

Discussion Paper Series – CRC TR 224

Discussion Paper No. 629  
Project B 07

# The Tragedy of the Common Heating Bill

Harald Mayr<sup>1</sup>  
Mateus Souza<sup>2</sup>

October 2025  
(First Version : January 2025)

<sup>1</sup>University of Zurich, Email: harald.mayr@econ.uzh.ch  
<sup>2</sup>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.

# The Tragedy of the Common Heating Bill

Harald Mayr<sup>a</sup> and Mateus Souza<sup>b</sup>

<sup>a</sup>University of Zurich, [harald.mayr@econ.uzh.ch](mailto:harald.mayr@econ.uzh.ch)

<sup>b</sup>University of Mannheim, [mateus.souza@uni-mannheim.de](mailto:mateus.souza@uni-mannheim.de)

October 5, 2025

## Abstract

We leverage quasi-experimental variation to study how group size influences free-riding behavior within a high-expense environment. When buildings lack apartment-specific heat meters, tenants use simple heuristics to split a common bill. We estimate that the staggered roll-out of a corrective technology, “submetering,” reduces heating expenses by 17%, on average. Machine learning techniques uncover substantial heterogeneity, consistent with strategic exit of free-riders and coordination failures in large buildings. Tenants in smaller buildings show minimal response and are surprisingly price elastic. Only a minority of households exploits the free-riding incentives. Targeted submetering policies can be much more cost-effective than universal mandates.

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

**JEL Codes:** D62, Q41, Q52

# 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. 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. The idea that cooperation is difficult to achieve in large groups goes back to Hume (1739) and the seminal work of Olson (1965). Our study provides novel evidence into this question by leveraging quasi-experimental variation in the context of residential heating energy consumption.

More specifically, we analyze energy consumption behavior for a large sample of renters in Swiss apartment buildings. We introduce a theoretical framework in which each household derives utility from heating their own apartment, yet, their energy bill payments may depend on the consumption of their neighbors. Approximately two thirds of centrally-heated buildings in Switzerland lack apartment-specific heat meters (SVW, 2022).<sup>1</sup> The implication is that heating energy can only be measured at the building level, such that total heating expenses have to be shared among all neighbors. This “common heating bill” regime leads to a market failure in which each household’s consumption imposes an externality on their neighbors via increased energy expenses.<sup>2</sup> This is analogous to the “diner’s dilemma” when splitting restaurant bills (Glance and Huberman, 1994; Gneezy, Haruvy, and Yafe, 2004).

We investigate this “dilemma” in a setting that is particularly insightful because it offers a reliable measure of cooperation (building-level energy consumption), variation in the benefits of cooperation (energy prices), and differences in group sizes (number of bill-sharing neighbors). Further, we observe behavior in naturally occurring groups with long-term interactions and with high stakes—given that households spend a significant portion of their income on heating energy. A straightforward implication of common billing is that it provides incentives to over-consume energy, as it effectively shrinks the marginal prices faced by consumers. Our theoretical framework is novel in that it reveals that this price distortion

---

<sup>1</sup>The lack of apartment-specific meters is widespread. 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). 17% of renters in the US face fixed heating expenses included in the rent (AHS, 2023).

<sup>2</sup>Note that this is in addition to the climate externality attributed to the carbon emissions from burning fossil fuels for heating.

depends on the number of bill-sharing neighbors, the energy efficiency of apartments, and household-specific preference parameters (such as intolerance to heat). Consequently, the effects of “submetering”—a technology that corrects the market failure—will also depend on those factors.<sup>3</sup>

Our framework reveals that, all else equal, price incentives become weaker in larger groups (more neighbors), leading to more over-consumption. This also implies stronger gains from submetering larger buildings (that have many apartments). The framework further provides a sufficient statistic for calculating building-specific welfare effects of submetering, depending both on the number of bill-sharing neighbors and on carbon externalities. A complementary result is that the price elasticity of heating energy demand should increase after submetering, again depending on the number of neighbors.

We test our framework’s predictions by leveraging the staggered introduction of heat submetering in 185 apartment buildings in Switzerland. This constitutes a difference-in-differences setting, where the “treated” buildings are those that switched to individual billing after the installation of submeters. The comparison group are buildings 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, covering the years 2006 to 2023. With this data, we begin by estimating event studies for the effects of submetering on heating expenses, using the heterogeneity-robust approach from [Sun and Abraham \(2021\)](#).

Our main identifying assumption is parallel trends. That is, in the absence of treatment, energy expenses from submetered and comparison apartments would have followed parallel trends/growth. We acknowledge that we cannot point to specific reasons why some building managers decided to switch to submetering. Nevertheless, the inclusion of apartment fixed effects attenuates concerns about potential selection into treatment driven by time-invariant factors (even if unobservable).<sup>4</sup> Still, time-varying factors may drive differential trends and selection. For this reason, we test and control for the influence of renovations and tenant turnover—to the best of our knowledge, these have not been explored in previous studies of submetering. We also show that our results hold for a restricted sub-sample of buildings matched on baseline covariates ([Appendix E.5](#)).

---

<sup>3</sup>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 a related problem where all members within a given household share a single heating bill (for their home), thus leading to incentives to over-consume ([Jack et al., 2024](#)).

<sup>4</sup>See [Ghanem, Sant’Anna, and Wüthrich \(2022\)](#) for a more complete discussion of the types of selection mechanisms that are allowed for unbiased estimation in the context of difference-in-differences. We invoke an assumption that selection depends only on fixed effects. In our context, this implies a time-homogeneity assumption for the conditional mean of idiosyncratic shocks to heating expenses.

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.<sup>5</sup> 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.

To explore heterogeneity, we use machine learning (ML) algorithms for direct imputation of untreated counterfactuals (Borusyak, Jaravel, and Spiess, 2024; Souza, 2019)—i.e., we predict energy expenses assuming that submetering had not happened. These are the building blocks for the estimation *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, even after controlling for apartment area and building vintage. Importantly, we document that treatment effects are stronger when we observe tenant turnover. This 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 (spillovers) between apartments do not significantly influence our results.<sup>6</sup>

The ML techniques also serve to estimate *building-specific* treatment and welfare effects. 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 86 Swiss francs (CHF), we find that submetering was cost-effective for only 9% of the treated buildings. This result rationalizes why the common heating bill is still prevalent in Switzerland and many other countries (see footnote 1). However, when we account for a SCC of 185 CHF/tCO<sub>2</sub> (Rennert et al., 2022), submetering is welfare-improving for 40% of treated buildings. Our results therefore suggest that submetering policies are justified for many apartment buildings across Switzerland, although universal submetering mandates might not be cost-effective.

---

<sup>5</sup>Anticipation effects have also been documented for submetering in other contexts (e.g., Elinder et al., 2024). In our setting, anticipation is plausible given that submetering often implies the installation of salient devices on radiators. Some misinformed renters might assume that submetered bills will be charged immediately after installation, while in fact they apply only in the following billing year.

<sup>6</sup>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.

In a complementary analysis, we investigate whether submetered households are more price elastic than those under common billing. We leverage energy price variations across regions, time, and the buildings’ heating fuel type in an empirical strategy similar to Myers (2019). Under the common heating bill, we find a price elasticity of  $-0.619$ , whereas households under the submetering regime are more price responsive ( $p < 0.01$ ), with an elasticity of  $-0.716$ .<sup>7</sup> Again, we find heterogeneity with respect to the number of neighbors in a building. The differential effect is driven by buildings with more than 12 neighbors, who are substantially less price responsive under the common heating bill. Submetering does not seem to change the price responsiveness of households in small apartment buildings.

Fully informed free-riding behavior alone cannot explain the magnitudes of our estimates for price elasticities and the effects of submetering on energy consumption. The relatively large price elasticity under the common heating bill is surprising, since these households face very low marginal heating costs. Conversely, the estimated elasticities imply that a switch from common billing to submetering should have induced a sharper decrease in energy consumption. These results can be jointly rationalized if some households do not exploit the free-riding incentive provided by the common heating bill. This behavior can be explained by various mechanisms, including “schmeduling” (Ito, 2014; Shaffer, 2020; Ito and Zhang, 2023), unawareness of the common heating bill, and pro-social behavior (Kaufmann, Andre, and Kőszegi, 2024). We discuss these mechanisms and provide survey evidence showing that many households in our setting are unaware of the common heating bill. Our results suggest that only about one-fifth of the households in our sample are fully free-riding under common billing.

This research relates to a sizeable literature on the effects of group size on cooperation. Evidence from experimental studies is mixed: some suggest that cooperation is easier in small groups (Marwell and Ames, 1979; Nosenzo, Quercia, and Sefton, 2015; Arifovic et al., 2023), while others find that cooperation *increases* with group size (Isaac, Walker, and Williams, 1994; Diederich, Goeschl, and Waichman, 2016; Nosenzo, Quercia, and Sefton, 2015). A related body of observational research explores the effect of group size in contexts such as fisheries, agriculture, and forestry commons (Ostrom, 1990; Agrawal, 2000; Poteete

---

<sup>7</sup> Note that we use annual data, such that estimated elasticities can be expected to be larger than those using monthly price variations, for example. Our strategy assumes that households are 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. Our estimates are also broadly in line with Filippini and Kumar (2021), who find an elasticity of  $-0.73$  for residential gas demand in Switzerland. These 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) estimate an elasticity of  $-0.7$  for district heating in South Korea, Schulte and Heindl (2017) and Rehdanz (2007) find elasticities close to  $-0.5$  for space heating in Germany, and Trotta, Hansen, and Sommer (2022) estimate a  $-0.53$  elasticity for district heating in Denmark.

and Ostrom, 2004; Casari and Tagliapietra, 2018). We contribute new insights from a quasi-experiment that corrects the free-riding problem, while analyzing an objective measure of cooperation across hundreds of groups of different sizes.

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; Carrillo, Contreras, and Scartascini, 2024), and hot water (Elinder et al., 2024) find that submetering leads to large reductions in consumption—between 18% and 36%. Substantial energy savings are also found in a related strand of literature analyzing the effects of switching from landlord-pay regimes (where tenants face zero marginal costs for heating) to tenant-pay regimes (e.g., Brewer, 2022). Direct 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. They find that the switch reduced heating energy consumption by 10%. Stronger reductions in our setting might be explained by the fact that households could not opt-out of submetered billing (except by leaving the building entirely). We further show 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 (Borenstein and Bushnell, 2022; Asker et al., 2024).

Finally, 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, information frictions, or split incentives between landlords and tenants (e.g., Gillingham, Harding, and Rapson, 2012; Myers, 2020).

## 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. We discuss deviations from free-riding in Section 5.

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  $c_i$ .<sup>8</sup> 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

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

The parameter  $\kappa_i$  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).<sup>9</sup> We focus on the context of heating (not cooling), such that  $T_0 \leq T(e_i)$  and  $e_i \geq 0$ . Combining these ingredients, the utility function of household  $i$  can be written as

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

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  $j = (1, \dots, N)$ , such that

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

---

<sup>8</sup>Quasi-linear utility is a standard assumption in the context of residential energy consumption (e.g., Allcott and Kessler, 2019; Ito, 2015).

<sup>9</sup>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.



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).

### Private solution

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

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

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  $\kappa_i$ , heat loss  $H$  and, importantly, the billing regime  $\theta$ . 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.<sup>10</sup> 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

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

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  $\theta$  from 0 to 1. The change in household  $i$ 's energy consumption in response to submetering will be:

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

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  $\kappa_i$ ), and at high energy prices  $p$ . Again,

---

<sup>10</sup>In practice,  $\theta$  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.

