The Tragedy of the Common Heating Bill

Harald Mayr[a] and Mateus Souza[b]

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

November 30, 2023

Latest version available here.

Abstract

We study the energy pricing distortion that arises when apartment buildings share a common heating bill, resembling a tragedy of the commons. Submetering corrects the distortion, but does it improve welfare? We propose a theoretical framework with shared billing and environmental externalities to derive a sufficient statistic for the welfare effect of submetering. We then analyze the staggered introduction of submetering in 265 apartment buildings in Switzerland. Event study estimates show that submetering reduces annual heating expenses by 9.6 percent, on average. Under benchmark assumptions, this average effect implies a welfare improvement only if the social cost of carbon is 228 CHF/tCO_2 or larger. However, machine learning estimates reveal substantial heterogeneity in treatment and welfare effects, suggesting a role for targeted interventions in this context.

Keywords: Submetering, individual billing, heating energy, tragedy of the commons, welfare

JEL Codes: D61, Q41, Q52

1 Introduction

The building sector is increasingly covered by carbon pricing policies. As almost 26% of global energy-related emissions are due to the operation of buildings (IEA, 2023), this sector has been earmarked as having great potential for abatement. The new EU Emissions Trading System for example, will cover emissions from buildings (European Comission, 2021). However, even if carbon pricing for buildings becomes widespread, inefficiencies remain due to other distortions of heating energy prices.[1] We study one such distortion: the common heating bill.

In many apartment buildings, tenants share a common heating bill. Each household's share of the bill is determined by a fixed key, often the area or volume of the apartment. This situation resembles the tragedy of the commons (Hardin, 1968). In contrast to the original allegory of herdsmen grazing their cattle on a common pasture, successful governance of the common heating bill is unlikely because households cannot monitor and sanction each other (Ostrom, 1990). Submetering[2] promises an immediate solution with potentially large welfare gains. Studies on electricity (Elinder et al., 2017), water (Ornaghi and Tonin, 2021), and hot water (Elinder et al., 2021) emphasize this promise with large reductions in consumption—between 22% and 36%. Whether these effect sizes apply to heating energy is an open question.

Residential heating is different from other utilities in at least three important aspects. First, heating likely exhibits satiation because excessive heating is uncomfortable. If households do not heat excessively in the common heating bill, submetering may not change consumption much. Second, heat flows between apartments can be substantial. Submetering may be less effective if heat can be "stolen". Third, effect heterogeneity is important,

[1] Distortions in retail energy prices might arise due to pre-existing taxes/subsidies, rigid regulatory pricing, or market power, for example. See Borenstein and Bushnell (2022) for estimates of the size these distortions for the case of the US.

[2] 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".

as buildings differ in energy efficiency. It is unlikely that the cost of submetering is worthwhile in the most energy efficient buildings. With all these caveats, the welfare effect of submetering on heating energy consumption is an open question.

Evidence on the effect of submetering on heating energy consumption is 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). Ito and Zhang (2022) study the introduction of a two-part heating tariff in China, where households can opt out of individual billing, and the pre-policy bill is entirely independent of energy consumption. This particular policy reduced heating energy consumption by 10%. To the best of our knowledge, we provide first quasi-experimental evidence on the switch from a common (i.e., building-level) heating bill to individual billing.

Effect heterogeneity is a crucial aspect of individual heat metering. The policy debate on individual heat metering is particularly active in the European Union (Canale et al., 2019), Russia (Korppoo and Korobova, 2012), and China (Lu et al., 2014). Policies vary widely, with some legislations mandating submetering for all apartment buildings, only a subset of buildings, or none at all. The Swiss context illustrates that policy makers seriously consider heterogenous treatment and welfare effects (Schweizerischer Verband für Energie- und Wasserkostenabrechnung, 2023). Submetering is mandated on the federal level, but various exemptions apply at the cantonal level. Exemptions include the existing building stock, buildings with high energy efficiency, and buildings with a small number of apartments. In all these cases, the cost of submetering is deemed too high relative to the expected energy savings. To base such policies on evidence, we aim to inform policy makers on the heterogeneous effects of submetering on heating energy consumption and welfare.

Our empirical strategy exploits the staggered introduction of submetering for heating in 265 apartment buildings in Switzerland between 2006 and 2021. Including treated and comparison groups, our sample includes 3,364 buildings with a total of 31,591 apartments. We implement event study regressions using the heterogeneity-robust estimator from Sun

and Abraham (2021). Our data includes two essential features that were mostly unavailable in previous studies on individual billing of heating energy. We observe heating energy consumption before the installation of submeters, which allows us to capture potential effects of announcement or installation. Moreover, we observe various types of renovations, document how they correlate with the introduction of submetering, and control for their effect on energy consumption.

We find that submetering reduces heating energy expenses by 9.6 %, on average. We find clear anticipation effects, as a large part of the decrease in energy consumption happens in the year of the installation of submeters, but before individual billing starts. We find that the effect size is slightly larger (11.8 %, albeit not statistically different from our main estimate) when we do not control for renovations and other potential confounders. In all specifications, the effects are persistent at least until the fourth year after installation.

We develop a theoretical framework to assess the welfare effects of submetering. Our framework incorporates two externalities: (i) the direct externality from the fact that, under shared billing, each household's consumption choice impacts all neighbors' heating bills; and (ii) the environmental externality from heating energy. The latter is important, as heating accounts for 80% of direct CO_2 emissions in the building sector (IEA, 2022). We show that submetering eliminates the direct externality while simultaneously reducing the environmental externality. The theoretical framework yields a sufficient statistic to evaluate the welfare effect of submetering.

The sufficient statistic derived in the theoretical framework allows us to assess the welfare effects of submetering in our empirical setting. In our sample, we find that submetering leads to an overall welfare improvement only under a relatively large social cost of carbon (SCC). Under our preferred set of assumptions, individual billing increases overall welfare only if the SCC is 228 Swiss Francs (CHF) per tonne of CO_2 or larger.[3] This estimate, however, is

[3]The mean SCC estimate from Rennert et al. (2022), of 185 USD/tCO_2, may be used as a benchmark for comparison. During the time period we study, the USD-CHF exchange rate between fluctuated around 1. At the time of writing this manuscript, one Swiss Franc is approximately equivalent to 1.1 US dollars.

sensitive to energy prices and the cost of submetering. Moreover, our theoretical framework suggests that there is reason to expect variation in welfare effects across buildings.

This motivates our analysis of heterogeneity in treatment and welfare effects. We use machine learning (ML) to estimate the *conditional* average treatment effects of submetering, while accounting for several confounding factors, such as renovations. Consistent with our framework, we find that submetering effects are more pronounced for apartments with a larger number of neighbors. This could be explained by the relatively sharper increase in implicit energy prices that these apartments face after submetering.

We use our ML approach to estimate *building-specific* welfare effects of submetering. This exercise allows us to identify the set of buildings in our sample for which submetering was cost-effective. We find that an SCC of 185 CHF/tCO_2 is sufficient to improve welfare in 45 percent of the buildings in our sample. We therefore argue that targeted, rather than uniform, policies in this context can lead to substantial cost-effective abatement. For example, our results suggest that more focus should be placed on the relatively larger apartment buildings.

Our study adds 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; Christensen et al., 2022; Gerarden and Yang, 2023). Our results also speak to 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 et al., 2017). Submetering can be viewed as an energy efficiency investment that could be profitably adopted in many buildings. Our main finding—that in most buildings, submetering only improves welfare with a positive SCC—suggests that there is no sizable energy efficiency gap with respect to submetering.

The remainder of this paper is organized as follows. Section 2 presents a theoretical framework to assess the welfare effects of submetering. Section 3 describes the data and

empirical strategy. Section 4 presents average treatment effects and Section 5 investigates effect heterogeneity. Section 6 discusses welfare implications and Section 7 concludes.

2 Theoretical framework

We propose a straightforward model incorporating the main elements necessary to illustrate the tragedy of the common heating bill. Let there be N households that share their heating bill (e.g., suppose that these households 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

$$U_i(e_i, c_i) = u_i(e_i) + c_i \ . \tag{1}$$

We specify $u_i(e_i)$ assuming that households do not derive utility from energy consumption per se, but from indoor temperature. Household i prefers temperature T_i and suffers disutility from the square of temperature deviations. The parameter κ_i captures how uncomfortable deviations from preferred temperature are for household i

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

The actual indoor temperature $T(e_i)$ is a function of energy consumption. We use the simplified heat loss formula $e_i = (T - T_0)H$, where T_0 is the outside temperature and H measures the building's heat loss.[4] We therefore obtain

$$U_i(e_i, c_i) = -\kappa_i(T_i - T_0 - \frac{e_i}{H})^2 + c_i \ . \tag{3}$$

[4]Heating energy can be written as $e_i = H(T - T_0) + S$, where H is the heat loss per degree of temperature difference (an inverse measure of energy efficiency that captures the thermal resistance of walls and windows, losses at thermal bridges, as well as ventilation losses), 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 and solar heat gains and other internal heat gains.

Each household faces a budget constraint that depends on its income y_i, the price of energy p, consumption of the numeraire, and the metering regime. A share $\theta \in [0, 1]$ of heating consumption is individually billed, while the remainder is equally split among N households. Hence, the budget constraint also depends on the average heating energy consumption of all neighbors

$$y_i \geq \theta e_i p + (1 - \theta)\frac{\sum^j e_j p}{N} + c_i \ . \tag{4}$$

Private solution

Note that the budget constraint in Equation 4 implies an externality if $\theta < 1$. The household considers only the effect of e_i on its own utility bill, but not the effect on its neighbors' bills. In particular, the household faces a price incentive

$$\tilde{p} = p\left(\theta + \frac{1 - \theta}{N}\right) \ . \tag{5}$$

With submetering and billing, $\theta = 1$, the price incentive is undistorted and $\tilde{p} = p$. When households' heating bills are fully shared, $\theta = 0$ and the price incentive is distorted to $\tilde{p} = \frac{p}{N}$. This distortion is particularly severe if N is large, i.e., the heating bill is shared among many households.

Combining Equations 3 and 4 (assuming the budget constraint holds with equality), the household chooses e_i to maximize

$$U_i(e_i) = -\kappa_i(T_i - T_0 - \frac{e_i}{H})^2 + y_i - \theta e_i p - (1 - \theta)\frac{\sum e_j p}{N} \ . \tag{6}$$

The solution to the household's problem is

$$e^*_{i,\theta} = (T_i - T_0)H - \frac{H^2}{2\kappa_i}\tilde{p} \ . \tag{7}$$

Therefore, the household's choice of energy consumption $e^*_{i,\theta}$ depends on the price incentive implied by θ, and heating requirements. Unsurprisingly, energy consumption is larger for high preferred indoor temperature T_i, low outside temperature T_0, and poor energy efficiency (i.e., high H). The household is more price sensitive if κ_i is small and H is high. In the absence of a price incentive ($\tilde{p} = 0$), the household chooses e^*_i such that $T = T_i$.

