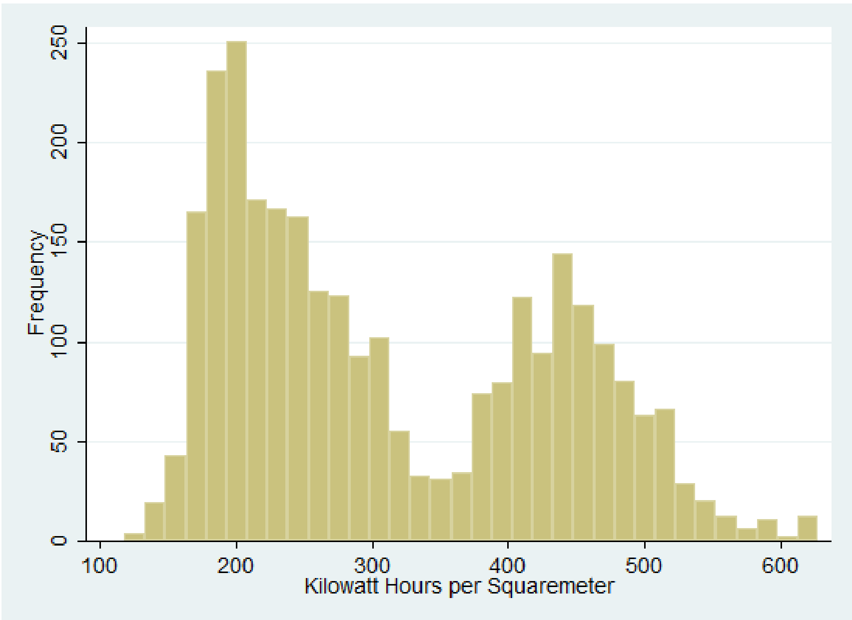
Clearly, the lack of required input data introduces noise into the generated variables. It will result in differences between the predicted fuel requirements and the actual consumption, beyond those intrinsic to the engineering model. If this error is correlated to explanatory variables in a systematic way, it will bias the resulting parameter estimates. This concern introduces an additional reason for the inclusion of a technical adjustment factor that can correct the efficiency variables for some of these effects where they exist. In the empirical analysis below I assume the remaining error after inclusion of the technical adjustment factor to be uncorrelated to explanatory variables.

[23] All files that are used for the calculation of the fuel requirements are provided in the supplementary material to this article. It also contains a detailed description how the files used for the generation are obtained based on the original program provided by Loga et al. (2005) as well as how it can be applied to the data.

I generate household specific fuel requirements for every household observed between 2002 and 2008. I apply two types of outlier corrections on the generated data, to avoid that empirical results are driven by extremely large or small values. A detailed discussion of the procedures used is provided in appendix A.3.1. Columns (7) to (10) of table A.5 in appendix A.1 verify that the outlier correction does not imply a selection on household or dwelling characteristics. Robustness checks reported in columns (6) and (7) of table A.3 in appendix A.1 confirm that the outlier corrections have no qualitative effect on the results of the empirical analysis conducted in section 1.6.

Figure 1.1 plots the distribution of the generated fuel requirements in the final sample between 2006 and 2008. The distribution of the generated fuel requirements is bimodal with a large density

Figure 1.1: Distribution of generated fuel requirements



of fuel requirements around 250 and 500 kWh/m^2 and very few predictions of fuel consumptions around 400 kWh/m^2. The bimodality results from a strong efficiency increase for dwellings constructed after 1969. More generally, the construction year is the most important predictor of the dwellings' efficiency. Appendix A.3.2 provides a detailed discussion of the generated fuel requirements.

1.6 Empirical analysis

Equations (1.20) and (1.19) are estimated using the OLS and Poisson estimator, respectively. Year fixed effects are included in all regressions and the standard errors are clustered on the household level. The empirical analysis considers households' demand conditional on the technology available to the household. That is, it takes a short-run perspective, in which the technology cannot be changed by the household.

Table 1.1: Coefficients of naive regressions [a]

Dep. Var: $\tau_{i,t}^{d}$	(1)	(2)	(3)
Income:			
$< 1,500$ €	-0.259***		-0.324***
	(0.052)		(0.053)
$\geq 3,500$ €	0.144***		0.171***
	(0.034)		(0.031)
log($c_{i,t}$)		-0.335***	-0.370***
		(0.025)	(0.026)
Constant	1.823***	3.556***	3.723***
	(0.024)	(0.131)	(0.138)
Observations	2,476	2,838	2,476
R-squared	0.054	0.094	0.159
Estimator	OLS	OLS	OLS

[a] Columns (1) to (3) of the table report estimation results from ordinary least squares (OLS) regressions on different specifications of equation (1.20). Year fixed effects are included in all regressions and the standard errors are clustered on the household level and reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

Table 1.1 reports results from OLS regressions of $\tau_{i,t}^{d}$ on households' income, marginal cost of producing thermal comfort, $c_{i,t}$, a constant and year fixed effects, respectively. Similar to the previous literature that has graphically analysed the ratio of actual and predicted consumption with respect to both economic variables (see discussion in section 1.2), the regressions do not control for socioeconomic characteristics of the household or potential errors in the engineering prediction. Not surprisingly, the results mirror those of the previous studies. They show a substantial impact of income and marginal heating cost on households' behavior. Column (1) indicates that households in the low income group consume a mean temperature level 25.9 % below those in the middle income group. Richer households in the high income group are found to demand mean indoor temperatures 14.4 % above the middle income group. Column (2) confirms that households reduce their consumption of indoor temperature when the marginal cost of production increase. The elasticity is estimated to be -0.34. The results do not change substantially if income and marginal heating cost are analysed jointly, as indicated by column (3).

While these results are intuitive, the regressions might be overly simplistic. For instance, preferences for thermal comfort might vary across different types of households. If they vary in a systematic way with household characteristics that are correlated with household income or the efficiency of the dwelling, the estimates reported in table 1.1 will be biased and inconsistent. Additionally, the discussion in section 1.4 has shown, that measurement error in the variable, $s_{i,t}$, is a potential problem that has to be investigated in the empirical analysis. A virtue of the regression framework developed in section 1.4 is, that it makes it straight forward to extend the analysis to deal with such problems.

In column (1) of table 1.2, I allow households' temperature choices to depend on the age, employment status, education level and income group of the survey respondent. I also include indicators for the number of adult persons and children living in the household and whether he owns the dwelling he lives in. Because households' preferences for high mean indoor temperatures might also vary with the size of the dwelling,[24] a categorical variable indicating whether the household lives in a dwelling that is smaller or larger than average is also added to the regression equation. The inclusion of these control variables has a substantial impact on the estimated coefficients on income and the marginal cost of production. The effect of income becomes substantially smaller. The low income group is now estimated to consume mean indoor temperatures 9.3 % below the average income group. The effect is significant at a 5 % level. The point estimate on the high income group now indicates only a 2.0 % increase in consumption for rich households, which is statistically insignificant. In contrast, the estimated price elasticity stays highly significant and almost doubles to -0.63. A household with marginal cost 1 % above an otherwise identical household is expected to consume a 0.63 % lower mean indoor temperature on average.

The inclusion of household characteristics substantially increases the variance explained by the regression model, which almost triples from 16 % in column (3) of table 1.1 to 45 % in the first column of table 1.2, as indicated by the R^2. The coefficients on household characteristics are intuitive and consistent with earlier results reported, e.g., by Meier and Rehdanz (2010) and Risch and Salmon (2017). For instance, older households are found to choose higher indoor temperatures

[24]While a low mean indoor temperature in a small dwelling implies, that households are likely to experience this lack of thermal comfort for a significant amount of time during the day, the household might have the opportunity to avoid some rooms that are not heated if he lives in a rather large dwelling. This would imply, that a high mean indoor temperature is valued less by a household living in a large dwelling.

Table 1.2: Estimated coefficients of the full regression model [a]

Dep. Var: $\tau^d_{i,t}$	(1)	(2)	(3)	(4)
Income:				
$< 1,500$ €	-0.093**	-0.076*	-0.099**	
	(0.044)	(0.041)	(0.042)	
$\geq 3,500$ €	0.020	0.011	0.024	
	(0.029)	(0.026)	(0.025)	
$\log(c_{i,t})$	-0.628***	-0.663***	-0.692***	
	(0.027)	(0.043)	(0.039)	
age:				
$30 - 39$	0.096	0.060	0.063	
	(0.065)	(0.062)	(0.061)	
$40 - 49$	0.116*	0.083	0.062	
	(0.064)	(0.059)	(0.059)	
$50 - 59$	0.222***	0.135**	0.118**	
	(0.063)	(0.059)	(0.058)	
≥ 60	0.269***	0.142**	0.123**	
	(0.067)	(0.062)	(0.061)	
# adults:				
2	0.181***	0.175***	0.138***	
	(0.039)	(0.033)	(0.033)	
3	0.227***	0.225***	0.181***	
	(0.048)	(0.041)	(0.041)	
≥ 4	0.269***	0.216***	0.154***	
	(0.057)	(0.053)	(0.053)	
# children:				
1	0.076*	0.048	0.031	
	(0.043)	(0.039)	(0.040)	
≥ 2	0.116***	0.073*	0.037	
	(0.042)	(0.038)	(0.039)	
Is employed	0.011	-0.036	-0.029	
	(0.034)	(0.030)	(0.029)	
Has Abitur	0.004	0.029	0.027	
	(0.026)	(0.024)	(0.023)	
Is owner	0.248***	0.035	0.047	
	(0.037)	(0.035)	(0.033)	
Size of the dwelling:				
Small: $<$ 1st tercile	-0.257***	-0.107***	-0.105***	-0.020
	(0.037)	(0.035)	(0.034)	(0.033)
Large: $>$ 2nd tercile	0.188***	0.166***	0.170***	-0.017
	(0.031)	(0.030)	(0.028)	(0.027)

Table 1.2 continued from previous page

Type of the dwelling:			
Row house	-0.181***	-0.185***	-0.066***
	(0.027)	(0.025)	(0.025)
Multi-family dwelling	-0.293***	-0.267***	-0.166***
	(0.055)	(0.057)	(0.055)
# apartments:			
$4-6$	-0.201***	-0.224***	-0.169***
	(0.057)	(0.059)	(0.055)
$7-12$	-0.225***	-0.280***	-0.196***
	(0.058)	(0.059)	(0.056)
≥ 13	-0.290***	-0.297***	-0.282***
	(0.072)	(0.076)	(0.066)
Construction year:			
≤ 1918	0.013	0.032	-0.283***
	(0.051)	(0.048)	(0.047)
$1919-1948$	-0.109**	-0.053	-0.374***
	(0.048)	(0.046)	(0.044)
$1949-1957$	-0.072	-0.043	-0.287***
	(0.053)	(0.051)	(0.051)
$1958-1968$	-0.069	-0.032	-0.278***
	(0.045)	(0.044)	(0.042)
$1978-1983$	0.091**	0.087**	0.170***
	(0.046)	(0.044)	(0.042)
$1984-1994$	0.038	0.018	0.222***
	(0.047)	(0.046)	(0.041)
$1995-2001$	0.018	-0.031	0.346***
	(0.048)	(0.046)	(0.039)
≥ 2002	-0.157***	-0.198***	0.201***
	(0.059)	(0.056)	(0.048)
Type of heating system:			
gas heating on the floor	-0.035	-0.050*	-0.031
	(0.029)	(0.029)	(0.029)
oven	0.013	0.070	0.036
	(0.094)	(0.083)	(0.085)
Has modernised:			
Thermal shell	0.038	0.034	0.309***
	(0.035)	(0.032)	(0.031)
Windows	0.040	0.028	0.075**
	(0.029)	(0.028)	(0.032)
Heating system	0.040	0.006	0.099***
	(0.028)	(0.026)	(0.029)

Table 1.2 continued from previous page

log(HDD)		0.064	0.030	0.106
		(0.111)	(0.111)	(0.111)
Fuel type:				
Oil		0.007	-0.002	-0.028
		(0.029)	(0.027)	(0.028)
Long distance heat		-0.150**	-0.148***	-0.103*
		(0.061)	(0.052)	(0.053)
Constant	4.571***	4.633***	5.175***	1.055
	(0.148)	(0.950)	(0.934)	(0.902)
$1/1+\hat{\beta}_{c_{i,t}}$		2.967***	3.250***	
		(0.378)	(0.417)	
Observations	2,310	2,254	2,256	2,778
R-squared	0.451	0.557	–	0.446
Estimator	OLS	OLS	Poisson	OLS

[a] The table reports estimation results from ordinariy least squares (OLS) and Poisson regressions of equations (1.20) and (1.19) respectively. Year fixed effects are included in all regressions and the standard errors are clustered on the household level and reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

than their younger counterparts. The effect is not linear. While the temperature choice is the same across young households, does the mean indoor temperature in the dwelling increase substantially when households older than 50 years are considered. Larger households are also found to consume higher mean indoor temperatures than smaller ones. The effect is strongest when moving from one to two person households. In that case, the mean indoor temperature is estimated to increase by 18.1 % in the considered specification. Further increases of the household size have substantially smaller effects on consumption. Also the presence of children has a weaker effect. Having one or at least two children is estimated to increase the mean indoor temperature by 7.6 % and 11.6 % respectively.

While ownership is found to have a strong significant effect on temperature consumption, the coefficients on the employment status and education level are insignificant.

In column (2) of table 1.2, I also allow for measurement errors in the generated production cost as described in section 1.4. The adjustment factor, $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t})$, is modelled as a function of the type of the dwelling (e.g., single home, detached vs. multi-family dwelling), the number of apartments, the construction year, the fuel and heating system type used, the logarithm of the heating degree days in the given year and whether the thermal insulation of the dwelling, the windows or the heating system have been modernised. Comparing the results of columns (1) and (2) clarifies the importance of accounting for potential errors in the generation of $s_{i,t}$. The impact of income decreases slightly, turning the coefficient insignificant at a 5 % confidence level. The estimated price elasticity decreases slightly to -0.66. A more substantial impact of including the adjustment factor can be observed for the estimated coefficients on household characteristics. While most coefficients keep the same sign and remain statistically significant at a 5 % level of confidence, the sizes of the coefficients are reduced. For example, the estimated effect of being older than 60 is reduced by almost one half to 0.14. The impact of the number of adults remains fairly stable, but no statistically significant effect can be found for children anymore. The point estimates for having one and two children is reduced to 4.8 and 7.3 % respectively. Employment status and education level remain insignificant. However, having a job is now estimated to have a negative effect on the mean temperature choice, consistent with employed people reducing their temperature level while they are at work. Most remarkably, the estimated coefficient on ownership collapses once the adjustment factor is included. While column (1) indicates that owners consume 24.8 % higher indoor temperatures, is the coefficient reduced to 3.5 % and statistically insignificant in column (2). That is, owners do not have stronger preferences for thermal warmth or more wealth allowing them a higher consumption level, as one would conclude from the regression results of column (1), but they tend to live in dwellings for which the overprediction of $s_{i,t}$ is smaller. The lower efficiency yields a higher implied indoor temperature, $\tau^d_{i,t}$, as the dependent variable. In regressions that do not account for the potential measurement error, this increase in the dependent variable is erroneously attributed to the ownership status instead of dwelling characteristics. This highlights the importance of accounting for potential measurement errors when using engineering predictions of efficiency states in such regression analyses. Approaches ignoring the possibility of systematic errors in the engineering model beyond the behavioral assumptions made are likely to obtain biased effects of household characteristics and economic variables on household behavior.

Column (3) reports results from a poisson estimation of equation (1.19), which has recently been argued preferable over the estimation of equations with log dependent variables by ordinary least squares (Silva and Tenreyro, 2006; Gould, 2011). The results indicate that the use of the Poisson estimator yields results qualitatively identical to those of obtained by OLS. While some changes in the estimated coefficients can be observed, such as a stronger effect and a resulting higher significance level for low income households, the results are broadly robust. I favor the linear specification for its simplicity and prevalence that allows many researchers to quickly understand and reproduce the results. Tables A.1 and A.2 in appendix A.1 provide results of all specifications using the Poisson estimator.[25]

[25] Note that the advantages of the Poisson estimator include its emphasis of the functional relationships imposed on $\tau^*(\boldsymbol{x}^{\boldsymbol{\tau}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\tau}}, \varepsilon^{\tau}_{i,t})$ and $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t})$, which I explicitly state in equations (1.17) and (1.18), as well as its ability to easily calculate correct partial effects on the levelized dependent variable. The fact that the regression model

Table 1.2 allows some analyses how the precision of the engineering prediction correlates with observable characteristics of the dwelling. It reports regression coefficients of variables that determine the size of the adjustment factor $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t})$.

As discussed in section 1.4, the mere coefficients do not reflect the partial effects of the respective dwelling characteristics on the adjustment factor $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t})$. They have to be adjusted by the term $1/1+\hat{\beta}_c$ to correct for the countervailing effect of $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t})$ on the dependent variable, τ^d, via households' rebound behavior. The estimates of the correction term, reported in the table, indicate that the size of the coefficients should almost be tripled, if they are to be interpreted in their effect on the adjustment factor $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t})$. Consider for instance how the adjustment of the predicted to the true efficiency state varies with the type of the dwelling. The base group is a single or two-family detached dwelling. The results indicate that the adjustment factor for a row house is substantially smaller than for the base group. The coefficient on row houses is -0.181, implying that the true reduction of the adjustment factor when considering a row house instead of a single or two-family detached dwelling decreases by $0.181 \cdot 2.967 \approx 52.7$ %. Since $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t}) = 1$ would imply that the efficiency state is not adjusted, the overprediction of the efficiency state is larger for the row house than for the detached dwelling.[26] Clearly, these are substantial effects. The adjustment factor, $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t})$, is even smaller for multi-family dwellings. Similarly, the overprediction of the efficiency state is more severe for dwellings with more apartments.

A reasonable explanation for these patterns lies in the simplified fuel requirement calculation applied in this study and described in section 1.5.2. It has been developed primarily for the use in single-family dwellings. Consequently, specifics of dwellings with multiple apartments, such as heat gains from neighbouring apartments, are not well considered. The result is a larger error in the predictions for these dwellings the model is not well suited for. This reasoning also explains the difference of the discussed results, to the findings by Knissel and Loga (2006), who identify the overprediction in dwellings with more apartments to be less severe than in smaller dwellings.[27] The crucial difference of their study is that they use fuel requirements calculated based on detailed audits by a certified expert for each individual site, instead of a simplified procedure that builds on

does not allow to identify level effects of covariates on the temperature choice and adjustment factor, discussed in section 1.4, substantially reduces the attractiveness of the latter property.

[26] Unfortunately, the model does not allow the determination of the absolute level of the adjustment. The reason is, that the intercept of the adjustment factor $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t})$ cannot be identified separately from the intercept of the optimal temperature choice, τ^*. See section 1.4 for details.

[27] Knissel and Loga (2006) study the ratio of predicted and actual fuel consumption. In contrast, I consider the inverted ratio in my analysis. Consequently, an increase of the adjustment term in their analysis is equivalent to a decrease of $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t})$ in the model presented here. In both cases, the gap between predicted and actual consumption is larger.

limited input data. The other additional variables included in both studies, the construction year of the dwelling and an indicator for past modernisations of the thermal shell, have qualitatively identical effects: The overprediction is more severe for dwellings constructed before the 1970's (the base group) and less severe thereafter as well as in dwellings that have received additional thermal insulation.[28] The results of column (4) reported in table 1.2 replicate the study of Knissel and Loga (2006) more closely, by keeping only variables affecting the overprediction of the efficiency state in the regression. Comparing the results of regressions (2) and (4) indicates, that also ignoring household characteristics in the estimation of the adjustment factor, $\lambda(\boldsymbol{x^{\lambda}_{i,t}}, \boldsymbol{\beta^{\lambda}}, \varepsilon^{\lambda}_{i,t})$, leads to biased estimates. The effects of dwellings' construction year and past modernisation activities get stronger, the dwelling type, the number of apartments and the size of the dwelling show weaker effects. This emphasizes again the importance of a regression framework that allows to jointly analyse households' behavior and the patterns of overpredictions of the efficiency state.

Finally, a limitation of the regression framework adopted, also emphasized in section 1.4, is that it does not allow to identify separate effects of one variable on the temperature choice and the adjustment factor. While most variables considered in the empirical analysis can plausibly be assumed to affect only either of the two, this is a problem for the interpretation of the estimated effect of the dwelling size, m. As argued before, it might affect household preferences for mean indoor temperatures. However, the positive relationship between the dependent variable and the dwelling size reported in table 1.2, is inconsistent with the idea that households living in larger dwellings have lower valuations of high mean indoor temperatures over the entire dwelling. In terms of the adjustment factor, $\lambda(\boldsymbol{x^{\lambda}_{i,t}}, \boldsymbol{\beta^{\lambda}}, \varepsilon^{\lambda}_{i,t})$, the estimated coefficient would imply that the engineering model predicts the efficiency levels more precisely for larger dwellings. The correct interpretation of the dwelling size thus remains obscure. It is an interesting question for future research to develop models that allow to separately identify the effects of one covariate on both, the temperature choice as well as the adjustment factor.

1.7 Conclusion

In this paper, I showed how engineering knowledge on the relationship between household behavior, technical characteristics of the dwelling and fuel demand can be used to improve econometric analyses of households' energy consumption. For this purpose, I made three major contributions. First, I proposed the use of a simplified fuel requirement calculator to generate state

[28] An exception are the most recent dwellings constructed after 2003. For these the overprediction of the efficiency state is particularly severe.

variables – indicating the amount of kilowatt hours per square metre the dwellings require to increase the mean indoor temperature by one degree – based on the limited dwelling characteristics observed in the survey data. The generation of these efficiency states enriches the available data with substantial information contained in the engineering model. It allows to relate technological efficiency to behavior in a succinct way and supersedes the need to estimate separate coefficients on imperfect proxies for energy efficiency. This greatly simplifies the regression equation and allows clear interpretations of the estimated parameters. Importantly, while the fuel requirement calculator of Loga et al. (2005) has been well suited for the empirical data at hand, the more recent "EPISCOPE-TABULA" project has developed consistent building typologies for 20 European countries Loga et al. (2016). This allows future researchers across Europe to use the dwelling information available in micro datasets in a similar way as suggested in the current project.

Second, I derived the concrete functional form, relating the efficiency state of the dwelling and household behavior to fuel consumption, from the engineering model used by Loga et al. (2005) and showed that it is linear in households' temperature choice. I thus provided a justification for the linearity assumption that has often been made implicitly in the prior literature. It is an interesting question for future research to explore whether and how this assumption can be relaxed.

Finally, the resulting regression framework can be extended to account for potential systematic errors in the engineering model. The results of the empirical analysis suggest, that it is important to allow for such errors, as estimates of household characteristics are sensitive to the inclusion of dwelling characteristics used to model them in the empirical application.

The empirical results indicate that the marginal cost of temperature consumption as well as household characteristics have a substantial effect on the consumption of thermal comfort. The price elasticity is estimated 0.66 in the preferred specification. There is only a weak positive effect of income on household behavior. In contrast, households' age and size are found to have strong positive effects.

There are several ways how future research can apply the presented framework and extend it to more detailed analyses. Researchers might use it to evaluate natural experiments and controlled field studies on the effect of improvements in energy efficiency on household behavior. From a modelling perspective, it might be interesting to introduce the framework into simultaneous equation models. Mertesacker (2020a) uses the framework to specify and estimate a full household production model of heat demand. Instead of specifying $\tau^*_{i,t}$ in a reduced form, he explicitly derives it from a structural model. This allows him to identify separate effects of variables on household behavior and the size of the adjustment factor and thus the partial effects on the levels of the respective quantities.

2 A structural analysis of households' heat consumption

2.1 Introduction

Residential energy demand for space heating is a major contributor to the overall energy demand in the economy. In 2018 the residential energy sector was responsible for 26.1 % of total final energy consumption in Europe. Roughly 63.3 % of the energy was consumed to heat residential dwellings (Eurostat, 2020). Consequently, understanding the factors that determine households' heat demand is important for policy makers interested in reducing energy consumption of the economy to meet climate targets.

The study of households' fuel demand for heating is complicated by the fact that it depends on the behavior of households living in a dwelling as well as on characteristics of the dwelling itself. While economists are experienced in the analysis of households' choices, it is a common exercise for engineers to predict the amount of fuel required to heat a dwelling given a (fixed) household behavior. This coexistence of different approaches to the study of residential heat demand has frequently raised demand for more interdisciplinary work (Lutzenhiser et al., 2010; Estiri, 2015), which has been addressed rather reluctantly to date.[1]

In this paper, I use the household production model as conceptual framework that links households' utility from the good "indoor temperature" to a technology that is used to produce that good using fuel as input in the production process. I specify households' utility such that they experience satiation effects in temperature consumption and never choose temperature levels above an ideal temperature level that yields blissful thermal comfort.[2] The input demand function for fuel for a given temperature level is taken from Mertesacker (2020b), who shows how a linear

[1]Some energy economy models have tried to combine both approaches in the attempt to predict nationwide energy consumption more reliably (Bataille et al., 2006; Hourcade et al., 2006). However, the economics included in these models to date is rather basic. A discussion of econometric models that have tried to use engineering knowledge in the empirical analysis is provided by Mertesacker (2020b).

[2]Throughout the article, I refer to the *ideal* temperature level whenever I mean the satiation point in households' temperature consumption, which is identical to the consumption level households would choose if they did not face monetary constraints. This is distinct from the *optimal* temperature levels, which reflect households' actual choices. See sections 2.2.2 and 2.2.3 for explicit treatments of the relationship between the two temperature levels.

functional form can be derived from engineering knowledge and be used in empirical applications. Combining both elements with a budget constraint, I thus obtain a theoretically consistent model that imposes structure on the household behavior, that is based on economic as well as engineering reasoning, respectively. I solve the theoretical model for households' optimal temperature choice to derive a regression equation from which the model parameters can be estimated. Following Mertesacker (2020b), I control for errors in the engineering model in the empirical application. The estimates of households' preferences for thermal comfort indicate, that larger and more educated households value high mean indoor temperatures more. In contrast no significant effect of income on household behavior can be found.

This paper makes two major contributions to the literature analysing households' demand for thermal heat: it develops a structural model of households' heat demand and provides a novel approach to the estimation of rebound effects. First, the model shows that a household production framework can be used to combine economic and engineering knowledge in a theoretically consistent way to study households' heat demand empirically. The structural model implies that the estimated parameters have a clear economic interpretation in terms of households' utility from thermal comfort. The model builds on a previous approach by Mertesacker (2020b), who shows how engineering knowledge can be used for the empirical analysis of households' temperature choices. In addition to the structure he imposes on the technology households use to produce thermal comfort, the approach followed in this research also models households' utility explicitly and in a way that satiation effects in thermal warmth consumption are introduced. This allows to explicitly derive households' optimal behavior from a theoretical model, instead of relying entirely on reduced form equations to specify households' energy demand. The additional structure helps to extend the model by Mertesacker (2020b). Concretely, effects of one variable on households' utility and on the size of an error in engineers' predicted fuel consumption can separately be identified, as well as the level of the temperature choice be predicted. Furthermore, the model allows to predict the level of households' temperature choice.

Several studies have recognized the relevance of the household production framework for the empirical analysis of residential heat demand before.[3] However, to my knowledge, this is the first paper to fully specify and estimate a household production model of heat demand. Most approaches have relied on reduced form equations to specify the demand and supply of heat (Scott, 1980) or only used engineering knowledge to impose some structure on the production technology (Hsueh and Gerner, 1993; Mertesacker, 2020b).[4] Dubin et al. (1986) define concrete

[3]See for instance Scott (1980) for a careful exposition of the interdependencies of households' utility for thermal comfort, the production technology, fuel prices and income.

[4]See Mertesacker (2020b) for a review of the literature introducing engineering knowledge into the econometric analysis of heat demand.

functional forms for the utility and input demand function, set up households' optimisation problem and derive the first order condition. However, instead of deriving a regression equation from the associated Marshallian demand for indoor temperature, they regress the ratio of actual and predicted fuel consumption, obtained from an engineering model, on households' income and a measure of the marginal heating cost. Most closely related to the approach of this paper, is the work by Anderson and Kushman (1987). They explicitly model satiation effects in households' utility from thermal warmth and combine it with a simple heat production function based on engineering principles. However, the resulting decision model cannot be solved explicitly. One advantage of the household production model presented in this paper is that it makes reasonable assumptions to obtain a model of household behavior that is easy to solve, estimate and interpret.

