

Khanna, Tarun M.; Bruns, Stephan; Miersch, Klaas; Minx, Jan C.

**Working Paper**

## Effects of electricity pricing schemes on household energy consumption. A meta-analysis of academic and non-academic literature

*Suggested Citation:* Khanna, Tarun M.; Bruns, Stephan; Miersch, Klaas; Minx, Jan C. (2023) : Effects of electricity pricing schemes on household energy consumption. A meta-analysis of academic and non-academic literature, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/270889>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Effects of electricity pricing schemes on household energy consumption

## A meta-analysis of academic and non-academic literature

Tarun M Khanna<sup>1,2</sup>, [khanna@mcc-berlin.net](mailto:khanna@mcc-berlin.net)

Stephan Bruns<sup>3</sup>, [stephan.bruns@uhasselt.be](mailto:stephan.bruns@uhasselt.be)

Klaas Miersch<sup>2</sup>, MCC-Berlin, [miersch@mcc-berlin.net](mailto:miersch@mcc-berlin.net)

Jan C Minx<sup>2</sup>, MCC-Berlin, [minx@mcc-berlin.net](mailto:minx@mcc-berlin.net)

---

<sup>1</sup> University of British Columbia, Vancouver, Canada

<sup>2</sup> Mercator Research Institute on Global Commons and Climate Change, Berlin, Germany

<sup>3</sup> Hasselt University, Belgium

# Abstract

Time-varying prices are thought to be critical for increasing economic efficiency of the power system. However, a rigorous assessment of the evidence from field trials that use time-of-use pricing, critical peak pricing, etc. in households is missing. This machine learning-assisted systematic review compiles the largest dataset till date of results from pricing pilots reported in both academic publications and electricity utility reports. This unique dataset enables us to deduce the presence of publication bias in peer-reviewed publications. Employing a multilevel meta-analysis, we estimate an average reduction of 8.7%-10.6% in peak consumption, 1.2%-1.5% in total consumption and no change in off-peak consumption across trials. Our heterogeneity analysis, using Bayesian Model Averaging, finds that a 10% increase in the peak-to-baseline price ratio is associated with a 0.47% reduction in peak consumption with marginal reduction in effects suggesting “scope effect” in household behavior. Overall consumption is not responsive to price ratio. Dynamic pricing thus seems to be effective in managing electricity demand but with limits.

# Keywords

Dynamic pricing, Energy demand, Households, Buildings emissions,  
Monetary incentives, Rebates, Time-of-use pricing, Demand  
response

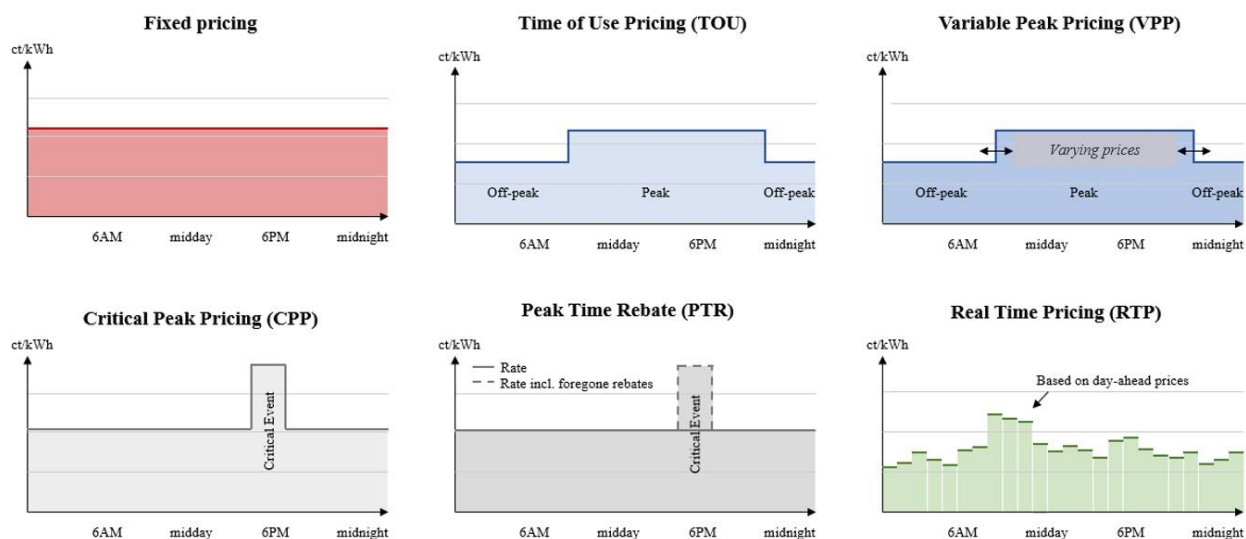
# 1. Introduction

Time-varying electricity prices of electricity, or dynamic pricing, for households, have long been recognized as more efficient than flat-rate electricity prices that are widely prevalent (S. Borenstein, 2009; Joskow & Wolfram, 2012). Wholesale prices for electricity vary sharply, reflecting changing supply and demand patterns, different marginal costs, and the inability to store electricity in large volumes. But households are often not exposed to this volatility. Under fixed-price electricity contracts, households are charged a constant price per kWh of electricity consumed—irrespective of the wholesale market price. Time-varying electricity prices correct this by aligning the prices faced by households to the underlying cost of supply, which are higher during peak demand periods (Sexton & Sexton, 1987). Other interventions reward consumers specifically for reducing peak period consumption. Households are expected to reduce or shift consumption if the financial savings from reduced consumption outweigh the costs of shifting or reducing consumption.

