Discussion Paper Series – CRC TR 224

Discussion Paper No. 629
Project B 07

The Tragedy of the Common Heating Bill

Harald Mayr[1]
Mateus Souza[2]

January 2025

[1]University of Zurich, Email: harald.mayr@econ.uzh.ch
[2]University of Mannheim, Email: mateus.souza@uni-mannheim.de

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

The Tragedy of the Common Heating Bill *

Harald Mayr[a] and Mateus Souza[b]

[a]University of Zurich, harald.mayr@econ.uzh.ch
[b]University of Mannheim, mateus.souza@uni-mannheim.de

December 17, 2024

Abstract

Without heat metering, households face strong free-riding incentives. Using data from Swiss households, we find that the staggered introduction of submetering reduced heating expenses by 17%, on average. Machine learning techniques reveal highly heterogeneous effects, consistent with coordination failure in larger buildings and strategic exit of free-riders. We find that households are price elastic even when they share a common heating bill. Our results suggest that most households do not exploit the free-riding incentive, especially in smaller buildings. "Schmeduling," inattention to the billing regime, and pro-social behavior can explain the low prevalence of free-riding. Nevertheless, submetering is welfare-improving for most buildings.

Keywords: Free-riding, submetering, individual billing, heating energy, tragedy of the commons, welfare

JEL Codes: D61, Q41, Q52

*Acknowledgements: We are very grateful for the comments and constructive feedback from Peter Christensen, Natalia Fabra, Andreas Gerster, David Hémous, Sébastien Houde, Kelsey Jack, Larry Karp, Erica Myers, Ulrich Wagner, Kathrine von Graevenitz, and from seminar participants at AERE 2023, AERE@OSWEET, EAERE 2023, EEA 2023, CESifo Munich, CEPR FGN Workshop, ETH Zurich, NBER Workshop on the "Economics of Decarbonizing the Built Environment," NOeG Winter Workshop 2023, NOeG Annual Meeting 2024, Toulouse School of Economics, Universidad Carlos III de Madrid, University of Antwerp, University of California Berkeley, University of California Santa Cruz, University of Innsbruck, University of Mannheim, University of Oregon, University of Salzburg, and ZEW Mannheim. Mayr acknowledges funding from the Swiss National Science Foundation (Project 201958). Souza acknowledges support from the German Research Foundation (DFG) through CRC TR 224 (Project B07). All errors are our own.

1 Introduction

Externalities and free-riding behavior are at the core of pressing global challenges. Climate change, antibiotic resistance, and overfishing are just a few examples where individual incentives are misaligned with social costs (Stern, 2008; Roope et al., 2019; Noack and Costello, 2024). Given the ubiquitous nature of these problems, it is not surprising that scholars have long been interested in the conditions under which groups can overcome them. Hume (1739) argues that large groups will find it difficult to provide public goods. Coase (1960) emphasizes the importance of low transaction costs, and Ostrom (1990) describes principles, including monitoring and sanctioning, that allow communities to effectively manage common-pool resources. We add to this body of work by studying free-riding behavior in the context of energy consumption in apartment buildings.

We study the implications of the "common heating bill," where each household pays a fixed share of building-level heating expenses. Households face a clear free-riding incentive, as the expenses from their marginal heat consumption will be split with their neighbors. This situation is analogous to the n-player prisoner's dilemma and the "diner's dilemma," where bill-splitting influences the consumption of diners sharing a table (Glance and Huberman, 1994; Gneezy, Haruvy, and Yafe, 2004). The common heating bill allows us to investigate free-riding behavior in a high stakes environment. Further, we contribute to understanding the role of bill-sharing group size, as we observe buildings that vary in number of apartments. Most importantly, we evaluate a policy that corrects the free-riding problem: submetering.[1]

The lack of heat submetering is widespread. For example, the majority of households in 23 of 37 economies surveyed in Central and Eastern Europe, Central Asia, and the Mediterranean have no heat meters (EBRD, 2023). A survey in China found that roughly 37% of households pay heating bills simply according to their dwelling's area, most of them in the colder North of the country (Guo, Huang, and Wei, 2015). According to the most recent version of the American Housing Survey (AHS, 2023), 17% of renters in the US face fixed heating expenses included in the rent. This can be problematic for worldwide decarbonization efforts in the residential sector, since the lack of consumption-based billing can lead to over-consumption and to muted carbon pricing signals.

Our study is carried out in three main steps. First, we model the common heating bill according to our empirical setting. In this framework, we evaluate how submetering changes energy demand, the price elasticity, and welfare. Second, we empirically test our model's

[1] Throughout this paper, we use the term "submetering" to describe the use of apartment-specific meters for billing. We use the term synonymous to "individual metering and billing". Submetering does not address the related problem within households, where individual household members may have an incentive to over-consume (Jack et al., 2024).

predictions by leveraging the staggered rollout of submetering in our sample, as well as energy price variation across time, region, and heating fuel type. We pay special attention to the heterogeneous effects of submetering. Third, we discuss potential deviations from our theoretical framework, including their implications for our welfare calculations.

Our theoretical framework incorporates two externalities: (i) the direct externality from the fact that, under the common heating bill, each household's consumption choice impacts all neighbors' heating bills; and (ii) the environmental externality from heating energy. Submetering eliminates the direct externality while simultaneously reducing the environmental externality. The latter is important, as heating accounts for 80% of direct CO_2 emissions in the building sector (IEA, 2022). Our framework yields a sufficient statistic to empirically evaluate the welfare effects of submetering. It also highlights that the effects of submetering depend on the building's heat loss, households' temperature preferences, the number of bill-sharing neighbors, and energy prices—motivating our exploration of effect heterogeneity.

Our main empirical strategy leverages the staggered introduction of heat submetering in 185 apartment buildings in Switzerland between 2008 and 2022. This constitutes a difference-in-differences setting, where the "treated" buildings are those that switched billing regimes after the installation of submeters. The comparison buildings are those that remained under common billing during the whole sample period (i.e., never-treated). Including treated and comparison groups, we have information from 4,124 buildings with a total of 44,415 apartments. Our data include three essential features that were mostly unavailable in previous studies on submetering of heating energy. We observe (i) heating energy expenses before the installation of submeters, which allows us to capture potential effects of announcement or installation. Moreover, we observe (ii) renovations and (iii) tenant changes, allowing us to document how these correlate with changes in energy consumption and the introduction of submetering. We implement event study regressions using the heterogeneity-robust estimator from Sun and Abraham (2021).

We find that submetering reduces heating energy expenses by 17%, on average. We document anticipation effects, as a significant portion of the energy savings happens in the year of the installation of submeters, but before submetered bills are charged. The effect sizes are larger (20.8%) when we do not control for renovations or turnover. In all specifications, the effects are significant and persistent at least until the fifth year after installation. Moreover, we find that submetering moderately increases tenant turnover.

We use machine learning (ML) algorithms to explore heterogeneity, primarily through a counterfactual imputation method (Borusyak, Jaravel, and Spiess, 2024; Souza, 2019). We estimate the *conditional* average effects of submetering, while accounting for several confounding factors. Consistent with our framework, we find that submetering effects are more pronounced in buildings with a larger number of apartments. Also in line with our framework, we find that submetering effects are stronger for large apartments (requiring more heating energy and facing stronger heat loss). Our analysis further reveals that treatment effects are larger when we observe tenant turnover. This finding is consistent with a scenario where high-consumption tenants (free riders under common billing) choose to leave the building because they anticipate higher heating bills under submetering (Brewer, 2022). We account for this in our welfare analysis, which would otherwise be based on inflated treatment effects. We do not find effect heterogeneity by apartment location in the building, suggesting that heat flows between apartments do not significantly influence our results.[2]

We also implement ML techniques to estimate *building-specific* treatment and welfare effects of submetering. This exercise allows us to identify the set of buildings in our sample for which submetering was cost-effective. Disregarding the social cost of carbon (SCC) and assuming annual submetering costs of CHF 86, we find that submetering was cost-effective in only 10% of the treated buildings. This result rationalizes why the common heating bill is still prevalent in Switzerland and many other countries. However, when we account for a SCC of 185 USD/tCO_2 (Rennert et al., 2022), submetering is welfare-improving for 63% of treated buildings. To the extent that our sample is similar to non-submetered buildings, our results suggest that submetering policies can improve welfare for many apartment buildings in Switzerland.

Our theoretical framework also predicts that submetered households are more price elastic than those under common billing. We empirically test that by comparing outcomes in submetered versus non-submetered buildings, leveraging energy price variation by region, over time, and depending on heating fuel type. Our estimates show that households in submetered buildings are more price elastic. This difference is primarily driven by the lower price-responsiveness of households who live in large apartment buildings under the common heating bill.

However, we note that our theoretical framework cannot quantitatively match our elasticity results. We find a relatively large price elasticity (around -0.6) under the common heating bill. If all households behaved according to our model, one would expect lower elasticities, since these households face a very low marginal heating cost. Conversely, the estimated elasticities imply that a switch from common billing to submetering should have induced a sharper decrease in energy consumption. We discuss that these discrepancies can

[2]The engineering literature suggests that heating energy expenses are lower for apartments located in "intermediate" floors (Ling, Li, and Xing, 2015), although measured indoor temperatures are uncorrelated with apartment location (Dahlblom, Nordquist, and Jensen, 2015). Taken together, these findings indicate that apartments in intermediate floors may benefit from heating from adjacent floors. However, in Section 4.2 we show that heat flows between apartments are unlikely to influence the effects of submetering.

be explained by "schmeduling" (responding to average, rather than marginal price changes), inattention to the common heating bill, and pro-social behavior.

We find that our results are consistent with approximately one quarter of the population free-riding according to our theoretical framework. The remaining three quarters can be explained by "schmeduling," inattention to the common heating bill, and/or pro-social behavior (e.g., Kaufmann, Andre, and Kőszegi, 2024). Prior literature suggests that schmeduling is prevalent in the context of residential energy consumption (Ito, 2014; Shaffer, 2020; Ito and Zhang, 2023). We provide novel survey evidence on inattention to the common heating bill. We asked 835 households under the common heating bill to state the billing regime they believed to be in. 19% of respondents mistakenly said that they were in the submetering regime. Our results indicate that pro-social behavior is relevant in this context. We find that buildings with fewer neighbors report better relationships, potentially explaining why the share of free-riders is substantially smaller in buildings with few neighbors. These findings are consistent with small groups managing the common heating bill more effectively. We argue that, given the behavioral mechanisms likely at play, our welfare estimates may be somewhat conservative.

Our study relates to several strands of economic literature. We contribute to a long-standing question on the effect of group size on cooperation. This question has been of interest since Hume (1739), and an important aspect of the seminal work of Olson (1965). A large experimental literature suggests that cooperation can either increase (Isaac, Walker, and Williams, 1994; Diederich, Goeschl, and Waichman, 2016; Nosenzo, Quercia, and Sefton, 2015) or decrease (Marwell and Ames, 1979; Nosenzo, Quercia, and Sefton, 2015; Arifovic et al., 2023) with group size in public good games, the diner's dilemma, and similar setups. A related literature studies the effect of group size in the context of fisheries, agriculture, and forestry commons (Ostrom, 1990; Agrawal, 2000; Poteete and Ostrom, 2004; Casari and Tagliapietra, 2018). Our study bridges experimental and observational approaches. Unlike lab studies, we analyze naturally occurring groups with long-term interactions and high stakes. Our setting allows for more controls than most prior observational studies, as we can leverage the staggered introduction of a policy intervention that corrects the free-riding problem. Using a rich dataset and quasi-experimental methodology, we provide new insights into free-riding behavior.

Our results are also directly relevant for a growing literature on the effects of submetering. Studies on electricity (Munley, Taylor, and Formby, 1990; Dewees and Tombe, 2011; Elinder, Escobar, and Petré, 2017; McRae, 2024), water (Ornaghi and Tonin, 2021), and hot water (Elinder et al., 2024) find that submetering leads to large reductions in consumption—between 18% and 36%. Evidence on the effects of heat submetering is somewhat limited. A number of case studies provide highly variable estimates, ranging from a 40% reduction to a 24% *increase* in energy consumption (Canale et al., 2019). More recently, Ito and Zhang (2023) provide quasi-experimental evidence on the switch from a fixed heating charge to a consumption-based heating tariff in China, finding that the switch reduced heating energy consumption by 10%. Our study contributes to this literature by revealing that it is important to take renovations into account, that submetering can lead to the exit of free-riders, that treatment effects are highly heterogeneous, that welfare effects of submetering depend crucially on the social cost of carbon, and that submetering increases households' price elasticity. The results on elasticity, in particular, suggest that the lack of submetering can be viewed as yet another energy price distortion, in addition to those caused by taxes, subsidies, regulatory pricing, or market power in the residential sector (Borenstein and Bushnell, 2022; Asker et al., 2024).

Our findings also speak to a similar strand of literature on the effects of switching from landlord-pay regimes (where tenants face zero marginal costs for heating) to tenant-pay regimes, finding substantial reductions in consumption (e.g., Brewer, 2022). We add to a recent body of evidence demonstrating heterogeneous effects (e.g., Knittel and Stolper, 2021; Christensen et al., 2023), and the potential gains from targeting in the context of residential sector energy interventions (e.g., Allcott and Kessler, 2019; Knittel and Stolper, 2019; Gerarden and Yang, 2023; Christensen et al., 2024). More broadly, we contribute to understanding the "energy efficiency gap," i.e., the relatively slow adoption of energy efficiency technologies, despite their apparent short payback periods (e.g., Allcott and Greenstone, 2012; Gillingham and Palmer, 2014; Gerarden, Newell, and Stavins, 2017). Submetering can be viewed as an energy efficiency investment that could be profitably adopted in some buildings, yet faces barriers such as coordination problems or split incentives between landlords and tenants (e.g., Gillingham, Harding, and Rapson, 2012; Myers, 2020).

The remainder of this paper is organized as follows. Section 2 presents a theoretical framework to assess the welfare effects of submetering. Section 3 describes the data and empirical strategy. Section 4 presents our empirical analysis, including average effect, effect heterogeneity, welfare estimates, and effects on the price elasticity of demand. Section 5 discusses alternative behaviors and Section 6 concludes.

2 Theoretical framework

We propose a straightforward model incorporating the main elements necessary to illustrate the "tragedy of the common heating bill." Our framework allows us to derive predictions which we empirically test in section 4. We also derive a sufficient statistic for the welfare effect of submetering. To this end, we begin with standard assumptions of free-riding behavior. In Section 5, we discuss plausible alternative behavioral mechanisms in our setting.

Let there be N households that live in separate apartments within the same building. We assume that each household i has a quasi-linear utility function that depends on heating energy e_i and a numeraire outside good c_i. Households do not derive utility from energy consumption per se, but from their apartment's indoor temperature $T(e_i)$. Household i prefers temperature T_i and suffers disutility from the square of temperature deviations

$$U_i(e_i, c_i) = -\kappa_i\big(T_i - T(e_i)\big)^2 + c_i \ .$$

The parameter κ_i captures the household's intolerance to deviations from its preferred temperature. We model indoor temperature as function of heating energy consumption, through a simplified heat loss formula $T(e_i) = T_0 + \frac{e_i}{H}$, where T_0 is the outside temperature and H measures the apartment's heat loss (inverse energy efficiency, capturing the thermal resistance of walls and windows, losses at thermal bridges, as well as ventilation losses).[3] Combining these ingredients, the utility function of household i can be written as

$$U_i(e_i, c_i) = -\kappa_i\left(T_i - T_0 - \frac{e_i}{H}\right)^2 + c_i \ . \tag{1}$$

The household faces a budget constraint that depends on its income y_i, the price of energy p, consumption of the numeraire good, and the billing regime for heating energy. A share $\theta \in [0, 1]$ of heating consumption is individually billed, while the remainder is equally split among all neighbors $j = (1, \ldots, N)$, such that

$$y_i \geq \theta e_i p + (1 - \theta)\frac{\sum^j e_j p}{N} + c_i \ . \tag{2}$$

This budget constraint implies an externality if $\theta < 1$, as household i's heating consumption e_i influences the energy bills of all neighbors (and vice-versa).

[3] We focus on the context of heating (not cooling), such that $T_0 \leq T(e_i)$ and $e_i \geq 0$. The engineering literature typically models heating energy requirements as $e_i = H(T - T_0) + S$, where H is the heat loss per degree of temperature difference, and S is the rate of heat storage within the structure (Johannesson et al., 1985). We omit S, because heat storage is negligible in the long term (Johannesson et al., 1985). We further abstract from solar heat gains and other internal heat gains.

Private solution

The household maximizes Equation 1 subject to Equation 2, resulting in

$$e_{i,\theta}^{*} = (T_i - T_0)H - \frac{H^2}{2\kappa_i} p \left(\theta + \frac{1-\theta}{N} \right) . \tag{3}$$

Unsurprisingly, optimal energy consumption $e_{i,\theta}^{*}$ is larger for high preferred indoor temperature T_i, low outside temperature T_0, and high heat loss H. The influence of energy prices p is modulated by the intolerance parameter κ_i, heat loss H and, importantly, the billing regime θ. When heating bills are fully shared among neighbors, then $\theta = 0$ and the price incentive is distorted to $\frac{p}{N}$. This distortion is particularly severe if N is large, i.e., the heating bill is shared among many households. Conversely, with individualized metering and billing, then $\theta = 1$ and the price incentive does not depend on the number of neighbors.[4] When energy prices are zero, households consume heating energy up to satiation, such that indoor temperature exactly matches their temperature preference.

The own-price elasticity of heating energy demand is

$$|\eta_{i,\theta}| = |\frac{\partial e_{i,\theta}^{*}}{\partial p} \frac{p}{e_{i,\theta}^{*}}| = \frac{H^2}{2\kappa_i} \frac{p}{e_{i,\theta}^{*}} \left(\theta + \frac{1-\theta}{N} \right) . \tag{4}$$

It is therefore clear that households should be more price sensitive under a submetering regime ($\theta = 1$). Under shared billing, the elasticity decreases as the number of neighbors N increases.

Submetering

Submetering increases θ from 0 to 1. The change in household i's energy consumption in response to submetering will be:

$$\Delta e_i^{*} = e_{i,\theta=1}^{*} - e_{i,\theta=0}^{*} = -\frac{H^2}{2\kappa_i} \frac{N-1}{N} p . \tag{5}$$

The negative sign implies a reduction in energy consumption after submetering. Energy savings are larger for apartments with high (squared) heat loss H, households that are more tolerant to temperature deviations (i.e., low κ_i), and at high energy prices p. Again, these effects are modulated by the number of bill-sharing neighbors N. All else equal, the energy savings from submetering are increasing in N. This framework also implies that *all*

[4] In practice, θ may be larger than zero but smaller than one when the heating bills include fixed service fees, or when only common-area heating costs are shared, for example. Appendix B.1 describes the conditions under which the simplified model with $\theta = 1$ is equivalent to a more complex model with fixed heating costs.

households i will over-consume when the energy bill is shared, resembling a classical diner's dilemma (e.g., Gneezy, Haruvy, and Yafe, 2004). Intuitively, this is because each household is paying for only a fraction of their marginal heating energy consumption.

While the direction of Δe_i^* is clear, submetering may either increase or decrease household i's utility:[5]

$$\begin{aligned}\Delta U_i &= U_i(e^*_{i,\theta=1}) - U_i(e^*_{i,\theta=0}) \\ &= \left(\frac{p^2}{N^2} - p^2\right)\frac{H^2}{4\kappa_i} + \left(\frac{\sum^j T_j}{N} - T_i\right)Hp \\ &\quad + p^2\frac{H^2}{2\kappa_i} - \frac{p^2}{N^2}\frac{H^2}{2}\sum^j \frac{1}{\kappa_j}\end{aligned} \tag{6}$$

There can be winners and losers from submetering, depending on households' preferences and the size of the price distortion. The first term of the right-hand side of Equation 6 captures the utility loss from lowering indoor temperature. That term will be small, for example, for households that are highly intolerant (high κ_i) to temperature deviations. The second term reflects distributional consequences that vary according preferences for indoor temperature T_i. This term is zero for the average household. However, it will be positive (negative) if the household prefers a relatively low (high) temperature, as compared to its neighbors. The remaining terms represent utility gains from removing the price distortion.

Finally, as already evident from Equation 4, our framework implies that submetering increases the own-price elasticity of heating energy demand:

$$\begin{aligned}\Delta \eta_i &= |\eta_{i,\theta=1}| - |\eta_{i,\theta=0}| \\ &= \frac{H^2}{2\kappa_i}\left(\frac{p}{e^*_{i,\theta=1}} - \frac{1}{N}\frac{p}{e^*_{i,\theta=0}}\right) > 0\end{aligned} \tag{7}$$

Welfare

Submetering is the preferred policy if it improves welfare. We model welfare from the perspective of a social planner who accounts for the utility of all households i that live in the same building. We abstract from any potential spillovers across buildings. In addition to affecting households' private utility, energy consumption is associated with an environmental

[5]The complete derivation is presented in Appendix B.2.

externality ϕ (to the extent it is not internalized by a carbon price). The welfare function also incorporates the per-household cost of submetering s, such that:

$$W = \sum^{i} U_i(e^*_{i,\theta}) - \phi \sum^{i} e^*_{i,\theta} - \mathbb{1}_{\theta>0} sN \ . \tag{8}$$

The welfare effect of submetering can be expressed in terms of empirical estimates. Note that the average treatment effect of submetering on energy expenses is $\tau = \frac{\sum^i \Delta e^*_i p}{N}$. Then the change in welfare will be:

$$\Delta W = -\frac{\tau N}{2}\left(1 - \frac{1}{N}\right) - \tau\frac{\phi}{p} - sN \ . \tag{9}$$

Therefore, an estimate of τ, together with the cost of submetering and the social cost of carbon are sufficient to estimate ΔW. In Section 4.3, we use this equation to calculate building-specific welfare effects.