The second contribution of this paper is to provide a novel approach to the estimation of households' elasticity with respect to changes in the marginal cost of heat consumption. Previous work has primarily estimated elasticities based on reduced form energy demand equations.[5] A finding of the literature is, that there is some heterogeneity in households' sensitivity to changes in the marginal cost of temperature production (Madlener and Hauertmann, 2011; Aydin et al., 2017). In particular, richer households are found to be less elastic than poorer households. The studies refer to the greater proximity of richer households to the ideal temperature level as one potential explanation for this pattern (Aydin et al., 2017). Implicitly, this assumes that because richer households have less scope to improve their temperature choice, they increase their temperature consumption (weakly) less after a cost reduction, such that the change in consumption relative to the pre-change consumption level gets smaller (implying a lower elasticity). In this paper, this idea is explicitly modelled by the introduction of a satiation point into households' utility function. I show that the size of the elasticity, that can be derived from the theoretical model, exclusively depends on households' temperature choice relative to the ideal temperature level. The heterogeneity in households' elasticity is thus understood as a natural outcome of the preferences and cost that determine their temperature choice. Given the estimated structural model, it is straight forward to predict individual temperature choices and consequently elasticities for every household in the sample. The results show substantial heterogeneity in the predicted price elasticities across households. The mean price elasticity is estimated to be -0.302, but the distribution is right-skewed with many predictions closer to zero and some substantially larger in absolute value. This implies that the median elasticity is only -0.214.

The paper proceeds as follows. Section 2.2 sets up and solves the household production model,

[5]See Risch and Salmon (2017) for a recent review of price elasticities estimated in energy demand equations. If changes in the efficiency level of the dwelling are considered, households' reactions are often interpreted as (direct) rebound effects (take-back effects). See Greening et al. (2000), Sorrell et al. (2009) and Madlener and Turner (2016) for reviews of this literature.

based on which the empirical specification is developed in section 2.3. The data is described in section 2.4, before section 2.5 presents the empirical analyses. Section 2.6 concludes.

2.2 A household production model of residential heat demand

The economic literature has long analysed households' heat demand using household production models. These models suggest to consider thermal comfort as a commodity households have to produce before consumption using (market) goods. Fuel, the quantity typically observed in empirical studies, is thus viewed as a mere input in the production of the desired commodity, that directly enters households' utility function.[6]

Formally, assume in every period, t, households, i, consume thermal comfort from the ambient temperature level and other features of the rooms they inhabit stored in a vector $\boldsymbol{\tau_{i,t}}$. Together with the consumption of a general good, $G_{i,t}$, this yields a period utility $u(\boldsymbol{\tau_{i,t}}, G_{i,t})$. They produce thermal comfort using the production function $W(F_{i,t}, \boldsymbol{D_{i,t}})$. In the short-run the characteristics of the dwelling, stored in the vector $\boldsymbol{D_{i,t}}$, are fixed, such that the amount of fuel, $F_{i,t}$, is the only flexible input factor. Accordingly, the optimal input demand of fuel, $I(\boldsymbol{\tau_{i,t}}, \boldsymbol{D_{i,t}})$, is easily obtained if the production function is invertible:

$$F_{i,t} = I(\boldsymbol{\tau_{i,t}}, \boldsymbol{D_{i,t}}) = W^{-1}(\boldsymbol{\tau_{i,t}}, \boldsymbol{D_{i,t}}). \tag{2.1}$$

The technology available to households constraints their choice set together with the budget constraint:

$$Y_{i,t} = p^F_{i,t} F_{i,t} + G_{i,t}, \tag{2.2}$$

where $Y_{i,t}$ denotes consumers' per period income, $p^F_{i,t}$ is the fuel price and $G_{i,t}$ is the numeraire good. Households maximise their utility subject to these constraints:

$$\begin{aligned} \max_{\boldsymbol{\tau_{i,t}}, G_{i,t}} \quad u(\boldsymbol{\tau_{i,t}}, G_{i,t}) \quad \text{s.t.} \quad & Y_{i,t} = p^F_{i,t} F_{i,t} + G_{i,t} \\ & F_{i,t} = I(\boldsymbol{\tau_{i,t}}, \boldsymbol{D_{i,t}}). \end{aligned} \tag{2.3}$$

It is straight forward to obtain the first order conditions characterising households' optimal choices. For example, let the variable $\tau_{i,t,k}$ from the vector $\boldsymbol{\tau_{i,t}}$ denote a measure of the ambient temperature level. The optimal trade-off between $\tau_{i,t,k}$ and the consumption of other goods, $G_{i,t}$,

[6]See for instance Willett and Naghshpour (1987) for a general household production model of residential demand for energy commodities. Gronau (1986) and Gronau (1997) provide surveys of the general theory of home production.

is described as

$$\frac{\partial u(\cdot)/\partial \tau_{i,t,k}}{\partial u(\cdot)/\partial G_{i,t}} = p^F_{i,t} \cdot \frac{\partial I(\boldsymbol{\tau_{i,t}}, \boldsymbol{D_{i,t}})}{\partial \tau_{i,t,k}} \equiv c_{i,t,k}, \tag{2.4}$$

where the left-hand side represents the marginal rate of substitution between the two goods and $c_{i,t,k}$ denotes the marginal cost of temperature consumption.

The framework is very useful to keep households' utility and technology conceptually distinct. Equation (2.4) emphasizes that households' technology enters the marginal cost of temperature consumption, $c_{i,t,k}$. Economic models that do not consider the role of technology include fuel prices as the only determinant of households' cost of temperature consumption and will therefore attribute changes in demand that are due to changes in technology to changes in tastes (see Pollak and Wachter, 1975)). This mixing up of preference and technology parameters substantially limits their interpretability and consequently the insights that can be gained from the analysis as well as the foundation it provides for the calculation of counterfactual policy scenarios.

The household production framework also emphasizes how economic and engineering knowledge can contribute to the understanding of households' energy demand, respectively. While economists are used to the idea that choices are the outcome of utility functions being optimised subject to constraints that agents face, engineers have substantial knowledge on the functional form of $I(\boldsymbol{\tau_{i,t}}, \boldsymbol{D_{i,t}})$. The subsequent sections refer to both disciplines to specify households' utility and input demand function and to derive an explicit solution for their optimal choices. Section 2.2.1 starts with a description of the input demand function, which follows the specification derived by Mertesacker (2020b). Section 2.2.2 then discusses the functional form of the utility function, before the model solution is derived and discussed in section 2.2.3.

2.2.1 The input demand for thermal comfort

Engineers have developed sophisticated models, that allow to predict the amount of fuel a dwelling requires under given assumptions on households' behavior. They use thermodynamic relationships to map dwelling characteristics and households' choices, stored in the vectors $\boldsymbol{D_{i,t}}$ and $\boldsymbol{\tau_{i,t}}$, into a fuel demand. In terms of the household production model described in section 2.2, the models thus provide very detailed descriptions of the input demand function $I(\boldsymbol{\tau_{i,t}}, \boldsymbol{D_{i,t}})$.

It is tempting to use this knowledge to specify the input demand function in an economic analysis. Yet, the large number of variables and their potentially nonlinear interactions generally make this a non-trivial task. In this paper, I follow the approach by Mertesacker (2020b) to obtain an input demand equation based on engineering knowledge that can be used for the empirical analysis of households' heat consumption. The author simplifies the representation of fuel requirement calculations by decomposing them into the generation of a measure $s_{i,t}$ – that summarizes all physical properties of the dwelling that are independent of household behavior – and the calculation of the total fuel demand given the variable $s_{i,t}$ and assumptions on household behavior. Since $s_{i,t}$ is independent of households' individual choices affecting the amount of heat that is demanded, it represents an indicator for the energetic efficiency of the dwelling that is comparable across dwellings. From an economic perspective, the aggregation of dwelling characteristics into the measure $s_{i,t}$ is part of the data generation. In contrast, the functional form $I(\boldsymbol{\tau_{i,t}}, s_{i,t})$ describes the consequences of households' choices on the amount of fuel that has to be used and thus constraints his behavior.

Mertesacker (2020b) shows how a simplified model by Loga et al. (2005) can be used to generate efficiency states based on limited input data frequently available in micro datasets.[7] Based on the physical relationships used in the same engineering model, he also derives a simple functional form of the input demand function:

$$I(\tau_{i,t}, s_{i,t}) = s_{i,t} \cdot m_{i,t} \cdot \tau_{i,t}. \tag{2.5}$$

The efficiency state $s_{i,t}$ indicates the amount of fuel in kilowatt hours (kWh) the dwelling requires per square metre, m, to increase the mean indoor temperature over the entire heating period and all rooms by one degree Celsius. The scalar variable, $\tau_{i,t}$, captures the mean number of degrees the entire dwelling is heated up over the entire heating period.[8] The product in equation (2.5) thus provides a prediction of the amount of fuel consumed in a dwelling as a function of a single measure of dwellings' energetic performance and and a single choice variable, which is easy to introduce into an economic model. To study the decision households make in greater detail, the next section lines out how the utility they receive from the consumption of a mean temperature increase can be modelled.

2.2.2 Households' utility from indoor temperature

The level of thermal comfort households experience is without doubt a function of numerous interdependent features of all the rooms they inhabit. These include the temperature level of the air and surfaces, the level of humidity and ventilation of the air as well as the degree to which radiation from warm sources can be enjoyed.[9] A formal specification of the utility function,

[7]A discussion how the efficiency states are generated is provided in section 2.4.

[8]Of course households' mean temperature choice is affected by their choice of the indoor temperature consumed in heated living areas as well as the fraction of the total area and daytime they heat, respectively. See Mertesacker (2020b) for a detailed discussion of the factors determining the mean temperature choice.

[9]See for instance the standard DIN EN ISO 7730:2006-05 (German Institute for Standardization, 2006) for a detailed characterization of conditions under which people enjoy thermal comfort. Greening et al. (2000) provide a discussion

that can be used to solve and estimate a theoretical model of heat demand, cannot capture all aspects of reality, but requires substantial simplifications over the true process that yields thermal comfort. In this model, I assume that households' utility function only depends on the mean indoor temperature inside the dwelling.[10]

There are good reasons for focussing on indoor temperatures in a first approximation of reality in economic analyses. First, it is of major importance for the level of thermal comfort households enjoy (see German Institute for Standardization, 2006).[11] Second, even though other variables certainly affect the level of thermal comfort households experience, is the ambient temperature level the crucial choice households' actively make in a static environment during the heating period.[12] By setting the thermostat in all rooms of the dwelling, they can actively control the level of thermal comfort consumed via the temperature level.[13] In contrast, the degree of ventilation or radiation cannot actively be adjusted in the short-run to achieve the desired level of thermal comfort.[14] Third, from a practical perspective, indoor temperatures are easy to comprehend and operationalise for formal and empirical analyses. It is intuitive to most people how indoor temperatures yield thermal comfort and which properties a utility function describing this process should arguably have. It is even quite simple to measure temperature levels in various rooms of the dwelling (or alternatively thermostat settings as a proxy), compared to the measurement of ventilation behavior or the amount of warmth radiating from warm sources.

The presence of satiation effects is a distinctive feature of the utility received from indoor temperatures. It implies that the marginal utility from the consumption of indoor temperatures is not only decreasing, but turns negative for temperature levels above a satiation level, $\overline{\tau}^{in}$, that yields blissful comfort.[15] Consequently, rational households never increase their indoor temperature above the satiation level. One goal of the structural model of households' decision

of factors affecting households' level of thermal comfort in the context of economic models.

[10]See Dubin et al. (1986) and Anderson and Kushman (1987) for two related approaches that focus on temperature choice as the main determinant of thermal comfort. Scott and Capper (1983) provide an early critique on such a focus on (mean) temperature choices.

[11]In particular, humidity only has a minor influence on thermal comfort in moderate climates (German Institute for Standardization, 2006, appendix F).

[12]This might be less clear during the summer period. In particular, ventilators might be used by households to make high temperature levels more enjoyable (German Institute for Standardization, 2006, appendix G). This article focuses on the consumption of thermal heat and thus on the heating period, when outside temperatures are generally low.

[13]Thermostat settings are not necessarily identical to internal temperatures (Vine and Barnes, 1989), but are certainly an important parameter for households to control the ambient temperature level

[14]In the long run, e.g., ventilation might be reduced to increase thermal comfort via modernisations of the thermal shell, for instance by sealing potential leakages.

[15]Note, that the focus on *mean* temperature choices as the variable yielding thermal comfort, does not reduce the validity of the satiation effects modelled. Households reach the point of blissful comfort, if the entire room is heated up to the optimal temperature level, $\overline{\tau}^{in}$, over the entire heating period. Only then would a further increase of mean temperature consumption imply a reduction in households' utility from thermal comfort.

problem, is to introduce these properties into the empirical analysis allowing the identification of additional parameters and a theoretically consistent explanation for heterogeneity in households' price elasticity.

Even though the presence of satiation effects is frequently acknowledged in the literature, it is rarely taken seriously in applied work. For instance, Scott (1980) observes that the marginal utility of thermal comfort turns negative for heat levels sufficiently large, but restricts his analysis on the domain of strictly positive marginal utility for which he considers a utility function with conventional properties. Similarly, Dubin et al. (1986) explicitly model households' utility to depend on the distance to an ideal temperature level, but only define the utility function for temperature choices weakly below the ideal level. The only other study that aims to explicitly introduce a utility function with satiation effects into the econometric estimation of household preferences for thermal comfort, that I am aware of, is provided by Anderson and Kushman (1987). They consider household production models that use utility functions with quadratic, translog and Diewert forms, respectively. Different to the approach of this paper, they do not assume temperature consumption and general goods to be additively separable in households' utility function and also allow general goods to enter households' utility in a quadratic way. In addition, Anderson and Kushman (1987) also estimate the bliss point empirically, instead of setting it to a predefined level. The drawback of the flexible functional forms they choose, is that they are not able to solve the resulting model for the Marshallian demands explicitly. As a result, the authors have to manipulate the first order conditions to obtain an estimable regression equation.

I model households to receive a quasilinear utility from the consumption of general goods, $G_{i,t}$, and the choice of a mean temperature increase, $\tau_{i,t}$, over the outdoor temperature level, $\tau_{i,t}^{out}$.[16] The utility households receive from the consumption of $\tau_{i,t}$ is assumed to be quadratic. Utility is increasing up to a temperature increase, $\overline{\tau}_{i,t}$, that results in the consumption of the ideal temperature level, $\overline{\tau}^{in}$, in the entire dwelling and over the full heating period:

$$u(\tau_{i,t}, G_{i,t}) = \beta^G G_{i,t} - \boldsymbol{\beta^\tau x_{i,t}^\tau}(\overline{\tau}_{i,t} - \tau_{i,t})^2. \tag{2.6}$$

[16] I specify households' utility as a function of temperature increases rather than temperature levels, for consistency with the formulation of the input demand equation (2.5). The utility function and all results that follow can be restated in terms of $\tau_{i,t}^{in}$ using the identity

$$\tau_{i,t} = \tau_{i,t}^{in} - \tau_{i,t}^{out}.$$

Analogously, the defined satiation level indicates the temperature increase when heating up to the ideal temperature level, $\overline{\tau}_{i,t}^{in}$. That is,

$$\overline{\tau}_{i,t} = \overline{\tau}^{in} - \tau_{i,t}^{out}.$$

The marginal utility of general goods, β^G, is constant and equals the marginal utility of income. Normalising it to unity therefore implies, that all utility values can be interpreted in monetary terms. The marginal utility of temperature consumption is determined by the product of the row and column vectors $\boldsymbol{\beta^\tau}$ and $\boldsymbol{x^\tau_{i,t}}$ of equal length. Households' preferences for thermal warmth are thus allowed to vary with observable characteristics in the vector $\boldsymbol{x^\tau_{i,t}}$. The estimation of the vector $\boldsymbol{\beta^\tau}$ containing the parameters $\beta^\tau_{i,t,k}$, measuring the effect of the k-ths variable in $\boldsymbol{x^\tau_{i,t}}$ on households' preferences for indoor temperature, is the essential purpose of the empirical analysis.

The use of a quasilinear function to model households' utility from indoor temperature consumption, provides a straight forward way to introduce satiation effects into the analysis. In the model, the ideal temperature level is considered a universal value of humans temperature sensation.[17] Households vary only with respect to the discomfort that they receive from deviations from that ideal level. This avoids identification problems and clarifies the interpretation estimated coefficients have in terms of the theoretical model.[18] The choice of a quadratic function is the most parsimonious way to introduce satiation effects in temperature consumption into the economic model. The resulting economic decision model is easy to solve, understand and interpret.

An implication of the quasilinear specification is that income effects are excluded from the theoretical model. This is generally considered a valid approximation if the expenditure for a good is only a small share of total income (Vives, 1987). In Germany households spent on average 2.9 % of their income on thermal heating in 2014 (Bach et al., 2018), such that absence of income effects should not be too important. To control for income as an important household characteristic, I include it in the vector of observable preference shifters, $\boldsymbol{x^\tau_{i,t}}$, in the empirical analysis.

2.2.3 Model Solution

Since the input demand function (2.5) as well as the utility function (2.6), depend on the mean temperature as the households' only choice variable, both can be combined into a household production model as described by equation (2.3). Households' decision problem reduces to the trade-off between two goods, $\tau_{i,t}$ and $G_{i,t}$, that is solved by the standard equation of their marginal rate of substitutions and the respective price ratio as indicated by equation (2.4). The marginal

[17] This is consistent with the general conditions under which people enjoy thermal comfort defined in the standard DIN EN ISO 7730:2006-05 (German Institute for Standardization, 2006). While these conditions vary with the context of peoples activities (e.g., the bath or the bedroom), they are independent of any personal characteristics (such as their age).

[18] Also note that while there is some consensus on general conditions under which people enjoy thermal comfort, used for instance to specify the standard DIN EN ISO 7730:2006-05 (German Institute for Standardization, 2006), no general rules describing the disutility experienced from deviating from an ideal temperature level exist. Accordingly, it is a reasonable approach to fix the ideal temperature level based on available knowledge and empirically estimate how households' utility is affected by indoor temperatures below the optimal level.

cost of a temperature increase, $c_{i,t}$, are determined by the product of the fuel price, $p^F_{i,t}$, and the marginal increase in fuel input associated to the additional consumption of $\tau_{i,t}$. Given the input demand equation (2.5), they have a simple form:

$$c_{i,t} = c(s_{i,t}) = p^F_{i,t} \cdot s_{i,t} \cdot m_{i,t}. \tag{2.7}$$

Households producing indoor temperature more efficiently have a lower state variable, $s_{i,t}$, and thus face lower marginal cost, $c(s_{i,t})$, in the consumption of $\tau_{i,t}$. The linearity of the input demand equation implies that $c(s_{i,t})$ is independent of the level of indoor temperature.

While the marginal utility of the linear good, $G_{i,t}$, is constant, the quadratic form of the utility households receive from the consumption of indoor temperature implies, that the respective marginal utility depends on the level of temperature consumption:

$$\frac{\partial u(\cdot)}{\partial \tau_{i,t}} \equiv MU^{\tau}_{i,t} = 2\boldsymbol{\beta^{\tau}} \boldsymbol{x^{\tau}_{i,t}} [\overline{\tau}_{i,t} - \tau_{i,t}] > 0 \qquad \forall \qquad \tau_{i,t} < \overline{\tau}_{i,t}. \tag{2.8}$$

Concretely, the marginal utility decreases at a constant rate in the level of temperature consumption. A marginal increase of the temperature level consumed in the dwelling thus increases households' utility only up to the ideal temperature increase, $\overline{\tau}_{i,t}$, and decreases utility if it exceeds that level. Since the marginal utility of general goods, $G_{i,t}$, is strictly positive, this implies that a rational household never consumes $\tau_{i,t} > \overline{\tau}_{i,t}$.

Assuming an interior solution, the Marshallian demand of the temperature choice is easily derived as

$$\tau^*_{i,t} = \overline{\tau}_{i,t} - \frac{\beta^G}{2\boldsymbol{\beta^{\tau}} \boldsymbol{x^{\tau}_{i,t}}} \cdot c_{i,t}. \tag{2.9}$$

Households' temperature consumption equals the ideal temperature increase minus some function that depends on the relative marginal utilities of the two goods and the marginal cost of a temperature increase. Households with lower marginal cost or a higher valuation of thermal comfort consume higher temperature levels. The ideal temperature level would only be consumed if $c_{i,t}$ was zero or $\boldsymbol{\beta^{\tau}} \boldsymbol{x^{\tau}_{i,t}}$ infinitely large.[19]

Considering the derivative of $\tau^*_{i,t}$ with respect to $c_{i,t}$ clarifies that decreases in the marginal

[19]Note that the assumption of an interior solution implies that $c_{i,t}$ must not become too large relative to $\boldsymbol{\beta^{\tau}} \boldsymbol{x^{\tau}_{i,t}}$, as otherwise equation (2.9) would imply $\tau^*_{i,t} < 0$.

cost of temperature consumption, increase the optimal temperature level at a constant rate:

$$\frac{\partial \tau^*_{i,t}}{\partial c_{i,t}} = -\frac{\beta^G}{2\boldsymbol{\beta^\tau x^\tau_{i,t}}} < 0. \tag{2.10}$$

For households with higher valuation of thermal comfort adjustments after cost reductions are smaller in magnitude than for households with a low valuation. Because their marginal utility decreases more for a given temperature increase, a smaller adjustment of the temperature choice is required to equate the marginal rate of substitution with the lower marginal cost in equation (2.4).

Furthermore, note that the size of the change in $\tau^*_{i,t}$ associated to a drop in $c_{i,t}$ is independent of the level of indoor temperature households consume. This directly results from the previous observation that the decrease in $MU^\tau_{i,t}$ associated to a temperature increase, is independent of the level of indoor temperature (i.e. $MU^\tau_{i,t}$ decreases at a constant rate). Accordingly, also the size of the adjustment in $\tau_{i,t}$ required to balance both sides of equation (2.4) after a change in the marginal cost of temperature consumption is independent of the level of indoor temperature. This implies that, for larger $\tau^*_{i,t}$, the change in temperature consumption as reaction to changes in $c_{i,t}$ decreases relative to the pre-change temperature level. Accordingly, households with higher $\tau^*_{i,t}$ are less elastic than households with low consumption levels.

Formally, the elasticity can be derived as

$$\varepsilon_{\tau_{i,t},c_{i,t}} = \frac{\partial \tau_{i,t}}{\partial c_{i,t}}\frac{c_{i,t}}{\tau_{i,t}} = -\frac{\beta^G}{2\boldsymbol{\beta^\tau x^\tau_{i,t}}} \cdot \frac{c_{i,t}}{\tau_{i,t}}. \tag{2.11}$$

Observing from equation (2.9) that $\frac{\beta^G}{2\boldsymbol{\beta^\tau x^\tau_{i,t}}} \cdot c_{i,t} = \overline{\tau}_{i,t} - \tau^*_{i,t}$, the elasticity can be restated as a function of the optimal temperature choice, $\tau^*_{i,t}$, and the ideal temperature level $\overline{\tau}_{i,t}$:

$$\varepsilon_{\tau_{i,t},c_{i,t}} = -\frac{\overline{\tau}_{i,t} - \tau^*_{i,t}}{\tau^*_{i,t}}. \tag{2.12}$$

That is, the elasticity equals the negative of the percentage change in the temperature level when moving from the current to the ideal temperature choice. This change will be small for households located close to $\overline{\tau}_{i,t}$ and gets the larger the less thermal comfort households consume.[20]

Intuitively, the decrease in households' elasticity in chosen temperature increases, $\tau^*_{i,t}$, results from the presence of a satiation point, $\overline{\tau}_{i,t}$, in households' temperature consumption. For the

[20] If the distance to the ideal temperature level is sufficiently large, households get very elastic. Concretely, the model implies the elasticity to be smaller than -1 if $\tau^*_{i,t} < \overline{\tau}_{i,t}/2$. In this case, the increase in temperature consumption after a cost reduction is so large, that in total more money is spent on the consumption of thermal warmth and the consumption of general goods, $G_{i,t}$, decreases. The substitution effect towards the consumption of indoor temperature outweighs the income effect from the cost reduction for the good $G_{i,t}$.

elasticity to be non-decreasing in the temperature choice, a necessary condition is that reactions to changes in the marginal cost are stronger for higher pre-change levels of $\tau_{i,t}$.[21] However, this is not reasonable to expect if a satiation point limits the extent to which temperature levels are increased. If anything, the size of adjustments after price changes should decline as the scope for improvements of the temperature level decreases. Consequently, an elasticity that declines in the temperature choice is a reasonable outcome in situations in which satiation points exist. In the presented model this is introduced by a utility function that implies adjustments of temperature consumption after price changes that are independent of the pre-change temperature choice, as outlined in the previous paragraphs.

Previous literature, studying rebound effects of households after efficiency increases, has found that richer households are less elastic towards changes in the price of heat consumption than their poorer counterparts (Madlener and Hauertmann, 2011; Aydin et al., 2017). It refers to the greater proximity of richer households to the ideal temperature level as one potential explanation for this pattern (Aydin et al., 2017). The presented model rationalizes this idea and emphasizes that the same reasoning also applies for any other factor that determines households' temperature choice, such that income might only be one of many variables able to create heterogeneity in households' price responsiveness.

Theoretically, there is a wide range of elasticities that can result. The empirical identification of the unobservable true elasticity requires a) the theoretical model to correctly describe households' behavior and b) the empirical model to identify the temperature level chosen by households. In the next section, I develop an empirical model based on equation (2.9) and outline the estimation procedure. The model allows to identify utility parameters and thus to predict households' temperature choices and consequently their elasticity towards price changes.

2.3 Empirical specification

Inserting households' optimal temperature choice, $\tau^*_{i,t}$, from equation (2.9) into the input demand function $I(\tau_{i,t}, s_{i,t}, m_{i,t})$ in equation (2.5), the optimal demand for fuel, $F^*_{i,t}$, can be derived. Relating this to the observed fuel consumption, $F^d_{i,t}$, a nonlinear regression equation can be

[21] This observation directly results from the definition of an elasticity, which is determined by the size of a temperature change *relative* to the choice before the price change.

obtained:

$$F_{i,t}^{d} = F_{i,t}^{*} = s_{i,t} \cdot m_{i,t} \cdot \tau_{i,t}^{*}(s_{i,t}; \beta^{G}, \boldsymbol{\beta^{\tau}}, \varepsilon_{i,t}^{\tau}) \tag{2.13}$$

$$= s_{i,t} \cdot m_{i,t} \cdot \left(\overline{\tau}_{i,t} - \frac{\beta^{G}}{2\boldsymbol{\beta^{\tau} x_{i,t}^{\tau}}} c(s_{i,t}) + \varepsilon_{i,t}^{\tau} \right). \tag{2.14}$$

The error $\varepsilon_{i,t}^{\tau}$ accounts for variations in $F_{i,t}$ that cannot be explained by observed preference shifters in the vector $\boldsymbol{x_{i,t}^{\tau}}$ or by changes in $s_{i,t}$, $m_{i,t}$ or $p_{i,t}^{F}$, determining the marginal cost of temperature consumption $c(s_{i,t})$. Since there are no parameters to be estimated on the interaction term $s_{i,t} \cdot m_{i,t}$, an algebraically equivalent representation of the regression equation is

$$\frac{F_{i,t}^{d}}{s_{i,t} m_{i,t}} \equiv \tau_{i,t}^{d} = \tau_{i,t}^{*}(s_{i,t}; \beta^{G}, \boldsymbol{\beta^{\tau}}, \varepsilon_{i,t}^{\tau}) \tag{2.15}$$

$$= \overline{\tau}_{i,t} - \frac{\beta_{G}}{2\boldsymbol{\beta^{\tau} x_{i,t}^{\tau}}} \cdot c(s_{i,t}) + \varepsilon_{i,t}^{\tau}. \tag{2.16}$$

This representation clarifies that the model implicitly regresses households' implied temperature choice – which would have led to the observed fuel consumption according to input demand function $I(\tau_{i,t}, s_{i,t}, m_{i,t})$ – on the temperature choice predicted by the economic model. It thus emphasizes the goal of the analysis to study households' temperature choice and the role of the input demand function derived from the engineering model to link these unobserved choices to observable fuel data.

In practical applications, efficiency states, $s_{i,t}^{e}$, obtained from an engineering model, are likely to systematically overestimate the amount of fuel dwellings require per degree temperature increase.[22] This introduces measurement error in the dependent variable, $\tau_{i,t}^{d}$, as well as the independent variable $c(s_{i,t}^{e})$ of equation (2.16). To avoid biases in the estimated coefficients, I follow Mertesacker (2020b) to introduce a proxy, $\lambda(\boldsymbol{x_{i,t}^{\lambda}}, \boldsymbol{\beta^{\lambda}}, \varepsilon_{i,t}^{\lambda})$, for the unobserved measurement error in the empirical application. Concretely, it is assumed that there exists a function of covariates stored in a vector, $\boldsymbol{x_{i,t}^{\lambda}}$, that can be used to adjust $s_{i,t}^{e}$ to the unobserved true efficiency state of the dwelling $s_{i,t}^{\circ}$:

$$s_{i,t}^{\circ} = \lambda(\boldsymbol{x_{i,t}^{\lambda}}, \boldsymbol{\beta^{\lambda}}, \varepsilon_{i,t}^{\lambda}) \cdot s_{i,t}^{e}. \tag{2.17}$$

[22] Engineering predictions of fuel consumption are known to systematically overpredict actual consumption (Sunikka-Blank and Galvin (2012), Laurent et al. (2013)). Mertesacker (2020b) argues that – besides erroneous engineering assumptions on household behavior – systematic overpredictions of $s_{i,t}$ are likely to contribute to the gap between actual and predicted consumption.