This ability to manage electricity demand through prices is also important for energy efficiency, system planning, and emissions mitigation. Reducing or shifting peak demand can substitute generation capacity at a given level of security of supply (Bushnell, 2005). Moving electricity demand to periods when electricity generation by wind and solar is abundant is necessary for the decarbonization of the energy system in long-term net-zero emissions scenarios (DeAngelo et al., 2021). Such flexibility reduces the cost of renewables integration (Hirth, 2013), avoids curtailment, and substitutes other, most costly flexibility options such as fossil power plants or energy storage (Khanna, 2022; Mier & Weissbart, 2020). With increasing electrification of transport and heating, it is important for the system to evolve so that the additional electricity demand can be directed to a time of the day when the marginal emissions intensity of electricity generation is low (Gagnon & Cole, 2022; Holland et al., 2022). Dynamic pricing can also reduce total energy consumption (Delmas et al., 2013; Khanna et al., 2021).

The most prevalent form of dynamic pricing is time-of-use (TOU) pricing. In such schemes, the day is broken into peak and off-peak periods to roughly approximate the typical daily distribution network load. Prices are different for peak and off-peak periods on each day but are constant within the periods across days. Variable Peak Pricing (VPP) is a particular type of TOU, with the difference that the peak period price fluctuates daily to reflect varying system conditions and costs. Under Critical Peak Pricing (CPP) customers pay the real-time spot market or any other significantly higher price in times of critical events, i.e., times of extreme system peaks, which are usually restricted to only a few hours. Similarly, Peak Time Rebates (PTR) are called in an event of extreme system peak. Here, utilities pay customers a rebate for reduced demand during critical event hours relative to a baseline. The most dynamic form of pricing is Real-Time Pricing (RTP), where wholesale prices are directly passed through to consumers. The retail price thus varies over very short intervals, e.g., every hour. Time-variant schemes can also be grouped together. For instance, CPP and PTR can be added on top of TOU (therefore also called TOU-CPP or TOU-PTR).

Figure 1: Taxonomy of schemes for dynamic pricing



Source: Modified from Montigel (2022)

Dynamic pricing schemes have been piloted extensively, especially in North America, but increasingly also in Asia and Europe. Despite the growing evidence, a rigorous assessment of the existing evidence from the pilots is missing. One of the first surveys of the trials was provided by Faruqui & Sergici, (2011) who described the results from 15 pilots in the US. Later reviews by Faruqui et al. (2013) and Faruqui et al. (2017) use a proprietary dataset (Arcturus) consisting of results from pilots in North America to estimate the effect of electricity pricing schemes on peak demand for electricity. While an excellent first step, these reviews suffer from several methodological and validation concerns. These studies do not attempt a systematic review of the scientific literature, nor do they employ accepted methodologies for meta-analysis but rather rely on descriptive statistics and multivariate regressions. They ignore the effects of pricing experiments on total electricity consumption and there is no attempt towards a critical appraisal of studies or identification of potential publication bias. Other reviews have focused on the effect of monetary incentives on total electricity consumption (Delmas et al., 2013; Khanna et al., 2021) but they ignore the peak demand for electricity and do not establish a relationship between electricity consumption and the size of the monetary incentive. Another review looks at the determinants of the success of demand response studies but is limited in scope and does not provide estimates of reliable effect sizes (Srivastava et al., 2018).

This paper presents a comprehensive, machine-learning-assisted systematic review of the empirical literature aimed at shifting or reducing the energy consumption of households in line with the latest standards for systematic reviews and the Open Synthesis Principles (Collaboration for Environmental Evidence, 2013; Haddaway, 2018; Haddaway et al., 2018). Our search of the published literature is supplemented by a novel data source: compliance reports submitted by utilities to electricity regulators assessing their dynamic pricing experiments. This results in a large dataset of around 85 studies and 650 observations from 14 countries which is several times larger than any existing public dataset. Our dataset includes the latest pilots and studies not only from the United States, UK, and Canada but also China, Japan, Asia, and

India. Unlike previous reviews, this dataset is open access. Beyond creating this dataset our analysis adds to the literature in several ways. First, we estimate the effect of dynamic pricing on not only the peak demand for electricity but on total electricity consumption and off-peak consumption. Second, we attempt to explain differences in its estimates reported in previous studies by coding over 27 parameters that capture study characteristics to assess when dynamic pricing schemes are most effective. Importantly, we are able to assess the role of the size of the monetary incentive in achieving the reductions. One obstacle that we face is the uncertainty over which of the study characteristics should be included in the model. To address this, we employ Bayesian model averaging: a method that estimates many regressions consisting of subsets of the potential explanatory variables and weights them by model fit and model complexity. Third, since about half of our data comes from compliance reports and not peer-reviewed literature, it provides us with a unique opportunity to perform an extensive analysis of the publication bias in this literature. Lastly, we supplement this analysis with a critical appraisal of studies and an assessment of the impact of study designs, out-of-sample bias, and scalability on the results of pricing pilots.

Our analysis shows that dynamic pricing seems to result in peak shaving rather than shifting. The average reduction in peak consumption was about 8.7%-10.6% across studies after adjusting for publication bias. The reported average reduction in total consumption is about 1.2%-1.5% and there is no statistically significant change in the average reported off-peak demand. This implies that subjecting households to dynamic pricing leads to negligible peak shifting but rather to peak shaving, with reductions in both peak and overall consumption. Our analysis of heterogeneity in the reported effect sizes across studies shows that the average peak-to-baseline price ratio of 4:1 reduces peak demand by about 6.5%, with decreasing marginal reduction as the peak prices are increased. There is no evidence that an increase in the size of the monetary incentive leads to marginal reductions in total consumption. The effectiveness of dynamic pricing varies only slightly across countries, but the effect seems to be higher for locations that

use electricity for cooling. The effectiveness is also enhanced by enabling technologies like in-home-display. Estimates of reductions are higher for studies where households are required to opt into experiments. Comparing the results from academic and grey literature shows a statistically significant small study bias and selective reporting of results in the peer-reviewed literature, while there is no detectable publication bias in the reports from grey literature.