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* 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  $\Delta e_i^*$  is clear, submetering may either increase or decrease household  $i$ 's utility:<sup>11</sup>

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

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  $\kappa_i$ ) 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

---

<sup>11</sup>The complete derivation is presented in Appendix B.2.

externality  $\phi$  (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:

$$W = \sum_i U_i(e_{i,\theta}^*) - \phi \sum_i e_{i,\theta}^* - \mathbf{1}_{\theta>0} sN . \quad (8)$$

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:

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

Therefore, an estimate of  $\tau$ , together with the cost of submetering and the social cost of carbon are sufficient to estimate  $\Delta W$ . 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  $\phi$ , 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 that operates across most of Switzerland. The company manages apartment buildings for institutional and private real

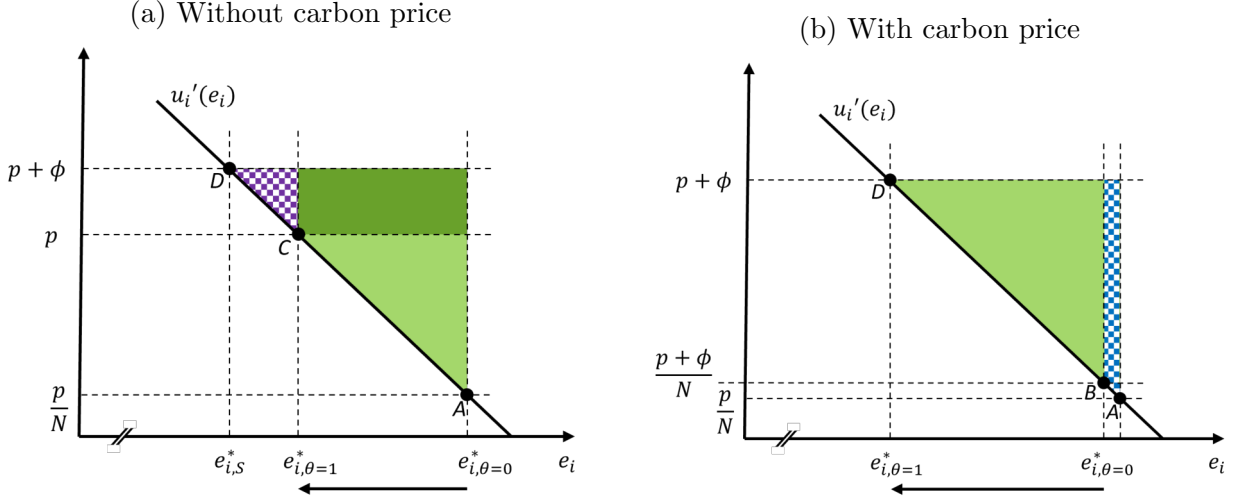


Figure 1: Private solutions and social optimum

estate owners. As part of their mandate, they provide a range of services, including the procurement and billing of heating energy. Buildings typically have a central heating system, which is used to heat the apartments and provide hot water. The 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.<sup>12</sup> 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<sub>2</sub> levy in 2008. The levy started out at 12 CHF/tCO<sub>2</sub> and was increased multiple times: to 36 CHF/tCO<sub>2</sub> in 2010, to 60 CHF/tCO<sub>2</sub> in 2014, to 84 CHF/tCO<sub>2</sub> in 2016, to 96 CHF/tCO<sub>2</sub> in 2018, and to 120 CHF/tCO<sub>2</sub> in 2022.

<sup>12</sup>Annual bill cycles are not unique to Switzerland. These are also common, for example, in Germany (e.g., Chlond, Goeschl, and Kesternich, 2025; Singhal, 2024). In California, households who install solar PV and adopt net energy metering (NEM) are also placed on a 12-month billing cycle (Borenstein, 2017). While the effects of bill frequency are outside the scope of this paper, prior literature suggests that consumption may decrease or even *increase* with higher billing frequency, depending on misperceptions about both quantities and prices (Wichman, 2017).

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.<sup>13</sup> Appendix Figure A.1 shows the geographic coverage of our sample (company market shares by Swiss cantons).<sup>14</sup> Although the share is relatively higher for the canton of Zurich, our sample covers most of the country. Also, the average rents and vintage of apartments in our sample are comparable to the national averages for residential buildings (Appendix Figure A.2).<sup>15</sup>

Our data includes 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, if we observe that all of the apartment weights are integers, then the corresponding building is assigned to the common billing regime. If at least 50% of the apartments have non-integer weights, then we assign the building to the submetering regime. Any buildings with a positive but small share ( $\leq 50\%$ ) of apartments with non-integer weights are completely dropped from our sample (less than 5% dropped). Finally, we

---

<sup>13</sup>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.

<sup>14</sup>Calculated with the same data used in our main analyses, compared to data on residential apartment buildings collected by the Swiss Statistical Office (BFS, 2016).

<sup>15</sup>We note over-sampling of buildings from the 60s and 70s, and under-sampling of older buildings (50s or before). Average rents are mostly comparable, except for smaller apartments, where our sample tends to exhibit higher costs per square meter (driven by the higher market share in Zurich).

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.

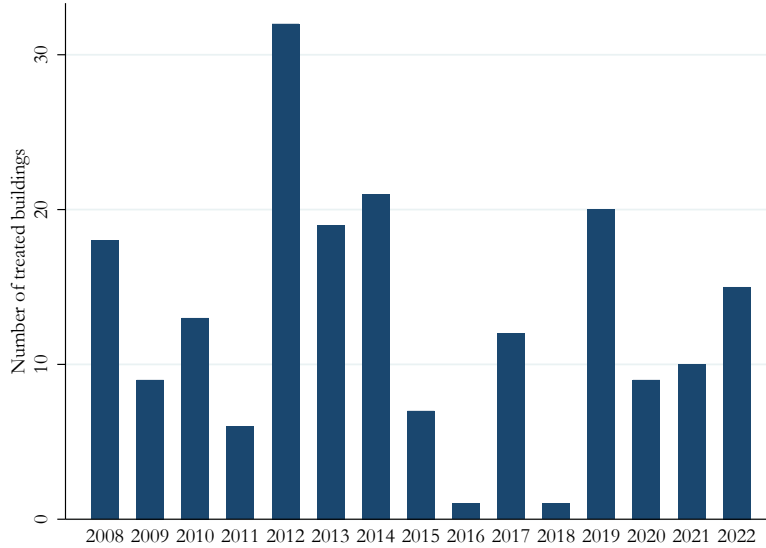


Figure 2: Year of installation of submeters

In practice, submetering is often introduced with other measures. This can include the replacement of the heating system, the insulation of the building’s facade, or 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, including the renovation year, as well as short texts describing what was performed. 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.

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. Price data for district heating and other heating sources is only available at 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;<sup>16</sup>  $S_i$  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

---

<sup>16</sup>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.

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 \(2021\)](#).<sup>17</sup> 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:

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

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;  $\gamma_i$  are apartment fixed effects;  $\gamma_t$  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.<sup>18</sup>

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.<sup>19</sup> 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  $ATT$ ).

<sup>17</sup>In Appendix E.1 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.

<sup>18</sup>Our sample contains 1,932 properties. On average, each property is constituted of 2 buildings. The maximum number of buildings per property is 31.

<sup>19</sup>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](#)).



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 our empirical strategy does *not* require quasi-random assignment of treatment dates (see [de Chaisemartin and D’Haultfoeuille, 2020](#); [Sun and Abraham, 2021](#)), and allows certain forms of selection into treatment ([Ghanem, Sant’Anna, and Wüthrich, 2022](#)). In particular, our *ATT* are still unbiased if selection is driven only by time-invariant factors, which are parsed out by the inclusion of apartment fixed effects.

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. Also, in Appendix [E.5](#) we show that our results hold for a sample of buildings matched based on pre-treatment variables. Another concern is that time-varying factors may drive differential trends and selection. We explore some potential confounders and the implication of selection more carefully in section [4.1](#). Further, we acknowledge that selection on time-fixed covariates may have implications for the external validity of our estimates. That is why we emphasize that our estimates should be interpreted as *ATT*, and not as unrestricted average treatment effects.

Finally, we 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;<sup>[20](#)</sup> and canton-by-year fixed effects. Henceforth, we refer 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 for ATT

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.

---

<sup>20</sup>In Appendix [E.4](#) we show that any systematic rent increases only happen when buildings are renovated. Therefore, by controlling for renovations we are also implicitly controlling for concurrent rent increases.

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 ( $r = 0$ ). That is the year when the intervention becomes visually salient to households: in many cases, individual devices are attached to each radiator of the apartment. Effects increase even further and stabilize once households are actually billed based on submetered consumption ( $r = 1$  and after). For our estimates of  $ATT$ , we pool the effects from periods from  $r = 1$  to  $r = 5$ , in which households have fully adjusted to the new regime. 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 [E.1](#) presents standard event study estimates obtained via two-way fixed effects. We find similar results to those in Figure [3](#). 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 \(2021\)](#).

### Selection and adjustment mechanisms

Selection on time-fixed covariates alone does not bias  $ATT$  estimated through difference-in-differences ([Ghanem, Sant’Anna, and Wüthrich, 2022](#)). Yet, this form of selection could lead to bias if it implies differential trends between treated and comparison buildings. We argue that this risk is minimal in our setting, given the flat pre-trends across several specifications. Also, we provide evidence that there is substantial covariate overlap between treated and untreated buildings in our sample. In Appendix Tables [A.1](#) and [A.2](#), we compare treatment cohorts to never-treated buildings across several characteristics. We cannot identify any systematic patterns that explain selection into treatment. Further, we implement matched differences-in-differences in Appendix [E.5](#). As shown in Appendix Table [E.2](#), our matching strategy was successful in achieving balance between groups. The resulting  $ATT$  estimates (Appendix Figure [E.7](#) and Table [E.1](#)) are remarkably similar to those from our main specifications using the full sample.

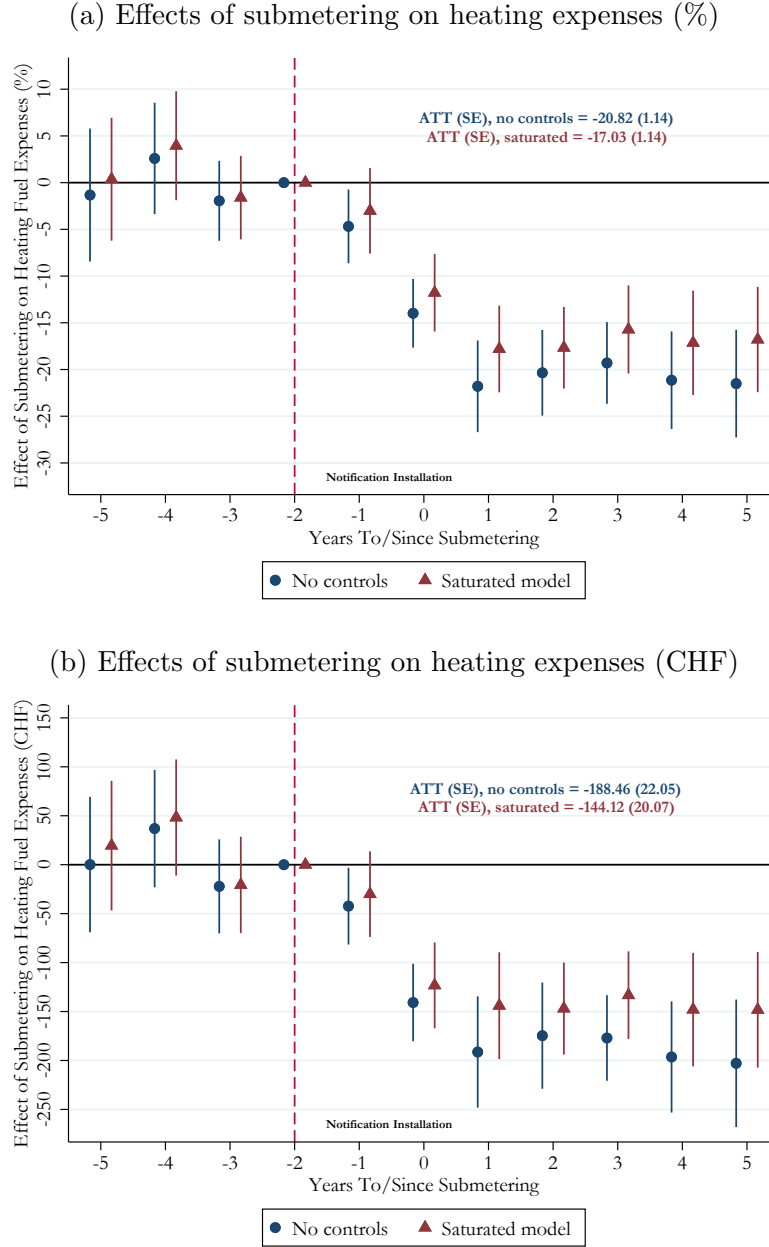


Figure 3: Heterogeneity-robust effects of submetering

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.

In terms of time-varying confounders, we assess the influence of turnover and renovations. To the best of our knowledge, these could not be controlled for in prior studies of the effects of submetering. We can expect increased turnover if some households face a substantial increase in their heating bill after submetering. 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 increased turnover after submetering. 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 after the new billing system has been implemented. This finding has 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 also be accompanied by other measures, such as energy efficiency renovations. In Appendix E.3, 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 E.4. 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

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

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

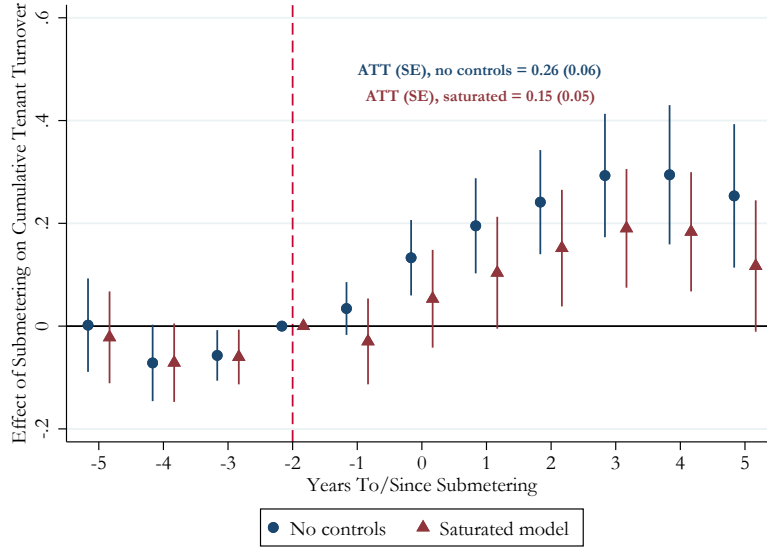


Figure 4: Heterogeneity-robust effects of submetering on cumulative turnover

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.