Since engineering models typically overpredict actual consumption, $\lambda(\boldsymbol{x_{i,t}^{\lambda}}, \boldsymbol{\beta^{\lambda}}, \varepsilon_{i,t}^{\lambda})$ is expected to lie in the interval between zero and one for most households. A simple intercept, β_0^{λ}, is able to capture the general size of the gap between actual and predicted consumption. In addition, dwelling characteristics are included in $\boldsymbol{x_{i,t}^{\lambda}}$ to capture potential systematic variations in the size of the overprediction. A failure to control for such systematic effects could imply household characteristics, $\boldsymbol{x_{i,t}^{\tau}}$, to be endogenous if they are correlated to elements in the vector $\boldsymbol{x_{i,t}^{\lambda}}$. In the empirical analysis, I control for many dwelling characteristics such that the remaining unobserved component of the measurement error $\varepsilon_{i,t}^{\lambda}$ is arguably exogenous to $\boldsymbol{x_{i,t}^{\tau}}$ and neglectable.

To introduce $\lambda(\boldsymbol{x_{i,t}^{\lambda}}, \boldsymbol{\beta^{\lambda}}, \varepsilon_{i,t}^{\lambda})$ into the empirical specification, I assume it can be represented by a linear function:[23]

$$\lambda(\boldsymbol{x_{i,t}^{\lambda}}, \boldsymbol{\beta^{\lambda}}, \varepsilon_{i,t}^{\lambda}) = \beta_0^{\lambda} + \boldsymbol{\beta^{\lambda}} \boldsymbol{x_{i,t}^{\lambda}} + \varepsilon_{i,t}^{\lambda}. \tag{2.18}$$

Introducing (2.18) and (2.17) into (2.14), the final regression equation is obtained:

$$F_{i,t}^{d} = s_{i,t} m_{i,t} \cdot \left(\beta_0^{\lambda} + \boldsymbol{\beta^{\lambda}} \boldsymbol{x_{i,t}^{\lambda}}\right) \cdot \left(\overline{\tau}_{i,t} - \frac{\beta_G}{2\boldsymbol{\beta^{\tau}} \boldsymbol{x_{i,t}^{\tau}}} \cdot c_{i,t} \cdot \left(\beta_0^{\lambda} + \boldsymbol{\beta^{\lambda}} \boldsymbol{x_{i,t}^{\lambda}}\right)\right) + \varepsilon_{i,t}. \tag{2.19}$$

where the error term, $\varepsilon_{i,t}$, is a function of the individual errors $\varepsilon_{i,t}^{\lambda}$ and $\varepsilon_{i,t}^{\tau}$.

Under the assumptions made, equation (2.19) allows to jointly identify both, the effects of dwelling characteristics on the size of the overprediction of the efficiency states as well as the heterogeneity in households' preferences for thermal comfort. The capabilities of the developed regression framework thus go beyond the approach suggested by Mertesacker (2020b). His approach required exclusion restrictions for covariates to be included either in the vector of variables affecting households' temperature choice, $\boldsymbol{x_{i,t}^{\tau}}$, or in the vector of variables affecting the size of the adjustment factor, $\boldsymbol{x_{i,t}^{\lambda}}$. Furthermore, it was restricted to the interpretation of relative effects of covariates on the dependent variables.

Here, the additional structure imposed on the model via the utility function helps to disentangle effects on the adjustment term from those on households' temperature choice. The satiation effects imply that the impact of variables in $\boldsymbol{x_{i,t}^{\tau}}$ on the dependent variable, $F_{i,t}^{d}$, is decreasing and converges to zero if preferences for thermal comfort are very high relative to the associated cost, implying that households consume temperature levels close to the satiation point. In contrast, the adjustment factor is modelled to have a constant impact. A variable included in $\boldsymbol{x_{i,t}^{\tau}}$ as well as $\boldsymbol{x_{i,t}^{\lambda}}$ thus affects the dependent variable differently through both mechanisms, implying that its

[23] In equation (2.18) the intercept, β_0^{λ}, is made explicit. With small abuse of notation the vectors $\boldsymbol{\beta^{\lambda}}$ and $\boldsymbol{x_{i,t}^{\lambda}}$ keep the same names even though their dimensionality is reduced by one variable (a constant) and parameter, respectively.

impact in both factors can be estimated in one regression.

2.4 Data

For the empirical analysis a dataset created by Mertesacker (2020b) is used. The dataset contains information on fuel consumption, household characteristics, dwelling characteristics and investment behavior, fuel prices and the efficiency states of the dwelling and thus all information required for the estimation of the structural model.

The main source behind the dataset is "The German Residential Energy Consumption Survey" (RWI and forsa, 2016).[24] The survey is based on a random sample of $6,715$ German households, that have been interviewed in 2010. The dataset includes information about household and dwelling characteristics as well as the energy consumption between 2006 and 2008.

The final analysis is restricted to households using natural gas, long distance heating or oil in their primary heating system. All fuel consumption data used is based on available billing data. Households that did not possess their energy bills anymore have been dropped from the analysis. Mertesacker (2020b) finds that this introduces a sampling bias towards owner-occupied dwellings, which are more likely to be able to report the amount of fuel consumed in the previous years. The results presented in section 2.5 should therefore be interpreted with some caution, as they might not generally apply to a representative population. However, robustness checks reported in appendix B.1 indicate, that the main results also hold in the subpopulations of owners and tenants only. This is consistent with more detailed analyses conducted by Mertesacker (2020b).

A caveat of the available household and dwelling characteristics is, that most variables of interest are only observed in 2010, at the time the interview was conducted. Besides the fuel consumption data, only the total number of persons living in the household as well as the heating degree days are available for the years 2006 to 2008. Another limitation of the survey is that individual specific information is only available for the survey respondent.[25] I use this information to proxy for characteristics of the entire household in the empirical analysis.

Household characteristics are either directly taken from the dataset, potentially after small modifications, or generated based on the available data. The age of the household is calculated based on the year of birth of the survey respondent. Similarly, the number of children living in the household is obtained from the reported dates of their births. This allows to track the number of adult persons in the household as the difference between the total number of household members

[24]The subsequent description of the data closely follows Mertesacker (2020b). He also provides additional detailed descriptions of the data cleaning and the generation of efficiency states, including a discussion of potential selection concerns.

[25]The exception is date of birth, which is also available for all children living in the households.

reported for a given year and the calculated number of children. For the empirical analysis, the number of adults and children is top-coded at four and two people respectively. The categorical income information in the data is used to obtain indicators for households being in a low, middle or high income group. Households are assigned to a low income group if they have less than $1,500$ euros, to a middle income group if they earn between $1,500$ and $3,500$ euros and a high income group if their income exceeds $3,500$ euros per month.[26] Finally, a variable whether a household has a job is generated based on occupation information available in the data. Binary indicator variables whether the household has a high-school degree (the German "Abitur") and possesses the dwelling he lives in, in the year the survey is conducted, can be obtained from the survey directly. I assume that these variables have stayed the same between 2006 and 2010. Fewer changes are required to prepare dwelling characteristics for the empirical analysis. An indicator of the relative size of the dwelling is generated based on its relative deviation from the mean dwelling size (in square metres) over the entire sample. Dwellings are assigned to the group of small, medium or large dwellings based on the terciles of the distribution. In addition, the number of construction year categories is reduced from nine, available in the primary data, to only five categories in the final analysis, to reduce the number of variables in the final estimation.

To obtain price data and information on efficiency states Mertesacker (2020b) combines the main data from the "The German Residential Energy Consumption Survey" with information about average fuel prices households had to pay between 2006 and 2008 and with information on dwellings' efficiency state, $s_{i,t}$. While the price data is obtained from the German ministry of economics (Bundesministerium für Wirtschaft und Energie, 2018), the efficiency states, indicating the amount of fuel (in $^{kWh}/_{m^2}$) required to increase the mean indoor temperature by one degree Celsius have to be generated. For this purpose, Mertesacker (2020b) uses an engineering calculation procedure developed by Loga et al. (2005). Their program has been developed with the intention to facilitate the creation of energy performance certificates for home owners and to provide guidance on potential savings that can be realized through modernisations. The program requires only few dwelling characteristics that are easily observable for home owners as inputs. Mertesacker (2020b) shows that this program can also be used to predict the efficiency states and fuel requirements for every household observed in the main dataset.

[26]The reduction of the number of income groups is convenient to a) ensure a sufficiently large number of observations in each income bin and b) reduce the number of variables included in the final regression. Unreported results confirm the robustness of the main empirical findings towards alternative definitions of income bins.

2.5 Empirical analysis

To obtain estimable regression equations, I normalise the ideal temperature level, by setting $\overline{\tau}^{in} = 21$.[27] Accordingly, the ideal temperature increase of household i in period t is normed to $\overline{\tau}_{i,t} = 21 - \tau^{out}_{i,t}$. The marginal utility of general goods, β_G, is set to unity, allowing utility to be interpreted in monetary values. I include several covariates into the regression to allow for the utility from the consumption of thermal warmth to vary with households' age, the number of adults and children living in the household, the employment status, education level, an indicator for ownership of the dwelling, the income level and the size of the dwelling. To control for systematic biases in the generated efficiency states, the adjustment factor includes an intercept and varies with indicators describing characteristics of the dwelling. In the preferred specification, the dwelling characteristics included in the adjustment term consist of the type of the dwelling, the number of apartments, the construction year, indicators for past modernisation investments, year and fuel type fixed effects, the number of heating degree days and the income level of the household as well as the dwelling size. All specified regression equations are estimated by nonlinear least squares.

I first estimate and discuss the estimated preference parameters of the utility function and show how the introduction of a technical adjustment factor affects the estimation results. I then use the estimated model to predict mean temperature choices in the sample and to obtain the estimate of the mean elasticity based on the theoretical model. The results indicate that it is important to control for measurement errors in the generated efficiency states. Finally, I use the structural model to provide interpretations of the estimates with respect to concrete economic variables. I show how heterogeneity in preferences for thermal comfort leads to heterogeneity in behavior including the temperature choice as well as the responsiveness to price changes.

2.5.1 Estimation of the structural model

Table 2.1 reports regression results of different variants of the empirical model developed in section 2.3. Columns (1) to (3) of the table differ in their treatment of the generated efficiency state. In column (1) no adjustment for potential measurement errors is included. Implicitly, the efficiency state generated by the engineering model is thus assumed to correctly represent the true efficiency level and the empirical model reduces to equation (2.16). Column (2) models the

[27] A temperature of 21 degrees Celsius is often considered as a temperature that maximises households' level of thermal comfort. For instance, it defines the indoor temperature in the main living area in the standard heating regime, which is typically used to identify fuel poverty (Harrington et al., 2005).

Tables B.1 and B.2 in appendix B.1 confirm that the main results of the empirical analysis are robust to local changes of the ideal temperature level.

adjustment factor, $\lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \boldsymbol{\beta}^{\boldsymbol{\lambda}}, \varepsilon^{\lambda}_{i,t})$, as an additional intercept, β^{λ}_0, to be estimated. It thus controls for the general overprediction of $s_{i,t}$ by engineering models, but not for potential systematic correlations of the size of the overprediction with dwelling characteristics potentially resulting in endogenity of the explanatory variables in $\boldsymbol{x}^{\boldsymbol{\tau}}_{\boldsymbol{i,t}}$. In column (3) the full adjustment factor is introduced, resulting in the main regression model of equation (2.19). All variants are estimated by nonlinear least squares.

A comparison of the regression results reported in columns (1) through (3) clarifies, that it is important to control for measurement errors in $s_{i,t}$ and that the mere inclusion of an additional intercept is not sufficient to raise confidence in the results. In fact, even though the estimate of β^{λ}_0 reported in column (2) is 0.69, implying a substantial downscaling of the efficiency states to 69 % of their predicted values, the estimated coefficients on households' preferences remain similar to those reported in column (1). Once additional covariates are added in column (3), to control for systematic variation of the measurement error with dwelling characteristics, substantial changes in the estimated coefficients are visible. Most remarkably, the estimated coefficient on the ownership of the dwelling a household inhabits collapses and turns from high economic and statistical significance to complete irrelevance. That is, owners do not have stronger preferences for thermal warmth, as one would conclude from the regression results of columns (1) and (2), but they tend to live in dwellings for which the overprediction of the efficiency state, $s_{i,t}$, is smaller. Similarly, substantial changes can be observed for the estimated coefficients on households' age, size, employment status, education level as well as the size of the dwelling they inhabit.

At the same time the adjustment factor is found to vary substantially with observable dwelling characteristics. For instance, it is 0.191 units smaller for a row house than for a single or two-family detached dwelling (the base group). This implies that the efficiency states are predicted more accurately for the latter, resulting in the adjustment factor being closer to unity. This is consistent with the results found by Mertesacker (2020b). As he points out, a reasonable explanation for this pattern is that the engineering model by Loga et al. (2005), used for the prediction of the efficiency states, has been primarily developed for the use in single-family dwellings. The size of the overprediction of the efficiency state also varies significantly with the number of apartments in the dwelling, its construction year, the existence of modernisation investments in previous years as well as the type of fuel used. At least some of these variables are likely to also correlate with variables affecting households' preferences for thermal comfort. Consequently, the estimates in columns (1) and (2) are likely to suffer from omitted variable bias.

Given the high relevance of controlling for measurement errors in the efficiency state, $s_{i,t}$, I focus on column (3) for interpretation. The estimates indicate a significant positive effect of

Table 2.1: Regression coefficients for different specifications of the adjustment term [a]

Dep. Var: $F^d_{i,t}$	(1)	(2)	(3)	(4)
Estimates of households' utility function parameters:				
Constant: β^τ_0	11.598***	10.047***	12.256***	12.572***
	(0.973)	(1.168)	(3.578)	(3.233)
age:				
$30-39$	1.480**	1.062*	1.262	1.393
	(0.604)	(0.643)	(3.077)	(2.772)
$40-49$	1.424**	1.149*	0.094	0.353
	(0.576)	(0.680)	(3.115)	(2.769)
$50-59$	2.143***	2.130***	3.918	3.954
	(0.620)	(0.738)	(3.356)	(3.038)
≥ 60	3.531***	3.890***	4.811	4.972
	(0.851)	(1.035)	(3.664)	(3.318)
# adults:				
2	0.866*	0.936*	3.014**	2.661***
	(0.484)	(0.536)	(1.216)	(0.982)
3	1.536**	2.351**	6.654**	6.355**
	(0.712)	(1.026)	(2.839)	(2.832)
≥ 4	1.479	2.723**	3.850*	3.253*
	(0.986)	(1.348)	(2.041)	(1.797)
# children:				
1	0.362	0.803	-0.368	-0.031
	(0.729)	(0.829)	(1.329)	(1.276)
≥ 2	0.584	0.918	2.964*	2.937*
	(0.729)	(0.904)	(1.718)	(1.718)
Is employed	0.542	0.413	-2.234	-2.286
	(0.586)	(0.586)	(2.109)	(2.005)
Has Abitur	0.141	0.302	3.271***	3.002**
	(0.423)	(0.517)	(1.262)	(1.222)
Is owner	1.771***	3.040***	-0.033	0.153
	(0.492)	(0.705)	(1.324)	(1.234)
Income:				
$< 1,500$ €	-0.500	-0.652	2.329	-0.010
	(0.541)	(0.600)	(2.151)	(0.025)
$\geq 3,500$ €	0.751	0.815	0.320	0.035
	(0.602)	(0.776)	(1.283)	(0.023)

Table 2.1 continued from previous page

Size of the dwelling:				
Small: $<$ 1st tercile	-4.045***	-4.138***	-1.103	-1.612
	(0.503)	(0.781)	(1.447)	(1.282)
Large: $>$ 2nd tercile	7.354***	7.156***	5.322***	5.524***
	(0.790)	(1.206)	(1.385)	(1.400)
Estimates of adjustment term:				
Constant: β_0^λ		0.690***	0.655***	0.656***
		(0.015)	(0.040)	(0.042)
Type of the dwelling:				
Row house			-0.191***	-0.193***
			(0.025)	(0.025)
Multi-family dwelling			-0.230***	-0.233***
			(0.038)	(0.039)
# apartments:				
$4-6$			-0.054*	-0.056*
			(0.031)	(0.032)
$7-12$			-0.081**	-0.084**
			(0.032)	(0.033)
≥ 13			-0.094***	-0.102***
			(0.034)	(0.035)
Construction year:				
$1919-1968$			-0.017	-0.021
			(0.028)	(0.029)
$1969-1977$			0.111***	0.109***
			(0.037)	(0.038)
$1978-1994$			0.230***	0.228***
			(0.037)	(0.037)
≥ 1995			0.160***	0.156***
			(0.035)	(0.035)
Has modernised:				
Windows			0.031	0.029
			(0.026)	(0.026)
Heating system			0.009	0.004
			(0.024)	(0.024)
Thermal shell			0.102***	0.104***
			(0.028)	(0.029)
log(HDD)			0.187*	0.190*
			(0.096)	(0.098)

Table 2.1 continued from previous page

Income:				
$< 1,500$ €			-0.025	
			(0.020)	
$\geq 3,500$ €			0.029	
			(0.022)	
Size of the dwelling:				
Small: $<$ 1st tercile			-0.033	-0.023
			(0.027)	(0.026)
Large: $>$ 2nd tercile			0.000	-0.002
			(0.025)	(0.025)
Fuel type:				
Oil			-0.043*	-0.043*
			(0.024)	(0.024)
Long distance heat			-0.058**	-0.059**
			(0.027)	(0.028)
Observations	2,314	2,314	2,256	2,256
R-squared	0.813	0.864	0.892	0.892

[a] The table reports estimation results of different variants of the empirical model developed in section 2.3. The parameter estimates in column (1) are obtained from a regression of equation (2.14). Colums (2) to (4) report estimates from regressions of equation (2.19) using different sets of covariates. Year fixed effects are included in the adjustment term in columns (3) and (4). In column (4) the income variables are constrained to have the same impact on household behavior and the adjustment of the efficiency state. The ideal temperature level, $\overline{\tau}_{i,t}^{in}$, is set to 21 degrees Celsius. All results are obtained by nonlinear least squares. Standard errors are reported in parantheses and clustered on the household level. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

household size on mean temperature consumption. The impact of the third adult living in the dwelling is slightly larger than the impact of the second adult. For households with four or more adult members no significant effect at a 5 % level is found, reflecting a rather distinct composition of these households, which applies to relatively few observations and might behave quite differently from more standard households. The estimates of the effects of children indicate, that the first child does not affect the mean temperature choice, but the presence of two or more children has an effect at a similar order of magnitude as the second adult household member. Generally, the results are consistent with larger households having stronger preferences for keeping indoor temperatures high in more rooms and for more hours during a day, as they occupy all rooms more. Similarly, the negative – though not significant – point estimate on the employment status is in line with the idea that households reduce thermostat settings when they are at work.

Households with a high-school degree (the German “Abitur”) are found to consume higher mean temperature levels. While the point estimates on households' age indicate a higher preference for high mean temperature levels for older households, these effects are statistically insignificant in the preferred specification.

To analyse the effect of income on households' fuel consumption, I allow it to affect households' preferences for thermal comfort as well as the size of the overprediction of the efficiency state. The estimated coefficients on the indicators for belonging to the low or high income group are statistically indistinguishable from zero.[28] The point estimate even indicates a higher consumption of mean indoor temperatures for poorer households in the preferred specification, which contrasts standard economic intuition as well as previous results of the empirical literature (Cayla et al., 2011). The estimated effect of income on the size of the overprediction of the efficiency state is also not significant. I consider this an indication that the dwelling characteristics included in the adjustment term successfully capture the systematic variation in the measurement error. The point estimate indicates that poorer households live in dwellings for which the overprediction of the efficiency states is larger. Column (4) of table 2.1 reports results of a regression estimating only one joint coefficient on the effects of income on fuel consumption via both channels. It is thus more similar to previous approaches, such as the one by Mertesacker (2020b), that are unable to separate both effects. The estimated coefficients are again statistically insignificant. Furthermore, no indication for a positive effect of lower income on fuel consumption is visible. A limitation of the data used in this study is that it undersamples low income households and thus precisely those that are constrained the most in their consumption of thermal warmth. Furthermore, the data only reports categorical income information and thus erases variation that might be useful to identify effects of income. The results suggest that more precise income data would be required to identify a positive effect of income on mean temperature consumption with the developed empirical framework in a relatively rich country like Germany with an extensive social security system.

It is interesting to consider the effect of the dwelling size (in square metres) in this respect. Again, I use the ability of the empirical framework to identify separate effects on the mean temperature choice and the size of the overprediction of the efficiency state, as dwelling size could have an effect on both. Households living in larger dwellings could have lower valuations of high mean indoor temperatures over the entire dwelling. At the same time, the precision of engineering predictions might be related to the size of the dwelling. The estimation results indicate, that the size of the dwelling only has a statistically significant impact on households' choice. However, the sign of the effect is different than expected. Households living in larger dwellings are actually found to have stronger preferences for high mean indoor temperatures. A potential explanation

[28] Unreported results confirm the robustness of this finding to alternative definitions of the income groups.

is that the size of the dwelling proxies income and wealth effects that are not captured by the income indicators. It is an interesting question for future research to use the empirical framework to disentangle these effects with more detailed data on households' income and wealth.

2.5.2 Predictions of mean temperature choices and elasticities

Table 2.2 illustrates how the empirical model allows to estimate temperature choices, $\tau^*_{i,t}$, based on observed fuel consumption data. It reports the mean values of the predicted optimal and implied temperature increases, $\hat{\tau}^*_{i,t}$ and $\hat{\tau}^d_{i,t}$, as well as statistics of further predicted model outcomes. The columns (1) to (4) represent the same regressions as the respective columns of table 2.1.

Table 2.2: Model Predictions [a]

	(1)	(2)	(3)	(4)
$\varnothing\ \hat{\tau}^*_{i,t}$	8.411***	10.021***	11.572***	11.518***
	(0.085)	(0.150)	(0.152)	(0.152)
$\varnothing\ \hat{\tau}^{in^*}_{i,t}$	14.843***	16.452***	17.996***	17.941***
	(0.085)	(0.150)	(0.152)	(0.152)
$\varnothing\ \hat{\tau}^d_{i,t}$	8.411***	6.913***	6.829***	6.833***
	(0.085)	(0.095)	(0.079)	(0.079)
$\varnothing\ \hat{\lambda}_{i,t}$	1.000	0.690***	0.600***	0.603***
	(0.000)	(0.015)	(0.012)	(0.012)
$\varnothing\ \varepsilon_{\tau^*_{i,t},c_{i,t}}$	-1.205	-0.596***	-0.302***	-0.308***
	(1.890)	(0.063)	(0.023)	(0.023)
$\varepsilon_{\tau^*,c_{i,t}}(\overline{\boldsymbol{x}}_{\boldsymbol{i,t}})$	-0.674***	-0.401***	-0.259***	-0.267***
	(0.018)	(0.021)	(0.018)	(0.018)

[a] Columns (1) to (4) of the table summarize predictions of model quantities based on estimates reported in the respective columns of table 2.1. The diameter symbol (∅) indicates the arithmetic mean over all households, i, and time periods, t, of the quantity that follows it. The predicted quantities are households' optimal temperature increase, $\hat{\tau}^*_{i,t}$, the resulting indoor temperature, $\hat{\tau}^{in^*}_{i,t}$, the temperature increase implied in the observed fuel data given the input demand function of equation (2.5), $\hat{\tau}^d_{i,t}$, the adjustment factor, $\hat{\lambda}_{i,t} = \lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \hat{\boldsymbol{\beta}}^{\boldsymbol{\lambda}})$, and the elasticity with respect to changes in the marginal cost of temperature consumption, $\varepsilon_{\tau^*_{i,t},c_{i,t}}$. In addition, the last row reports the elasticity of a hypothetical household with the mean characteristics in the sample stored in the vector $\overline{\boldsymbol{x}}_{\boldsymbol{i,t}}$. Standard errors are reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

In column (1) no adjustment for measurement errors in the efficiency state, $s_{i,t}$, is included in the regression. As indicated by equation (2.16), the optimal temperature increase, $\tau^*_{i,t}$, is thus directly fitted to the implied temperature increase, $\tau^d_{i,t}$. Consequently, the mean predictions of both quantities are identical. They equal 8.411 degrees Celsius implying an average predicted indoor temperature in the dwelling, $\hat{\tau}^{in^*}_{i,t}$, of 14.843 degrees Celsius. Since the efficiency states are not adjusted, the mean value of the adjustment term, $\hat{\lambda}_{i,t} = \lambda(\boldsymbol{x^{\lambda}_{i,t}}, \boldsymbol{\hat{\beta}^{\lambda}})$, is trivially equal to unity and requires no estimation. In column (2) a constant adjustment of the efficiency state is introduced and estimated to be 0.690. The additional variable implies that the optimal temperature increase is not directly fitted to the implied temperature increase anymore and allows $\hat{\tau}^*_{i,t}$ to deviate from $\hat{\tau}^d_{i,t}$. As a consequence, the average predicted temperature choice in the sample increases to 16.452 degree Celsius. At the same time, the average prediction of implied temperature choices, indicating the temperature choice consistent with the observed fuel demand according to the input demand equation (2.5), decreases to 6.917. This is much closer to the mean implied temperature choice of 6.913 observed in the data, emphasizing the superior ability of the regression to fit the equation to the dependent variable. Allowing the adjustment factor to vary with observable dwelling characteristics, results in a stronger adjustment and increases the average predicted temperature choice further to 17.996 degree Celsius.

The model predictions reported in table 2.2 illustrate how the introduction of the adjustment term allows to identify households' temperature choices. Without its inclusion, the systematic measurement error in the dependent variable results in a downward bias of the estimated temperature choice. Given the relationship between households' temperature choice and their sensitivity to price changes, stated in equation (2.12), it is hence not surprising that the inclusion of the adjustment factor into the regression also has a substantial impact on the estimated elasticities. In column (1) the mean of the predicted elasticities, $\varepsilon_{\tau^*_{i,t}, c_{i,t}}$, is estimated extremely small at -1.205. The downward bias in the predicted temperature choice implies that households are predicted very elastic. The effect is exacerbated by the nonlinearity of the elasticity in the temperature choice. The smaller the temperature level, the more does the sensitivity towards price changes increase if the temperature level is decreased further.[29] The increase in predicted temperature levels in columns (2) and (3) also implies a decrease in the estimated price elasticity (in absolute values). In the preferred specification, the estimated price elasticity is -0.302 and highly statistically significant.

Table 2.2 also reports the predicted elasticity evaluated at mean values $\boldsymbol{\overline{x}^{\tau}_{i,t}}$ and $\boldsymbol{\overline{x}^{\lambda}_{i,t}}$, which makes it less sensitive to outliers. The estimate is therefore less extreme in column (1) and generally below the mean elasticities in the sample.[30] Both estimates become the closer, the more detailed

[29] This sensitivity towards small changes in the temperature level also results in large standard errors making the elasticity estimate insignificant despite its large absolute value.

[30] Also the elasticity is estimated much more precisely and therefore statistically significant in column (1).

the adjustment factor is modelled. In the preferred specification the predicted elasticity at mean values is -0.259 and thus just slightly smaller in absolute value than the mean elasticity in the sample.

The estimated elasticities are consistent with estimates of direct rebound effects obtained in earlier research (Aydin et al., 2017; Sorrell et al., 2009; Greening et al., 2000). Besides its rigorous theoretical foundation, one advantage of the novel approach to the estimation of rebound effects presented in this paper is, that it allows to analyse heterogeneity in households' elasticity in a very natural way. This is considered in detail in the next section.

2.5.3 Economic interpretation and heterogeneity analysis

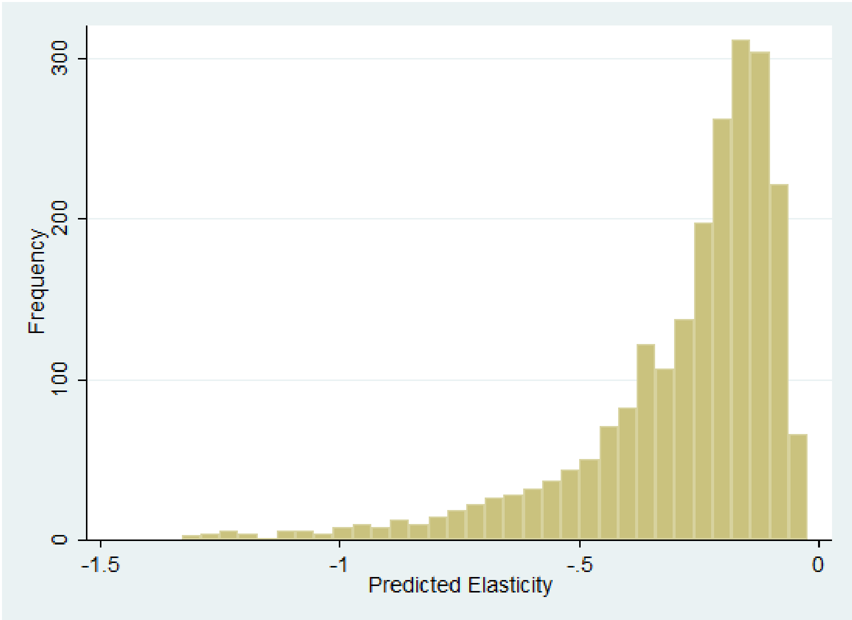
Given the estimated structural model, it is straight forward to predict the temperature choice and consequently the elasticity with respect to changes in the marginal cost of a temperature increase for every household in the sample. The empirical model thus provides a natural framework to move beyond the mere analysis of mean elasticity values and to consider the heterogeneity in households' sensitivity towards price changes in detail.

