TI 2023-082/V
Tinbergen Institute Discussion Paper

Energy-efficient homes: effects on poverty, environment and comfort

Vincent P. Roberdel[1]
Ioulia V. Ossokina[2]
Vladimir A. Karamychev[3]
Theo A. Arentze[4]

1 Eindhoven University of Technology

2 Eindhoven University of Technology

3 Erasmus University Rotterdam and Tinbergen Institute

4 Eindhoven University of Technology

Energy-efficient homes: effects on poverty, environment and comfort*

Vincent P. Roberdel,[#] Ioulia V. Ossokina,[#] Vladimir A. Karamychev,[##]
Theo A. Arentze [#]
[#]Eindhoven University of Technology, [##]Erasmus University Rotterdam

December 18, 2023

Abstract

Energy efficiency improvements in low income housing are increasingly used as a policy instrument to alleviate poverty. Our paper shows that this may come at the expense of reduced environmental benefits. We follow 125,000 Dutch low-income households during eight years and exploit a quasi-experimental policy that diminished the heat losses in their homes. We pay specific attention to the policy effects at the very left tail of the income distribution. While the average after-policy reduction in natural gas consumption for heating amounts to 22%, the poorest only save 16%. We build and calibrate a microeconomic model explaining this pattern from substitution between thermal comfort and other goods, and use it to compute welfare trade-offs of the policies.

JEL Codes: D12, Q4, Q48, Q5

Keywords: Energy-efficient homes, Social housing, Poverty, Quasi-experiment, Retrofit, Welfare effects

*We are grateful to housing associations Bazalt Wonen, Elan Wonen, Pre Wonen, Woonbedrijf and engineering bureau Atriensis for sharing expertise and data; to energy and microdata experts from Statistics Netherlands for support; to seminar participants in Eindhoven, Rotterdam, Rimini, Limassol, and especially Jos van Ommeren and Dinand Webbink, for valuable comments. The nonpublic microdata used in the paper are available via remote access to the Microdata services of Statistics Netherlands. We acknowledge support from the Dutch Science Foundation (NWO) grant 403.19.230 and the Dutch enterprise agency (RVO).

1 Introduction

Many countries subsidize energy efficiency upgrades in low income housing (e.g. insulation to reduce heat losses, solar panels and heat pumps for renewable energy generation). These policies hinge on two interconnected goals. One is carbon emission reduction and, as a consequence, environmental quality improvement. This was for example the driving force behind the 2012 energy retrofitting agreements within the Dutch social housing sector (Ministry of the Interior and Kingdom Relations, 2012). The other goal is reducing the energy costs for low-income households, and, as a consequence, reducing poverty. This second goal has increased in importance lately, due to the high and peaking energy prices. It is explicit in e.g. the Weatherization Assistance Program in the US, or the UK's Warm Front Home Energy Efficiency Scheme (see e.g. Fowlie et al. (2018) and Sovacool (2015)). In this paper we exploit a large-scale quasi-experiment and a microeconomic model to show that the two goals are competing: prioritizing energy efficiency upgrades for the poor comes at the expense of lower environmental benefits.

Existence of a trade-off between poverty reduction and environmental savings hinges on the hypothesis that poor households' energy consumption is less responsive to home efficiency upgrades. To illustrate this hypothesis, take consumption of heating services in the winter and the respective natural gas spending as an example. If a house suffers from large heat losses and the marginal heating costs are high, residents optimally reduce their consumption of heating services, by choosing for a lower indoor temperature in the winter. Due to the binding credit constraints, the poor might sacrifice thermal comfort more than average by accepting uncomfortably low temperatures or even not heating at all. A heating efficiency upgrade (e.g. home insulation) reduces heat losses in the house and lowers the marginal price of the heating. Following this upgrade, all households re-optimize the consumption pattern towards an increase in the indoor temperature. The poor will likely do this relatively more, because their marginal benefits from one degree temperature increase are higher. A higher increase in heating services consumption for the poor implies lower than average gas and environmental savings.

The re-optimization of the energy consumption pattern after home upgrades is generally known in the literature as the 'rebound effect' (see e.g. Sorrell and Dimitropoulos (2008) for a definition). The likely income heterogeneity in this rebound was pinpointed as early as 2000 by Milne and Boardman (2000). Later studies provided some empirical support to it by documenting lower energy savings for the poor (Aydin et al., 2017; McCoy and Kotsch, 2021; Liang et al., 2018). However, up to now there have been no large-scale studies that would allow to gain insight into the welfare effects of the resulting trade-off between poverty reduction, comfort improvement and environmental savings. Our paper aims to fill this gap.

In this paper, we follow 125,000 Dutch low- and median income households during eight years (2014-2021) and exploit a quasi-experimental policy that reduced heat losses in their homes. Heat loss reduction occurred through house insulation retrofits: adding extra material to the walls and the roof. The quasi-experiment takes advantage of a unique institutional setting involving a conditionally random treatment assignment (conditional on observable housing characteristics) and absence of self-selection. The households in our data live in social housing, owned by the so-called housing associations - non-commercial entities whose statutory duty is to provide housing to people with lower incomes.[1] Since 2012 Dutch housing associations have been required by law to improve the energy efficiency of their housing stock.[2] Due to the large size of the stock, it was physically and economically infeasible to tackle all the houses simultaneously. The dwellings to be treated in each year were thus selected by housing providers based on observable housing characteristics (i.e. age, type and energy efficiency) and on internal organization reasons, which were explicitly formulated and uncorrelated with the expected outcomes.[3] Further, again by law, the insulation

[1]The social housing sector in the Netherlands is large en includes 2.2 million dwellings (30% of the Dutch housing stock). It offers housing to people below the median income, at regulated rent levels. In 2020 the income threshold to be eligible for social housing was around 40.000 euro gross yearly income.

[2]As a result the average energy efficiency in the Dutch social housing sector improved by 20% between 2016 and 2022 (Aedes 2016, 2020).

[3]For example, retrofits were synchronized with the timing of regular painting works in buildings. In Section 3 we discuss the selection process at length and also provide a formal test

retrofits were compulsory for all tenants of each tackled housing block, so that people could not opt-out.

Using a two-way fixed effect panel regression (Angrist, 2008), we derive that the treated households reduced natural gas consumption by 22% on average, following the policy. This effect gradually increased in the first two years after retrofit and leveled out after. For the households on the left tail of the income distribution, which we are mostly interested in, estimated savings are up to one third lower than the average.

To better understand the behavioural mechanisms and welfare implications of the documented lower savings for the low-income people, we develop and calibrate a quantitative consumer choice model. The key model features are: (i) Households derive utility from thermal comfort and a composite good consumption; (ii) Households spend income on natural gas for heating and on a consumption good; (iii) There is a satiety level of thermal comfort. The model predicts that the chosen indoor temperature and the resulting gas consumption positively depend on income: in houses with bad heating efficiency, the poor choose for an uncomfortably low level of thermal comfort. When the heating efficiency of a home goes up due to a retrofit, households re-optimize their consumption by increasing the optimal temperature. They substitute part of the possible monetary savings for a higher thermal comfort. The marginal effect of a unit degree temperature increase is higher the lower the income. So the temperature increase after retrofit negatively depends on income while the resulting gas savings positively depend on income.

We calibrate the model parameters to the observed pre-retrofit gas consumption by income and the estimated gas savings from the quasi-experiment. The calibrated model yields elasticities and outcomes that are in line with known empirical stylized facts. The model is then applied to value the gas savings from retrofits as well as comfort increase that treated households in the quasi-experiment receive through re-optimizing temperature consumption. Results suggest that, for the lowest incomes, up to 20% of the private benefits from insulation occur through comfort improvement; for higher incomes it is only 5%.

of conditional randomness.

Still, the total of the private benefits from reduction in gas consumption and comfort increase is considerably smaller for the poor households as compared to average.

The estimation and calibration results are robust to a host of different parameter values and modeling assumptions. We also provide additional insight into other possible behavioural adjustments of the poor people after retrofit, such as e.g. substitution between different heating sources, and do not find evidence of these adjustments, at least in the four years after retrofit. Finally, we use econometric and causal forest techniques (Athey et al., 2019) and a large number of household and house variables present in our data, to study other possible determinants of heterogeneity in the response to the retrofits.

Our paper is related to several streams of literature. First, we contribute to the growing quasi-experimental literature on the effect evaluation of home energy efficiency programs, see Gillingham et al. (2018) and Saunders et al. (2021) for reviews. There are only few large scale quasi-experimental evaluations so far, and even fewer focus on heating efficiency improvements: Fowlie et al. (2018) and Allcott and Greenstone (2017) for US; Webber et al. (2015), Peñasco and Anadón (2023), McCoy and Kotsch (2021) and Adan and Fuerst (2016) for UK.[4] We add to the literature by performing a large-scale evaluation of the effects of heating efficiency retrofits for a continental European country, the Netherlands, in a setting that faces very few if any endogeneity concerns.

Our paper is further related to the literature on the energy efficiency gap between the ex-ante engineering forecasts and the actual savings from retrofits, and the possible behavioural explanations for this gap (the rebound effects). The gaps were documented by, among others, Sorrell and Dimitropoulos (2008); Gerarden et al. (2015); Allcott and Greenstone (2017); Aydin et al. (2017); Fowlie et al. (2018). See also a review in Peñasco and Anadón (2023). In this paper we expand this literature with a theoretical model that not only offers an explanation for the gap from the microeconomic premises, but also predicts income heterogeneity in the size of the gap. We offer an empirical test for the suggested

[4]Liang et al. (2021) for Arizona, US and Davis et al. (2014) for Mexico are large scale evaluations of the effects of electric appliances. Davis et al. (2020); Aydin et al. (2017); Hancevic and Sandoval (2022) perform small size effect evaluations for Mexico, Netherlands, US.

model, calibrate it and use it to compute broader welfare effects of the retrofits.

Yet another relevant stream of literature studies heterogeneity of energy spending responses and the factors driving this heterogeneity. Several papers support the hypothesis that low-income households' energy savings from retrofits are lower than average (e.g. Davis et al., 2014; Aydin et al., 2017; Liang et al., 2018; McCoy and Kotsch, 2021),[5] while Doremus et al. (2022) documents lower responses of the poor to extreme temperatures. On the other hand, a frequently found determinant of high responses is high energy use (Liang et al., 2018; McCoy and Kotsch, 2021). Most of these studies however have one of the two following limitations: either no household level data are available and/or only one type of heterogeneity is studied. In addition, people (and especially low-income) tend to self-select out of retrofit programs (e.g. Fowlie et al., 2018). Our paper uses a rich longitudinal and granular household dataset with data available at household and individual level. The rich content of the data allows to test for heterogeneity of the gas consumption responses in a large number of household and home characteristics. Last but not least, the institutional setting of our study excludes self-selection and opt-out.

Finally, we contribute to the public discussion on the welfare effects of the energy-efficiency policies aiming to reduce the energy burden of low-income households in social housing. Various studies in Western and Southern Europe argue that insulation retrofits can have positive effects on social welfare and should receive more attention from decision makers (Avanzini et al., 2022; Sdei et al., 2015; Walker et al., 2014). We provide new evidence on the costs and benefits of targeting poor households with energy efficient home upgrades. We show that policies targeting heating efficiency improvements to poor households result in lower than average gas savings and thus forego possible environmental benefits. Comfort improvements that low income households experience due to retrofits compensate part of this loss.

The structure of the article is as follows. Section 2 introduces the theoretical model explaining why the poor have lower responses to energy retrofits. Section 3

[5] On the contrary Hammerle and Burke (2022) finds that vulnerable households experience a higher gas reduction after switching to electric heating.

describes the institutional background, the sample and the data. Section 4 discusses the empirical methodology and identification and reports the main results. Section 5 investigates heterogeneity of the treatment effect. Section 6 describes the calibration and calculates welfare effects. Section 7 concludes.

2 Theoretical Framework

In this section, we develop a simple consumer choice model in which a household spends its income on consumption and heating. We solve the model, deriving the optimal household consumption and the resulting indirect utility function. Then, we decompose the welfare gain from an energy efficiency upgrade into a thermal comfort component and a consumption component.

2.1 Model

Household utility is given by the following constant elasticity of substitution, CES, specification:

$$u(x,\theta) \stackrel{\text{def}}{=} \left((f_1(x))^{\frac{\sigma-1}{\sigma}} + (f_2(\theta))^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}},$$

where $x \geq 0$ is the consumption of a composite good and $f_1(x)$ is a consumption utility component, whereas $\theta \in [\theta_0, \bar{\theta}]$ is a thermal comfort measured by the indoor winter temperature and $f_2(\theta)$ is a thermal comfort utility component. Parameter σ is the elasticity of substitution between the consumption utility and the thermal utility components.

We assume that thermal comfort utility $f_2(\theta)$ is an increasing and concave function that reaches its maximum at some temperature $\bar{\theta}$. The idea behind this assumption is that not only too low but also too high indoor temperatures negatively affect individual wellbeing. As a result, the household never chooses a value of θ beyond $\bar{\theta}$. We operationalize $f_2(\theta)$ as a second degree concave polynomial as follows:

$$f_2(\theta) \stackrel{\text{def}}{=} (2\bar{\theta} - \theta)\theta.$$

For consumption utility, we assume

$$f_1(x) \stackrel{\text{def}}{=} x$$

for simplicity. Thus, the household utility is defined for $x \geq 0$ and $\theta \in [\theta_0, \bar{\theta}]$ by

$$u(x, \theta) = \left(x^{\frac{\sigma-1}{\sigma}} + \left((2\bar{\theta} - \theta)\theta \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}. \tag{1}$$

The household earns income w and spends it on (x, θ). We normalize the price for consumption to unity, $p_x = 1$. Thermal comfort is produced from natural gas according to the production function:

$$\theta = \theta_0 + qg,$$

where $g > 0$ is gas usage for heating, Parameter $\theta_0 \leq \theta$ is a natural indoor winter temperature, i.e., the temperature achieved without any additional heating and $q \geq 0$ is a home heating efficiency parameter. The higher is q, the less gas is needed to increase the indoor temperature by one degree. We use an increase in q from $q = q_L$ to $q = q_H > q_L$ for modeling home heating efficiency upgrades.