The remainder of the paper is structured as follows. Section 2 describes the approach we use to collect estimates and our dataset including the summary statistics. Section 3 investigates the sources of heterogeneity in the results of dynamic pricing experiments. Section 4 estimates the average treatment effect and the influence of publication bias. Section 5 provides a discussion of the results. The appendix provides the correlation matrix of the variables used, shows diagnostics of the Bayesian model averaging exercise, tests for publication bias in the literature, provides a robustness check using an alternative set of priors, and contains the list of studies included in the data set.

#### Collecting the evidence on dynamic pricing of electricity

Our data collection strategy involved string-based searches of bibliographic databases, a review of relevant existing reviews and the studies identified by them, and searches for grey literature on Google. We searched a broad set of bibliographic databases (Web of Science Core Collections Citation Indexes, Scopus, JSTOR), the working paper repository RePEc, and the web-based academic search engine Google Scholar, based on a comprehensive search string (available in supplementary material). We screened the search results using a machine learning algorithm that ordered search results in the order of relevance (see supplementary material for details). In addition to academic research, many electric power utilities, especially in the US, have implemented pilot programs to reduce consumers' electricity consumption during peak hours. These programs are regularly evaluated, and results are published in the form of annual reports and are not published in academic journals. Therefore, in addition to searching academic



literature, we manually searched for utility reports which are not commonly included in academic databases. We identified five main databases that contain relevant reports, the California Measurement Advisory Council (CALMAC), The Department of Energy's Office of Electricity (OE) Resource Library, the Public Utilities Commission California, the Association of Energy Services Professionals (AESP) Resource Library, and Illinois Commerce Commission (ICC) Database. While most of the utility reports are from the United States, some are also from Canada and Europe. We then searched these databases for relevant policy evaluation reports. Overall, a team of four reviewers then manually screened 300 research papers and 153 utility reports at the full-text level. Further details of the search and screening, including inclusion-exclusion criteria, the search query, and the ROSES flowchart are available in the Appendix.

The resulting dataset includes information extracted from 85 research papers and utility reports. It records the change in peak demand, off-peak demand, and total consumption in response to the particular pricing scheme that was analyzed. In addition, the dataset also includes information on the size of the monetary incentive, study characteristics, study design, and methodology among others. There is frequently more than one relevant effect size per study since a pilot may consist of multiple pricing schemes, their impact on the peak, off-peak or total demand, and multiple monetary incentive size levels. Our dataset consists of a total of 648 effect sizes or an average of about eight effect sizes per study. These effect sizes represent pilot studies across 14 countries. The summary statistics of the variables recorded in our dataset are presented in Table 1 and in the appendix.

The way consumers respond to price signals differs by the pricing scheme. Under fixed pricing, consumers do not have the incentive to react to prices as they are constant. Under critical peak pricing (CPP) and peak time rebates (PTR), consumers are likely to respond by shifting demand away from the few hours of “critical events”. Under time-of-use (TOU) pricing, they are expected to shift demand within the day between peak and off-peak blocks. Under real-time

pricing (RTP), consumers can shift demand on an hourly basis. But consumers may not shift demand completely from peak to off-peak hours and may end up reducing their overall energy consumption. About 60% of observations in our dataset deal with TOU pricing with CPP being the next most studied intervention. RTP has been implemented and studied relatively infrequently and we have only 13 observations coming from only four studies. The proportion of effects from these different schemes is roughly similar across the academic literature and the utility reports. The pilots included in our dataset record the impact of the pricing schemes on peak demand, off-peak demand, or total consumption of energy. Since most of the pilots aim at reducing the peak demand for electricity, most of the observations in our dataset record changes in peak demand. Changes in the total electricity consumption and off-peak demand are less relevant to some peak load pricing experiments but are also of less interest as they are not the stated goal of pricing pilots. Only about 22% and 19% of observations in our dataset record changes in total consumption and off-peak demand respectively. In general, utility reports tend to report information about off-peak and total consumption more consistently than academic literature. The average percentage change in total consumption reported in peer-reviewed papers is -2.61% vs. -1.35% in utility reports. Peer-reviewed papers have an average increase of 1.62% in off-peak consumption while utility reports have an average decrease of 0.44%. This suggests systematic differences in reporting of effects by published and unpublished sources and potential publication bias (see section 4. ).

*Table 1: Number of observations by intervention and type of demand measured.*

	Total	Peak	Off-Peak	Mid-Peak	Row Total
CPP	7	105	41	6	159
Rebates/ Rewards	30	41	23	2	96
RTP	7	6	0	0	13
TOU	100	198	64	18	380
Column Total	144	350	128	26	648

Table 2: Mean change in energy consumption (%) by intervention and type of demand measured.

%	Total	Peak	Off-Peak	Mid-Peak	Row Total
CPP	-2.05	-14.59	0.76	-4.50	-9.70
Rebates/ Rewards	-6.05	-14.00	-0.46	-0.47	-7.99
RTP	-0.52	-11.42	NA	NA	-5.55
TOU	-1.21	-8.38	1.42	-2.22	-4.55
Column Total	-2.23	-10.95	0.87	-2.61	-6.34

Notes: Mean in Table 2 refers to the weighted mean using sample size as weights. CPP: Critical Peak Pricing, RTP: Real Time Pricing, TOU: Time-of-use pricing

The two datasets, from academic articles and utility reports, also differ across some other metrics. The treatment and control sample size for utility reports (mean = 6175 and 10620 respectively) is greater for the utility reports than for the peer-reviewed literature (mean = 636 and 2095 respectively), which in turn is greater than the unpublished academic articles (mean = 178 and 272 respectively). The intervention duration is somewhat higher for utility reports (mean = 50 weeks) than peer-reviewed studies (mean = 41 weeks). The proportion of estimates from the US (65%) is much larger in utility reports than in peer-reviewed literature (50%), while the share of estimates from Europe is higher in the case of peer-reviewed literature. There is greater diversity in the study design and methods employed in the peer-reviewed literature with studies using pre-post (8%) and control treatment (48%) designs more often than utility reports that mostly use a difference-in-difference design (64%). Similarly, more peer-reviewed studies use means differences than utility reports that use more OLS methods. The proportion of estimates using panel methods is similar across the two datasets. The diversity of methods also reflects that the peer-reviewed literature is somewhat older. On other metrics, the two datasets are somewhat more similar. The distribution of an important moderator variable—the peak-to-baseline price ratio—is similar across peer-reviewed studies and utility reports (mean = 4.4). The proportion of effects that come from studies where the households opted into the experiment is also similar (30%). The proportion of effects that control for weather, season, residence type, and other characteristics are also similar across datasets.