Figure 2.1 shows the distribution of estimated elasticities in the sample, excluding only observations with elasticities below the first percentile of the distribution.[31] The histogram indicates quite some heterogeneity in the estimated elasticities. As stated in column (3) of table 2.2 the mean elasticity is -0.302. However, the distribution is right skewed with a high probability mass at its right tail. This results in the median, -0.214, to differ substantially from the mean value, indicating a lower sensitivity to price changes. The first and second quartile of the distribution are -0.370 and -0.139, respectively.

Clearly, the distribution indicates that heterogeneity is not neglectable. While most households are less elastic than the mean estimate would suggest, some are substantially more sensitive to changes in the marginal cost of heating. A focus on the mean values thus misses important information on how individual households are likely to respond, e.g., to changes in the level of thermal insulation of their dwelling. While some of the previous literature has estimated separate elasticities, for instance for owners and tenants or low and high income households (see Aydin et al., 2017), are these approaches limited to the inspection of heterogeneity with respect to some few variables that are typically interacted with the elasticity estimate in a regression model. In contrast, in the present structural model the heterogeneity of estimated elasticities is a natural outcome of households' preferences for thermal comfort. Table 2.3 allows to track the mechanisms

[31] These observations are excluded to narrow the range of the presented histogram. In total, 23 observations are excluded. The minimum and maximum of the elasticities excluded are -5.072 and -1.347, the mean and median -2.004 and -1.789 respectively.

Figure 2.1: Distribution of predicted elasticities



resulting in variation of predicted elasticities across households in detail and provides economic interpretations of the estimated coefficients reported in table 2.1.

Column (1) states how households' utility gain after a marginal temperature increase – expressed in euro – varies with the variables included in the vector $x_{i,t}^{T}$. For example, it indicates that a household consisting of two instead of one adult person is willing to pay 38.721 euro more to increase the mean indoor temperature in the dwelling (over the heating period) by one degree over his current choice. In general, the relative magnitudes and statistical significance of the estimated utility gains of a marginal temperature increase are similar to the coefficients analysed in section 2.5.1. However, column (1) of table 2.3 provides a way to interpret the economic significance of the heterogeneity in households' preferences. It is the structural framework developed in this paper and specifically the use of a quasilinear function to model households' utility from temperature consumption that allows to give the estimates a concrete economic meaning.

Column (2) of table 2.3 analysis, how the heterogeneity in households' preferences results in heterogeneity in temperature choices. A household with higher valuation of thermal comfort is expected, ceteris paribus, to consume a higher mean indoor temperature over the heating period. For instance, the estimates indicate that the two person household consumes an indoor temperature 0.53 degree over the one person household. Ultimately, it is this heterogeneity in the

Table 2.3: Heterogeneity of households in their preferences for and choices of temperature increases

	$\frac{\partial^2 u(\tau_{i,t},\hat{\boldsymbol{\beta}}^{\tau}\overline{\boldsymbol{x}}^{\tau}_{i,t})}{\partial \tau_{i,t}\partial x^{\tau}_{i,t,k}}\Big\vert_{\tau_{i,t}=\hat{\tau}^*_{i,t}}$	$\frac{\partial \tau^*(\hat{\boldsymbol{\beta}}^{\tau}\overline{\boldsymbol{x}}^{\tau}_{i,t})}{\partial x^{\tau}_{i,t,k}}$	$\frac{\partial \varepsilon_{\tau^*,c_{i,t}}(\hat{\boldsymbol{\beta}}^{\tau}\overline{\boldsymbol{x}}^{\tau}_{i,t})}{\partial x^{\tau}_{i,t,k}}$
age: $30-39$	16.215	0.234	0.027
	(39.531)	(0.607)	(0.071)
age: $40-49$	1.203	0.019	0.002
	(40.021)	(0.62)	(0.073)
age: $50-59$	50.338	0.636	0.071
	(43.119)	(0.642)	(0.075)
age: ≥ 60	61.801	0.750	0.083
	(47.075)	(0.677)	(0.079)
#adults$=2$	38.721**	0.53**	0.06**
	(15.627)	(0.233)	(0.027)
#adults$=3$	85.489**	0.99***	0.108***
	(36.469)	(0.35)	(0.038)
#adults≥ 4	49.465*	0.65**	0.073**
	(26.22)	(0.323)	(0.036)
#children$=1$	-4.734	-0.058	-0.006
	(17.076)	(0.213)	(0.024)
#children≥ 2	38.077*	0.401*	0.043*
	(22.077)	(0.216)	(0.023)
Is employed	-28.698	-0.327	-0.035
	(27.09)	(0.294)	(0.032)
Has Abitur	42.018***	0.502***	0.055***
	(16.21)	(0.185)	(0.02)
Is Owner	-0.424	-0.005	-0.001
	(17.006)	(0.199)	(0.022)
Income: $< 1,500$ €	29.927	0.434	0.046
	(27.634)	(0.275)	(0.028)
Income: $\geq 3,500$ €	4.110	-0.096	-0.011
	(16.478)	(0.243)	(0.027)
Dwelling size: $<$ 1st tercile	-14.164	-0.198	-0.023
	(18.586)	(0.274)	(0.032)
Dwelling size: $\geq$ 2nd tercile	68.371***	0.702***	0.074***
	(17.79)	(0.194)	(0.021)

[a] The three main columns of the table report discrete change effects of predicted model quantities when moving from a base category to a respective household characteristic, $x^{\tau}_{i,t,k}$. All characteristics that are not subject to a discrete change are evaluated at their respective sample means stored in the vectors $\overline{\boldsymbol{x}}^{\tau}_{i,t}$ and $\overline{\boldsymbol{x}}^{\lambda}_{i,t}$. The predictions are based on estimates reported in column (3) of table 2.1.
There are three different quantities that are predicted. First, the change in marginal utility with changes in the respective household characteristic, $x^{\tau}_{i,t,k}$. This is evaluated at the predicted temperature choice of the considered household. Second, the change in the optimal temperature choice and third the change in the price elasticity with respect to changes in the respective household characteristic, $x^{\tau}_{i,t,k}$. Note that even though the column titles indicate marginal changes for notational convenience, all effects reported are actually discrete change effects, as only discrete variables are considered. Standard errors are reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

level of indoor temperature chose by households, that drives heterogeneity in their elasticity to price changes. As derived in section 2.2.3, the theoretical model implies that households consuming a higher indoor temperature are less elastic towards price changes. Consequently, households with higher preferences for thermal warmth, resulting in the choice of a higher indoor temperature, are less elastic than otherwise identical households that value high indoor temperatures less. Comparing again the one- with the two-person household, the latter is estimated to have an elasticity 0.06 percentage points below the former (in absolute values).

2.6 Conclusion

In this paper, I used a household production framework to develop a structural econometric model for the empirical analysis of households' demand for thermal heat. In the model, households receive a quadratic disutility from the distance of their ambient temperature to an ideal temperature level, introducing satiation effects in temperature consumption. The amount of fuel required for the production of indoor temperature is determined by a linear input demand function that Mertesacker (2020b) derived from a fuel requirement calculation procedure developed by engineers. I solve the theoretical model for the optimal temperature choice of the household and the associated fuel consumption. Using consumption data of $2,256$ households from "The German Residential Energy Consumption Survey" (RWI and forsa, 2016), I estimate households' preferences for thermal comfort using nonlinear least squares. The results indicate that larger and more educated households have stronger preferences for high mean temperature levels. In contrast, no significant effect of income on temperature consumption can be found.

The presented research makes two major contributions to the previous literature. First, it shows how a household production framework can be used to combine economic and engineering knowledge in a theoretically consistent way to estimated households' demand for thermal heat. To my knowledge this is the first paper to fully specify, solve and estimate a household production model of heat demand. The theoretical foundation implies that the estimated parameters have a clear interpretation in terms of the economic model. The structure imposed by the specification of the utility as well as the input demand function furthermore allows to identify mean temperature levels chosen by households, even though the respective information is unobserved in the data.

Second, the structural model provides a novel approach to estimate rebound effects after efficiency increases. Instead of estimating the price elasticity of temperature consumption in a reduced form framework, it can explicitly be derived theoretically from the Marshallian demand of temperature consumption. Given the estimates of households' preferences for thermal comfort, it is straight forward to predict the individual elasticity for every household in the sample. The individual elasticities are thus obtained as a natural outcome of households' preferences for thermal comfort. Households with stronger preferences for thermal comfort consume temperature levels closer to the ideal temperature level, which makes them less sensitive towards price changes. The mean price elasticity is estimated to be -0.302 in the preferred specification and thus confirms rebound effects found in previous research. However, the individual elasticities show substantial heterogeneity. A majority of households is less elastic than indicated by the mean value and some are (substantially) more elastic, resulting in a median elasticity of only -0.214.

The developed framework indicates interesting research opportunities for the future. First, the results emphasize that the heterogeneity in households' rebound behavior after efficiency increases should be considered in more detail. Only a precise understanding how individual households are likely to react to increases in energy efficiency, allows an efficient design of policies to reduce carbon emissions in the heating sector. Second, may the model framework well be used to improve existing energy economy models (Swan and Ugursal (2009), Kavgic et al. (2010), Mundaca et al. (2010)), which forecast nationwide energy consumption. The model provides the structural parameters of households' decision process that are required to conduct counterfactual policy scenarios that these studies aim to address. The ability to link fuel consumption to a utility value households' receive, might furthermore be of interest for researchers who want to model a period utility in a dynamic model of modernisation investments. Finally, it is certainly interesting to further develop the model and apply it to other datasets. In particular, households' utility function from thermal comfort might be adjusted to introduce income effects explicitly into the analysis. Richer datasets that include metered data on indoor temperatures, humidity levels and potentially other factors determining households' level of thermal comfort might furthermore allow more detailed analyses of the different dimensions that yield thermal comfort in the future.

3 Households' dynamic investment in domestic energy efficiency

3.1 Introduction

The European Commission takes strong actions to combat climate change. It has set Europe-wide goals to reduce greenhouse gas emission by 40 % compared to 1990 levels and to realize energy efficiency improvement by 32.5 % until 2030 (European Commission, 2020a).[1] Numerous regulatory frameworks and subsidy programs are in place to trigger household investment into energy saving technology as means to meet climate targets. In Germany alone they include mandatory prescriptions for new products, subsidy programs and tax credits for the adoption of energy efficient technology, and information campaigns of different forms and scale.

Many of these programs focus on setting incentives for the retrofitting of domestic housing, for example the modernization of the thermal shell and heating system of dwellings. One reason for the relevance of retrofitting is that residential space heating is a major contributor to the overall energy consumption in the economy. In the European Union households have accounted for roughly 26 % of the total final energy consumption in 2018 of which almost 64 % was used for heating (Eurostat, 2020).

Engineering calculations show, that the energy saving potentials from retrofitting are substantial. Energy demand for residential heating can be reduced by half with the appropriate retrofitting measures (Becchio et al., 2012). In Germany roughly 60 % of gas-fired and 70 % of oil-fired installed heating systems have been more than 20 years old in 2019 (Bundesverband Des Schornsteinfegerhandwerks, 2019). Additionally, only 50.4 % of all dwellings in Germany had received some thermal insulation of the outer walls in 2016 (Cischinsky and Diefenbach, 2018, p. 44) , which further indicates the large potential that could be leveraged, if the necessary investments were conducted.

Households are found to be fairly reluctant to retrofit their homes. The empirical observation that even (seemingly) profitable investments into energy efficiency remain undone, has extensively

[1] In September 2020 it proposed to further raise the reduction target to a 55 % reduction of greenhouse gas emissions over 1990 levels until 2030 (European Commission, 2020a).

been discussed in the literature and is commonly referred to as the "Energy Paradox" (Hirst and Brown, 1990; Jaffe and Stavins, 1994; Gerarden et al., 2017). Reasons for the low investment rate are manifold. They include market failures, financial constraints, behavioral biases in consumers' decision process, and misconceptions about the level and heterogeneity of actual cost and potential savings that households face (see Gerarden et al., 2017).

To assess how effective government policies are to promote investments in energy saving technology, a precise understanding of how they affect households' decision process is required. We contribute to the understanding of households' retrofit decisions by developing a dynamic structural model of their decision to modernise the thermal shell, windows or heating system of the dwelling they inhabit.

In the model, households choose in every period a mean indoor temperature in the dwelling to maximize period utility. The amount of fuel required to produce the desired temperature level depends on the thermal efficiency level of the dwelling. Households invest in energy saving technology to improve the energy efficiency standard and thus the amount of fuel required to heat the dwelling to the desired temperature level. This investment choice has dynamic implications: By improving the domestic energy efficiency, households benefit from savings in energy consumption or from higher thermal consumption levels in the future. Households only invest if the expected resulting gain in lifetime utility, discounted to the current period, exceeds the one-time fixed costs that arise at the time of investment. Using fuel data from "The German Residential Energy Consumption Survey" (RWI and forsa, 2016) we first estimate the parameters of households' period utility function using a framework developed by Mertesacker (2020a). In a second step, the investment costs are then estimated given the increase in lifetime utility that we calculate based on the estimates of households' preferences for thermal comfort obtained in step 1. Using maximum likelihood the investment cost are chosen such that they rationalize investments observed in the data, given the developed economic model.

The empirical analysis of households' retrofitting decisions using the proposed structural dynamic investment model has a couple of advantages over standard regression analyses, such as logit or probit, that rely on static utility models to estimate the relationship between household and dwelling characteristics and the propensity to invest. First, the lifetime utility gain from investing is derived from a sound economic model of the period utility that households receive from the consumption of thermal comfort. Different to mere engineering estimates of the benefits of modernising the dwelling, our model explicitly accounts for the possibility that households may benefit from increased efficiency levels by reducing their expenditures for thermal heating as well as by increasing the temperature level in their dwelling. While the possibility of households to rebound after an efficiency increase is undesirable from a policy perspective that aims to reduce fuel consumption to mitigate carbon emissions, it increases the potential benefits of investments to households. Our model allows to explicitly consider both effects. Conveniently, the chosen functional form of the period utility function implies that lifetime utility gains can be expressed in monetary values even though they also include non-monetary benefits to households. In our sample, the mean expected increase in lifetime utility that results from investing is $4,368$ euro.

Second, the framework allows to estimate and analyse the benefits and costs of investing separately. Standard regressions of observed investments on household and dwelling characteristics only estimate the net impact of both factors behind the investment decision. In contrast, our model assigns a clear structural interpretation to every estimated parameter. Since investment costs are estimated in relationship to the lifetime benefits of investing, they are also expressed in monetary equivalents, providing an intuitive quantification of all impediments that might hinder more investment to occur. The cost estimates thus provide a convenient alternative to standard calculations of high discount rates to characterize households' low investment activity despite large potential savings that might be realized.

Third, given the estimates of all parameters of the dynamic structural model, we can predict and thus quantify model quantities such as the temperature choice and the associated fuel consumption, the period utility, the expected gain in lifetime utility associated to an investment and the resulting investment probability for every household in the sample. In contrast to estimates of average impacts of covariates on the investment probability, provided by standard regression approaches, this allows a much more detailed study of households' incentives to invest and how these might change if conditions in the economic environment are altered.

Finally, the model allows to explicitly analyse the consequences of different policy scenarios – that are designed to facilitate investments or reduce energy consumption – on households' decisions. Simulating a public policy that aims at reducing households' costs of investing via a direct subsidy, we find that the investment rate is increased, but that this does only little to decrease average fuel consumption. Similar effects are found for a policy that increases the effectiveness of modernisations, e.g., by funding research and development. In contrast, an increase in energy prices, for instance via a tax, creates high energy saving incentives for households. It leads to an increase in the investment rate by 22.1 % and a reduction of households' mean temperature choice by 4.8 %.

In the next section we introduce the theoretical model of households' energy demand and dynamic investment decision. Section 3.3 describes the data used for the estimation. Section 3.4 discusses the estimation procedure and empirical results. Section 3.5 provides concluding remarks.

3.2 Theoretical model

This section develops a theoretical model of households' dynamic decision to improve their domestic energy efficiency through retrofitting. The model is structured in three steps. First, households decide whether or not to retrofit their home. This can be the insulation of walls, installation of double glazing (two or three glass window panes) or the adoption of a more efficient heating system. The second step describes the impact of these retrofitting measures on the energy required for heating. Dwellings with lower energy requirements are considered more energy efficient. In the third step, the changes in domestic energy efficiency affect households' optimal consumption of thermal warmth and can lead to improvements in households' overall utility level.

In the dynamic model households invest in improvements of their domestic energy efficiency level to maximizes the discounted sum of expected future utility, while taking into account the impact of retrofitting measures on domestic energy requirement and the resulting improvements in their utility level. The next subsections develop the theoretical model for each stage. We first analyse the link between the energy efficiency level, households' consumption of warmth and the utility they receive, before we move to the dynamic retrofit decision.

3.2.1 Energy efficiency and thermal heat consumption

To model households' consumption of thermal comfort, we make use of a structural empirical model developed by Mertesacker (2020a). The model considers households to consume thermal comfort by choosing a mean indoor temperature in their dwelling. If the temperature level reaches a satiation point of 21 degrees Celsius, they enjoy blissful thermal comfort. Deviations from this ideal temperature level create discomfort. To avoid this disutility, households spend parts of their income on the consumption of heating energy. The remaining part of income is spent on the consumption of all other goods. Household i's period utility function is given by

$$u(\tau_{i,t}) = \beta^G \left(Y_{i,t} - p^F_{i,t} \cdot F(\tau_{i,t}, s_{i,t}, m_{i,t})\right) - \boldsymbol{\beta^\tau x^\tau_{i,t}} \cdot (\bar{\tau}_{i,t} - \tau_{i,t})^2. \tag{3.1}$$

Household i's income in period t is denoted by $Y_{i,t}$. The function $F(\tau_{i,t}, s_{i,t}, m_{i,t})$ determines the amount of heating energy consumed by households, measured in kilowatt hours (kWh), and $p^F_{i,t}$ denotes the energy price per kilowatt hour. The amount of fuel consumed, $F_{i,t}$ depends on households' decision by how many degrees to increase the indoor temperature, $\tau_{i,t}$, on a measure for the efficiency of the dwelling, $s_{i,t}$, and on the size of the living area, $m_{i,t}$, measured in squared metres. The variable $s_{i,t}$ measures the amount of energy in kilowatt hours per square metre required to increase the mean indoor temperature of the dwelling by one degree Celsius over the entire heating period. It is larger for very inefficient dwellings and the smaller the less fuel is required to heat the dwelling to a desired indoor temperature.[2] Thus, the first term of the utility function is the utility level households receive from their available income net heating expenditures.

Depending on the outdoor temperature, the parameter $\bar{\tau}_{i,t}$ denotes the maximal temperature increase households can choose by heating the entire dwelling up to 21 degrees Celsius over the entire heating period. It is the difference between the ideal and the outdoor temperature level. The actual temperature increase chosen by households is $\tau_{i,t} \in [0, \bar{\tau}_{i,t}]$. Thus, the second term of the utility function describes their disutility resulting from the deviation between the ideal temperature increase and the actual choice of $\tau_{i,t}$. Additionally, we allow for the thermal disutility to vary by household characteristics stored in the column vector $\boldsymbol{x^{\tau}_{i,t}}$. The parameters stored in the row vector $\boldsymbol{\beta^{\tau}}$ indicate how differences in variables in $\boldsymbol{x^{\tau}_{i,t}}$ affect households' valuation of thermal comfort and will be estimated in the empirical analysis. The heterogeneity in the marginal utility of indoor temperature leads to heterogeneity in temperature choices and thus energy consumption.

Overall, households can reduce their thermal discomfort by choosing a temperature increase, $\tau_{i,t}$, that is close to $\bar{\tau}$. This however involves higher spending on fuel consumption.

Following Mertesacker (2020b) and Mertesacker (2020a), we assume that fuel consumption can be related to households' temperature choice by a linear function:

$$F_{i,t} = F(\tau_{i,t}, s_{i,t}, m_{i,t}) = s_{i,t} m_{i,t} \tau_{i,t}. \tag{3.2}$$

Solving households' period utility maximization problem, the optimal temperature choice can be derived:[3]

$$\tau^*_{i,t}(s_{i,t}) \quad = \quad \bar{\tau}_{i,t} - \frac{\beta^G}{2\boldsymbol{\beta^{\tau}} \boldsymbol{x^{\tau}_{i,t}}} \cdot p^F_{i,t} s_{i,t} m_{i,t}. \tag{3.3}$$

Households' optimal temperature increase is a function of the utility function parameters (β^G, $\boldsymbol{\beta^{\tau}}$), dwelling characteristics ($s_{i,t}, m_{i,t}$), the energy price, $p^F_{i,t}$, and household characteristics $\boldsymbol{x^{\tau}_{i,t}}$. The ratio ${}^{\beta^G}/_{\boldsymbol{\beta^{\tau}} \boldsymbol{x^{\tau}_{i,t}}}$ indicates households' valuation for thermal comfort relative to other goods. A high valuation of thermal comfort is reflected by small values of ${}^{\beta^G}/_{\boldsymbol{\beta^{\tau}} \boldsymbol{x^{\tau}_{i,t}}}$, implying the consumption of high temperature levels according to equation (3.3). Furthermore, as $s_{i,t}$ decreases, the amount of fuel required to increase the room temperature is reduced. This effectively lowers the marginal cost for heating and results in a higher temperature choice. This rebound effect – documented

[2]See Mertesacker (2020b) for a very detailed discussion how the variable $s_{i,t}$ can be obtained from an engineering model and be interpreted.

[3]See Mertesacker (2020a) for a very detailed discussion of the entire theoretical model.

in previous studies (see, e.g., Aydin et al., 2017) – reflects utility maximising behavior and is an important part of the benefits associated to retrofit investments. At the same time it reduces the amount of fuel that is saved through a modernisation. Our model explicitly incorporates both effects.

Inserting the optimal temperature choice, $\tau^*_{i,t}$, into equation (3.1), it is straight forward to obtain households' utility as a function of the efficiency level, $s_{i,t}$, and other state variables:

$$u(\tau^*_{i,t}(s_{i,t})) \quad = \quad \beta^G Y_{i,t} - \beta^G p^F_{i,t} s_{i,t} m_{i,t} \bar{\tau}_i + \frac{(\beta^G p^F_{i,t} s_{i,t} m_{i,t})^2}{4\boldsymbol{\beta^\tau x^\tau_{i,t}}}. \tag{3.4}$$

Equation (3.4) allows to directly calculate the period utility households receive given different efficiency levels of the dwelling. It provides the basis to explore the benefits households may receive from retrofitting their dwellings. In the dynamic estimation we simplify equation (3.4) by dropping the term $\beta^G Y_{i,t}$, which neither affects households' optimal temperature choice nor their benefits from investing. The resulting period utility $\tilde{u}(\tau^*_{i,t}(s_{i,t}))$ strictly smaller than zero.

To calculate the utility gains associated to actual modernisations, their impact on dwellings' efficiency levels has to be modelled. We model the domestic energy efficiency level to follow a first order Markov process, that can be shifted by retrofitting investments. Using discrete indicators of retrofitting investments, $r_{i,t}$, the evolution process of energy efficiency can be characterized as follows:

$$\begin{aligned} s_{i,t+1} &= g(s_{i,t}, r_{i,t}) + \varepsilon_{i,t} \\ &= \lambda_0 + \lambda_1 s_{i,t} + \lambda_2 s^2_{i,t} + \lambda_3 s^3_{i,t} + \lambda_4 r_{i,t} + \varepsilon_{i,t}. \end{aligned} \tag{3.5}$$

The realized energy efficiency level in period $t+1$ depends on its lagged values, the modernisation activities in the last period and an error term $\varepsilon_{i,t}$. Generally, the energy efficiency level is a combination of several dwelling characteristics, for instance the materials used for construction. While the building can deteriorate over time, the main characteristics do not change vastly and previous energy efficiency levels will generally be carried over to future periods. The parameters $\lambda_1, \lambda_2, \lambda_3$ jointly determine the persistence of the energy efficiency level. If households invest into retrofitting measures, energy requirements for heating can be reduced and the level of $s_{i,t}$ will be shifted by the amount of λ_4. In the empirical model we also distinguish the impact of different retrofitting types and their interaction with the energy efficiency level to capture the heterogeneity in the overall impact. The random term $\varepsilon_{i,t}$ captures the uncertainty in the efficiency process. It allows for households with the same efficiency levels and investment decisions to have different energy efficiency realization in the next period. This can be due to the ability of the construction companies that implement the retrofitting, the products they use and specifics of the dwelling that affect the exploitation of the energy saving potentials. We assume $\varepsilon_{i,t}$ to be i.i.d normally distributed with mean zero and variance σ_ε^2, according to a distribution function $\Omega(0, \sigma_\varepsilon^2)$.

3.2.2 Households' dynamic investment choice

This section develops households' dynamic investment decision in domestic retrofitting. The majority of the empirical energy literature aims at measuring the correlation between households' investment decisions and their socioeconomic characteristics. We take another approach and model households' optimal investment decision structurally. In our model the investment decision results from a comparison of the long-run benefit from investing and the one-time fixed cost associated to it. Households' long-run benefit from investing is the potential gain in period utility through improved domestic energy efficiency in all future periods. These gains might be achieved through lower total cost for the consumption of indoor temperature or by realising a higher level of thermal comfort in the dwelling. The cost of retrofitting can be interpreted as the sum of all costs households encounter when conducting the energy efficiency improvement. Most obviously, this contains the monetary spending for the installation of retrofitting measures. It however, also includes non-monetary impediments to the investment such as behavioral cost that may arise from the necessity of gathering information about the investment alternatives, the existence of a construction site within the dwelling or other inconveniences related to the installation. In our model, we capture the effect of all costs from retrofitting on households' choice by a variable, $C_{i,t}$, that directly measures the total utility households sacrifice in exchange for the modernisation of the dwelling. It indicates the total loss in (lifetime) utility associated to an investment discounted to period t when the retrofit decision is made. The investment cost vary across households. Different households work with different construction companies and enjoy different prices for the retrofitting. Households can also have idiosyncratic differences in their behavioral impediments to invest. To allow for the cost heterogeneity, we model the investment cost $C_{i,t}$ as a random draw from an exponential distribution with mean γ, $C_{i,t} \sim \Phi_C^\gamma$.

In each period, t, households observe their energy efficiency level and choose the temperature increase that maximises period utility according to equation (3.3). Then, they learn their investment cost and make an investment decision $r_{i,t} \in \{0, 1\}$. Note, that even though households observe their investment cost in the current period, they remain uncertain about the cost they will face in future periods. In period $t + 1$, the new energy efficiency level is realized based on the evolution process stated in equation (3.5).

Households' value function before observing the investment cost is:[4]

$$V(s_{i,t}) = \quad u(\tau^*_{i,t}(s_{i,t})) + \max_{r_{i,t}} \int_{C_{i,t}} \Big\{ \delta EV(s_{i,t+1}|s_{i,t}, r_{i,t} = 1) - C_{i,t}; \\ \delta EV(s_{i,t+1}|s_{i,t}, r_{i,t} = 0) \Big\} d\Phi^{\gamma}(C), \tag{3.6}$$

where δ denotes the discount factor. The term $EV(s_{it+1}|s_{i,t}, r_{i,t})$ denotes households' expected value of future utility given the current energy efficiency level, $s_{i,t}$, and the retrofitting decision, $r_{i,t}$. The expectation about future utility is taken with respect to the realization of $s_{i,t+1}$. That is,

$$EV(s_{i,t+1}|s_{i,t}, r_{i,t}) = \int_{s_{i,t+1}} V(s_{it+1}) d\Omega(s_{i,t+1}|s_{i,t}, r_{i,t}).$$

Households' decision to modernize the dwelling, implies that the evolution process follows a different, more favourable, path than if they decide not to invest. Consequently, the expected stream of future period utilities given that an investment has occurred, $EV(s_{i,t+1}|s_{i,t}, r_{i,t} = 1)$, exceeds the respective expectation without an investment, $EV(s_{i,t+1}|s_{i,t}, r_{i,t} = 0)$. The difference between the two future value streams determines the expected long-run gain of a retrofit investment.

$$\Delta EV(s_{i,t}) = EV(s_{i,t+1}|s_{i,t}, r_{i,t} = 1) - EV(s_{it+1}|s_{i,t}, r_{i,t} = 0).$$

Retrofitting occurs if and only if the expected discounted lifetime utility gain of investing, $\Delta EV(s_{i,t})$, is larger than the investment cost, $C_{i,t}$,

$$r(s_{i,t}) = \begin{cases} 1 & \text{if} \quad \delta \Delta EV(s_{i,t}) > C_{i,t} \\ 0 & \text{else} \end{cases} \tag{3.7}$$

The expected gain from investing in energy saving technology depends on the current efficiency level of the dwelling and further socioeconomic characteristics such as households' age and size. Differences in $s_{i,t}$ and the variables in the vector $\boldsymbol{x^{\tau}_{i,t}}$ across households result in variation in the expected gain from investing. Together with the variation in the cost that households draw in every period, this allows for heterogeneity in households' choices.