The household buys natural gas at a market price $p_g > 0$ so that the household budget constraint is given by:

$$x + \frac{p_q}{q}(\theta - \theta_0) = w. \tag{2}$$

where $\frac{1}{q}(\theta - \theta_0)$ is the annual gas use. Thus, all households face the same gas price p_g, and each household is characterised by parameters $(q, w, \theta_0, \bar{\theta}, \sigma)$ and chooses $x \geq 0$ and $\theta \in [\theta_0, \bar{\theta}]$ to maximize utility (1) subject to budget constraint (2).

2.2 Household Optimal Behaviour

The household maximizes its utility (1) over $x \geq 0$ and $\theta \in [\theta_0, \bar{\theta}]$ subject to the budget constraint (2). In the following proposition, we provide the solution (x^*, θ^*) to this utility maximization problem, UMP.

Proposition 1 *Let the critical income level $\underline{w}(q)$ be defined by:*

$$\underline{w}(q) \stackrel{\text{def}}{=} (2\bar{\theta} - \theta_0)\theta_0 \left(\frac{p_g}{2q(\bar{\theta} - \theta_0)} \right)^{\sigma}. \tag{3}$$

Then:

1. *If $w \leq \underline{w}(q)$, then $x^*(q, w) = w$ and $\theta^*(q, w)) = \theta_0$.*

2. *If $w > \underline{w}(q)$, then $\theta^*(q,w)$ is uniquely defined by:*

$$0 = w - (2\bar{\theta} - \theta^*)\theta^* \left(\frac{p_g}{2q(\bar{\theta} - \theta^*)}\right)^{\sigma} - \frac{\theta^* - \theta_0}{q} p_g, \tag{4}$$

and

$$x^*(q,w) = w - \frac{\theta^* - \theta_0}{q} p_g. \tag{5}$$

3. *For $w \geq \underline{w}(q)$, $x^*(q,w)$ and $\theta^*(q,w)$ increase in w.*

4. *$\theta^*(q,w)$ increases in q, and it approaches $\bar{\theta}$ when q or w increase unboundedly.*

The proof of the proposition is in Appendix A. For low income levels below $\underline{w}(q)$, the optimal consumption is a corner solution where the household consumes no gas and stays at the natural house temperature $\theta^*(q,w) = \theta_0$. The household spends then all its income w on the composite good, $x^*(q,w) = w$. For higher income levels, $w > \underline{w}(q)$, the optimal consumption θ^* is an interior solution satisfying $\theta^* \in (\theta_0, \bar{\theta})$. Both thermal comfort θ and composite consumption x are normal goods so that their consumption increases with income w.

Figure 1a illustrates Proposition 1. It shows the optimal indoor temperature θ^* as a function of income w for two values q_L and q_H of the heating efficiency parameter q, with $q_L < q_H$. For the lowest income levels, the optimal thermal comfort is at its natural level θ_0. With rising income, the optimal thermal comfort also rises and converges in the limit to the satiety threshold $\bar{\theta}$. With the increase in q, the optimal thermal comfort starts to increase at lower income levels.

The optimal gas consumption $g^*(q,w)$ is determined by $\theta^*(q,w)$:

$$g^*(q,w) = \frac{1}{q}(\theta^*(q,w) - \theta_0). \tag{6}$$

In the following proposition, we provide a characterization of $g^*(q,w)$.

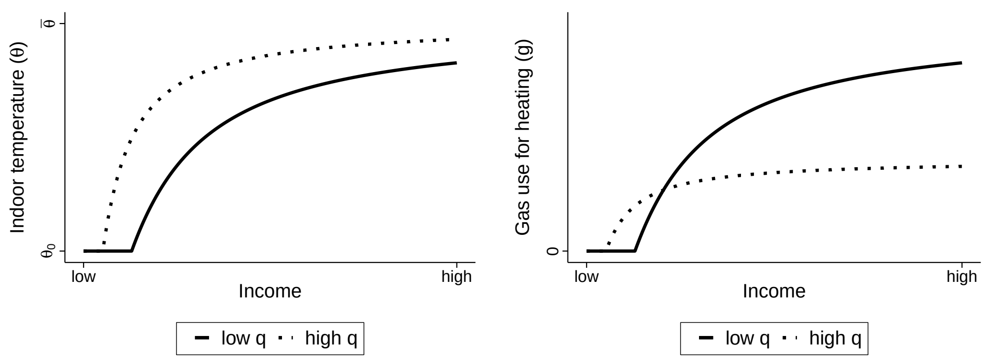
Proposition 2 *Optimal gas consumption $g^*(q,w)$ has the following properties:*

1. *$g^*(q,w)) = 0$ for $w \leq \underline{w}(q)$.*

2. *$g^*(q,w)$ increases in w for $w > \underline{w}(q)$ and and converges to $\frac{1}{q}(\bar{\theta} - \theta_0)$ when w increases unboundedly.*

Figure 1: Optimal thermal comfort θ^* and gas consumption g^*.

(a) Thermal comfort (b) Gas consumption



Notes: The lines show the optimal levels of indoor temperature respectively natural gas consumption as a function of income ($\theta^*(w)$ respectively $g^*(w)$) as implied by the utility maximization problem Equations (1) and (2). See Appendix A for the derivations.

The proof of the proposition is a straightforward application of the results of Proposition 1 and is, therefore, omitted. Figure 1b illustrates Proposition 2. It shows the optimal gas consumption g^* as a function of income w for two values q_L and q_H of the heating efficiency parameter q, with $q_L < q_H$. Since the graph of $g^*(q_H, w)$ starts to increase at a lower income level $\underline{w}(q_H)$ and converges to a lower limit $\frac{1}{q_H}(\bar{\theta} - \theta_0)$ than the graph of $g^*(q_L, w)$ does, the graphs necessarily intersect. For income levels below $\underline{w}(q_H)$, the optimal gas consumption is zero, $g^*(q_H, w) = g^*(q_L, w) = 0$. For income levels $w \in (\underline{w}(q_H), \underline{w}(q_L))$, $g^*(q_H, w) > g^*(q_L, w) = 0$. For sufficiently large income levels, $g^*(q_H) < g^*(q_L)$.

Summarising, when the heating efficiency of a house increases, all households re-optimize their consumption patterns, trading-off potential natural gas savings against an increase in the level of thermal comfort. Households with a sufficiently low income can even increase their gas consumption because they are further away from the satiety threshold and, therefore, face a larger marginal benefit of a unit temperature increase. High-income households, to the contrary, decrease their gas consumption because for them, the marginal benefit of a unit temperature increase is low. This results in lower gas savings for the poor, as compared to the rich.

2.3 Household Welfare

Let $V(q,w)$ be indirect utility of the household:

$$V(q,w) \stackrel{\text{def}}{=} u(x^*(q,w),\theta^*(q,w)),$$

where u, θ^*, and x^* are defined respectively by Equations (1), (4) and (6). Then, the compensating variation CV for the change in heating efficiency q from $q = q_L$ to $q = q_H$, is implicitly defined by:

$$V(q_H, w - CV) = V(q_L, w). \tag{7}$$

Compensating variation CV is the household's willingness to pay for the heating efficiency improvement and is, therefore, the monetary measure of the corresponding welfare gain. In other words, CV is the income effect of the heating efficiency improvement. It is negative to the so-called Hicksian compensation of the heating efficiency change:

$$\Delta^H \stackrel{\text{def}}{=} -CV.$$

An alternative, yet imprecise measure of the same income effect is the Slutsky compensation Δ^S, which is defined as follows:

$$\begin{aligned}\Delta^S &\stackrel{\text{def}}{=} p_x x^*(q_L,w) + \frac{p_g}{q_H}(\theta^*(q_L,w) - \theta_0) - w \\ &= x^*(q_L,w) - x^*(q_H,w) + \frac{p_g}{q_H}(\theta^*(q_L,w) - \theta^*(q_H,w)).\end{aligned}$$

By construction, $(-\Delta^S)$ equals the income of the household that remains after the thermal upgrade from q_L to q_H if the household maintains the pre-upgrade consumption levels $x^*(q_L,w)$ and $\theta^*(q_L,w)$. Despite that $-\Delta^S$ is an imprecise measure of CV, it can readily be decomposed into the effect on composite good consumption:

$$-\Delta^S_x \stackrel{\text{def}}{=} x^*(q_H,w) - x^*(q_L,w), \tag{8}$$

and the effect on thermal comfort consumption:

$$-\Delta^S_\theta \stackrel{\text{def}}{=} \frac{p_g}{q_H}(\theta^*(q_H,w) - \theta^*(q_L,w)). \tag{9}$$

In the following sections, we exploit quasi-experimental improvements in the heating efficiency of Dutch houses to estimate $g^*(q,w)$ as a function of income and heating efficiency. Then, in Section 7, these results are used to calibrate the model Equations (1) and (2) and to compute CV and its approximate decomposition into $(-\Delta_x^S)$ and $(-\Delta_\theta^S)$. According to theory, $\Delta^S \geq \Delta^H$ and $\Delta^H = -CV$. Hence, the sum of the effects $(-\Delta_x^S)$ and $(-\Delta_\theta^S)$ does not exceed CV:

$$(-\Delta_x^S) + (-\Delta_\theta^S) = -\Delta^S < -\Delta^H = CV,$$

so that the sum of these effects underestimates the exact income effect CV.

3 Quasi-experiment, data and sample

Before discussing the empirical model, we first introduce the quasi-experiment and the data. We start with describing the institutional background of the Dutch social housing as this is crucial for our identification strategy.

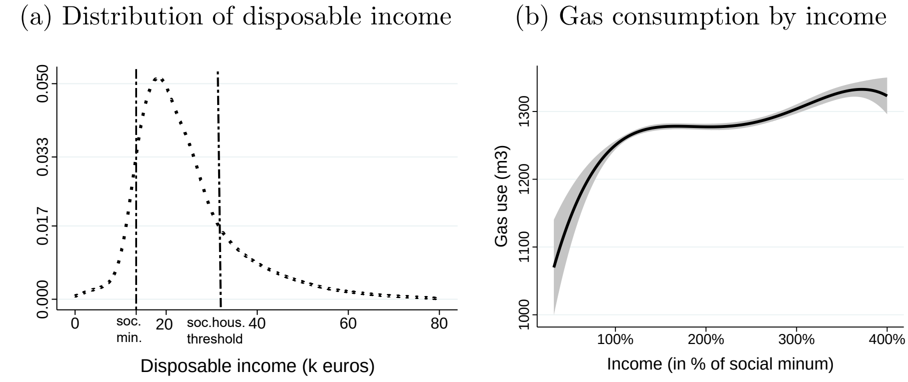
3.1 Dutch social housing: residents and dwellings

This study focuses on the households living in Dutch social housing. The social housing sector in the Netherlands is large en includes 2.2 million dwellings (one third of the Dutch housing stock). It offers housing at regulated rent levels to households with an income below the median. In 2020 the threshold to be eligible for social housing was around 40.000 euro yearly gross income per household (this amounts to some 33.000 euro disposable income). However the income check is only done once, when the renter signs a contract for a new dwelling. Therefore, although the majority of social renters are low income people, also households with incomes higher than the threshold live in the social dwellings. Figure 2a shows the distribution of the social housing residents by income; our data offers considerable variation by income on both tails (below the social minimum and above the threshold), which we will use in our study.

Figure 2b plots households' yearly natural gas consumption against their disposable incomes, for the same households as in Figure 2a. As 75% of natural gas consumed by a household per year, is spent on space heating and another 20% on hot water (Eurostat (2023)), we conclude that the insights of the Figure are in line with the theoretical conclusions of the previous Section.[6] On the one

[6] The theoretical conclusions from Section 2 hold for thermal comfort from both space heating

Figure 2: Income and gas consumption in social housing 2016



Notes: Figure (a) shows the distribution of the disposable income, for the residents of social houses. The vertical line of the left indicates the median income of households below the social minimum (this social minimum is computed by Statistics Netherlands, its value depends on the household type). The vertical line of the right indicates the maximum income threshold to enter social housing. Figure (b) shows a polynomial fit (of degree 4) of household's gas use against income, whereby income is measured in percent of social minimum, in 2016.

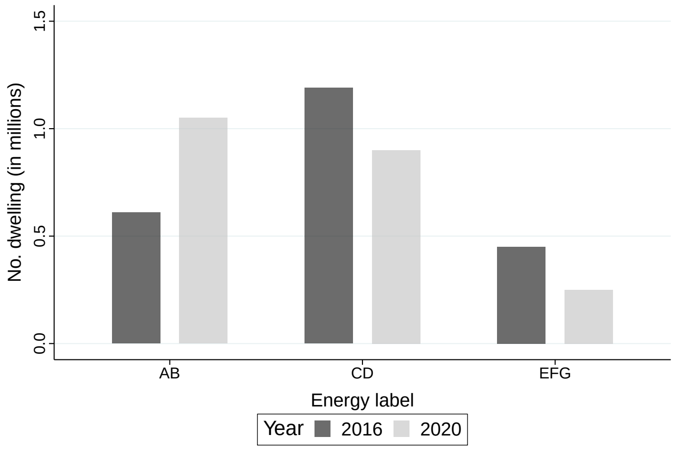
hand, gas use increases in income. On the other hand, there is a diminishing marginal effect. While the median gas consumption in the social housing lies around 1270 m^3/year, the poorest consume up to 10% less.

We turn now from the residents of social houses to the dwellings they live in. The potential for energy and environmental savings in the social housing sector is high. About two-thirds of the stock was built before 1993, according to the low energy efficiency building standards of that time. Social housing owners - the so-called housing associations - are required by the government to improve the energy efficiency of these properties. Energy retrofits started with the 2012 convenant Social Sector which aimed at 33% CO_2 savings by 2020. Until 2020 half a million homes was improved, still leaving one million homes to go.

Figure 3 plots the distribution of the social dwellings by energy efficiency in 2016 and 2020, as measured by the European energy label. This label is

and hot water usage. Space heating comfort can be measured with indoor temperature at winter while hot water comfort can be measured with water temperature. The model for hot water comfort would be identical to the model for space heating, with a natural water temperature and a satiety water temperature level.