The demand elasticity is

$$|\varepsilon_{i,\theta}| = |\frac{\partial e^*_{i,\theta}}{\partial p}\frac{p}{e^*_{i,\theta}}| = \frac{H^2}{2\kappa_i}\left(\theta + \frac{1-\theta}{N}\right)\frac{p}{e^*_{i,\theta}} \ . \tag{8}$$

Submetering

Submetering increases θ from 0 to $\bar{\theta}$. The price incentive increases from $\frac{p}{N}$ to p if $\bar{\theta} = 1$.[5] In response to submetering, household i chooses a lower $e^*_{i,\theta=1}$:

$$\Delta e^*_i = e^*_{i,\theta=1} - e^*_{i,\theta=0} = -\frac{H^2}{2\kappa_i}\frac{N-1}{N}p \ . \tag{9}$$

This implies that the effect of submetering on energy consumption is larger with poor energy efficiency (i.e., high H), flexible temperature preferences (i.e., low κ_i), large N, and a high energy price p.

The effect on utility of household i is ambiguous. Utility may increase or decrease, depending on the effect of submetering on the household's energy expenses.

$$\begin{aligned}\Delta U_i &= U_i(e^*_{i,\theta=1}) - U_i(e^*_{i,\theta=0}) \\ &= \left(\frac{p^2}{N^2} - p^2\right)\frac{H^2}{4\kappa_i} + \left(\frac{\sum^j T_j}{N} - T_i\right)Hp \\ &+ p^2\frac{H^2}{2\kappa_i} - \frac{p^2}{N^2}\frac{H^2}{2}\sum^j\frac{1}{\kappa_j}\end{aligned} \tag{10}$$

[5] In practice, $\bar{\theta}$ is lower than 1, to account for fixed heating costs like service fees, heat losses in pipes, or common heated areas. Appendix A.1 describes the conditions under which the simplified model with $\bar{\theta} = 1$ is equivalent to a more complex model with fixed heating costs.

This result has an intuitive interpretation. The first term captures the negative effect on utility from lower indoor temperature. The second term reflects distributional consequences resulting from preferences for indoor temperature. Household i will benefit from submetering if it prefers a relatively low temperature, as compared to its neighbors. This term is zero for the average household. The remaining terms represent gains from removing the price distortion. Overall, ΔU_i can be positive or negative depending on households' preferences and the size of the price distortion, i.e., there can be winners and losers from submetering.

This framework suggests that submetering increases the demand elasticity, as shown by equation 11 below.

$$\begin{aligned} \Delta\varepsilon_i &= |\varepsilon_{i,\theta=\bar{\theta}}| - |\varepsilon_{i,\theta=0}| \\ &= \frac{H^2}{2\kappa_i}\left(\frac{p}{e^*_{i,\theta=1}} - \frac{1}{N}\frac{p}{e^*_{i,\theta=0}}\right) > 0 \end{aligned} \tag{11}$$

Welfare

Submetering is the preferred policy if it improves welfare. In addition to households' private utility, we consider an environmental externality from energy consumption, ϕ, as well as the per-household cost of submetering s. Consequently, welfare is given by

$$W = \sum^{i} U_i(e^*_{i,\theta}) - \phi \sum^{i} e^*_{i,\theta} - \mathbb{1}_{\theta>0} sN \ . \tag{12}$$

The welfare effect of submetering is

$$\Delta W = \sum^{i} \Delta U_i - \phi \sum^{i} \Delta e^*_i - sN \ . \tag{13}$$

This equation can be expressed in terms of empirical estimates. We estimate effects on energy expenses. Using $\tau = \sum^i \Delta e_i^* p$, we get

$$\Delta W = -\tau \frac{N-1}{2N} - \tau \frac{\phi}{p} - sN \ . \tag{14}$$

Building-level treatment effects τ, together with estimates for the cost of individual billing and the social cost of carbon are sufficient to estimate ΔW. We provide welfare estimates for different scenarios in Section 6.

Figure 1 illustrates the welfare loss from shared metering. 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., the point 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 individual billing, the household also internalizes the environmental externality ϕ, it consumes the social optimum $e^*_{i,S}$, i.e., the point on the upper left. The central insight from this illustration is that submetering partially addresses welfare losses due to the environmental externality. The welfare gain from submetering, without internalizing the environmental externality, is represented by the sum of areas A_i and B_i.

3 Setting and data

We use data from a large real estate management company in Switzerland. Our dataset includes household-level utility bills for 3,364 buildings with a total of 31,591 apartments, from the year 2006 to 2022.[6] We can identify apartments in each building. The data include yearly heating and auxiliary expenses that are billed by the real estate management company. Expenses include heating fuels, water and sewage bills, as well as various service fees.

[6]These numbers already represent a restricted sample for which we drop "always-treated" apartments (i.e., those that were already submetered in the first year we observe them). The earliest treated buildings in our regression samples were submetered in 2007, and the latest were submetered in 2021.

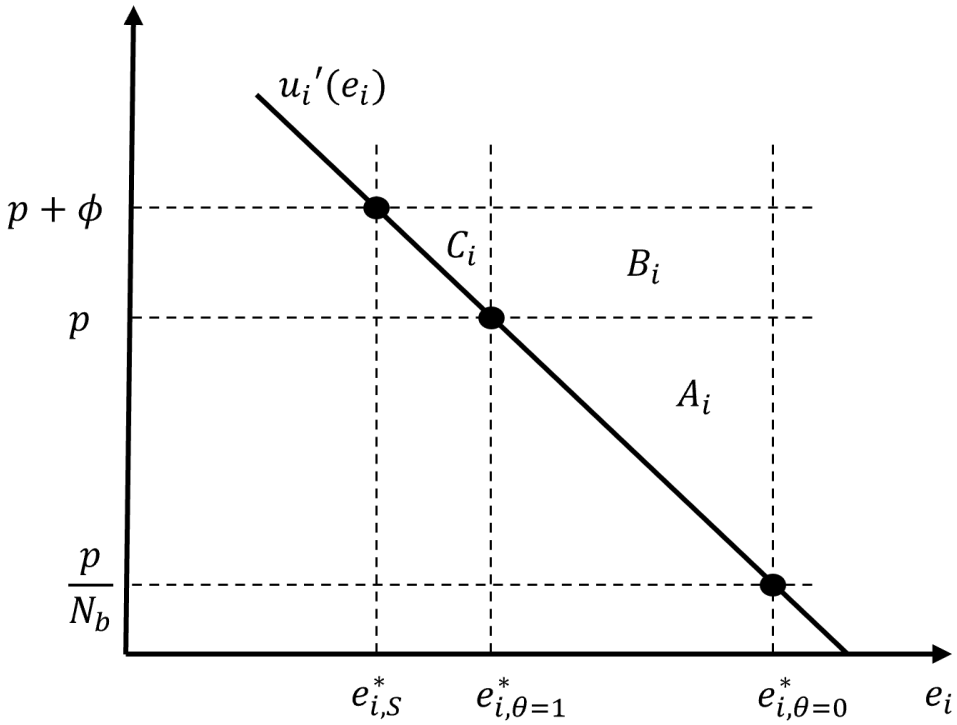


Figure 1: Private solutions and social optimum

For each household in our sample, we observe the yearly heating energy bill in Swiss Francs (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 or volume 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 971 in a year.

We infer submetering from a particular expense type called "Cost of metering/billing services." This expense type reflects service charges for submeters or heat cost allocators that are paid by tenants. Out of the 3,364 buildings that we observe, 265 received submetering during our sample period. Figure 2 shows the staggered rollout of submetering over the years in our sample.

In practice, submetering is often introduced with other measures. This can include the installation of thermostatic valves, the replacement of the heating system, or the insulation of the building's facade, among other improvements. We observe renovations in our sample.

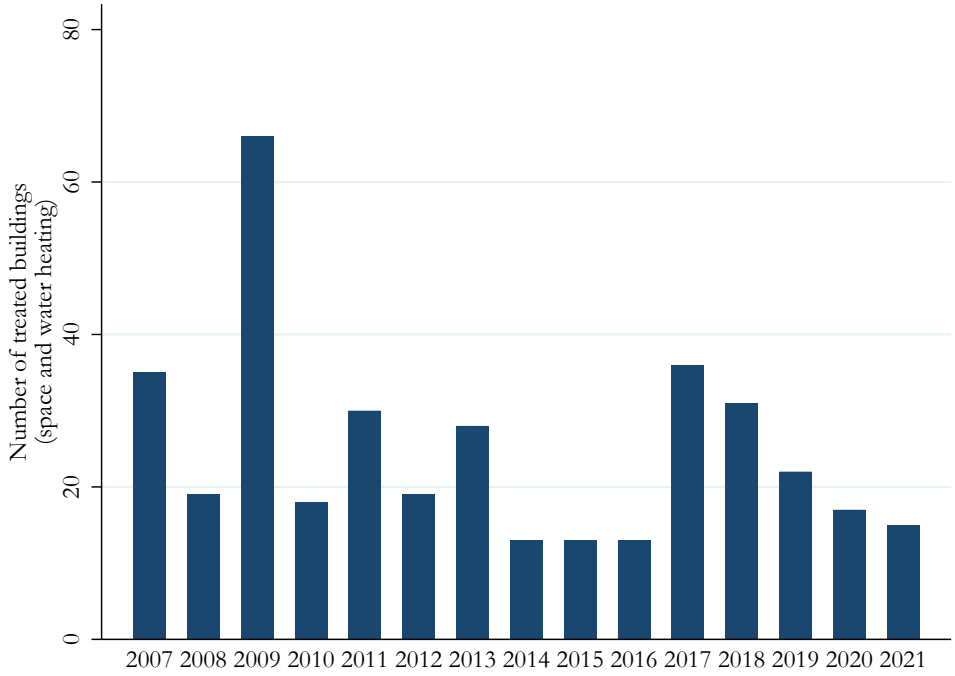


Figure 2: Year of submeter installation

For each renovation, we know the respective year and a short text description. We use the text descriptions to categorize renovations. This information allows us to control for the effect of renovations on energy consumption.

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

4 Average effects of submetering

4.1 Two-way fixed effects

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

$$\begin{aligned} ATT &= \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0) | S_i = 1; Post_t = 1] \ , \\ ATT(r) &= \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0) | S_i = 1; Post_t = 1; t - q_i = r], \text{ for } r \geq 0 \ , \end{aligned}$$

where $Y(1)_{i,t}$ and $Y(0)_{i,t}$ are potential outcomes for apartment i in year t under the treated (individually metered) and untreated scenarios, respectively; 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 submetering starts in apartment i, such that r represents years since the start of submetering.