As there is only a limited number of papers that analyze real-time pricing (RTP) we cannot include these articles in the meta-analysis. The limited evidence does not provide any conclusive evidence regarding the effectiveness of RTP. Fabra et al. (2021) analyze the nationwide RTP rollout in Spain, which made RTP the default option for residential consumers, and find the price elasticity of demand to be zero. The authors indicate that low consumer awareness, costly information acquisition, and low potential financial gains are to blame for the ineffectiveness of the program. In a randomized control trial in the District of Columbia with 1,245 participants, CPP lead to a reduction in peak electricity demand by 9% on average compared to only around 3% for RTP (Wolak, 2011). The author notes that the peak-to-baseline price ratios were on average three times higher for CPP than for the RTP meaning that these effects are not directly comparable. A study on a randomized experiment with voluntary customers in Chicago, Illinois estimated the price elasticity of demand for RTP at 0.1. The study further finds that a price light that indicates particularly high prices further increases the effectiveness of real-time pricing (Allcott, 2011). A report led by the Pacific Northwest National Laboratory with voluntary customers found that both CPP and RTP reduce peak demand significantly (Hammerstrom et al., 2008). Further research on the usefulness of RTP is therefore required to understand the price responsiveness under such schemes though it seems that some of the conclusions in this meta-analysis regarding enabling technologies may hold.

## 2. Reduction in peak, off-peak and total consumption

### 2.1. Estimation strategy

This section employs meta-analysis to estimate the average treatment effect across primary studies. We first standardized the effects by converting the estimates reported by each primary study to the percentage change in energy consumption (either total, peak, or off-peak hours). Many studies in literature do not provide standard errors. Therefore, we use sample size as a

measure of precision instead of the standard error (Fernández-Castilla et al., 2021; Stanley, 2005). Specifically, we use the inverse square root of the effective sample size as a substitute for the standard error (see Appendix for details) as the effective sample size accounts for unbalanced sampling between treatment and control groups (Nakagawa et al., 2022). For ease of exposition, we refer throughout the paper to “standard error”. We probe the robustness of our approach by comparing the results obtained using the effective sample size with the results obtained through reported standard errors for a subset of studies for which standard errors are available (see Appendix). The results obtained using the two approaches are consistent.

We used both a multilevel model (ML) and a fixed effects (FE) model to aggregate the effect sizes from the primary studies. The multilevel model introduces a random effect to account for heterogeneity in the genuine effect whereas the fixed effects model assumes that all primary studies estimate exactly the same genuine effect (M. Borenstein et al., 2009). While genuine heterogeneity is likely to be present in our sample, meta-analyses with random effects have been shown to be more vulnerable to publication bias compared to fixed-effects models (e.g., Stanley & Doucouliagos (2015)). Therefore, we discuss the results from both models.

The ordinary random effects model is however inadequate when the effect sizes included are not independent (Ringquist, 2013). Effect sizes are likely to be dependent in our sample as, following the best practices in meta-analysis, we extracted multiple effect sizes from each study (Havránek et al., 2020; Ringquist, 2013). In addition, several of the studies in our set employ multiple treatments, and some use data from the same underlying experiments. We address this dependence by using a multilevel model that adds further random effects at the level of primary studies. We present the results of the multilevel model as our main results and discuss the results of the fixed effects model to explore robustness.

We used the *metafor* package in R for implementing the random effects model using the restricted maximum likelihood (REML) estimator (Viechtbauer, 2010). We use clustered

standard errors for both the multilevel and fixed effects model using the *club sandwich* package in R to estimate the variance-covariance matrix (Cameron & Miller, 2015; Pustejovsky & Tipton, 2018). To check that no single study exerts undue influence on the aggregate effect sizes measured, we calculated the Cook’s distance and cov ratio for each observation for the base RE and FE models using the *metafor* package. Influential observations were excluded from the analyses, and we provide the code used to identify the observations and the list in the online repository (Havránek et al., 2020).