Figure 3: Energy efficiency in social housing



Notes: The Figure reports the number of social houses in the Netherlands in millions, by energy efficiency label, in 2016 and 2020. Source: Aedes (2016,2020).

derived from the thermal quality of the dwelling and is assigned to dwellings by trained professionals after a technical inspection. The label takes elements such as insulation quality, heating installation, (natural) ventilation and indoor air climate, solar systems, and built-in lighting into account. The label is based on a simple universal indicator of the energy consumption – the energy index, which reflects the engineering projection of primary energy consumption under average conditions. Labels 'A-B' are considered good, labels 'E-F-G' are considered bad and need to be improved in the first place. Figure 3 shows that the share of the labels 'C' to 'G' (medium to poor energy efficiency) fell between 2016 and 2020, and the share of the labels 'A' and 'B' grew. This mostly happened through retrofits - heating and electricity system upgrades.[7] In this paper we will study the effects of the heating efficiency upgrades applied to the dwellings of labels ranging from 'C' to 'G'.

3.2 Heating efficiency retrofits; quasi-experiment

One of the most frequent heating-efficiency retrofits in the social housing is *insulation of the building*, whereby materials are added to the walls and the roof in order to reduce the heat losses and the natural gas quantity required for

[7]New construction was another factor that affected this shift.

heating. Insulation is often seen as a prerequisite for many other energetic improvements. In this and next Sections we study the effects of the insulation retrofits undertaken by the Dutch social housing associations in 2017-2019, on the natural gas consumption of the social housing residents.[8] Two characteristics of these retrofits are important for our identification strategy and allow for a quasi-experimental approach; we highlight these here. First, as discussed above, the total number of old and energy-inefficient dwellings that qualified for an insulation upgrade was very large in 2016. These houses could not be tackled all simultaneously because of the financial and physical constraints. Therefore, a selection rule to prioritize some houses above other was necessary. From discussions with renovation managers of a number of Dutch housing associations[9] we learned that, during the study period, targeting was largely based on observable building characteristics (e.g. construction period, energy label), on the one hand, and on organizational considerations, on the other hand. The latter generally implied synchronizing the retrofit with the regular maintenance like painting of exterior walls, replacement of lighting, pipes and tubes in the building.[10] Regular maintenance is a cyclical process for which planning is known for many years to go (e.g. painting is usually scheduled every 6 years, etc.) It is performed by *complex* - a block of adjacent houses sharing the same building year and similar technical characteristics. The timing of regular maintenance can thus be assumed independent of and uncorrelated with the potential outcomes of insulation retrofits.[11] As a result, the assignment of the houses to treatment can arguably be considered random, conditional on a few observable building characteristics such as: construction year, energy efficiency and dwelling type.

The second useful feature of the social housing insulation retrofits is that

[8]Insulation retrofits in our data include roof, floor, facade insulation as well as replacing window frames and glass for energy-efficient ones.

[9]We are grateful for these discussions to the experts of Bazalt Wonen, Elan Wonen, PreWonen, Woonbedrijf.

[10]Recently, due to the rising energy prices, other criteria - like tackling poor households first - have also been used in prioritizing insulation retrofits. This change is outside the (time) scope of our study.

[11]Note that replacement of the boiler - an intervention that does affect gas usage - does not fall under regular maintenance and follows an own cycle, which is often dwelling-specific.

self selection in or out of retrofit was next to impossible for the tenants. By Dutch law, if 70% tenants of a complex agree with the retrofit plans (and this was mostly the case in social housing), individual tenants do not have a right to opt out any more, even if they wish so. This means that the assignment to treatment of households can be seen as conditionally random, and the treated sample is representative for the social renters population in the country.[12]

The randomness of the treatment conditional on observed building characteristics is an important identification assumption in the quasi-experimental fixed effects panel regression method we aim to use. We will also formally test this assumption in Section 3.4.

3.3 Sample and data

We exploit information on insulation retrofits performed by 128 Dutch social housing associations in 2017-2019. The housing associations in the sample collectively own about 1 million dwellings located in all regions of the country. Our sample covers 40% of the total social housing stock and is representative for the Dutch social housing sector.

We combine two data sets. The first one includes longitudinal dwelling-level data on building characteristics, energy efficiency indicators and insulation retrofit attributes for the years 2016-2021.[13] The second, also longitudinal, dataset contains restricted access microdata on household level made available by Statistics Netherlands. These include socio-economic characteristics of the households as well as their yearly consumption of gas and electricity for the years 2012-2021. Two datasets are merged on address level. This yields, for one million houses, information on (1) structural house characteristics 2012-2021, (2) retrofit incidence and retrofit characteristics 2016-2021, (3) resident household characteristics 2012-2021 and (4) energy use 2012-2021.

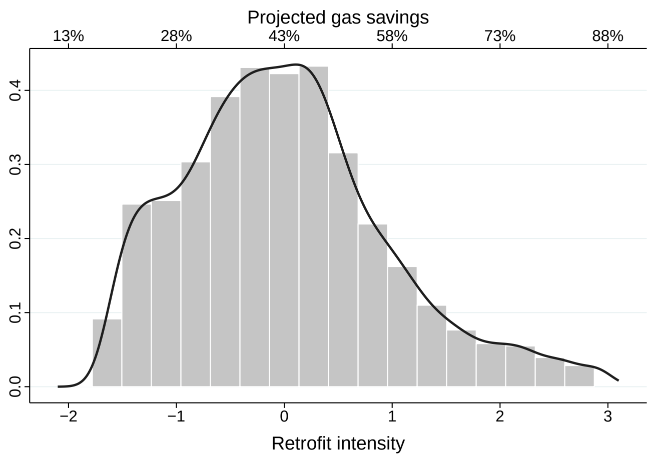
Our main outcome variable is yearly gas consumption per dwelling in cubic

[12] We note that people could vote with their feet and relocate to another house if they did not agree with the insulation retrofit. We will show formally that this did not happen.

[13] We thank engineering bureau Atriensis for sharing with us their Energy Monitor data, and social housing associations Bazalt Wonen, Elan Wonen, PreWonen, Woonbedrijf for sharing their expertise and additional data on retrofits.

meters. The main covariate is the binary indicator of whether a dwelling was retrofitted in or before a specific year. Further, as the type and size of the insulation retrofit may differ by house, we derive a retrofit intensity index and include it - in a standardized form - as a control.[14] The intensity is a continuous variable based on the engineering projections of the change in dwelling heating efficiency after retrofit (i.e. change in the engineering projected log gas consumption). Engineering projections are conventionally made under the NEN 7120 guidelines by the building performance software VABI, which is used by all housing associations in our data. Figure 4 reports a histogram of the standardized retrofit intensity. Other covariates included as controls are: house and household characteristics (dwelling type, construction period, surface, energy efficiency of the house, household type, number of persons, education, income, etc.) as well as energy consumption before retrofit.

Figure 4: Distribution of retrofit intensity



Notes: The figure shows the distribution of the retrofit intensity in the data sample used in this paper. We define the retrofit intensity as the standardized projected gas savings. The projected gas savings are the difference between pre- and post-retrofit projected log gas consumption. Projected gas consumption is computed by the engineering building performance model VABI.

To test the hypothesis about divergent responses of households on the left tail of the income distribution, we make use of the poverty line and social minimum

[14] To fully grasp the effect of retrofit intensity, we will include the retrofit intensity as a different order flexible polynomial in the regression, see Appendix D.

indicators defined by Statistics Netherlands. The definition of the social minimum is the 'minimal amount one needs in order to cover basic personal needs' (Statistics Netherlands). The amount is determined yearly and is derived from the size of the social welfare benefits. It therefore depends on the composition of the household. The poverty line is an amount that represents the same purchasing power as the social welfare benefits in year 1979. It also depends on the composition of the household. E.g. in 2017 the poverty line equaled a monthly disposable income of 1 040 euro for a single person, 1 380 euro a one-parent family with one child and 1 960 euro for a couple with two children. Using the two indicators, we will distinguish four strata of poor households: those below (i) 100%, (ii) 130%, (iii) 150% of the social minimum and (iv) households below the poverty line.

3.4 Treatment and control group, descriptives

In the main analysis we will focus on single family dwellings that qualified for an insulation retrofit in 2016, according to two criteria: building year before 1993, energy label 'C' to 'G'. We drop dwellings with missing data on energy efficiency and energy use, student condominiums and dwellings without individual natural-gas-based heating during the study period. The resulting study sample contains 124,300 single family dwellings, of which 13,409 belong to the treatment group and 110,891 to the control group. The treatment group is defined as houses which got an insulation retrofit between 2017 and 2019 and did not change tenant between one year before and one year after the retrofit. The control group is defined as dwellings that did not experience an energy efficiency upgrade between 2000 and 2021.

Table 1 reports the descriptive statistics for the treatment and control groups in 2016, the year before the first treatment in the sample. We distinguish three groups of characteristics: dwelling (panel A), household socio-economics (panel B) and energy usage (panel C) and report the balancing tests. The socio-economics are balanced well between treatment and control groups, while the dwelling characteristics and energy usage are not. This is in line with the assumption of the conditional random assignment to treatment based on dwelling characteristics (see Section 3.2). To test this assumption formally, we perform a randomization test for covariate imbalance as suggested in e.g. Hennessy et al. (2016). First, we regress the observed gas consumption in 2016 on the building covariates. The residuals from this regression we call "adjusted gas consumption". Then, we carry out the randomization test by calculating the test statistic - the difference of means of adjusted gas consumption between treatment and control groups. The test yields a statistic of $-3m^3$ with a p-value of 0.35 (calculated over 10,000 random permutations), suggesting that the gas consumption adjusted for building covariates is well-balanced. We therefore cannot reject the null hypothesis that the (adjusted) gas consumption does not differ between the control and treatment groups. In sum, socio-economics and energy use covariates are balanced. This is consistent with the assumption that the treatment assignment is determined by observed building characteristics only. We will account for the imbalance in dwelling characteristics by controlling for them explicitly in the empirical model.

4 Gas savings from retrofits: average and poor

4.1 Empirical model and identification

Our main empirical method is a two-way fixed-effect panel regression with year and household/dwelling fixed effects. As the sample is defined to only include households that lived in the dwelling at the time of the retrofit, the dwelling and household fixed effects coincide. The baseline econometric specification is:

$$g_{i,t} = R_{i,t}\left(\alpha + \beta S_i\right) + \delta X_{i,t} + \gamma_i + \phi T_t + u_{i,t}. \tag{10}$$

Here $g_{i,t}$ is the (log) yearly gas consumption of household/dwelling i in year t. The binary treatment variable $R_{i,t}$ takes value 1 in the years following retrofit and value 0 before;[15] S_i is the retrofit intensity (see Section 3.3 for the definition); X_i controls for time-varying observable characteristics of the household (e.g. size) and dwelling (e.g. new boiler installed); γ_i are household/dwelling time invariant fixed effects; T_t are year fixed effects and $u_{i,t}$ is the idiosyncratic error term.

[15] We will control separately for the retrofit year self because of the noise in the data - we do not know in which month the retrofit was performed.

Table 1: Comparison of treatment and control groups

	Treatment	Control	p-value	SMD	VR
Panel A: Socio-economics					
No. persons	2.13	2.09	0.00	0.03	1.10
No. children	0.63	0.58	0.00	0.06	1.11
No. seniors	0.51	0.51	0.68	0.00	1.00
Income (k euro)	26.63	27.43	0.00	0.07	0.89
Education high (0/1)	0.10	0.10	0.15	0.01	0.97
Migration background foreign (0/1)	0.22	0.20	0.00	0.05	1.07
Below 100% social min. (0/1)	0.03	0.03	0.62	0.00	0.98
Below 130% social min. (0/1)	0.27	0.26	0.46	0.01	1.01
Below 150% social min. (0/1)	0.38	0.37	0.06	0.02	1.01
Below poverty line (0/1)	0.08	0.08	0.85	0.00	1.01
Panel B: House characteristics					
Surface (m^2)	94.79	94.27	0.00	0.03	0.87
Constr. Period 1906-1939 (0/1)	0.06	0.07	0.00	0.05	0.83
Constr. Period 1940-1965 (0/1)	0.53	0.30	0.00	0.48	1.18
Constr. Period 1966-1976 (0/1)	0.38	0.32	0.00	0.13	1.08
Constr. Period 1977-1992 (0/1)	0.03	0.31	0.00	0.79	0.15
Energy label EFG (0/1)	0.45	0.26	0.00	0.40	1.28
Panel C: Energy use					
Electricity (kWh)	2538.12	2601.39	0.00	0.05	0.95
Gas (m^3)	1371.25	1270.60	0.00	0.21	1.07
Heating burden	0.05	0.04	0.00	0.18	1.10
No. houses	13409	110891			
No. complexes	980	9957			
No. housing associations	113	96			

Notes: The table reports a balancing test between treatment and control dwellings. The columns *mean treated* and *mean control* report the mean values of selected covariates. The column *p-value* reports the p-value of a mean equality test between treatment and control group. The column *SMD* reports the standardised mean difference between the treatment and the control group. The column *VR* reports the variance ratio. $SMD = |\bar{X}_{treated} - \bar{X}_{control}|/\sqrt{\left(S^2_{treated} + S^2_{control}\right)/2}$ and $VR = S^2_{treated}/S^2_{control}$, where $\bar{X}$ is the sample mean and S^2 is the sample variance. The balancing is considered good for *SMD* smaller than 0.25 *VR* between 0.5 and 2 (Rubin, 2001; Stuart, 2010).

We are in the first place interested in the coefficient α, which can be interpreted as the effect of a retrofit of *average intensity* in the post-insulation years. This interpretation is possible because we standardized the retrofit intensity variable. To allow for slow adjustment (e.g. because of learning), we also estimate a model that includes dynamic yearly effects. For this we use a distributive lag specification as, e.g. in Ossokina et al. (2022):

$$g_{i,t} = \sum_{L=-5}^{4} R_{i,t-L} \left(\alpha_L + \beta_L S_i\right) + \delta X_{i,t} + \gamma_i + \phi T_t + u_{i,t}. \tag{11}$$

Finally, to study heterogeneity in the effects of insulation retrofist, we use a two-way interaction to allow the coefficients to differ by stratum $j \in J$:

$$g_{i,t} = \sum_{j=1}^{J} R_{i,j,t} \left(\alpha_j + \beta_j S_i\right) + \delta X_{i,t} + \gamma_i + \phi T_t + u_{i,t}. \tag{12}$$

4.2 Identification

Our main identification strategy is based on a fixed effects panel regression (Angrist, 2008). To derive a causal effect of a heating efficiency improvement on natural gas consumption, we use a treatment and a control group as defined in Section 3.4. The internal validity of this approach hinges on the assumption that the treatment assignment was random, conditional on observed dwelling characteristics. Section 3.1 provided institutional arguments and Section 3.4 a formal test to support the assumption. Including in the regression dwelling fixed effects and dwelling time varying controls accounts for the imbalance in dwelling characteristics.