Overall, our model endogenizes the retrofitting decision of households and links it to the evolution of energy efficiency and their choice of thermal comfort. The key structural components

[4]Besides the energy efficiency level, households' value function also depends on other dwelling and household characteristics, which we omit here for notational convenience.

of the model which we estimate from the data are: (i) parameters of the utility function in equation (3.4), (ii) parameters of the energy efficiency evolution process stated in equation (3.5) and (iii) the parameter γ of the investment cost distribution. The model can be estimated using data on households' investment decisions, $r_{i,t}$, fuel consumption, $F_{i,t}$, dwelling characteristics $m_{i,t}$ and $s_{i,t}$ and demographic characteristics $\boldsymbol{x}_{i,t}^{\tau}$. The next sections describe the data, estimation procedure and discuss the results.

3.3 Data

For the empirical analysis a dataset created by Mertesacker (2020b) is used. The dataset contains information on fuel consumption, household characteristics, dwelling characteristics and investment behavior, fuel prices and the efficiency states of the dwelling and thus all information required for the estimation of the structural model.

The main source behind the dataset is "The German Residential Energy Consumption Survey" (RWI and forsa, 2016).[5] The survey is based on a random sample of $6,715$ German households, that have been interviewed in 2010. The dataset includes information about household and dwelling characteristics as well as the energy consumption between 2006 and 2008.

Households' investment activities are in the centre of our analysis. The survey provides data on the investments that occurred between 2002 and 2008. Households separately report if and when investments occurred into thermal insulation, new windows and new heating systems during this time period, respectively. In our empirical model we define the investment variable to equal 1 in years households have undertaken any of these modernisation measures and 0 otherwise. The investment rate according to this definition amounts to approximately 6.154 % in our sample.

We also use fuel consumption data in the estimation of the parameters of the period utility function. The survey provides data on households' fuel consumption between 2006 and 2008. Furthermore, it contains data on the number of household members and children living in the households as well as the income, age, education and employment status, which we use to analyse heterogeneity in households' preferences for thermal comfort.

To obtain price data and information on efficiency states Mertesacker (2020b) combines the main data from the "The German Residential Energy Consumption Survey" with information about average fuel prices households had to pay between 2006 and 2008 and with information on dwellings' efficiency state, $s_{i,t}$. While the price data is obtained from the German ministry of economics (Bundesministerium für Wirtschaft und Energie, 2018), the efficiency states, indicating

[5] The subsequent description of the data closely follows Mertesacker (2020b). He also provides additional detailed descriptions of the data cleaning and the generation of efficiency states, including a discussion of potential selection concerns.

the amount of fuel (in $^{kWh}/_{m^2}$) required to increase the mean indoor temperature by one degree Celsius have to be generated.

For this purpose, Mertesacker (2020b) uses an engineering calculation procedure developed by Loga et al. (2005). Their program has been developed with the intention to facilitate the creation of energy performance certificates for home owners and to provide guidance on potential savings that can be realized through modernisations. The program requires only few dwelling characteristics, that are easily observable for home owners, as inputs. Mertesacker (2020b) shows that this program can also be used to predict the efficiency states and fuel requirements for every household observed in the main dataset. The average constructed efficiency level amounts to roughly 24.97 $^{kWh}/_{m^2}$ per year. Overall, the efficiency measure varies in building characteristics, it ranges between 13.39 $^{kWh}/_{m^2}$ and 43.93 $^{kWh}/_{m^2}$ at the 5th and 95th percentiles of the efficiency distribution, with less modern buildings exhibiting higher energy requirements.[6]

In our empirical analysis we focus on households using natural gas, oil or long-distance heat for primary heating. Furthermore, tenants as well as dwellings with more than two apartments are excluded from the dataset.[7] This ensures that the investment decisions are made by the households studied in the empirical analysis and that the type of investments that are observed are of a broadly similar kind of magnitude.

To allow for households' period utility to vary with demographic characteristics, we construct a couple of discrete variables to include in the empirical analysis. These include three categorical variables indicating the number of adults and children living in the household as well as the age cohort of the survey respondent.[8] Finally, we construct dummy variables indicating whether the survey respondent has a high-school degree (the German "Abitur"), is employed, belongs to a high or low income group and lives in a relatively small or large dwelling, respectively.[9]

[6]See Mertesacker (2020b) for a very detailed discussion how the calculation procedure by Loga et al. (2005) can be applied to micro datasets to predict fuel requirements as well as an analysis of the predicted efficiency states.

[7]To ensure reliability of the fuel consumption data and the generated efficiency states outlier corrections have been conducted. Furthermore, households' for which some fuel consumption data is unobserved, some inputs required to generate the efficiency states are missing or whose children have already moved out are excluded. See Mertesacker (2020b) for further details.

[8]The number of persons living in the household is top-coded at four adults and two children, respectively

[9]Households are assigned to a low income group, if their income is below $1,500$ euro per month and to a high income group if it exceeds $3,500$ euro. The reduction of the number of income groups is ensures that a sufficiently large number of observations in each income bin and reduces the number of variables included in the final regression. The main results are robust towards alternative definitions of income bins.

3.4 Estimation and empirical results

Our model can be estimated using data on households' investment decisions, $r_{i,t}$, households' energy consumption $F_{i,t}$, dwelling characteristics, $m_{i,t}$ and $s_{i,t}$, and household characteristics $\boldsymbol{x^{\tau}_{i,t}}$. The estimation of the model involves three steps. In the first step, we estimate the parameters of the period utility function stated in equation (3.4). The second step estimates the evolution process of the efficiency state, $s_{i,t}$, defined in equation (3.5). We use this evolution process to model agents' expectations about future realisations of the state variable in dependence of their investment behavior. In the third step, we estimate the parameter of the investment cost distribution using maximum likelihood. The key structural components of the model which we estimate from the data are: (i) the vector of utility function parameters $\boldsymbol{\beta^{\tau}}$, (ii) parameters of the energy efficiency evolution process $(\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and the variance the error term σ^2_{ε}, and (iii) the parameter γ of the investment cost distribution.

3.4.1 Estimation of the period utility function

Following Mertesacker (2020a), we estimate the parameters of the utility function by relating the optimal temperature choice derived from the theoretical model, $\tau^*_{i,t}$, to observed fuel demand, $F_{i,t}$, using the fuel consumption function of equation (3.2). This yields a regression equation that is nonlinear in the parameters to be estimated:

$$F_{i,t} = s_{i,t} \cdot m_{i,t} \cdot \left(\overline{\tau}_{i,t} - \frac{\beta^G}{2\boldsymbol{\beta^{\tau}} \boldsymbol{x^{\tau}_{i,t}}} \cdot p^F_{i,t} s_{i,t} m_{i,t} + \varepsilon^{\tau}_{i,t} \right). \tag{3.8}$$

The error $\varepsilon^{\tau}_{i,t}$ accounts for variations in $F_{i,t}$ that cannot be explained by observed preference shifters in the vector $\boldsymbol{x^{\tau}_{i,t}}$ or by changes in $s_{i,t}$, $m_{i,t}$ or $p^F_{i,t}$.

Equation (3.8) clarifies that preferences for general goods and for thermal comfort cannot separately be identified. A larger fuel consumption can equally be explained by a low marginal utility of income – indicated by a small parameter β^G – or by a strong preference for thermal comfort, indicated by a larger value of the product $\boldsymbol{\beta^{\tau}} \boldsymbol{x^{\tau}_{i,t}}$. We therefore normalize the parameter β^G to unity. This implies that the parameters in $\boldsymbol{\beta^{\tau}}$ as well as households' period and long-run utility can be interpreted in monetary terms.

Previous work by Mertesacker (2020b) and Mertesacker (2020a) has shown that efficiency states, $s_{i,t}$, generated from an engineering model, are likely to suffer from measurement errors. Concretely, they mostly overestimate the amount of fuel a dwelling requires. The size of the overprediction systematically correlates with dwelling characteristics. To avoid that this introduces biases in our estimates on household characteristics from the vector $\boldsymbol{x^{\tau}_{i,t}}$, we follow Mertesacker

(2020a) and model an adjustment term, $\lambda(\boldsymbol{x^{\lambda}_{i,t}}, \boldsymbol{\beta^{\lambda}}, \varepsilon^{\lambda}_{i,t})$, that captures systematic overpredictions in the generated efficiency state, $s^{e}_{i,t}$, over the unobserved true efficiency level, $s^{\circ}_{i,t}$:

$$s^{\circ}_{i,t} = \lambda(\boldsymbol{x^{\lambda}_{i,t}}, \boldsymbol{\beta^{\lambda}}, \varepsilon^{\lambda}_{i,t}) \cdot s^{e}_{i,t}. \tag{3.9}$$

$$= (\beta^{\lambda}_{0} + \boldsymbol{\beta^{\lambda}} \boldsymbol{x^{\lambda}_{i,t}} + \varepsilon^{\lambda}_{i,t}) \cdot s^{e}_{i,t}. \tag{3.10}$$

The adjustment term is allowed to vary with observable dwelling characteristics stored in the vector $\boldsymbol{x^{\lambda}_{i,t}}$. The linear specification for the adjustment term of equation (3.10) can be easily included in regression equation (3.8):

$$F^{d}_{i,t} = s_{i,t} m_{i,t} \cdot \left(\beta^{\lambda}_{0} + \boldsymbol{\beta^{\lambda}} \boldsymbol{x^{\lambda}_{i,t}}\right) \cdot \left(\overline{\tau}_{i,t} - \frac{\beta^{G}}{2\beta^{\tau}(\boldsymbol{x^{\tau}_{i,t}})} \cdot p^{F}_{i,t} s_{i,t} m_{i,t} \cdot \left(\beta^{\lambda}_{0} + \boldsymbol{\beta^{\lambda}} \boldsymbol{x^{\lambda}_{i,t}}\right)\right) + \varepsilon_{i,t}. \tag{3.11}$$

The effect of dwelling characteristics on the size of the overprediction is then estimated along with households' preferences for thermal comfort. The error term $\varepsilon_{i,t}$ is a function of $\varepsilon^{\tau}_{i,t}$ and $\varepsilon^{\lambda}_{i,t}$. Given that we control for many dwelling characteristics in the empirical estimation, we are confident that the remaining unexplained systematic measurement error, $\varepsilon^{\lambda}_{i,t}$, is small and uncorrelated to variables in $\boldsymbol{x^{\tau}_{i,t}}$, such that biases can effectively be avoided.

Table 3.1 reports results obtained from estimating equation (3.11). These indicate that there is some heterogeneity in the utility households receive from the consumption of thermal warmth and thus in the choices they make. Older, larger and more educated households are found to have stronger preferences for high mean indoor temperatures. Increases in income have no statistically significant impact on thermal heat consumption of home owners, conditional on the other controls included in the regression equation. However, there is a substantial significant impact of dwelling size on households' temperature consumption, which might capture income and wealth effects that the available income data is not able to pick up.[10]

It is straight forward to derive households' elasticity towards changes in the marginal cost of heating from their optimal temperature consumption stated in equation (3.3). This elasticity can be predicted for every household given the estimates of the utility function parameters reported in table 3.1. We obtain a mean elasticity of households in the estimation sample of -0.377.

The results also confirm that there is some overprediction of the efficiency state that varies with

[10] Unfortunately, income data is only available as a categorical variable in the primary dataset, which limits the level of detail in which it can be analysed. In addition, the dataset undersamples low income households. See Mertesacker (2020b) and Mertesacker (2020a) for more details on the role of income on households' temperature choice and problems in the identification of these effects with the given dataset.

Table 3.1: Estimates of utility function and adjustment term parameters [a]

Dep. Var: $F_{i,t}^{d}$	Coefficients	Standard Errors
Estimates of households' utility function parameters:		
Constant: β_0^{τ}	12.790***	(2.333)
Age ≥ 50	4.071**	(1.858)
# adults:		
2	3.045**	(1.404)
3	6.778**	(3.128)
≥ 4	5.119**	(2.445)
# children:		
1	-1.315	(1.603)
≥ 2	2.928	(1.806)
Is employed	-2.304	(2.000)
Has Abitur	3.673**	(1.478)
Income:		
$< 1,500$ €	3.520	(2.243)
$\geq 3,500$ €	0.898	(1.360)
Size of the dwelling:		
Small: $<$ 1st tercile	-1.487	(1.761)
Large: $>$ 2nd tercile	4.373***	(1.520)
Estimates of adjustment term:		
Constant: β_0^{λ}	0.661***	(0.069)
Row house	-0.207***	(0.029)
Construction year:		
$1919-1968$	0.026	(0.061)
$1969-1977$	0.125*	(0.067)
$1978-1994$	0.240***	(0.065)
≥ 1995	0.159**	(0.062)
Has modernised:		
Windows	0.046	(0.038)
Heating system	-0.024	(0.034)
Thermal shell	0.115***	(0.039)
Income:		
$< 1,500$ €	0.029	(0.058)
$\geq 3,500$ €	0.048	(0.031)
Size of the dwelling:		
Small: $<$ 1st tercile	-0.053	(0.042)
Large: $>$ 2nd tercile	-0.015	(0.031)
Predicted mean elasticity	-0.377***	(0.031)
Observations	1,365	
R-squared	0.893	

[a] The table reports results from an estimation of equation (3.11) by nonlinear least squares. The ideal temperature increase is set to $\bar{\tau}_{i,t} = 21 - \tau_{i,t}^{out}$ degrees Celsius. Year and fuel type fixed effects are included as adjustment term parameters in the regression. Standard errors are reported in parantheses and clustered on the household level. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

characteristics of the dwelling. The intercept, β_0^λ, indicates that the efficiency state of a dwelling from the base group is adjusted to 66.1 % of its predicted value.[11] The size of the adjustment is smaller for newer dwellings and those that have received additional thermal insulation. For row houses the adjustment is stronger. We use estimates from the lower panel of table 3.1 to adjust the efficiency states in our sample and use these for the subsequent analyses. Working directly with the adjusted $s_{i,t}$ greatly reduces the number of state variables that have to be included in the dynamic estimation, which substantially increases calculation time. Before moving to the dynamic estimation, the next section estimates the evolution process using the adjusted data.

3.4.2 Estimation of the evolution process

Using the efficiency states that have been adjusted based on the estimation results from section 3.4.1, we estimate the evolution process of $s_{i,t}$ as a function of the lagged efficiency state and past investment behavior by ordinary least squares. The estimated evolution processes provides an approximation to the full fuel requirement calculation conducted based on the model by Loga et al. (2005). The estimates provide a simple generalised rule to calculate the efficiency states based on the average impact of modernization in our observed sample.

Table 3.2 reports the estimation results for four different specifications of the evolution process. Since no fuel data is required, the entire time span from 2002 to 2008, for which investments are observed, is used for the estimation. All specifications show that the dwelling experiences some depreciation if households do not invest. If they do, the investment is generally successful, as been illustrated by column (1). On average, the amount of fuel required to increase the effective indoor temperature by one degree is reduced by approximately 10.366 $^{kWh}/_{m^2}$. This is quite a substantial improvement, equivalent to roughly 43 % of the mean efficiency state of 23.96 $^{kWh}/_{m^2}$.

Column (3) differentiates between the different investment alternatives that households have. Investing into thermal insulation has by far the largest impact on the overall efficiency level of the dwelling, reducing it by 16.419 $^{kWh}/_{m^2}$. The impact of new windows and a new heating system is smaller, but still statistically and economically significant. While combined investments into windows and thermal insulations have a reinforcing effect, the joint investment into a new heating system and windows or thermal insulation mediates the total impact. The reason for this is also intuitive. The lower the heat loss due to an improved insulation of the dwelling, the lower the absolute benefit that can be realised by a more efficient heating system.

Finally, it is also important to allow for decreasing marginal impacts of investments in our model. Ceteris paribus, the realisation of (large) efficiency gains should be harder for very efficient

[11] The base group is a single family detached dwelling of average size, constructed before 1919, without modernisation investments in the recent years heated with natural gas and inhabited by a household with average income.

Table 3.2: Estimates of evolution process [a]

Dep. Var.: $s_{i,t+1}$	(1)	(2)	(3)	(4)
$s_{i,t}$	1.071***	1.100***	1.054***	1.056***
	(0.025)	(0.039)	(0.020)	(0.013)
$s_{i,t}^2$	-0.003***	-0.003**	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.000)
$s_{i,t}^3$	1.62 e-5*	2.58 e-5*	8.71 e-6	1.32 e-5***
	(9.1 e-6)	(1.36 e-5)	(7.26 e-6)	(3.93 e-6)
$r_{i,t}^{iso}$			-16.419***	2.927***
			(0.300)	(0.351)
$r_{i,t}^{win}$			-3.056***	-1.520***
			(0.134)	(0.375)
$r_{i,t}^{iso} * r_{i,t}^{win}$			-1.587*	0.823
			(0.830)	(0.890)
$r_{i,t}^{heat}$			-5.859***	1.864***
			(0.187)	(0.478)
$r_{i,t}^{iso} * r_{i,t}^{heat}$			3.579***	-0.861
			(1.147)	(1.542)
$r_{i,t}^{win} * r_{i,t}^{heat}$			0.747	3.983
			(0.941)	(2.761)
$r_{i,t}^{iso} * r_{i,t}^{win} * r_{i,t}^{heat}$			0.395	-1.663
			(2.067)	(3.409)
$r_{i,t}^{iso} * s_{i,t}$				-0.638***
				(0.012)
$r_{i,t}^{win} * s_{i,t}$				-0.059***
				(0.015)
$r_{i,t}^{iso} * r_{i,t}^{win} * s_{i,t}$				0.026
				(0.029)
$r_{i,t}^{heat} * s_{i,t}$				-0.317***
				(0.023)
$r_{i,t}^{iso} * r_{i,t}^{heat} * s_{i,t}$				0.174***
				(0.053)
$r_{i,t}^{win} * r_{i,t}^{heat} * s_{i,t}$				-0.077
				(0.113)
$r_{i,t}^{iso} * r_{i,t}^{win} * r_{i,t}^{heat} * s_{i,t}$				0.013
				(0.129)
$r_{i,t}$	-10.366***	4.494***		
	(0.195)	(0.589)		
$r_{i,t} * s_{i,t}$		-0.532***		
		(0.024)		
Constant	0.244	-0.394	0.239	-0.024
	(0.203)	(0.301)	(0.164)	(0.102)
Observations	27,756	27,756	27,756	27,756
R-squared	0.927	0.945	0.953	0.968

[a] The table reports results from estimations of several specifications of the evolution process characterized in equation (3.5) by ordinary least squares. The binary variables $r_{i,t}^{iso}$, $r_{i,t}^{win}$ and $r_{i,t}^{heat}$ equal one if an investment into the thermal insulation of the dwelling, new windows or a new heating system has occured and zero otherwise. The dummy variable $r_{i,t}$ equals one if any of these investments have occured. Standard errors are reported in parantheses and clustered on the household level. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

dwellings than for inefficient ones. Column (2) confirms that this is indeed the case. The general impact of investing in any type of modernisation measure is the stronger the higher the fuel requirement of a dwelling is before the investment. In column (4) all investment opportunities analysed in column (3) are interacted with the efficiency state, $s_{i,t}$, providing the estimates used in the estimation of the dynamic model.

3.4.3 Estimation of investment cost

Given that we have successfully estimated the parameters of the period utility function and the evolution process, we are now also able to estimate the fixed investment cost using our dynamic model of household investment. We first obtain the value function, $V(s_{i,t})$, of equation (3.6) by value function iteration, which we then use to calculate the expected gain from investment, $\Delta EV(s_{i,t}|\boldsymbol{x^{\tau}_{i,t}})$, for every household in the sample. Given the decision rule in equation (3.7) and our distributional assumption on the investment cost, $C_{i,t}$, we can then determine the investment probability for every household as

$$Pr(r_{i,t}=1|s_{i,t},\boldsymbol{x^{\tau}_{i,t}}) = Pr\left(C_{i,t} \leq \delta\Delta EV(s_{i,t}|\boldsymbol{x^{\tau}_{i,t}})\right). \tag{3.12}$$

We assume that households' state variables are independent of the cost draws and furthermore that all cost draws, across households and time periods, are i.i.d. from the same distribution Φ^{γ}_{C}, conditional on the observable characteristics. In the main estimation, we allow the distribution of modernisation costs to vary with the size of the dwelling. The goal of the dynamic estimation is to estimate the mean values of the different distributions stored in the vector $\boldsymbol{\gamma}$. The likelihood function for households' investment data is

$$\begin{aligned} L(\boldsymbol{\gamma}|s_{i,t},r_{i,t},\boldsymbol{\beta^{\tau}},\boldsymbol{x^{\tau}_{i,t}}) = \prod_{i}^{N}\prod_{t}^{T_i} \quad & [Pr(r_{i,t}=1)\cdot(r^{d}_{i,t}=1) \\ & + (1-Pr(r_{i,t}=1))\cdot(r^{d}_{i,t}=0)], \end{aligned} \tag{3.13}$$

where $r^{d}_{i,t}$ is a binary indicator equal to 1 if household i has conducted a retrofitting investment in period t, N denotes the total number of households in the sample and T_i the number of periods household i is observed in the data. In the estimation we assume the discount factor $\delta = 0.95$ and the ideal temperature increase in the dwelling to be $\bar{\tau}_{i,t} = 21 - \tau^{out}_{i,t}$, where $\tau^{out}_{i,t}$ denotes the mean outdoor temperature.

The estimation procedure thus chooses the mean value of the investment cost distribution such that it rationalises the investments observed in the data, given the expected utility gain from investing, which is determined by our model based on the current efficiency state, $s_{i,t}$, and observable household characteristics, $\boldsymbol{x}_{i,t}^{\tau}$. Table (3.3) reports estimated mean values of the investment cost distribution based on two separate models. In the upper panel, a single mean value has been estimated that applies to all observations in the sample. The mean modernisation cost households encounter is found to equal $66,233$ euros. The columns in the right part of the table report percentiles of the associated exponential distribution. Since the exponential distribution is left-skewed a high probability mass is assigned to modernisation costs below the mean value. This is illustrated by the densities of exponential distributions plotted in figure (3.1). The large cost estimate mirrors households' reluctance to invest, despite substantial potential increases in the dwellings' energetic performance, that has been observed in the previous literature.[12] It captures all factors that prevent households from conducting retrofit investments.[13] These include the actual monetary costs of the material and installation service, but also for instance inconveniences related to the purchase, the risk of failure in achieving the desired efficiency improvement, problems to acquire the necessary capital or just households' unawareness of the efficiency gains that could be realized. The magnitude of the cost estimate indicates that the impediments to investments are large and likely hard to address even with well designed policies.

The lower panel of table (3.3) reports results of a separate estimation, that allowed modernisation costs to vary by the size of the dwelling. It indicates that the modernisation costs are larger in the group of households living in larger dwellings. An intuitive explanation for this pattern is that the size of the heating system, the number of windows or the area that has to be covered by better insulating materials get larger and therefore more expensive as the size of the dwelling increases.

Given the estimated investment cost distribution, the investment probability can be calculated for every household in the sample. The mean investment probability equals 6.305 % and 6.194 % for the estimates from panel A and B, respectively. Both estimates thus provide a good fit to the data in which the investment rate is 6.154 %.

A distinctive feature of our model is that it allows us to structurally separate the fixed cost that arise at the time of the investment from the (expected) long-run gains that are associated to it. The separation of the two effects requires a dynamic modelling of the investment decision. Static utility models of the investment decision can, in contrast, only estimate the net effect of

[12]The previous literature has often stated households' reluctance to invest in terms of high estimated discount rates. Our estimate of high investment costs is an equivalent way to express the same pattern in the data. To see this more clearly, note that the discount rate and the investment cost draw are two sides of the same coin in equation (3.7). An alternative estimation could follow our approach, but fix the mean of the investment cost distribution and estimate the discount factor using maximum likelihood.

[13]Because the investment cost are estimated in direct relation to the infinite stream of expected future period utilities, they also have to be interpreted in monetary terms. That is, with respect to the marginal utility of income which has been normalised to unity.

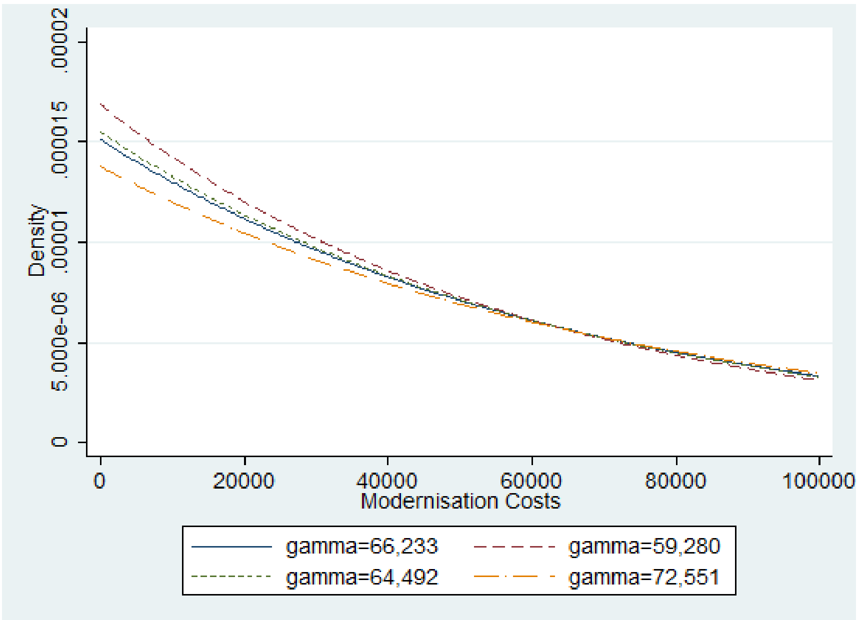
Table 3.3: Estimates of modernisation costs [a]

	$\hat{\gamma}$	SE	Percentiles				
			5th	25th	50th	75th	95th
Single Cost Estimate							
	66,233	(6,624)	3,397	19,054	45,909	91,818	198,416
Estimated costs by dwelling size [b]							
Small	59,280	(15,708)	3,040	17,053	41,089	82,180	177,587
Medium	64,492	(17,233)	3,308	18,553	44,702	64,492	193,201
Large	72,551	(0,579)	3,721	20,872	50,289	100,577	214,343

[a] The table reports estimates of mean values of the exponential distribution of modernisation costs, $\hat{\gamma}$, based on two separate models. In the upper panel, a single mean value has been estimated that applies to all observations in the sample. In the lower panel three different values have been estimated based on the size of the dwelling. The estimates are obtained by maximizing the likelihood function in equation (3.13). Standard errors of the estimates are reported in parantheses. The right part of the table reports some percentiles of the exponential distributions associtated to the reported mean values.

[b] Dwellings have been assigned to the group of small, medium and large dwellings based on the terciles of the dwelling size distribution. Dwellings below the first tercile of the distribution are considered small, those between the first and second tercile are medium-sized and those above the second tercile are large.

Figure 3.1: Density of an exponential distribution with mean γ



the two different factors entering the decision process. This makes the use of such models for counterfactual policy scenarios problematic, since it is generally unclear, whether the hypothetical policy change can be incorporated correctly into the decision process of households. Changes in investment cost and expected utility gains are, however, very likely to trigger substantially different economic behavior and outcomes.

The role of the dwelling size nicely illustrates the value of the dynamic estimation framework. The static results reported in table (3.1) indicate that the benefits from investing increase as dwelling size increases, because households living in larger dwellings are found to value higher mean indoor temperatures more than those living in smaller dwellings. In addition, the absolute level of savings from lower values of $s_{i,t}$, is naturally larger for larger dwellings. The dynamic estimates reveal that these effects are counteracted by higher costs that are associated to investments in larger dwellings. A static regression of investment on dwelling size (e.g., by logit or probit) would only estimate the net effect of both influences on households' decision to retrofit.

The crucial advantage of the developed dynamic framework is that it allows to quantify the temperature consumption, the associated level of period utility and the lifetime benefit of investing for every household in the sample. It also allows to investigate the changes in households' incentives and behavior if certain variables are varied. This enables us to study the two principle forces that govern households' decision process and their interplay in a comprehensive model.