To estimate these parameters, 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. For the ATT, we can consider a two-way fixed effects (TWFE) regression, as follows:

$$Y_{i,t} = \beta S_i \times Post_t + \gamma_i + \gamma_t + \varepsilon_{i,t} \ , \tag{15}$$

where $Y_{i,t}$ is the observed outcome of interest; the interaction term $S_i \times Post_t$ is equal to one for individually metered apartments, only for years after the start of submetering (i.e., $t \geq q_i$), zero otherwise; γ_i are apartment fixed effects; γ_t are year fixed effects; and $\varepsilon_{i,t}$ is an idiosyncratic error term.

As outcome variables, we focus on heating expenses, measured in Swiss francs (CHF), as well as the log transformation of those expenses. The coefficient estimate $\hat{\beta}$ can thus be interpreted as the average change in heating expenses caused by submetering, under the parallel trends assumption (Angrist and Pischke, 2009). That is, the outcomes for both the individually metered and the comparison apartments would have followed parallel paths in the absence of treatment. Note that, if this assumption holds, our empirical strategy does not require quasi-random assignment of treatment dates. Nevertheless, in Appendix Table B.1 we show outcome and covariate averages across treatment cohorts, and find no systematic differences in patterns when comparing early- versus late-treated buildings.

In order to assess pre-treatment trends and to estimate $ATT(r)$, we implement the following event study specification:

$$Y_{i,t} = \sum_{r \neq -2} \beta_r S_i \times \mathbb{1}[r = t - (q_i)] + \gamma_i + \gamma_t + \varepsilon_{i,t} \ , \tag{16}$$

where S_i is the submetering indicator defined as above; $\mathbb{1}[r = t - q_i]$ are indicators for years relative to the treatment dates q_i; and the other parameters are defined as above. More specifically, q_i represents the year in which the submetering technology was installed. However, submetered bills are only charged one 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 submetering (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. Similarly, the never-treated apartments are set at $r = -2$.

For $r \geq -1$, the estimated coefficients $\hat{\beta}_r$ can be interpreted as the post-treatment coefficients. These will also provide insight on whether the effects of submetering change over time. With the specification above, we can also assess the parallel trends assumption by analyzing the pre-treatment coefficients (with $r < -2$). For inference, we cluster standard errors at the building level, which is effectively the level at which treatment occurs.

We also implement variants of specifications (15) and (16) including time-varying controls. Specifically, we add: a continuous variable for the share of months in a given year that an apartment was vacant; a variant of this vacancy share that is aggregated at the building level; heating degree days (HDD) and squared HDD; energy prices (heating gas and oil prices by NUTS2 region and year, district heating and other energy sources by year) and squared energy prices; dummies for the number of tenant changes; and indicators for renovation and construction years. 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 conditional parallel trends assumptions.

4.2 Heterogeneity-robust specifications

Note that standard TWFE specifications, as in equations (15) and (16), will correctly estimate *ATT* only under an assumption of *homogeneous treatment effects*. This is one of the insights from recent econometric literature on difference-in-differences (e.g., Borusyak et al., 2021; Goodman-Bacon, 2021). When treatment effects are expected to be heterogeneous (e.g., changing across apartments or over time), then standard TWFE can be biased. For staggered rollout designs, in particular, standard TWFE can lead to near-term bias (Goodman-Bacon, 2021). This is because standard regressions place more weight in sample regions with higher treatment variance (i.e., typically where there is more switching from an untreated to a treated status).

To take this into account, we implement the alternative estimator from Sun and Abraham (2021). 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:

$$Y_{i,t} = \sum_{r \neq -1} \sum_{c=2007}^{2022} \beta_{r,c} S_i \times \mathbb{1}[r = t - (q_i)] \times \mathbb{1}[q_i = c] + \gamma_i + \gamma_t + \varepsilon_{i,t} \ , \tag{17}$$

where $\mathbb{1}[q_i = c]$ are the cohort c indicators, and the other terms are defined as above. Note that specification (17) estimates separate results $\hat{\beta}_{r,c}$ for each relative time 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. This results in heterogeneity-robust estimates of both ATT and $ATT(r)$.

4.3 Results for ATT and $ATT(r)$

Figure 3 presents results for the heterogeneity-robust specification (16). The outcome is annual heating expenses measured in Swiss Francs (CHF). Blue dots represent point estimates of $ATT(r)$ for the model with no controls other than year and apartment fixed effects. Red triangles are point estimates for the saturated model with time-varying controls. Visual inspection of the pre-treatment coefficients ($r < -2$) in both panels suggests that pre-trends are limited. None of the pre-treatment coefficients are statistically significant. This supports the parallel trends assumption and causal interpretation of the post-treatment coefficients.

The post-treatment coefficients suggest a drop in annual heating expenses starting already in $r = -1$. The implication is that households may have reacted to the mere announcement of submetering. Announcement effects are consistent with the results in Elinder et al. (2017). The drop in energy expenses becomes stronger in the year of submeter installation ($r = 1$) and increases further in the first year of individual billing ($r = 1$). In $r = 1$, energy expenses decrease by 110 CHF. The effects for subsequent years are slightly smaller, yet statistically indistinguishable from that in $r = 1$. For our estimates of ATT, we pool the effects from periods from $r = 1$ to $r = 4$, which we consider to be more "stable" effects after households have already fully-adjusted to individual billing. The resulting pooled ATT represents a CHF 85 reduction in energy expenses in the specification without controls, and a CHF 85 reduction in the saturated model.

Figure 4 presents estimates with the log of heating as outcome variable. These results are qualitatively similar to those in Figure 3. Overall, for the period from $r = 1$ to $r = 4$, we estimate a 11.8 % reduction in energy expenses in the specification without controls, and

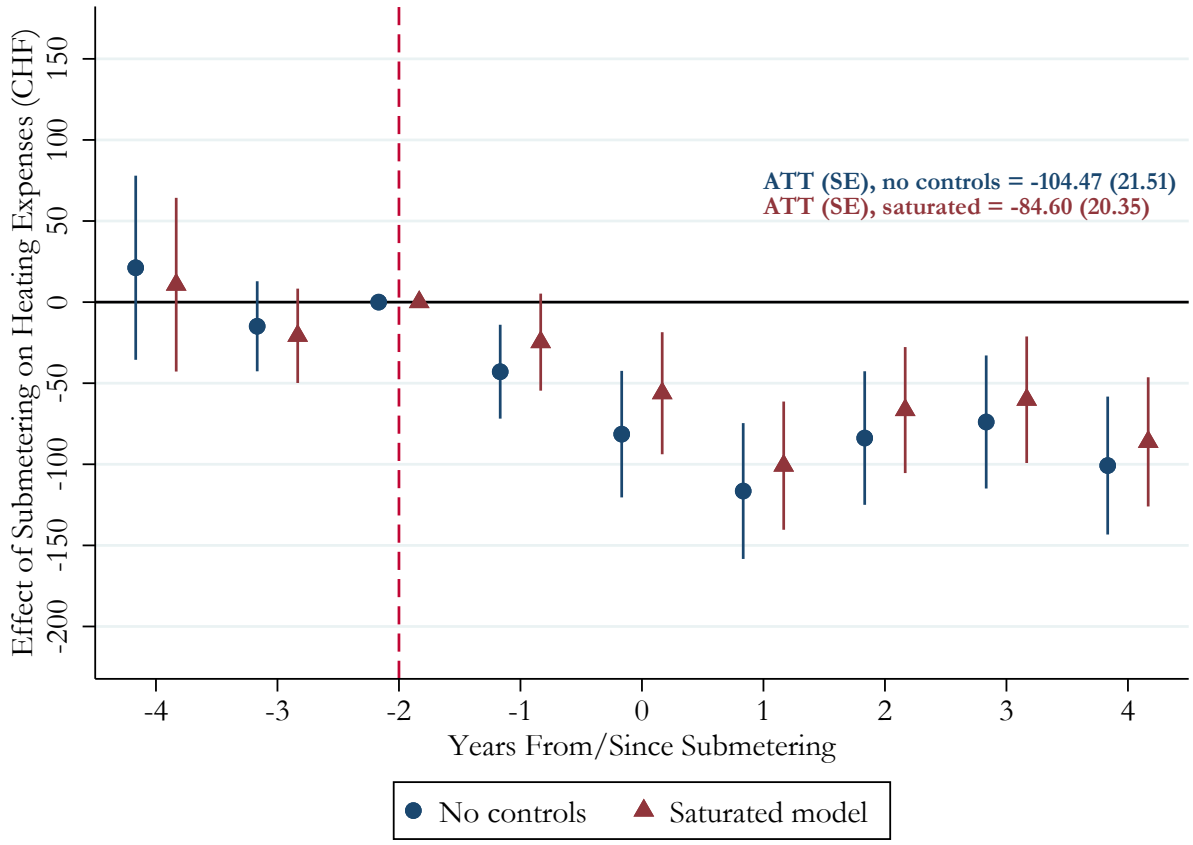


Figure 3: Heterogeneity-robust Effects of Submetering

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 4.2. The outcome variables are heating expenses in levels (Swiss Francs). The ATT estimates and their standard errors, included as text within the graphs, pool the effects from periods $r = 1$ to $r = 4$. The $ATT(r)$ estimates are presented graphically. All specifications include year and apartment fixed effects. The saturated specifications also control for apartment vacancies, heating degree days, energy prices, tenant changes, and renovations. Standard errors are clustered at the building level. The vertical lines around the point estimates represent 95% confidence intervals.

a 9.6 % reduction in the saturated model. Given the 971 median heating expenses in our sample, these effects are also quantitatively similar to those in Figure 3.

Figure 5 presents standard TWFE estimates. 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.