Below we discuss a number of other possible identification concerns. The first concern is related to the retrofit intensity. We argued above that, conditional on the building characteristics, the assignment to the treatment can be seen as random. The retrofit intensity S in Equation (10) is however not random. The 2012 energy agreements in the Dutch social housing sector (Ministry of the Interior and Kingdom Relations, 2012) prescribed an improvement of energy efficiency at least to a (high) energy label B. Consequently, the lower the initial energy efficiency, the larger the assigned retrofit intensity would be, ceteris paribus. In the econometric model, we account for the effect of the retrofit intensity by including it explicitly in the regression equation. The baseline specification Equation (10) includes the retrofit intensity in a linear way. We also run alternative specifications with a flexible higher order polynomial in retrofit intensity, to allow for a non-linear relationship, as well as a specification without retrofit intensity.

Still, in Equation (12) which studies heterogeneity of the treatment effect by (socio-economic) group, one may be concerned about the possible correlation between the retrofit intensity and specific socio-economic characteristics of the household. If, for example, people with lowest income systematically get larger retrofits, the treatment coefficient for this group may be biased. To tackle this concern, we show that retrofit intensity is not correlated to socio-economic variables nor to initial gas consumption. Table B1 in Appendix B reports the estimation results from regressing the retrofit intensity on dwelling, income and energy consumption characteristics of the households. As expected, the pre-retrofit energy efficiency of the dwelling is negatively correlated with the retrofit intensity. Socio-economic and gas consumption variables show either no statistically significant or next to zero relationship with the retrofit intensity.

The second concern is related to the self-selection into/ out of the treatment group. As discussed in Section 3.1, by law tenants could not opt out of the insulation retrofit program while living in the dwelling. They could, however, avoid the retrofit by moving out of the dwelling. Based on information obtained from the experts working for housing associations and on institutional knowledge, we do not expect that moving decisions of tenants are endogenous on the treatment, for three reasons. First, social houses in the Netherlands are offered at a considerable discount compared to the market rents, and the waiting lists are long (Van Ommeren and Van Der Vlist, 2016). Second, rents are tenure-related: only limited rent increases are allowed for incumbent tenants, while this rule does not apply for new tenants. Moving house thus generally implies a considerable upward jump in rent. Third, retrofits never have negative financial consequences for the tenants: by law, rent increases associated to retrofits may not be larger than the energy bill savings (see e.g. Ossokina et al. (2021)).

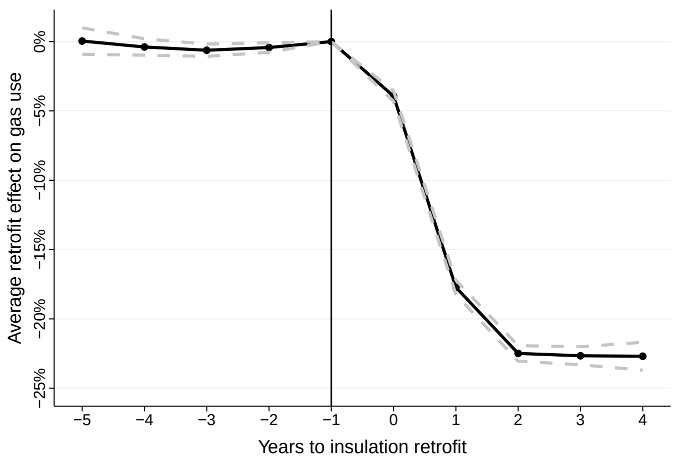
Still, we test whether there are indications that the decision to move is endogenous on the treatment. We estimate a logit model with as dependent variable the indicator of moving out of the dwelling and as main independent the treatment indicator. Controls are also included. The results are in Table B2 in Appendix B. They show no evidence that retrofits are associated with higher relocation frequencies. On the contrary, if anything, there might be a very small negative effect of retrofits on the probability to move.

4.3 Empirical results

4.3.1 Average treatment effect

We start with reporting the yearly effects of insulation retrofits from Equation (11). These are plotted in Figure 5. As expected, the Figure shows no statistically significant effect in five years before retrofit, and a gradual increase in the absolute size of the effect after, from 19% gas savings in the first year to 22% in the years two to four. Note that the effect in the year of retrofit is not informative, because we do not know the exact month in which the retrofit was performed. In sum, households need (some) time to adjust their behaviour; this adjustment process reaches its equilibrium quickly however.

Figure 5: Gas savings from retrofit: event study by year



Notes: Plotted values are the coefficients of the interaction effect of the treatment indicator with the year-to-retrofit, see Equation (11). Year -1 (vertical line) is the last pre-retrofit year. The dashed lines represent the 95% confidence interval. Standard errors are clustered at household level.

Table 2 reports the estimated average treatment effect (ATE) from Equation (10). To account for the slow adjustment found in Figure 5, we control for the year of retrofit and the year after with separate dummies. The results in the table should thus be interpreted as the estimated effect in the years two-four after retrofit. Columns (1) to (4) report different specifications: with or without household/dwelling fixed effects and with or without household time-varying controls.[16] Our preferred specification (4) includes household and year fixed effect, as well as household controls.

Table 2: Average effects of insulation retrofit on gas consumption

Dependent: log of yearly natural gas use	(1)	(2)	(3)	(4)
Retrofit (year $\geq$ 2)	-0.149***	-0.228***	-0.228***	-0.218***
	(0.004)	(0.003)	(0.004)	(0.003)
Retrofit (year $\geq$ 2) $\times$ Retrofit index	-0.058***	-0.077***	-0.100***	-0.078***
	(0.004)	(0.003)	(0.004)	(0.003)
No. obs.	963459	963459	959073	959073
No. treatment houses	13409	13409	13409	13409
No. control houses	110891	110891	110891	110891
R^2 Adj.	0.021	0.822	0.144	0.826
Year fixed-effect	X	X	X	X
Household fixed-effect		X		X
Controls			X	X

Notes: The table shows estimates of four separate regressions. The dependent variable is the log of gas consumption. Standard errors in parentheses are clustered at household level. Statistical significance: *p<0.1; **p<0.05; ***p<0.01.

The main finding is that an insulation of average intensity reduces natural gas consumption of households by about 22%. One standard deviation increase in retrofit intensity reduces gas consumption by another eight percentage points.[17]

[16] Table C1 in Appendix C reports the full set of coefficients for the four specifications.

[17] In Appendix E we include retrofit intensity in different functional specifications, including a flexible polynomial. The retrofit intensity effect is robust, the higher order terms are not statistically significant.

4.3.2 Effects for the poor and underlying mechanisms

Table 3 reports the estimated average treatment effect (ATE) for the poor households. Here Equation (12) was run four times with $J = 2$, including a two-way interaction of the retrofit indicator with each time another poverty dummy indicator, as defined in Section 3.3. In line with the theoretical model, we find that the magnitude of the gas savings falls with income, more so on the very left tail of the income distribution. The poorest (below 100% social minimum) show one third smaller savings than the average; those below 130% of the social minimum one tenth lower savings.

Table 3: Effects of retrofits for poor households

	Baseline
Retrofit (year ≥ 2)	-0.218 (0.003)***
$\times$ Below poverty line	0.024 (0.011)**
$\times$ Below 100% soc.min.	0.062 (0.016)***
$\times$ Below 130% soc.min.	0.025 (0.006)***
$\times$ Below 150% soc.min.	0.018 (0.006)***

Notes: The table shows estimates of Equation (12) for 5 separate regressions. Coefficients reported are two-way interactions. The symbol $\times$ indicates an effect as compared to the reference level (non-poor). The dependent variable is log of gas. Each regression includes controls, household fixed-effects and year fixed-effects. The sample size is 13409 treated and 110891 control units. Standard errors in parentheses are clustered at household level. Statistical significance: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

The above analysis provides empirical support for the hypothesis that the lowest incomes realize smaller gas savings after an insulation retrofit. The underlying mechanism we hypothesized in the theoretical model is that poor households reoptimize their heating consumption patterns after retrofit more than others, because their pre-retrofit heating consumption was relatively far from the satiety threshold. Re-optimization of the heating consumption can however take place through other channels too, next to gas heating. An obvious candidate is electricity consumption. Our data allow to test for the existence of substitution effects between gas and electricity after insulation retrofit. Some 5300 dwellings in the treatment sample got solar panels (amounting, on average, to 2 000 kWh renewable electricity per year), simultaneously with the insulation. We test whether these households differently responded to insulation than households without solar panels. If there is substitution between gas and electricity in heating, we should see larger gas savings for the solar-households, as they can make use of additional free solar energy at their disposal. Also, we should see a rise in grid electricity consumption for the non-solar households. Table 4 reports the results of running Equation (12) with two-way and three-way interactions of treatment, solar and poverty indicators. We use as outcome variables both the log gas consumption and the log grid electricity consumption.

Table 4: Effects of retrofits on gas and electricity, by solar panel availability

	Dependent: log gas		Dependent: log electricity	
	No solar	Yes solar	No solar	Yes solar
Retrofit (year $\geq$ 2)	-0.223 (0.003)***	-0.237 (0.004)***	0.010 (0.004)***	-0.286 (0.007)***
$\times$ Below poverty line	-0.005 (0.015)	0.053 (0.014)***	0.059 (0.013)***	0.033 (0.028)
$\times$ Below 100% soc.min.	0.054 (0.019)***	0.063 (0.025)**	0.037 (0.022)*	0.037 (0.039)
$\times$ Below 130% soc.min.	0.013 (0.008)*	0.041 (0.009)***	0.014 (0.008)*	0.006 (0.016)
$\times$ Below 150% soc.min.	0.011 (0.007)	0.027 (0.008)***	0.010 (0.007)	-0.011 (0.015)

Notes: The table shows estimates of Equation (12) for 10 separate regressions. Coefficients reported are two- and three-way interactions. The symbol $\times$ indicates an effect as compared to the reference level (non-poor). The combination of the column and row name indicates the interaction (e.g. below poverty line $\times$ Yes solar). The dependent variable is log of gas or log of electricity. Each regression includes controls, household fixed-effects and year fixed-effects. The sample size is 13409 treated and 110891 control units. Standard errors in parentheses are clustered at household level. Statistical significance: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Availability of solar electricity does not seem to have much effect on gas consumption: the gas savings after retrofit are practically the same in the solar and no-solar dwellings. In terms of grid electricity consumption however, in no-solar dwellings we observe a small increase equal to 1 to 4%. A likely explanation for this is the additional electricity demand due to the installation of mechanical ventilation that is necessary to ensure sufficient air quality in well-insulated dwellings. The solar-dwellings, on the other hand, reduce grid electricity consumption by almost 30% on average, which is in line with the literature. Concluding, we do not find convincing evidence of large substitution effects between gas and electricity for heating purposes after insulation.[18]

Results of Table 4 provide additional insights into the working of the income and substitution effects after insulation retrofits. The income effect implies that households can use the monetary savings from retrofits to increase their consumption of other goods. We do not observe evidence of this for electricity consumption. The substitution effect implies that when electricity becomes more affordable, households may start to obtain part of their thermal comfort through electricity instead of gas, effectively reducing further their gas consumption (e.g. by buying electric space heaters). Households may perceive that solar panels make electricity more affordable. Our results however show that solar panel installations hardly change the effect of insulation retrofits on gas use.

4.4 Robustness checks

We have subjected the results of Table 3 to a range of sensitivity analyses, see Appendix D. First, we re-estimate the model of Equation (12) for various subsamples, allowing the retrofit effect to differ by: (i) year in which insulation retrofit took place (2017, 2018 and 2019), Table D1; (ii) pre-retrofit energy-efficiency as defined by the energy label (C, D, E, F, G), Table D2; (iii) socio-economic characteristics of households, Table D3; (iv) pre-retrofit gas use quintile, Table D4. Results are robust across all the year and energy label subsamples. The effect of insulation however differs by household type. For instance, singles reach larger savings, while households with migration background reduce gas consumption less than average. The effect also differs by pre-retrofit gas use: households with low gas demand experience almost half lower savings than average. The low-income specific response to insulation is however robust in all the subsamples. In the next Section we will dig deeper into the heterogeneities in the effect.

[18]We note that our data only include 4 years after retrofit. It might be that such substitution effects take a longer time to manifest themselves. On the other hand, substituting gas for electricity often requires an investment upfront (e.g. buying an electric space heater). Low-income households we are studying might face binding credit constraints prohibiting such investments.

We also rerun the model using alternative model specifications, see Appendix E. These inlcude: (i) various functional form specifications to include the retrofit intensity in the model, Table E1; (ii) Sun and Abraham estimator to account for a possible bias due to the staggered treatment (Callaway and Sant'Anna (2021); Sun and Abraham (2021)), Table E2; (iii) including group-specific time trends, Table E3. The analysis suggests that the specific low-income response shown in Table 3 holds under numerous modelling specifications.

5 Heterogeneity in treatment

5.1 Determinants of the size of gas savings

In the previous Section we documented up to one third lower savings from retrofits for households on the very left tail of the income distribution. Below we provide insight into other possible heterogeneities in the treatment effect and their size. We aim to compare the importance of socio-economic and dwelling characteristics in explaining the size of the gas savings from insulation retrofits. We start by running Equation (13) for all the covariates in Table 1.

$$g_{i,t} = (1 + \kappa_1 h_i)R_{i,t}\alpha + (1 + \kappa_2 h_i)R_{i,t}\beta S_i + \delta X_{i,t} + \gamma_i + \phi T_t + u_{i,t}. \quad (13)$$

Here, h_i is a heterogeneity covariate, included in a standardized form to make the estimates mutually comparable. The covariate h is time-independent and takes 2016 values, the year before any retrofits occur in our sample. We substitute different covariates in the model one by one.