Figure 1 depicts the potential welfare loss associated with the common heating bill. Panel a shows the situation without a carbon price. Privately optimal energy consumption in the common heating bill regime $e^*_{i,\theta=0}$ is characterized by $u'_i(e^*_{i,\theta=0}) = \frac{p}{N}$, i.e., point A on the lower right. Submetering increases the implicit price from $\frac{p}{N}$ to p and decreases energy consumption to $e^*_{i,\theta=1}$. If, in addition to submetering, the household also internalizes the environmental externality ϕ, it consumes the social optimum $e^*_{i,S}$, i.e., the point on the upper left. The central insight from this illustration is that submetering partially addresses welfare losses due to the environmental externality. The welfare gain from submetering is represented by the green areas in Panel a of Figure 1. The light green triangle represents the first term of Equation 9, as households save more on their heating bill than they lose in terms of thermal comfort. The dark green rectangle represents the second term of Equation 9, as emissions are reduced to the climate's benefit. The welfare loss represented by the tiled purple area cannot be resolved by submetering alone. Panel b of Figure 1 shows the situation with a carbon price. Given the common heating bill, a carbon price only addresses a fraction of the welfare loss, represented by the tiled blue area. Submetering resolves the remaining welfare loss, represented by the light green area. Submetering is welfare improving if the benefits of submetering exceed the cost, i.e., the third term of Equation 9.

3 Setting and data

We use data from a large real estate management company in Switzerland. The company manages apartment buildings for institutional and private real estate owners. As part of their

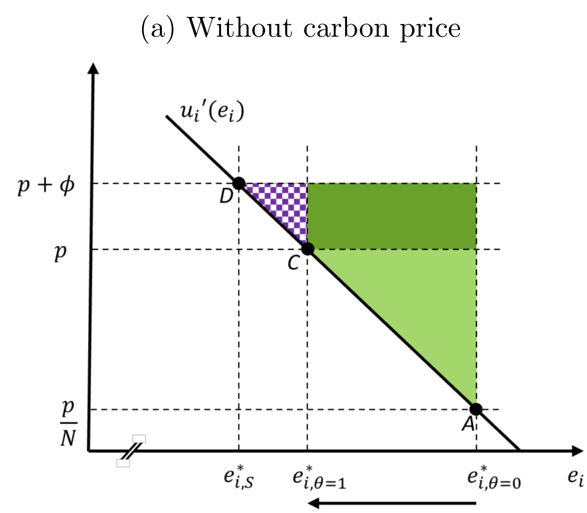


Figure 1: Private solutions and social optimum

mandate, the company provides a range of services to the buildings they manage, including the provision and billing of heating energy. Buildings typically have a central heating system, which is used to heat the apartments and provide hot water. The real estate management company is responsible for the maintenance and operation of the heating system. Depending on the type of heating system, it buys heating oil, natural gas, or district heating from energy suppliers. The company bills tenant households annually for their share of heating expenses. Under the common heating bill regime, the company divides the total heating expenses among households according to a fixed key, usually the area of the apartment. In the submetering regime, the company bills households based on their individual consumption of heating energy. Households receive an annual bill that includes heating expenses and other utilities.

Providers of natural gas and district heating are local monopolies, whereas local heating oil providers operate in a free market. There is no increasing block pricing and the price of energy is largely independent of consumption. Households have no influence on the price of energy, as contracts are arranged between the real estate management company and energy providers. Switzerland introduced a CO_2 levy in 2008. The levy started out at 12 CHF/tCO_2 and was increased multiple times: to 36 CHF/tCO_2 in 2010, to 60 CHF/tCO_2 in 2014, to 84 CHF/tCO_2 in 2016, to 96 CHF/tCO_2 in 2018, and to 120 CHF/tCO_2 in 2022.

Since they are all tenants, households in our setting also have limited influence regarding any structural changes to the apartments. Most decisions are at the discretion of the real estate owners, not the real estate management company. This includes the decision to install submeters, the choice of heating system and fuel type, and energy efficiency renovations. Households are usually neither in direct contact with owners, nor involved in the decision making process. They can influence energy bills through their heating behavior, but they cannot influence the billing regime or the energy efficiency of the building.

Our dataset includes household-level utility bills for 4,124 buildings with a total of 44,415 apartments, from the years 2006 to 2023.[6] We can identify which apartments belong to each building. The data include yearly heating and auxiliary expenses that are billed by the real estate management company. Expenses include heating fuels, water and sewage bills, as well as various service fees. For each household in our sample, we observe the yearly heating energy bill in CHF. In the common heating bill regime, this amount is a fixed share of the buildings' heating energy expenses. This fixed share usually depends on the area of the apartment in relation to the whole building. In the submetering regime, each household's heating bill depends on its own consumption of heating energy. The median household in our sample consumes heating energy worth CHF 980 in a year.

We infer submetering from how a buildings' total heating energy expenses are shared. In the common heating bill regime, each apartment's share of the bill depends on a weight that is an integer (usually its area in square meters). In the submetering regime, apartments pay according to their consumption, which implies that their weight is unlikely to be an integer. In each year, we assign a building to the submetering regime if at least 50% of its apartments have non-integer weights. All remaining buildings, with a positive but small share ($\leq$ 50%) of apartments with non-integer weights, are dropped from our sample (less than 5% dropped). Finally, we restrict the sample such that we observe the treated buildings for at least two years before and two years after submetering. Our final sample consist of 4,124 buildings, out of which 185 received submetering during our sample period (corresponding to 1,564 apartments). Figure 2 shows the staggered rollout of submetering over the years.

In practice, submetering is often introduced with other measures. This can include the replacement of the heating system, or the insulation of the building's facade, and other improvements that may influence the building's energy efficiency. We have access to the real estate company's database where any such improvements are recorded. This includes the renovation year, as well as short texts describing the renovation. We observe 1,473 renovations in our sample. Using the OpenAI API (OpenAI, 2024), we categorize renovations into comprehensive renovations, non-comprehensive renovations that change the energy efficiency of the building, and other renovations. Our empirical specifications control for the effect of renovations on energy consumption. Appendix C describes the renovation classification procedure in detail.

[6] These numbers already represent a restricted sample for which we drop "always-treated" apartments (i.e., those that were already submetered in the first year we observe them). The earliest treated buildings in our regression samples were submetered in 2008, and the latest were submetered in 2022.

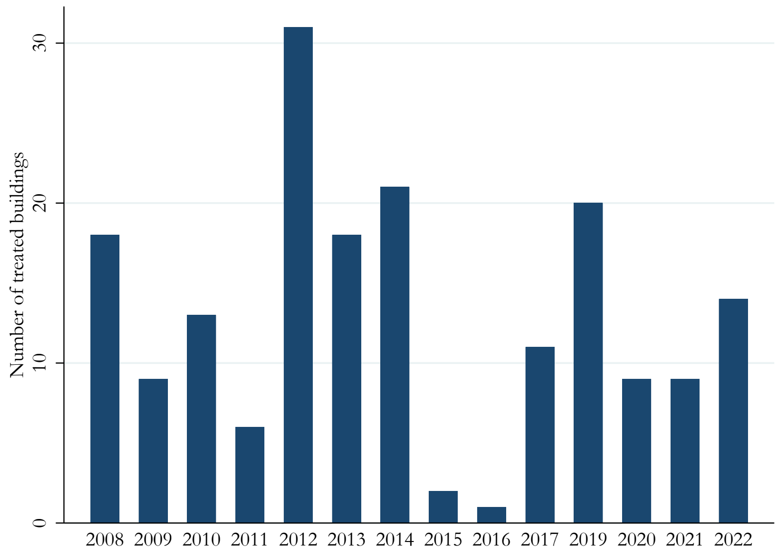


Figure 2: Year of submeter installation

We use several other data sources to account for potential confounders. For each apartment in our sample, we observe vacancy spells and tenant changes. In addition to energy expenses, for a subset of our data, we observe quantities in kWh. This information allows us to calculate heating oil and natural gas prices by NUTS2 region, as well as the price of district heating and other heating sources on the national level. Weather data from the Swiss Federal Office for Meteorology and Climatology allows us to control for local heating degree days.

4 Empirical analysis

4.1 Average effect of submetering

Within the setting described above, we estimate the change in heating expenses attributable to submetering. Our parameters of interest are the average treatment effect on the treated (ATT), and the "dynamic" $ATT(r)$ for r periods after submetering is installed. Formally, these parameters can be defined as

$$ATT = \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0)|S_i = 1; Post_t = 1],$$
$$ATT(r) = \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0)|S_i = 1; Post_t = 1; t - q_i = r],$$

where $Y_{i,t}(1)$ and $Y_{i,t}(0)$ are *potential* outcomes for apartment i in year t under the treated (submetered) and untreated (common heating billing) scenarios, respectively;[7] S_i is a treatment indicator, equal to one for all apartments that have been individually metered during our sample period, zero otherwise; $Post_t$ is an indicator equal to one for post-treatment years, zero otherwise; q_i is the year in which the submetering technology was installed in apartment i, such that r represents years since the installation.

To estimate ATT and $ATT(r)$, we leverage the staggered introduction of submetering within a difference-in-differences framework. Both not-yet-treated and never-treated buildings serve as a comparison group to the buildings that were individually metered during our sample period.

Heterogeneity-robust specifications

As shown in Equation 5 from our theoretical framework, we expect heterogeneity in submetering treatment effects in our setting. Also, recall that we observe buildings that were submetered at different points in time (Figure 2). Put together, these facts pose a challenge for estimation of the ATT and $ATT(r)$ described above. For example, recent econometric literature on difference-in-differences (e.g., Borusyak, Jaravel, and Spiess, 2024; Goodman-Bacon, 2021) advises against the implementation of standard two-way fixed effects (TWFE) in these types of settings. That is because standard TWFE specifications require an assumption of *homogeneous treatment effects*. For staggered rollout designs, in particular, standard TWFE can lead to near-term bias (Goodman-Bacon, 2021), since estimates would over-weight sample regions with higher treatment variance (i.e., typically where there is more switching from an untreated to a treated status).

To avoid this type of bias, we implement the estimator from Sun and Abraham (2021).[8] The building block of their approach is a "fully dynamic" regression specification for *cohort-specific* effects, where cohorts are defined based on the timing of treatment (Figure 2). Within our setting, the fully dynamic specification can be written as:

$$Y_{i,t} = \sum_{r \neq -2} \sum_{c=2008}^{2022} \beta_{r,c} S_i \times \mathbb{1}[r = t - (q_i)] \times \mathbb{1}[q_i = c] + \gamma_i + \gamma_t + \varepsilon_{i,t} \ , \tag{10}$$

[7] Note that we use the subscript i to refer interchangeably to apartments and households. We acknowledge that this will only be accurate for a setting without tenant changes (i.e., without turnover). That is not the case in our data. In Section 4.2 we discuss the potential impacts of turnover in detail. We show results for a sub-sample of apartments without turnover. We also test the extent to which submetering itself may induce tenant changes.

[8] In Appendix D we present results from standard two-way fixed effects specifications. Those generally suggest stronger treatment effects than our preferred heterogeneity-robust procedure. We refrain from interpreting estimates from TWFE, as they might be biased.

where $Y_{i,t}$ is the observed outcome of interest, heating expenses, measured in Swiss Francs (CHF) or in logs; S_i is a time-fixed variable, equal to one for apartments that will be submetered, zero otherwise; $\mathbb{1}[r = t - q_i]$ are indicators for years r relative to the treatment dates q_i; $\mathbb{1}[q_i = c]$ are the cohort c indicators; γ_i are apartment fixed effects; γ_t are year fixed effects; and $\varepsilon_{i,t}$ is an idiosyncratic error term. For inference, we cluster standard errors at the property level (which can include multiple buildings), because the decision to introduce submetering may occur at this level.

Note that q_i represents the year in which the submetering technology was installed. However, submetered bills are only charged in the billing year after that installation (in $q_i + 1$, or $r = 1$). Also, letters or notifications of submetering might have been sent in the year before installation (in $q_i - 1$, or $r = -1$), such that there is scope for households to react in anticipation of submetering. For this reason, in our regression specifications we set $r = -2$ as the omitted comparison year.

Equation 10 estimates separate coefficients $\hat{\beta}_{r,c}$ for each relative year r and for each cohort c. The procedure from Sun and Abraham (2021) re-weights and linearly combines these coefficients according to the shares of each cohort in each time period.[9] This results in average estimates $\hat{\beta}_r$ for each relative time. For $r \geq -1$, the estimated coefficients can be interpreted as the post-treatment effects (corresponding to $ATT(r)$). By taking a simple average those time-dependent effects, we can estimate the average submetering effects for the whole post-treatment sample $\hat{\beta}$ (corresponding to ATT).

The coefficients for $r < -2$ (pre-treatment) can be used to assess our main identifying assumption: parallel trends. That is, the outcomes for both the individually metered and the comparison apartments would have followed parallel paths in the absence of treatment. If pre-treatment $\hat{\beta}_r$ are close to zero, then we have evidence that this holds at least for the pre-treatment sample. We assume that paths would have remained parallel also during the post-treatment periods if submetering had not happened. Note that this assumption implies that our empirical strategy does *not* require quasi-random assignment of treatment dates (see de Chaisemartin and D'Haultfoeuille, 2020; Sun and Abraham, 2021). Nevertheless, in Appendix Tables A.1 and A.2 we show outcome and covariate averages across treatment cohorts, and find no systematic differences in patterns when comparing early- versus late-treated buildings.

We also implement a variant of Equation 10 including time-varying controls. Specifically, we add: heating degree days (HDD) and squared HDD; flexible controls for the time since the last renovation, by type of renovation; and canton-by-year fixed effects. Henceforth, we refer

[9]The weights are obtained from an auxiliary regression that estimates propensity of treatment, depending only on the relative time indicators and the size of each cohort (Sun and Abraham, 2021).

to the specifications that include this full set of controls as the "saturated" specifications or models. For specifications with added controls, we rely on a conditional parallel trends assumption.

Results

Figure 3a presents results for the heterogeneity-robust specification in Equation 10. The outcome is the log of annual heating expenses. Blue dots represent point estimates of $ATT(r)$ for the model with no controls other than year and apartment fixed effects. Red triangles are point estimates for the saturated model with time-varying controls. Visual inspection of the pre-treatment coefficients ($r < -2$) in both panels suggests that pre-trends are limited. None of the pre-treatment coefficients are statistically significant. This supports the parallel trends assumption and causal interpretation of the post-treatment coefficients.

The post-treatment coefficients suggest a drop in annual heating expenses starting already in $r = -1$, at least in the model with no controls. The implication is that households may have reacted to the mere announcement of submetering. Announcement effects are consistent with the results in Elinder, Escobar, and Petré (2017). The drop in energy expenses becomes stronger in the year of submeter installation ($r = 0$) and increases further in the first year of submetered billing ($r = 1$). For our estimates of ATT, we pool the effects from periods from $r = 1$ to $r = 5$, which we consider to be more "stable" effects after households have fully adjusted to submetered billing. The resulting pooled ATT represents a 20.8% reduction in energy expenses in the specification without controls, and a 17% reduction in the saturated model.

Figure 3b presents estimates with heating expenses in CHF as outcome variable. These results are qualitatively similar to those in Figure 3a. Overall, for the period from $r = 1$ to $r = 5$, we estimate a CHF 188 reduction in energy expenses in the specification without controls, and a CHF 144 reduction in the saturated model. Given the 972 average heating expenses in our treated sample, these effects are somewhat smaller than those in Figure 3.

Appendix Figure D.1 presents standard event study estimates obtained via two-way fixed effects. We find similar results to those in Figure 3a. The point estimates for $ATT(r)$ are slightly larger in magnitude, and the confidence intervals are wider. The potential bias in TWFE estimates does not appear to be large in our setting. Nevertheless, we draw our main conclusions from the heterogeneity-robust estimates from Sun and Abraham (2021).

Submetering may increase tenant turnover. One reason for increased turnover could be that some households see a substantial increase in their heating bill. Figure 4 shows results from heterogeneity-robust specifications (Equation 10), but with cumulative tenant turnover as the outcome variable (i.e., an apartment-level count variable that increases by one unit each time that there is a tenant change in a given apartment). These results indeed suggest that submetering increases turnover. In the model without controls, we estimate a 0.26 effect on cumulative turnover. This effect is reduced to 0.15 in the saturated model, which suggests that approximately one in seven households leaves the building because of submetering. This finding may have important consequences for the interpretation of our results and for our welfare analysis. We revisit this question in Section 4.2, where we discuss the potential implications of turnover in detail.

The introduction of submetering can be accompanied by other measures, such as energy efficiency renovations. In Appendix F.2, we assess how often submetering is introduced in conjunction with various types of renovations. We find that energy related renovations are 20 percentage points more likely in $r = -1$. This pattern may explain why our saturated model finds smaller effects on heating energy than the model without controls—concurrent renovations inflate the effects of submetering. We further explore renovations in Section 4.2, where we investigate whether treatment effects differ between renovated buildings and other buildings. Renovations are also important when we consider effects of submetering on rent in Appendix F.3. As renovations are usually more expensive than the installation of submeters, we find sizable rent increases when submetering concurs with renovations. When submeters are installed without a concurrent renovation, we do not find statistically significant rent increases.

4.2 Heterogeneous effects of submetering

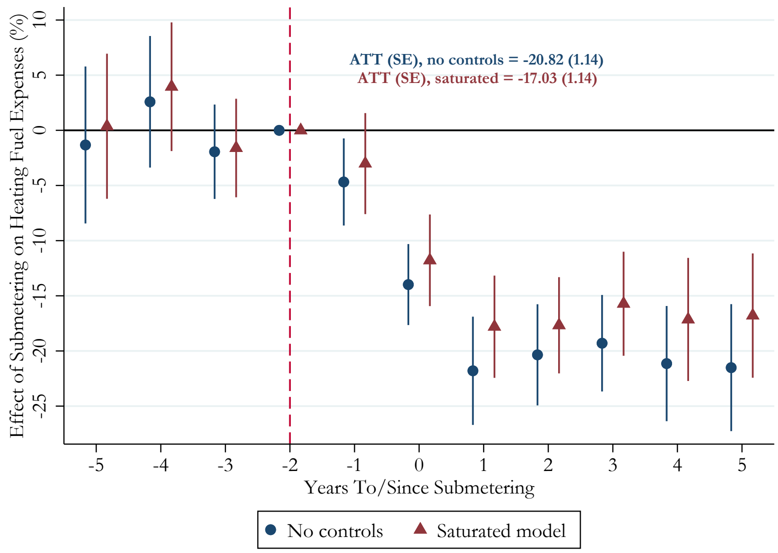
To explore heterogeneity and the mechanisms through which submetering might lead to reduced heating expenses, we analyze *conditional* average treatment effects on the treated ($CATT$). That is, we are interested in understanding how the effects of submetering may depend on some key covariates. Let $CATT$ be defined as

$$CATT(\mathbf{c}) = \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0) | S_i = 1; \mathbf{X}_{i,t} = \mathbf{c}; Post_t = 1] \,, \tag{11}$$

where $\mathbf{c}$ denotes a set of conditions or restrictions on some covariates $\mathbf{X}_{i,t}$ of interest. Our theoretical framework from Section 2 motivates our choice of $\mathbf{X}_{i,t}$ and $\mathbf{c}$. For example, our framework suggests that the effect of submetering depends on the number of neighbors N that previously shared a common heating bill.

One natural approach for estimating $CATT$ would be to run a regression as in Equation 10, but restricted to a sub-sample of apartments that have many neighbors (large N), for example. Such an approach has a few drawbacks. First, the comparison (not submetered) group of buildings in that regression would also be restricted, thus different from the original comparison group used to estimate ATT. This may not necessarily be an issue per se, but needs to be taken into account when interpreting the sizes of coefficients (as the average out-

(a) Effects of Submetering on Heating Expenses (%)



(b) Effects of Submetering on Heating Expenses (CHF)

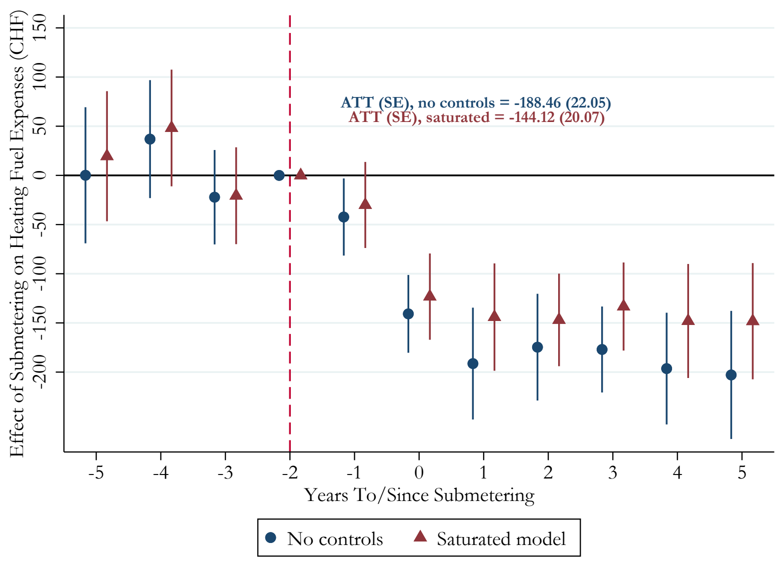


Figure 3: Heterogeneity-robust Effects of Submetering

Notes: This figure presents estimates of both ATT and $ATT(r)$ for the effects of submetering according to the procedure from Sun and Abraham (2021) described in Section 4.1. For Panel (a) the outcome variable is heating expenses in logs, while for Panel (b) it is in levels (CHF). The ATT estimates and their standard errors, included as text within the graphs, pool the effects from periods $r = 1$ to $r = 5$. The $ATT(r)$ estimates are presented graphically. For Panel (a), estimates are transformed to represent percent effects $(100 \times (exp(\beta_1) - 1))$. All specifications include year and apartment fixed effects. The saturated specifications also control for heating degree days, renovations, and Canton-by-year fixed effects. Standard errors are clustered at the property level. The vertical lines around the point estimates represent 95% confidence intervals.