Table 3.4 reports model predictions of households' optimal temperature increase, $\tau^*_{i,t}$, and fuel consumption, $F_{i,t}$, as well as the associated period utility, $u(\tau^*_{i,t})$, at different values of important model variables. Panel A considers the first, second and third quartile of the efficiency state distribution. It clarifies that households' energy consumption rebounds after an increase in dwellings' efficiency level resulting from a modernisation. A household living in a more efficient dwelling (i.e., having a lower $s_{i,t}$), faces lower marginal cost of consuming thermal warmth and therefore consumes a higher level of $\tau_{i,t}$. As the energy required to increase the mean indoor temperature by one degree Celsius decreases from 27.275 to 19.514 kWh, households' chosen mean indoor temperature rises from 7.620 to 9.464 degree Celsius. The larger mean indoor temperature in the dwelling might be due to a general increase of the ambient temperature level or due to adjustments in the number of rooms being heated or the length of time periods the desired temperature level is reached. The higher temperature consumption implies an increase in period utility, $\tilde{u}(\tau^*_{i,t})$, from -2.712 to -2.128, resulting from reduced thermal discomfort. Despite the increase of temperature consumption, the amount of fuel households consume, $F_{i,t}$, declines, such that their overall expenses for the consumption of thermal warmth decline, which also contributes to the increase in period utility. At the first quartile of the efficiency state distribution, period utility is -1.657 and thus reduced further by the same mechanisms.

The remaining three columns allow to inspect households' incentives to invest at the considered efficiency states. They report the long-run benefit of investing, ΔEV, in thousand euros, as well as the resulting investment probability, $Pr(r_{i,t} = 1)$, and the mean costs households face if they invest, $E[C_{i,t}^{med}|r_{i,t} = 1]$, assuming all households to draw their investment costs from the distribution associated to medium-sized dwellings. The expected gain in lifetime utility associated to an investment gets the smaller, the more efficient the dwelling is. At the first and third quartile of the efficiency state distribution ΔEV equals $6,098$ and $2,792$ euros, respectively. Intuitively, as the level of discomfort from temperature levels below the ideal level declines, also the absolute gain that can be realized through modernisations gets smaller. In addition, the improvements in the energetic performance of the dwelling get the smaller the more efficient the dwelling was before the investment.[14] The smaller long-run benefit of investing, implies that households living in more efficient dwellings are less likely to invest. While the investment rate is 9 % at the third quartile of the efficiency distribution, it only equals 4.2 % at the first quartile. At the median efficiency state the investment probability is 7.4 %. The decision rule of equation (3.7) implies that households only invest if their one-time investment costs are below the lifetime gain in utility. The left-skewed form of the exponential distribution ensures that mean investment costs conditional on investment, reported in the last column of table 3.4, are substantially smaller than ΔEV. They equal $3,999$ and $1,386$ euros at the third and first quartile of the efficiency distribution respectively.

It is interesting to compare these values to actual monetary costs households encounter when modernising their dwellings. For this purpose appendix C.1 provides summary statistics of the monetary cost of actual modernisation investments conducted by German home owners between 2010 and 2015. The data is obtained from the 34th version of "The German Socioeconomic Panel Study" (SOEP v34), which is a large representative household panel of the German population. The reported investment cost have been deflated to 2007 euros. An advantage of the modernisation cost reported in the SOEP is that they relate to a very similar – relatively broad – measure of investment activity as the information in our main dataset. Households are asked whether they have invested into a new heating system, new thermal insulation or new windows in the last year and what the associated cost have been. The mean cost of households conducting any of the three modernisation investments are 6731.87 euros.[15] This is substantially below the mean

[14] See the discussion of the estimated evolution process in section 3.4.2.

[15] Actual monetary cost of retrofitting are hard to measure and vary depending on the extent and type of the modernisation conducted. The conformity of the questions in the "Residential Energy Consumption Survey" and the SOEP is therefore a valuable opportunity to obtain an impression of the monetary cost of the type of retrofit investment considered in this study. Studies providing estimates of investment cost include Palmer et al. (2017), who consider the housing market in the United Kingdom. Their results indicate an average monetary installation cost

Table 3.4: Model predictions [a]

PANEL A:

$s_{i,t}$	$\tau^*_{i,t}$	$F_{i,t}$	$\tilde{u}(\tau^*_{i,t})$	ΔEV	$Pr(r_{i,t}=1)$	$E[C^{med}_{i,t} \vert r_{i,t}=1]$ [b]
14.343	10.718	1399.983	-1.657	2.792	0.042	1.386
	(0.046)	(10.139)	(0.014)	(0.017)	(0.000)	(0.009)
19.514	9.464	1653.395	-2.128	4.946	0.074	2.440
	(0.063)	(12.090)	(0.017)	(0.031)	(0.000)	(0.015)
27.275	7.620	1798.381	-2.712	6.098	0.090	2.999
	(0.083)	(17.520)	(0.020)	(0.041)	(0.001)	(0.020)

PANEL B:

# adults	$\tau^*_{i,t}$	$F_{i,t}$	$\tilde{u}(\tau^*_{i,t})$	ΔEV	$Pr(r_{i,t}=1)$	$E[C^{med}_{i,t} \vert r_{i,t}=1]$ [b]
1	7.776	1218.316	-1.996	3.346	0.050	1.654
	(0.121)	(17.135)	(0.025)	(0.065)	(0.001)	(0.032)
2	8.792	1454.114	-2.133	3.829	0.057	1.890
	(0.064)	(10.382)	(0.018)	(0.044)	(0.001)	(0.021)
3	10.414	1859.583	-2.344	4.161	0.062	2.052
	(0.059)	(16.555)	(0.028)	(0.062)	(0.001)	(0.030)
4	9.206	1553.459	-2.188	4.055	0.060	2.000
	(0.104)	(17.682)	(0.032)	(0.080)	(0.001)	(0.039)

PANEL C:

Age	$\tau^*_{i,t}$	$F_{i,t}$	$\tilde{u}(\tau^*_{i,t})$	ΔEV	$Pr(r_{i,t}=1)$	$E[C^{med}_{i,t} \vert r_{i,t}=1]$ [b]
< 50	8.436	1372.983	-2.085	3.602	0.054	1.779
	(0.066)	(10.658)	(0.016)	(0.040)	(0.001)	(0.019)
≥ 50	9.735	1686.961	-2.257	4.121	0.061	2.032
	(0.050)	(10.701)	(0.018)	(0.043)	(0.001)	(0.021)

[a] The table reports the mean predictions of several model quantities setting the different model variables for all households to predefined values. Panel A considers different values of the efficiency state, $s_{i,t}$, evaluating households' choices at the first, second and third quartile of the efficiency state distribution. In panel B and panel C mean values are reported for varying size and age of the household. The chosen temperature increase is denoted by $\tau^*_{i,t}$, $\tilde{u}(\tau^*_{i,t})$ provides the period utility from thermal warmth consumption. The fuel consumption is denoted $F_{i,t}$, ΔEV is the long-run benefit from investing, $Pr(r_{i,t}=1)$ the investment probability and $E[C^{med}_{i,t}|r_{i,t}=1]$ the mean cost of those households that the model predicts to invest. The quantities $\tilde{u}(\tau^*_{i,t})$, ΔEV and $E[C^{med}_{i,t}|r_{i,t}=1]$ are stated in thousand euros, $F_{i,t}$ in thousand kilowatt hours. Standard errors are reported in parantheses.

[b] The term $C^{med}_{i,t}$ indicates draws from the cost distribution that applies to medium-sized dwellings.

unconditional costs reported in table 3.3. The results of table 3.4 indicate, that at such monetary cost investments would be profitable for many households at the third quartile of the efficiency state distribution. In contrast, for households living in more efficient dwellings the monetary costs alone might often make investments unattractive.

Panel B and C of table 3.4 consider the effects of households' preferences for thermal comfort on their temperature and investment choices. As discussed in section 3.4.1 larger households as well as those older than 50 value higher indoor temperatures more. Panel B and C clarify that this is actually associated to increases in temperature levels and fuel consumption that are statistically and economically significant. Compared to a one person household, a household with two adult members consumes a more than one degree higher mean indoor temperature, resulting in an average increase of yearly fuel consumption by 235.718 kWh. The stronger preferences for thermal comfort results in a lower overall period utility level for the larger household. This implies larger potential benefits that can be realized through investments and accordingly a higher investment rate. The mechanisms and qualitative effects are the same when comparing households form above 50 to younger counterparts.

3.4.4 Simulation and policy analysis

Using the estimates of our model, we simulate the effects of changes in households' economic environment on temperature consumption and the probability to invest. Table 3.5 summarises the percentage changes of the investment probability, $Pr(r_{i,t} = 1)$, the long-run utility gain from investing, ΔEV, the period utility, $\tilde{u}(\tau^*_{i,t})$, the optimal temperature choice, $\tau^*_{i,t}$, and fuel expenditure, $p^F_{i,t} \cdot F_{i,t}$, five years after three different policy scenarios compared to the status quo scenario without exogenous changes in the economic environment.

In the first experiment, we consider the impact of a fuel price increase by 10 % (e.g., through the introduction of a new tax). The price increase raises the marginal cost of consuming thermal warmth, leading to a lower period utility of the household due to lower consumption of thermal comfort and higher cost associated with its remaining consumption. The decrease in period utility implies larger potential utility gains to be realised by investing into energy efficient technology. The simulation results in the first row of table 3.5 indicate, that the incentive to invest, ΔEV, increases by 5.1 % on average going along with an increase of the investment probability by 22.1 %. The efficiency increases due to the additional investments, counteract the fuel tax's impact on average consumption of thermal warmth and period utility. However, simulation results

around 8,000 euros as a plausible benchmark. For the German market Thema et al. (2018) calculate the complete modernisation of an apartment building (including the retrofit of all walls, windows and the heating system) to cost roughly 61,500 euros. However, note that while this helps to rationalise the occasional occurence of extremely high investment cost, households hardly ever conduct investments of that scale at once.

Table 3.5: Simulation of counterfactual policy scenarios [a]

Scenario	$Pr(r_{i,t}=1)$	ΔEV	$\tilde{u}(\tau^*_{i,t})$	$\tau^*_{i,t}$	$p^F_{i,t}\cdot F_{i,t}$
Fuel price increase [b]	0.221	0.051	-0.079	-0.048	0.042
	(0.054)	(0.003)	(0.001)	(0.001)	(0.002)
More effective modernisation [c]	0.201	0.009	0.006	0.004	-0.005
	(0.054)	(0.003)	(0.001)	(0.001)	(0.001)
Subsidy on investment costs [d]	0.463	-0.002	0.003	0.004	-0.001
	(0.064)	(0.003)	(0.001)	(0.001)	(0.001)

[a] The table reports the percentage changes of the investment probability, $Pr(r_{i,t}=1)$, the long-run utility gain from investing, ΔEV, the period utility, $\tilde{u}(\tau^*_{i,t})$, the optimal temperature choice, $\tau^*_{i,t}$, and expenditure for for fuel, $p^F_{i,t}\cdot F_{i,t}$, five years after three different policy scenarios compared to the status quo scenario without exogenous changes in the economic environment. The relative changes are calculated as ${}^{\Lambda^{ps}-\Lambda^{sq}}/{}_{|\Lambda^{sq}|}$, where Λ and ps denote the considered quantity and policy scenario respectively and sq indicates the status quo. Standard errors are reported in parantheses.

[b] The scenario considers a general fuel price increase by 10 %.

[c] The scenario considers an increase of the impact of the modernisation by 10 %.

[d] The scenario considers a reduction of modernisation costs the household as to pay by 20 % of the average cost.

indicate that still the average consumed level of indoor temperature decreases by 4.8 %. Yet, the temperature reduction does not offset the negative monetary affect of the fuel price increase on total spending. On average households have to spend 4.2 % more on the consumption of fuel. Together with the lower level of thermal comfort, this determines the period utility to decline by 7.9 % on average.

While the impact of a fuel price increase by 10 % thus allows a significant reduction in temperature and thus fuel consumption, the impact of the other two policies we have simulated are less effective. Improving the effectiveness of the retrofitting by 10 % and subsidising investments by 20 % of average construction cost increases investment by 20.1 % and 46.3 %, respectively. However, the policies have no economically significant impact on households' choice of indoor temperature and energy consumption. The effective changes in the efficiency states of the dwellings triggered by the policies are too small to generate significant changes in the average temperature choice and fuel consumption across households. This indicates that policies designed to exclusively increase domestic retrofitting without setting incentives to reduce fuel consumption in households' static decision environment, will therefore have problems to reduce the amount of fuel consumed in the economy by an amount that helps to meet climate targets.

3.5 Conclusion

In this article we develop and estimate a dynamic structural model of households' investment decision in energy efficient technology. In our model, households are forward looking. When making the decision to invest, they trade-off one-time fixed cost in the current period against long-run gains in utility. By explicitly modelling and estimating the utility households receive in every period from the consumption of thermal warmth and other goods in a first step, we can predict the change in fuel consumption as well as the lifetime utility gains that result from an investment. The investment costs are then estimated in relation to the potential gains in a second step. They are chosen such that observed investments can be rationalized by the developed dynamic decision model. All estimated parameters thus have a structural interpretation in terms of the developed economic decision model. This allows us to analyse households' retrofit decision in greater detail than related approaches that rely on static utility models to estimate the relationship between household and dwelling characteristics and the propensity to invest. Using the structural parameter estimates of the model, we can conduct counterfactual experiments on how households would respond in terms of their energy consumption and investment behavior given changes in the economic environment.

We estimate the model using a subsample from "The German Residential Energy Consumption Survey". We find that households' valuation for thermal comfort as well as the marginal cost associated to its consumption matter for their temperature and investment choice. A household living in a less efficient dwelling consumes lower temperature levels and has, ceteris paribus, higher incentives to invest. A household with stronger preferences for high temperature levels, e.g. an older household, chooses a higher mean indoor temperature and has stronger incentives to retrofit the dwelling to further decrease the cost of temperature consumption. The results thus clarify the importance of understanding the sources of heterogeneity in households' static temperature choice to achieve a detailed understanding of the mechanisms that determine their investment choices.

In our simulations of counterfactual policies, we find that government subsidy programs that reduce investment costs have a positive impact on the investment decision. In our case, they increase the investment rate by 46.3 %. Yet, the resulting increase in the economy wide level of efficiency is not large enough to reduce households' average energy consumption on a significant scale. Similarly, increasing the effectiveness of modernisation measures, for instance by supporting research and development, increases the rate of investment, but has only a small impact on the fuel consumption in the economy. In contrast, a tax on energy prices induces a higher investment rate, which increases by 22.1 percent, and at the same time is effective in the reduction of temperature and thus fuel consumption. According to our simulation, ten percent higher fuel prices reduce the mean indoor temperature in dwellings by 4.8 %. The results thus emphasize the strength of direct taxes on fuel consumption in incentivising households to reduce their consumption, both, by reducing temperature choices as well as increasing their engagement to retrofit.

Appendices

A Appendix to chapter 1

A.1 Robustness analyses and descriptives

Table A.1: Coefficients of naive regression using the Poisson estimator [a]

Dep. Var: $\tau_{i,t}^{d}$	(1)	(2)	(3)
Income:			
$< 1,500$ €	-0.238***		-0.300***
	(0.051)		(0.052)
$\geq 3,500$ €	0.122***		0.150***
	(0.031)		(0.029)
log($c_{i,t}$)		-0.337***	-0.367***
		(0.024)	(0.025)
Constant	1.956***	3.685***	3.823***
	(0.022)	(0.126)	(0.132)
Observations	2,483	2,846	2,483
Estimator	POI	POI	POI

[a] Columns (1) to (3) of the table report estimation results from Poisson regressions (POI) on different specifications of equation (1.19). Year fixed effects are included in all regressions and the standard errors are clustered on the household level and reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

Table A.2: Estimated coefficients of the full regression model using the Poisson estimator [a]

Dep. Var: $\tau^d_{i,t}$	(1)	(2)	(3)	(4)
Income:				
$< 1,500$ €	-0.101**	-0.099**	-0.076*	
	(0.044)	(0.042)	(0.041)	
$\geq 3,500$ €	0.022	0.024	0.011	
	(0.028)	(0.025)	(0.026)	
$\log(c_{i,t})$	-0.606***	-0.692***	-0.663***	
	(0.025)	(0.039)	(0.043)	
age:				
$30-39$	0.091	0.063	0.060	
	(0.065)	(0.061)	(0.062)	
$40-49$	0.092	0.062	0.083	
	(0.063)	(0.059)	(0.059)	
$50-59$	0.200***	0.118**	0.135**	
	(0.063)	(0.058)	(0.059)	
≥ 60	0.240***	0.123**	0.142**	
	(0.066)	(0.061)	(0.062)	
# adults:				
2	0.166***	0.138***	0.175***	
	(0.038)	(0.033)	(0.033)	
3	0.199***	0.181***	0.225***	
	(0.048)	(0.041)	(0.041)	
≥ 4	0.220***	0.154***	0.216***	
	(0.056)	(0.053)	(0.053)	
# children:				
1	0.050	0.031	0.048	
	(0.043)	(0.040)	(0.039)	
≥ 2	0.083*	0.037	0.073*	
	(0.043)	(0.039)	(0.038)	
Is employed	0.009	-0.029	-0.036	
	(0.032)	(0.029)	(0.030)	
Has Abitur	0.006	0.027	0.029	
	(0.025)	(0.023)	(0.024)	
Is owner	0.244***	0.047	0.035	
	(0.037)	(0.033)	(0.035)	
Size of the dwelling:				
Small: $<$ 1st tercile	-0.221***	-0.105***	-0.107***	-0.014
	(0.038)	(0.034)	(0.035)	(0.035)
Large: $>$ 2nd tercile	0.183***	0.170***	0.166***	-0.024
	(0.030)	(0.028)	(0.030)	(0.027)

Table A.2 continued from previous page

Type of the dwelling:			
Row house	-0.185***	-0.181***	-0.075***
	(0.025)	(0.027)	(0.024)
Multi-family dwelling	-0.267***	-0.293***	-0.129**
	(0.057)	(0.055)	(0.062)
# apartments:			
$4-6$	-0.224***	-0.201***	-0.186***
	(0.059)	(0.057)	(0.061)
$7-12$	-0.280***	-0.225***	-0.257***
	(0.059)	(0.058)	(0.061)
≥ 13	-0.297***	-0.290***	-0.308***
	(0.076)	(0.072)	(0.072)
Construction year:			
≤ 1918	0.032	0.013	-0.261***
	(0.048)	(0.051)	(0.052)
$1919-1948$	-0.053	-0.109**	-0.309***
	(0.046)	(0.048)	(0.048)
$1949-1957$	-0.043	-0.072	-0.236***
	(0.051)	(0.053)	(0.054)
$1958-1968$	-0.032	-0.069	-0.239***
	(0.044)	(0.045)	(0.044)
$1978-1983$	0.087**	0.091**	0.166***
	(0.044)	(0.046)	(0.043)
$1984-1994$	0.018	0.038	0.207***
	(0.046)	(0.047)	(0.043)
$1995-2001$	-0.031	0.018	0.316***
	(0.046)	(0.048)	(0.040)
≥ 2002	-0.198***	-0.157***	0.165***
	(0.056)	(0.059)	(0.050)
Type of heating system:			
gas heating on the floor	-0.050*	-0.035	-0.036
	(0.029)	(0.029)	(0.029)
oven	0.070	0.013	0.086
	(0.083)	(0.094)	(0.082)
Has modernised:			
Thermal shell	0.034	0.038	0.315***
	(0.032)	(0.035)	(0.028)
Windows	0.028	0.040	0.083***
	(0.028)	(0.029)	(0.032)
Heating system	0.006	0.040	0.085***
	(0.026)	(0.028)	(0.028)

Table A.2 continued from previous page

log(HDD)		0.030	0.064	0.065
		(0.111)	(0.111)	(0.114)
Fuel type:				
Oil		-0.002	0.007	-0.040
		(0.027)	(0.029)	(0.028)
Long distance heat		-0.148***	-0.150**	-0.096**
		(0.052)	(0.061)	(0.045)
Constant	4.574***	5.175***	4.633***	1.464
	(0.144)	(0.934)	(0.950)	(0.922)
$\frac{1}{1+\hat{\beta}_{c_{i,t}}}$		3.250***	2.967***	
		(0.417)	(0.378)	
Observations	2,314	2,256	2,254	2,784
Estimator	POI	POI	OLS	POI

[a] The table reports estimation results from ordinary least squares (OLS) and Poisson regressions of equations (1.20) and (1.19) respectively. Year fixed effects are included in all regressions and the standard errors are clustered on the household level and reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

Table A.3: Robustness checks using alternative data restrictions [a]

Dep. Var: $\tau^d_{i,t}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income:								
$< 1,500$ €	0.015	-0.117**	-0.101*	-0.038	-0.061	-0.083**	-0.076*	-0.076*
	(0.056)	(0.056)	(0.052)	(0.044)	(0.042)	(0.042)	(0.041)	(0.041)
$\geq 3,500$ €	0.012	-0.037	-0.005	-0.051*	0.004	0.008	0.011	0.011
	(0.029)	(0.057)	(0.032)	(0.028)	(0.024)	(0.027)	(0.026)	(0.026)
$\log(c_{i,t})$	-0.761***	-0.463***	-0.669***	-0.695***	-0.646***	-0.685***	-0.663***	-0.663***
	(0.046)	(0.089)	(0.052)	(0.046)	(0.041)	(0.044)	(0.043)	(0.043)
age:								
$30 - 39$	-0.139	0.092	-0.003	0.084	0.048	0.078	0.060	0.060
	(0.091)	(0.081)	(0.065)	(0.066)	(0.057)	(0.062)	(0.062)	(0.062)
$40 - 49$	-0.136	0.179**	0.063	0.096	0.095*	0.089	0.083	0.083
	(0.091)	(0.076)	(0.063)	(0.060)	(0.054)	(0.059)	(0.059)	(0.059)
$50 - 59$	-0.038	0.144*	0.113*	0.149**	0.134**	0.131**	0.135**	0.135**
	(0.089)	(0.081)	(0.062)	(0.060)	(0.055)	(0.058)	(0.059)	(0.059)
≥ 60	-0.029	0.204**	0.137**	0.138**	0.143**	0.155**	0.142**	0.142**
	(0.090)	(0.094)	(0.067)	(0.066)	(0.059)	(0.061)	(0.062)	(0.062)
# adults:								
2	0.166***	0.167***	0.176***	0.158***	0.167***	0.165***	0.175***	0.175***
	(0.046)	(0.048)	(0.042)	(0.034)	(0.032)	(0.034)	(0.033)	(0.033)
3	0.221***	0.196***	0.208***	0.180***	0.228***	0.215***	0.225***	0.225***
	(0.053)	(0.072)	(0.053)	(0.043)	(0.039)	(0.043)	(0.041)	(0.041)
≥ 4	0.188***	0.248**	0.196***	0.221***	0.243***	0.227***	0.216***	0.216***
	(0.062)	(0.123)	(0.069)	(0.059)	(0.057)	(0.057)	(0.053)	(0.053)
# children:								
1	0.067	0.031	0.123***	0.016	0.062*	0.049	0.048	0.048
	(0.047)	(0.066)	(0.045)	(0.043)	(0.036)	(0.041)	(0.039)	(0.039)
≥ 2	0.090**	0.090	0.083*	0.040	0.104***	0.082*	0.073*	0.073*
	(0.042)	(0.094)	(0.049)	(0.044)	(0.035)	(0.043)	(0.038)	(0.038)

Table A.3 continued from previous page

Is employed	-0.024	-0.021	-0.041	-0.015	-0.041	-0.030	-0.036	-0.036
	(0.035)	(0.055)	(0.038)	(0.034)	(0.029)	(0.031)	(0.030)	(0.030)
Has Abitur	0.041	-0.006	0.022	0.034	0.021	0.017	0.029	0.029
	(0.028)	(0.045)	(0.029)	(0.025)	(0.023)	(0.026)	(0.024)	(0.024)
Is owner			0.039	-0.010	0.037	0.054	0.035	0.035
			(0.043)	(0.034)	(0.033)	(0.038)	(0.035)	(0.035)
Size of the dwelling:								
Small: $<$ 1st tercile	-0.127***	-0.071	-0.096**	-0.226***	-0.114***	-0.120***	-0.107***	-0.107***
	(0.039)	(0.078)	(0.043)	(0.039)	(0.034)	(0.039)	(0.035)	(0.035)
Large: $>$ 2nd tercile	0.189***	0.241***	0.206***	0.166***	0.154***	0.185***	0.166***	0.166***
	(0.032)	(0.085)	(0.037)	(0.031)	(0.029)	(0.032)	(0.030)	(0.030)
Type of the dwelling:								
Row house	-0.210***	-0.067	-0.186***	-0.133***	-0.193***	-0.216***	-0.181***	-0.181***
	(0.028)	(0.090)	(0.033)	(0.030)	(0.025)	(0.028)	(0.027)	(0.027)
Multi-family dwelling	-0.340***	-0.166*	-0.292***	-0.280***	-0.287***	-0.254***	-0.293***	-0.293***
	(0.080)	(0.100)	(0.067)	(0.055)	(0.051)	(0.061)	(0.055)	(0.055)
# apartments:								
$4-6$	-0.182**	-0.200***	-0.216***	-0.155***	-0.186***	-0.211***	-0.201***	-0.201***
	(0.091)	(0.077)	(0.074)	(0.054)	(0.054)	(0.064)	(0.057)	(0.057)
$7-12$	-0.292***	-0.194**	-0.216***	-0.161***	-0.223***	-0.262***	-0.225***	-0.225***
	(0.088)	(0.082)	(0.067)	(0.057)	(0.053)	(0.062)	(0.058)	(0.058)
≥ 13	-0.290**	-0.231**	-0.320***	-0.227***	-0.298***	-0.321***	-0.290***	-0.290***
	(0.141)	(0.092)	(0.094)	(0.069)	(0.068)	(0.077)	(0.072)	(0.072)
Construction year:								
≤ 1918	0.090	-0.191**	0.043	-0.013	-0.008	0.013	0.013	0.013
	(0.055)	(0.092)	(0.062)	(0.054)	(0.048)	(0.054)	(0.051)	(0.051)
$1919-1948$	-0.046	-0.242***	-0.067	-0.112**	-0.109**	-0.114**	-0.109**	-0.109**
	(0.055)	(0.086)	(0.058)	(0.049)	(0.045)	(0.049)	(0.048)	(0.048)
$1949-1957$	-0.031	-0.177*	-0.047	-0.114**	-0.083	-0.070	-0.072	-0.072
	(0.063)	(0.093)	(0.065)	(0.057)	(0.051)	(0.055)	(0.053)	(0.053)

Table A.3 continued from previous page

1958 – 1968	0.053	-0.270***	-0.060	-0.093*	-0.074*	-0.090*	-0.069	-0.069
	(0.052)	(0.076)	(0.061)	(0.048)	(0.043)	(0.047)	(0.045)	(0.045)
1978 – 1983	0.107**	0.053	0.110*	0.038	0.100**	0.090*	0.091**	0.091**
	(0.051)	(0.091)	(0.064)	(0.050)	(0.043)	(0.048)	(0.046)	(0.046)
1984 – 1994	0.013	0.136	0.024	-0.005	0.044	0.058	0.038	0.038
	(0.053)	(0.107)	(0.061)	(0.054)	(0.045)	(0.051)	(0.047)	(0.047)
1995 – 2001	-0.045	0.163*	0.007	0.015	0.014	0.054	0.018	0.018
	(0.057)	(0.090)	(0.061)	(0.050)	(0.046)	(0.053)	(0.048)	(0.048)
$\geq$ 2002	-0.242***	-0.040	-0.167**	-0.193***	-0.143**	-0.095	-0.157***	-0.157***
	(0.065)	(0.109)	(0.069)	(0.062)	(0.056)	(0.071)	(0.059)	(0.059)
Type of heating system:								
gas heating on the floor	-0.044	0.001	-0.033	-0.032	-0.047*	-0.032	-0.035	-0.035
	(0.034)	(0.055)	(0.030)	(0.031)	(0.027)	(0.030)	(0.029)	(0.029)
oven	0.072	-0.116	-0.089	-0.006	-0.023	-0.007	0.013	0.013
	(0.101)	(0.126)	(0.139)	(0.098)	(0.089)	(0.097)	(0.094)	(0.094)
Has modernised:								
Thermal shell	0.015	0.058	0.033	0.046	0.047	0.050	0.038	0.038
	(0.037)	(0.089)	(0.041)	(0.038)	(0.034)	(0.037)	(0.035)	(0.035)
Windows	0.007	0.121*	0.020	0.051	0.042	0.046	0.040	0.040
	(0.032)	(0.065)	(0.036)	(0.033)	(0.028)	(0.030)	(0.029)	(0.029)
Heating system	-0.013	0.192***	0.046	0.081***	0.038	0.041	0.040	0.040
	(0.031)	(0.052)	(0.034)	(0.030)	(0.026)	(0.030)	(0.028)	(0.028)
log(HDD)	0.004	0.170	0.138	-0.096	0.086	0.068	0.064	0.064
	(0.127)	(0.232)	(0.136)	(0.124)	(0.107)	(0.119)	(0.111)	(0.111)
Constant	5.870***	2.632	4.095***	6.086***	4.391***	4.711***	4.633***	4.633***
	(1.101)	(1.932)	(1.156)	(1.052)	(0.912)	(1.005)	(0.950)	(0.950)

Table A.3 continued from previous page

$^1/_{1+\hat{\beta}_{c_{i,t}}}$	4.192***	1.862***	3.019***	3.278***	2.826***	3.175***	2.967***	2.967***
	(0.817)	(0.308)	(0.476)	(0.494)	(0.330)	(0.441)	(0.378)	(0.378)
Observations	1,563	691	1,498	1,806	2,447	2,332	2,254	2,254
R-squared	0.493	0.523	0.584	0.583	0.558	0.548	0.557	0.557
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Owners included	Yes	**No**	Yes	Yes	Yes	Yes	Yes	Yes
Tenants included	**No**	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fuel types used[b]	All	All	**HNG**	All	All	All	All	All
Exclude households								
with secondary heating system	No	No	No	**Yes**	No	No	No	No
with adult children	Yes	Yes	Yes	Yes	**No**	Yes	Yes	Yes
reporting no consumption	No	No	No	No	No	No	No	**Yes**
Exclude outliers in[c]								
generated fuel requirements	Yes	Yes	Yes	Yes	Yes	**Yes**	**No**	Yes
implied indoor temperatures	Yes	Yes	Yes	Yes	Yes	**No**	**Yes**	Yes

[a] The table reports robustness checks of the preferred specification, reported in column (2) of table (1.2), towards changes in various restrictions on the dataset. The lower panel of the table summarizes the restrictions imposed in the individual regressions. Differences to the restrictions in the preferred specification are printed in bold.
The estimates are obtained from ordinary least squares regressions of equation (1.20). Year and fuel type fixed effects are included in all regressions and the standard errors are clustered on the household level and reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.
[b] The cells in this row indicate, whether all fuel types used in the preferred specification - natural gas, oil and long distance heat - or only natural gas (hng) are included in the estimation.
[c] The results from a regression excluding both types of outlier corrections are identical to the results in column (6) of the table and are therefore not reported separately.