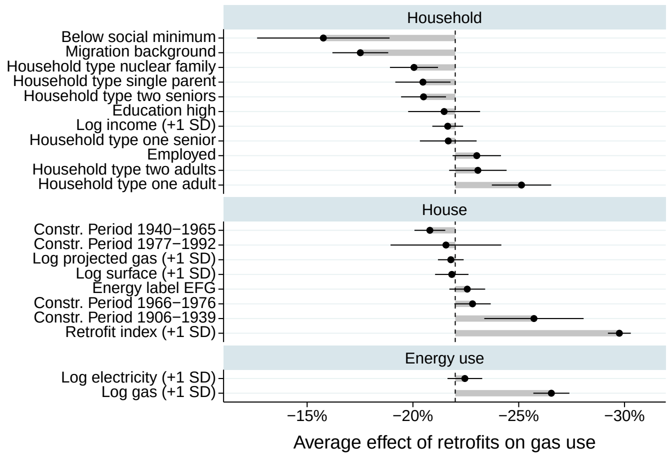
Figure 6 shows estimates of the main effect of the interaction terms. For continuous variables - indicated with "+1 SD" - the Figure shows the average retrofit effect when the respective interaction variable increases with one standard deviation (SD).[19] For the remaining (binary) variables, the Figure shows the average effect in the respective subgroup from Equation (12).

First, notice that the retrofit intensity is the most important determinant of the heterogeneity: one standard deviation increase leads to a 8 percentage point larger treatment effect. Housing characteristics do not seem to play a large role,[20]

[19]We obtain this effect with the transformation $(h_i - mean(h_i))\,/sd\,(h_i) - 1$

[20]Note that if we do not control for retrofit size in the model, low energy label E,F,G of

Figure 6: Covariate importance in explaining gas savings



Notes: Grey bars show the estimated coefficients of the interaction terms $\kappa_1 \alpha$ in Equation (13), black lines indicate the 95% confidence interval. For continuous variables - indicated with "+1 SD" - the Figure shows the average retrofit effect when the respective interaction variable increases with one standard deviation. For the remaining (binary) variables, the Figure shows the average effect in the respective subgroup. All coefficients should be interpreted for an "average" retrofit (except for the retrofit intensity coefficient).

while socio-economics and especially pre-retrofit gas usage do. Large savers are those with a high pre-retrofit gas use and single households (4 respectively 3 percentage point larger savings), while households with migration background and families save between 6 respectively 3 percentage point less than average. While the lowest incomes are small savers, in other income strata income does not affect the treatment effect much.

For the covariates from Figure 6 that show important effect on retrofit effect (pre-retrofit gas use, type of household, migration background and energy label), we check the robustness of low-income response. Appendix D reports the heterogeneity by households type, migration background (binary indicator taking value 1 when all household members are born outside the Netherlands, first and second generation.), energy label and pre-retrofit gas use. The specific response of low-income is robust across these covariates.

a house becomes a good predictor of the savings, with a 4 percentage point larger treatment effect than average.

5.2 Causal forest

To offer additional insights and robustness checks on the heterogeneity in the size of the treatment effect, we exploit the causal forest machine learning approach (Wager and Athey, 2018). The algorithm identifies (predicts) treatment effects *by observation*, based on a non-linear estimation using a large number of predictors.[21] It also yields information about the contribution of each attribute in growing the forest from which the treatment effects are derived.

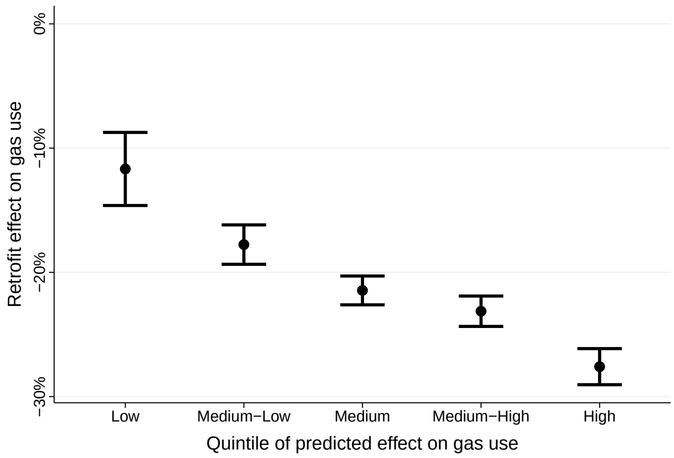
We use causal forest to divide observations into quintiles of predicted treatment effect. Then, we estimate econometrically the heterogeneity between the ATE in each of the quintiles, see Figure 7 for the estimated coefficients. Note that these are *not* the forest predictions, but coefficients from a two-way fixed effect estimation of Equation (12), where $J = 5$ are the five treatment effect quintiles to which causal forest assigned the observations.

Figure 7 allows to test whether there are other heterogeneity determinants besides those identified in Section 5.1. The result confirms the existence of considerable heterogeneity in the effect of insulation retrofits on gas use. Conditional on an average retrofit size, gas savings range from 11% (lowest quintile) to 29% (highest quintile), an almost threefold difference. Figure F1 in Appendix F reports the covariate importance from the forest. Same covariates as distinguished above account for most of heterogeneity, with on the top the pre-retrofit gas consumption.[22]

[21] No ex-ante assumptions about these drivers need to be imposed as is the case with the conventional regression techniques with interaction effects. Under the unconfoundeness assumption, causal forest estimates treatment effects that are consistent and asymptotically Gaussian distributed (Athey et al., 2019; Wager and Athey, 2018).

[22] Our linear model Section 5.1 found that house characteristics such as surface and projected gas use do not explain heterogeneity in treatment effect while Figure F1 causal forest suggests that house characteristics are important. This difference is due to the retrofit intensity being included in our linear model but not in causal forest. House characteristics become important predictors of insulation effects when the retrofit intensity is not controlled for in our linear model.

Figure 7: Heterogeneity of retrofit effects: combination of causal forest and fixed effects panel regression



Notes: This figure reports coefficients from Equation (12) with $J = 5$. Coefficients are two-way interaction effects of the treatment with the causal forest quintile dummies. Bars indicate 95% confidence interval. Standard errors clustered at household level.

6 Welfare effects

In this section, we develop a calibrated version of the consumer choice model Equations (1) and (2). The calibrated model is then used to assess the welfare effects of the heating efficiency retrofits that took place in the Dutch social housing sector between 2017 and 2019.

6.1 Model calibration

To calibrate the model, we need to choose a number of exogenous parameters. First, the gas price is set to the 2016 level of consumer gas price according to Statistics Netherlands, p_g = € 0.65 per cubic meter. Second, the income distribution is approximated by a set of ten income deciles, which are defined as follows. We split our study sample of 124,300 households (treatment and control) into ten deciles $d = 1, \ldots, 10$, based on household income expressed in percentage of the social minimum. For each decile d, we assign the median disposable income w_d of that decile. Third, we derive from data the median gas consumption g_d before the home heating efficiency upgrade, this for each income decile d. Finally, the empirically estimated retrofit effect r_d from Equation (12) is assigned to each income decile. This retrofit effect is measured as the change in cubic meters of gas consumption due to the heating efficiency upgrade.

We assume that all households face the same values of the preference parameters $(\bar{\theta}, \sigma)$, natural temperature θ_0, pre-retrofit heating efficiency $q = q_L$ and post-retrofit heating efficiency $q = q_H$. We compute calibrated values of these parameters $(q_L, q_H, \theta_0, \bar{\theta}, \sigma)$ by using a non-linear least squares method that minimizes the weighted sum S:

$$S = W_1 \sum_d (g^*(q_L, w_d) - g_d)^2 + W_2 \sum_d (g^*(q_H, w_d) - g^*(q_L, w_d) - r_d)^2,$$

subject to the following constraints:

$$q_L > 0, q_H > 0, \theta_0 \geq 10, \bar{\theta} \in [18, 24], \sigma \geq 0.$$

The weights $W_1 = \mathrm{SD}\,(g_d)^{-2}$ and $W_2 = \mathrm{SD}\,(r_d)^{-2}$ are chosen in such a way that both sums in S have comparable scales.

Effectively, we choose parameters $(q_L, q_H, \theta_0, \bar{\theta}, \sigma)$ to match the observed pre-retrofit gas consumption g_d in 2016 with the model prediction $g^*(q_L, w_d)$ and the empirically estimated change r_d in gas consumption due to retrofit with the model prediction $g^*(q_H, w_d) - g^*(q_L, w_d)$. The calibration balances the goodness of fit before and after the retrofit for ten income deciles d.

Table 5 reports the exogenous and calibrated parameter values. Note that the calibrated value of the elasticity of substitution parameter σ is very close to unity, which implies that the household utility is close to the Cobb-Douglas utility specification.

6.2 Model validation

We perform a number of validation tests for the calibrated model. First, Figure 8 plots the observed g_d and the calibrated gas usage $g^*(q_L, w_d)$ (left panel), as well as the estimated r_d and the calibrated retrofit effect on gas consumption $(g^*(q_H, w_d) - g^*(q_L, w_d))$ (right panel), per income decile d. Visual inspection suggests a good overall fit, although we note that the absolute size of the calibrated retrofit effect is overestimated for the lowest income decile and underestimated for the highest income decile.

Table 5: Calibrated parameters

Description	Parameter	Value
Exogenously chosen parameters		
Price of gas (euro/m^3)	p_g	0.65
Price of other consumption	p_x	1.00
Calibrated parameters		
Indoor temperature at g=0 (^{o} C)	θ_0	11.10
Elasticity of substitution	σ	1.00
Satiety level of thermal comfort (^{o} C)	$\bar{\theta}$	23.80
Energy efficiency before retrofit (^{o} C/m^3)	q_L	$\frac{1}{103}$
Energy efficiency after retrofit (^{o} C/m^3)	q_H	$\frac{1}{80}$

Second, we aim to validate the calibrated increase in the optimal temperature $\theta^*(q_H, w_d) - \theta^*(q_L, w_d)$ that shows how much households re-optimize their consumption towards a higher thermal comfort after the heating efficiency improvement. The temperature increase generated by the model ranges from 0.1 to 0.6 degrees Celsius, this value decreases with income. The values are consistent with the earlier findings from small scale empirical studies, e.g., Fisk et al. (2020).

Third, we look at the share of potential gas savings that is foregone because households re-optimize consumption towards higher thermal comfort. This so-called rebound effect can be computed as the monetary value of the thermal comfort improvement divided by the retrofit income effect:

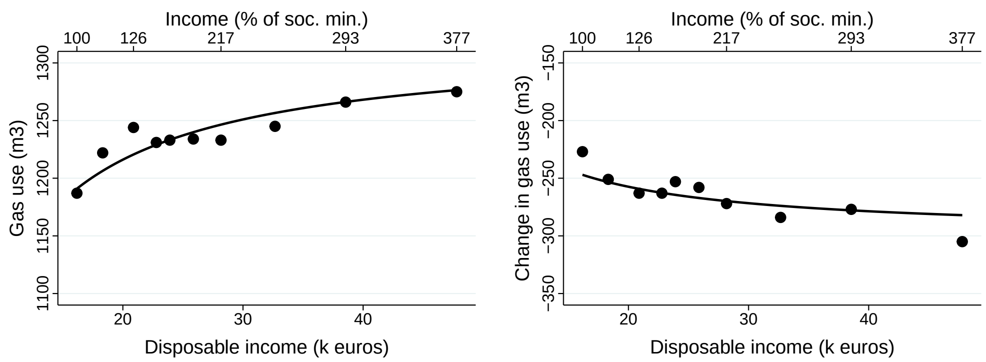
$$\frac{-\Delta_\theta^S}{-\Delta^S}.$$

The average rebound effect over the income deciles amounts to 5.3% (in 2016 prices) in our calibrated model. This value compares well to recent findings on the size of the rebound, see, e.g., Christensen et al. (2023).

Fourth, we compute the implied price elasticity of gas consumption and compare it with the values form the existing literature. Resulting price elasticities range from -0.04 to -0.27, which is in line with the earlier findings, see, e.g., Asche et al. (2008).

Figure 8: Observed versus calibrated outcomes

(a) Pre-retrofit gas use (m^3) (b) Retrofit effect on gas use (m^3)



Notes: The dots are observed respectively estimated values, the lines are outcomes of the calibrated model. Panel (a) depicts the median gas use, across income deciles. Panel (b) depicts the average retrofit effect on gas use across income deciles. Both figures use the data from the baseline sample of 124,300 households (treatment + control).

6.3 Welfare outcomes

The calibrated model is applied to value the benefits from the heating efficiency retrofits performed in the Dutch social housing in 2017-2019. The retrofit is described as a change from q_L (low heating efficiency) to q_H (high heating efficiency), see the values in Table 5. Two scenario's are defined: (i) a reference scenario, for which the model was calibrated (Table 5); (ii) a counterfactual, in which the gas price is set to the high level $p_g = 1.36$ euro/m$_3$ it reached in 2022. Table 6 reports the retrofit outcomes for the reference and the counterfactual, for three income groups: low (below the social minimum), average (median of the income distribution in our study sample) and high (75 percentile of the same income distribution).

Columns (2)-(3) of Table 6 describe the effects of the retrofits on the households' optimal consumption of temperature and natural gas. Note first that the effect on the temperature is larger and the effect on gas consumption is smaller in the counterfactural as compared to the reference. Reason is that higher gas prices make heating services more expensive, so that more households choose for low thermal comfort and uncomfortably low temperatures when $q = q_L$. The resulting high marginal utility of one degree temperature increase leads to larger adjustments in temperature consumption after the retrofit. Lower gas savings follow. Second, in line with the theoretical insights of Section 2, low income households realize considerably larger temperature increase and smaller absolute reduction in gas use than higher incomes. For example, for the poor, the temperature increase after retrofit reaches 0.3 grades in the reference scenario and 0.7 grades in the counterfactual; this is thrice as much as for the higher income households.