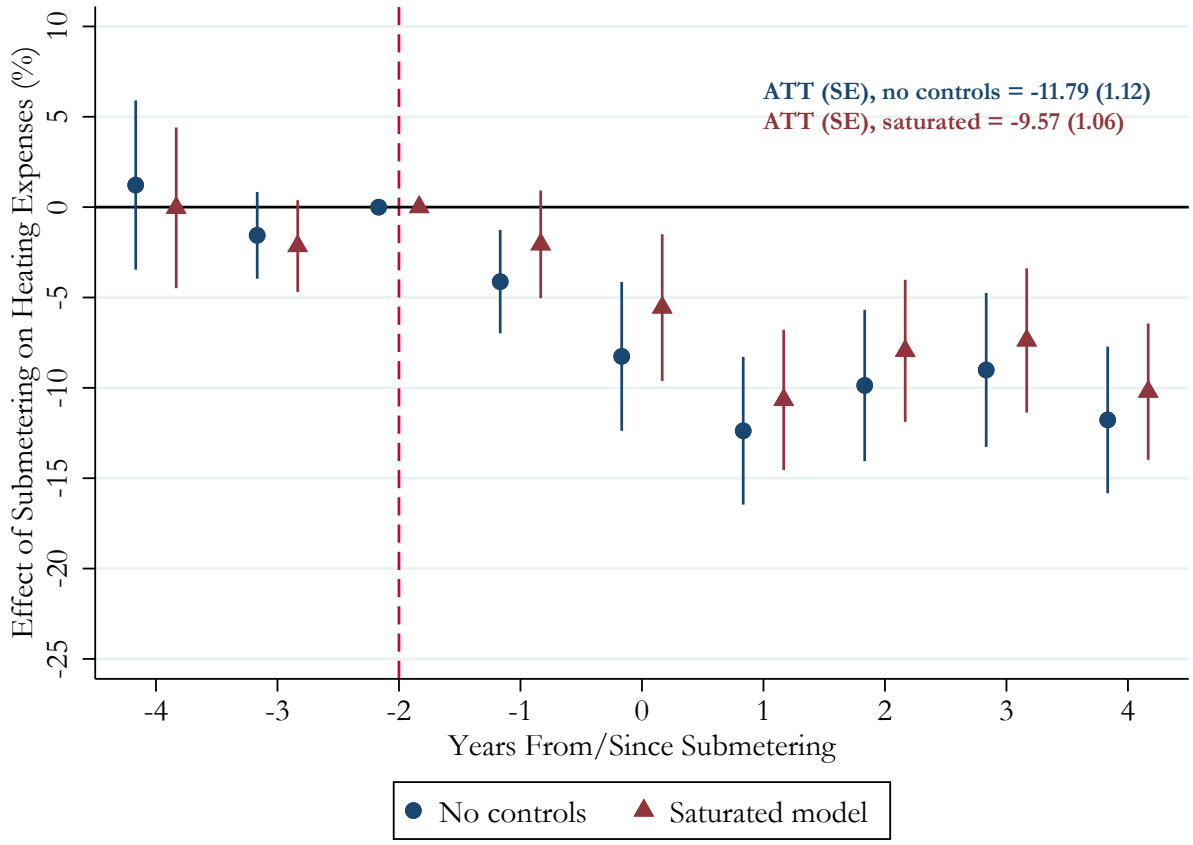


Figure 4: Heterogeneity-robust Effects of Submetering (log)

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 4.2. The outcome variable is heating expenses in logs. The ATT estimates and their standard errors, included as text within the graphs, pool the effects from periods $r = 1$ to $r = 4$. The $ATT(r)$ estimates are presented graphically. Estimates are transformed to percent effects. All specifications include year and apartment fixed effects. The saturated specifications also control for apartment vacancies, heating degree days, energy prices, tenant changes, and renovations. Standard errors are clustered at the building level. The vertical lines around the point estimates represent 95% confidence intervals.

Nevertheless, we rely on the heterogeneity-robust estimates from Sun and Abraham (2021) in the remainder of this paper.

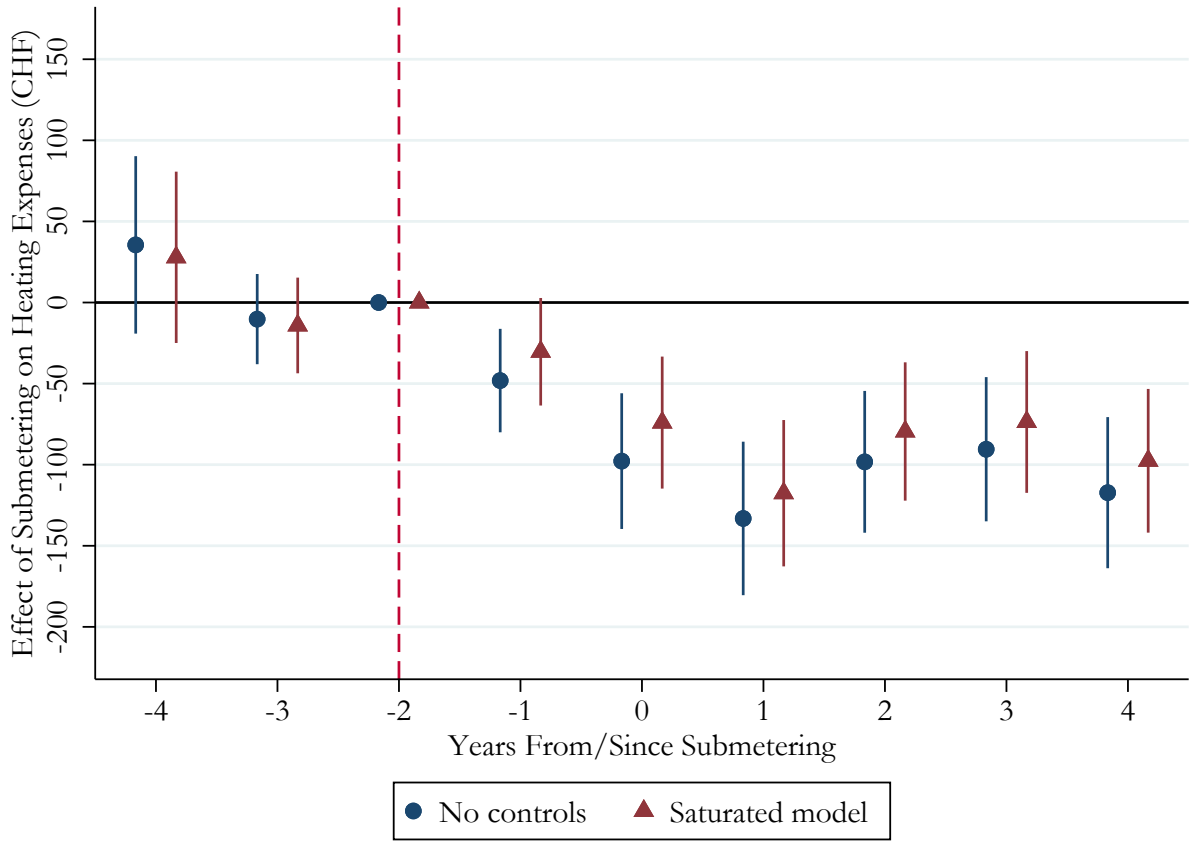


Figure 5: Two-way Fixed Effects Estimates

Notes: This figure presents two-way fixed effects estimates of both ATT and $ATT(r)$ for the effects of submetering. The outcome variable is heating expenses in levels (Swiss Francs). The $ATT(r)$ estimates are presented graphically. All specifications include year and apartment fixed effects. The saturated specifications also control for apartment vacancies, heating degree days, energy prices, tenant changes, and renovations. Standard errors are clustered at the building level. The vertical lines around the point estimates represent 95% confidence intervals.

5 Heterogeneous effects of submetering

To explore heterogeneity and the mechanisms through which individual billing 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 a few key covariates. Let $CATT$ be defined as

$$CATT(\mathbf{c}) = \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0)|S_i = 1; \mathbf{X}_{i,t} = \mathbf{c}; Post_t = 1] \ , \tag{18}$$

where $\mathbf{c}$ denotes a set of conditions or restrictions on a few covariates $\mathbf{X}_{i,t}$ of interest. Our theoretical framework from Section 2 motivates our choice of $\mathbf{X}_{i,t}$ and $\mathbf{c}$. In particular, our framework suggests that the effect of submetering depends on the number of neighbors N that previously shared a common heating bill. Henceforth, we focus on estimating heterogeneity depending on the number of neighbors N.

One natural approach for estimating $CATT$ would be to run a TWFE regression as in equation 15, but restricted to a subsample 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 comparison group from the original sample 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 in levels (as the average outcome of the comparison group might differ from the original sample). Second, subsampling 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 TWFE regression that interacts the covariate of interest (or bins of that covariate) with the submetering treatment indicator. However, that specification does not address another key concern, which is when our covariate of interest N might be correlated with other covariates that also drive heterogeneity in effects.

For example, it could be that apartment size is correlated with the number of neighbors in buildings. This may bias naive estimates of heterogeneity by number of neighbors, in case apartment size also drives heterogeneity of treatment effects. This is particularly expected for estimates in levels, as larger apartments typically require more heating energy and therefore might experience a larger reduction in heating expenses after submetering.

To account for that type of bias, as well as other potential confounders in our setting, we turn to machine learning-based solutions for estimation of conditional average effects. In particular, we implement the method from Souza (2019), which proposes direct prediction of outcomes under a counterfactual (untreated) scenario. The first step, within our setting, 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.

5.1 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 the predictor variables that are not affected by treatment. These include the area of the apartments, the number of rooms per apartment, heating degree days, indicators for construction and renovations years, zip codes of the buildings, as well as the number of neighbors in the buildings (our key covariate of interest).

We use a machine learning 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 split the pre-treatment data into 4 equally-sized subsamples. We use 3 of those subsamples to train the algorithm, then predict energy expenses on the 4th subsample ("validation set"). By recursively repeating this process 4 times, each time changing the validation set, we obtain out-of-sample predictions for all the apartments. Finally, those predictions are compared to

the true pre-treatment expenses, which gives rise to prediction performance metrics such as root-mean-square error (RMSE).

The "best-performing" algorithm is that which achieves the lowest cross-validation RMSE. To begin, we trained 18 configurations of gradient boosted trees (implemented through XGBoost, Chen and Guestrin (2016)). The cross-validation RMSE of those configurations are presented in Appendix Table C.2. Based on those, we selected a model with 3 thousand trees, max tree depth of 10, minimum of 20 observations per node, and shrinkage of 0.2 (details in the Appendix). In future work, we will also test the performance of other algorithms, such as LASSO or random forests.

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

$$\hat{\varepsilon}^{cv}_{i,t} = \sum_{r<-1} \beta_r S_i \times \mathbb{1}[r = t - (q_i)] + u_{i,t} \ . \tag{19}$$

The coefficients β_r will thus represent average pre-treatment residuals at different event times. Estimates from equation 19 are presented in Figure 6 below. We cannot reject that these coefficients are 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 $t \geq -1$.[7] 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).