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 outcome 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 modulates 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.<sup>21</sup>

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 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).<sup>22</sup>

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 D.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.<sup>23</sup>

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

---

<sup>21</sup>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.

<sup>22</sup>More details are presented in Appendix D.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.

<sup>23</sup>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 D.1.

in difference-in-differences settings. For this, again we use a specification based on Souza (2019). Let  $\hat{\varepsilon}_{i,t}^{cv} = Y_{i,t} - \hat{Y}_{i,t}^{cv}$  be the pre-treatment cross-validation residuals. We then regress these residuals on event-time indicators as follows:

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

The coefficients  $\beta_r$  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 sample:  $\hat{Y}_{i,t}(0) = \hat{g}(\mathbf{X}_{i,t})$  for  $t \geq -1$ .<sup>24</sup> 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 E.2, 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 E.3 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:

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

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

The machine learning approach is flexible enough to allow 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-

---

<sup>24</sup>Recall that we allow for some anticipation effects by defining “post-treatment” as starting one year before submetering installation.



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. For example, one year after submetering, reductions in expenses range from 103 CHF to 238 CHF (95% confidence interval). 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.<sup>25</sup>

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’ heat loss ( $H$ ), tenants’ sensitivity to deviations from preferred temperature ( $\kappa_i$ ), 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 D.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 D.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

---

<sup>25</sup>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).

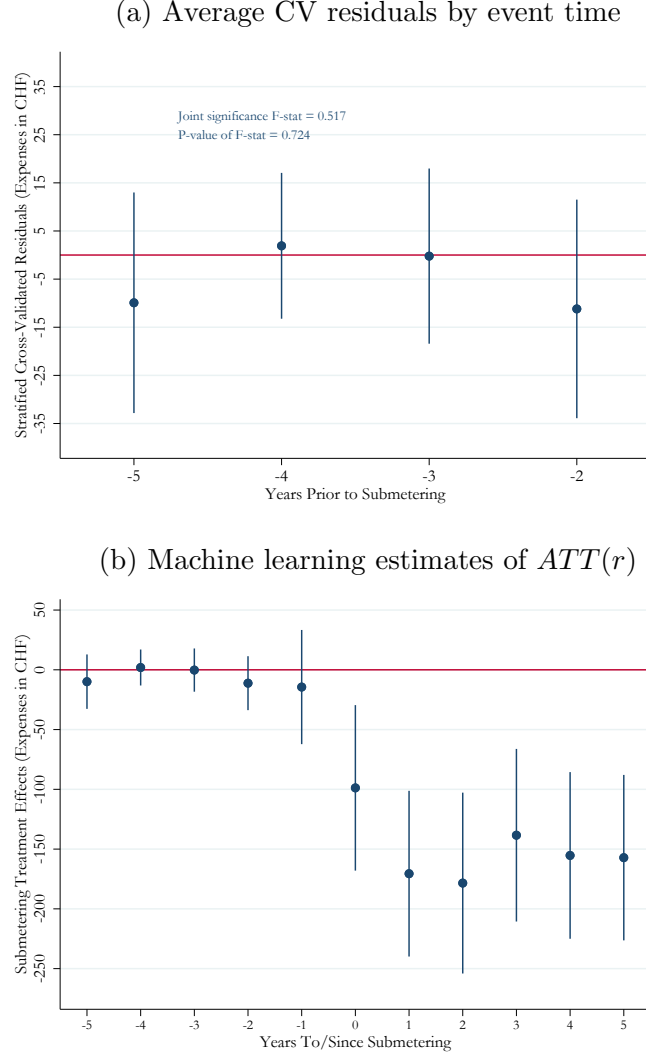


Figure 5: Machine learning average residuals and  $ATT(r)$  estimates

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.

determine the optimal number of bins for each variable (details in Appendix D.3).<sup>26</sup> Additionally, we include canton fixed effects ( $\delta_m$ ), to parse out heterogeneity that may be attributed to regional variations in energy prices. The full specification can be written as:

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

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. 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 147 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 103 CHF compared to small apartments (up to 71 sq. meters). The difference for medium apartments (72 to 88 sq. meters) is close to 41 CHF (although not statistically significant). Again, this is consistent with our framework, assuming that large apartments have higher heat loss. Regarding building's vintage, the naive estimates (Appendix Figure D.5) suggest that treatment effects were stronger for older buildings. However, those differences vanish when we jointly control for other factors. This may be due to correlations between building vintage and number of neighbors, for example.<sup>27</sup>

<sup>26</sup>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.

<sup>27</sup>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 Appendix F.1, where we explore price elasticities.

We test the robustness of our heterogeneity estimates ( $\beta^n$ ,  $\beta^a$ , and  $\beta^v$ ) to changes in Equation 14. In columns (2) to (5) 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.<sup>28</sup> We find that apartments that experienced tenant changes save an additional 64 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 (38 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 building. 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 D.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 D.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 D.2 we show that the prediction errors are uncorrelated with several other covariates.<sup>29</sup>

### Strategic turnover

We observe results that are consistent with high-user tenants leaving the building because of submetering. This is rational in case submetering increases their heating bill by more than their moving costs. We find that submetering increases turnover (see Figure 4), and that the induced turnover appears to influence heating expenses: submetering treatment effects are

<sup>28</sup>Here we tag apartments that experienced any tenant changes during submetering event times -2 to +2.

<sup>29</sup>Average errors across these other covariates are presented as a form of robustness check in Appendix D.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)
<i>Number of Neighbors (comparison group: 8 or less)</i>					
9 to 12	-31.47 (54.61)	-26.82 (52.50)	-81.98 (54.03)	-71.42 (51.25)	-69.42 (50.75)
more than 12	-147.11*** (49.34)	-138.36*** (48.38)	-144.37*** (45.51)	-131.10*** (40.73)	-130.12*** (40.32)
<i>Area of Apartment (sq. meters; comparison group: 71 or less)</i>					
72 to 88	-41.05 (33.10)	-43.70 (32.67)	-72.01* (37.55)	-68.19* (35.00)	-68.34* (35.27)
more than 88	-103.17*** (37.36)	-104.48*** (36.51)	-137.18*** (41.48)	-117.06*** (41.99)	-115.42*** (42.36)
<i>Building's Vintage (comparison group: built in the 60s or earlier)</i>					
70s or 80s	-3.05 (61.09)	-3.61 (60.74)	-27.64 (81.98)	-28.40 (73.28)	-27.45 (73.80)
90s or later	-67.09 (97.08)	-60.35 (96.70)	-139.91 (107.98)	-166.14* (93.53)	-166.43* (93.41)
Post-treat. Turnover (0/1)		-63.97*** (16.43)	-62.07*** (17.63)	-38.29** (16.21)	-37.68** (16.42)
Pre-treat. Turnover Share (%)			-3.86** (1.91)	-3.80** (1.86)	-3.86** (1.86)
Energy-related Renovations (0/1)				-69.86 (53.15)	-71.34 (53.54)
<i>Duration of Vacancies (comparison group: no vacancies)</i>					
Vacant 1 to 183 days				-88.55*** (20.68)	-89.03*** (20.75)
Vacant 184 to 366 days				-290.53*** (44.71)	-289.12*** (44.67)
<i>Location of Apartment in the Building (comparison group: ground floor)</i>					
Intermediate Floor					-12.62 (15.39)
Top Floor					-16.64 (22.74)
Regression Constant	-52.00 (54.31)	-26.15 (55.18)	48.32 (64.84)	78.90 (53.91)	88.61 (55.12)
Number of observations	6,667	6,667	5,620	5,620	5,602

Notes: Column (1) presents results from the regression in Equation 14. The outcome variable is apartment-by-year treatment effects ( $\hat{b}_{i,t}$ ). 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. We drop buildings (less than 5%) for which vintage information was not available. All specifications control for canton fixed effects. Standard errors (in parentheses) are clustered at the property level. Significance levels denoted by: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , and \*  $p < 0.1$ .

larger for apartments that experience tenant changes (see Table 1). These effects cannot be explained by renovations and the rent increase that comes with them (see Appendix E.4). 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:

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

where  $e_{.,\theta}^*$  denotes energy consumption for each type of tenant, also depending on the billing regime  $\theta$ . 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 of them might decide to leave the building.<sup>30</sup> 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  $\kappa_i$ ), 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:

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

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.

---

<sup>30</sup>A similar argument can be made about households with high  $\kappa_i$ .

## 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.<sup>31</sup> We aggregate those to the building level by summing the resulting *CATT* from all apartments that belong to the same building.<sup>32</sup> 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. We focus on interpreting the effects for stayers. The Figure reveals substantial heterogeneity, with savings ranging from 30 to 270 CHF per household annually. This heterogeneity has implications for welfare, which we discuss next.

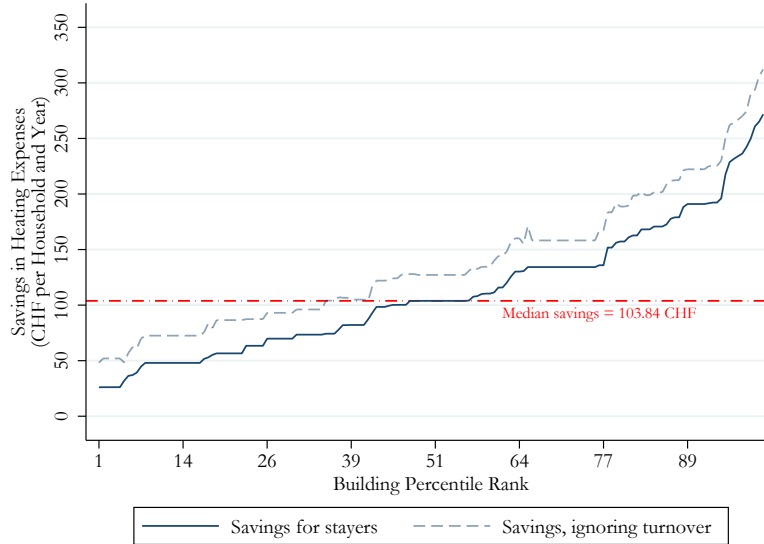


Figure 6: Building-specific treatment effects

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).

<sup>31</sup>For this exercise, we ignore any heterogeneity implied by canton fixed effects.

<sup>32</sup>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.

### 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  $\phi$ , 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 104 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.<sup>33</sup> We use the SCC estimate from Rennert et al. (2022) of 185 CHF/tCO<sub>2</sub> to value reduced CO<sub>2</sub> emissions.<sup>34</sup> We also account for the carbon tax implemented in Switzerland, which had a median value 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  $-\frac{\hat{\tau}_{stay}}{2}(1 - \frac{1}{N}) = \frac{104}{2}(1 - \frac{1}{7}) \approx 45$  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  $-\hat{\tau}_{stay} \frac{\phi}{p} = 104 \frac{(185-36)0.00022718}{0.0883} \approx 40$  CHF shows that reduced CO<sub>2</sub> emissions are a substantial benefit of submetering.<sup>35</sup> The third term of Equation 9 represents the cost of submetering  $s = 86$ . Putting all terms together ( $45+40-86 = -1$ ), we find that submetering generates a welfare loss of one CHF per household and year for our hypothetical building. However, this calculation ignores important heterogeneity.

As shown in Section 4.2, 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.

---

<sup>33</sup>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.

<sup>34</sup>[Rennert et al. \(2022\)](#) present a SCC measured in US dollars per ton of CO<sub>2</sub>. 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.

<sup>35</sup>Note that the factor 0.00022718 tCO<sub>2</sub>/kWh represents the carbon intensity of heating energy in our treatment sample.



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 is 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 submetering is privately cost-effective for only 9% of buildings in our sample, 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 40%. Since the SCC substantially changes the welfare assessment, there is a role for policy in this context. Real estate owners may not internalize the social costs of carbon, and therefore may not invest in submetering even if it is socially optimal.

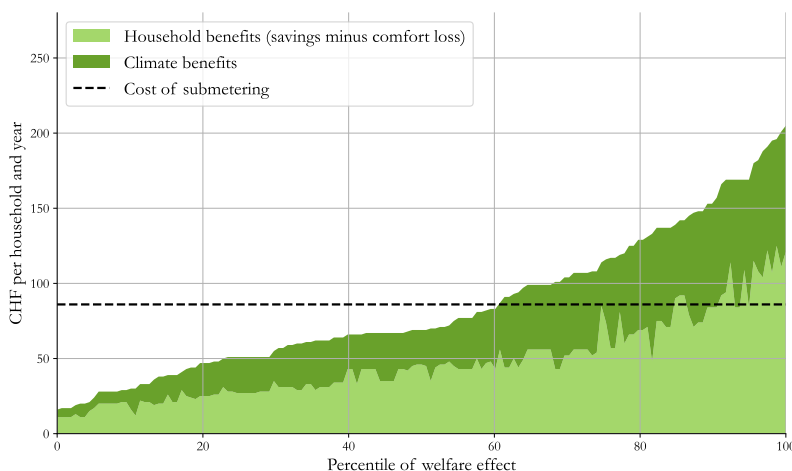


Figure 7: Welfare effects without turnover (CHF per household)

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 is the cost of submetering. Welfare effects are positive where the green areas exceed the dashed line.