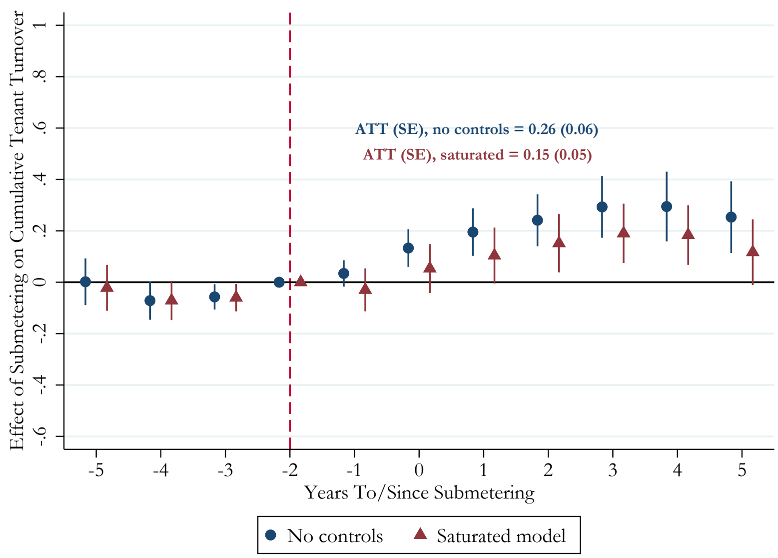


Figure 4: Heterogeneity-robust Effects of Submetering on Cumulative Turnover

Notes: This figure presents estimates of both ATT and $ATT(r)$ for the effects of submetering according to the procedure from Sun and Abraham (2021) described in Section 4.1. The outcome variable is cumulative turnover (i.e., the cumulative number of tenant changes observed for a given apartment over our sample period). The ATT estimates and their standard errors, included as text within the graphs, pool the effects from periods $r = 1$ to $r = 5$. The $ATT(r)$ estimates are presented graphically. All specifications include year and apartment fixed effects. The saturated specifications also control for heating degree days, renovations, and canton-by-year fixed effects. Standard errors are clustered at the property level. The vertical lines around the point estimates represent 95% confidence intervals.

come of the comparison group might differ from the original sample). Second, sub-sampling can lead to losses in statistical efficiency, as it reduces the number of observations used for estimation.

Both of the aforementioned drawbacks can be partially addressed with the implementation of a regression that interacts the covariate of interest (or bins of that covariate) with the submetering treatment indicator. However, that specification would not address another key concern which arises when we are interested in multiple correlated covariates that drive heterogeneity. For example, it could be that apartment size is correlated with the number of neighbors in buildings. This may bias "naive" estimates of heterogeneity by number of neighbors, in case apartment size also drives heterogeneity in treatment effects.

To simultaneously account for several factors that can potentially drive heterogeneity of submetering effects, we turn to ML-based techniques for estimation of *conditional average effects*. In particular, we implement a method that is based on counterfactual imputation (Borusyak, Jaravel, and Spiess, 2024; Souza, 2019). The first step is to predict $[Y_{i,t}(0)|S_i = 1; Post_t = 1]$, which represents post-treatment heating expenses under a counterfactual scenario of no submetering.

Prediction of untreated counterfactuals

Following Souza (2019), we use data from never-treated and not-yet-treated apartments to build a model $g(\mathbf{X}_{i,t})$ to predict untreated energy expenses $Y_{i,t}(0)$. We estimate $\hat{g}(\mathbf{X}_{i,t})$ using all observations from the never-treated buildings, as well as pre-treatment data from the not-yet-treated buildings. Importantly, $\mathbf{X}_{i,t}$ are predictor variables that are not affected by treatment. These include: area of the apartments; number of neighbors in the building; number of rooms per apartment; location of the apartment (floor) within the building; presence of commercial establishment in building; billing cycle indicators; heating degree days, indicators for construction and renovations years; zipcode and canton fixed effects; year fixed effects; and numeric building IDs.[10]

We use ML algorithms to estimate $\hat{g}(\mathbf{X}_{i,t})$. The *out-of-sample* prediction performance of the algorithms is assessed via 4-fold cross-validation. That is, first we randomly split the pre-treatment data into 4 equally-sized sub-samples. We use a "stratified" sub-sampling procedure, such that all (annual) observations from a given building are contained within the same sub-samples. Then, we use 3 of the partitions to train the ML algorithms, and predict

[10]Our preferred prediction algorithm essentially combines many "simple" regression trees, each constructed with only a subset of the candidate covariates. For this reason, it is possible to include nested fixed effects (e.g., zipcode and canton) without collinearity issues. Building IDs are included as a continuous variable (not as a fixed effects), to prevent the algorithm from building too many uninformative trees that contain only fixed effects. Also, note that we do not include any tenant characteristics, which would otherwise contaminate our counterfactual predictions for apartments that experience tenant turnover.

energy expenses on the 4th "hold-out" sub-sample. By recursively repeating this process 4 times, each time changing the hold-out sub-sample, we obtain out-of-sample predictions for all observations. Finally, those predictions are compared to the true pre-treatment expenses, which allows us to calculate prediction performance metrics such as root-mean-square error (RMSE).[11]

The "best-performing" algorithm is that which achieves the lowest cross-validation RMSE. We perform a grid search varying parameter configurations of gradient boosted trees, implemented through XGBoost (Chen and Guestrin, 2016). Since XGBoost is a tree-based algorithm, it automatically accounts for nonlinear relationships between the outcome and the covariates based on the branch splits. It also captures interactions between covariates, with deeper trees allowing for more complex interactions. In Appendix Table E.2 we present the cross-validation RMSE for the 18 algorithm configurations in our grid search. Our preferred configuration is an "ensemble," combining the predictions from three of those configurations, highlighted in gray.[12]

The RMSE provides a good metric for the average performance of the algorithm. However, we are also interested in checking if predictions are not biased at different time periods relative to submetering treatment. This is akin to checking the "parallel trends" assumption in difference-in-differences settings. For this, again we use a specification based on Souza (2019). Let $\hat{\varepsilon}^{cv}_{i,t} = Y_{i,t} - \hat{Y}^{cv}_{i,t}$ be the pre-treatment cross-validation residuals. We then regress these residuals on event-time indicators as follows:

$$\hat{\varepsilon}^{cv}_{i,t} = \sum_{r<-1} \beta_r S_i \times \mathbb{1}[r = t - (q_i)] + u_{i,t} \; . \tag{12}$$

The coefficients β_r will thus represent average pre-treatment residuals at different event times. Estimates from Equation 12 are presented in Figure 5a. Based on an F-test (shown in graph), we cannot reject that these coefficients are jointly zero, such that the ML algorithm is unlikely to be biased in the pre-treatment sample. We can then proceed to use this algorithm to predict (untreated) counterfactual energy expenses in the post-treatment

[11]More details are presented in Appendix E.1. We also implement an alternative CV strategy, without the stratified sub-sampling. We argue that performance metrics obtained from the stratified CV procedure are better representative of potential errors for counterfactual prediction. This is because the stratified procedure guarantees that predictions for a given building are always obtained based on a model trained with data from *other* buildings. This is analogous to the intuition behind canonical difference-in-differences. Regardless, in our setting, similar algorithm configurations would be selected with standard k-fold CV or stratified CV procedures.

[12]Ensemble methods combine the outcomes of several algorithms, resulting in predictions that are more accurate than those obtained for each algorithm separately (Dong et al., 2020; Van der Laan, Polley, and Hubbard, 2007). We use non-negative least squares to determine the ensemble weights in our setting. Details are described in Appendix E.1.

sample: $\hat{Y}_{i,t}(0) = \hat{g}(\mathbf{X}_{i,t})$ for $t \geq -1$.[13] This requires an implicit assumption that the model $g(\mathbf{X}_{i,t})$ for untreated energy expenses would have remained stable over time, in the absence of submetering (Souza, 2019).

In Appendix F.1, we test this "stability" assumption with a placebo treatment exercise. That is, we take the sample of never-treated buildings and randomly allocate some of them to a placebo treatment condition. We then perform the ML training/predictions as described above, but with this new semi-synthetic data. Appendix Figure F.2 presents the prediction errors across event times. We find that the errors are also close to zero both before and after the placebo treatment. This suggests that our model can accurately predict future consumption patterns in a counterfactual scenario without submetering.

Returning to the real data/treatment, our primary interest is on estimating treatment effects. We obtain those by subtracting the untreated counterfactual predictions from the realized post-treatment energy expenses:

$$\hat{b}_{i,t} = Y_{i,t} - \hat{Y}_{i,t}(0)\ . \tag{13}$$

The estimates $\hat{b}_{i,t}$ correspond to apartment-by-year treatment effects. These can be used to summarize submetering treatment effects across sub-samples of interest.

Summarizing heterogeneous treatment effects

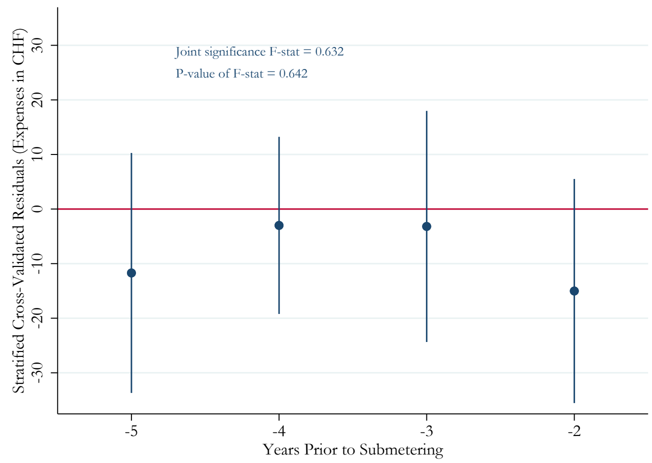
Note that the ML approach allows estimation of how treatment effects change over time (i.e., estimation of $ATT(r)$). This can be done by calculating the average of $\hat{b}_{i,t}$ at different times relative to treatment. We summarize these estimates in Figure 5b. For event times -2 or before, the figure presents average residuals for pre-treatment observations, while average treatment effects are estimated for event times -1 and later (the post-treatment sample). We note that these ML estimates are within range of those obtained with the Sun and Abraham (2021) procedure. Reductions in expenses are statistically significant, ranging from 110 CHF (on the year of submetering installation), to 190 CHF (one year after submetering). The somewhat wide confidence intervals ("whiskers" around the point estimates) can be attributed to our conservative clustering approach (at the property level), but might also be related to substantial heterogeneity in the effects.[14]

To further explore how heterogeneity may depend on covariates, we summarize $\hat{b}_{i,t}$ across different sub-samples of interest. As highlighted in the theoretical framework, Equation 5, we may expect heterogeneity depending on the number of neighbors (N), the apartments'

[13] Recall that we allow for some anticipation effects by defining "post-treatment" as starting one year before submetering installation.

[14] This would be in line with our theoretical framework. Also, while average expenses are reduced, one could expect that some tenants face an increase in expenses (e.g., those who prefer higher temperatures).

(a) Average CV residuals by event time



(b) Machine learning estimates of $ATT(r)$

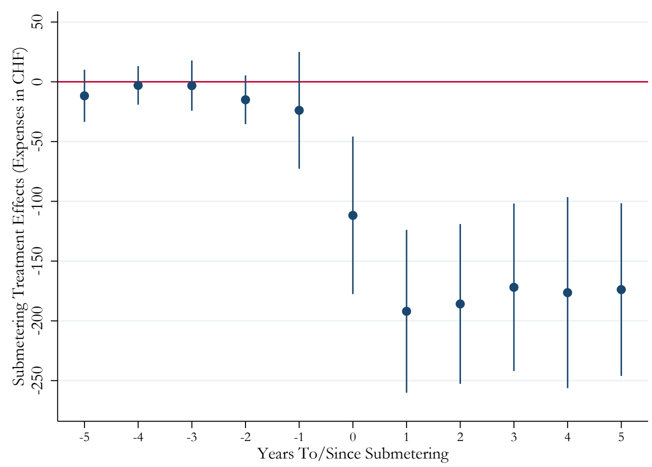


Figure 5: Machine learning average residuals and $ATT(r)$ estimates

Notes: Panel (a) presents coefficient estimates obtained from regression Equation 12. These represent the average cross-validated residuals for time periods prior to submetering. The reported F-statistic (and corresponding p-value) tests the joint significance of the coefficients. Panel (b) summarizes $ATT(r)$ according to the machine learning estimation strategy. We plot average residuals for event times -2 or before, and average effects for event times -1 or later. Standard errors are clustered at the property level. The "whiskers" around point estimates correspond to 95% confidence intervals.

heat loss (H), tenants' sensitivity to deviations from preferred temperature (κ_i), and energy prices (p). In this sub-section, we explore the first two dimensions because they can easily be used to define submetering policies. While N is directly observable, we use apartment area and building vintage as proxies for heat loss.

Heterogeneity estimates may be obtained through a sub-sampling approach, such as taking the average of $\hat{b}_{i,t}$ for buildings with small, medium, or large number of neighbors, as shown in Appendix Figure E.4. Alternatively, we may regress the treatment effects $\hat{b}_{i,t}$ on a constant plus the binned indicators for number of neighbors, omitting one of the groups to serve as a benchmark for comparison. That directly allows us to test whether any differences between groups are statistically significant. The resulting coefficients are represented by the blue triangles in Appendix Figure E.5. However, those should still be considered "naive" estimates, as they ignore other potential drivers of heterogeneity, such as apartment area, that might be correlated with the number of neighbors. For this reason, we refrain from interpreting those coefficients. Rather, as described at the top of Section 4.2 we take a holistic approach for heterogeneity, summarizing effects across sub-samples while controlling for other factors.

To simultaneously account for the multiple dimensions of heterogeneity in this setting, we regress the submetering treatment effects $\hat{b}_{i,t}$ on a constant, plus binned indicators for number of neighbors, apartment area, and building vintage, as shown in Equation 14 below. We include binned (rather than continuous) versions of these covariates in order to capture nonlinear effects. We use a data-driven approach developed by Cattaneo et al. (2024) to determine the optimal number of bins for each variable (details in Appendix E.3).[15] Additionally, we include canton fixed effects (δ_m), to parse out heterogeneity that may be attributed to regional variations in energy prices. The full specification can be written as:

$$\begin{aligned}\hat{b}_{i,t} = \alpha + \sum_{n=2}^{3} \beta^n \mathbb{1}[\text{Neighbor Bin} = n]_i + \sum_{a=2}^{3} \beta^a \mathbb{1}[\text{Area Bin} = a]_i + \\ \sum_{v=2}^{4} \beta^v \mathbb{1}[\text{Building's Vintage} = v]_i + \delta_m + \varepsilon_{i,t}\ .\end{aligned} \tag{14}$$

The coefficients of interest are $\hat{\beta}^n$, $\hat{\beta}^a$, and $\hat{\beta}^v$, which estimate how the effects of submetering depend on number of neighbors, apartment area, and building's vintage, respectively.

[15] Note that there is a bias-variance tradeoff when defining bin splits (Cattaneo et al., 2024). With more splits, the researcher may be able to uncover more nuanced heterogeneity patterns. However, this comes at a risk: in case bins are too small, some spurious effects may be mistaken for true heterogeneity patterns. The procedure from Cattaneo et al. (2024) takes that into account, resulting in (not too small) bins that contain approximately the same number of (post-treatment) observations.

Importantly, these are jointly estimated with the same regression, allowing us to understand the contribution of each covariate while holding other factors fixed. Results are presented in Table 1, column (1). We note that there is no significant difference between apartments with up to 8 neighbors versus those with 9 to 12 neighbors. Conversely, there is a strong and statistically significant difference for apartments with more than 12 neighbors. For those, the point estimate suggests that submetering induced an *additional* reduction in expenses of approximately 111 CHF. A larger treatment effect in buildings with many neighbors is in line with our theoretical framework, as the implied marginal price change is larger (see Equation 5).

Our results from Table 1 also reveal a relationship between submetering treatment effects and apartment area. Large apartments (more than 88 sq. meters) reduced heating expenses by additional 115 CHF compared to small apartments (up to 71 sq. meters). The difference for medium apartments (72 to 88 sq. meters) is close to 56 CHF. Again, this is consistent with our framework, assuming that large apartments have higher heat loss. Regarding building's vintage, the naive estimates (Appendix Figure E.5) suggest that treatment effects were stronger for older buildings. However, those differences vanish when we control jointly control for other factors. This may be due to correlations between building vintage and number of neighbors, for example.[16]

We test the robustness of our heterogeneity estimates (β^n, β^a, and β^v) to changes in Equation 14. In columns (2) to (4) of Table 1 we successively add controls. We find that the coefficients for number of neighbors and apartment area are stable across specifications. Further, this exercise reveals other potentially important drivers of heterogeneity.

Column (2) of Table 1 presents results controlling for post-treatment turnover.[17] We find that apartments that experienced tenant changes save an additional 58 CHF, compared to apartments without turnover. Column (3) shows that effect remains significant even after controlling for pre-treatment turnover rates (event years -2 and before). The influence of turnover is somewhat weaker (39 CHF) once we additionally control for energy-related renovations and vacancies, as shown in column (4). Although the coefficient on vacancies is strong, it is only relevant for a small fraction of the treated apartments (vacancies happen in only 8% of the post-treatment sample). The result on turnover has important implications for our welfare calculations, which we discuss below. Finally, in column (5) we test whether submetering treatment effects vary depending on the apartment's location within the build-

[16]The large buildings (with many neighbors) in our treated sample tend to be older. Newly constructed large buildings may have already been submetered from the beginning, so they drop out from our sample. These constitute an "always-treated" group that is not used in the estimation procedures presented so far. However, the always-treated buildings are used in Section 4.4, where we explore price elasticities.

[17]Here we tag apartments that experienced any tenant changes during submetering event times -2 to +2.

ing. We do not find statistically significant differences between apartments on the ground floor versus those on intermediate and top floors. This finding suggests that heat flows between apartments are not a major concern in our setting. If heat flows between apartments were substantial, we would expect larger treatment effects in apartments on intermediate floors.

Note that the validity of the estimates described above relies on an assumption that the ML algorithms are unbiased along the relevant heterogeneity dimensions. Essentially, we need to rule out that prediction errors from our ML algorithm are driving the estimated heterogeneity shown in Figure E.5. To do so, we run a regression analogous to Equation 14, but with ML cross-validated residuals as the outcome. We test whether these residuals are systematically larger for the sub-samples of interest. Results are presented in Appendix Figure E.1. Since most coefficients are statistically indistinguishable from zero, we conclude that our ML algorithm is unlikely to be biased for our covariates of interest. In Appendix Figures E.2 we show that the prediction errors are uncorrelated with several other covariates.[18]

Strategic turnover

We observe results that are consistent with tenants leaving the building if submetering increases their heating bill by more than their moving costs. Not only does submetering increase turnover (see Figure 4), but the induced turnover appears to influence heating expenses: submetering treatment effects are larger for apartments that experiecne tenant changes (see Table 1). These effects cannot be explained by renovations and the rent increase that comes with them (see Appendix F.3). This finding has implications for the interpretation of our results and for our welfare analysis.

To model strategic tenant exit, we consider three types of tenants: stayers (s), leavers (l), and newcomers (n). Our empirical estimate $\hat{\tau}$ is the weighted average of: (i) the treatment effect on stayers; and (ii) the difference between newcomers under the submetering regime and leavers under the common heating bill regime. Formally, we can write this as:

$$\hat{\tau} = \frac{\sum^{s}(e^{*}_{s,\theta=1} - e^{*}_{s,\theta=0}) + \sum^{n}(e^{*}_{n,\theta=1} - e^{*}_{l,\theta=0})}{N} \; , \tag{15}$$

where $e^{*}_{\cdot,\theta}$ denotes energy consumption for each type of tenant, also depending on the billing regime θ. We note that $\hat{\tau}$ is subject to a selection effect. In particular, $e^{*}_{n,\theta=1}$ can be lower than $e^{*}_{l,\theta=0}$ not only because of submetering, but also because leavers might be a selected sample. In particular, households with high T_i will lose from submetering, such that some

[18] Average errors across these other covariates are presented as a form of robustness check in Appendix E.2. We also do not find that these covariates are significant drivers of heterogeneity in our setting.