Table 6: Welfare outcomes

Income	Change in thermal comfort		Private benefits (euro)			
	Δ Temp.	Δ Gas	Slutsky valuation			Hicksian val.
	$^{\circ}C$	m^3	$-\Delta^S$	$-\Delta^S_\theta$	$-\Delta^S_x$	CV
(1)	(2)	(3)	(4)	(5)	(6)	(7)
	In prices 2016 (0.65 euro/m^3)					
low	0.30	-245	175	15	159	177
average	0.19	-265	182	10	172	183
high	0.10	-280	187	5	182	188
	In prices 2022 (1.36 euro/m^3)					
low	0.65	-182	320	71	249	329
average	0.41	-225	352	45	307	358
high	0.22	-258	377	24	353	381

Notes: The Table reports the effects of the average retrofit in our data, computed with the calibrated model. This is done for two scenarios (reference with gas prices of 2016 and counterfactual with gas prices of 2022) and for three income groups (low, average and high, respectively 16keuro, 24keuro and 43keuro in disposable yearly income). Columns (2) and (3) document the changes in consumption of temperature respectively other goods, following the retrofits. Column (4) reports the valuation of the private benefits of the retrofits using the Slutsky compensation. Column (7) reports the same valuation using the Hicksian compensation (compensating variation). Columns (5) and (6) decompose the Slutsky compensation into the parts that arise due to the change in temperature consumption respectively the change in consumption of other goods. The Table shows yearly outcomes.

Columns (4)-(7) of Table 6 report the private welfare gains from the retrofit: total and decomposed into the benefits of increased temperature respectively of other consumption, as derived in Equations (7) to (9). The comfort benefits from temperature increase make a considerable part of the total gains, more so for the low income households. In the reference scenario with low gas prices, the comfort benefits amount to 9% of the total utility increase for the social minima and only 3% for the richer households. In the counterfactual with high gas prices, the comfort benefits make 22% respectively 6% of the total gains for the two groups. The driving force behind these differences is that the marginal benefit of a one degree temperature increase is higher for the poor and when gas prices are high, so that residents spend a larger share of the potential gas savings on comfort increase.

The benefits are distributed unevenly among income levels: the gains for the poor are 6% to 14% lower in comparison with the higher income peers. This is intuitive. Because low income households consume less gas before retrofit, their potential savings from heating efficiency upgrades are also smaller. By trading off potential gas savings for a comfort increase, households improve their welfare, but the resulting gains still stay below the benefits that higher incomes can obtain. From a policy perspective, this insight points at a trade-off that accompanies policies subsidizing heating efficiency improvements for low income households. Lower gas savings of the poor translate one-to-one to lower environmental (CO_2) benefits. Alleviating poverty and increasing living comfort for the poor comes at the expense of lower environmental benefits.

It is instructive to compute the net present value of the discussed welfare benefits and to compare it with the costs of the heating efficiency retrofits in the social housing. We use a discount rate of 2.25%, which is prescribed for the Dutch cost-benefit analyses and take a time horizon of 50 years, which is technically feasible for home insulation investments. The net present value (NPV) of the private welfare benefits in the reference scenario with low gas prices amounts to 5.5 respectively 5.8 thousand euro NPV per household, for low and higher incomes respectively. In the counterfactual with high gas prices it rises to 10 respectively 12 thousand euro per household. We also test the sensitivity of the result for a shorter, 30 years time horizon, which may be more realistic given the the Netherlands' goal to become greenhouse gas-neutral in 2050. The resulting NPV's are then one quarter lower. Mot et al. (2023) reports the average cost of insulation retrofits in Dutch social housing to equal 11 thousand euro per dwelling in 2020. Our analysis suggests therefore that the benefits from gas savings and comfort increase likely fall short of the costs of the heating efficiency upgrades, even at high gas prices. Obviously, one should expect heating efficiency retrofits to yield other welfare benefits as well, besides those studied in our paper. These are, among other things: health improvement due to reduced exposure to draught and extreme temperatures (Maidment et al. (2014)), climate benefits due to lower CO_2 emission, poverty alleviation gains (Banerjee et al. (2021)), etc. In computing the societal returns to the heating efficiency investments, these benefits need to be taken into account as well.

7 Conclusion

Many countries subsidize energy efficiency upgrades in low income housing. The goal of these policies is twofold: reducing CO_2 emissions and alleviating poverty. This paper showed that the two goals are competing: prioritizing energy efficiency upgrades for the poor may come at the expense of lower environmental benefits. We conducted a large-scale evaluation of the effects of heating efficiency retrofits that inhabitants of Dutch social (low-income) housing received from their housing providers in 2017-2019. Our study followed 125,000 households during 2014-2021, leveraging considerable variation in income in the sample (from below the social minimum to above the population median). We exploited a unique conditional random assignment to retrofit in the Dutch social housing sector in the study years. The evaluation used quasi-experimental two-way fixed effects econometrics on the one hand, and, on the other hand, a calibrated microeconomic consumer choice model, in which people choose between thermal comfort and other goods. We specifically focused on the retrofit-induced benefits from lower gas use and from higher comfort.

Four primary findings of our study should be emphasized. First, we documented empirically that lowest-income households realize considerably smaller than average natural gas savings from home heating-efficiency retrofits. The quasi-experimental estimates suggest that, after a heating efficiency upgrade, the social minima reduced their gas consumption by 16%, while the average gas savings in the sample were 22%. In absolute terms, this means up to one third lower gas and environmental savings for the poor, when compared with their more well-off peers. Second, this heterogeneity in gas savings can be explained from income-specific behavioural responses to the retrofit. Our calibrated consumer choice model suggests that the poor reinvest up to 20% of the potential monetary savings from a heating efficiency upgrade into thermal comfort improvement, i.e. a higher temperature in house. The more well-off peers only reinvest 5%, because their thermal comfort was already high before retrofit. Third, even accounting for the benefits from comfort improvement, the monetary value of the private welfare gain from retrofits is lower for the poor, as compared to their richer peers. Fourth, also when gas prices are high, the size of the studied private welfare benefits falls short of the costs of an average heating efficiency retrofit.

Our study provides novel evidence into the benefits and trade-offs of using heating efficiency retrofits as an instrument to alleviate poverty. We also contribute to the literature and public discussion about the returns to such policies.[23] Obviously, the welfare effects we computed are an underestimation of the society's benefits due to the heating efficiency upgrades. For instance, we looked from the household perspective only and fail to recognize the benefits of greenhouse gas and local pollutant emissions reductions. Further, private benefits of households involve more aspects than the financial savings and comfort improvement we included in the analysis. Among other things, insulation-induced reduction in draught and extreme temperatures in house will likely have a positive impact on the inhabitants' health (Maidment et al. (2014)). Moreover, specifically for the left tail of the income distribution, additional societal gain may be achieved through poverty alleviation (Banerjee et al. (2021)). In this paper, we find the environmental benefits and monetary savings from reduced gas consumption to be smaller for the poorest. Comfort gains are however higher, so will be poverty alleviation benefits and - possibly - the health effects. Further research into these latter aspects is desirable to facilitate a cost-benefit test of heating efficiency upgrades by income group. Our paper suggests a methodology to make the welfare trade-offs explicit and quantify them.

[23] For instance, Fowlie et al. (2018) found negative returns for weatherization retrofits in Michigan, US, despite their much lower cost of 4585 dollar per household. The Michigan retrofits involved attic and wall insulation, infiltration reduction and furnace replacement.

References

Adan, H. and Fuerst, F. (2016). Do energy efficiency measures really reduce household energy consumption? A difference-in-difference analysis. *Energy Efficiency*, 9(5):1207–1219.

Allcott, H. and Greenstone, M. (2017). Measuring the Welfare Effects of Residential Energy Efficiency Programs. page 97.

Angrist, J. D. (2008). Mostly Harmless Econometrics: An Empiricist's Companion. page 290.

Asche, F., Bjarte Nilsen, O., and Tveteras, R. (2008). Natural Gas Demand in the European Household Sector. *The Energy Journal*, 29(3).

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2).

Avanzini, M., Pinheiro, M. D., Gomes, R., and Rolim, C. (2022). Energy retrofit as an answer to public health costs of fuel poverty in Lisbon social housing. *Energy Policy*, 160:112658.

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.

Banerjee, A., Duflo, E., and Sharma, G. (2021). Long-term effects of the targeting the ultra poor program. *American Economic Review: Insights*, 3(4):471–86.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Christensen, P., Francisco, P., Myers, E., and Souza, M. (2023). Decomposing the Wedge between Projected and Realized Returns in Energy Efficiency Programs. *Review of Economics and Statistics*, 105(4):798–817.

Davis, L. W., Fuchs, A., and Gertler, P. (2014). Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico. *American Economic Journal: Economic Policy*, 6(4):207–238.

Davis, L. W., Martinez, S., and Taboada, B. (2020). How effective is energy-efficient housing? Evidence from a field trial in Mexico. *Journal of Development Economics*, 143:102390.

Doremus, J. M., Jacqz, I., and Johnston, S. (2022). Sweating the energy bill: Extreme weather, poor households, and the energy spending gap. *Journal of Environmental Economics and Management*, 112:102609.

Eurostat (2023). Disaggregated final energy consumption in households. https://ec.europa.eu/eurostat/databrowser/view/{NRG}_D_HHQ__custom_8936568/default/table?lang=en.

Fisk, W. J., Singer, B. C., and Chan, W. R. (2020). Association of residential energy efficiency retrofits with indoor environmental quality, comfort, and health: A review of empirical data. *Building and Environment*, 180:107067.

Fowlie, M., Greenstone, M., and Wolfram, C. (2018). Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program*. *The Quarterly Journal of Economics*, 133(3):1597–1644.

Gerarden, T., Newell, R. G., and Stavins, R. N. (2015). Deconstructing the Energy-Efficiency Gap: Conceptual Frameworks and Evidence. *American Economic Review*, 105(5):183–186.

Gillingham, K., Keyes, A., and Palmer, K. (2018). Advances in Evaluating Energy Efficiency Policies and Programs. *Annual Review of Resource Economics*, 10(1):511–532.

Hammerle, M. and Burke, P. J. (2022). From natural gas to electric appliances: Energy use and emissions implications in Australian homes. *Energy Economics*, 110:106050.

Hancevic, P. I. and Sandoval, H. H. (2022). Low-income energy efficiency programs and energy consumption. *Journal of Environmental Economics and Management*, 113:102656.

Hennessy, J., Dasgupta, T., Miratrix, L., Pattanayak, C., and Sarkar, P. (2016). A Conditional Randomization Test to Account for Covariate Imbalance in Randomized Experiments. *Journal of Causal Inference*, 4(1):61–80.

Knittel, C. R. and Stolper, S. (2021). Machine Learning about Treatment Effect Heterogeneity: The Case of Household Energy Use. *AEA Papers and Proceedings*, 111:440–444.

Liang, J., Qiu, Y., James, T., Ruddell, B. L., Dalrymple, M., Earl, S., and Castelazo, A. (2018). Do energy retrofits work? Evidence from commercial and residential buildings in Phoenix. *Journal of Environmental Economics and Management*, 92:726–743.

Liang, J., Qiu, Y., and Xing, B. (2021). Social Versus Private Benefits of Energy Efficiency Under Time-of-Use and Increasing Block Pricing. *Environmental and Resource Economics*, 78(1):43–75.

Maidment, C. D., Jones, C. R., Webb, T. L., Hathway, E. A., and Gilbertson, J. M. (2014). The impact of household energy efficiency measures on health: A meta-analysis. *Energy Policy*, 65:583–593.

McCoy, D. and Kotsch, R. A. (2021). Quantifying the Distributional Impact of Energy Efficiency Measures. *The Energy Journal*, 42(01).

Milne, G. and Boardman, B. (2000). Making cold homes warmer: the effect of energy efficiency improvements in low-income homes A report to the Energy Action Grants Agency Charitable Trust. *Energy Policy*, 28(6-7):411–424.

Ministry of the Interior and Kingdom Relations (2012). Convenant Huursector (https://www.rijksoverheid.nl/documenten/convenanten/2012/06/28/convenant-huursector).

Mot, E., Schippers, V., Phan, N., Schulenberg, R., Griffioen, E., Mulder, P., Tigchelaar, C., and Zwamborn, A. (2023). Inkomenseffecten van woningisolatie naar de isolatiestandaard. Technical report, CPB and TNO.

Ossokina, I. V., Kerperien, S., and Arentze, T. A. (2021). Does information encourage or discourage tenants to accept energy retrofitting of homes? *Energy Economics*, 103:105534.

Ossokina, I. V., van Ommeren, J., and van Mourik, H. (2022). Do highway widenings reduce congestion? *Journal of Economic Geography*, lbac034.

Peñasco, C. and Anadón, L. D. (2023). Assessing the effectiveness of energy efficiency measures in the residential sector gas consumption through dynamic treatment effects: Evidence from England and Wales. *Energy Economics*, 117:106435.

Rubin, D. B. (2001). Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. *Health Services and Outcomes Research Methodology*, 2(3/4):169–188.

Saunders, H. D., Roy, J., Azevedo, I. M., Chakravarty, D., Dasgupta, S., de la Rue du Can, S., Druckman, A., Fouquet, R., Grubb, M., Lin, B., Lowe, R., Madlener, R., McCoy, D. M., Mundaca, L., Oreszczyn, T., Sorrell, S., Stern, D., Tanaka, K., and Wei, T. (2021). Energy Efficiency: What Has Research Delivered in the Last 40 Years? *Annual Review of Environment and Resources*, 46(1):135–165.

Sdei, A., Gloriant, F., Tittelein, P., Lassue, S., Hanna, P., Beslay, C., Gournet, R., and McEvoy, M. (2015). Social housing retrofit strategies in England and France: A parametric and behavioural analysis. *Energy Research & Social Science*, 10:62–71.

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

Sovacool, B. K. (2015). Fuel poverty, affordability, and energy justice in England: Policy insights from the Warm Front Program. *Energy*, 93:361–371.

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, 25(1).

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Tibshirani, J., Athey, S., Sverdrup, E., and Wager, S. (2022). grf: Generalized Random Forests, R package version 2.2.1.

Van Ommeren, J. N. and Van Der Vlist, A. J. (2016). Households' willingness to pay for public housing. *Journal of Urban Economics*, 92:91–105.

Wager, S. and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Walker, S., Lowery, D., and Theobald, K. (2014). Low-carbon retrofits in social housing: Interaction with occupant behaviour. *Energy Research & Social Science*, 2:102–114.

Webber, P., Gouldson, A., and Kerr, N. (2015). The impacts of household retrofit and domestic energy efficiency schemes: A large scale, ex post evaluation. *Energy Policy*, 84:35–43.