## 5 Discussion

Here we briefly discuss some reasons why households may not fully exploit the free-riding incentive provided by the common heating bill. For example, some households may react to changes in average prices, rather than marginal prices (e.g., Ito, 2014). Others may have pro-social motivations to limit their consumption (e.g., Frey and Meier, 2004), or may simply be uninformed about their own billing regime. In Appendix B.3, we extend our theoretical framework to more formally assess the implications of these aforementioned behavioral factors. Their main consequence is that some households may not free-ride.

With some simplifying assumptions, it is possible to quantify the prevalence of free-riding. Let there be a share  $\rho$  of the population that is free-riding according to our theoretical framework from Section 2. If  $\rho < 1$ , the remaining share  $(1 - \rho)$  of the population does not free-ride. Following the arguments in Appendix B.3, we assume that households who do not free-ride consume energy as if submetered ( $e_{\theta=1}^*$ ), irrespective of the actual billing regime. The treatment effect  $\tau$  of submetering on energy consumption can therefore be written as:

$$\tau = \rho(e_{\theta=1}^* - e_{\theta=0}^*). \quad (17)$$

Also, from equations (3) and (4) we can write common-bill consumption  $e_{\theta=0}^*$  as a function of submetered consumption, elasticity, and the number of bill-sharing neighbors:

$$e_{\theta=0}^* = e_{\theta=1}^* \left( 1 + \eta_{\theta=1} \frac{N-1}{N} \right). \quad (18)$$

Now, equations (17) and (18) allow us to pin down the share of free-riders  $\rho$  as:

$$\rho = \frac{\tau}{e_{\theta=1}^* \eta_{\theta=1} \frac{N-1}{N}}. \quad (19)$$

We use our estimates from the sections above to calculate  $\rho$ . The *ATT* of CHF 104 corresponds to a heating energy reduction of 1,178 kWh.<sup>36</sup> The median treated household consumes 11,262 kWh of energy under submetering, and the median building has 7 neighbors. With these numbers, we obtain  $\rho \approx \frac{0.122}{\eta_{\theta=1}}$ . The last step is to plug in estimates for demand elasticity under submetering. In Appendix F.1, we estimate elasticities by leveraging heating energy price variations across regions, time, and buildings' heating fuel type (i.e., oil, gas, or district heating), in a strategy similar to Myers (2019). One key assumption is that tenants are price takers (which we argue is feasible in the context of Switzerland, as described in Section 3).

Our specifications include region-by-time fixed effects, such that elasticities can only be identified because of differential energy price movements depending on the buildings' main heating fuel. Our elasticity estimates are broadly in line with prior literature (see footnote 7). In particular, we find a price elasticity of  $-0.619$  without submetering, and  $-0.716$  with submetering. Note that these are estimated with annual data. Shorter-term (i.e., monthly) elasticities are expected to be lower (e.g., Rubin and Auffhammer, 2024). In Table 2 we test the sensitivity of our results to varying elasticity assumptions.

---

<sup>36</sup>With a simplifying assumption that households face the average energy price in our sample: CHF 0.0883 per kWh.

With a submetered elasticity of 0.716, the calculation described above yields  $\rho \approx \frac{0.122}{0.716} \approx 17\%$ . This implies that, for a median building, only one-fifth of households free-ride under the common heating bill.<sup>37</sup> This share may seem surprisingly low at first glance. It is not so surprising, however, if we consider that schmeduling and altruism seem to be common in the population (Rees-Jones and Taubinsky, 2020; Fehr and Schmidt, 2006). Unawareness of the common heating bill is also prevalent. We document this fact presenting new survey evidence in Appendix F.2. We surveyed 835 tenants under common billing. Almost 33% did not know their billing regime, and 19% thought that they were submetered.

Equation (19) can also be used to calculate the share of free-riders for different subsamples, or with different assumptions about elasticities. We illustrate this point in Table 2, where we calculate  $\rho$  with estimated (-0.7), moderate (-0.5), and low (-0.3) elasticities, and across building sizes (6, 10, or 14 neighbors). Again, we use the *CATT* estimates according to Table 1, and submetered energy consumption based on predictions from section 4.2. As expected, the share of free-riders increases with the number of neighbors. This is especially noticeable for the larger buildings (14 neighbors), with a share that is primarily driven by the stronger estimated *CATT*.

The calculated share of free-riders is sensitive to assumptions about energy price elasticities. With a low elasticity (-0.3) and for large buildings, the share reaches almost 100%. However, we argue that full free-riding is unlikely in our setting. As shown through survey evidence, many households do not even know their billing regime, for example. These findings therefore serve as complementary evidence in favor of moderate to high heating energy price elasticities in the Swiss context. When using moderate elasticities (-0.5) that are in line with prior literature (footnote 7), we find that the share of free-riders is unlikely to exceed 60%.

In summary, our results suggest that most households do not free-ride under the common heating bill. This would be consistent with the prevalence of alternative behaviors documented in the literature (schmeduling and altruism), and in our survey (unawareness of the common heating bill). It is also consistent with moderate to high demand elasticities. Households might behave as if they were submetered, thereby reacting more strongly to price variation. As such, the modest treatment effect of submetering and the relatively large elasticity under the common heating bill may be two sides of the same coin—a large share of the population does not exploit the free-riding incentive of the common heating bill.

---

<sup>37</sup>This result also implies that submetering reduces free-riders' consumption by 38%, a large effect roughly corresponding to a 5°C decrease in indoor temperature (BFE Bundesamt für Energie, 2014). This decrease 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.

Table 2: Share of free-riders across sub-samples, varying assumptions

	Number of Neighbors		
	6	10	14
Elasticity	Share of Free-Riders		
Estimated [-0.7]	0.157	0.179	0.419
Moderate [-0.5]	0.226	0.257	0.600
Low [-0.3]	0.376	0.428	0.999

Notes: Share of free-riders calculated according to Equation (19). Submetered energy consumption is based on predictions from section 4.2. The *CATT* across sub-samples are calculated using the coefficients from column (2) of Table 1 (no turnover). The first column presents assumptions on the price elasticity of energy demand. Details for “Estimated” elasticity are presented in Appendix F.1. We assume all households face the average energy price in our sample: CHF 0.0883 per kWh.

## 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. One key result 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 9% 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 40% 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.

## Acknowledgments

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, Kathrine von Graevenitz, Ulrich Wagner, and from seminar participants at AERE 2023, AERE@OSWEET, CESifo Munich, CEPR FGN Workshop, EAERE 2023, EEA 2023, ESPE 2025, 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, World Congress of the Econometric Society 2025, and ZEW Mannheim.

## Funding and conflict of interest

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). The authors declare no conflict of interest.

## 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. 2017. “Private net benefits of residential solar PV: The role of electricity tariffs, tax incentives, and rebates.” *Journal of the Association of Environmental and Resource Economists* 4 (S1):S85–S122.
- 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>.

- Carrillo, Paul E., Ivette Contreras, and Carlos Scartascini. 2024. “Turn off the faucet: Can individual meters reduce water consumption?” *Journal of Environmental Economics and Management* 128:103065. URL <https://www.sciencedirect.com/science/article/pii/S0095069624001396>.
- 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* .
- Chlond, Bettina, Timo Goeschl, and Martin Kesternich. 2025. “Money versus procedures—Evidence from an energy efficiency assistance program.” *Journal of Environmental Economics and Management* 130:103080.
- 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 and the U.S. Census Bureau (AHS). 2023. “American Housing Survey 2023.” URL <https://www.census.gov/programs-surveys/ahs/data.html>.
- Deweese, 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.
- 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.”
- European Bank for Reconstruction and Development (EBRD). 2023. “Transition Report 2023-24: Transitions Big and Small – Chapter 4. Houses, Homes and Heating.” URL <https://2023.tr-ebrd.com>. Responsibility for the content of the report is taken by the Office of the Chief Economist.
- 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](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.
- Ghanem, Dalia, Pedro HC Sant’Anna, and Kaspar Wüthrich. 2022. “Selection and parallel trends.” *arXiv preprint arXiv:2203.09001* .
- 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*.
- Iacus, Stefano M, Gary King, and Giuseppe Porro. 2012. “Causal inference without balance checking: Coarsened exact matching.” *Political analysis* 20 (1):1–24.
- 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.
- . 2015. “Asymmetric incentives in subsidies: Evidence from a large-scale electricity rebate program.” *American Economic Journal: Economic Policy* 7 (3):209–237.
- 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 Köszegi. 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](http://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: Public goods and the theory of groups*. 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.
- Rees-Jones, Alex and Dmitry Taubinsky. 2020. “Measuring “Schmeduling”.” *The Review of Economic Studies* 87 (5):2399–2438. URL <https://doi.org/10.1093/restud/rdz045>.
- Rehdanz, Katrin. 2007. “Determinants of residential space heating expenditures in Germany.” *Energy Economics* 29 (2):167–182.
- 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 CO<sub>2</sub>.” *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.
- Rubin, Edward and Maximilian Auffhammer. 2024. “Quantifying heterogeneity in the price elasticity of residential natural gas.” *Journal of the Association of Environmental and Resource Economists* 11 (2):319–357.
- 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.
- Singhal, Puja. 2024. “Inform me when it matters: Cost salience, energy consumption, and efficiency investments.” *Energy Economics* 133:107484.
- 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](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>.
- Swiss Association for Energy and Water Cost Accounting (SVW). 2022. “Wärme/Kälte.” URL <https://svw-asc.ch/energie-wasserkosten/waerme-kaelte/>.
- Swiss Federal Statistical Office (BFS). 2016. “Data on Buildings.” URL <https://www.bfs.admin.ch/bfs/en/home/statistics/construction-housing/buildings.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).
- Wichman, Casey J. 2017. “Information provision and consumer behavior: A natural experiment in billing frequency.” *Journal of Public Economics* 152:13–33.

## A Additional descriptive statistics

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

Treatment Cohort	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Heating Degree Days	Energy-Related Renovation Rate	Number of Neighbors	Years since Construction	Shares of Heating Fuel Type			
					District Heating	Gas	Oil Heating	Other/Unclear
2008	3,394.60 (235.94)	0.00 (.)	7.53 (3.50)	38.95 (15.69)	0.05 (0.23)	0.58 (0.51)	0.26 (0.45)	0.11 (0.32)
2009	4,241.21 (572.21)	0.00 (.)	7.55 (2.94)	47.82 (16.69)	0.00 (.)	0.27 (0.47)	0.64 (0.50)	0.09 (0.30)
2010	3,112.40 (504.72)	0.12 (0.34)	7.27 (4.94)	37.95 (16.90)	0.18 (0.39)	0.32 (0.48)	0.23 (0.43)	0.27 (0.46)
2011	3,180.90 (402.98)	0.33 (0.49)	15.35 (20.72)	38.12 (15.83)	0.24 (0.44)	0.35 (0.49)	0.41 (0.51)	0.00 (.)
2012	3,232.54 (385.34)	0.01 (0.09)	8.03 (3.75)	44.26 (13.24)	0.08 (0.27)	0.24 (0.43)	0.61 (0.50)	0.08 (0.27)
2013	2,900.48 (375.45)	0.02 (0.12)	10.73 (6.94)	51.45 (11.90)	0.14 (0.35)	0.45 (0.51)	0.32 (0.48)	0.09 (0.29)
2014	3,180.03 (419.04)	0.00 (.)	8.34 (3.92)	42.75 (5.95)	0.03 (0.18)	0.47 (0.51)	0.44 (0.50)	0.06 (0.25)
2015	3,043.68 (349.93)	0.00 (.)	9.26 (11.58)	31.21 (20.12)	0.26 (0.45)	0.47 (0.51)	0.05 (0.23)	0.21 (0.42)
2016	2,400.50 (.)	1.00 (.)	12.00 (.)	41.00 (.)	0.00 (.)	0.00 (.)	1.00 (.)	0.00 (.)
2017	3,774.80 (851.32)	0.03 (0.18)	13.50 (5.59)	46.29 (15.91)	0.00 (.)	0.86 (0.36)	0.00 (.)	0.14 (0.36)
2018	2,829.05 (10.96)	0.00 (.)	8.67 (6.53)	50.83 (4.49)	0.00 (.)	0.17 (0.41)	0.83 (0.41)	0.00 (.)
2019	3,263.60 (277.30)	0.10 (0.30)	6.18 (2.59)	40.86 (9.97)	0.45 (0.51)	0.14 (0.35)	0.32 (0.48)	0.09 (0.29)
2020	3,116.67 (200.37)	0.03 (0.17)	9.90 (6.33)	45.80 (7.51)	0.90 (0.32)	0.10 (0.32)	0.00 (.)	0.00 (.)
2021	3,044.63 (318.87)	0.06 (0.25)	9.67 (5.84)	51.50 (9.87)	0.00 (.)	0.42 (0.51)	0.33 (0.49)	0.25 (0.45)
2022	3,339.18 (510.33)	0.03 (0.17)	12.85 (11.31)	46.70 (22.60)	0.05 (0.22)	0.50 (0.51)	0.05 (0.22)	0.40 (0.50)
Never Treated	3,198.53 (560.85)	0.03 (0.17)	11.01 (8.75)	49.84 (13.50)	0.12 (0.33)	0.37 (0.48)	0.42 (0.49)	0.09 (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).