Table 1: Summarizing treatment effect heterogeneity across subgroups

	(1)	(2)	(3)	(4)	(5)
Number of Neighbors (comparison group: 8 or less)					
9 to 12	-8.63	-4.26	-59.10	-51.56	-48.20
	(52.57)	(50.58)	(55.96)	(54.47)	(54.19)
more than 12	-110.99**	-103.30**	-127.16**	-112.22**	-108.91**
	(45.45)	(44.37)	(47.73)	(44.72)	(44.32)
Area of Apartment (sq. meters; comparison group: 71 or less)					
72 to 88	-55.96	-57.92*	-85.78**	-81.43**	-81.43**
	(34.61)	(34.36)	(38.57)	(36.69)	(37.00)
more than 88	-115.08***	-115.58***	-152.12***	-134.32***	-134.08***
	(35.15)	(34.57)	(43.50)	(44.87)	(44.83)
Building's Vintage (comparison group: built in the 90s or later)					
No data	-82.90	-91.12	-90.98	-100.31	-102.18
	(95.37)	(94.50)	(109.23)	(108.86)	(109.38)
60s or earlier	-41.67	-45.41	-47.88	-24.78	-27.64
	(65.90)	(64.39)	(107.33)	(103.25)	(104.12)
70s or 80s	-25.80	-30.54	-34.23	-18.37	-19.32
	(53.49)	(50.81)	(62.69)	(65.29)	(66.16)
Post-treat. Turnover (0/1)		-57.62***	-59.32***	-39.40**	-39.26**
		(16.04)	(17.58)	(16.56)	(16.84)
Pre-treat. Turnover Share (%)			-3.81*	-3.59*	-3.62*
			(2.15)	(2.10)	(2.10)
Energy-related Renovations (0/1)				-48.68	-49.09
				(48.26)	(48.62)
Duration of Vacancies (comparison group: no vacancies)					
Vacant 1 to 183 days				-77.90***	-78.18***
				(20.64)	(20.79)
Vacant 184 to 366 days				-286.99***	-286.51***
				(43.86)	(43.82)
Location of Apartment in the Building (comparison group: ground floor)					
Intermediate Floor					-16.56
					(15.31)
Top Floor					-5.89
					(26.36)
Regression Constant	-56.64	-29.02	34.65	37.73	48.17
	(50.80)	(50.56)	(73.39)	(74.07)	(76.61)
Number of observations	6,856	6,856	5,748	5,748	5,730

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Column (1) presents results from the regression in Equation 14. Column (2) additionally controls for post-treatment turnover. Column (3) additionally controls for pre-treatment turnover rates. Column (4) additionally controls for renovations and vacancies. Column (5) tests the influence of apartments' locations in the building. Vintage information was not available for some buildings (rather than dropping those, we group them into a separate category labeled as "No data"). All specifications control for canton fixed effects. Standard errors (in parentheses) are clustered at the property level.

of them might decide to leave the building.[19] The resulting selection effect increases our estimate of the treatment effect.

Our results from Table 1 suggest that leavers have higher T_i than newcomers. Note that if stayers, leavers, and newcomers had similar preferences (T_i and κ_i), we would not expect a significant coefficient on turnover, because tenant turnover would not matter on average. Our specifications allow us to account for this selection effect by focusing on conditional average treatment effect for stayers:

$$\hat{\tau}_{stay} = \frac{\sum^{s}(e^{*}_{s,\theta=1} - e^{*}_{s,\theta=0})}{N_s} \ . \tag{16}$$

Focusing on $\hat{\tau}_{stay}$ has two main advantages compared to the unconditional ATT. First, it is not subject to the selection effect induced by strategic tenant exit. This is important for our welfare analysis. Within our framework, the selection effect does not improve welfare—it merely shifts consumption outside of our sample, which we cannot observe or evaluate. Second, it allows us to abstract from the prevalence of submetering in the overall housing market, which would matter for newcomers and leavers. For example, if all buildings in the market had submetering, then leavers would receive submetering also in their new building. If no other buildings had submetering, then newcomers will receive submetering, and leavers will not. To abstract from these complications, we focus on $\hat{\tau}_{stay}$ in our welfare analysis.

Building-specific effects

We use the heterogeneity coefficient estimates $\hat{\beta}^a$, $\hat{\beta}^n$, and $\hat{\beta}^v$ from Equation 14 to calculate *building-specific* submetering effects. This is essentially an aggregation step. Note that the model in Equation 14 yields apartment-specific effects that vary depending on covariates.[20] We aggregate those to the building level by summing the resulting $CATT$ from all apartments that belong to the same building.[21] We then divide that sum by the number of neighbors in the building, resulting in average savings per household.

Results are presented in Figure 6. The solid blue curve presents a ranking of the building-specific savings, according to the model for stayers ($\hat{\tau}_{stay}$, based on Table 1, column 2). The dashed blue curve is for a model that ignores the issue of turnover (Table 1, column 1). The comparison of both curves again reveals that newcomer tenants in our context are likely to have lower T_i preferences than tenants that left the buildings after submetering.

[19] A similar argument can be made about households with high κ_i.

[20] For this exercise, we ignore any heterogeneity implied by canton fixed effects.

[21] We perform this aggregation for a restricted sub-sample for which we observe the same set of apartments within their respective buildings over time. This corresponds to about 80% of the treated sample.

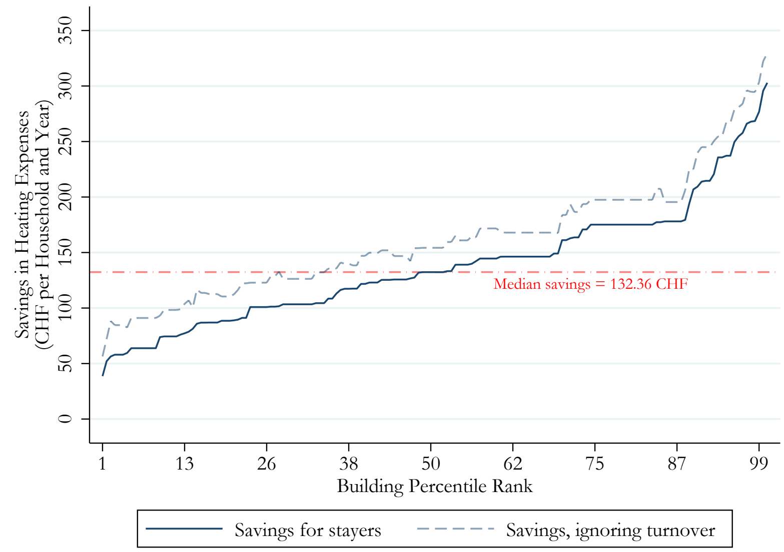


Figure 6: Building-specific treatment effects

Notes: This Figure plots ranked building-specific treatment effects. Buildings are ranked from lowest savings (left) to highest savings (right). The x-axis represent percentiles, the y-axis savings in energy expenses (CHF per household).

We focus on interpreting the effects for stayers. The Figure reveals substantial heterogeneity, with savings ranging from 40 to 300 CHF per household annually. This heterogeneity has implications for welfare, which we discuss next.

4.3 Welfare evaluation

We use the sufficient statistic in Equation 9 to assess the welfare effects of submetering. The relevant parameters are the treatment effect for stayers $\hat{\tau}_{stay}$, the number of neighbors N, the social cost of carbon ϕ, the energy price p, and the cost of submetering s.

We start with a hypothetical building that is representative of our sample. The median treatment effect for stayers $\hat{\tau}_{stay}$ is CHF 132 per year, the median number of neighbors N is 7, and the average energy price in our treatment sample is CHF 0.0883 per kWh. We assume a cost of submetering of CHF 86 per year.[22] We use the SCC estimate from Rennert et al. (2022) of 185 CHF/tCO_2 to value reduced CO_2 emissions.[23] We also account for the

[22]The cost of submetering consists of investment costs (hardware and installation) as well as running service fees. We observe service fees in our data. On average, service fees amount to CHF 41 per apartment and year. Investment costs are harder to assess. We use CHF 45 annualized investment costs as an approximation, based on EnergieSchweiz (2023) who approximate the annualized cost for both heating *and water* at CHF 90. Our cost estimate assumes that heat submetering is responsible for 50% of these costs.

[23]Rennert et al. (2022) present a SCC measured in US dollars per ton of CO_2. During the time period that we study, the exchange rate between US dollars and Swiss francs fluctuated around 1. At the time of writing this manuscript, one Swiss franc is approximately equivalent to 1.1 US dollars.

carbon tax implemented in Switzerland, which had a median value of of CHF 36 during our sample period.

Given these values for our hypothetical building, we use Equation 9 to calculate the welfare effect of submetering. The first term of Equation 9 represents household benefits. The resulting value of $-\frac{\hat{\tau}_{stay}}{2}(1-\frac{1}{N}) = \frac{132}{2}(1-\frac{1}{7}) \approx 57$ CHF shows that it is important to account for the loss in thermal comfort, which is generally close to half of the treatment effect on the heating bill. The second term of Equation 9 represents climate benefits. The resulting value of $-\hat{\tau}_{stay}\frac{\phi}{p} = 132\frac{(185-36)0.00022718}{0.0883} \approx 51$ CHF shows that reduced CO_2 emissions are a substantial benefit of submetering.[24] The third term of Equation 9 represents the cost of submetering $s = 86$. Putting all terms together, we find that submetering generates a welfare gain of $57 + 51 - 86 = 22$ CHF per household and year for our hypothetical building.

The welfare estimate for our hypothetical building ignores important heterogeneity. For instance, in Section 4.2, we show that buildings with more neighbors also experience larger treatment effects. The implication is that we might underestimate welfare gains for those buildings, if we simply assume that they experience the unconditional average treatment effect. To more thoroughly investigate how welfare effects differ by building, we take the building-specific expense reductions shown in Figure 6. For each building in our treatment sample, we calculate a welfare effect—just like we did with the hypothetical building above.

Figure 7 shows a ranking of welfare effects in our treatment sample. The light green area represents household benefits, the dark green area climate benefits, and the dashed line the cost of submetering. Welfare effects are positive where the green areas exceed the dashed line. If we abstract from climate benefits, we find that only 10% of buildings in our sample benefit from submetering, with the light green area exceeding the dashed line. When we take climate benefits into account, the share of buildings with positive welfare effects increases to 63%, with a mean welfare effect of CHF 18 per household and year.

4.4 Effect on price elasticity of demand

Our theoretical framework predicts that submetering changes the price elasticity of heating demand. In particular, households who pay for their individual consumption should be more responsive to price changes, as shown in Equation 7. In this section, we test the extent to which this prediction holds empirically.

Empirical strategy

To estimate own-price elasticities of heating energy demand, we use an empirical strategy similar in spirit to Myers (2019). We leverage heating energy price variations across regions,

[24]Note that the factor 0.00022718 tCO_2/kWh represents the carbon intensity of heating energy in our treatment sample.

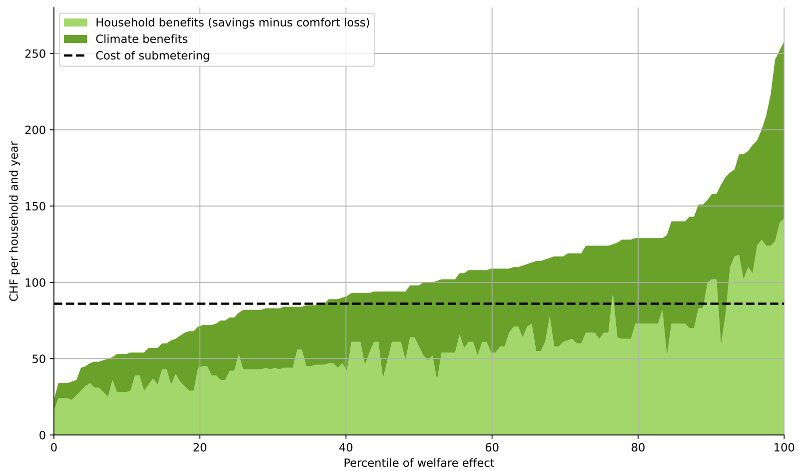


Figure 7: Welfare effects without turnover (CHF per household)

Notes: This Figure plots ranked building-specific welfare effects. Buildings are ranked from lowest (left) to highest (right) welfare effect. The x-axis represents percentiles, and the y-axis measures benefits and costs in CHF. The light green area represents household benefits, the dark green area climate benefits, and the dashed line the cost of submetering. Welfare effects are positive where the green areas exceed the dashed line.

time, and buildings' heating fuel type (i.e., oil, gas, or district heating). We implement the following regression specification:

$$log(Y_{i,t}) = \beta_1 log(p_{r,t,f}) + \beta_2 log(p_{r,t,f}) \times S_i + \mathbf{X}'_{i,t}\delta + \gamma_i + \gamma_{r,t} + \varepsilon_{i,t} \ , \qquad (17)$$

where $log(Y_{i,t})$ is the log of heating energy expenses for household i in year t. We are particularly interested in the coefficients associated with the log of energy prices $log(p_{r,t,f})$.[25] Importantly, energy prices vary by region r, year t, and heating fuel type f. The specification also includes NUTS2-by-year fixed effects $\gamma_{r,t}$, which control for potential demand confounders such as regional changes in environmental awareness or energy saving campaigns. Those fixed effects also imply that price fluctuations would be completely absorbed if there were no variations in heating fuel type within regions. The apartment fixed effects γ_i control for time-invariant apartment characteristics. $\mathbf{X}_{i,t}$ comprises the same control variables used in the saturated model in Section 4.1. Another key regressor is the interaction of the submetering indicator S_i with log energy prices. β_1 estimates the own-price elasticity of energy expenses for households under the common billing regime. β_2 captures the additional

[25] This specification assumes that households are essentially price takers with respect to heating energy fuels. We argue that this is realistic in the case of Switzerland, since these fuels are often supplied by local monopolies and because we study tenant households, as described in Section 3.

price sensitivity of submetered households. We subtract 1 from those coefficients, to reflect elasticities of demand, not expenses.[26]

Different from our analyses in the previous sections, we estimate Equation 17 with a sample consisting of apartments that were either *always* submetered or *never* submetered. That is, we exclude apartments that have changed their submetering status during our sample period. This is to abstract from dynamic treatment effects and adjustment mechanisms that can be expected shortly after submetering. In essence, this section focuses on estimating price elasticities under a "steady state" where households have already adjusted to the billing regime they face.

Finally, we are also interested in estimating how elasticities vary depending on the number of neighbors in the buildings. Equation 4 from our framework predicts that the number of neighbors N only influences elasticities of households under the common billing regime. We test that with a variant of specification 17, where we intereact the log price and the submetering indicator with binned indicators for number of neighbors.

Results

Table 2 reports our elasticity estimates. Standard errors (in parentheses) are clustered at the building level. Column (1) reports results for the model in Equation 17. We find a price elasticity of -0.619 without submetering, and -0.716 with submetering.[27] The difference between the two groups is statistically significant at the 1% level. This implies that submetered apartments are more responsive to price changes than non-submetered apartments, confirming the prediction of our theoretical framework.

Further, Column (2) from Table 2 reveals that the elasticity decreases substantially for large buildings under the common bill. The difference in elasticities between small (8 or fewer neighbors) and large buildings (more than 12 neighbors) is 0.19 (statistically significant with $p < 0.01$). As predicted in our framework, households under the common heating bill are less responsive to changes in heating energy prices when they live in buildings with many neighbors. The number of neighbors does not affect the price-responsiveness of households in submetered buildings. There, the estimated difference in elasticities is at most 0.04 (not statistically significant). This finding is consistent with our theoretical framework, because

[26]The elasticity of energy expenses can be defined as $\eta_{ep} = \frac{\partial ep}{\partial p}\frac{p}{ep} = (\frac{\partial e}{\partial p}p + \frac{\partial p}{\partial p}e) \cdot \frac{1}{e} = \frac{\partial e}{\partial p} \cdot \frac{p}{e} + 1 = \eta_e + 1$.

[27]These estimates are broadly in line with Filippini and Kumar (2021), who find an elasticity of -0.73 for residential gas demand in Switzerland. They are also similar to estimates for other countries. For example, Alberini, Gans, and Velez-lopez (2011) find elasticities for gas ranging from -0.566 to -0.693 in the United States, Lim, Kim, and Yoo (2016) find a short-run elasticity of -0.7 for district heating in South Korea, Schulte and Heindl (2017) find a -0.5 elasticity for space heating in Germany, and Trotta, Hansen, and Sommer (2022) estimate a -0.53 elasticity for district heating in Denmark.

the price incentive of households under submetering does not depend on the number of neighbors.

Table 2: Price elasticities

	(1) Baseline	(2) Number of Neighbors
Common Bill	-0.619***	
	(0.029)	
Common Bill: 8 or fewer		-0.725***
		(0.036)
Common Bill: 9 to 12		-0.679***
		(0.037)
Common Bill: more than 12		-0.537***
		(0.035)
Submetering	-0.716***	
	(0.043)	
Submetering: 8 or fewer		-0.695***
		(0.049)
Submetering: 9 to 12		-0.735***
		(0.067)
Submetering: more than 12		-0.727***
		(0.058)
Number of observations	618,936	618,936

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Dependent variable: log Heating expenses. Results from regression specification 17, which also controls for apartment fixed effects, region-by-year fixed effects, heating degree days, and renovations. Reported coefficients were transformed to reflect price elasticities of heating energy demand (see Footnote 26). Standard errors (in parentheses) are clustered at the building level.

5 Behavioral mechanisms

Fully informed free-riding behavior, as described in our theoretical framework in Section 2, cannot adequately explain our results. Why do households in the common heating bill react to price variation if they have little incentive to do so? We would expect a very low price elasticity under the common heating bill, but find substantial elasticities exceeding 0.5. Given such price elasticities, why do households not react more strongly to submetering? Recall that we estimate that submetering reduced heating energy expenses by 17%. As submetering increases *marginal* heating energy prices by a factor of N, and the median building in our sample has 7 apartments, we may expect a demand elasticity of 0.028—substantially smaller than any of the estimates from Table 2.

In this section, we discuss behavioral mechanisms that rationalize our results: "schmeduling," inattention to the common heating bill, and pro-social behavior. We start with extensions of the theoretical model that accommodate these mechanisms, and discuss their implications for our welfare estimates. We then combine our results with the theoretical framework in Section 2 to estimate the prevalence of free-riding. Finally, we show survey evidence for inattention and pro-social behavior in the common heating bill.

5.1 Extensions of the theoretical framework

We start from the theoretical framework in Section 2. The framework assumes fully rational, informed, and selfish behavior. In practice, households may deviate from these assumptions. They may not equate marginal benefits and costs of heating, they may not be fully informed about their billing regime, or they may care about their neighbors. We describe how these deviations from standard free-riding behavior can be accommodated in our theoretical framework, what they imply for our welfare estimate, and how they impact households' price elasticity.

Schmeduling, inattention to the common heating bill, and pro-social behavior can rationalize our empirical results. Each of these mechanisms implies that households under the common heating bill consume less, and are more price elastic, than predicted by standard free-riding behavior. The implications for our welfare estimates depend on the precise behavioral mechanism at play. Our welfare estimates are correct under standard free-riding behavior, inattention to the common heating bill, and altruism; they are a lower bound where schmeduling, moral cost, and social norms apply.

Schmeduling

Prior work documents that some households do not respond to marginal price changes, but rather to average prices (Ito, 2014). This behavior is referred to as "schmeduling" (Liebman and Zeckhauser, 2004). Ito and Zhang (2023) evaluate a heating price reform in China and show that schmeduling is relevant in the context of heat submetering.

Schmedulers use the average price $\bar{p} = p\frac{\sum^j e_j}{Ne_i}$, rather than the marginal price p, to make consumption decisions under the common heating bill. This changes Equation 3 to

$$e^*_{i,\theta} = (T_i - T_0)H - \frac{H^2}{2\kappa_i}\left(\theta p + (1-\theta)\bar{p}\right) . \tag{18}$$

Schmeduling does not change consumption decisions under submetering, where households face a constant price. Under the common heating bill, low-consumers and high-consumers pay the same heating bill. Hence, low-consumption schmedulers face a high average price and consume "too little," as compared to their private optimum. Conversely, high-consumption schmedulers face a low average price and consume more than their private optimum.

Figure 8 shows the implications of schmeduling in our theoretical framework. Under the common heating bill, all households in a building face the same energy bill, regardless of their relative consumption. Households with low consumption, represented by household l in Figure 8, pay an average price that exceeds p. Those with high consumption, represented by household h, pay an average price below p. Submetering sets the average price equal to the marginal price. As a consequence, schmedulers with low consumption increase their consumption, while those with high consumption decrease theirs. The overall effect on consumption is close to zero.[28]

Schmeduling is relevant for our welfare estimates. If the energy price is sufficiently below social marginal cost, then the consumption increase of low-consumers will decrease welfare (illustrated by the dark red triangle in Figure 8), while the consumption decrease among high-consumers will increase welfare (illustrated by the green triangle). The overall welfare effect is positive. Intuitively, the environmental externality is unaffected because emissions do not change if overall energy consumption remains constant.[29] The positive welfare effect arises because households are not choosing their private optimum under the common heating bill. Submetering corrects this inefficiency. Hence, if households are schmeduling, our welfare estimates from Section 4.3 are a lower bound.

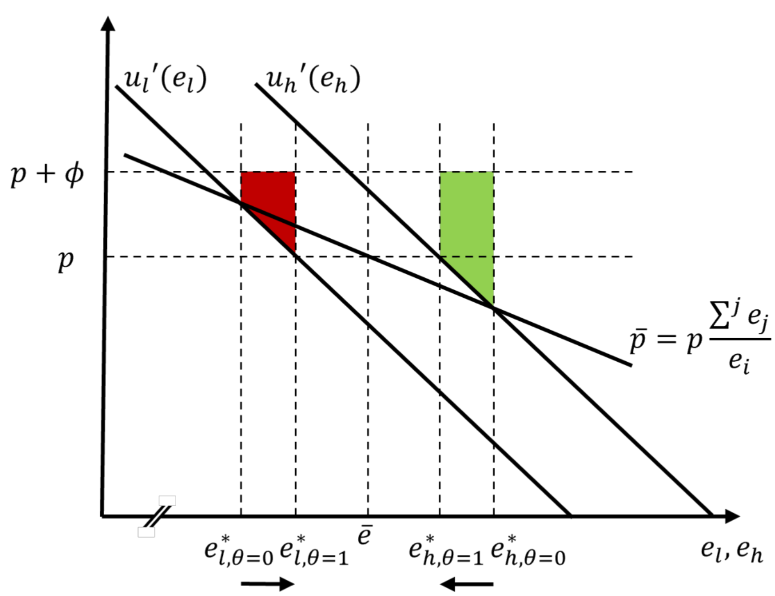


Figure 8: Schmeduling in our theoretical framework

[28]The overall consumption effect is *exactly* zero in Figure 8. The figure imposes two simplifying assumptions. First, household heterogeneity is only driven by T_i (rather than κ_i), implying parallel demand curves. Second, $\bar{p}$ decreases linearly, which can be viewed as a local approximation.