## 2.2. Results

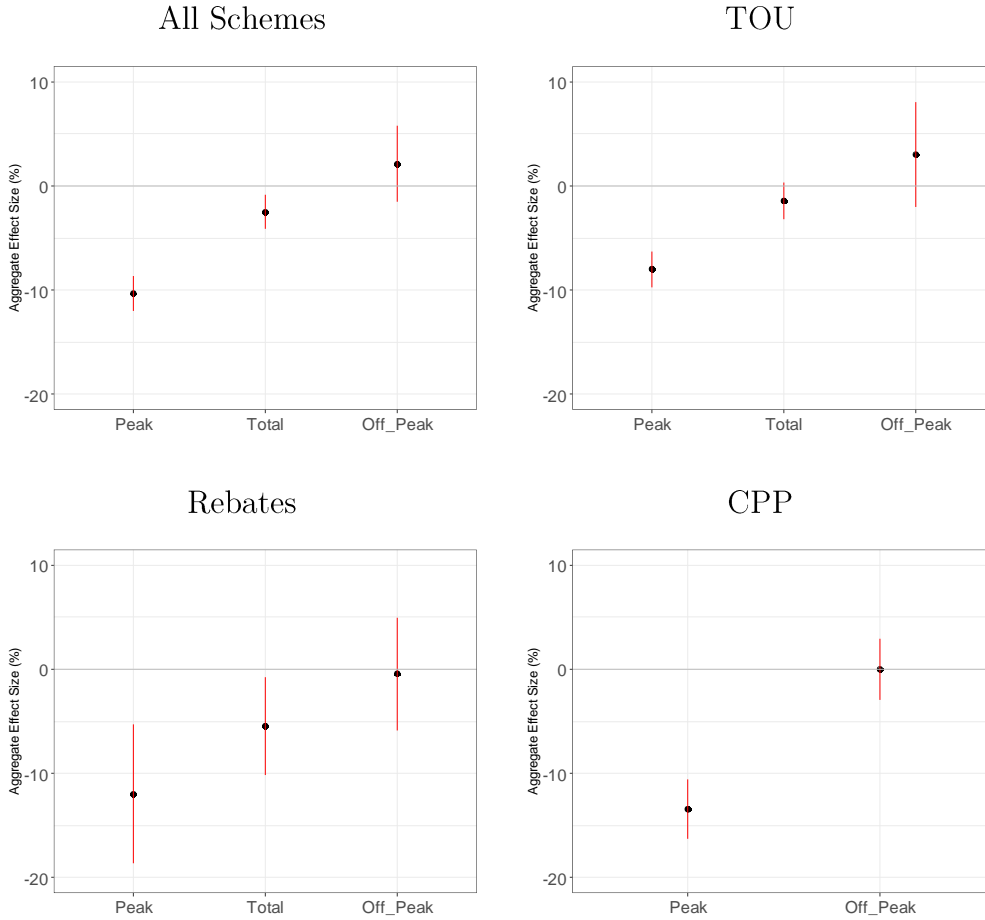
**Aggregate effect sizes.** We calculated the average treatment effect using the multilevel model for the different types of consumption: peak, off-peak and total. First for the complete dataset and then separately for each dynamic pricing scheme. We find that time-varying pricing schemes on average lead to a substantial reduction in energy consumption by households during peak hours, and also to a statistically significant but moderate reduction in total energy consumption. There seems to be no statistically significant effect on off-peak consumption. The average change is estimated to be -10.44% (95% CI = [-12.1; -8.7]) in peak consumption, -2.61% (95% CI = [-4.2; -1.0]) in total consumption, while there is no statistically significant change in the average off-peak consumption (95% CI = [-2.3; 4.7]).

**Scheme-wise treatment effects.** The average estimated treatment effect for peak consumption is higher for CPP and rebates (including peak time rebates) pilots than for TOU, though the estimates for rebates show a high amount of uncertainty. CPP schemes seem very effective in reducing peak demand: the average estimated change is -13.4% (95% CI = [-16.2; -10.6]). But it should be noted that CPP events are usually not applied on a daily basis. Rebates also seem to have a strong effect on peak consumption, with the average effect being -12% (95% CI = [-18.2; -5.6]) but the confidence interval is larger. While TOU experiments show a somewhat lower average reduction in peak demand at -8.0% (95% CI = [-18.6; -5.2]), such

reductions can be sustained over longer periods of time and could lead to permanent shifts in demand. We find a somewhat lower effect of TOU in reducing total consumption. The average estimated effect is -1.4% (95% CI = [-3.15; 0.29]) and is significant at higher confidence intervals. Targeted rebates seem more effective in inducing overall energy consumption with the average effect being -5.5% (95% CI = [-0.7; -10.1]) but there is substantial variation in outcomes and the confidence intervals are large. We do not have sufficient observations that capture the effect of CPP on total consumption. We further explore the effectiveness of the schemes on peak and total consumption in the heterogeneity analysis (see section 3.3).

**Robustness of estimates.** The average treatment effect for peak, off-peak and total consumption presented above is robust to the choice of the meta-analysis model. The results from the fixed effects model are presented in the Appendix. The estimated change in peak and total consumption using the fixed effects model is -9.2% (95% CI = [-12.25; -7.17]) and -1.72% (95% CI = [-2.41, -1.03]) respectively. The change in off-peak consumption is statistically insignificant. While these estimates are somewhat lower than those obtained using the multilevel model, the confidence intervals overlap. The standard errors of the estimates are larger than in the case of the multilevel model, which is expected as the multilevel better models the data, resulting in reduced variance. The results for the scheme-wise estimates are also similar to those from the multilevel model, except for the estimates for rebates, which have much larger confidence intervals such that the effect size for total consumption is not statistically significant.

Figure 2: Average effect sizes and 95% confidence intervals across interventions and the type of demand measured.



Notes: The dependent variable is % change in electricity consumption. Effect  $< 0$  means a reduction in energy demand. The black dots provide the average estimated effect and the red lines show the 95% confidence intervals around the average effect size. Results were obtained using a multilevel model with standard errors clustered at the study level.

### 3. Explaining the heterogeneity in effect sizes

#### 3.1. Potential factors explaining heterogeneity

There is considerable unexplained heterogeneity in effect sizes that we extract from primary studies. A previous meta-analysis by Faruqui et al. (2017) claims that much of the variance in results across studies is explained by differences in the peak-to-baseline price ratio of the analyzed dynamic pricing scheme. But the authors do not control for other study characteristics,



nor do they account for publication bias or study quality. To investigate the heterogeneity among the effect sizes from various dynamic pricing studies more systematically, we code 27 variables that capture several characteristics of the studies. Estimates of the effect size may vary both because of the variation in the underlying response to dynamic pricing (e.g., due to the size of the monetary incentive used in the experiment) or because of differences in estimation methods (e.g., due to differences in the control variables used in the regressions). The complete list of the study characteristics that we code as potential explanatory variables are described below and their summary statistics are presented in the appendix.

#### *Underlying variation in effect sizes*

**Pricing scheme and demand type.** The estimates we extract from primary studies record the response of households to various pricing schemes—TOU, CPP, and monetary rebates including peak time rebates—either separately or in conjunction with each other. Further, the type of energy consumption captured in each estimate can be either peak consumption, total consumption, or off-peak consumption. We include demand type and pricing scheme as moderator variables as it allows us to use the entire sample for the meta-regressions enabling a higher sample size and precision in the estimation of the effects. We also perform analysis for sub-samples with peak and total consumption as the dependent variable, though this leads to a relatively small sample size for total consumption.