Market Share in Percent by Canton in 2016

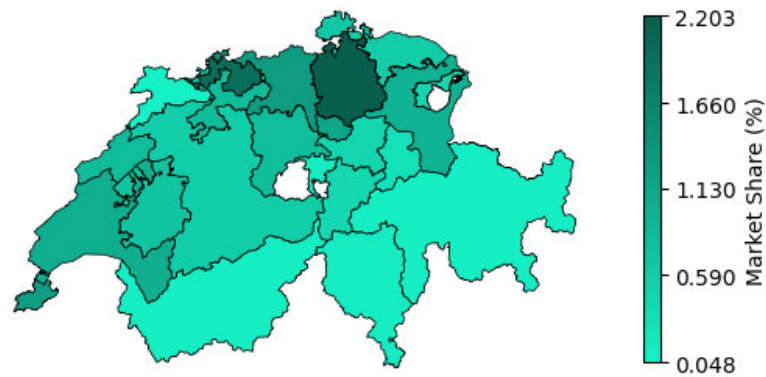


Figure A.1: Market share of property management company, by canton

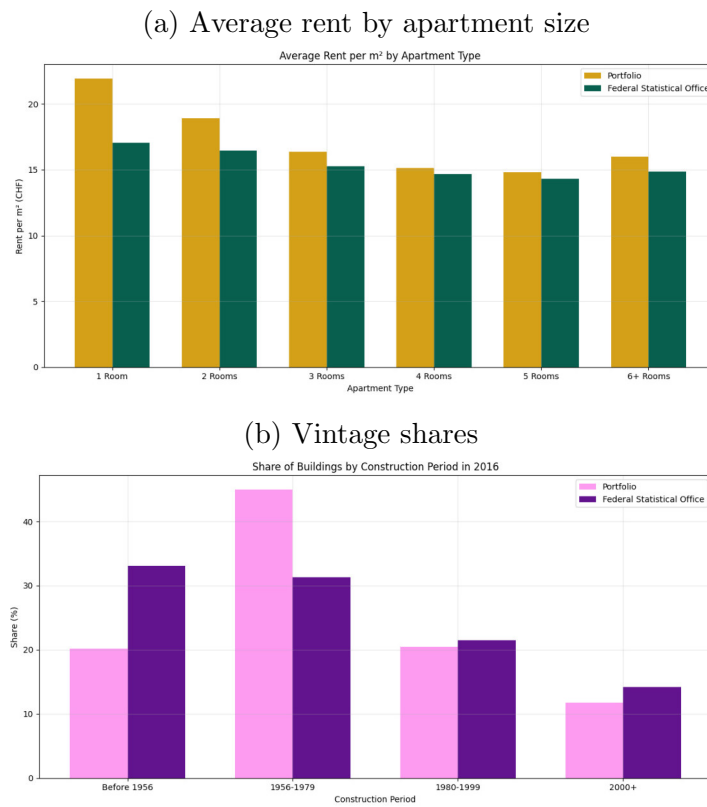


Figure A.2: Comparison of sample buildings with national averages

Notes: Comparison of buildings in our sample against averages according to data from the Swiss Federal Statistical Office (BFS, 2016). The plots use data for residential apartment buildings from the year 2016.

## 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

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

The marginal cost of energy consumption in this setting is

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

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

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

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  $N \rightarrow \infty$ . 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  $\theta$  from 0 to  $\bar{\theta}$ . In response to submetering, household  $i$  chooses a lower  $e_{i,\theta}^*$ :

$$\Delta e_i^* = e_{i,\theta=\bar{\theta}}^* - e_{i,\theta=0}^* = -\frac{H^2}{2\kappa_i} \frac{N-1}{N} p \bar{\theta}. \quad (\text{B.4})$$



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

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

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.

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

### B.3 Alternative behavioral mechanisms

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, unawareness of the common heating bill, and pro-social behavior can explain a lack of free-riding behavior. 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, unawareness of 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.

Schedulers 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

$$e_{i,\theta}^* = (T_i - T_0)H - \frac{H^2}{2\kappa_i} \left( \theta p + (1 - \theta)\bar{p} \right). \quad (\text{B.8})$$

Figure B.1 shows the implications of scheduling in our theoretical framework. Scheduling does not change consumption decisions under submetering, where households face a constant price. 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 B.1, 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, schedulers with low consumption increase their consumption, while those with high consumption decrease theirs. The overall effect on consumption is close to zero.<sup>38</sup>

Scheduling 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 B.1), 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.<sup>39</sup> 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 scheduling, our welfare estimates from Section 4.3 are a lower bound.

Schedulers 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, schedulers under the common heating bill react to price changes as if they were submetered.

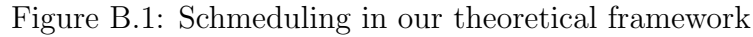
### Unawareness of the common heating bill

Inattention is well documented in the domain of energy consumption (e.g., Sallee, 2014; Jessoe and Rapson, 2014). It is plausible that some households are unaware of their billing regime. We focus on the case where households are unaware of the common heating bill. Households may also be unaware of 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.

unawareness of the common heating bill implies that some households act as if  $\theta = 1$ , independent of the true value of  $\theta$ . Hence, they behave as if they were submetered. This behavior can explain our results because the price elasticity under the common heating bill

<sup>38</sup>The overall consumption effect is *exactly* zero in Figure B.1. The figure imposes two simplifying assumptions. First, household heterogeneity is only driven by  $T_i$  (rather than  $\kappa_i$ ), implying parallel demand curves. Second,  $\bar{p}$  decreases linearly, which can be viewed as a local approximation.

<sup>39</sup>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 schedulers is unambiguously positive.



a.7

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.

## 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 Additional details on machine learning and heterogeneity analyses

### D.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 [D.1](#) and [D.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 [D.1](#) and [D.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 [D.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 0.656, 0.282, and 0.062, respectively, for Model ID 7, 9, and 11. These optimal weights were estimated via Non-Negative Least Squares ([Chen and Guestrin, 2016](#)).

Table D.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	81.058	88.661
2	3000	5	10	0.20	62.187	73.096
3	3000	5	20	0.10	81.989	90.358
4	3000	5	20	0.20	64.225	75.138
5	3000	5	30	0.10	83.649	91.797
6	3000	5	30	0.20	66.122	76.834
7	3000	10	10	0.10	32.269	56.742
8	3000	10	10	0.20	26.706	57.691
9	3000	10	20	0.10	36.277	58.074
10	3000	10	20	0.20	30.124	57.352
11	3000	10	30	0.10	39.395	59.523
12	3000	10	30	0.20	32.550	57.549
13	3000	20	10	0.10	21.901	64.876
14	3000	20	10	0.20	21.670	66.662
15	3000	20	20	0.10	23.271	62.669
16	3000	20	20	0.20	21.924	65.153
17	3000	20	30	0.10	25.059	61.090
18	3000	20	30	0.20	22.518	63.931

Notes: Cross-validated RMSE was obtained through standard 4-fold cross-validation. The algorithms selected for the final ensemble predictor are highlighted in gray.

Table D.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	88.255	187.279
2	3000	5	10	0.20	70.853	191.410
3	3000	5	20	0.10	89.025	189.240
4	3000	5	20	0.20	71.755	191.455
5	3000	5	30	0.10	89.602	188.884
6	3000	5	30	0.20	72.943	191.506
7	3000	10	10	0.10	45.796	186.670
8	3000	10	10	0.20	43.166	188.725
9	3000	10	20	0.10	47.668	185.113
10	3000	10	20	0.20	44.443	190.098
11	3000	10	30	0.10	49.687	185.631
12	3000	10	30	0.20	45.556	189.578
13	3000	20	10	0.10	41.709	196.061
14	3000	20	10	0.20	41.692	199.623
15	3000	20	20	0.10	41.936	195.659
16	3000	20	20	0.20	41.707	199.811
17	3000	20	30	0.10	42.369	195.307
18	3000	20	30	0.20	41.785	197.872

Notes: Cross-validated RMSE was obtained through “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 algorithms selected for the final ensemble predictor are highlighted in gray.

## D.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 D.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 D.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.



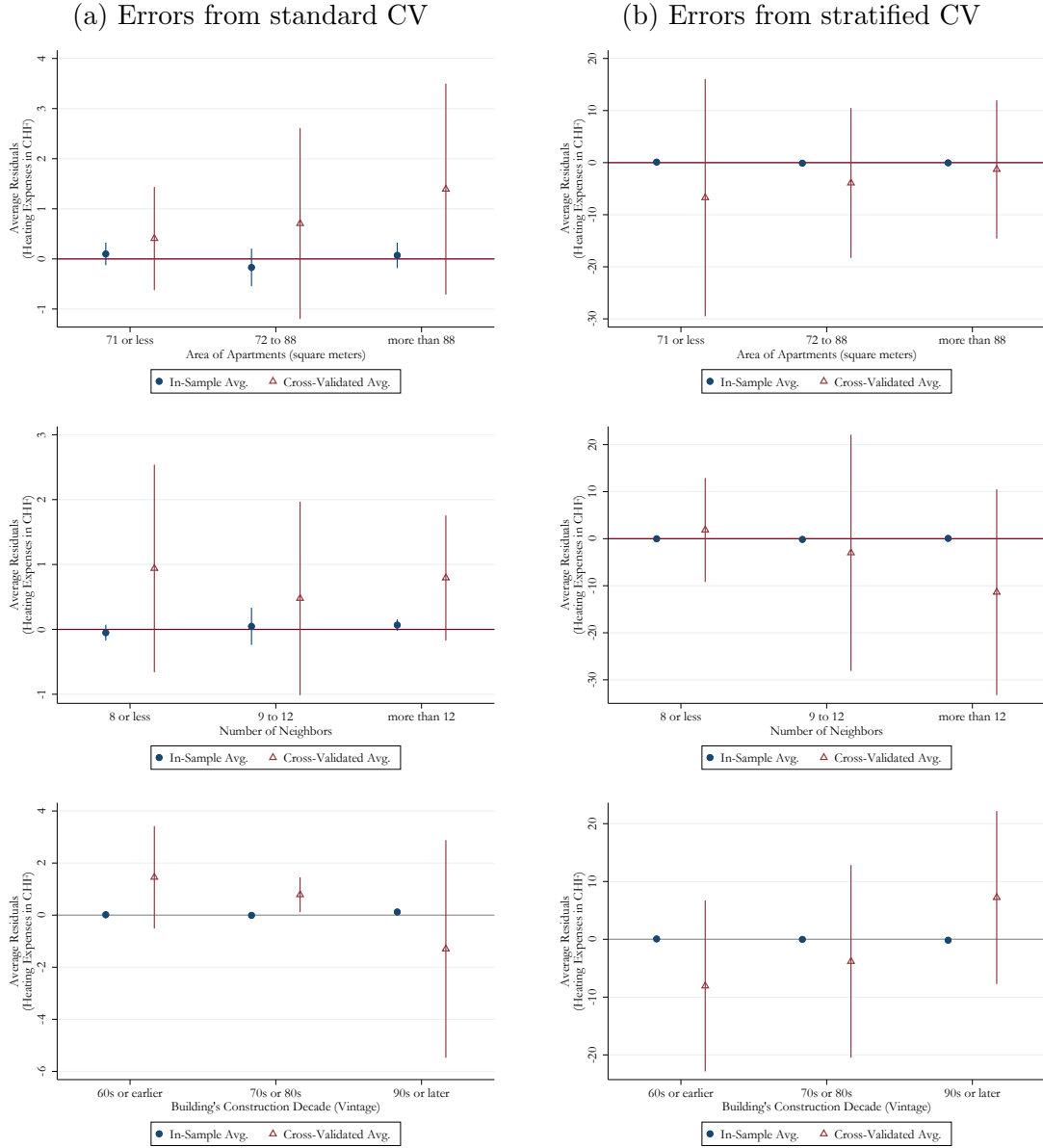


Figure D.1: ML prediction errors by covariates of interest

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.