[29]This argument applies if the price is well below social marginal cost. If the price equals the social marginal cost, the welfare effect of submetering for schmedulers is unambiguously positive.

Schmedulers can also explain the large elasticities we find under the common heating bill, as they are more price elastic than free-riders. In the common heating bill (and also under submetering), any $x\%$ change in energy prices changes the average price by $x\%$. Hence, schmedulers under the common heating bill react to price changes as if they were submetered.

Inattention to the common heating bill

Some households may be inattentive to the common heating bill. Inattention is well documented in the domain of energy consumption (e.g., Sallee, 2014; Jessoe and Rapson, 2014), and it is plausible that some households are inattentive to their billing regime. We focus on the case where households are inattentive to the common heating bill. Households may also be inattentive to being submetered, but this is unlikely to be a major concern for newly treated households. The installation of submetering is typically accompanied by information letters, which should make households aware of the change in their billing regime.

Inattention to the common heating bill implies that some households act according to $\hat{\theta} = 1$, independent of the true value of θ. Hence, they behave as if they were submetered. This behavior can explain our results because the price elasticity under the common heating bill is relatively large, and submetering does not change energy consumption. Inattention to the common heating bill does not bias our welfare assessment, as it is reflected in a lower treatment effect of submetering.

Pro-social behavior

Pro-social behavior can have different motivations and implications. We discuss three types of pro-social behavior (altruism, moral cost, and social norms) and acknowledge that this discussion is by no means exhaustive (e.g., Fehr and Gächter, 2000; Frey and Meier, 2004; Bénabou and Tirole, 2006; Falk and Fischbacher, 2006; Kaufmann, Andre, and Kőszegi, 2024).

Altruism changes Equation 1 to include neighbors' utility (Simon, 1993; Fehr and Schmidt, 2006). A fully altruistic household realizes that their heating decision affects their neighbors' utility. Taking this into account, the altruistic household chooses the same heating energy consumption under the common heating bill as under submetering. Submetering does not change the altruistic household's consumption or price elasticity. Atruism does not bias our welfare assessment because, to the extent of their altruism, households value their own consumption as much as their neighbors'.

Some households may pay a moral cost for heating. This is plausible in the common heating bill, where heating has a clear externality on neighbors, but could also reflect a "moral cost of carbon" (Houde, Faure, and Schleich, 2024). If we assume a moral cost of $\frac{N-1}{N}p$ under the common heating bill, submetering neither changes consumption nor the price elasticity. Submetering increases welfare, however, as the household no longer pays a moral cost. Our welfare estimate would not capture this benefit.

Building communities may share a social norm to not exploit the free-riding incentive provided by the common heating bill. In addition, they may find ways to observe each other, to punish free-riders, and achieve low consumption levels (Ostrom et al., 1999). If building communities manage to cooperate, submetering may not change consumption, but it makes enforcement obsolete. Our welfare estimate would be a lower bound, as it does not capture enforcement costs.

5.2 Prevalence of free-riding

We can use our theoretical framework to assess the prevalence of free-riding. We are interested in the share ρ of the population that is free-riding according to our theoretical framework in Section 2. The remaining share $(1-\rho)$ of the population is schmeduling, inattentive to the common heating bill, or pro-social. Following the arguments in Section 5.1, we assume that $(1-\rho)$ households always behave as if they were under the submetering regime, irrespective of their actual billing regime. This analysis abstracts from any other dimensions of heterogeneity. Hence, all households behave identical under submetering. $(1-\rho)$ households have the same behavior under the common heating bill, and ρ households free-ride under the common heating bill.

We can use our estimates to pin down the demand curve in Equation 3 and calculate ρ. Expenses under submetering and the energy price determine point C in Figure 1(a). The estimated price elasticity of demand under submetering determines the slope of the demand curve. $(1-\rho)$ households consume $e^*_{i,\theta=1}$ regardless of billing regime. The remaining ρ households consume the same amount $e^*_{i,\theta=1}$ under submetering, but a larger amount $e^*_{i,\theta=0}$ under the common heating bill. Given the demand curve and the number of neighbors N, we can calculate $e^*_{i,\theta=0}$. The estimated treatment effect of submetering equals $\rho(e^*_{i,\theta=0} - e^*_{i,\theta=1})$.

We use the following values to calculate ρ. Treated households have average expenses of CHF 972 before submetering and an average energy price of CHF 0.0883 per kWh. The median building has a treatment effect of CHF 132. These values imply $9,513$ kWh of energy consumption under submetering. The estimated demand elasticity under submetering of -0.716 implies that a one cent price decrease increases consumption by 771 kWh. The median building has 7 neighbors, implying a price change of 7.6 cents and $5,838$ kWh higher consumption in the common heating bill. As the average treatment effect corresponds to $1,495$ kWh, we find $\rho = \frac{1,495}{5,838} = 26\%$.

We find that approximately one quarter of the population exploits the free-riding incentive under the common heating bill. Submetering reduces their consumption by 38%, a large effect roughly corresponding to a 5°C decrease in indoor temperature (BFE Bundesamt für Energie, 2014). It is not implausible because free-riders start from a high consumption level, and recommended practices (e.g., turning down the thermostat when away, intensive airing) can reduce energy consumption with minimal comfort loss.

Our results from Section 4 suggest that free-riding depends on the number of neighbors in the building. We use the results from Table 1 to calculate treatment effects for each number of neighbors bin (8 or fewer, 9 to 12, and more than 12 neighbors), and use the number of neighbors of the median building in each bin. We use the same elasticity under submetering (-0.716) for all bins because (i) the behavioral mechanisms presented above predict identical behavior under submetering and (ii) Table 2 supports this assumption. The results imply relatively few free-riders (23%) in small and medium-sized buildings (up to 12 neighbors). However, the share of free-riders is almost twice as large (44%) in buildings with more than 12 neighbors.

A large share of households in our sample, particularly in smaller buildings, may not free-ride. As described above, not exploiting the free-riding incentive is consistent with schmeduling, inattention to the common heating bill, and pro-social explanations. Next, we provide survey evidence to discuss the importance of these behavioral mechanisms in more detail.

5.3 Survey evidence and discussion

We provide new survey evidence to inform the behavioral mechanisms behind our empirical findings. Again partering with the real estate management company that provided data for the quasi-experiment, we invited 6,111 tenants in the common heating bill to participate in a survey. Survey participation was incentivized with a lottery of vouchers from a Swiss online retailer. Between September 26, 2024, and October 24, 2024, 835 tenants (14%) provided complete responses. The survey focuses on tenants' support for different submetering policies. Here, we describe a subset of questions that may inform our discussion of behavioral mechanisms. We show results for 495 tenants who passed attention checks and watched information videos as instructed.

We find that many respondents are inattentive to their billing regime. After seeing a short video on the difference between the common heating bill and submetering, survey participants were asked about their own billing regime. 44% of respondents in the common heating bill correctly stated that they were under the common heating bill. The remaining participants reported that they do not know (33%), thought they were in the submetering regime (19%), or thought they were in another billing regime (3%). Figure G.1 in Appendix G shows that inattention to the common heating bill is similar in small/medium ($N \leq 12$) and large ($N > 12$) buildings. These results suggests that inattention to the common heating bill is highly prevalent in our sample, but unlikely to explain the higher share of free-riders in large buildings.

We assess altruism using a modified question from the preference module designed by Falk et al. (2023). The question asks participants "How willing are you to share with your neighbors without expecting anything in return?" on a five-point scale. 10% of respondents report that they are very willing to share with their neighbors, 23% are willing, 38% are neutral, 18% are unwilling, and 11% are completely unwilling. Figure G.2 in Appendix G shows that altruism is similar in small and large buildings. The share of respondents who are very willing to share appears somewhat higher in buildings with 12 or fewer neighbors, but the difference is not statistically significant ($p > 0.1$).

Prior literature suggests that strong relationships among group members facilitate cooperation (Shrestha, 2018). Survey participants were asked "How would you describe the relationship with your neighbors?". We find that most respondents report neutral (34%) or positive (59%) relationships. The strength of relationships varies depending on building size. Figure G.3 in Appendix G shows that buildings with 12 or fewer neighbors are more likely to report being "friends with your neighbors." The share of households choosing this option is similar in buildings with 8 or fewer neighbors (29%) and medium-size buildings (27%), but lower in buildings with more than 12 neighbors (18%; p-values of the differences are $p = 0.07$ and $p = 0.02$, respectively). This pattern matches our treatment effects and free-riding shares, which suggest that buildings with 12 or fewer neighbors do better under the common heating bill.

Taking stock of this section, we find that multiple behavioral mechanisms play a role in the common heating bill. Previous literature provides compelling evidence for schmeduling in energy consumption (Ito, 2014; Shaffer, 2020; Ito and Zhang, 2023). We document that inattention to the common heating bill is prevalent. Together, these two mechanisms can explain why a majority of households does not exploit the free-riding incentive. We also document that buildings with few neighbors are relatively price elastic under the common heating bill and react little to the introduction of submetering. The effect of N on the price incentive cannot fully explain this pattern, as our results suggest a lower prevalence of free-riding in smaller buildings. Better relationships among small building communities may facilitate cooperation and explain this pattern.

The welfare estimates based on the theoretical framework in Section 2 apply to free-riders, those who are inattentive to the common heating bill, and altruists. Notwithstanding, households who schmedule, face a moral cost of heating, or enforce a social norm, may see additional benefits from submetering. To the extent that these mechanisms play a role, our welfare estimates should be viewed as conservative.

6 Conclusions

We study free-riding behavior in the context of heating energy consumption in apartment buildings. In particular, we estimate how heat consumption is affected by a switch from a common billing regime to submetered billing. We use a sample of 4,124 buildings, 185 of which introduced submetering between 2008 and 2022. Our main finding is that submetering leads to a persistent reduction of heating energy consumption of 17%. This is somewhat lower than estimates for the effects of submetering on electricity (Elinder, Escobar, and Petré, 2017), water (Ornaghi and Tonin, 2021), and hot water (Elinder et al., 2024). It is, however, substantially larger than the 10% effect attributed to switching from fixed charges to two-part tariffs in Ito and Zhang (2023). A larger effect is plausible in our setting because it does not allow for opt-out and our estimate accounts for anticipation effects due to announcement and installation.

We use recent advances in ML to estimate heterogeneity in submetering effects. This exercise provides insights regarding the mechanisms that drive (or prevent) energy savings in this context. For example, consistent with our theoretical framework, we find that the effect of submetering is larger in buildings with more neighbors. We also find substantial heterogeneity depending on the floor area of the submetered apartments. This has important implications for welfare, which would have been masked by estimates of unconditional average effects.

Our results suggest that submetering is privately cost-effective only for 10% of buildings in our sample. Widespread adoption of submeters in Switzerland is therefore unlikely in the absence of policy. Once we account for the social cost of carbon, submetering is welfare-improving for 63% of buildings. As carbon prices rise for the residential context, then so do the incentives for the adoption of submeters. However, these investments might still be delayed or even prevented by coordination problems and split incentives between landlords and tenants (e.g., Gillingham, Harding, and Rapson, 2012). In this context, policy intervention has the potential to improve welfare. Our findings imply that targeted policies might be more cost-effective than universal submetering mandates. Finally, our results reveal the extent to which price signals are muted under shared-billing regimes. This has implications for carbon pricing policies in this context. Submetering might be necessary to unlock the full benefits of carbon pricing, especially for large apartment buildings.

References

Agrawal, Arun. 2000. "Small Is Beautiful, but Is Larger Better? Forest-Management Institutions in the Kumaon Himalaya, India." *People and Forests: Communities, Institutions, and Governance* :57.

Alberini, A., Will Gans, and Daniel Velez-lopez. 2011. "Residential Consumption of Gas and Electricity in the U.S.: The Role of Prices and Income." *Energy Economics* 33:870–881.

Allcott, Hunt and Michael Greenstone. 2012. "Is There an Energy Efficiency Gap?" *Journal of Economic Perspectives* 26 (1):3–28.

Allcott, Hunt and Judd B. Kessler. 2019. "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons." *American Economic Journal: Applied Economics* 11 (1):236–76. URL https://www.aeaweb.org/articles?id=10.1257/app.20170328.

Arifovic, Jasmina, Cars Hommes, Anita Kopányi-Peuker, and Isabelle Salle. 2023. "Ten isn't large! Group size and coordination in a large-scale experiment." *American Economic Journal: Microeconomics* 15 (1):580–617.

Asker, John, Allan Collard-Wexler, Charlotte De Canniere, Jan De Loecker, and Christopher R Knittel. 2024. "Two Wrongs Can Sometimes Make a Right: The Environmental Benefits of Market Power in Oil." Tech. rep., National Bureau of Economic Research.

Bénabou, Roland and Jean Tirole. 2006. "Incentives and prosocial behavior." *American economic review* 96 (5):1652–1678.

BFE Bundesamt für Energie. 2014. "Heizen mit Köpfchen. Bundespublikationen."

Borenstein, Severin and James B Bushnell. 2022. "Do two electricity pricing wrongs make a right? Cost recovery, externalities, and efficiency." *American Economic Journal: Economic Policy* 14 (4):80–110.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies* URL https://doi.org/10.1093/restud/rdae007. Forthcoming.

Brewer, Dylan. 2022. "Equilibrium sorting and moral hazard in residential energy contracts." *Journal of Urban Economics* 129:103424. URL https://www.sciencedirect.com/science/article/pii/S0094119022000018.

Canale, L., M. Dell'Isola, G. Ficco, T. Cholewa, S. Siggelsten, and I. Balen. 2019. "A comprehensive review on heat accounting and cost allocation in residential buildings in EU." *Energy and Buildings* 202:109398. URL https://www.sciencedirect.com/science/article/pii/S0378778819316883.

Casari, Marco and Claudio Tagliapietra. 2018. "Group size in social-ecological systems." *Proceedings of the National Academy of Sciences* 115 (11):2728–2733.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng. 2024. "On Binscatter." *American Economic Review* 114 (5):1488–1514. URL https://www.aeaweb.org/articles?id=10.1257/aer.20221576.

Chen, Tianqi and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." *arXiv:1603.02754* .

Christensen, Peter, Paul Francisco, Erica Myers, Hansen Shao, and Mateus Souza. 2024. "Energy efficiency can deliver for climate policy: Evidence from machine learning-based targeting." *Journal of Public Economics* 234:105098. URL https://www.sciencedirect.com/science/article/pii/S0047272724000343.

Christensen, Peter, Paul Francisco, Erica Myers, and Mateus Souza. 2023. "Decomposing the wedge between projected and realized returns in energy efficiency programs." *Review of Economics and Statistics* 105 (4):798–817.

Coase, RH. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3:1–44.

Dahlblom, Mats, Birgitta Nordquist, and Lars Jensen. 2015. "Distribution and variation of indoor temperatures in apartment blocks with individual metering and billing of space heating costs—on building, apartment, and room level." *Energy Efficiency* 8:859–880.

de Chaisemartin, Clément and Xavier D'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9):2964–96. URL https://www.aeaweb.org/articles?id=10.1257/aer.20181169.

Department of Housing and Urban Development (HUD) and the U.S. Census Bureau (AHS). 2023. "American Housing Survey 2023." URL https://www.census.gov/programs-surveys/ahs/data.html.

Dewees, Donald and Trevor Tombe. 2011. "The impact of sub-metering on condominium electricity demand." *Canadian Public Policy* 37 (4):435–457.

Diederich, Johannes, Timo Goeschl, and Israel Waichman. 2016. "Group size and the (in) efficiency of pure public good provision." *European Economic Review* 85:272–287.

Dong, Xibin, Zhiwen Yu, Wenming Cao, Yifan Shi, and Qianli Ma. 2020. "A survey on ensemble learning." *Frontiers of Computer Science* 14:241–258.

EBRD. 2023. "Transition Report 2023-24: Transitions Big and Small." URL https://2023.tr-ebrd.com. Responsibility for the content of the report is taken by the Office of the Chief Economist.

Elinder, Mikael, Sebastian Escobar, and Ingel Petré. 2017. "Consequences of a price incentive on free riding and electric energy consumption." *Proceedings of the National Academy of Sciences* 114 (12):3091–3096. URL https://www.pnas.org/doi/abs/10.1073/pnas.1615290114.

Elinder, Mikael, Xiao Hu, Che-Yuan Liang, and Shane Minckley. 2024. "Mind the tap–how volumetric pricing affects residential hot water consumption." *Journal of the Association of Environmental and Resource Economists* .

EnergieSchweiz. 2023. "VEWA Modell zur verbrauchsabhängigen Energie- und Wasserkostenabrechnung."

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2023. "The preference survey module: A validated instrument for measuring risk, time, and social preferences." *Management Science* 69 (4):1935–1950.

Falk, Armin and Urs Fischbacher. 2006. "A theory of reciprocity." *Games and economic behavior* 54 (2):293–315.

Federal Office for Housing (FOH). 2022. "Living in Switzerland–Everything you need to know about renting an apartment." URL https://www.bwo.admin.ch/dam/bwo/de/dokumente/02_Wie_wir_wohnen/27_Infoblatt-Wohnen/informationsblatt/englisch.pdf.download.pdf/englisch.pdf.

Fehr, Ernst and Simon Gächter. 2000. "Fairness and retaliation: The economics of reciprocity." *Journal of economic perspectives* 14 (3):159–182.

Fehr, Ernst and Klaus M Schmidt. 2006. "The economics of fairness, reciprocity and altruism–experimental evidence and new theories." *Handbook of the economics of giving, altruism and reciprocity* 1:615–691.

Filippini, Massimo and Nilkanth Kumar. 2021. "Gas demand in the Swiss household sector." *Applied Economics Letters* 28 (5):359–364.

Frey, Bruno S and Stephan Meier. 2004. "Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment." *American economic review* 94 (5):1717–1722.

Gerarden, Todd D., Richard G. Newell, and Robert N. Stavins. 2017. "Assessing the Energy-Efficiency Gap." *Journal of Economic Literature* 55 (4):1486–1525. URL http://www.aeaweb.org/articles?id=10.1257/jel.20161360.

Gerarden, Todd D and Muxi Yang. 2023. "Using targeting to optimize program design: Evidence from an energy conservation experiment." *Journal of the Association of Environmental and Resource Economists* 10 (3):687–716.

Gillingham, Kenneth, Matthew Harding, and David Rapson. 2012. "Split incentives in residential energy consumption." *The Energy Journal* 33 (2):37–62.

Gillingham, Kenneth and Karen Palmer. 2014. "Bridging the Energy Efficiency Gap: Policy Insights from Economic Theory and Empirical Evidence." *Review of Environmental Economics and Policy* 8 (1):18–38. URL https://doi.org/10.1093/reep/ret021.

Glance, Natalie S and Bernardo A Huberman. 1994. "The dynamics of social dilemmas." *Scientific American* 270 (3):76–81.

Gneezy, Uri, Ernan Haruvy, and Hadas Yafe. 2004. "The inefficiency of splitting the bill." *The Economic Journal* 114 (495):265–280.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics* 225 (2):254–277. URL https://www.sciencedirect.com/science/article/pii/S0304407621001445.

Guo, Jin, Ying Huang, and Chu Wei. 2015. "North–South debate on district heating: Evidence from a household survey." *Energy Policy* 86:295–302. URL https://www.sciencedirect.com/science/article/pii/S0301421515300288.

Houde, Sébastien, Corinne Faure, and Joachim Schleich. 2024. "The moral cost of carbon." Tech. rep., Working Paper.

Hume, David. 1739. *A Treatise of Human Nature*.

International Energy Agency (IEA). 2022. "Technical Report: Heating." URL https://www.iea.org/reports/heating.

Isaac, R Mark, James M Walker, and Arlington W Williams. 1994. "Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups." *Journal of Public Economics* 54 (1):1–36.

Ito, Koichiro. 2014. "Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing." *American Economic Review* 104 (2):537–563.

Ito, Koichiro and Shuang Zhang. 2023. "Do consumers distinguish fixed cost from variable cost? "Schmeduling" in two-part tariffs in energy." *American Economic Journal: Economic Policy* forthcoming.

Jack, B. Kelsey, Seema Jayachandran, Flavio Malagutti, and Sarojini Rao. 2024. "Environmental externalities and free-riding in the household." *Journal of Development Economics* 170:103294. URL https://www.sciencedirect.com/science/article/pii/S0304387824000439.

Jessoe, Katrina and David Rapson. 2014. "Knowledge is (less) power: Experimental evidence from residential energy use." *American Economic Review* 104 (4):1417–1438.

Johannesson, G., L. Agnoletto, G. Anderlind, B.R. Anderson, R.D. Godfrey, K. Kimura, O. Lyng, C. Roulet, H.C. Sorensen, and H. Werner. 1985. "The Calculation of Space-Heating Requirements for Residential Buildings." *Buildings III Conference Proceedings* 89.

Kaufmann, Marc, Peter Andre, and Botond Kőszegi. 2024. "Understanding Markets with Socially Responsible Consumers." *The Quarterly Journal of Economics* 139 (3):1989–2035. URL https://doi.org/10.1093/qje/qjae009.