Table A.4: Regression coefficients using an alternative dependent variable [a]

Dep. Var: $\tilde{\tau}^d_{i,t}$	Coeff.	SE
Income:		
$< 1,500$ €	-0.086**	(0.037)
$\geq 3,500$ €	0.014	(0.023)
log($c_{i,t}$)	-0.639***	(0.038)
	(0.038)	
age:		
$30-39$	0.069	(0.051)
$40-49$	0.084*	(0.049)
$50-59$	0.136***	(0.050)
≥ 60	0.141***	(0.052)
# adults:		
2	0.145***	(0.029)
3	0.193***	(0.036)
≥ 4	0.183***	(0.047)
# children:		
1	0.043	(0.036)
≥ 2	0.067**	(0.034)
Is employed	-0.034	(0.026)
Has Abitur	0.028	(0.021)
Is owner	0.036	(0.030)
Size of the dwelling:		
Small: $<$ 1st tercile	-0.115***	(0.030)
Large: $>$ 2nd tercile	0.159***	(0.026)
Type of the dwelling:		
Row house	-0.160***	(0.024)
Multi-family dwelling	-0.239***	(0.048)
# apartments:		
$4-6$	-0.162***	(0.050)
$7-12$	-0.185***	(0.050)
≥ 13	-0.254***	(0.064)
Construction year:		
≤ 1918	-0.023	(0.045)
$1919-1948$	-0.115***	(0.042)
$1949-1957$	-0.088*	(0.046)
$1958-1968$	-0.085**	(0.040)
$1978-1983$	0.082**	(0.040)
$1984-1994$	0.021	(0.042)
$1995-2001$	-0.022	(0.043)
≥ 2002	-0.127**	(0.051)

Table A.4 continued from previous page

Type of heating system:		
gas heating on the floor	0.054**	(0.025)
oven	0.167**	(0.080)
Has modernised:		
Thermal shell	0.045	(0.030)
Windows	0.018	(0.025)
Heating system	0.052**	(0.024)
log(HDD)	0.066	(0.098)
Fuel type:		
Oil	-0.005	(0.026)
Long distance heat	-0.131***	(0.050)
Constant	4.637***	(0.834)
$\frac{1}{1+\hat{\beta}_{c_{i,t}}}$	2.769***	(0.289)
Observations	2,256	
R-squared	0.581	
Estimator	OLS	

[a] The table reports a robustness check of the preferred specification, reported in column (2) of table (1.2), towards the use of an alternative dependent variable, $\tilde{\tau}^d_{i,t}$, that is obtained if the term $Q_A \cdot \xi$ is not dropped during the derivation of the input demand equation (1.9) in section 1.3.2. The dependent variable is defined as

$$\tilde{\tau}^d_{i,t} = \frac{F^d_{i,t} + Q_A \cdot \xi}{s^d_{i,t} m_{i,t}}$$

The estimates are obtained by ordinary least squares regression. The regression includes year fixed effects and the standard errors are clustered on the household level and reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

Table A.5: Descriptive statistics over different variable restrictions for the year 2008 [a]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
∅[b] F^d	78.75	67.47	135.03	130.45	170.53	170.62	172.21	164.36	164.36	164.55
∅ F^e [c]	314.66	316.14	304.74	302.85	304.56	303.42	300.68	304.96	304.96	308.55
∅ Household Age	44.93	45.66	48.87	49.48	50.30	50.39	50.75	50.77	50.77	51.19
∅ Household Size	2.49	2.47	2.62	2.56	2.59	2.59	2.62	2.60	2.60	2.49
Share of households with Abitur	0.51	0.51	0.53	0.53	0.53	0.53	0.52	0.53	0.53	0.53
Share of households with university degree	0.29	0.30	0.33	0.34	0.34	0.34	0.34	0.34	0.34	0.35
Share of households employed	0.77	0.76	0.73	0.71	0.69	0.68	0.68	0.68	0.68	0.66
∅ Income[d]	2557.29	2592.33	2801.33	2805.95	2858.45	2859.26	2887.41	2883.12	2883.12	2841.88
Share of owners	0.54	0.54	0.69	0.68	0.69	0.70	0.71	0.70	0.70	0.69
∅ Size of dwelling (in m^2)	107.34	106.77	120.14	116.75	118.29	118.42	119.93	119.19	119.19	117.95
Share of households living in multi-family dwellings	0.41	0.43	0.29	0.32	0.28	0.28	0.27	0.27	0.27	0.28
Number of households observed	6714	5607	2371	2089	1221	1202	1145	1098	1098	1016
Excluding households										
using other fuel types than gas, oil or long distance heat	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
unable to report their fuel consumption[e]	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
estimating their oil consumption	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
w/o data on some fuel types used in their dwelling[f]	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
living in "non-regular" dwelling types[g]	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
with outliers in F^e	No	No	No	No	No	No	Yes	No	Yes	Yes
with outliers in τ^d	No	No	No	No	No	No	No	Yes	Yes	Yes
with grown up children	No	No	No	No	No	No	No	No	No	Yes

[a] The table reports descriptive statistics over different variable restrictions for the year 2008. Column (1) represents the unconstrained representative panel of households provided by RWI and forsa (2016). Subsequent columns consecutively add the restrictions on the data. Column (10) represents the dataset used in the final analysis.
[b] The diameter symbol (∅) indicates the arithmetic mean over all observations in the sample of the quantity that follows it.
[c] F^e denotes the fuel consumption predicted by the engineering model: $F^e = I^e(\psi^e(\boldsymbol{D^e}))$.
[d] The dataset only contains categorical income information. For the calculation of mean income values in monetary terms, the mean of the lower and upper bound defining each income category is used. Since income cannot be expected to be uniformly distributed within each income category, the resulting calculated mean is unlikely to represent the true mean in the sample.
[e] That is, households that are unable to report the amount of fuel used for the primary heating system in the dwelling.
[f] That is, households that are unable to report at least one of the fuel types they use for heating their rooms (including additional heating systems like ovens, etc.).
[g] Excluded are households that live in old industrial dwellings, cottages or some unknown or “other” dwelling type.

A.2 Theorems

Theorem 1 1. *Suppose the amount of fuel, F, a household has to use to produce an increase of the mean indoor temperature by τ degrees, given a state variable of the dwellings overall efficiency, s, can be represented by an input demand function*

$$F = I(s, m, \tau).$$

Let $\varepsilon_{F,s}$ and $\varepsilon_{\tau,s}$ denote the elasticity of fuel and temperature consumption with respect to the variable s, respectively. Then the following relationship holds if and only if $I(s, m, \tau)$ is linear in τ:

$$\varepsilon_{F,s} = 1 + \varepsilon_{\tau,s}.$$

Proof.

$$\begin{aligned}
& \varepsilon_{F,s} = 1 + \varepsilon_{\tau,s} \\
\Leftrightarrow \quad & \frac{\partial ln(F)}{\partial ln(s)} = 1 + \frac{\partial ln(\tau)}{\partial ln(s)} && \mid \text{Integrating both sides} \\
\Leftrightarrow \quad & ln(F) = ln(s) + ln(\tau) + k && \text{, for some } k \in \mathbb{R} \\
\Leftrightarrow \quad & F = s \cdot \tau \cdot e^k && \mid \text{Define: } m \equiv e^k \\
\Leftrightarrow \quad & F = s \cdot m \cdot \tau
\end{aligned}$$

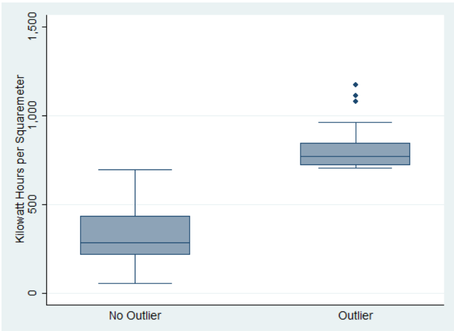
□

A.3 Details on the generation of efficiency states

A.3.1 Detailed descriptions of the outlier corrections applied to the generated efficiency states

The range of the distribution is affected by two types of outlier corrections that have been conducted to avoid that regression results are driven by extremely large or small values, which might indicate very specific types of dwellings or errors in the calculation procedure.[1] First, fuel requirements outside an interval of three standard deviations around the mean value are dropped in an iterative procedure. After every iteration the procedure is repeated until no values are excluded anymore. The choice of an interval of three standard deviations around the mean is a conservative choice made to ensure that no plausible values are dropped.[2] Figure A.1 visualises the distribution of the observations that are dropped on in the outlier correction. All dropped

Figure A.1: Boxplot of outliers in predicted fuel requirements



observations are at the right tail of the fuel requirement distribution and thus represent extremely inefficient dwellings. Column (7) of table A.5 clarifies that the outlier correction implies no selection on household or dwelling characteristics.

In a second step also the relation of the predicted fuel requirements to observed fuel consumption is used to identify implausible observations. For this purpose, I generate the empirical indoor temperatures, $\tau^d_{i,t}$ and repeat the same outlier identification procedure described before on them. Since I need plausible consumption data against which I can compare the generated fuel requirements, the correction is conducted under the same restrictions on reported fuel consumption used in the final analysis. Figure A.2 clarifies that the procedure only drops observations with

[1] In addition, also observations with unclear statements on crucial input variables are dropped, including households who do not know the construction year, number of apartments and type of their dwelling or whether they own the dwelling or not.

[2] If the true fuel requirements follow a normal distribution 99,73 % of all values that belong to that distribution should lie within the considered interval.

a very *large* implied indoor temperature. Large implied indoor temperatures result from fuel requirements that are *small* relative to the observed fuel consumption. In contrast to the sorting out of very large fuel requirements in the previous correction procedure, observations are thus now dropped because the fuel requirements are very small or the reported fuel consumption large respectively. Figure A.3 illustrates the idiosyncracy of the variables dropped in this procedure. In contrast to the usual pattern in the data, the actual fuel consumption tends to be larger than the predicted one. The outlier corrections on the mere fuel consumption and fuel requirements respectively therefore cannot identify these observations as special in the data. The sorting out of observations based on the level of the implied indoor temperature is therefore complementary to these alternative procedures.

Figure A.4 emphasizes that the correction procedure eliminates observations that are intuitively problematic. All dropped observations have implied temperature levels above 21 degrees Celsius, which is typically considered an ideal temperature level, e.g., in engineering calculations. In the figure these observations appear as having a negative distance to the ideal temperature level. At the same time, the observations that remain after outliers have been dropped only rarely exceed the ideal temperature level and if they do only slightly. I therefore add the correction procedure to my constraints on the dataset. Column (8) of table A.5 indicates that the outlier correction reduces the size of the dataset available for the year 2008 by 104 observations. The exclusion implies that the mean of actual fuel consumption decreases and the mean of predicted fuel consumption increases slightly. There is no visible selection on household or dwelling characteristics. A robustness check of the preferred regression equation reported in column (6) of table A.3 clarifies, that the outlier correction has no impact on the qualitative results either.

Figure A.2: Boxplot of implied indoor temperatures

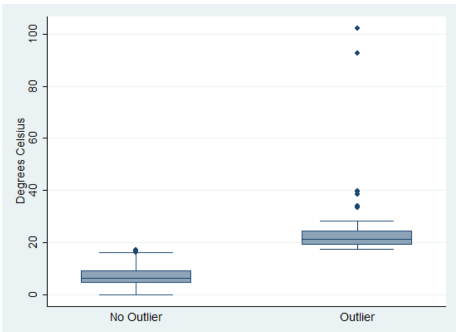


Figure A.3: Boxplot of fuel consumption and predicted fuel requirements

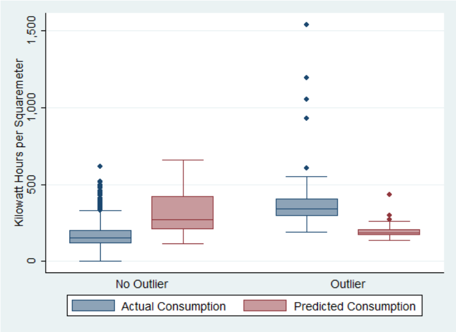
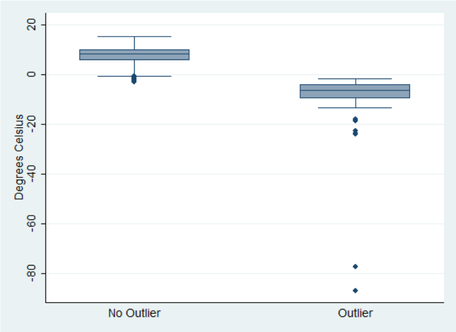


Figure A.4: Boxplot of distances to the ideal temperature level

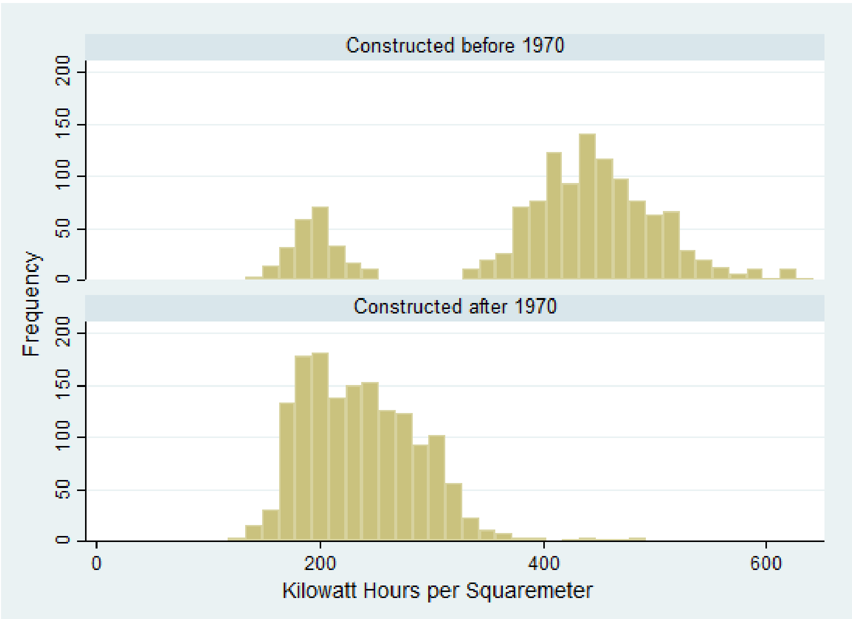


A.3.2 Detailed analyses of the generated fuel requirements

Figure 1.1 illustrates, that the distribution of generated fuel requirements is bimodal with a large density of fuel requirements around 250 and 500 $^{kWh}/_{m^2}$ and very few predictions of fuel consumptions around 400 $^{kWh}/_{m^2}$. Regression results of the observed dwelling characteristics on the generated fuel requirements reported in column (3) of table A.6, provide an impression how the available data is mapped into efficiency states by applying the calculation procedure of Loga et al. (2005). Clearly, the construction year is the most important predictor of the dwellings' efficiency. More recent dwellings require less fuel to be heated to a standardised indoor temperature than older dwellings. A dramatic shift occurs for dwellings constructed after 1969, which are substantially more efficient than those constructed earlier. As figure A.5 illustrates, this discrete shift in the calculation procedure also produces the bimodal shape of the efficiency state distribution discussed before. Dwellings constructed after 1969 have efficiency states between 200 and 400 $^{kWh}/_{m^2}$. In contrast, the fuel requirements of dwellings constructed before are primarily located in the range between 400 and 800 $^{kWh}/_{m^2}$. Only relatively few older dwellings reach efficiency levels in the left part of the distribution by modernisation investments.

The regression results also indicate that larger dwellings require less fuel per square metre than smaller ones. While the existence of more than one apartment in the dwelling reduces efficiency, semi-detached and multi-family dwellings have significantly lower predicted fuel consumptions

Figure A.5: Generated fuel requirements of dwellings constructed before and after 1969



than detached singly-family dwellings. Modernisation investments significantly increase the overall efficiency, of which the impact of investments into thermal insulation is the largest. The impact into new heating systems and windows is substantially smaller. Investing the former leads to larger efficiency reductions than investments in the latter. Finally, the results also indicate that central heating systems and long distance heating are less efficient than other heating types. Columns (1) and (2) of table A.6 again clarify that the data restrictions have no substantial impact on the qualitative results.

Table A.6: Coefficients of regression of predicted fuel consumption on dwelling characteristics [a]

Dep. Var: $F_{i,t}^{e}$	(1)	(2)	(3)
Dwelling Size im m^2	-0.348***	-0.339***	-0.549***
	(0.107)	(0.108)	(0.057)
Construction year:			
≤ 1918	136.508***	136.377***	141.793***
	(4.249)	(4.253)	(8.507)
$1919 - 1948$	145.406***	143.360***	143.696***
	(4.160)	(3.955)	(6.885)
$1949 - 1957$	122.637***	123.624***	115.496***
	(4.175)	(4.186)	(8.706)
$1958 - 1968$	112.155***	111.391***	112.375***
	(3.136)	(3.093)	(6.255)
$1978 - 1983$	-37.360***	-37.305***	-31.164***
	(2.390)	(2.397)	(4.971)
$1984 - 1994$	-66.874***	-66.902***	-60.525***
	(2.556)	(2.541)	(4.625)
$1995 - 2001$	-108.437***	-108.556***	-102.567***
	(2.505)	(2.464)	(4.503)
≥ 2002	-116.900***	-117.350***	-115.466***
	(3.271)	(3.254)	(5.813)
# apartments:			
2	8.269***	9.573***	10.437**
	(2.657)	(2.675)	(5.246)
3	0.124	-1.922	-3.051
	(4.924)	(4.573)	(10.215)
$4 - 6$	8.642	5.203	-2.195
	(5.883)	(5.640)	(9.589)
$7 - 12$	9.315	7.015	0.428
	(6.119)	(5.960)	(11.349)
$13 - 20$	12.071	7.852	0.458
	(7.541)	(7.171)	(14.972)
≥ 20	23.439***	18.931***	17.289

Table A.6 continued from previous page

	(7.536)	(7.135)	(12.848)
Type of the dwelling:			
Row house	-32.193***	-31.768***	-35.898***
	(2.636)	(2.595)	(4.120)
Multi-family dwelling	-30.532***	-27.252***	-32.964***
	(5.138)	(4.974)	(9.928)
Has modernised:			
Thermal shell	-97.185***	-95.390***	-96.114***
	(3.052)	(3.064)	(6.230)
Windows	-12.593***	-12.218***	-7.985
	(2.970)	(2.909)	(6.090)
Heating system	-25.285***	-25.158***	-29.434***
	(2.738)	(2.678)	(5.060)
Type of Heating System:			
Gas heating on the floor	-26.041***	-24.409***	-22.744***
	(2.394)	(2.365)	(4.658)
Oven	6.707	4.975	-15.247
	(5.218)	(5.070)	(14.665)
Electric Heating	-73.907***	-73.784***	
	(4.534)	(4.538)	
Heat Pump	-146.517***	-144.940***	
	(6.363)	(6.631)	
Fueltype:			
Liquid Gas	0.472	1.025	
	(6.015)	(6.094)	
Oil	5.389**	7.010***	9.854**
	(2.392)	(2.370)	(4.560)
Wood Pellets	29.620**	43.720***	
	(11.795)	(12.033)	
Logs	-1.564	1.755	
	(8.170)	(8.126)	
Wood Chips	-8.389	9.512	
	(67.830)	(99.062)	
Wood Briquettes	29.846*	23.103*	
	(17.569)	(13.459)	
Coal	-4.816	-2.719	
	(23.475)	(25.439)	
Lignite	41.039***	40.839***	
	(14.622)	(14.277)	
Long Distance Heat	-40.397***	-40.090***	-47.948***
	(3.348)	(3.031)	(6.566)

Table A.6 continued from previous page

Constant	380.252^{***}	378.016^{***}	401.312^{***}
	(14.643)	(14.695)	(9.972)
Observations	15,585	15,181	2,784
R-squared	0.735	0.742	0.739
Drop Outliers	No	Yes	Yes
Impose restrictions on fuel data	No	No	Yes

[a] The table reports estimation results from ordinariy least squares (OLS) regressions of fuel requirements, predicted by the engineering model from Loga et al. (2005), on dwelling characteristics. Column (1) reports results from regressions on the full sample. In column (2) the sample is restricted to predicted fuel requirements that are not marked as outliers in the outlier correction procedure. In column (3) additional restrictions used in the preferred specification in the main analysis are added. In particular, only households that are able to report their energy consumption and use natural gas, oil or long distance heat for their primary heating system are considered. Standard errors are clustered on the household level and reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

B Appendix to chapter 2

B.1 Robustness analyses

Table B.1: Regression coefficients for different calibrations of $\overline{\tau}^{in}$ [a]

Dep. Var: $F^d_{i,t}$	(1)	(2)	(3)	(4)
Estimates of households' utility function parameters:				
Constant: β_0^τ	23.503***	16.537***	9.442***	7.494***
	(7.225)	(4.931)	(2.714)	(2.129)
age:				
$30-39$	2.914	1.843	0.915	0.692
	(6.186)	(4.234)	(2.336)	(1.834)
$40-49$	0.409	0.189	0.048	0.024
	(6.257)	(4.285)	(2.366)	(1.858)
$50-59$	8.071	5.443	2.956	2.309
	(6.768)	(4.623)	(2.546)	(1.998)
≥ 60	10.088	6.734	3.607	2.805
	(7.433)	(5.060)	(2.775)	(2.174)
# adults:				
2	6.124**	4.162**	2.284**	1.791**
	(2.466)	(1.679)	(0.922)	(0.723)
3	13.496**	9.184**	5.044**	3.955**
	(5.848)	(3.943)	(2.142)	(1.673)
≥ 4	7.801*	5.311*	2.920*	2.291*
	(4.140)	(2.817)	(1.547)	(1.213)
# children:				
1	-0.946	-0.569	-0.252	-0.180
	(2.674)	(1.829)	(1.010)	(0.793)
≥ 2	5.997*	4.087*	2.248*	1.764*
	(3.467)	(2.367)	(1.304)	(1.024)
Is employed	-4.624	-3.111	-1.681	-1.310
	(4.337)	(2.928)	(1.591)	(1.242)

Table B.1 continued from previous page

Has Abitur	6.764***	4.552***	2.463**	1.921**
	(2.565)	(1.743)	(0.956)	(0.749)
Is owner	-0.124	-0.065	-0.015	-0.004
	(2.717)	(1.837)	(0.999)	(0.781)
Income:				
$< 1,500$ €	4.937	3.279	1.738	1.345
	(4.361)	(2.970)	(1.629)	(1.277)
$\geq 3,500$ €	0.458	0.387	0.265	0.221
	(2.592)	(1.768)	(0.973)	(0.763)
Size of the dwelling:				
Small: $<$ 1st tercile	-1.568	-1.332	-0.914	-0.764
	(3.046)	(2.029)	(1.083)	(0.842)
Large: $>$ 2nd tercile	9.847***	7.074***	4.147***	3.321***
	(2.818)	(1.914)	(1.048)	(0.821)
Estimates of adjustment term:				
Constant: β_0^λ	0.898***	0.758***	0.577***	0.515***
	(0.055)	(0.047)	(0.036)	(0.032)
Type of the dwelling:				
Row house	-0.258***	-0.219***	-0.169***	-0.151***
	(0.034)	(0.029)	(0.022)	(0.020)
Multi-family dwelling	-0.311***	-0.265***	-0.204***	-0.183***
	(0.051)	(0.044)	(0.033)	(0.030)
# apartments:				
$4-6$	-0.074*	-0.062*	-0.047*	-0.042*
	(0.042)	(0.035)	(0.027)	(0.024)
$7-12$	-0.112**	-0.094**	-0.071**	-0.064**
	(0.044)	(0.037)	(0.028)	(0.025)
≥ 13	-0.130***	-0.109***	-0.083***	-0.074***
	(0.047)	(0.039)	(0.030)	(0.027)
Construction year:				
$1919-1968$	-0.019	-0.018	-0.016	-0.015
	(0.038)	(0.032)	(0.025)	(0.022)
$1969-1977$	0.157***	0.130***	0.097***	0.086***
	(0.051)	(0.043)	(0.033)	(0.029)
$1978-1994$	0.306***	0.262***	0.204***	0.184***
	(0.050)	(0.043)	(0.033)	(0.030)
≥ 1995	0.192***	0.175***	0.147***	0.136***
	(0.047)	(0.040)	(0.031)	(0.028)

Table B.1 continued from previous page

Has modernised:				
Windows	0.040	0.035	0.028	0.025
	(0.035)	(0.029)	(0.023)	(0.020)
Heating system	0.014	0.011	0.007	0.006
	(0.032)	(0.027)	(0.021)	(0.019)
Thermal shell	0.125***	0.113***	0.093***	0.085***
	(0.037)	(0.032)	(0.025)	(0.023)
log(HDD)	0.251*	0.214*	0.166*	0.149*
	(0.131)	(0.111)	(0.085)	(0.076)
Income:				
$< 1,500$ €	-0.034	-0.029	-0.021	-0.019
	(0.028)	(0.024)	(0.018)	(0.016)
$\geq 3,500$ €	0.039	0.033	0.025	0.023
	(0.029)	(0.025)	(0.019)	(0.017)
Size of the dwelling:				
Small: $<$ 1st tercile	-0.047	-0.039	-0.029	-0.025
	(0.037)	(0.032)	(0.024)	(0.022)
Large: $>$ 2nd tercile	0.002	0.001	-0.000	-0.001
	(0.034)	(0.029)	(0.022)	(0.020)
$\overline{\tau}^{in}$	17	19	23	25
Observations	2,256	2,256	2,256	2,256
R-squared	0.893	0.893	0.892	0.892

[a] The table reports estimates of equation (2.19) for different ideal temperature levels, $\overline{\tau}^{in}_{i,t}$. Year and fuel type fixed effects are included in the adjustment term in all regressions. All results are obtained by nonlinear least squares. Standard errors are reported in parantheses and clustered on the household level. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by by three (***), two (**) and one (*) asterisks, respectively.