A Appendix Utility maximization problem solutions

In this Appendix we offer the Proof of Proposition 1. Write out the Lagrangian for the household utility maximization problem, UMP:

$$L = \left(x^{\frac{\sigma-1}{\sigma}} + \left((2\bar{\theta}-\theta)\theta\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} + \lambda\left(w - x - \frac{\theta-\theta_0}{q}p_g\right)$$

Due to the strict monotonicity of $u(x,\theta)$ w.r.t. x, it follows that $\lambda > 0$ and the F.O.C.s for an interior solution are:

$$\begin{cases} 0 = \left(x^{\frac{\sigma-1}{\sigma}} + \left((2\bar{\theta}-\theta)\theta\right)^{\frac{\sigma-1}{\sigma}}\right)^{-\frac{1}{\sigma}} x^{-\frac{1}{\sigma}} - \lambda \\ 0 = 2\left(x^{\frac{\sigma-1}{\sigma}} + \left((2\bar{\theta}-\theta)\theta\right)^{\frac{\sigma-1}{\sigma}}\right)^{-\frac{1}{\sigma}} \left((2\bar{\theta}-\theta)\theta\right)^{-\frac{1}{\sigma}}(\bar{\theta}-\theta) - \frac{p_g}{q}\lambda \\ 0 = w - x - \frac{\theta-\theta_0}{q}p_g \end{cases}$$

The first two equations imply:

$$x = (2\bar{\theta}-\theta)\theta\left(\frac{p_g}{2q(\bar{\theta}-\theta)}\right)^{\sigma}.$$

Then, the third equation implies that θ^* satisfies Equation (4), which can be written as

$$F(\theta^*, w, q) = 0, \tag{14}$$

where

$$F(\theta, w, q) \stackrel{\text{def}}{=} w - (2\bar{\theta}-\theta)\theta\left(\frac{p_g}{2q(\bar{\theta}-\theta)}\right)^{\sigma} - \frac{\theta-\theta_0}{q}p_g. \tag{15}$$

It can be seen that for $w \geq 0$, $q > 0$, and $\theta \in [\theta_0, \bar{\theta})$, F increases with w and q and decreases with θ, because its derivatives are:

$$\begin{aligned} F_\theta &= -\left(\frac{p_g}{2q}\right)^{\sigma}\left(2(\bar{\theta}-\theta)^2 + \sigma(2\bar{\theta}-\theta)\theta\right)\left(\bar{\theta}-\theta\right)^{-\sigma-1} - \frac{p_g}{q} < 0, \\ F_w &= 1 > 0, \\ F_q &= \frac{\sigma}{q}\left(\frac{p_g}{2q}\right)^{\sigma}(2\bar{\theta}-\theta\theta)(\bar{\theta}-\theta)^{-\sigma} > 0. \end{aligned}$$

Therefore, if Equation (14) has a solution $\theta^*(w,q)$, it is monotone increasing.

Since $F(\theta_0, \underline{w}, q) = 0$, where $\underline{w}$ is defined in Equation (3), it follows that for $w < \underline{w}$, Equation (14) has no solution satisfying $\theta \geq \theta_0$. For such low income levels, the UMP has a corner solution in which $\theta^* = \theta_0$ and $x^* = w$. This proves part 1 of the proposition.

For $w > \underline{w}$, Equation (14) defines a unique solution $\theta^*(w, q)$. The solution always exists because for any income $w > \underline{w}$:

$$F(\theta_0, w, q) > 0,$$
$$\lim_{\theta \uparrow \bar{\theta}} F(\theta, w, q) = -\infty,$$

and F continuously decreases with θ. This proves part 2 of the proposition. The monotonicity properties of θ^* follow form the monotonicity properties of F:

$$\theta^*_w = -\frac{F_w}{F_\theta} = -\frac{1}{F_\theta} > 0,$$
$$\theta^*_q = -\frac{F_q}{F_\theta} > 0.$$

The monotonicity of x^* can be seen from:

$$x^*_w = 1 - \frac{p_g}{q}\theta^*_w > 0.$$

This proves part 3 of the proposition. Finally, since F in unbounded in w and θ, the solution θ^* approaches $\bar{\theta}$ when w increases unboundedly. Similarly, for any $\theta < \bar{\theta}$ and $w > 0$, $F(\theta, w, q)$ converges to $w > 0$ when q increases unboundedly. Therefore, in the limit, it must be that the solution θ^* converges to $\bar{\theta}$. This proves part 4 of the proposition.

B Appendix Identification

Table B1 reports the estimation results from an OLS regression of the retrofit intensity on pre-retrofit dwelling, income and energy consumption characteristics of the households. The sample includes all 13409 retrofitted houses from our baseline sample. The results indicate that among energy use and socio-economics, only gas use and income have a statistically significant effect on the retrofit intensity. However, this correlation is very small: a one standard deviation increase in gas use or income leads to a retrofit intensity up to 0.054 smaller, i.e. projected gas savings 0.83 percentage points smaller - this is negligible as compared to the 43% average projected gas savings.

Table B2 reports the estimation results from a logit model relating the probability of moving house to the retrofit incidence, observed pre-retrofit socio-economic, dwelling characteristics and energy use ($\ln \frac{p_i}{1-p_i} = \beta_0 + \beta_1 R_i + \beta X_i + \epsilon_i$; X_i are observed controls in 2016, p_i is the probability of relocation and R_i indicates dwellings treated between 2017 and 2019). Numeric variables are standardised so that the inverse log-odds of the coefficient β_0 "Constant" can be interpreted as the relocation rate for an average household in the control group (e.g. $\exp(-1.129)/(1+\exp(-1.129)) = 24.4\%$). The sample consists of all households in our study sample for which there is no missing observed characteristics. In columns (3) and (4) treated dwellings are matched to 3 control dwellings on observed characteristics. The results indicate that the relocation rate after treatment remains largely unchanged, e.g. $\exp(-1.129 - 0.054)/(1 + \exp(-1.129 - 0.054)) = 23.5\%$ vs. 24.4% for the non-treated.

Table B1: Determinants of retrofit intensity

	(1)	(2)
(Intercept)	-0.004 (0.009)	0.110* (0.061)
Log gas (standardized)	0.046*** (0.009)	-0.054*** (0.007)
Log income (standardized)	-0.036*** (0.009)	-0.035*** (0.010)
No. children (standardized)		-0.050 (0.067)
No. persons (standardized)		0.039 (0.074)
No. persons squared (standardized)		0.052 (0.091)
No. senior squared (standardized)		-0.024 (0.039)
No. children squared (standardized)		-0.018 (0.053)
No. seniors (standardized)		0.024 (0.049)
No. females (standardized)		0.019 (0.021)
No. females squared (standardized)		-0.019 (0.021)
Log surface (standardized)		-0.241*** (0.008)
Log projected gas (standardized)		0.516*** (0.008)
Log construction year (standardized)		0.214*** (0.016)
Employed 0/1		0.031* (0.018)
Solar panels 0/1		0.153** (0.064)
Boiler changed 0/1		0.311*** (0.025)
Household type one adult (ref)		
Household type nuclear family 0/1		0.076 (0.060)
Household type one senior 0/1		0.039 (0.046)
Household type single parent 0/1		0.149*** (0.051)
Household type two adults 0/1		0.031 (0.049)
Education high (ref)		
Education low 0/1		0.025 (0.024)
Education medium 0/1		0.031 (0.025)
Education unknown 0/1		0.009 (0.025)
Energy label C (ref)		
Energy label D 0/1		0.207*** (0.018)
Energy label E 0/1		0.459*** (0.019)
Energy label F 0/1		0.514*** (0.026)
Energy label G 0/1		0.733*** (0.028)
Constr. Period 1906-1940 (ref)		
Constr. Period 1940-1965 0/1		-0.616*** (0.051)
Constr. Period 1966-1976 0/1		-0.418*** (0.066)
Constr. Period 1977-1992 0/1		-0.794*** (0.085)
Num.Obs.	13401	13401
R2	0.003	0.443
R2 Adj.	0.003	0.442

Notes: The tables shows estimates of two separate OLS regressions. The dependent variable is the retrofit intensity. The independent variables are all pre-retrofit observed controls. Statistical significance: *p<0.1; **p<0.05; ***p<0.01.

Table B2: Effect of retrofits on household's relocation

	Dependent variable: household relocated before 2021 (0/1)			
	Logit models			
	(1)	(2)	(3)	(4)
Constant (β_0)	-1.063^{***}	-1.052^{***}	-1.129^{***}	-1.124^{***}
	(0.013)	(0.003)	(0.029)	(0.007)
Retrofit (β_1)	-0.027^{**}	-0.121^{***}	-0.054^{***}	-0.048^{***}
	(0.014)	(0.013)	(0.015)	(0.015)
Controls	X		X	
Matching			X	X
Observations	630,692	630,692	137,676	137,676
Log Likelihood	−349,024.400	−359,846.700	−74,444.580	−76,333.970

Notes: the dependent variable is a binary indicator for household relocation between 2017 and 2021. Controls include socio economics, house characteristics and energy use in 2016. Numeric variables are standardised so that the inverse log-odds of the "Constant" coefficient can be interpreted as the relocation rate for an average household in the control group (e.g. $\exp(-1.124)/(1+\exp(-1.124)) = 24.5\%$). The sample consists of all households living in our study sample and for which there is no missing observed characteristics. In columns (3) and (4) treated dwellings are matched to 3 control dwellings on observed characteristics. Statistical significance: *p<0.1; **p<0.05; ***p<0.01.

C Appendix Main results - full table

Table C1 reports the full set the coefficients behind Table 2 estimating Equation (10).

Table C1: Average effects of insulation retrofit

Dependent: log of yearly natural gas use	(1)	(2)	(3)	(4)
Retrofit (year $\geq$ 2)	-0.149***	-0.228***	-0.228***	-0.218***
	(0.004)	(0.003)	(0.004)	(0.003)
Retrofit (year $<$ 2)	-0.028***	-0.109***	-0.105***	-0.100***
	(0.004)	(0.002)	(0.004)	(0.002)
Retrofit (year $\geq$ 2):Retrofit index	-0.058***	-0.077***	-0.100***	-0.078***
	(0.004)	(0.003)	(0.004)	(0.003)
Retrofit (year $<$ 2):Retrofit index	-0.024***	-0.039***	-0.065***	-0.038***
	(0.004)	(0.002)	(0.003)	(0.002)
No. children			-0.019**	0.018***
			(0.008)	(0.004)
No. persons			0.115***	0.022***
			(0.008)	(0.004)
No. persons squared			-0.014***	-0.002**
			(0.002)	(0.001)
No. senior squared			0.002	-0.002
			(0.003)	(0.002)
No. children squared			0.011***	0.000
			(0.002)	(0.001)
No. seniors			0.020**	-0.001
			(0.008)	(0.005)
No. females			0.050***	0.061***
			(0.008)	(0.007)
No. females squared			-0.009	-0.024***
			(0.006)	(0.004)
Household type nuclear family			0.084***	0.059***
			(0.007)	(0.004)
Household type one senior			0.055***	0.001
			(0.006)	(0.004)
Household type single parent			0.094***	0.047***
			(0.007)	(0.004)
Household type two adults			0.033***	0.042***
			(0.006)	(0.003)
Household type two seniors				0.035***
				(0.003)
Employed			-0.013***	-0.001

			(0.002)	(0.001)
Log income			0.034***	0.044***
			(0.003)	(0.002)
Education low			0.051***	0.016*
			(0.004)	(0.009)
Education medium			0.035***	0.005
			(0.004)	(0.008)
Education unknown			0.050***	0.035***
			(0.004)	(0.010)
Boiler changed			-0.019***	-0.042***
			(0.005)	(0.003)
Solar installation			-0.036***	-0.035***
			(0.003)	(0.002)
Log proj. gas use			0.257***	
			(0.004)	
Constr. Period 1940-1965			0.017***	
			(0.005)	
Constr. Period 1966-1976			0.006	
			(0.005)	
Constr. Period 1977-1992			-0.021***	
			(0.005)	
Log surface			0.215***	
			(0.007)	
Energy label D			0.018***	
			(0.003)	
Energy label E			0.034***	
			(0.003)	
Energy label F			0.042***	
			(0.005)	
Energy label G			0.049***	
			(0.005)	
No. obs.	963459	963459	959073	959073
No. treatment houses	13409	13409	13409	13409
No. control houses	110891	110891	110891	110891
R^2 Adj.	0.021	0.822	0.144	0.826
Year fixed-effect	X	X	X	X
Household fixed-effect		X		X
Controls			X	X

Notes: The table shows estimates of four separate regressions. The dependent variable is the log of gas consumption. Standard errors in parentheses are clustered at household level. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

D Appendix Sensitivity checks

In this Appendix we subject the results of Table 3 to a range of sensitivity analyses. We re-estimate the model Equation (12) for various subsamples, allowing the retrofit effect to differ by: (i) year in which insulation retrofit took place (2017, 2018 and 2019), Table D1; (ii) pre-retrofit energy-efficiency as defined by the energy label (C, D, E, F, G), Table D2; (iii) socio-economic characteristics of households, Table D3; (iv) pre-retrofit gas use quintile, Table D4. Results are robust across all the year and energy label subsamples. The average effect of insulation however differs by household type. For instance, singles reach larger savings, while households with migration background reduce gas consumption less than average. The average effect also differs by pre-retrofit gas use: households with low gas demand experience almost half lower savings than average. The low-income specific response to insulation is however robust in all the subsamples.