Knittel, Christopher R and Samuel Stolper. 2019. “Using machine learning to target treatment: The case of household energy use.” *NBER Working Paper* 26531.

———. 2021. “Machine learning about treatment effect heterogeneity: The case of household energy use.” *AEA Papers and Proceedings* 111:440–444.

Liebman, Jeffrey B and Richard J Zeckhauser. 2004. “Schmeduling.” URL www.hks.harvard.edu/jeffreyliebman/schmeduling.pdf.

Lim, Seul-Ye, Hyo-Jin Kim, and S. Yoo. 2016. “The demand function for residential heat through district heating system and its consumption benefits in Korea.” *Energy Policy* 97:155–160.

Ling, Jihong, Qiang Li, and Jincheng Xing. 2015. “The influence of apartment location on household space heating consumption in multi-apartment buildings.” *Energy and Buildings* 103:185–197.

Marwell, Gerald and Ruth E Ames. 1979. “Experiments on the provision of public goods. I. Resources, interest, group size, and the free-rider problem.” *American Journal of Sociology* 84 (6):1335–1360.

McRae, Shaun. 2024. “Efficiency and equity effects of electricity metering: Evidence from Colombia.” *Available at SSRN 4801287* .

Munley, Vincent G, Larry W Taylor, and John P Formby. 1990. “Electricity demand in multi-family, renter-occupied residences.” *Southern Economic Journal* :178–194.

Myers, Erica. 2019. “Are Home Buyers Inattentive? Evidence from Capitalization of Energy Costs.” *American Economic Journal: Economic Policy* 11(2):165–188. URL https://www.aeaweb.org/articles?id=10.1257/pol.20170481.

———. 2020. “Asymmetric information in residential rental markets: Implications for the energy efficiency gap.” *Journal of Public Economics* 190:104251.

Noack, Frederik and Christopher Costello. 2024. “Credit markets, property rights, and the commons.” *Journal of Political Economy* 132 (7).

Nosenzo, Daniele, Simone Quercia, and Martin Sefton. 2015. “Cooperation in small groups: the effect of group size.” *Experimental Economics* 18:4–14.

Olson, Mancur. 1965. “The logic of collective action.” *Harvard University Press* .

OpenAI. 2024. “OpenAI Platform.” URL https://platform.openai.com/docs/overview.

Ornaghi, Carmine and Mirco Tonin. 2021. “The effects of the universal metering programme on water consumption, welfare and equity.” *Oxford Economic Papers* 73 (1):399–422.

Ostrom, Elinor. 1990. *Governing the commons: The evolution of institutions for collective action*. Cambridge university press.

Ostrom, Elinor, Joanna Burger, Christopher B Field, Richard B Norgaard, and David Policansky. 1999. "Revisiting the commons: local lessons, global challenges." *science* 284 (5412):278–282.

Poteete, Amy R and Elinor Ostrom. 2004. "Heterogeneity, group size and collective action: The role of institutions in forest management." *Development and Change* 35 (3):435–461.

Rennert, Kevin, Frank Errickson, Brian C Prest, Lisa Rennels, Richard G Newell, William Pizer, Cora Kingdon, Jordan Wingenroth, Roger Cooke, Bryan Parthum et al. 2022. "Comprehensive evidence implies a higher social cost of CO2." *Nature* 610 (7933):687–692.

Roope, Laurence SJ, Richard D Smith, Koen B Pouwels, James Buchanan, Lucy Abel, Peter Eibich, Christopher C Butler, Pui San Tan, A Sarah Walker, Julie V Robotham et al. 2019. "The challenge of antimicrobial resistance: what economics can contribute." *Science* 364 (6435):eaau4679.

Sallee, James M. 2014. "Rational inattention and energy efficiency." *The Journal of Law and Economics* 57 (3):781–820.

Schulte, Isabella and Peter Heindl. 2017. "Price and income elasticities of residential energy demand in Germany." *Energy Policy* 102:512–528. URL https://www.sciencedirect.com/science/article/pii/S0301421516307194.

Shaffer, Blake. 2020. "Misunderstanding nonlinear prices: Evidence from a natural experiment in electricity demand." *American Economic Journal: Economic Policy* 12 (3).

Shrestha, Manoj K. 2018. "Network structure, strength of relationships, and communities' success in project implementation." *Public Administration Review* 78 (2):284–294.

Simon, Herbert A. 1993. "Altruism and economics." *The American Economic Review* 83 (2):156–161.

Souza, Mateus. 2019. "Predictive Counterfactuals for Treatment Effect Heterogeneity in Event Studies with Staggered Adoption." *SSRN Working Paper* 3484635. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3484635.

Stern, Nicholas. 2008. "The economics of climate change." *American Economic Review* 98 (2):1–37.

Sun, Liyang and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* 225 (2):175–199. URL https://ideas.repec.org/a/eee/econom/v225y2021i2p175-199.html.

Trotta, Gianluca, Anders Rhiger Hansen, and Stephan Sommer. 2022. "The price elasticity of residential district heating demand: New evidence from a dynamic panel approach." *Energy Economics* 112:106163.

Van der Laan, Mark J, Eric C Polley, and Alan E Hubbard. 2007. "Super learner." *Statistical applications in genetics and molecular biology* 6 (1).

A Additional Tables

Table A.1: Averages by Treatment Cohort—variables at the apartment-by-year or apartment level

Treatment Cohort	(1) Heating Bill (CHF per year)	(2) Rent (CHF per month)	(3) Turnover Rate	(4) Number of Vacancy Days	(5) Apartment Area (sq. meters)	(6) Rooms per Apartment
2008	954.80	1,320.48	0.15	3.99	88.07	3.81
	(269.64)	(546.17)	(0.36)	(15.62)	(24.52)	(1.22)
2009	927.19	1,162.26	0.12	3.10	89.42	3.87
	(250.54)	(529.48)	(0.33)	(13.34)	(25.93)	(0.94)
2010	875.15	1,196.49	0.16	6.79	84.47	3.63
	(288.73)	(411.23)	(0.37)	(31.03)	(27.67)	(1.02)
2011	1,156.77	1,238.75	0.17	2.81	88.75	3.75
	(384.65)	(457.32)	(0.38)	(12.95)	(20.95)	(0.80)
2012	972.00	1,040.27	0.15	7.04	73.04	3.30
	(316.65)	(363.62)	(0.36)	(34.72)	(20.89)	(1.10)
2013	1,069.76	1,253.86	0.13	5.16	84.12	3.43
	(342.09)	(424.10)	(0.34)	(32.40)	(34.72)	(1.29)
2014	1,167.18	1,138.21	0.11	1.97	90.49	3.74
	(362.70)	(274.69)	(0.31)	(18.00)	(17.58)	(0.76)
2015	1,006.20	1,104.71	0.16	11.29	69.35	2.75
	(363.04)	(390.22)	(0.36)	(50.12)	(40.93)	(1.37)
2016	954.80	1,320.48	0.15	3.99	104.33	4.50
	(269.64)	(546.17)	(0.36)	(15.62)	(4.70)	(.)
2017	1,203.56	1,141.02	0.18	21.80	74.54	3.34
	(334.63)	(334.05)	(0.38)	(72.36)	(17.80)	(0.80)
2018	901.93	1,229.16	0.20	7.03	81.19	3.68
	(257.40)	(504.39)	(0.40)	(26.15)	(23.59)	(0.93)
2019	1,245.25	1,283.43	0.14	4.79	89.41	3.74
	(371.00)	(292.22)	(0.35)	(22.20)	(21.17)	(1.00)
2020	955.09	1,341.54	0.14	4.81	76.66	3.01
	(330.23)	(391.68)	(0.35)	(24.13)	(26.30)	(1.13)
2021	992.84	1,279.99	0.14	10.58	69.93	3.04
	(283.14)	(464.62)	(0.35)	(43.38)	(21.91)	(1.34)
2022	1,201.23	1,019.47	0.10	5.50	77.53	3.38
	(329.56)	(434.30)	(0.30)	(32.43)	(20.76)	(0.90)
Never Treated	1,034.33	1,226.33	0.13	8.40	74.96	3.32
	(413.84)	(423.54)	(0.33)	(43.05)	(25.41)	(1.13)
Observations	362,991	351,304	362,991	362,991	40,956	43,030

Notes: Averages and standard deviations (in parentheses) for variables at the apartment-by-year level are presented in columns 1 through 4. These use all observations from the never-treated apartments, but only pre-treatment observations for the treated apartments. Columns 5 and 6 are for variables fixed at the apartment level (i.e., that do not change over time).

Table A.2: Averages by Treatment Cohort—variables at the building-by-year or building level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					Shares of Heating Fuel Type			
Treatment Cohort	Heating Degree Days	Energy-Related Renovation Rate	Number of Neighbors	Years since Construction	District Heating	Gas	Oil Heating	Other/Unclear
2008	3,394.60	0.00	7.53	38.95	0.05	0.58	0.26	0.11
	(235.94)	(.)	(3.50)	(15.69)	(0.23)	(0.51)	(0.45)	(0.32)
2009	4,241.21	0.00	7.55	47.82	0.00	0.27	0.64	0.09
	(572.21)	(.)	(2.94)	(16.69)	(.)	(0.47)	(0.50)	(0.30)
2010	3,112.40	0.12	7.27	38.86	0.18	0.32	0.23	0.27
	(504.72)	(0.34)	(4.94)	(17.54)	(0.39)	(0.48)	(0.43)	(0.46)
2011	3,180.90	0.33	15.35	38.59	0.24	0.35	0.41	0.00
	(402.98)	(0.49)	(20.72)	(16.44)	(0.44)	(0.49)	(0.51)	(.)
2012	3,232.54	0.01	8.03	48.11	0.08	0.24	0.61	0.08
	(385.34)	(0.09)	(3.75)	(12.56)	(0.27)	(0.43)	(0.50)	(0.27)
2013	2,900.48	0.02	10.73	51.95	0.14	0.45	0.32	0.09
	(375.45)	(0.12)	(6.94)	(12.21)	(0.35)	(0.51)	(0.48)	(0.29)
2014	3,180.03	0.00	8.34	44.41	0.03	0.47	0.44	0.06
	(419.04)	(.)	(3.92)	(8.99)	(0.18)	(0.51)	(0.50)	(0.25)
2015	3,043.68	0.00	9.26	53.74	0.26	0.47	0.05	0.21
	(349.93)	(.)	(11.58)	(9.45)	(0.45)	(0.51)	(0.23)	(0.42)
2016	2,400.50	1.00	12.00	41.00	0.00	0.00	1.00	0.00
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
2017	3,774.80	0.03	13.50	55.86	0.00	0.86	0.00	0.14
	(851.32)	(0.18)	(5.59)	(6.32)	(.)	(0.36)	(.)	(0.36)
2018	2,829.05	0.00	8.67	60.00	0.00	0.17	0.83	0.00
	(10.96)	(.)	(6.53)	(.)	(.)	(0.41)	(0.41)	(.)
2019	3,263.60	0.10	6.18	45.32	0.45	0.14	0.32	0.09
	(277.30)	(0.30)	(2.59)	(5.34)	(0.51)	(0.35)	(0.48)	(0.29)
2020	3,116.67	0.03	9.90	41.70	0.90	0.10	0.00	0.00
	(200.37)	(0.17)	(6.33)	(15.52)	(0.32)	(0.32)	(.)	(.)
2021	3,044.63	0.06	9.67	51.50	0.00	0.42	0.33	0.25
	(318.87)	(0.25)	(5.84)	(9.87)	(.)	(0.51)	(0.49)	(0.45)
2022	3,339.18	0.03	12.85	49.05	0.05	0.50	0.05	0.40
	(510.33)	(0.17)	(11.31)	(13.16)	(0.22)	(0.51)	(0.22)	(0.50)
Never Treated	3,198.53	0.03	11.01	54.94	0.12	0.37	0.42	0.09
	(560.85)	(0.17)	(8.75)	(10.05)	(0.33)	(0.48)	(0.49)	(0.29)
Observations	30,934	33,139	3,978	3,978	3,978	3,978	3,978	3,978

Notes: Averages and standard deviations (in parentheses) for variables at the building-by-year level are presented in columns 1 and 2. These use all observations from the never-treated apartments, but only pre-treatment observations for the treated apartments. Columns 3 through 8 are for variables fixed at the building level (i.e., that do not change over time).

B Extensions of the theoretical framework

B.1 Fixed and variable heating costs

The theoretical framework in the main text applies to full submetering ($\bar{\theta} = 1$) of variable energy costs. In practice, the heating bill includes fixed costs that cannot be reduced by household behavior. These include service fees for the heating system and potentially a chimney, heat losses in pipes serving individual apartments, and the heating of common areas. To account for fixed heating costs, submetering usually only applies to a certain share of the heating bill, $\bar{\theta} < 1$. We extend the theoretical framework to account for fixed heating costs.

Note that fixed heating costs and $\bar{\theta}$ change the framework through the budget constraint. We show how the budget constraint changes with fixed heating costs and $\bar{\theta} < 1$. Then, we discuss the conditions under which this budget constraint, and hence the theoretical framework, is equivalent to the simplified version with $\bar{\theta} = 1$ and no fixed heating costs.

The budget constraint with fixed heating costs F is

$$y_i \geq \bar{\theta} e_i p + (1-\bar{\theta}) \frac{\sum^j e_j p}{N} + \bar{\theta} F \frac{e_i}{\sum^j e_j} + (1-\bar{\theta}) \frac{F}{N} + c_i \ . \tag{B.1}$$

The marginal cost of energy consumption in this setting is

$$\frac{\partial y_i}{\partial e_i} = \bar{\theta} p + (1-\bar{\theta}) \frac{p}{N} + \bar{\theta} F \frac{\sum^j e_j - e_i}{(\sum^j e_j)^2} \ . \tag{B.2}$$

Note that the marginal cost of energy in the simple case with $\bar{\theta} = 1$ and $F = 0$ is simply p. The marginal cost of energy in both settings is equivalent if

$$\bar{\theta} = \frac{\sum^j e_j p}{\sum^j e_j p + F \frac{N}{N-1} \frac{\sum^j e_j - e_i}{\sum^j e_j}} \ . \tag{B.3}$$

We see that this condition is satisfied under two conditions. First, household i consumes approximately $\frac{1}{N}$ of the variable heating costs, i.e. $\frac{1}{N} \approx \frac{e_i}{\sum^j e_j}$. This condition is likely to hold, in particular if N is large. Concretely, it holds when e_i is finite and $N \to \infty$. Second, $\bar{\theta}$ equals the share of variable heating costs in the total heating bill. Since the consideration of fixed costs is precisely the rationale behind $\bar{\theta} < 1$, this may be approximately true.

B.2 Partial submetering

Submetering increases θ from 0 to $\bar{\theta}$. In response to submetering, household i chooses a lower $e^*_{i,\theta}$:

$$\Delta e^*_i = e^*_{i,\theta=\bar{\theta}} - e^*_{i,\theta=0} = -\frac{H^2}{2\kappa_i} \frac{N-1}{N} p \bar{\theta} \ . \tag{B.4}$$

Utility may increase or decrease, depending on the effect of submetering on the household's energy expenses.

$$
\begin{aligned}
\Delta U_i &= U_i(e^*_{i,\theta=\bar{\theta}}) - U_i(e^*_{i,\theta=0}) \\
&= \left(\frac{p^2}{N^2} - p^2\left(\theta + \frac{1-\theta}{N}\right)^2\right)\frac{H^2}{4\kappa_i} + \left(\frac{\sum^j T_j}{N} - T_i\right)Hp\bar{\theta} \\
&+ \bar{\theta}p^2\left(\theta + \frac{1-\theta}{N}\right)\frac{H^2}{2\kappa_i} + (1-\bar{\theta})\frac{p^2}{N}\left(\theta + \frac{1-\theta}{N}\right)\sum^j \frac{1}{\kappa_j} - \frac{p^2}{N^2}\frac{H^2}{2}\sum^j \frac{1}{\kappa_j}
\end{aligned}
\tag{B.5}
$$

We also find that submetering increases the demand elasticity, as shown in Equation 7 below.

$$
\begin{aligned}
\Delta \eta_i &= |\eta_{i,\theta=\bar{\theta}}| - |\eta_{i,\theta=0}| \\
&= \frac{H^2}{2\kappa_i}\left(\left(\bar{\theta} + \frac{1-\bar{\theta}}{N}\right)\frac{p}{e^*_{i,\theta=\bar{\theta}}} - \frac{1}{N}\frac{p}{e^*_{i,\theta=0}}\right) > 0
\end{aligned}
\tag{B.6}
$$

The welfare effect can be expressed in terms of empirical estimates.

$$
\Delta W = -\tau\frac{N-1}{N}\left(1 - \frac{\bar{\theta}}{2}\right) - \tau\frac{\phi}{p} - sN \ . \tag{B.7}
$$

C Renovation Classification

We use data on renovations provided by the real estate management company. For each property, we have a list of renovations that contains the year of the renovation and a text description. The descriptions are written in German, French, or Italian.

We use the OpenAI API (model version `gpt-4o-2024-05-13`) to classify renovations into categories with the following prompt:

```
You are tasked with classifying descriptions of real estate renovations (given
     in German, French, or Italian). For each description, provide the
    following details:

1. A short description of your understanding of the renovation in English.

2. Investment cost.

3. Does this renovation improve the energy efficiency of the windows? Note:
    This concerns the windows' ability to retain heat. Do not account for
    renovations that merely block sunlight.

4. Does this renovation improve the energy efficiency of the exterior walls?

5. Does this renovation improve the energy efficiency of the roof?

6. Does this renovation concern the heating system?

7. If the description mentions addresses or IDs of apartments or buildings,
    list them, separated by commas. Otherwise answer "none".

Note: If a renovation description does not specify the type of renovation,
    assume it is large and comprehensive, including energy efficiency
    improvements to windows, exterior walls, and roof. This does not apply if
    you cannot make sense of the description.

The output should only contain your answers, separated by line breaks (no
    enumeration, no repeated questions). The answer to 2 should be either "
    small", "medium", or "large". Answers to 3-6 should be either "yes" or "no
    ".

Now, classify the following renovation description:
```

We use the output from this prompt to distinguish between comprehensive renovations, non-comprehensive renovations that change the energy efficiency of the building, and other renovations. We define a comprehensive renovation as one that includes energy efficiency improvements to windows, exterior walls, and the roof. We classify a renovation as energy efficiency improving if it is not comprehensive and includes at least one of improvements to windows, exterior walls, the roof, or the heating system. All remaining renovations are classified as other.

D Standard event study via two-way fixed effects

To estimate $ATT(r)$, we alternatively consider a two-way fixed effects (TWFE) regression, as follows:

$$Y_{i,t} = \sum_{r \neq -2} \beta_r S_i \times \mathbb{1}[r = t - (q_i)] + \gamma_i + \gamma_t + \varepsilon_{i,t} \ , \tag{D.1}$$

where S_i is the submetering indicator; $\mathbb{1}[r = t - q_i]$ are indicators for years relative to the treatment dates q_i; and the other parameters are defined as in the main text. As with the heterogeneity-robust specifications, we implement a variant of Equation D.1 including time-varying controls (i.e., the "saturated" specification).

Results are presented in Figure D.1. These are mostly in line with the heterogeneity-robust specifications. If anything, estimates from TWFE seem slightly stronger.

(a) Effects of Submetering on Heating Expenses (%)

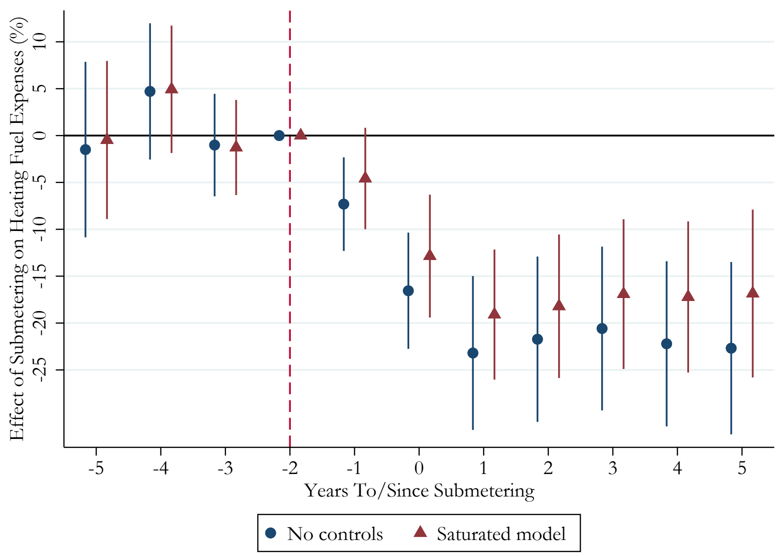


Figure D.1: Effects of Submetering According to TWFE

Notes: This figure presents estimates of $ATT(r)$ for the effects of submetering according to two-way fixed effects regressions. The outcome variable is heating expenses in logs. Estimates are transformed to represent percent effects ($100 \times (exp(\beta_1) - 1)$). All specifications include year and apartment fixed effects. The saturated specification also controls for heating degree days, renovations, and Canton-by-year fixed effects. Standard errors are clustered at the property level. The "whiskers" around the point estimates represent 95% confidence intervals.

E Additional Details on Machine Learning and Heterogeneity Analyses

E.1 Cross-validation strategies

We consider two cross-validation strategies to assess the performance of the ML algorithms. The first strategy is standard k-fold cross-validation, where the validation folds are defined at random. Given that we have panel data (apartment-by-year), standard k-fold cross-validation implies that (pre-treatment) observations for a given apartment might be present in multiple folds. In this case, the CV residuals can be interpreted as "out of sample" in the sense that they reflect errors when the specific apartment-by-year observations were not included in the training sample (although observations for the same apartments, but for *different years*, might have been included).