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

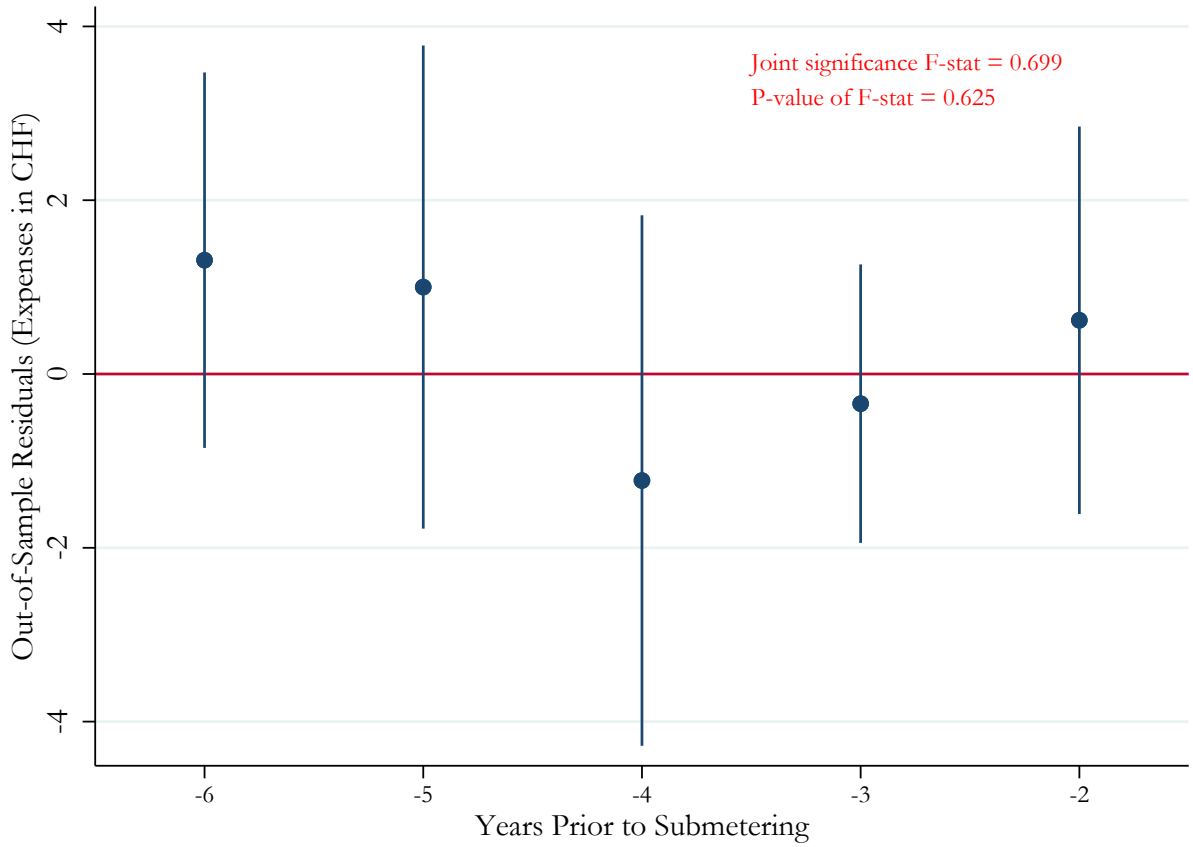


Figure 6: Average CV residuals by event time

Notes: This figure presents coefficient estimates obtained from regression equation 19. These represent the average cross-validated residuals for time periods prior to submetering. Standard errors were clustered at the building level. The reported F-statistic (and corresponding p-value) tests the joint significance of the coefficients.

After we predict the untreated counterfactuals, we subtract them from realized post-treatment energy expenses to obtain a distribution of treatment effects:

$$\hat{b}_{i,t} = Y_{i,t} - \hat{Y}_{i,t}(0) \ . \tag{20}$$

The estimates $\hat{b}_{i,t}$ are the building blocks that can be used to summarize submetering treatment effects across subsamples of interest.

5.2 Summarizing heterogeneous treatment effects

Note that the machine learning approach allows estimation of how treatment effects change over time (i.e., estimation of $ATT(r)$). This can be done by calculating the average of $\hat{b}_{i,t}$ at different times relative to treatment. We summarize these estimates in Figure 7 below. The left-hand side of the figure presents average residuals for pre-treatment observations, while the right-hand side presents average treatment effects in the post-treatment sample. We note that these ML estimates are within range of those obtained with standard TWFE. Reductions in expenses range from 20 CHF (one year before submetering), to 120 CHF (one year after submetering). Importantly, the "whiskers" around the point estimates suggest substantial heterogeneity of effects.

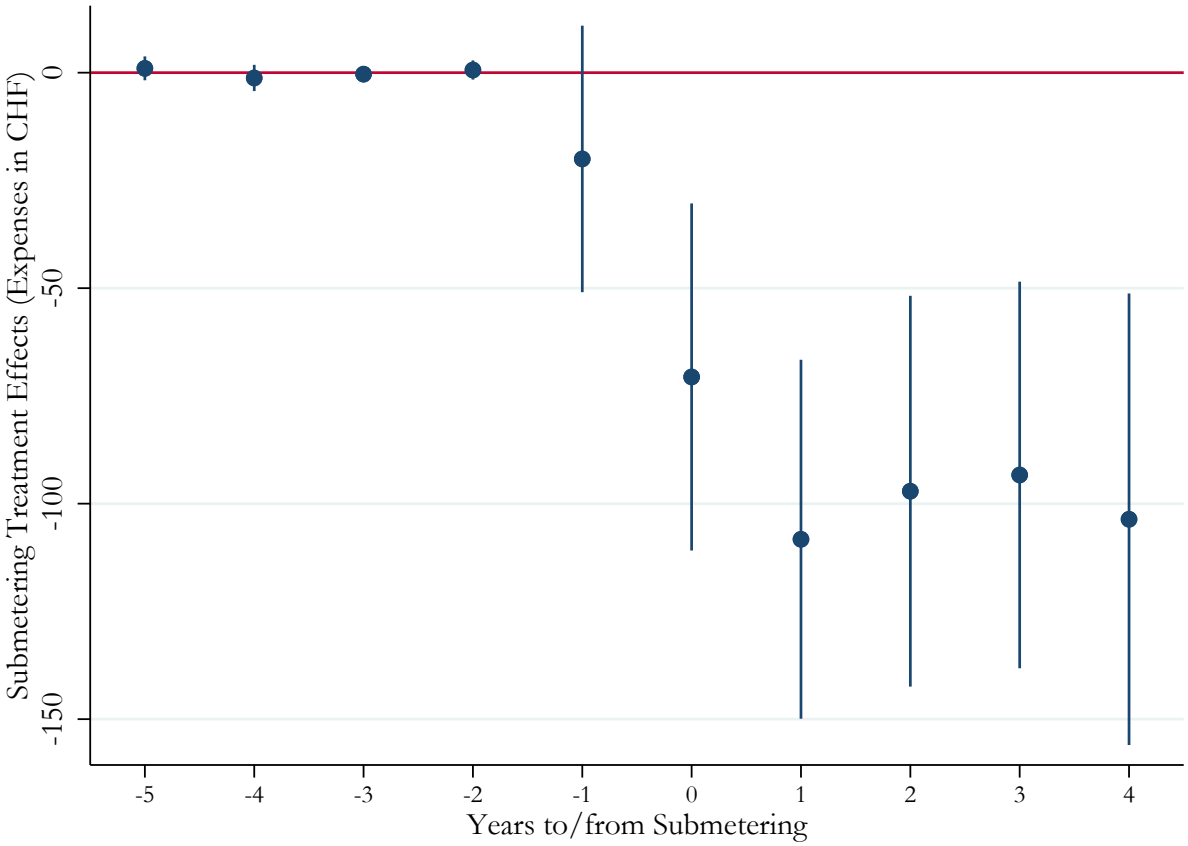


Figure 7: Machine learning estimates of $ATT(r)$

To further explore how heterogeneity may depend on covariates, we summarize $\hat{b}_{i,t}$ across different subsamples of interest. As highlighted in the theoretical framework, we are particularly interested in how the effects change depending on the number of neighbors that shared a common bill. Naive estimates may be obtained by taking the average of $\hat{b}_{i,t}$ for buildings with small, medium, or large number of neighbors, as shown in Appendix Figure C.2. That figure does reveal substantial heterogeneity: the average reduction in heating expenses is CHF 75 for buildings with 6 neighbors or fewer, while this reduction is almost CHF 125 for buildings with more than 14 neighbors. These point estimates are virtually unchanged when we restrict our analysis to a subsample of apartments that had not been recently renovated (also shown in Appendix Figure C.2). However, those naive estimates ignore the fact that there might be other correlated drivers of heterogeneity, such as apartment area.

To account for a more complex heterogeneity structure, we regress the expense changes $\hat{b}_{i,t}$ on a constant, plus quartile indicators for number of neighbors, as well as quartile indicators for apartment area:[8]

$$\hat{b}_{i,t} = \alpha + \sum_{a=1}^{4} \beta^a \mathbb{1}[\text{Area Bin} = a]_i + \sum_{n=1}^{4} \beta^n \mathbb{1}[\text{Neighbor Bin} = n]_i + \varepsilon_{i,t} \ . \tag{21}$$

The coefficients β^a will estimate how the effects of submetering depend on apartment area, holding the number of neighbors fixed. Conversely, β^n capture the heterogeneous effects by number of neighbors, while holding apartment area fixed. Our results, shown in Figure 8 below, reveal substantial heterogeneity along those dimensions.

Panel b of Figure 8 suggests that, holding apartment size fixed, submetering leads to stronger reductions in energy expenses as the number of neighbors increases. The differential effect ranges from 25 to about 60 CHF. This is in line with our theoretical predictions from Section 2. Also as expected, in Panel a of Figure 8 we find that submetering effects are larger (by about 110 CHF) for larger apartments (with more than 100 square meters).

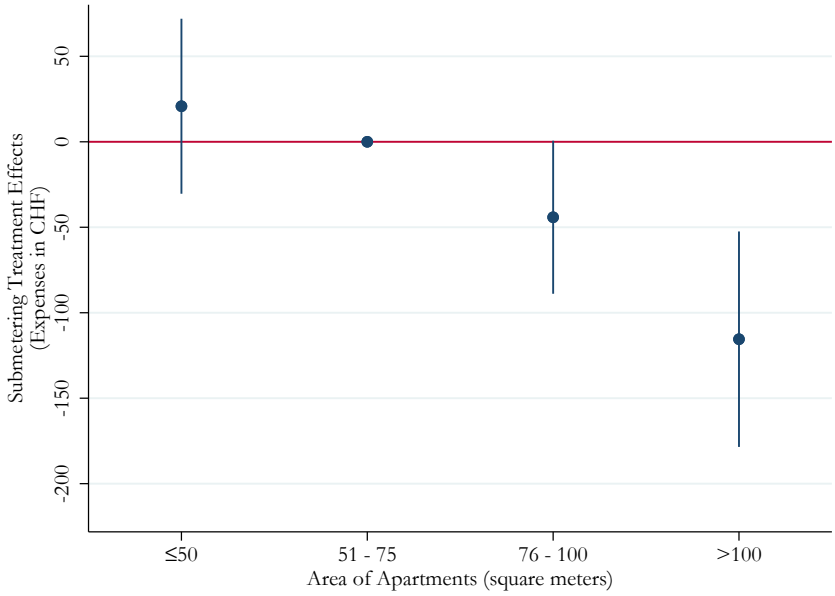
[8]For ease of interpretation, we create indicators based on approximate, not exact quartiles of the covariates of interest.