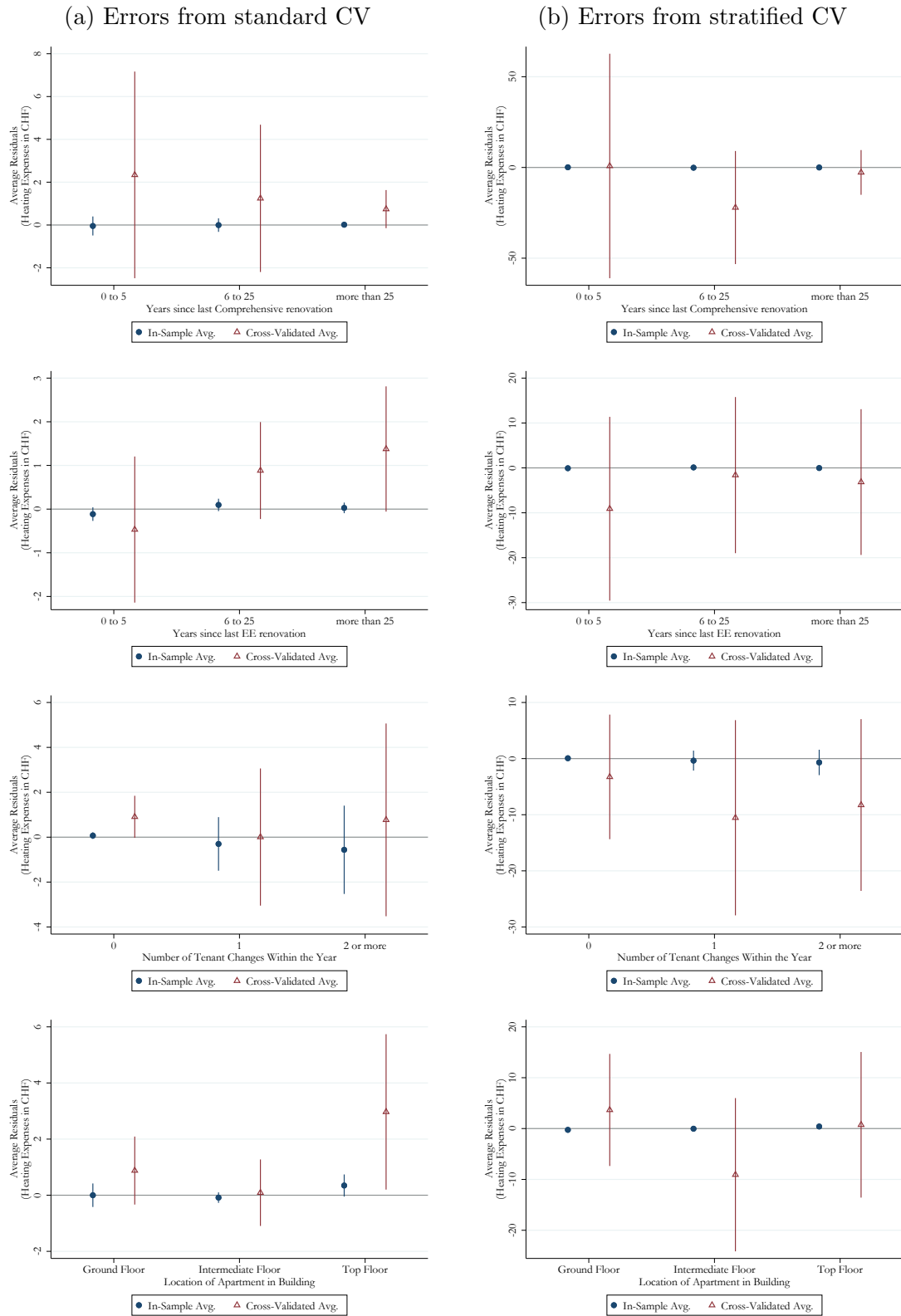


Figure D.2: ML prediction errors by other covariates

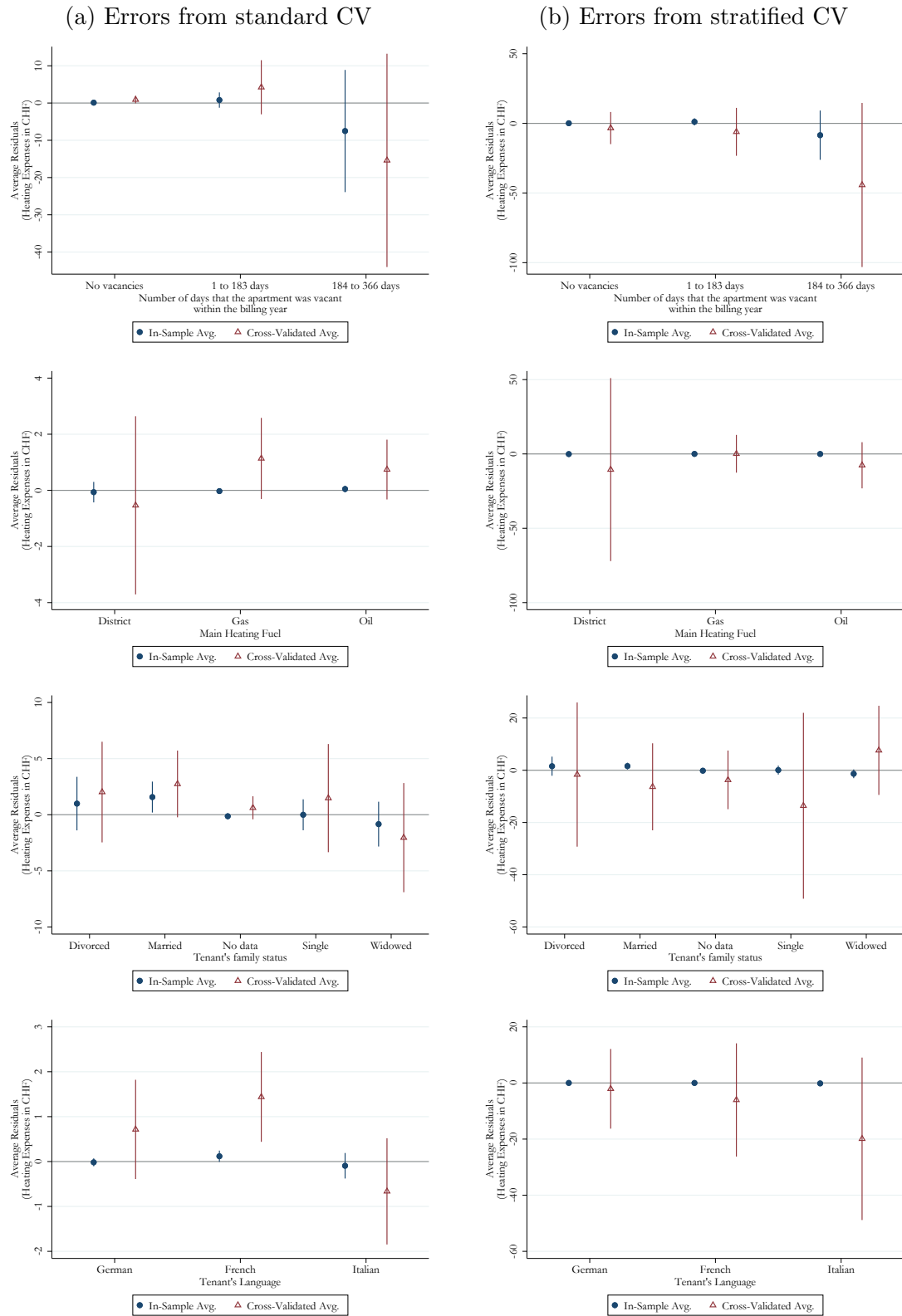


Figure D.2 (continued): ML prediction errors by other covariates

### D.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 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.

## D.4 Summarizing treatment effect heterogeneity

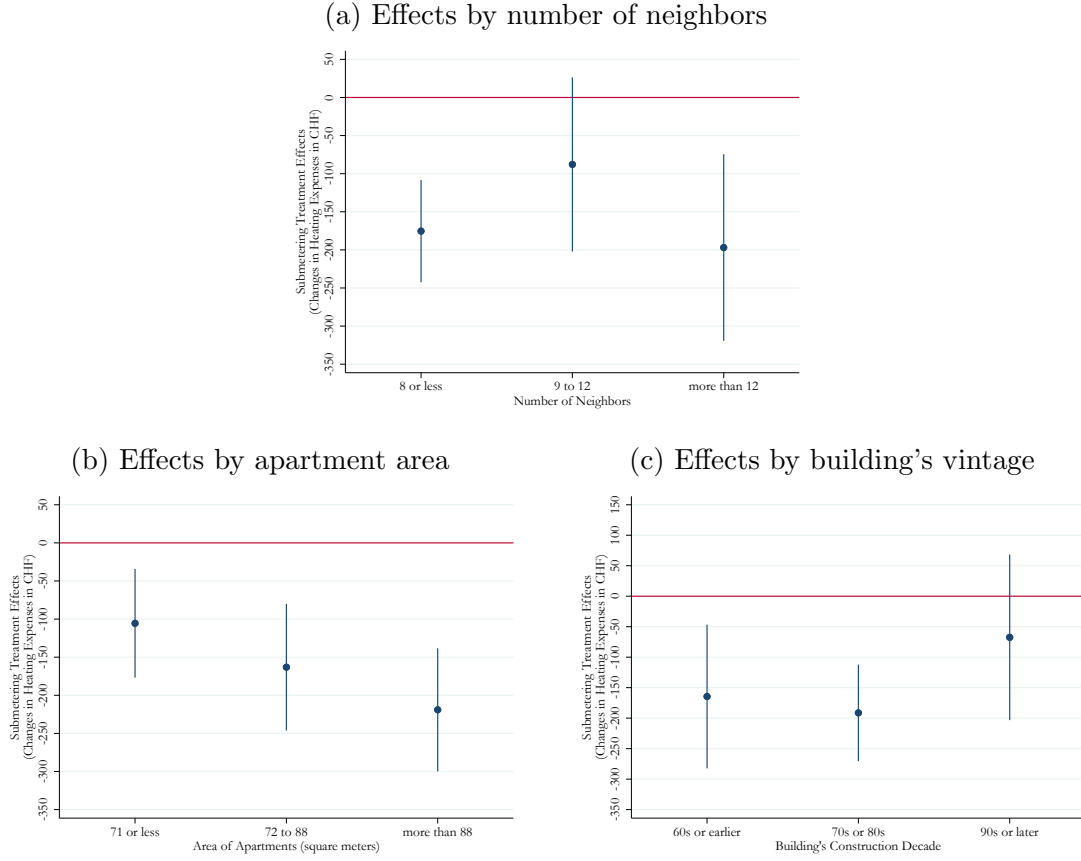
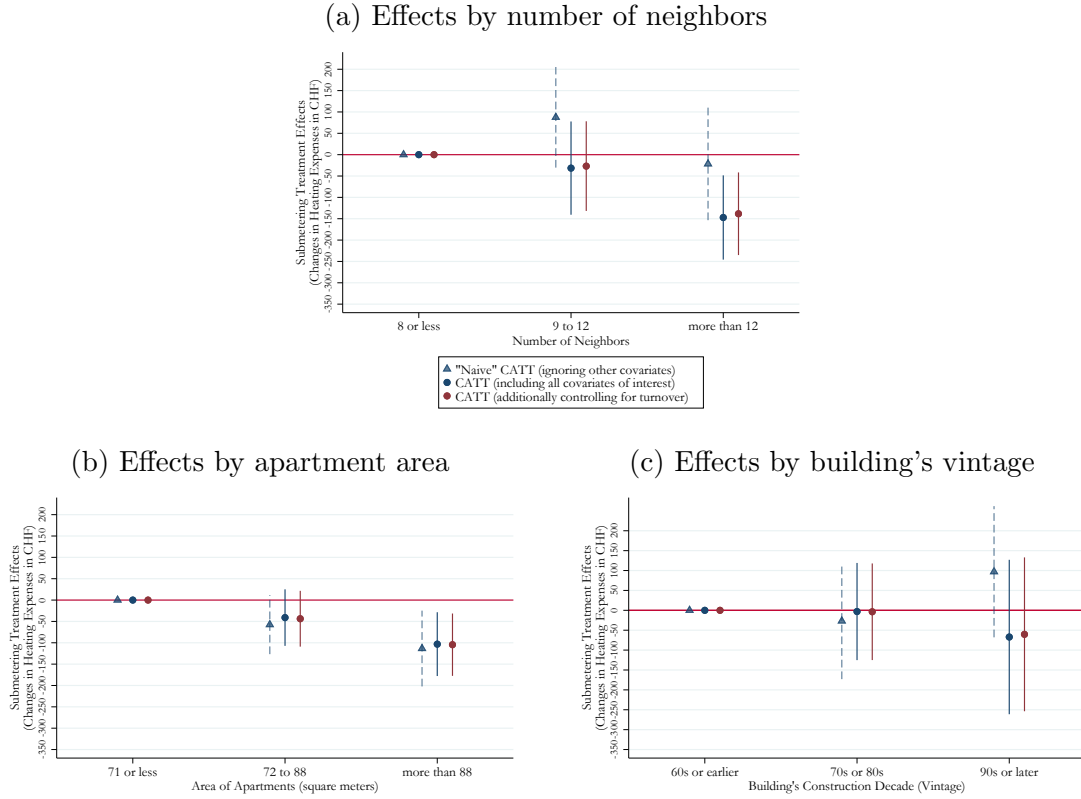


Figure D.4: ML-based “naive” estimates of submetering treatment effect heterogeneity

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.