We also implement an alternative CV strategy which can be considered stricter and potentially better suited for ex-ante analyses (e.g., for ex-ante targeting of treatment). We refer to this strategy as "stratified CV" because we impose that all observations from a given building must be contained within the same CV fold. The implication is that the CV residuals would represent fully out-of-sample errors, where the training sample excludes all observations from the buildings contained in the validation samples. A drawback of this strategy is that it does not allow the inclusion of building or apartment fixed effects as predictors.

Tables E.1 and E.2 summarize the average predictive performance of several XGBoost hyperparameter configurations. Both in-sample and cross-validated root-mean-square errors (RMSE) are presented. We have trained XGBoost algorithms with 3 thousand trees, but which vary in terms maximum tree depth (5, 10, or 20), minimum observations per terminal node (10, 20, or 30), and shrinkage rate (0.1 or 0.2). A total of 18 configurations is obtained by varying these hyperparameters.

We focus on interpreting the CV RMSE, which are a proxy for out-of-sample performance. By comparing Tables E.1 and E.2, it can be noted that prediction errors are substantially larger under the stratified CV strategy. This is expected because, in that case, the algorithms are trained with no information about the apartments for which they attempt to predict energy expenses. The tables also reveal the "best-performing" hyperparameter configurations (that have lower CV RMSE). Particularly from Table E.2, it is clear that the best configurations have a maximum tree depth of 10, and a shrinkage rate of 0.1. However, the tables diverge in terms of the optimal number of observations per terminal node. For this reason, rather than picking just one configuration, we base our predictions on an "ensemble" of three configurations highlighted in gray. The ensemble weights are 0.685, 0.283, and 0.032, respectively, for Model ID 7, 9, and 11. These optimal weights were estimated via Non-Negative Least Squares (Chen and Guestrin, 2016).

Table E.1: Predictive Performance of XGBoost Configurations—Standard 4-fold CV

	N Trees	Max Tree Depth	Min Obs per Node	Shrinkage	In-Sample RMSE	Cross-Validated RMSE
1	3000	5	10	0.10	83.514	90.895
2	3000	5	10	0.20	63.845	74.253
3	3000	5	20	0.10	84.984	92.437
4	3000	5	20	0.20	65.890	76.293
5	3000	5	30	0.10	86.233	93.932
6	3000	5	30	0.20	67.737	78.110
7	3000	10	10	0.10	33.327	57.161
8	3000	10	10	0.20	27.642	58.398
9	3000	10	20	0.10	37.794	58.591
10	3000	10	20	0.20	31.008	57.732
11	3000	10	30	0.10	40.629	60.119
12	3000	10	30	0.20	33.510	57.905
13	3000	20	10	0.10	21.958	65.723
14	3000	20	10	0.20	21.672	67.512
15	3000	20	20	0.10	23.646	63.307
16	3000	20	20	0.20	21.990	66.293
17	3000	20	30	0.10	25.703	61.514
18	3000	20	30	0.20	22.700	64.696

Notes: Cross-validated RMSE was obtained through standard 4-fold cross-validation. The algorithm with the lowest CV RMSE is highlighted in gray.

Table E.2: Predictive Performance of XGBoost Configurations—“Stratified” Cross-Validation

	N Trees	Max Tree Depth	Min Obs per Node	Shrinkage	Stratified In-Sample RMSE	Stratified CV RMSE
1	3000	5	10	0.10	90.543	192.242
2	3000	5	10	0.20	72.581	192.184
3	3000	5	20	0.10	91.246	191.602
4	3000	5	20	0.20	73.620	193.450
5	3000	5	30	0.10	92.627	190.921
6	3000	5	30	0.20	74.794	193.002
7	3000	10	10	0.10	46.252	186.275
8	3000	10	10	0.20	43.506	191.809
9	3000	10	20	0.10	48.664	186.211
10	3000	10	20	0.20	44.997	190.583
11	3000	10	30	0.10	50.300	187.080
12	3000	10	30	0.20	46.087	190.819
13	3000	20	10	0.10	41.727	198.970
14	3000	20	10	0.20	41.706	200.645
15	3000	20	20	0.10	42.010	196.572
16	3000	20	20	0.20	41.730	200.694
17	3000	20	30	0.10	42.567	195.790
18	3000	20	30	0.20	41.831	199.108

Notes: Cross-validated RMSE was obtained through standard “stratified” cross-validation, where all observations from a given building must be contained within the same CV fold. Models do not include building or apartment fixed effects. The algorithm with the lowest CV RMSE is highlighted in gray.

E.2 Detailed predictive performance by covariates

Next, we turn to analyzing prediction errors by our key covariates of interest: number of neighbors in the buildings, area of the apartments, and building's vintage. These are the covariates that we focus on for heterogeneity (i.e., estimation of CATT). For our CATT estimates to be valid, then our ML predictions of counterfactuals need to be unbiased for those covariates. As stated in the main text, we test this unbiasedness assumption by regressing the cross-validated residuals on binned indicators for these covariates (i.e., we run a variant of Equation 14, with cross-validated residuals as the outcome).

Results are presented in Figure E.1. We plot average residuals according to both cross-validation strategies (standard and stratified). We note that, regardless of the CV strategy, average residuals are small and not statistically significant across the relevant covariate bins, suggesting that the selected algorithm is unlikely to be biased along those dimensions. Similarly, in Figure E.2 we show that prediction errors are not strongly correlated with renovations, tenant changes (turnover), or apartment's location in the building. That figure also reveals small prediction errors for other tenant/building characteristics that are unlikely to drive treatment effect heterogeneity.

(a) Errors from standard CV (b) Errors from stratified CV

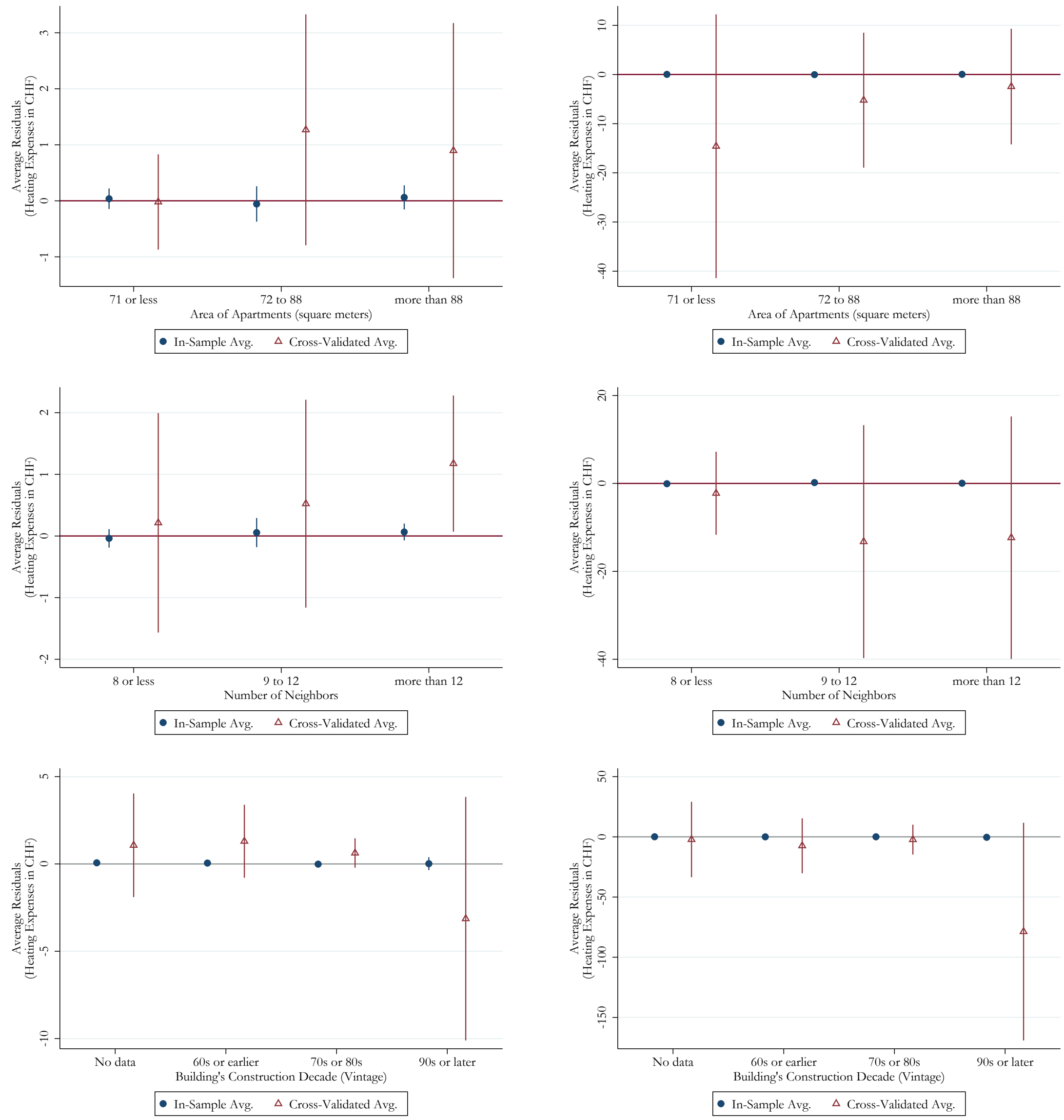


Figure E.1: ML prediction errors by covariates of interest

Notes: This Figure plots coefficient estimates and 95% confidence intervals from a regression of cross-validated ML residuals on bins for selected covariates, namely apartment area, number of neighbors, and building's vintage. Graphs on the left use residuals from a standard 4-fold cross-validation strategy. Graphs on the right use residuals from stratified cross-validation.

(a) Errors from standard CV (b) Errors from stratified CV

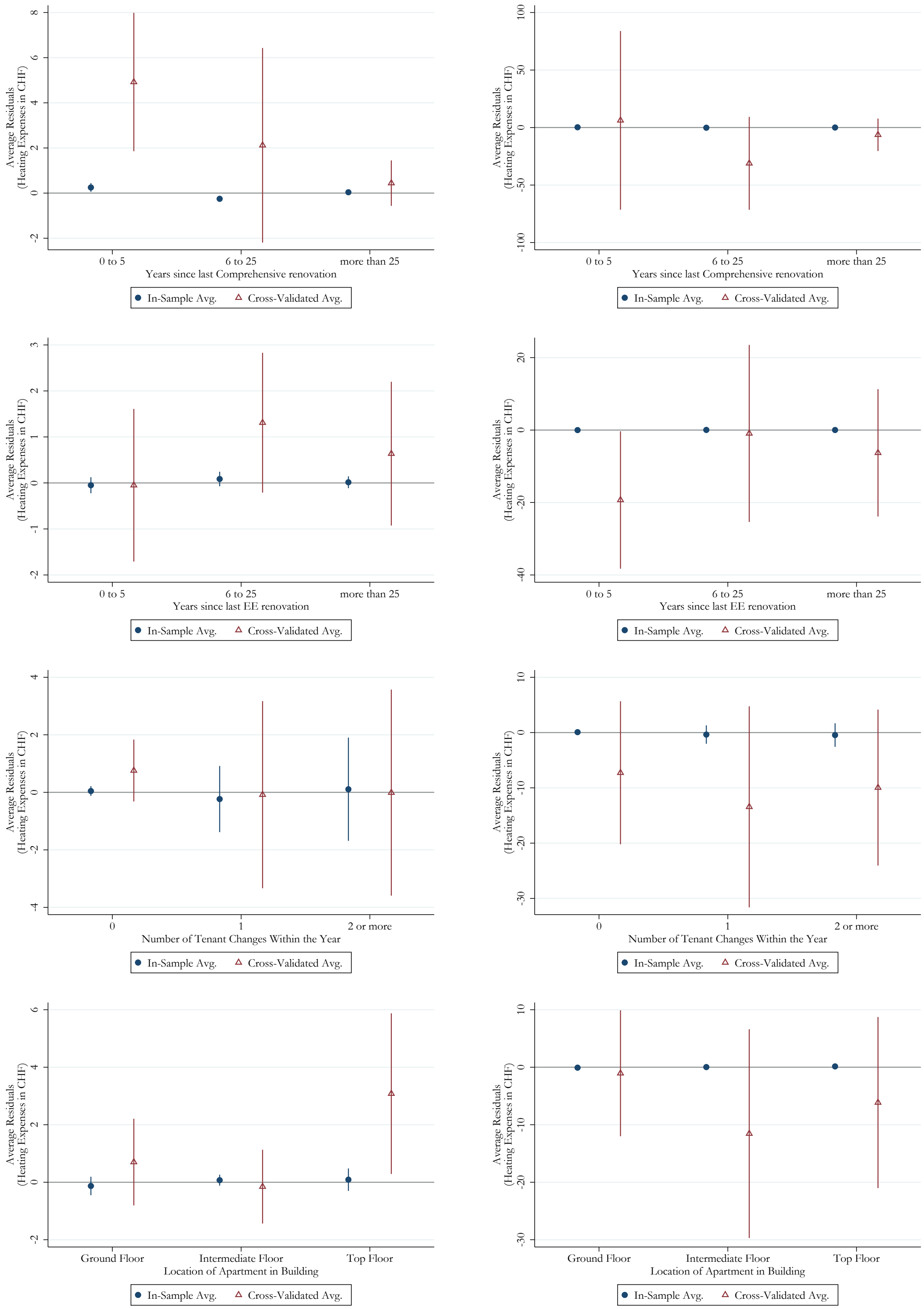
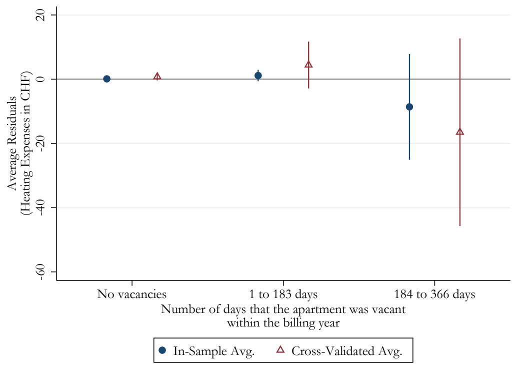
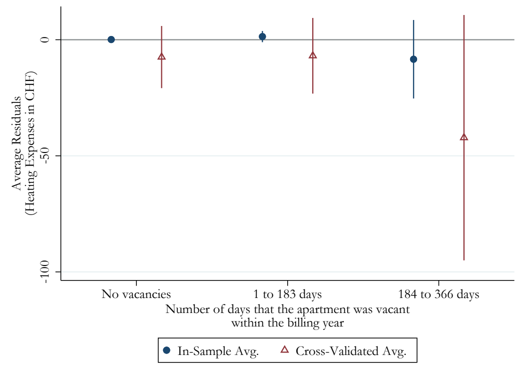


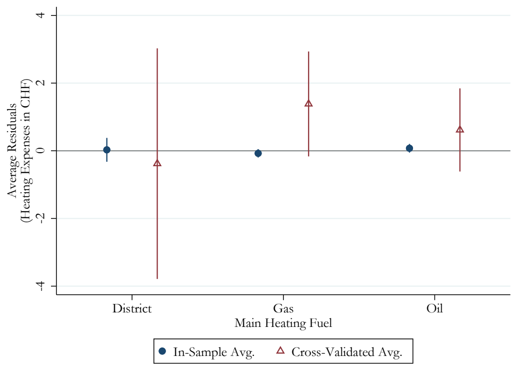
Figure E.2: ML prediction errors by other covariates

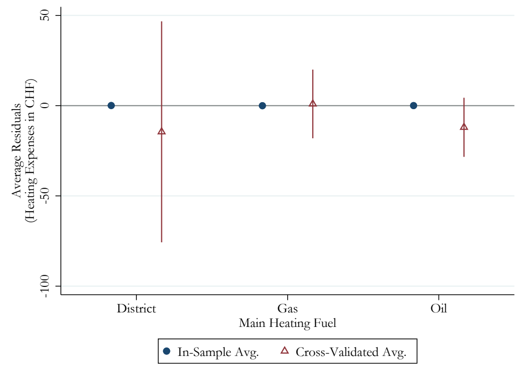


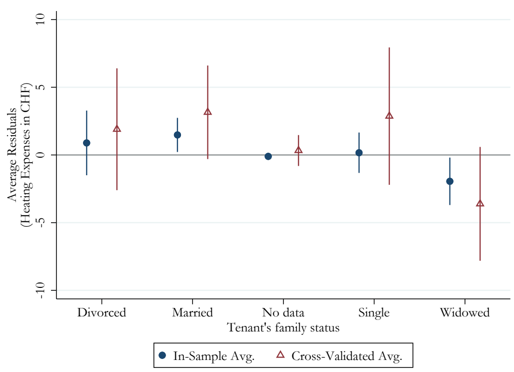
(a) Errors from standard CV

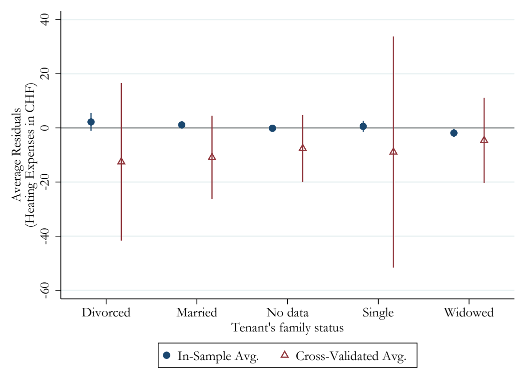


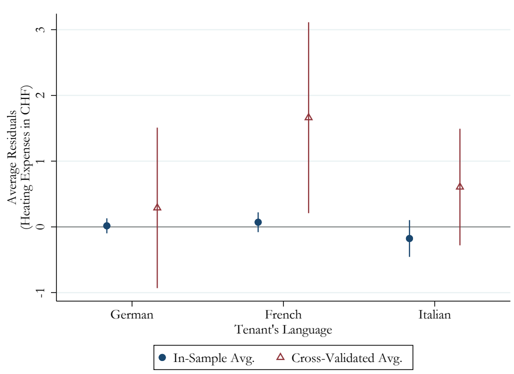
(b) Errors from stratified CV











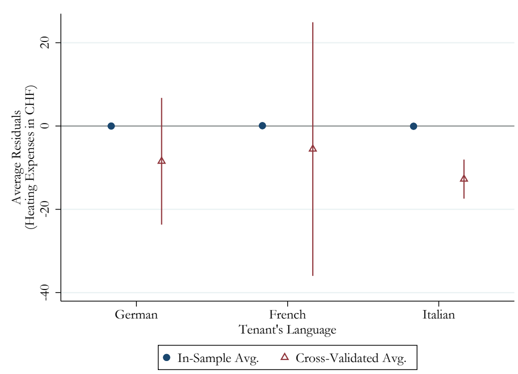


Figure E.2 (continued): ML prediction errors by other covariates

E.3 Bin selection for continuous variables of interest

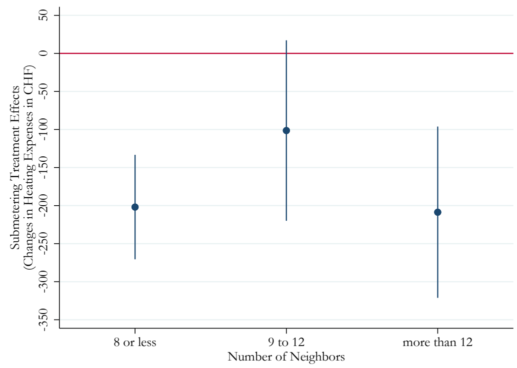
We observe three covariates of interest (for the heterogeneity analyses) that are recorded as continuous variables, namely: apartment area, number of neighbors, and building vintage. By directly including those in the heterogeneity specifications, it would be possible to estimate a form of "marginal effect" of how the benefits of submetering change for each marginal change in the covariates. However, this could mask any potential nonlinear effects. Also, we argue that a categorical interpretation of these covariates is more tractable for real-world settings. For these reasons, we discretize these variables (create bins) prior to including them in the heterogeneity regressions.

We implement a data-driven binning procedure developed by Cattaneo et al. (2024), via the the R package binsreg. More specifically, the "optimal" number of bins and the bin knots (or cutoffs) are selected based on integrated mean square error (IMSE), balancing the bias-variance tradeoff in this context. For example, if we split the covariates into too many bins, then the number of observations in each bin would be too small, potentially leading to biased (spurious) effects for some bins. Conversely, if bin sizes are too large, then any meaningful heterogeneity may be masked, essentially through a mean reversion process (we revert to the average effect for the whole sample, which defeats the purpose of a heterogeneity analysis).

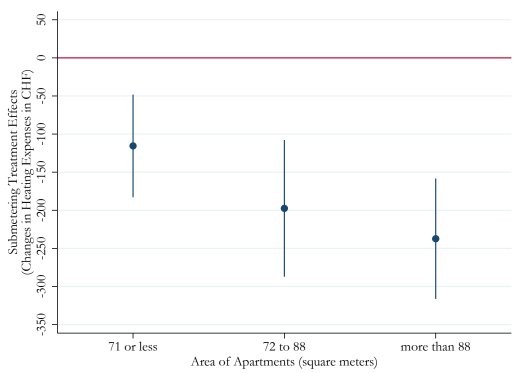
The specification used for bin selection is analogous to the heterogeneity Equation 14 from the main text. Therefore, optimal bins are selected for each covariate while simultaneously adjusting for the other covariates (including canton fixed effects, as well as indicators for turnover and renovations). We use quantile-spaced binning, rather than evenly-spaced spaced binning, such that the gaps between knots are not necessarily constant.