**Size of the monetary incentive.** The extent to which consumers respond to electricity pricing schemes is expected to depend on the size of the monetary incentive. Households are expected to reduce consumption if the financial savings from reduced consumption outweigh the costs of shifting or reducing consumption (S. Borenstein, 2009). The ratio between peak-to-baseline prices, or peak-to-baseline prices, can be used to compare the relative size of the monetary incentive across studies or within a study. Indeed, several experiments employ varying peak-to-baseline price ratios (Aigner & Lillard, 1984a; Battalio et al., 1979; Jessoe & Rapson,

2014; Matsukawa, 2018). We expect the reductions in energy consumption to increase with the size of the monetary incentive though it is unclear if the same relationship would hold for peak consumption and total consumption.

**Duration of the peak period.** The duration of the peak period, during which the prices are higher, can affect the potential reduction in energy consumed by households. It may only be possible for households to delay consumption by only a few hours. For example, Hofmann & Lindberg (2021) introduce seven different combinations of two monetary incentive sizes, with different peak time durations and distributions over the hours of the day. They find that the biggest peak load reduction comes with the two tariffs that concentrate the peak pricing in just two hours though these tariffs also have the lowest off-peak load reduction. They also find evidence to show that households postpone consumption to many later times in the day rather than to shoulder hours of the peak period so that the length of the treatment period would not affect behavior.

**Electricity for heating or cooling.** Weather is a key determinant of energy demand. Since most of our sample deals with the pricing of electricity, whether households use electricity for heating or cooling can influence their response to dynamic pricing. In their analysis of a CPP field experiment in Texas, Burkhardt et al. (2019) attribute up to 74% of the reduction in electricity consumption to changes in the usage of air conditioning. Ito (2015) shows that air conditioning captures part of the heterogeneity in treatment effects. Air-conditioning has also been studied as an interruptible load that can be shifted through the right incentives. The evidence on heating load is less prevalent. A few studies like Wolak (2011, p. 86) find that the effects of dynamic pricing are stronger for households that have electric heating than households that don't. Previous meta-analyses have identified that dynamic pricing seems to be more successful in areas with cooling needs (Srivastava et al., 2018) but have not commented upon heterogeneity in effect sizes due to the use of heating or cooling.

**Infrastructure or other complementary interventions.** Dynamic pricing schemes have been enabled in the last decades through smart metering technologies, in-home displays, etc. that allow households to better manage their electricity consumption. Faruqui & Sergici (2011) found that enabling technologies, such as real-time in-home-displays and automated control of air conditioning significantly reduce peak demand. Carroll et al. (2014), in a TOU experiment conducted by an Irish utility, find that in-home displays are most effective in reducing peak demand, but that monthly feedback is most effective in reducing total and off-peak demand. Other primary studies focus on comparing the effect of behavioral interventions like information and feedback with monetary incentives like pricing, or with a combination of approaches including dynamic pricing along with energy savings feedback. For example, Chen et al. (2017) show that non-monetary messages that framed electricity consumption in terms of environmental and health impacts were more effective than messages emphasizing the monetary gain associated with reduced electricity consumption. Moreover, they find that combining environmental messages with monetary incentives led to lower electricity demand reductions than messages alone. This intuition also found support in the meta-analysis by Khanna et al. (2021) who show that some combinations of interventions are complementary but others tend to lower effectiveness.

**Regional variation.** It is possible that the responsiveness of consumers to dynamic pricing schemes varies with the region due to the underlying regulatory and cultural factors, even after accounting for the size of the monetary incentive, use of electricity for heating and cooling, and other observable study characteristics. Effectiveness may differ due to cultural preferences, the density of urban populations, and the ability of populations to shift load. For example, Capitán et al. (2021) find that implementing time-varying pricing may increase total electricity consumption in non-rich countries. Regional variation in types of housing could be another important factor. Apartments, for example, are consistently found to be less energy-demanding

than single-family houses and have more flexible heating/cooling due to better thermal inertia (Pon, 2017).

#### *Variation in estimation and study methodology*

**Study design.** The primary studies in our dataset either compared the electricity consumption of the households before and after an intervention (pre–post) (Herter et al., 2013; Lifson & Miedema, 1981; Slavin et al., 1981), across treatment–control groups (Aigner & Lillard, 1984b; Hartway et al., 1999; Matsukawa, 2018; Midden et al., 1983; Pellerano et al., 2017; Wolak, 2011), or both before and after intervention and across treatment groups (difference–in–difference design, DID) (Andersen et al., 2019; Battalio et al., 1979; Belenguer et al., 2019; Ito, 2015). The robustness of the study design could potentially result in systematic differences in the estimated effect sizes.

**Opt-in.** We also expect that whether households select into treatment impacts study results. We expect the demand response to be lower when households are not required to opt into the experiment. Potter et al. (2014) find that the effect of the treatment on total consumption is significantly larger for households that opt into treatment than for households that are treated by default. The difference is less pronounced for the effect of the treatment on total consumption. Cappers & Todd-Blick (2021) also show that own price elasticity estimates are higher for voluntary TOU customers than for default TOU customers. On the other hand, Fowlie et al. (2021) find that individual responses are higher for opt-in programs but overall savings can be higher by defaulting people into programs. This has important implications for scaling up pricing experiments (Löschel et al., 2020).

**Duration of the experiment.** It is possible that the effect of pricing experiments decreases over time as households adapt to different price levels and ignore the price incentives. For example, Pon (2017) finds that TOU with real-time usage information reduces electricity usage by up to 8.7% during peak times at the start of the trial but the effect becomes indistinguishable

from other treatment groups after the first three months. However, some pilots in our set find that effects increase over time (Elevate Energy, 2015). This could be because the pilot runs over many years with consumers being allowed to opt-in and opt-out. Self-selection by consumers over time could lead to stronger effects over time.