Notes: This Figure plots  $\beta$  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. We drop buildings (less than 5%) for which vintage information was not available.

## E Additional robustness checks

### E.1 Standard event study via two-way fixed effects

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

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

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 E.1 including time-varying controls (i.e., the “saturated” specification).

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

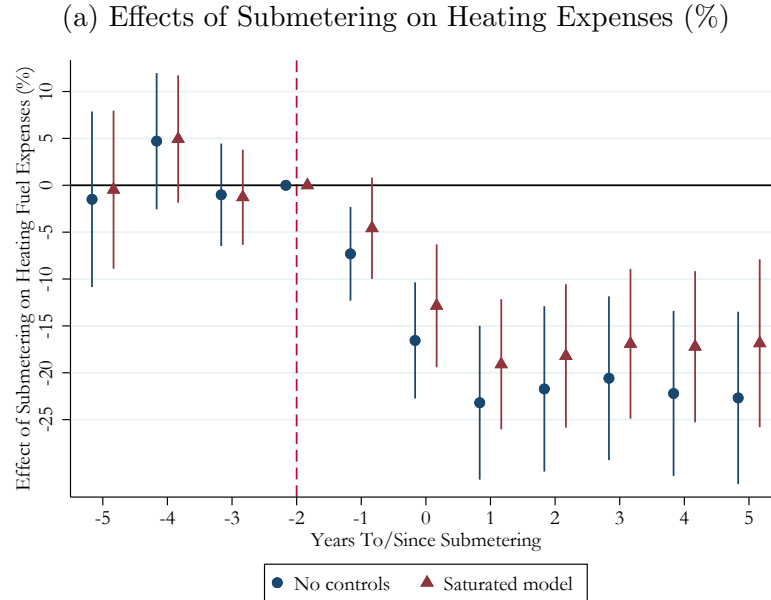


Figure E.1: Effects of Submetering According to TWFE

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.2 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 randomly 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. Accordingly, 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 E.2 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 E.3. 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.

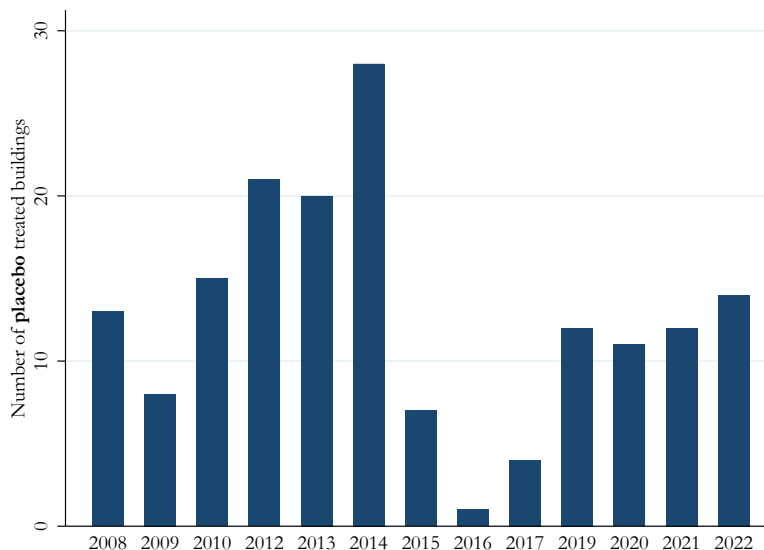


Figure E.2: Placebo treatment rollout



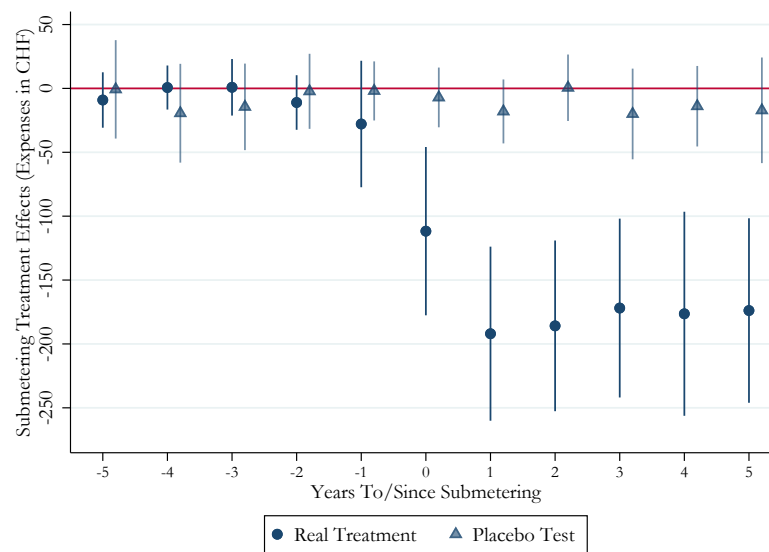


Figure E.3: Real and placebo treatment effects

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.

### E.3 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 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 E.4. 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 E.4, 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.

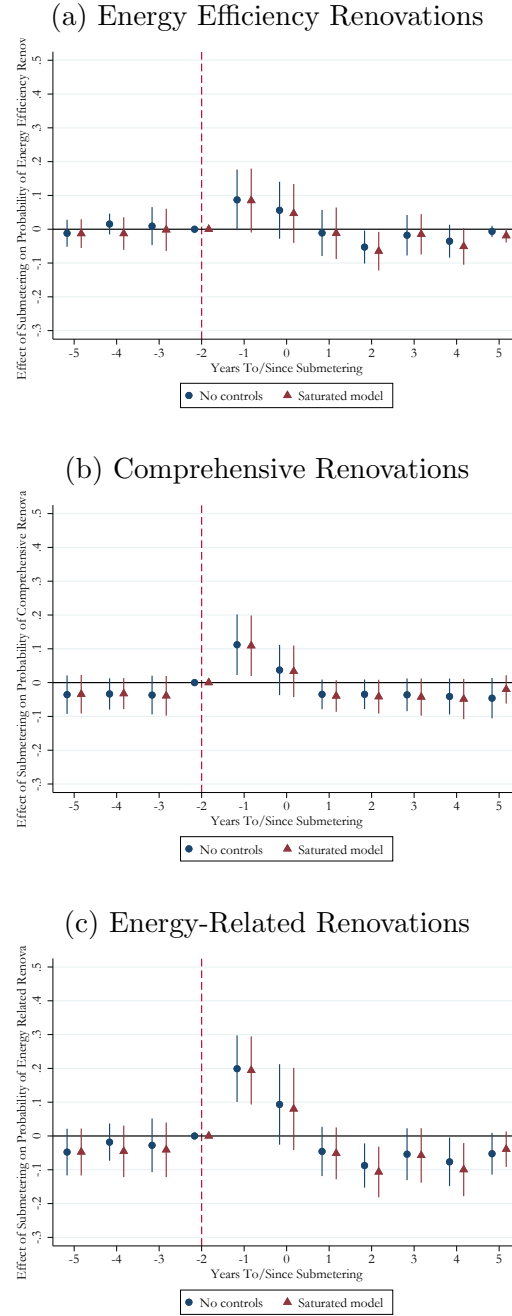


Figure E.4: Does submetering affect the probability of renovations?

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.

## E.4 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 E.1, including interactions for the confounders of interest as follows:

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

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 E.2 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 E.5 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.

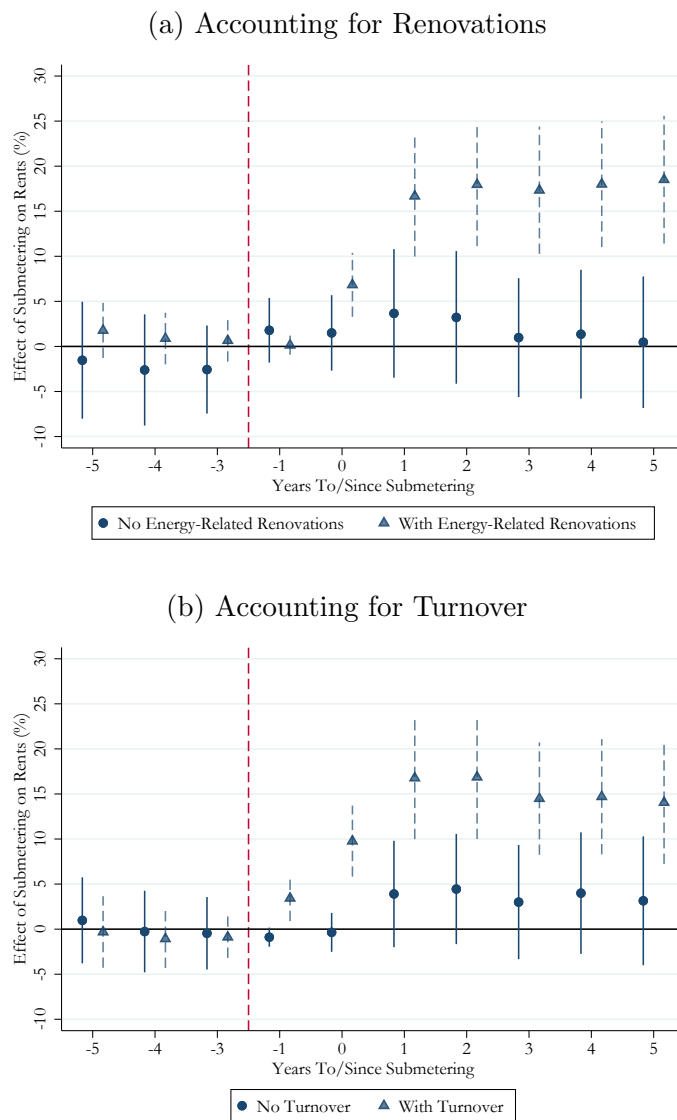


Figure E.5: Submetering Effects on Rent

Notes: This figure presents estimates  $ATT(r)$  for the effects of submetering on rents, according to specifications E.2. 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.

## E.5 Event studies with matched sample

Descriptive statistics (Appendix A) suggest that there may be observable differences between submetered versus comparison buildings. This does not necessarily bias our *ATT* estimates, as long as there is overlap in covariates and if parallel trends hold (Ghanem, Sant’Anna, and Wüthrich, 2022). While our event study results suggest that parallel trends are likely to hold, so far we still have not directly tested the degree of overlap between treated and comparison buildings. To do so, we begin by estimating a probit model for propensity of treatment, and assess resulting propensity scores across groups.

More specifically, we run the following probit model, using only pre-treatment data, aggregated to the building level:

$$Pr(\text{treat} = 1|X_i) = F(\beta_1 \text{Heat}_i + \beta_2 \text{Rent}_i + \beta_3 \text{YearBuilt}_i + \beta_4 \text{Neighbors}_i + \beta_5 \text{Area}_i + \gamma_r)$$

where  $Pr(\text{treat} = 1|X_i)$  is building  $i$ ’s probability of being submetered;  $F()$  is the cumulative distribution function of the standard normal;  $\text{Heat}_i$  are annual heating expenses;  $\text{Rent}_i$  are monthly rents;  $\text{YearBuilt}_i$  is year of construction;  $\text{Neighbors}_i$  is the number of neighbors;  $\text{Area}_i$  is apartment size; and  $\gamma_r$  are canton fixed effects.

The resulting predicted propensity scores are shown in Figure E.6. There is substantial overlap in the distributions for treatment and comparison groups. Yet, there are a few submetered buildings with high treatment propensity that may not have an appropriate comparison building. We next test whether our *ATT* estimates are sensitive to excluding those buildings. This is done via Coarsened Exact Matching (Iacus, King, and Porro, 2012), followed by difference-in-differences. We require exact matching on cantons, bins for number of neighbors, and bins for apartment area. For average rent, heating bill, and year of construction, matched buildings need to be within a bandwidth of half of the pre-treatment sample standard deviation. We search for at least 1 and no more than 5 comparison buildings for each treated unit. Note that this procedure implies dropping from the estimation sample both treated and comparison buildings for which it was not possible to find a match. As such, this is a conservative approach that favors internal validity of our estimates, sacrificing sample size and thereby statistical power. The final matched estimation sample consists of 276 buildings, 65 of which were submetered (see sample sizes in Table E.1).

The event study and *ATT* results after matching are presented in Figure E.7. These are for regressions with log energy expenses as the outcome. The effects of submetering according to this approach are remarkably similar to those estimated with the full sample. After matching and controlling for covariates (saturated model), we find that submetering reduces heating energy expenses by approximately 17.5%, on average. Results for expenses in levels (Table E.1) are also similar. Point estimates for the pre-treatment sample are not statistically significant, lending support to the parallel trends assumption. Finally, in Table E.2 we show that our matching procedure was successful in balancing groups. After matching, there are no statistically significant differences between groups in terms of observable covariates.