Table D1: Effects of retrofits by retrofit year

		Retrofit year		
	Baseline	2017	2018	2019
Retrofit (year ≥ 2)	-0.218 (0.003)***	-0.217 (0.005)***	-0.221 (0.004)***	-0.212 (0.005)***
$\times$ Below poverty line	0.024 (0.011)**	0.025 (0.020)	0.028 (0.017)*	0.004 (0.021)
$\times$ Below 100% soc.min.	0.062 (0.016)***	0.088 (0.031)***	0.046 (0.024)*	0.040 (0.027)
$\times$ Below 130% soc.min.	0.025 (0.006)***	0.022 (0.011)*	0.021 (0.010)**	0.032 (0.011)***
$\times$ Below 150% soc.min.	0.018 (0.006)***	0.022 (0.010)**	0.012 (0.008)	0.017 (0.010)

Notes: The table shows estimates of Equation (12) for 10 separate regressions. Coefficients reported are two- and three-way interactions. The symbol $\times$ indicates an effect as compared to the reference level (non-poor). The combination of the column and row name indicates the interaction. The dependent variable is log of gas. Each regression includes controls, household fixed-effects and year fixed-effects. The sample size is 13409 treated and 110891 control units. Standard errors in parentheses are clustered at household level. Statistical significance: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table D2: Effects of retrofits by energy label

		Energy label				
	Baseline	C	D	E	F	G
Retrofit (year $\geq$ 2)	-0.218^{***}	-0.223^{***}	-0.218^{***}	-0.228^{***}	-0.228^{***}	-0.225^{***}
	(0.003)	(0.007)	(0.005)	(0.005)	(0.008)	(0.012)
$\times$ Below poverty line	0.024^{**}	0.036	0.039^{*}	0.020	0.034	-0.022
	(0.011)	(0.032)	(0.022)	(0.017)	(0.028)	(0.078)
$\times$ Below 100% soc.min.	0.062^{***}	0.083^{*}	0.071^{**}	0.050^{*}	0.027	0.169^{*}
	(0.016)	(0.044)	(0.030)	(0.028)	(0.066)	(0.088)
$\times$ Below 130% soc.min.	0.025^{***}	0.035^{**}	0.022^{*}	0.025^{**}	-0.011	0.020
	(0.006)	(0.017)	(0.013)	(0.012)	(0.019)	(0.031)
$\times$ Below 150% soc.min.	0.018^{***}	0.023	0.010	0.020^{*}	-0.004	0.011
	(0.006)	(0.015)	(0.011)	(0.011)	(0.017)	(0.026)

Notes: The table shows estimates of Equation (12) for 10 separate regressions. Coefficients reported are two- and three-way interactions. The symbol $\times$ indicates an effect as compared to the reference level (non-poor). The combination of the column and row name indicates the interaction. The dependent variable is log of gas. Each regression includes controls, household fixed-effects and year fixed-effects. The sample size is 13409 treated and 110891 control units. Standard errors in parentheses are clustered at household level. Statistical significance: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table D3: Effects of retrofits by household type and migration background

	Mig. BG	Household type					
	Yes	1 adult	nuclear fam.	1 senior	single parent	2 adults	2 seniors
Retrofit (year $\geq$ 2)	-0.176^{***}	-0.252^{***}	-0.193^{***}	-0.223^{***}	-0.202^{***}	-0.225^{***}	-0.209^{***}
	(0.006)	(0.007)	(0.006)	(0.005)	(0.007)	(0.007)	(0.005)
$\times$ Below poverty line	0.044^{***}	0.020	0.063^{***}	-0.009	0.057^{***}	0.041	-0.009
	(0.018)	(0.020)	(0.018)	(0.026)	(0.023)	(0.034)	(0.053)
$\times$ Below 100% soc.min.	0.090^{***}	0.083^{***}	0.047	0.064^{***}	0.008	0.126^{*}	0.120^{***}
	(0.026)	(0.034)	(0.043)	(0.027)	(0.031)	(0.075)	(0.047)
$\times$ Below 130% soc.min.	0.028^{**}	0.005	0.044^{***}	0.028^{***}	0.028^{**}	0.032	0.053^{***}
	(0.013)	(0.017)	(0.015)	(0.011)	(0.014)	(0.026)	(0.013)
$\times$ Below 150% soc.min.	0.023^{**}	-0.009	0.047^{***}	0.021^{**}	0.019^{*}	0.027	0.033^{***}
	(0.012)	(0.015)	(0.013)	(0.011)	(0.013)	(0.022)	(0.010)

Notes: The table shows estimates of Equation (12) for 10 separate regressions. Migration background (Mig. Bg) is "Yes" when all household members are born outside the Netherlands (first and second generation). Nuclear family stands for 2 adults with children. Coefficients reported are two- and three-way interactions. The symbol $\times$ indicates an effect as compared to the reference level (non-poor). The combination of the column and row name indicates the interaction (e.g. below poverty line $\times$ Yes Migration background). The dependent variable is log of gas. Each regression includes controls, household fixed-effects and year fixed-effects. The sample size is 13409 treated and 110891 control units. Standard errors in parentheses are clustered at household level. Statistical significance: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table D4: Effects of retrofits by pre-retrofit gas use quintiles

		Pre-retrofit gas use quintile	
	Baseline	first quintile	last quintile
Retrofit (year $\geq$ 2)	-0.218^{***}	-0.128^{***}	-0.265^{***}
	(0.003)	(0.010)	(0.005)
$\times$ Below poverty line	0.024^{**}	0.015	0.006
	(0.011)	(0.034)	(0.019)
$\times$ Below 100% soc.min.	0.062^{***}	0.152^{***}	0.047^{*}
	(0.016)	(0.049)	(0.027)
$\times$ Below 130% soc.min.	0.025^{***}	0.047^{**}	0.014
	(0.006)	(0.022)	(0.011)
$\times$ Below 150% soc.min.	0.018^{***}	0.031	0.013
	(0.006)	(0.020)	(0.010)

Notes: The table shows estimates of Equation (12) for 15 separate regressions. Coefficients reported are two- and three-way interactions. The symbol $\times$ indicates an effect as compared to the reference level (non-poor). The combination of the column and row name indicates the interaction (e.g. below poverty line $\times$ first quintile). The dependent variable is log of gas. Each regression includes controls, household fixed-effects and year fixed-effects. The sample size is 13409 treated and 110891 control units. Standard errors in parentheses are clustered at household level. Statistical significance: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

E Alternative model specifications

E.1 Functional form retrofit intensity

Table E1 shows the retrofit effect on poor households for various specifications of the retrofit intensity: the first column is the baseline specification Equation (10), the second column excludes the largest retrofits (retrofit intensity > 2), the third column allows for non-linear effects of the retrofit intensity and the last column discard the retrofit intensity. Low-income response is robust across all these specifications.

Table E1: Effects of retrofits across various retrofit intensity specifications

	Specification of retrofit intensity			
	Baseline	Linear and $\leq$2	Polynomial	Not controlled for
Retrofit (year $\geq$ 2)	-0.218 (0.003)***	-0.222 (0.006)***	-0.223 (0.003)***	-0.217 (0.003)***
$\times$ Below poverty line	0.024 (0.011)**	0.023 (0.012)**	0.024 (0.014)*	0.026 (0.011)**
$\times$ Below 100% soc.min.	0.062 (0.016)***	0.064 (0.017)***	0.062 (0.022)***	0.063 (0.017)***
$\times$ Below 130% soc.min.	0.025 (0.006)***	0.023 (0.007)***	0.013 (0.008)	0.023 (0.006)***
$\times$ Below 150% soc.min.	0.018 (0.006)***	0.017 (0.006)***	0.009 (0.007)	0.016 (0.006)***

Notes: The table shows estimates of Equation (12) for 20 separate regressions. In the first column, the retrofit intensity enters the model linearly. In second column, observations with the retrofit intensity larger than 2 are discarded. In the third column, the retrofit intensity and its second and third orders enter the model. In the last column, the retrofit intensity is discarded from the model. Coefficients reported are two-way interactions. The symbol $\times$ indicates an effect as compared to the reference level (non-poor). The dependent variable is log of gas. Each regression includes controls, household fixed-effects and year fixed-effects. The sample size is 13409 treated and 110891 control units. Standard errors in parentheses are clustered at household level. Statistical significance: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

E.2 Sun and Abraham estimator

The coefficient of treatment effect can be biased in studies where the treatment timing differs across units, as is shown in (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). Table E2 reports the retrofit effect on poor households, using the estimator from Sun and Abraham (2021) that corrects for the variation in treatment timing (staggered treatment). Low-income response is robust to this correction.

Table E2: Effects of retrofits, Sun and Abraham estimator ("Sunab")

	Not sunab	Sunab
Retrofit (year $\geq$ 2)	-0.218 (0.003)***	-0.225 (0.003)***
Below poverty line	-0.188 (0.012)***	-0.198 (0.012)***
Below 100% soc.min.	-0.155 (0.017)***	-0.161 (0.017)***
Below 130% soc.min.	-0.209 (0.006)***	-0.224 (0.006)***
Below 150% soc.min.	-0.215 (0.005)***	-0.228 (0.005)***

Notes: The table shows estimates of Equation (12) for 10 separate regressions."Sunab" stands for Sun and Abraham estimator (Sun and Abraham, 2021). All coefficients (except "Retrofit") are estimated on the sub-samples of poor households. The dependent variable is log of gas. Each regression includes controls, household fixed-effects and year fixed-effects. The baseline sample size is 13409 treated and 110891 control units. Standard errors in parentheses are clustered at household level. Statistical significance: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

E.3 Heterogeneous time trends

Table E3 shows the retrofit effect on poor households where the time-fixed effect is allowed to differ between poor and non-poor households. Low-income response is robust to this specification.

Table E3: Effects of retrofits, allowing heterogenous time trends

	Baseline	Heterogeneous time trends
Retrofit (year $\geq$ 2)	-0.218 (0.003)***	-0.218 (0.003)***
$\times$ Below poverty line	0.024 (0.011)**	0.030 (0.011)***
$\times$ Below 100% soc.min.	0.062 (0.016)***	0.054 (0.017)***
$\times$ Below 130% soc.min.	0.025 (0.006)***	0.010 (0.007)
$\times$ Below 150% soc.min.	0.018 (0.006)***	0.002 (0.006)

Notes: The table shows estimates of Equation (12) for 10 separate regressions where the time fixed-effect can differ between the poor and non-poor households. Coefficients reported are two-way interactions. The symbol $\times$ indicates an effect as compared to the reference level (non-poor). The dependent variable is log of gas. Each regression includes controls, household fixed-effects and year fixed-effects. The sample size is 13409 treated and 110891 control units. Standard errors in parentheses are clustered at household level. Statistical significance: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

F Appendix Causal forest

The causal forest algorithm we apply, is based on random forests (Breiman, 2001), while additionally enabling consistency and asymptotic validity of the heterogenous treatment estimates and providing valid confidence intervals for them. The basic building block of the causal forest is a *regression tree*, which employs recursive partitioning to split a sample into subgroups that maximize heterogeneity across splits.[24] We employ the generalized random forest (grf) R package by Tibshirani et al. (2022). Taking advantage of the panel data structure, we redefine the dependent variable as the *difference* between the yearly gas usage in the year 2021 and the year 2016.[25] The predictors are all observed house and household characteristics in 2016.

F.1 Tree growing routine

First we explain the tree growing routine which lies at the bottom of the causal forest algorithm.

1. Randomly draw (i) a sample of households (50% of the original data) and (ii) a subset of available covariates.

2. Use the sample and the subset to grow a tree, by splitting the sample iteratively in branches. (See Knittel and Stolper (2021) for an example tree.) A split in two branches is performed when the resulting branches maximize heterogeneity, under the constraint that each branch should contain at least 10 treated and control units. The formal criterion to be maximized is defined in Athey et al. (2019) and is proportional to the squared difference of treatment effects between the two branches (treatment effect is equal to mean outcome of the treatment units minus mean outcome of control units).

[24]Each tree starts with a single root node, which is split in child nodes, which are split further recursively to form a tree. To maximize heterogeneity in subgroup ATE, penalties for within-node variance in ATE and treatment-control imbalance are applied. When splitting a specific node cannot result in an improved fit, that node forms a 'leaf' of the final tree. A forest is formed by a collection of a large number of such trees.

[25]Recall that 2016 is a pre-treatment year in our data and 2021 is a post-treatment year. No retrofits occur in the data in these two years.

3. Match households in the sample to leaves of the trees, according to observed characteristics.

4. Estimate ATEs in each leaf using the matched observations in that leaf. The within-leaf ATE estimation is implemented as a cross-sectional, difference-in-means comparison between treatment and control group.

5. Repeat for next tree.

F.2 Running causal forest

The procedure followed in the causal forest, can then be described by four steps:

1. Tree-growing routine, repeated 40,000 times to train the causal forest. See above.

2. Predict the insulation retrofit effect on the outcome of interest, for every house in our data (both, treatment and control group), using the trained model. Fro each dwelling, the 40,000 predictions are aggregated into a single, central estimate of a household's treatment effect using adaptive neighbourhood estimation (Tibshirani et al., 2022).

3. Distinguish houses into 5 quintiles, by size of the predicted effect. We call the quintiles low, medium-low, medium, medium-high, high.

4. Use Equation (12) to estimate the average retrofit effect within each quintile and to test the hypothesis that responses to retrofits differ between the quintiles.

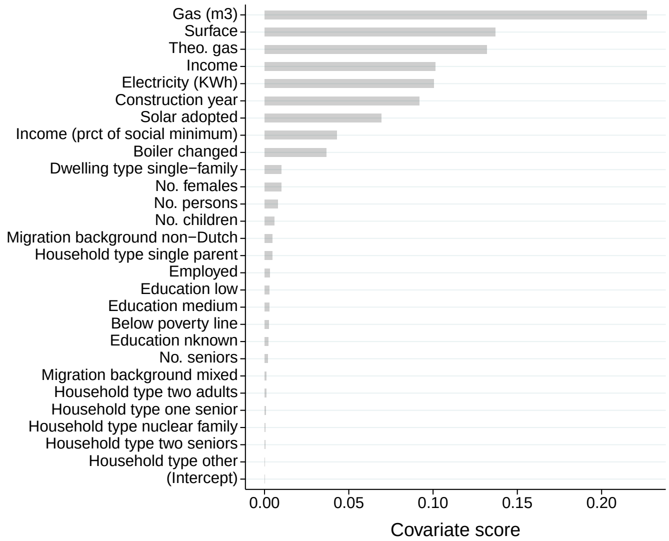
Summarizing, in our analysis, the causal forest predictions are instrumental to identify the possible range of heterogenous responses to the treatment. We do not use the forest predictions directly; all effect estimations are based on econometric models like Equation (12).

To get a first idea of the degree of heterogeneity in the treatment effect we use causal forest to order the observations into five quintiles, by increasing size of the predicted treatment effect. Afterwards we estimate Equation (12) with $J = 5$, including two-way interaction effects of the treatment with the quintiles dummies. In this way we allow the treatment effect to differ for each of the forest quintiles.

F.3 Covariates importance

Figure F1 reports the covariate importance.

Figure F1: Causal forest covariate importance



Notes: The covariate score in the figure indicates how frequently the covariate was selected by causal forest to split the sample. In the causal forest model specification, the dependent variables is the change of log gas. The control covariates are taken in 2016, before any renovation occur in our sample.