**Control variables.** Primary studies often control for the effect of weather, season, household characteristics, and demographics while estimating the effect size. However, these control variables are not uniformly used across studies and their omission could result in differences in the estimated treatment effect. For example, controlling for weather has been shown to be important in estimating the impact of pricing experiments (Khanna et al., 2021).

**Type of publication.** About half of our sample is derived from academic literature, both peer-reviewed studies and working papers and the other half comes from reports filed by electricity utilities to regulators. It is possible that the results from the academic literature and the utility reports systematically differ, even after accounting for variations in study design, region, and other characteristics discussed above.

**Publication bias.** Economic theory suggests that peak pricing and time-of-use tariffs decrease the consumption of energy by households both during peak periods and overall (S. Borenstein, 2009). Therefore, researchers, reviewers, or editors may suppress effect sizes that are counter-intuitive, i.e., are positive. Moreover, authors may not submit effect sizes that are small or statistically insignificant for peer-review publication. The editors of scientific journals may also be unwilling to publish articles with insignificant findings. In the presence of such publication bias or small study bias, we expect the precision of the estimates to be correlated with the size of the estimates. We discuss publication bias in detail in section 4. To account for such bias in the heterogeneity analysis, we also include the variable ‘standard error’, which is calculated as the inverse square root of the effective sample size, as a moderator variable.

### 3.2. Estimation strategy

The heterogeneity among the reported reduction in energy consumption can be explained by any of the 27 variables identified with aspects of the studies in the previous section. We use the following regression to estimate the relationship between the potential explanatory factors discussed above, and the energy consumption outcomes reported by the studies in our dataset:

$$\hat{y}_{ij} = \alpha + \sum_{k=1}^{27} \theta_k Z_{k,ij} + \varepsilon_{ij}$$

where  $\hat{y}_{ij}$  is the  $i$ -th estimate from the  $j$ -th study, and  $Z_{k,ij}$  is the corresponding value of the  $k$ -th explanatory variable.  $\alpha$  is the intercept and  $\theta$  is a vector of  $k$  parameters,  $\varepsilon$  is a vector of error terms.

It is, however, difficult to specify the best model that contains all the relevant explanatory variables. One approach is to include all variables. But such an estimation would ignore model uncertainty inherent in meta-analysis: while we have a strong rationale to include some of the variables (e.g., monetary incentive size), others are considered mainly as controls for which there is no theory on how they could affect the results of studies (e.g., study design, location). But including all variables in the regression would inflate the standard errors and yield inefficient estimates even though some variables might prove to be redundant. Given the large number of potential model combinations ( $2^{27}$ ), choosing a single model specification by leaving out some of the variables could result in erroneous inference.

To address these issues, we employ Bayesian model averaging (BMA)—an approach successfully used in previous meta-analyses in economics (e.g. Bajzik et al., 2020; Havranek et al., 2012, 2017). BMA runs many regressions with different possible combinations of explanatory variables and does not require selecting one individual specification. We do not estimate all combinations of model specifications possible given our explanatory variables but rather employ the Markov

chain Monte Carlo (specifically, the Metropolis-Hastings algorithm of the *bms* package for R by Zeugner & Feldkircher, 2015), which walks through the most likely models. In the Bayesian setting, the likelihood of each model is represented by the posterior model probability. The estimated BMA coefficients for each variable are represented by posterior means and are weighted across all models by their posterior probability. Each coefficient is then assigned a posterior inclusion probability that reflects the probability of the variable being included in the underlying model and is calculated as the sum of posterior model probabilities across all models in which the variable is included (Eicher et al., 2011; Raftery et al., 1997).

In the baseline specification, we employ the unit information g-prior that the regression coefficient is zero the same weight as one observation of the data (Eicher et al., 2011). This agnostic prior reflects our lack of knowledge regarding the probability of individual parameter values. To test the robustness of our estimates we use the dilution prior that adjusts model probabilities by multiplying them by the determinant of the correlation matrix of the variables included in the model (Havranek et al., 2012, 2017). Furthermore, as another robustness check, we follow Fernandez et al. (2012) who use the so-called BRIC g-prior. As another simple robustness check of our baseline BMA specification, we run a hybrid frequentist-Bayesian model. We employ variable selection based on BMA (specifically, we only include the variables with PIPs above 50%) and estimate the resulting model using OLS with clustered standard errors. This is the “frequentist check” of the baseline BMA exercise. We run the BMA on the total sample and the sample with only peak consumption. We also run the BMA on total consumption estimates but the small sample size (100 observations) leads to relatively high standard errors.

### 3.3. Results

The results of the BMA are shown in Table 3. Inference in BMA is based on a weighted average of individual regressions that include different combinations of explanatory variables; the

weights reflect the posterior inclusion probabilities (PIPs) of the corresponding individual specifications. When the sign of the estimated average coefficient is negative (positive), it implies that an increase in the value of the variable on average leads to higher (lower) reductions in electricity consumption. The column ‘Cond. positive sign’ shows the proportion of individual regressions in which the variable is included with a positive sign.

Results from the BMA support some of the observations that we make in other sections. The reduction in peak consumption is significantly higher than the total. The peak variable has a PIP of 1.00 and the mean posterior estimate is about an 8.94% reduction in energy consumption, which is approximately equal to the average effect size estimated by us (see section 4. ). The regression on the subset of estimates on peak consumption also shows that the reduction in peak demand is lower in TOU schemes as compared to rebates or CPP. Further, the variable ‘standard error’ has a PIP of 51% in the sub-sample of peak consumption and the sign of the variable is unambiguously negative across all the models in which it is included. This indicates the presence of publication bias that we also confirm using other tests (see section 4. ). We do not find any significant differences between estimates from the academic literature and the utility reports, controlling for the effect of other moderator variables including publication bias (see section 4. ). The dummy variable which indicates that we extracted them from a utility report rather than an academic paper has a PIP of only 5%.