Importantly, we need to rule out that prediction errors from our ML algorithm are driving the estimated heterogeneity shown in Figure 8. To do so, we run a regression analogous to equation 21, but with ML cross-validated residuals as the outcome. In essence, we test whether these residuals are systematically larger for certain subsamples of interest. Residuals by apartment size and number of neighbors are presented in Appendix Figure C.1. Since all coefficients are close to zero and are not statistically significant, we conclude that our ML algorithm is not biased along those dimensions.

Finally, we use the heterogeneity estimates β^a and β^n from equation 21 to calculate building-specific submetering effects. This is essentially an aggregation step. Note that from 21 we obtain apartment-specific effects that vary depending on the size of the apartments and the number of neighbors. We aggregate those to the building level by summing the resulting $CATT$ from all apartments that belong to the same building.[9] Results are presented in Figure 9. We note substantial heterogeneity, with the majority of buildings saving less than 1,000 CHF annually, while approximately 5% of buildings save more than 2,000 CHF annually. This can have implications for welfare and may lead to distributional considerations, which are discussed in the following section.

[9]We perform this aggregation for a restricted subsample for which we observe all apartments within their respective buildings. This corresponds to about 50% of the sample, and concentrates a larger share of relatively smaller apartment buildings.

(a) Submetering treatment effects by apartment area



(b) Submetering treatment effects by number of neighbors

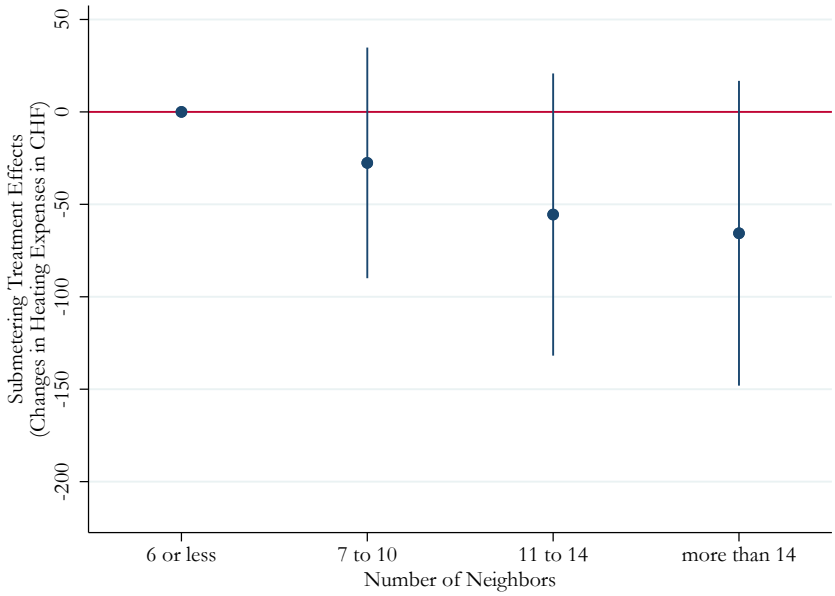
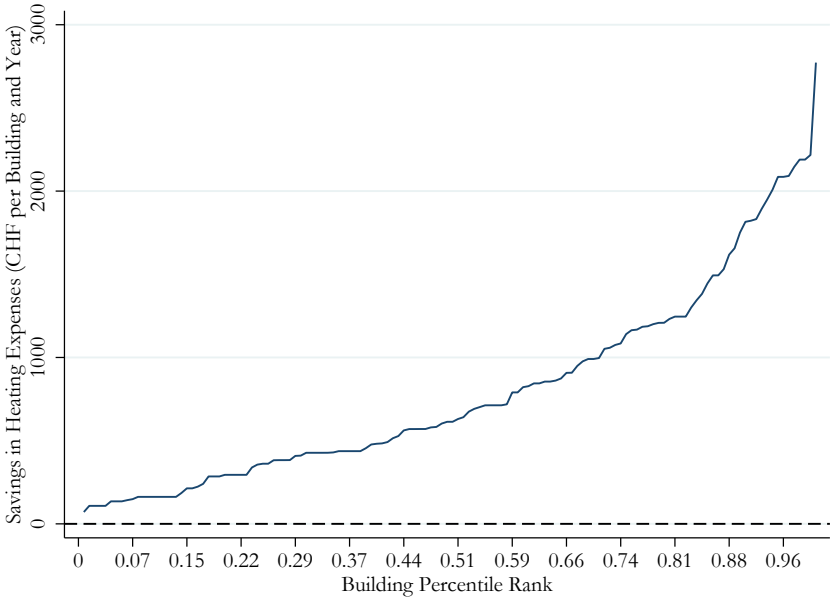


Figure 8: Heterogeneity of treatment effects by selected covariates

Notes: This Figure plots β coefficient estimates and 95% confidence intervals regression equation 21. These coefficients capture treatment effect heterogeneity by variables of interest, while controlling for several confounding factors. Coefficients should be interpreted compared to the omitted comparison categories (zero point estimate in the figures).

(a) Savings in heating expenses (CHF)



(b) Savings in heating expenses (CHF per household)

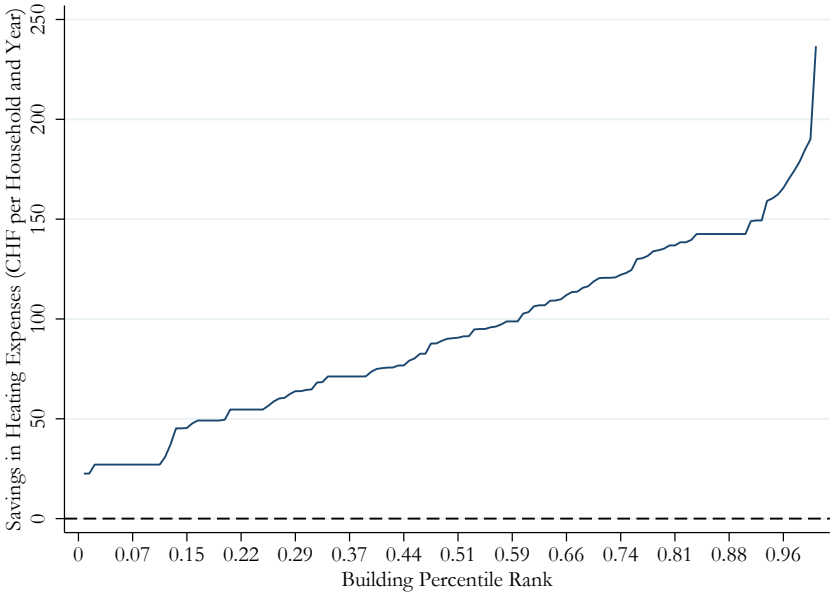


Figure 9: Building-specific treatment effects

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

6 Welfare estimation

We use the sufficient statistic in Equation 14 to assess welfare effects. The relevant parameters are the treatment effect τ, the number of neighbors N, the SCC ϕ, the energy price p, and the cost of individual billing s.

To assess the aggregate welfare effect of submetering in our sample, we use the treatment effect from the saturated model in Figure 3 and the average number of neighbors in our treated sample. We assume unchanged energy prices and a cost of individual billing of CHF 80 per year.[10] We then calculate the SCC that would make the welfare effect of submetering zero. Using the benchmark assumptions mentioned above, we find that submetering is welfare improving if the SCC exceeds 228 CHF/tCO_2.

Table 1 shows how this welfare assessment varies with the energy price p, the cost of individual billing s, and the average treatment effect τ. We consider a scenario where energy prices are 25% higher than in the benchmark scenario. In this case, the SCC that makes submetering welfare improving is substantially lower at 83 CHF/tCO_2. In times of energy crises, it is important to note that our welfare assessment is therefore sensitive to energy prices. The reason is that energy prices influence the welfare assessment twice: through the mechanic effect on energy expenses, and because the treatment effect is linear in energy prices. Next, we consider a scenario where the cost of individual billing is only CHF 70 per year. In this case, the SCC that makes submetering welfare improving is 172 CHF/tCO_2. Finally, we consider a scenario where the treatment effect is twice as large as the average effect from the benchmark scenario. In this large effect scenario, an SCC of 4 CHF/tCO_2 is sufficient to make submetering welfare improving. This finding illustrates that heterogeneity in treatment effects is crucial for welfare assessment, an aspect we discuss in the following.

[10] The cost of individual billing 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 35 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 heating *and water* at CHF 90. Our cost estimate assumes that heat submetering is responsible for 50% of these costs.

Table 1: Welfare Estimates Under Different Scenarios

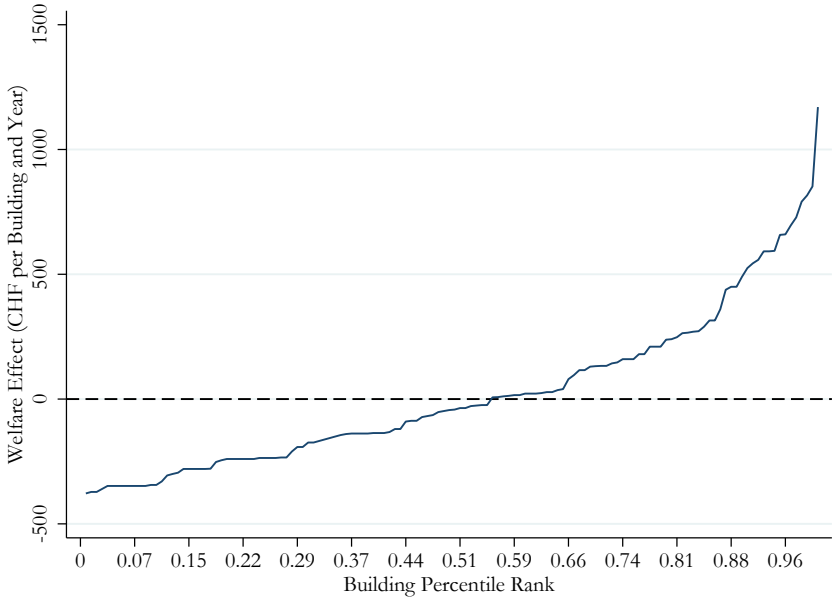
	Scenario			
Parameter	Benchmark	Expensive energy	Inexpensive submeters	Large effect
p	unchanged	+25%	unchanged	unchanged
s [CHF]	80	80	70	80
τ [CHF]	-85	-85	-85	-170
ϕ_{req} [CHF/tCO_2]	228	83	172	4

The aggregate welfare estimates from Table 1 ignore important heterogeneity. For example, in section 5, we showed that buildings with more bill-sharing 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 may differ by building, we therefore take the building-specific expense reductions shown in Figure 9, and convert those into welfare gains/losses, assuming a fixed SCC at 185 CHF/tCO_2.[11] This allows the policymaker, for example, to decide on submetering building-by-building.