E.4 Summarizing treatment effect heterogeneity

(a) Effects by number of neighbors



(b) Effects by apartment area



(c) Effects by building's vintage

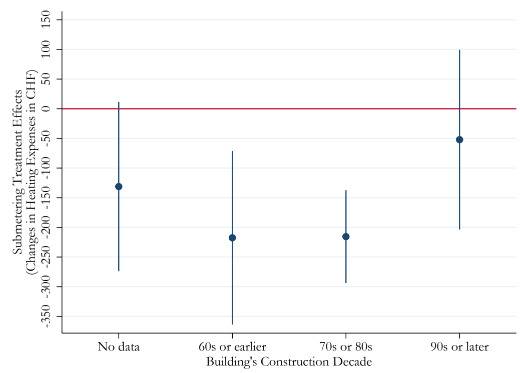
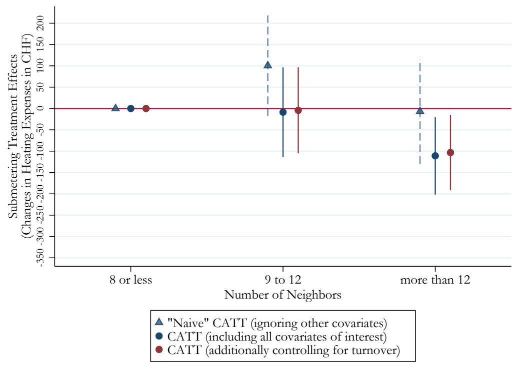


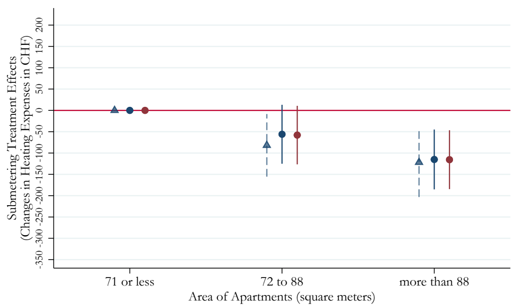
Figure E.4: ML-based "naive" estimates of submetering treatment effect heterogeneity

Notes: This figure summarizes ML-based average submetering treatment effects for sub-samples defined by (a) apartment area, (b) number of neighbors, and (c) building's vintage. These may be considered naive estimates, as they ignore the fact that these variables may be correlated and simultaneously driving heterogeneity. The blue dots represent point estimates for heterogeneity specifications without sample restrictions, other than those defined by the covariates of interest.

(a) Effects by number of neighbors



(b) Effects by apartment area



(c) Effects by building's vintage

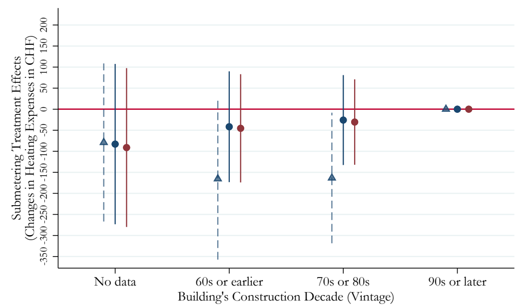


Figure E.5: Heterogeneity of submetering treatment effects by selected covariates

Notes: This Figure plots β coefficient estimates and 95% confidence intervals regression Equation 14. The non-naive coefficients capture treatment effect heterogeneity by variables of interest, while controlling for several confounding factors, including canton fixed effects. Coefficients should be interpreted compared to the omitted comparison categories (zero point estimate in the figures). Standard errors are clustered at the property level. Vintage information was not available for some buildings (rather than dropping those, we group them into a separate category labeled as "No data").

F Additional Robustness Checks

F.1 Placebo treatment effects

To further validate our machine learning approach, we implement a placebo analysis with the sample of the never-treated (not submetered) buildings. The key idea is to ramdomly allocated some of those buildings to a placebo treatment, then carry out the ML procedure (as described in the main text) and check the model's performance in the placebo post-treatment sample. Note that these buildings were not submetered, such that we do not expect substantial changes in consumption patterns in the placebo post-treatment. Accrodingly, the ML predictions from this exercise should closely match real consumption in the placebo post-treatment. This would provide further credibility that our model results in accurate prediction of a counterfactual in the absence of submetering.

We aim to closely mimic our true empirical setting, for which there was staggered rollout of submetering. For this reason, the placebo treatment allocation is not completely random. Rather, we impose that treatment is more likely to happen in certain years. Figure F.1 presents the rollout of placebo treatment, which closely resembles the true rollout patterns. The total number of placebo treated buildings is 186.

Finally, with this new semi-synthetic sample, we implement a machine learning procedure where the training sample consists of observations from the never-treated buildings plus the placebo pre-treatment. Results from the placebo treatment effects are presented in Figure F.2. For comparison, we also present the true treatment effects. As expected, we find no significant errors in the pre-treatment sample, since those data were used to train the model. Most importantly, the errors are also close to zero also in the placebo post-treatment sample. We therefore conclude that our model can accurately predict future consumption patterns in a counterfactual scenario without submetering.

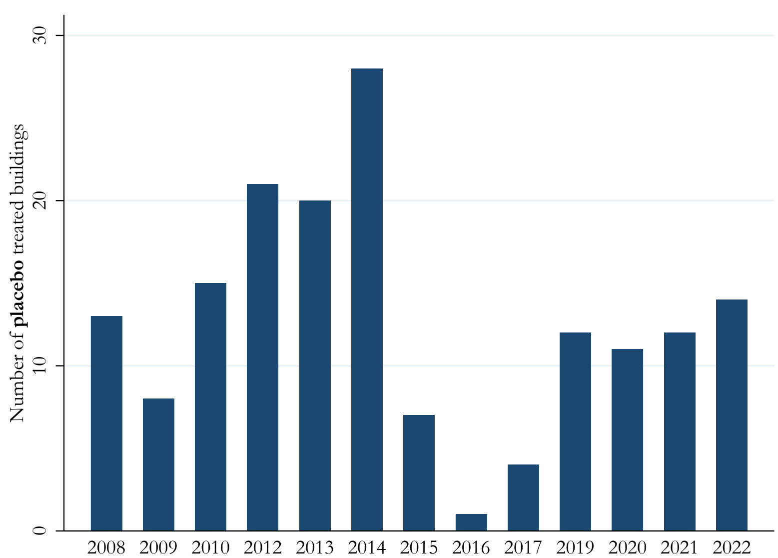


Figure F.1: Placebo treatment rollout

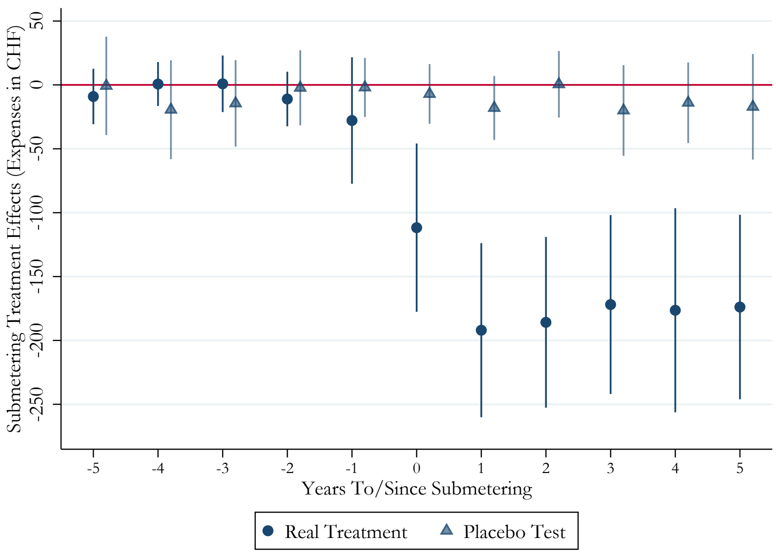


Figure F.2: Real and placebo treatment effects

Notes: The blue dots reproduce the true treatment effects as shown in the main text. The blue triangles correspond to predictions for the placebo treatment exercise described in this Appendix. Standard errors are clustered at the property level. The "whiskers" around point estimates correspond to 95% confidence intervals.

F.2 Concurrent Renovations

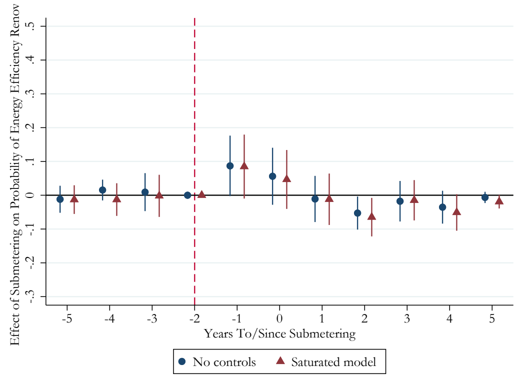
Our data provider (the real estate management company) keeps a record of all major renovations carried out in the properties they manage. These records include the calendar year of the renovation, as well as some text describing key changes. As described in Appendix C, we use the OpenAI API to to distinguish between comprehensive renovations, non-comprehensive renovations that change the energy efficiency of the building, and other renovations. In this section, we focus on energy-related renovations, which include comprehensive renovations and energy efficiency renovations. We investigate how often submetering is introduced in conjunction with these types of renovations.

To provide further insight regarding the timing of renovations, we implement variants of our main event study specification in Equation 10, replacing the outcome by indicators for each of the renovation categories. These indicators are equal to one in the year that a renovation happened, zero otherwise. We therefore estimate linear probability models to test whether submetering affects the probability of apartment renovations. Results are presented in Figure F.3. We find evidence suggesting that both energy efficiency and comprehensive renovations happen concurrently with (or just before) submetering. The event studies suggest a 20 percentage point increase in energy related renovation probabilities in $r = -1$, with comprehensive and energy efficiency renovations contributing approximately 10 percentage points each.

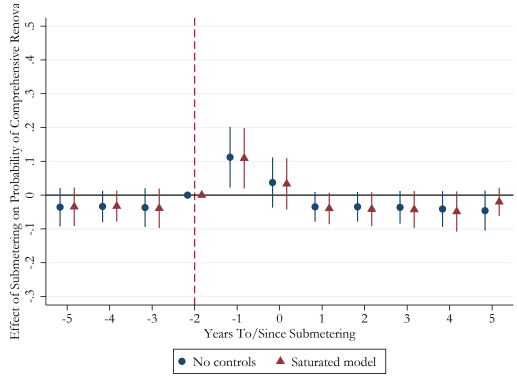
In our setting it is unlikely that renovations are a consequence of submetering, for two reasons. First, as shown in Figure F.3, renovations seem to happen concurrently or before submetering, not after submetering. Second, households in our sample are tenants, and they have limited influence regarding structural changes to their building. The most likely explanation for the concurrence of submetering and renovations is that submetering is sometimes implemented as part of a broader renovation plan.

We account for concurrent renovations in our average treatment effect and heterogeneity analyses. In Section 4.1, we show results from two models: one that controls only for year and apartment fixed effects, and another that includes controls for renovations. We find that controlling for renovations matters, as the saturated model finds smaller effects on heating energy consumption (17%) than the model without these controls (20.8%). As we are primarily interested in the effect of submetering (and not in the effect of renovations), we present the smaller effect from the saturated model as our main estimate. Also, in Table 1, we investigate whether energy-related renovations influence submetering treatment effects, within our ML framework. We find that renovated buildings reduce expenses by an additional 26 CHF after submetering, but that difference is not statistically significant. We therefore conclude that buildings with and without concurrent renovations are similar in their treatment effects.

(a) Energy Efficiency Renovations



(b) Comprehensive Renovations



(c) Energy-Related Renovations

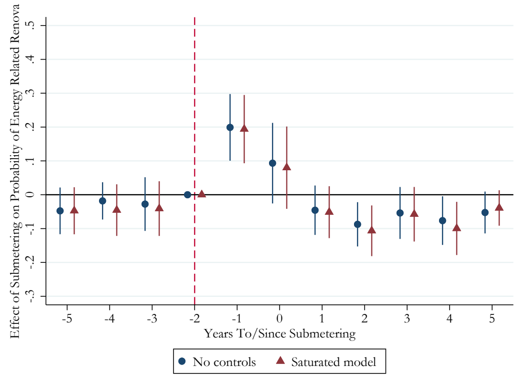


Figure F.3: Does submetering affect the probability of renovations?

Notes: This figure presents estimates of both ATT and $ATT(r)$ for the effects of submetering according to the procedure from Sun and Abraham (2021) described in Section 4.1. The outcome variables are indicators for renovations. The ATT estimates and their standard errors, included as text within the graphs, pool the effects from periods $r = 1$ to $r = 5$. The $ATT(r)$ estimates are presented graphically. All specifications include year and apartment fixed effects. The saturated specifications also control for heating degree days and canton-by-year fixed effects. Standard errors are clustered at the property level. The "whiskers" around the point estimates represent 95% confidence intervals.

F.3 Rents

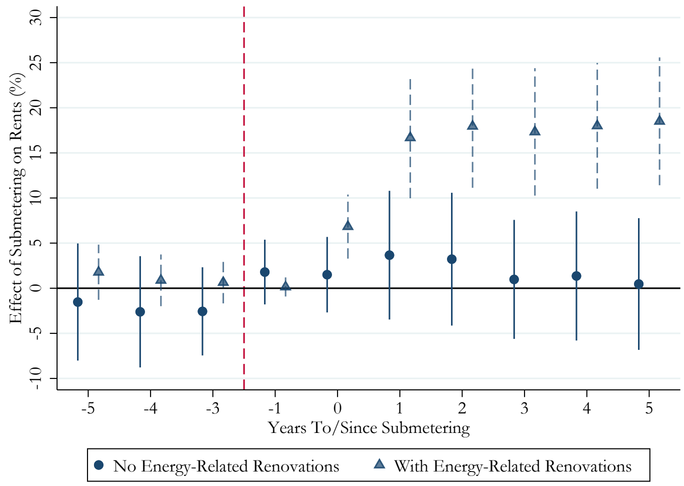
We also investigate the effects of submetering on rents. One complication is that rents might concurrently be affected by renovations or turnover. Tenancy regulations in Switzerland allow for rent increases in both cases (FOH, 2022). We should therefore control for these confounding factors. For this, we implement variants of the event study Equation D.1, including interactions for the confounders of interest as follows:

$$\begin{aligned} Y_{i,t} &= \sum_{r\neq -2}\sum_{g=0}^{1}\beta_{r,g}S_i \times \mathbb{1}[r = t-(q_i)] \times \mathbb{1}[\text{Renovations} = g] + \gamma_i + \gamma_t + \varepsilon_{i,t} \ , \\ Y_{i,t} &= \sum_{r\neq -2}\sum_{g=0}^{1}\beta_{r,g}S_i \times \mathbb{1}[r = t-(q_i)] \times \mathbb{1}[\text{Turnover} = g] + \gamma_i + \gamma_t + \varepsilon_{i,t} \ , \end{aligned} \tag{F.1}$$

where the outcome variable $Y_{i,t}$ is log monthly rents. Note the inclusion of indicator variables $\mathbb{1}[\text{Renovations} = g]$ and $\mathbb{1}[\text{Turnover} = g]$ for energy-related renovations and turnover, respectively. These are defined to be time-fixed. We tag apartments that experienced renovations during *any* event times from -5 to +5. For the case of turnover, we restrict that window to -2 to +2, assuming that turnover in earlier/later time periods are unlikely to be related to the submetering implementation.

The specifications F.1 above allow us to estimate the effects of submetering $\beta_{r,g}$ on rent, separately for apartments that experienced renovations or turnover (i.e., when $g = 1$) versus those that did not (i.e., when $g = 0$). Results are presented in Figure F.4 below. We find that rents increase substantially (15%, on average) one year after submetering, but only for apartments that were renovated or that had tenants changes (point estimates illustrated by blue triangles in the Figure). Conversely, the rent increase is smaller than 5% and not statistically significant for non-renovated/no-turnover apartments. It is therefore unlikely that the submetering technology itself is capitalized into rent increases.

(a) Accounting for Renovations



(b) Accounting for Turnover

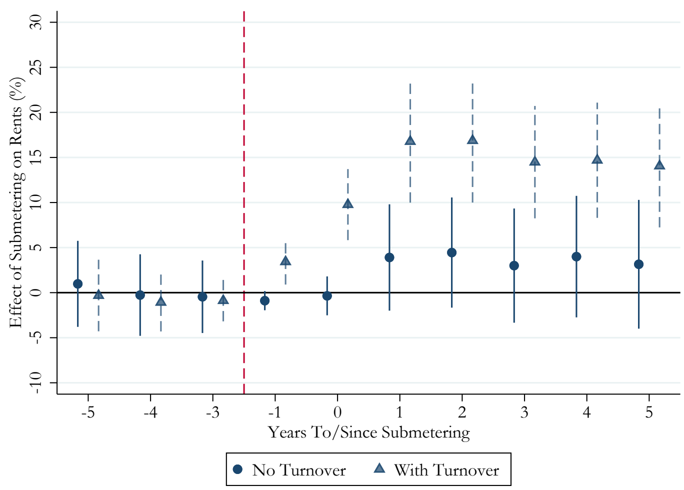


Figure F.4: Submetering Effects on Rent

Notes: This figure presents estimates $ATT(r)$ for the effects of submetering on rents, according to specifications F.1. The outcome variable is month rent in logs. The $ATT(r)$ estimates are presented graphically. All specifications include year and apartment fixed effects. Standard errors are clustered at the property level. The "whiskers" around the point estimates represent 95% confidence intervals.

G Survey

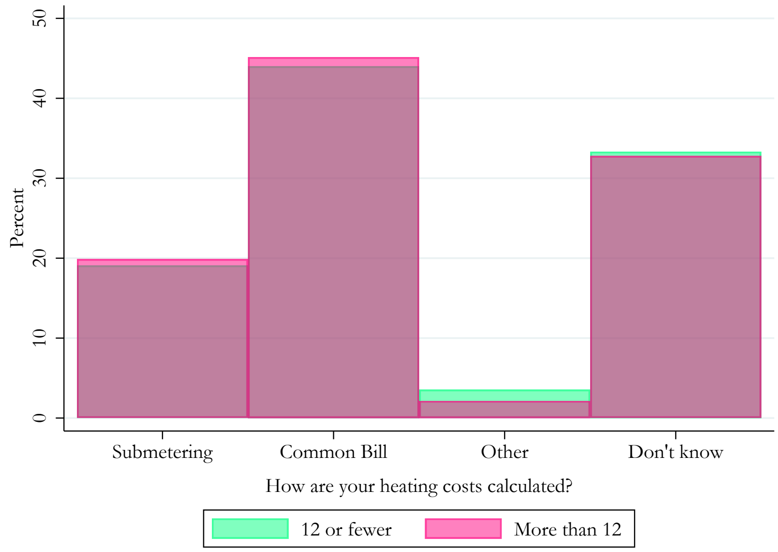


Figure G.1: Beliefs about Billing Method

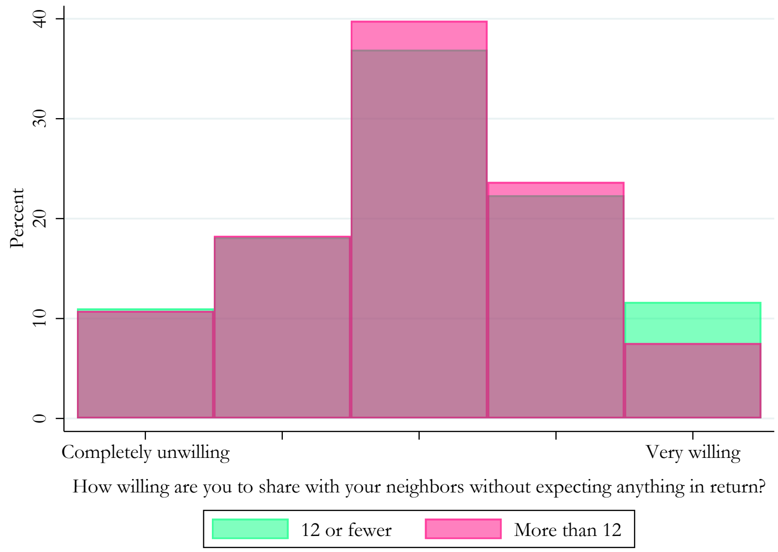


Figure G.2: Altruism toward Neighbors

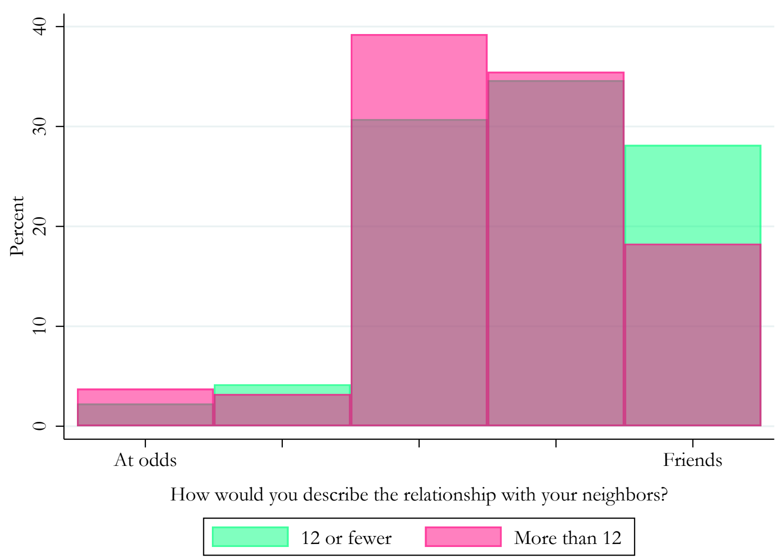


Figure G.3: Relationship with Neighbors