Table B.2: Model Predictions for different calibrations of $\overline{\tau}^{in}$ [a]

	(1)	(2)	(3)	(4)
$\varnothing\ \hat{\tau}^{*}_{i,t}$	8.482***	10.027***	13.117***	14.662***
	(0.111)	(0.132)	(0.172)	(0.193)
$\varnothing\ \hat{\tau}^{in^*}_{i,t}$	14.906***	16.451***	19.541***	21.086***
	(0.111)	(0.132)	(0.172)	(0.193)
$\varnothing\ \hat{\tau}^{d}_{i,t}$	6.832***	6.831***	6.828***	6.827***
	(0.078)	(0.078)	(0.079)	(0.079)
$\varnothing\ \hat{\lambda}_{i,t}$	0.818***	0.692***	0.530***	0.474***
	(0.016)	(0.014)	(0.010)	(0.009)
$\varnothing\ \varepsilon_{\tau^*_{i,t},c_{i,t}}$	-0.288***	-0.296***	-0.306***	-0.310***
	(0.022)	(0.023)	(0.023)	(0.023)
$\varepsilon_{\tau^*,c_{i,t}}(\overline{\boldsymbol{x}}_{\boldsymbol{i,t}})$	-0.245***	-0.253***	-0.264***	-0.268***
	(0.018)	(0.018)	(0.018)	(0.018)
$\overline{\tau}^{in}$	17	19	23	25

[a] Columns (1) to (4) of the table summarize predictions of model quantities based on estimates reported in the respective columns of table B.1. The diameter symbol ($\varnothing$) indicates the arithmetic mean over all households, i, and time periods, t, of the quantity that follows it. The predicted quantities are households' optimal temperature increase, $\hat{\tau}^{*}_{i,t}$, the resulting indoor temperature, $\hat{\tau}^{in^*}_{i,t}$, the temperature increase implied in the observed fuel data given the input demand function of equation (2.5), $\hat{\tau}^{d}_{i,t}$, the adjustment factor, $\hat{\lambda}_{i,t} = \lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \hat{\boldsymbol{\beta}}^{\boldsymbol{\lambda}})$, and the elasticity with respect to changes in the marginal cost of temperature consumption, $\varepsilon_{\tau^*_{i,t},c_{i,t}}$. In addition, the last row reports the elasticity of a hypothetical household with the mean characteristics in the sample stored in the vector $\overline{\boldsymbol{x}}_{\boldsymbol{i,t}}$. Standard errors are reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

Table B.3: Model Predictions for regressions under different data restrictions [a]

	(1)	(2)	(3)
$\varnothing\ \hat{\tau}^*_{i,t}$	11.649***	11.268***	9.750***
	(0.167)	(0.162)	(0.317)
$\varnothing\ \hat{\tau}^{in*}_{i,t}$	18.007***	17.607***	16.364***
	(0.167)	(0.162)	(0.317)
$\varnothing\ \hat{\tau}^d_{i,t}$	6.993***	7.654***	5.239***
	(0.099)	(0.102)	(0.119)
$\varnothing\ \hat{\lambda}_{i,t}$	0.611***	0.684***	0.584***
	(0.014)	(0.015)	(0.020)
$\varnothing\ \varepsilon_{\tau^*_{i,t},c_{i,t}}$	-0.307***	-0.353***	-0.915
	(0.025)	(0.025)	(0.642)
$\varepsilon_{\tau^*,c_{i,t}}(\overline{\boldsymbol{x}}_{\boldsymbol{i,t}})$	-0.254***	-0.290***	-0.454***
	(0.021)	(0.019)	(0.046)
Data Restr.	Only HNG	Only Owner	Only Tenants

[a] Columns (1) to (3) of the table summarize predictions of model quantities based on estimates reported in the respective columns of table B.4. The diameter symbol ($\varnothing$) indicates the arithmetic mean over all households, i, and time periods, t, of the quantity that follows it. The predicted quantities are households' optimal temperature increase, $\hat{\tau}^*_{i,t}$, the resulting indoor temperature, $\hat{\tau}^{in*}_{i,t}$, the temperature increase implied in the observed fuel data given the input demand function of equation (2.5), $\hat{\tau}^d_{i,t}$, the adjustment factor, $\hat{\lambda}_{i,t} = \lambda(\boldsymbol{x}^{\boldsymbol{\lambda}}_{\boldsymbol{i,t}}, \hat{\boldsymbol{\beta}}^{\boldsymbol{\lambda}})$, and the elasticity with respect to changes in the marginal cost of temperature consumption, $\varepsilon_{\tau^*_{i,t},c_{i,t}}$. In addition, the last row reports the elasticity of a hypothetical household with the mean characteristics in the sample stored in the vector $\overline{\boldsymbol{x}}_{\boldsymbol{i,t}}$. Standard errors are reported in parantheses. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

Table B.4: Regression coefficients obtained under different data restrictions [a]

Dep. Var: $F^d_{i,t}$	(1)	(2)	(3)
Estimates of households' utility function parameters:			
Constant: β^τ_0	11.173***	13.210**	13.916***
	(3.487)	(6.378)	(1.455)
age:			
$30-39$	-1.199	0.456	0.366
	(2.218)	(6.011)	(0.808)
$40-49$	-1.081	-1.675	0.347
	(2.296)	(6.077)	(0.680)
$50-59$	3.370	2.562	-0.273
	(2.601)	(6.042)	(0.704)
≥ 60	5.759*	3.272	2.129**
	(3.432)	(6.248)	(0.875)
# adults:			
2	4.522***	3.375***	1.836***
	(1.362)	(1.208)	(0.486)
3	3.138	6.995***	0.410
	(2.114)	(2.692)	(0.586)
≥ 4	6.891**	5.545**	1.046
	(2.750)	(2.153)	(1.675)
# children:			
1	1.352	-0.720	0.384
	(1.676)	(1.452)	(0.678)
≥ 2	4.918**	3.193*	-0.215
	(2.106)	(1.825)	(1.852)
Is employed	-2.026	-2.295	0.753
	(2.773)	(2.205)	(0.478)
Has Abitur	3.534**	3.725***	-1.140**
	(1.576)	(1.379)	(0.527)
Is owner	0.339		
	(1.428)		
Income:			
$< 1,500$ €	5.245	3.559*	-0.369
	(3.790)	(2.067)	(0.526)
$\geq 3,500$ €	-0.381	1.529	0.389
	(1.453)	(1.352)	(1.299)

Table B.4 continued from previous page

Size of the dwelling:			
Small: < 1st tercile	-1.576	-1.704	-6.396***
	(1.332)	(1.388)	(0.890)
Large: > 2nd tercile	4.715***	4.466***	11.582***
	(1.733)	(1.454)	(4.254)
Estimates of adjustment term:			
Constant: β_0^λ	0.691***	0.691***	0.506***
	(0.048)	(0.055)	(0.088)
Type of the dwelling:			
Row house	-0.203***	-0.199***	-0.092
	(0.032)	(0.027)	(0.077)
Multi-family dwelling	-0.292***	-0.266***	-0.076
	(0.050)	(0.054)	(0.112)
# apartments:			
$4-6$	-0.033	-0.035	-0.114
	(0.041)	(0.052)	(0.083)
$7-12$	-0.056	-0.048	-0.138
	(0.039)	(0.061)	(0.089)
≥ 13	-0.069	-0.044	0.645***
	(0.043)	(0.067)	(0.146)
Construction year:			
$1919-1968$	-0.031	-0.011	0.001
	(0.029)	(0.046)	(0.044)
$1969-1977$	0.045	0.093*	0.175**
	(0.048)	(0.051)	(0.074)
$1978-1994$	0.194***	0.210***	0.330***
	(0.044)	(0.050)	(0.085)
≥ 1995	0.112***	0.127***	0.256***
	(0.040)	(0.047)	(0.064)
Has modernised:			
Windows	0.029	0.022	0.038
	(0.031)	(0.033)	(0.047)
Heating system	0.018	-0.009	0.135**
	(0.028)	(0.031)	(0.055)
Thermal shell	0.075**	0.115***	-0.023
	(0.034)	(0.035)	(0.044)
log(HDD)	0.170	0.275**	-0.006
	(0.116)	(0.127)	(0.184)

Table B.4 continued from previous page

Income:			
$< 1,500$ €	-0.049**	0.000	-0.087**
	(0.024)	(0.044)	(0.037)
$\geq 3,500$ €	0.036	0.023	-0.019
	(0.029)	(0.027)	(0.040)
Size of the dwelling:			
Small: $<$ 1st tercile	-0.020	-0.038	0.052
	(0.031)	(0.033)	(0.050)
Large: $>$ 2nd tercile	0.031	-0.004	0.047
	(0.032)	(0.027)	(0.063)
Observations	1,498	1,564	692
R-squared	0.898	0.893	0.879
Data Restr.	Only HNG	Only Owner	Only Tenants

[a] The table reports estimates of equation (2.19) that have been obtained under different data restrictions. In column (1) only households heating with natural gas (HNG) have been included in the analysis. In column (2) and (3) the analysis has focussed on owners and tenants respectively. Year and fuel type fixed effects are included in the adjustment term in all regressions. The ideal temperature level, $\overline{\tau}^{in}_{i,t}$, is set to 21 degrees Celsius. All results are obtained by nonlinear least squares. Standard errors are reported in parantheses and clustered on the household level. Statistical significance at a 1 %, 5 % and 10 % level of confidence is indicated by three (***), two (**) and one (*) asterisks, respectively.

C Appendix to chapter 3

C.1 Descriptives of actual monetary cost of retrofitting in Germany

We use data from the 34th version of "The German Socioeconomic Panel Study" (SOEP v34) to gain insights on the monetary cost associated to actual modernisations conducted by German home owners.

The SOEP is the largest and oldest multi-disciplinary household panel in Germany (Goebel et al., 2019). It contains questions regarding households' modernisation investment since its start in 1984. In the time period from 2010 to 2015 the surveyed households have explicitly been asked for the monetary cost they encountered when modernising their dwellings.[1] This provides an opportunity to study the monetary cost related to modernisation investments directly and separated from additional non-monetary cost that households might encounter.

The modernisations relevant in the context of our research include investments into the thermal shell, windows or heating systems of a dwelling, which are able to increase the overall energy efficiency of the dwelling. The survey questions asking for the respective investments are very similar in the SOEP and the "Residential Energy Consumption Survey". This makes us confident, that the monetary cost associated to similar types of household decisions as studied in our dynamic model can be captured form the SOEP data.

We deflate the available cost data to 2007 levels, to make them comparable to the time period considered in our empirical analysis.[2] Table C.1 provides summary statistics of the monetary cost encountered by households that modernised their dwelling. Panel A of the table focuses on the cost that were reported by households that invested into energy efficiency improvements only. It indicates that the mean cost of investing into heating systems and thermal insulation are at a similar order of magnitude, even though the distribution of cost associated to the latter

[1] Home owners were asked for general maintenance cost during most years since the start of the survey. Only in the time period between 2010 and 2015 an additional question explicitly asking for the cost of conducted modernisations has been added.

[2] We used inflation rates on consumer prices for the maintenance and repair of dwellings (classification code CC13-043) obtained from the German statistical agency (Statistisches Bundesamt, 2019) to deflate the modernisation cost reported by home owners to 2007 levels.

Table C.1: Descriptives of monetary modernisation costs of home owners based on SOEP data between 2010 and 2015 [a]

	Obs.	Mean	Stand. Dev.	Percentiles				
				5th	25th	50th	75th	95th
PANEL A								
Single investment into								
heating system	408	8,306.42	6,737.88	1,638.00	4,101.85	5,984.76	10,098.19	22,207.16
windows	937	4,762.79	5,647.17	532.97	1,522.93	2,866.50	6,391.30	14,608.70
thermal insulation	536	7,970.56	11,852.68	327.60	1,369.57	3,652.17	9,130.43	30,173.68
Combined investment into								
heating system and windows	38	12,287.12	11,255.01	1,724.21	6,391.30	8,924.38	12,931.58	50,040.73
heating system and thermal insulation	28	18,430.77	18,869.38	1,826.09	5,880.30	10,399.95	23,923.23	62,550.92
Total	1,947	6,731.87	8,728.50	499.18	1,724.21	4,095.00	8,340.12	22,207.16
PANEL B								
Any investment into								
energy efficiency	1,947	6,731.87	8,728.50	499.18	1,724.21	4,095.00	8,340.12	22,207.16
other modernisations	3,664	8,589.41	12,118.53	560.37	2,283.50	5,076.45	10,345.26	25,863.16
energy efficiency and other modernisations	1,347	29,251.37	41,710.49	2,001.63	6,847.83	15,846.23	33,360.49	115,477.23
Total	6,958	12,069.58	22,511.01	655.20	2,502.04	5,733.00	12,691.12	41,700.61

[a] The table reports descriptive statistics of the monetary cost associated to actual modernisations conducted by German home owners between 2010 and 2015. The data is obtained from the 34th version of "The German Socioeconomic Panel Study" (SOEP v34). The reported cost have been deflated to 2007 levels using data from the German statistical agency (Statistisches Bundesamt, 2019). Panel A of the table reports statistics of the cost encountered by households that have invested into measures to increase the energetic efficiency of the dwelling only. These can be investments into the heating system, the windows or the thermal shell of the dwelling. Panel B of the table compares the cost of investing into any of the efficiency measures and of investing into some non-efficiency measures only, to the cost that arise if the two types of modernisations are conducted jointly within a year.

is wider. Investments into windows are cheaper than the other two types of energy efficiency modernisations. Combined investments are rare, but occur for investments into heating systems and windows as well as heating systems and thermal insulation.

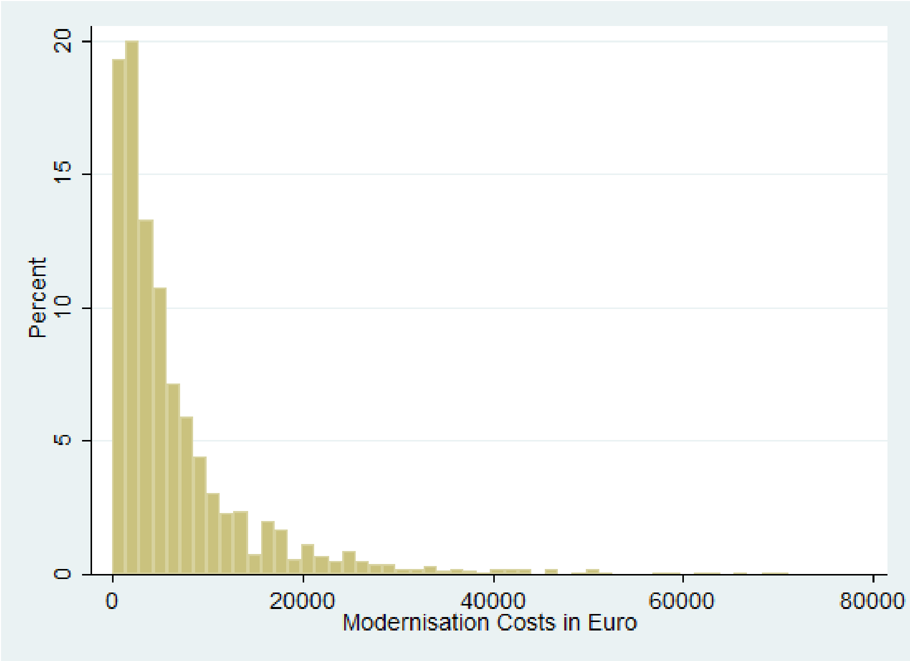
The primary information of interest in the context of our empirical analysis is the mean investment cost over all types of efficiency modernisations that households can conduct. The last row of Panel A reports their mean to be 6731.87 euros. There is quite some variation in the cost households face, resulting from the variety of investments that can be conducted. Figure C.1 plots the distribution of the modernisation cost of interest. It is left-skewed implying a lower probability mass at very large cost. Consistent with this the median of the distribution is $4,095$. The 75th and 95th percentiles are $8,340.122$ and $22,207,16$ euros, respectively.

A limitation of the cost data available from the SOEP is that it does not contain separate reports by modernisation type. In case that households invest in energy efficiency and other types of modernisation, such as a new kitchen or bath, within the same year, the associated cost cannot be distinguished. This is a problem for the assessment of the monetary cost related to energy efficiency retrofits, if particularly large, and therefore expensive, activities are more likely to be conducted jointly with other modernisation measures. The focus on cost reported from households that have only invested in energy efficiency of the dwelling (taken in Panel A of table C.1) would then imply that particularly expensive modernisations are undersampled in the data used to calculate mean values and other statistics.

Panel B of table C.1 considers this potential problem more closely. It provides summary statistics of modernisation cost separated by investments into energy efficiency measures, other modernisations or both. It is easy to see that the sum of the mean cost of individual investments into energy efficiency and other modernisations respectively, is substantially below the mean cost if both investment types are conducted jointly. The difference of the two values is $13,930.09$ euros, indicating that expensive retrofits are likely substantially undersampled in Panel A.

To gauge the size of the resulting bias, we assume that the relative magnitude of the cost associated to efficiency related investments and other investments is the same, whether the measures are conducted independently or jointly within a year. The cost for efficiency retrofits are a bit smaller, being responsible for 43.94 % of the combined cost of individual investments. This relationship is very stable over the full range of the distributions, supporting the main assumption of the analysis. If this cost relationship is fixed, the counterfactual modernisation cost associated to efficiency retrofits if other modernisations are conducted in the same year can be approximated to be $12,852.48$ euros. The weighted average of cost of energy efficiency retrofits when other modernisations are conducted in the same year or not can then be calculated to be $9,234.74$ euros. Accordingly, the mean modernisation cost when focussing on households that have only invested into energy efficiency would underestimate the cost over all investments by $2,502.87$ euros or 37.18 %. Obviously, this indicates a substantial bias in the mean cost reported in Panel A implying that the stated mean values should be interpreted with some caution.

Figure C.1: Distribution of monetary modernisation cost of home owners based on SOEP data between 2010 and 2015



Bibliography

Anderson, J. G. and Kushman, J. E. (1987). A Model of Household Heating Demand: Home Production with Satiety and an Endowment. *Journal of Consumer Affairs*, 21(1):1–20.

Aydin, E., Kok, N., and Brounen, D. (2017). Energy efficiency and household behavior: The rebound effect in the residential sector. *The RAND Journal of Economics*, 48(3):749–782.

Bach, S., Harnisch, M., and Isaak, N. (2018). Verteilungswirkungen der Energiepolitik–Personelle Einkommensverteilung. Retrieved from https://www.bmwi.de/Redaktion/DE/Publikationen/Studien/verteilungswirkungen-der-energiepolitiken.pdf?__blob=publicationFile&v=8.

Bakaloglou, S. and Charlier, D. (2018). The role of individual preferences to explain the energy performance gap. FAERE-French Association of Environmental and Resource Economists. Retrieved from http://faere.fr/pub/Conf2017/FAERE2017_Bakaloglou.pdf.

Bataille, C., Jaccard, M., Nyboer, J., and Rivers, N. (2006). Towards general equilibrium in a technology-rich model with empirically estimated behavioral parameters. *The Energy Journal*, (Special Issue# 2).

Becchio, C., Corgnati, S. P., Ballarini, I., and Corrado, V. (2012). Energy saving potential by retrofitting residential buildings in Europe. *REHVA Journal*, 6:34–38.

Bundesministerium für Wirtschaft und Energie (2018). Zahlen und Fakten Energiedaten: Nationale und Internationale Entwicklung. Data retrieved 8th February 2010, from https://www.bmwi.de/Redaktion/DE/Artikel/Energie/energiedaten-gesamtausgabe.html.

Bundesverband Des Schornsteinfegerhandwerks (2019). Erhebungen des Schornsteinfegerhandwerks für 2019. Retrieved from https://www.schornsteinfeger.de/erhebungen.aspx.

Cayla, J.-M., Maizi, N., and Marchand, C. (2011). The role of income in energy consumption behaviour: Evidence from French households data. *Energy Policy*, 39(12):7874–7883.

Cischinsky, H. and Diefenbach, N. (2018). Datenerhebung Wohngebäudebestand 2016: Datenerhebung zu den Energetischen Merkmalen und Modernisierungsraten im Deutschen und Hessischen Wohngebäudebestand. Retrieved from https://www.iwu.de/fileadmin/user_upload/dateien/gebaeudebestand/prj/Endbericht_Datenerhebung_Wohngeb%C3%A4udebestand_2016.pdf.

Dubin, J. A. and McFadden, D. L. (1983). A heating and cooling load model for single-family detached dwellings in energy survey data. Working Papers 469, California Institute of Technology, Division of the Humanities and Social Sciences. Retrieved from https://core.ac.uk/download/pdf/211364834.pdf.

Dubin, J. A., Miedema, A. K., and Chandran, R. V. (1986). Price Effects of Energy-Efficient Technologies: A Study of Residential Demand for Heating and Cooling. *The RAND Journal of Economics*, 17(3):310.

Estiri, H. (2015). A structural equation model of energy consumption in the united states: Untangling the complexity of per-capita residential energy use. *Energy research & social science*, 6:109–120.

European Commission (2020a). Climate strategies & targets: 2030 climate & energy framework. Retrieved on 4th December 2020 from https://ec.europa.eu/clima/policies/strategies/2030_en.

European Commission (2020b). Climate strategies & targets: 2050 long-term strategy. Retrieved on 4th December 2020 from https://ec.europa.eu/clima/policies/strategies/2050_en.

Eurostat (2020). Statistics Explained: Energy consumption in households. Retrieved 19th December 2020, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_consumption_in_households. Planned article update: June 2021.

Galvin, R. (2014). Estimating broad-brush rebound effects for household energy consumption in the EU 28 countries and Norway: Some policy implications of Odyssee data. *Energy Policy*, 73:323–332.

Gerarden, T. D., Newell, R. G., and Stavins, R. N. (2017). Assessing the Energy-Efficiency Gap. *Journal of Economic Literature*, 55(4):1486–1525.

German Institute for Standardization (2006). DIN EN ISO 7730:2006-05, Ergonomics of the thermal environment - Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria (ISO 7730:2005); German version EN ISO 7730:2005.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel Study (SOEP). *Journal of Economics and Statistics*, 239(2):345–360. Retrieved from: `https://doi.org/10.1515/jbnst-2018-0022`.

Gould, W. (2011). Use poisson rather than regress; tell a friend. *The Stata Blog*. Retrieved from `https://blog.stata.com/2011/08/22/use-poisson-rather-than-regress-tell-a-friend/`.

Greening, L. A., Greene, D. L., and Difiglio, C. (2000). Energy efficiency and consumption — the rebound effect — a survey. *Energy Policy*, 28(6-7):389–401.

Gronau, R., editor (1986). *Home production—a survey*, volume 1, 273-304 of *Handbook of Labor Economics*. Elsevier.

Gronau, R. (1997). The theory of home production: The past ten years. *Journal of Labor Economics*, 15(2):197–205.

Haas, R., Auer, H., and Biermayr, P. (1998). The impact of consumer behavior on residential energy demand for space heating. *Energy and Buildings*, 27(2):195–205.

Haas, R. and Biermayr, P. (2000). The rebound effect for space heating Empirical evidence from Austria. *Energy Policy*, 28(6-7):403–410.

Harrington, B. E., Heyman, B., Merleau-Ponty, N., Stockton, H., Ritchie, N., and Heyman, A. (2005). Keeping warm and staying well: findings from the qualitative arm of the Warm Homes Project. *Health & social care in the community*, 13(3):259–267.

Hausman, J. A. (1979). Individual discount rates and the purchase and utilization of energy-using durables. *The Bell Journal of Economics*, pages 33–54.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1):153.

Hirst, E. and Brown, M. (1990). Closing the efficiency gap: Barriers to the efficient use of energy. *Resources, Conservation and Recycling*, 3(4):267–281.

Hourcade, J.-C., Jaccard, M., Bataille, C., and Ghersi, F. (2006). Hybrid modeling: new answers to old challenges introduction to the special issue of the energy journal. *The Energy Journal*, (Special Issue# 2).

Hsueh, L.-M. and Gerner, J. L. (1993). Effect of Thermal Improvements in Housing on Residential Energy Demand. *Journal of Consumer Affairs*, 27(1):87–105.

Jaffe, A. B. and Stavins, R. N. (1994). The energy-efficiency gap What does it mean? *Energy Policy*, 22(10):804–810.

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., and Djurovic-Petrovic, M. (2010). A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment*, 45(7):1683–1697.

Knissel, J. and Loga, T. (2006). Vereinfachte Ermittlung von Primärenergiekennwerten. *Bauphysik*, 28(4):270–277.

Laurent, M.-H., Allibe, B., Oreszczyn, T., Hamilton, I., and Galvin, R. (2013). Back to reality: How domestic energy efficiency policies in four European countries can be improved by using empirical data instead of normative calculation. In *eceee Summer Study proceedings*, pages 2057–2070, Stockholm, Sweden. European Council for an Energy Efficient Economy (ECEEE).

Loga, T., Diefenbach, N., Knissel, J., and Born, R. (2005). *Kurzverfahren Energieprofil. Ein vereinfachtes, statistisch abgesichertes Verfahren zur Erhebung von Gebäudedaten für die energetische Bewertung von Gebäuden*, volume 72 of *Bauforschung für die Praxis*. Fraunhofer IRB-Verlag, Stuttgart. Available at `https://www.iwu.de/forschung/energie/kurzverfahren-energieprofil/`.

Loga, T., Stein, B., and Diefenbach, N. (2016). TABULA building typologies in 20 European countries—Making energy-related features of residential building stocks comparable. *Energy and Buildings*, 132:4–12.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester conference series on public policy*, 1(1).

Lutzenhiser, L., Moezzi, M., Hungerford, D., and Friedmann, R. (2010). Sticky Points in Modeling Household Energy Consumption. *Center for Urban Studies Publications and Reports*, 152. Retrieved from `http://archives.pdx.edu/ds/psu/22746`.

Madlener, R. and Hauertmann, M. (2011). Rebound effects in german residential heating: Do ownership and income Matter? *FCN Working Paper*, 2011(2). Available at SSRN: http://dx.doi.org/10.2139/ssrn.1887030.

Madlener, R. and Turner, K. (2016). After 35 years of rebound research in economics: where do we stand? In *Rethinking Climate and Energy Policies*, pages 17–36. Springer.

Majcen, D., Itard, L., and Visscher, H. (2013). Actual and theoretical gas consumption in Dutch dwellings: What causes the differences? *Energy Policy*, 61:460–471.

Meier, H. and Rehdanz, K. (2010). Determinants of residential space heating expenditures in Great Britain. *Energy Economics*, 32(5):949–959.

Mertesacker, S. (2020a). A structural analysis of households' heat consumption. *unpublished document.*

Mertesacker, S. (2020b). Introducing engineering knowledge into the econometric analysis of households' heat demand. *unpublished document.*

Mundaca, L., Neij, L., Worrell, E., and McNeil, M. (2010). Evaluating energy efficiency policies with energy-economy models. *Annual review of environment and resources*, 35:305–344.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., and Dasgupta, P. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change.* Ipcc, Geneva, Switzerland, 151 pp.

Palmer, J., Livingstone, M., and Adams, A. (2017). What does it cost to retrofit homes? Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/656866/BEIS_Update_of_Domestic_Cost_Assumptions_031017.pdf.

Pollak, R. A. and Wachter, M. L. (1975). The relevance of the household production function and its implications for the allocation of time. *Journal of political Economy*, 83(2):255–278.

Rehdanz, K. (2007). Determinants of residential space heating expenditures in germany. *Energy Economics*, 29(2):167–182.

Risch, A. and Salmon, C. (2017). What matters in residential energy consumption: Evidence from France. *International Journal of Global Energy Issues*, 40(1/2):79.

RWI and forsa (2011). Erhebung des Energieverbrauchs der privaten Haushalte für die Jahre 2006-2008. *RWI Projektberichte*. Retrieved from: https://www.rwi-essen.de/media/content/pages/publikationen/rwi-projektberichte/PB_Energieverbrauch-priv-HH-2006-2008.pdf.

RWI and forsa (2016). German Residential Energy Consumption Survey. 2006-2008. RWI-GRECS. Version: 1. RWI – Leibniz-Institut für Wirtschaftsforschung. Datensatz. http://doi.org/10.7807/EVpH:06-08:V1.

Scott, A. (1980). The economics of house heating. *Energy Economics*, 2(3):130–141.

Scott, A. and Capper, G. (1983). Production functions for house temperatures: Problems and issues. *Energy Economics*, 5(4):275–278.

Silva, J. M. C. S. and Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Socio-Economic Panel (SOEP) (nd). Data for years 1984-2017, version 34, SOEP, 2019, doi: 10.5684/soep.v34.

Sorrell, S. and Dimitropoulos, J. (2008). The rebound effect: Microeconomic definitions, limitations and extensions. *Ecological Economics*, 65(3):636–649.

Sorrell, S., Dimitropoulos, J., and Sommerville, M. (2009). Empirical estimates of the direct rebound effect: A review. *Energy Policy*, 37(4):1356–1371.

Statistisches Bundesamt (Destatis), Genesis-Online (2019). Verbraucherpreisindex: Deutschland, Jahre, Klassifikation der Verwendungszwecke des Individualkonsums (COICOP 2-5-Steller Hierarchie). Verfügbarer Zeitraum 1991-2019. Retrieved on 21 December 2020 from https://www-genesis.destatis.de/genesis//online?operation=table&code=61111-0003#astructure. Datenlizenz by-2-0 .

Sunikka-Blank, M. and Galvin, R. (2012). Introducing the prebound effect: The gap between performance and actual energy consumption. *Building Research & Information*, 40(3):260–273.

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

Thema, J., Tholen, L., Adisorn, T., Lütkehaus, H., Braungardt, S., Schumacher, K., and Hünecke, K. (2018). Erweiterung von Kosten-Nutzen-Analysen zu ausgewählten Energieeffizienzmaßnahmen um Erkenntnisse zum Markt für Energieeffizienzdienstleistungen und zum Aufwand von Investitionen in Energieeffizienz: Endbericht. Retrieved from `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/656866/BEIS_Update_of_Domestic_Cost_Assumptions_031017.pdf`.

Vine, E. and Barnes, B. K. (1989). Monitored indoor temperatures and reported thermostat settings: How different are they? *Energy*, 14(5):299–308.

Vives, X. (1987). Small Income Effects: A Marshallian Theory of Consumer Surplus and Downward Sloping Demand. *The Review of Economic Studies*, 54(1):87.

Willett, K. D. and Naghshpour, S. (1987). Residential demand for energy commodities. *Energy Economics*, 9(4):251–256.

Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach.* South-Western Cengage Learning, Mason, OH, USA, 5th edition edition.