Figure 10 shows the welfare effect by building. We find substantial variation in welfare: from -378 to $+1,170$ CHF, with a mean of CHF 18. The majority of buildings would not benefit from submetering, but the welfare effect is positive for 45% of our sample. It is worth noting that this assessment is conditional on underlying assumptions. If a policymaker applies an SCC that exceeds 185 CHF/tCO_2, submetering is welfare improving in a larger share of buildings. The assessment will also be more positive if the cost of individual billing is lower than 80 CHF per year or if higher energy prices are taken into account.

[11] We use the mean SCC estimate from Rennert et al. (2022), of 185 USD/tCO_2 as a benchmark. During the time period we study, the USD-CHF exchange rate between fluctuated around 1. At the time of writing this manuscript, one Swiss Franc is approximately equivalent to 1.1 US dollars.

(a) Welfare effects (CHF)



(b) Welfare effects (CHF per household)

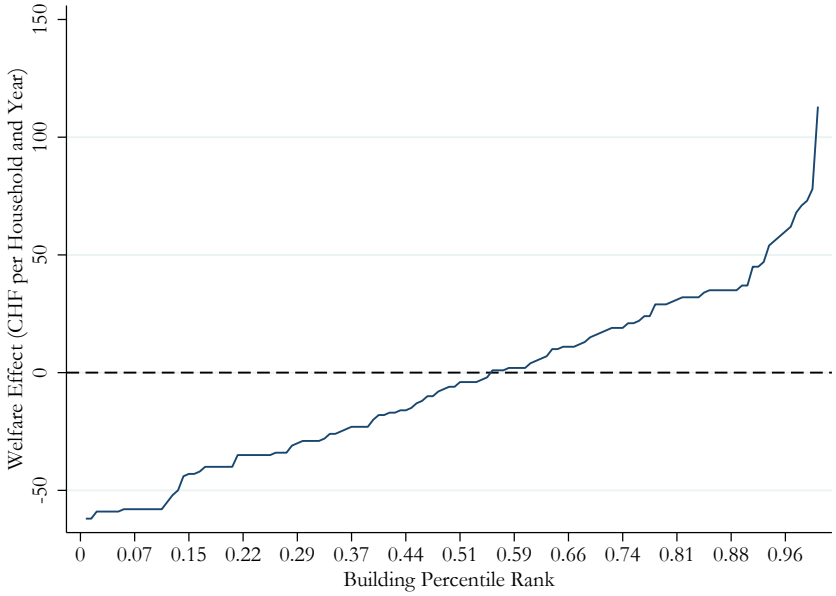


Figure 10: Building-specific welfare effects

Notes: This Figure plots ranked building-specific welfare effects. Buildings are ranked from lowest (left) to highest (right) welfare effect. The x-axes represent percentiles, the y-axes welfare effects. The top panel depicts building-specific effects in CHF per building. The bottom panel depicts building-specific effects in CHF per household.

7 Conclusions

This study investigates the impacts of submetering of heating energy. We use a sample of 3,364 buildings, 265 of which introduced submetering between 2007 and 2021. Our main finding is a persistent reduction of heating energy consumption of 9.6 %. This effect is substantially lower than the effects found in studies on individual billing of electricity (Elinder et al., 2017), water (Ornaghi and Tonin, 2021), and hot water (Elinder et al., 2021). This has important policy implications as, in our sample and under benchmark assumptions, submetering is welfare improving only if the social cost of carbon exceeds 228 CHF/tCO_2.

The richness of our data allows us to study potential mechanisms behind the effects of submetering. To this end, we use recent advances in machine learning to estimate heterogeneity in treatment effects. Consistent with our theoretical framework, we find that the effect of submetering is larger in buildings with more neighbors. Incorporating treatment effect heterogeneity into our welfare analysis is important. We find large differences in welfare effects across buildings. The welfare effect is positive for 45% of buildings in our sample. This suggests that submetering should be implemented building-by-building, rather than as a blanket policy.

Our work raises important questions for future research. We need more evidence on the mechanisms behind the effects of submetering. In particular, we need to understand why the effects of submetering are so much smaller for heating energy than for other utilities. The reason could be a lack of salience of the new billing method, generally inelastic demand for room heating, or heat losses between apartments that render individual savings efforts futile. Energy prices may also be too low to induce larger behavioral change in a rich country like Switzerland. Future studies in different contexts and climates may inform about the external validity of our results. Little is known about tenants' preferences for submetering. Understanding whether, and under which conditions, tenants prefer submetering is also important for policy design. Finally, our theoretical framework suggests that submetering

can increase the own-price elasticity of energy demand. We aim to test that prediction in future work.

References

Allcott, H. and Greenstone, M. (2012). Is there an energy efficiency gap? *Journal of Economic Perspectives*, (26(1)):3–28.

Allcott, H. and Kessler, J. B. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics*, 11(1):236–76.

Angrist, J. D. and Pischke, J. (2009). Chapter 5: Parallel worlds. In *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Borenstein, S. and Bushnell, J. B. (2022). Headwinds and tailwinds: Implications of inefficient retail energy pricing for energy substitution. *Environmental and Energy Policy and the Economy*, 3:37–70.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv Working Paper 2108.12419*.

Canale, L., Dell'Isola, M., Ficco, G., Cholewa, T., Siggelsten, S., and Balen, I. (2019). A comprehensive review on heat accounting and cost allocation in residential buildings in EU. *Energy and Buildings*, 202:109398.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *arXiv:1603.02754*.

Christensen, P., Francisco, P., Myers, E., Shao, H., and Souza, M. (2022). Energy efficiency can deliver for climate policy: Evidence from machine learning-based targeting. *NBER Working Paper*, 30467.

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

Elinder, M., Escobar, S., and Petré, I. (2017). Consequences of a price incentive on free riding and electric energy consumption. *Proceedings of the National Academy of Sciences*, 114(12):3091–3096.

Elinder, M., Hu, X., Liang, C.-Y., et al. (2021). Water conservation and the common pool problem: Can pricing address free-riding in residential hot water consumption? *Center for Environmental and Resource Economics (CERE) Working Paper*.

EnergieSchweiz (2023). VEWA Modell zur verbrauchsabhängigen Energie- und Wasserkostenabrechnung.

European Comission (2021). Revision of the EU trading system. *Document 52021PC0551*.

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

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

Gillingham, K. and Palmer, K. (2014). Bridging the energy efficiency gap: Policy insights from economic theory and empirical evidence. *Review of Environmental Economics and Policy*, 8(1):18–38.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859):1243–1248.

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

International Energy Agency (IEA) (2023). Tracking clean energy progress 2023. https://www.iea.org/reports/tracking-clean-energy-progress-2023.

Ito, K. and Zhang, S. (2022). Do consumers distinguish fixed cost from variable cost? "Schmeduling" in two-part tariffs in energy. *NBER Working Paper 30830*.

Johannesson, G., Agnoletto, L., Anderlind, G., Anderson, B., Godfrey, R., Kimura, K., Lyng, O., Roulet, C., Sorensen, H., and Werner, H. (1985). The calculation of space-heating requirements for residential buildings. *Buildings III Conference Proceedings*, 89.

Knittel, C. R. and Stolper, S. (2019). Using machine learning to target treatment: The case of household energy use. *NBER Working Paper*, 26531.

Knittel, C. R. and Stolper, S. (2021). Machine learning about treatment effect heterogeneity: The case of household energy use. *AEA Papers and Proceedings*, 111:440–444.

Korppoo, A. and Korobova, N. (2012). Modernizing residential heating in russia: End-use practices, legal developments, and future prospects. *Energy Policy*, 42:213–220.

Lu, S., Feng, W., Kong, X., and Wu, Y. (2014). Analysis and case studies of residential heat metering and energy-efficiency retrofits in China's northern heating region. *Renewable and Sustainable Energy Reviews*, 38:765–774.

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

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

Rennert, K., Errickson, F., Prest, B. C., Rennels, L., Newell, R. G., Pizer, W., Kingdon, C., Wingenroth, J., Cooke, R., Parthum, B., et al. (2022). Comprehensive evidence implies a higher social cost of co2. *Nature*, 610(7933):687–692.

Schweizerischer Verband für Energie- und Wasserkostenabrechnung (2023). Verbrauchsabhängigen Heiz- und Warmwasserkostenabrechnung: Übersicht kantonale Energiegesetze / Umsetzung MuKEn 2014.

Souza, M. (2019). Predictive counterfactuals for treatment effect heterogeneity in event studies with staggered adoption. *SSRN Working Paper*, 3484635.

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

A Extensions of the theoretical framework

A.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 can not be reduced by household behavior. These include service fees for the heating system and potentially a chimney, heat losses in pipes serving individual apartments, and the heating of common areas. To account for fixed heating costs, submetering usually only applies to a certain share of the heating bill, $\bar{\theta} < 1$. We extend the theoretical framework to account for fixed heating costs.

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

The budget constraint with fixed heating costs F is

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

The marginal cost of energy consumption in this setting is

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

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

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

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

A.2 Partial submetering

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

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

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

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

We also find that submetering increases the demand elasticity, as shown by equation 11 below.