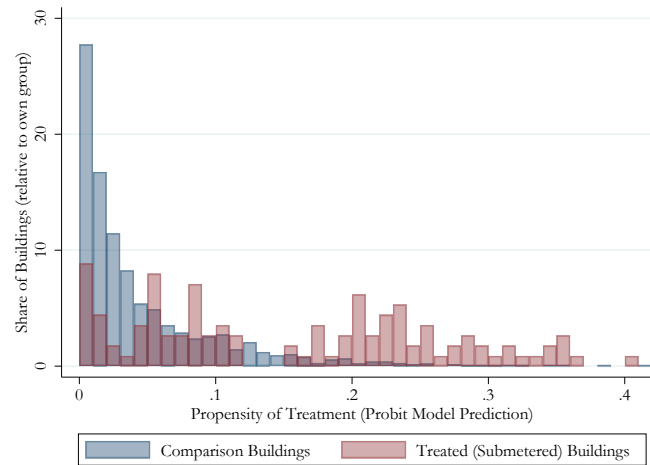


Figure E.6: Distribution of propensity scores

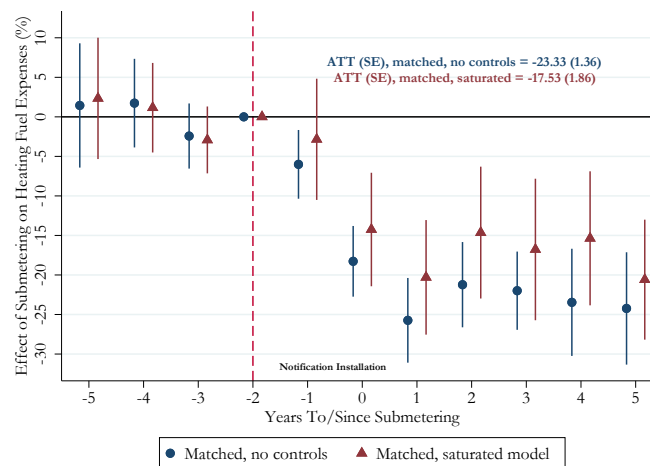


Figure E.7: Event study estimates after matching

Notes: Event study results for the effect of submetering on heating expenses in logs. Standard errors clustered at the property level.

Table E.1: Comparison of submetering effects for matched and unmatched samples

ATTr(-5)	0.13 (35.30)	19.50 (33.74)	22.21 (44.71)	28.48 (44.75)
ATTr(-4)	36.90 (30.60)	48.22 (30.27)	31.48 (32.38)	17.32 (34.97)
ATTr(-3)	-22.16 (24.48)	-20.65 (25.12)	-27.32 (24.94)	-37.13 (26.15)
ATTr(-1)	-42.30** (20.01)	-30.06 (22.32)	-48.79** (23.60)	-12.87 (45.94)
ATTr(0)	-140.79*** (20.18)	-123.26*** (22.38)	-194.10*** (22.12)	-144.35*** (44.53)
ATTr(1)	-191.34*** (29.00)	-144.05*** (27.84)	-242.12*** (31.46)	-168.93*** (50.98)
ATTr(2)	-174.65*** (27.68)	-146.94*** (24.01)	-200.65*** (32.36)	-110.90** (55.26)
ATTr(3)	-177.01*** (22.30)	-133.32*** (22.85)	-215.44*** (26.54)	-133.66** (60.59)
ATTr(4)	-196.42*** (28.98)	-148.03*** (29.58)	-216.27*** (44.33)	-122.56** (51.76)
ATTr(5)	-202.90*** (33.23)	-148.26*** (30.15)	-211.41*** (52.18)	-164.09*** (59.38)
Controls	No	Yes	No	Yes
Matched	No	No	Yes	Yes
Adjusted R-squared	.799	.848	.758	.825
Number of observations	367,230	341,775	28,172	26,775
Number of apartments	37,568	36,713	2,468	2,468
Treated apartments	1,777	1,745	564	564
Number of buildings	3,466	3,398	276	276
Treated buildings	194	192	65	65

Notes: Event study results for the effect of submetering on heating expenses in levels (CHF per year). Standard errors, in parentheses, are clustered at the property level.



Table E.2: Covariate balance after matching

	(1) Untreated	(2) Treated	(3) Difference
<i>Apartment-by-year covariates</i>			
Heating Bill (CHF per year)	1092.11 (31.89)	1148.61 (39.31)	56.50 (50.15)
Rent (CHF per month)	1197.93 (34.19)	1110.89 (68.08)	-87.04 (75.27)
Turnover Rate	0.10 (0.01)	0.12 (0.01)	0.02 (0.01)
Number of Vacancy Days	5.16 (1.95)	6.32 (2.20)	1.16 (2.92)
<i>Apartment-level covariates</i>			
Apartment Area (sq. meters)	78.51 (1.66)	78.68 (3.32)	0.17 (3.66)
Rooms per Apartment	3.45 (0.05)	3.36 (0.19)	-0.09 (0.19)
<i>Building-by-year covariates</i>			
Heating Degree Days	3164.12 (18.27)	3193.92 (64.93)	29.80 (66.53)
Energy-Related Renovation Rate	0.04 (0.01)	0.03 (0.02)	-0.01 (0.02)
<i>Building-level covariates</i>			
Number of Neighbors	8.75 (0.44)	8.69 (0.79)	-0.05 (0.90)
Years since Construction	43.90 (1.42)	43.83 (2.40)	-0.07 (2.77)
Share of District Heating	0.18 (0.05)	0.17 (0.10)	-0.02 (0.11)
Share of Gas Heating	0.39 (0.07)	0.34 (0.11)	-0.05 (0.12)
Share of Oil Heating	0.34 (0.06)	0.46 (0.13)	0.12 (0.14)
Share of Other/Unclear Heating	0.09 (0.02)	0.03 (0.02)	-0.06 (0.03)

Notes: Averages and standard errors calculated using only the matched sample. Standard errors, in parentheses, are clustered at the property level.

## F Evidence informing the discussion

### F.1 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 for estimating elasticities

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

$$\log(Y_{i,t}) = \beta_1 \log(p_{r,t,f}) + \beta_2 \log(p_{r,t,f}) \times S_i + \mathbf{X}'_{i,t} \delta + \gamma_i + \gamma_{r,t} + \varepsilon_{i,t}, \quad (\text{F.1})$$

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  $\log(p_{r,t,f})$ . 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  $\gamma_i$  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.  $\beta_1$  estimates the own-price elasticity of energy expenses for households under the common billing regime.  $\beta_2$  captures the additional price sensitivity of submetered households. We subtract 1 from those coefficients, to reflect elasticities of demand, not expenses.<sup>40</sup>

Our specification assumes that households are price takers with respect to heating energy fuels. This is plausible for Switzerland, since these fuels are often supplied by local monopolies. Also, landlords or property managers are the decision-makers with respect to energy services. Different from our analyses in the previous sections, we estimate Equation F.1 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 F.1, where we interact the log price and the submetering indicator with binned indicators for number of neighbors.

<sup>40</sup>The elasticity of energy expenses can be defined as  $\eta_{ep} = \frac{\partial ep}{\partial p} \frac{p}{ep} = (\frac{\partial e}{\partial p} p + \frac{\partial p}{\partial p} e) \cdot \frac{1}{e} = \frac{\partial e}{\partial p} \cdot \frac{p}{e} + 1 = \eta_e + 1$ .

## Results for elasticities

Table F.1 reports our elasticity estimates. Standard errors (in parentheses) are clustered at the building level. Column (1) reports results for the model in Equation F.1. We find a price elasticity of  $-0.619$  without submetering, and  $-0.716$  with submetering.<sup>41</sup> 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 F.1 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 the price incentive of households under submetering does not depend on the number of neighbors.

## F.2 Survey evidence

We provide new survey evidence to inform the behavioral mechanisms behind our empirical findings. Again partnering 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 unaware of 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 F.1 shows that unawareness of the common heating bill is similar in small/medium ( $N \leq 12$ ) and large ( $N > 12$ ) buildings. These results suggest that unawareness of the common heating bill is highly prevalent in our sample, but unlikely to explain the higher share of free-riders in large buildings.

---

<sup>41</sup>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.

Table F.1: 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 F.1, 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 40). Standard errors (in parentheses) are clustered at the building level.

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 F.2 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 co-operation (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 F.3 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 scheduling in energy consumption (Ito, 2014; Shaffer, 2020; Ito and Zhang, 2023). We document that unawareness of 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.

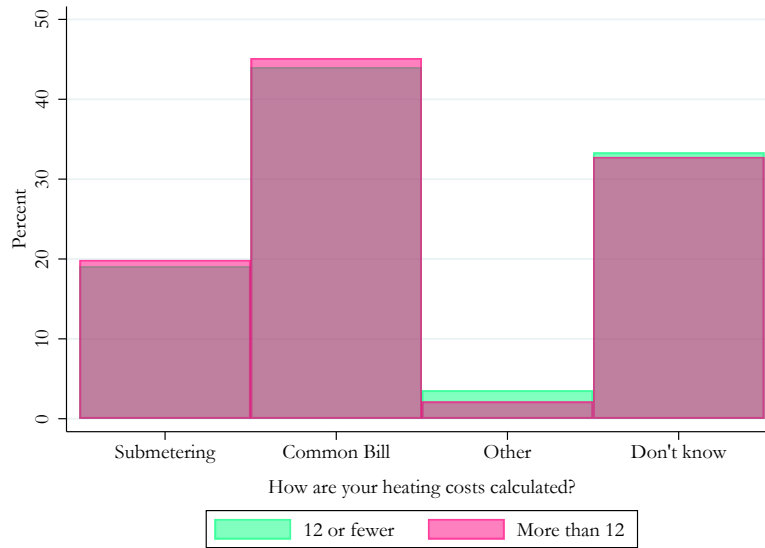


Figure F.1: Beliefs about Billing Method

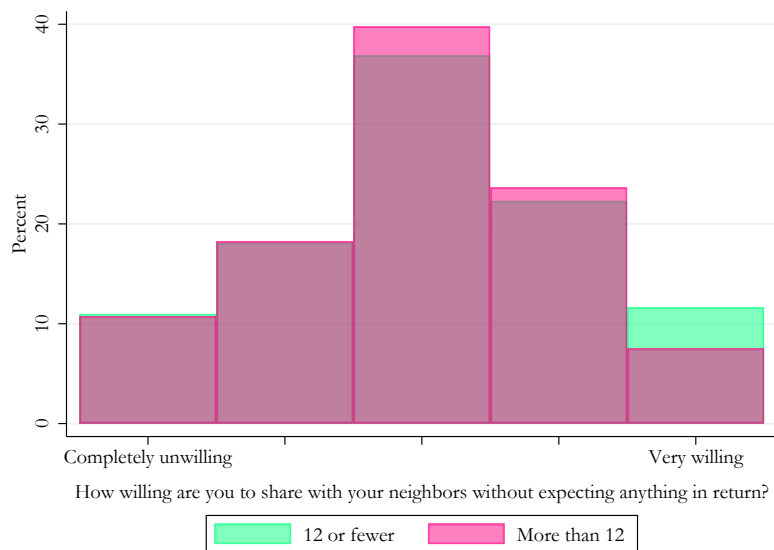


Figure F.2: Altruism toward Neighbors

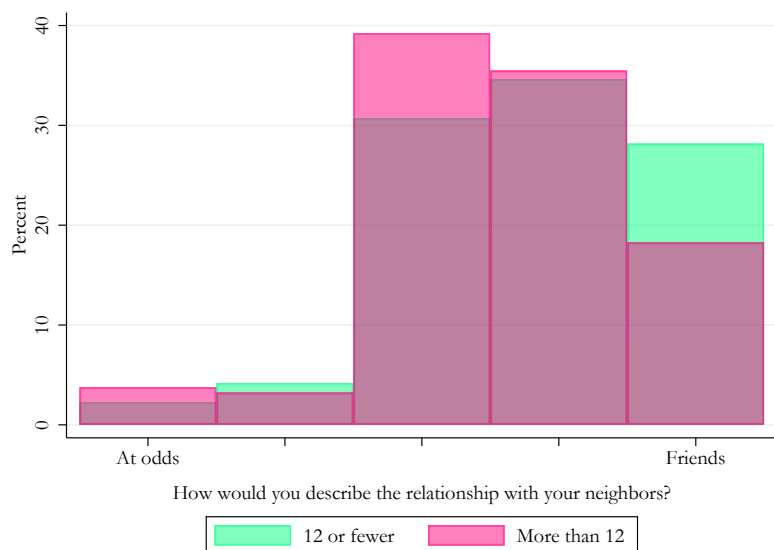


Figure F.3: Relationship with Neighbors