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

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

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

B Data Appendix

Table B.1: Average Outcomes and Covariates by Treatment Cohort

Treatment Cohort	Heating Bill (CHF per year)	Rent (CHF per month)	HDD	Rooms per Apartment	Apartment Area (sq. meters)	Number of Neighbors
2007	1,096.80	1,197.63	3,669.86	3.90	92.72	7.03
	(29.68)	(33.48)	(47.13)	(0.07)	(1.67)	(1.13)
2008	860.45	846.80	3,299.65	3.38	72.14	9.00
	(32.03)	(35.19)	(50.86)	(0.09)	(2.27)	(1.53)
2009	1,109.85	1,117.11	2,925.25	3.44	81.05	10.95
	(9.55)	(10.34)	(15.16)	(0.04)	(0.98)	(0.82)
2010	1,117.46	1,165.18	3,200.58	3.57	82.52	7.78
	(20.60)	(22.32)	(32.71)	(0.09)	(2.11)	(1.57)
2011	1,040.61	1,162.57	3,022.94	3.35	76.10	9.30
	(12.40)	(13.36)	(19.68)	(0.07)	(1.56)	(1.22)
2012	1,211.16	1,290.77	3,319.95	3.51	85.51	8.74
	(18.38)	(20.08)	(29.18)	(0.09)	(2.10)	(1.53)
2013	1,023.99	1,061.77	3,048.00	3.14	72.50	10.64
	(11.26)	(12.23)	(17.87)	(0.07)	(1.54)	(1.26)
2014	1,139.50	1,003.80	3,038.09	3.99	94.72	9.00
	(19.06)	(20.85)	(30.27)	(0.10)	(2.31)	(1.85)
2015	846.15	1,077.81	5,272.09	3.36	80.83	11.08
	(20.74)	(22.53)	(32.92)	(0.10)	(2.23)	(1.85)
2016	1,241.08	1,130.18	2,758.02	3.58	80.48	13.15
	(16.28)	(17.63)	(25.84)	(0.08)	(1.98)	(1.85)
2017	1,277.18	1,107.14	2,984.31	3.44	75.70	9.89
	(10.68)	(12.07)	(16.95)	(0.06)	(1.34)	(1.11)
2018	1,224.20	1,054.81	2,992.47	3.28	74.53	10.29
	(10.00)	(11.05)	(15.87)	(0.06)	(1.47)	(1.20)
2019	1,010.55	1,255.33	3,061.33	3.05	76.04	9.82
	(13.39)	(14.53)	(21.26)	(0.07)	(1.64)	(1.42)
2020	1,074.83	1,256.50	3,104.77	3.32	77.44	8.88
	(19.51)	(21.59)	(30.97)	(0.08)	(1.94)	(1.62)
2021	1,401.87	1,228.74	3,674.80	3.65	83.13	10.73
	(13.52)	(14.81)	(21.46)	(0.09)	(2.14)	(1.72)
Never Treated	1,012.30	1,252.06	3,229.61	3.26	74.14	9.82
	(0.92)	(1.01)	(1.46)	(0.01)	(0.16)	(0.12)
Observations	203,833	196,851	203,295	34,332	31,065	3,326

Notes: This table present averages and standard errors (in parentheses) obtained by regressing the variables of interest (columns) on submetering treatment cohort indicators (rows). The regressions for Heating Bill, Rent, and Heating Degree Days (HDD) use all observations from the never-treated apartments, but only pre-treatment observations only for the treated apartments.

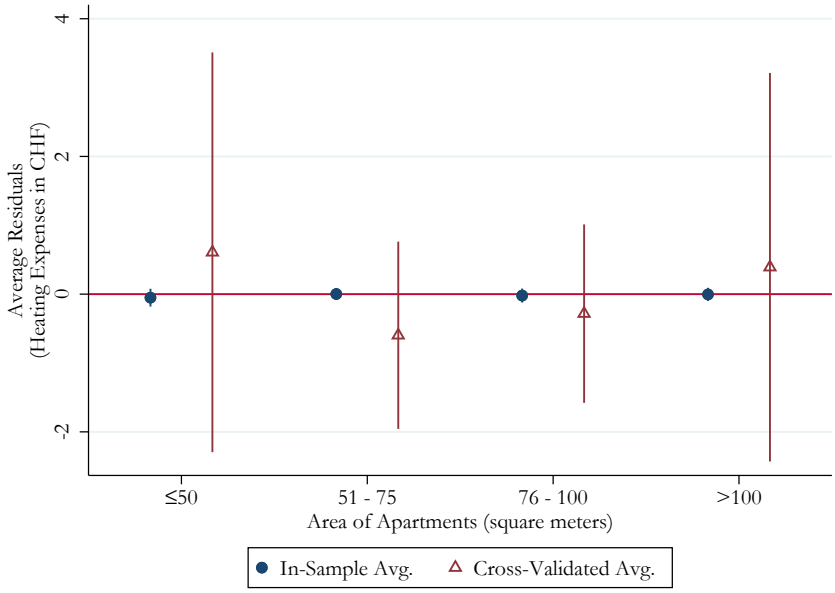
C Machine learning performance

Table C.2: Predictive Performance of XGBoost Configurations

Model ID	N Trees	Max Tree Depth	Min Obs per Node	Shrinkage	In Sample RMSE	Cross-Validated RMSE
1	3000	5	10	0.10	74.785	87.727
2	3000	5	10	0.20	52.161	69.205
3	3000	5	20	0.10	77.175	90.456
4	3000	5	20	0.20	55.932	72.086
5	3000	5	30	0.10	79.749	92.399
6	3000	5	30	0.20	58.209	74.905
7	3000	10	10	0.10	24.679	53.153
8	3000	10	10	0.20	23.023	53.673
9	3000	10	20	0.10	27.448	54.154
10	3000	10	20	0.20	23.901	52.979
11	3000	10	30	0.10	29.403	55.486
12	3000	10	30	0.20	24.898	53.094
13	3000	20	10	0.10	22.531	64.433
14	3000	20	10	0.20	22.524	65.874
15	3000	20	20	0.10	22.641	60.715
16	3000	20	20	0.20	22.530	62.145
17	3000	20	30	0.10	22.986	58.925
18	3000	20	30	0.20	22.559	59.937

Notes: Cross-validated RMSE was obtained through 4-fold cross-validation. Model ID 10 was selected as "best-performing" as it has the lowest CV RMSE.

(a) Errors by apartment area



(b) Errors by number of neighbors

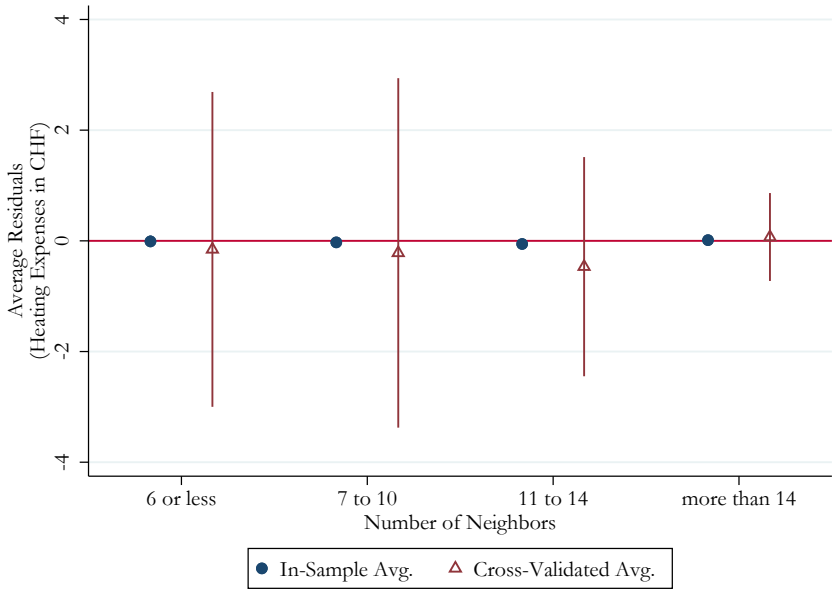
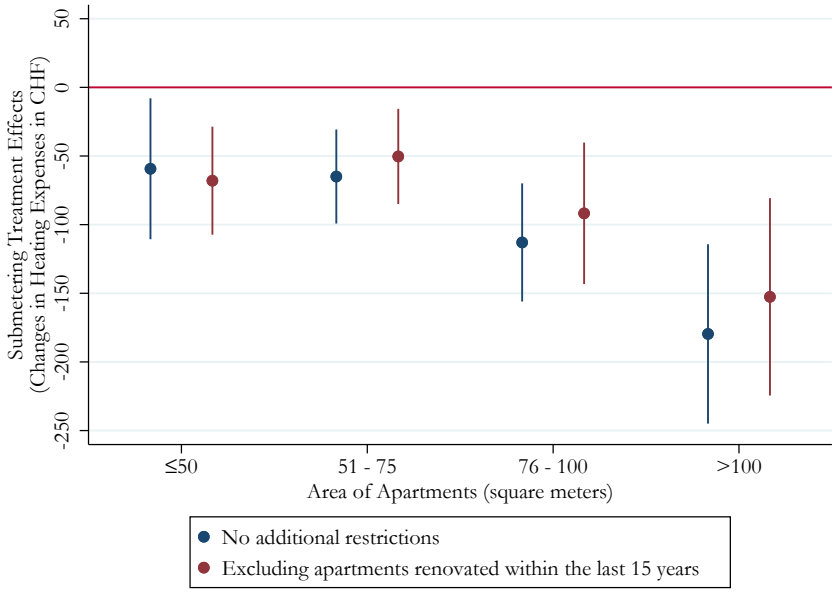


Figure C.1: ML prediction errors by selected covariates

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 and number of neighbors. These serve to test whether the ML algorithm may result in biased predictions for observations within these selected bins. The resulting cross-validated errors are small and not statistically significant across virtually all bins, suggesting that the model is not biased along those dimensions.

(a) Submetering treatment effects by apartment area



(b) Submetering treatment effects by number of neighbors

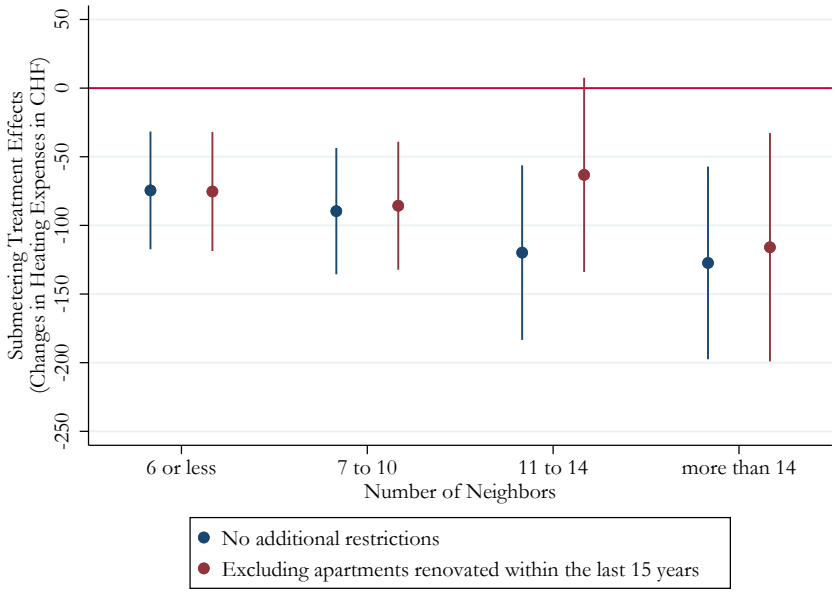


Figure C.2: ML-based naive estimates of submetering treatment effect heterogeneity

Notes: This figure summarizes ML-based average submetering treatment effects for subsamples defined by apartment area (Panel a) or number of neighbors (Panel b). 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. The red dots are point estimates for specifications that exclude apartments that were renovated within the past 15 years.