The meta-regression also helps us gain new insights into the literature. BMA confirms that the reduction in peak demand is proportional to the size of the monetary incentive measured as the peak-to-baseline price ratio. The log of monetary incentive size has a consistently negative coefficient and is included in all the models. The frequentist check confirms that the coefficient is statistically significant. The estimate implies a non-linear relationship with larger marginal reductions in peak consumption at lower levels of the monetary incentive size but diminishing marginal reductions as the peak-to-baseline price ratio increases. Controlling for other factors, a 10% increase in the price ratio leads to a 0.47% reduction in peak demand (Table 3, column



with Peak Consumption =1). In substantive terms, starting from a uniform price for all hours, as the peak price is increased so that it is double the baseline price, there is a reduction of 3.2% in peak demand. When the price ratio increases to four (the average in our sample), the reduction in peak demand is about 6.4%. Increasing the price ratio brings marginal gains. Our estimates are however lower than those of Faruqui et al. (2017). The regressions also show that there is no discernable relationship between the peak-to-baseline price ratio and total consumption.

The reduction in energy consumption is enhanced by the presence of enabling technologies like "in-home displays" and "smart meters". The variable is included in virtually all the models and always with a negative sign. In contrast, the variable denoting the provision of specific feedback on energy savings has a PIP of about 30%, but the sign of the coefficient is always positive when the variable is included. The direction of the coefficient indicates that the provision of such feedback is associated with a lower reduction in energy consumption and is consistent with results from previous meta-analyses (Khanna et al., 2021). But it is difficult to draw clear conclusions due to the low probability of inclusion of the variable.

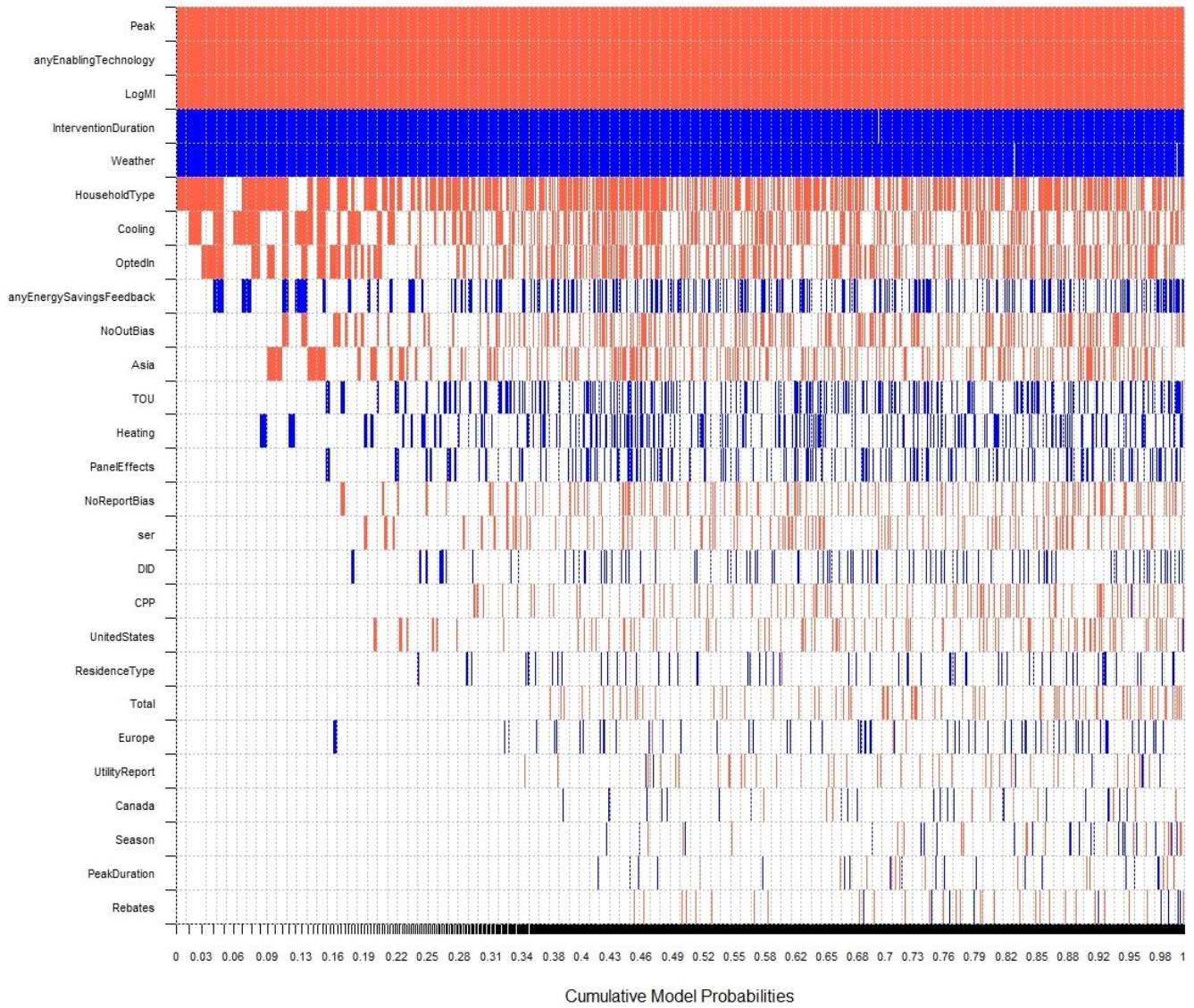
The estimates reported by studies are quite consistent across the regions, except the estimates reported from Asia that seem to vary from other regions even after accounting for other characteristics of the schemes. Studies from Asia report a higher reduction in peak demand. On other hand, the reduction in total energy consumption is not as high but the estimates are less consistent due to the smaller sample size. Peak shifting also seems to be more feasible in regions that use electricity for cooling. The coefficient for cooling in the analysis on the subset with peak consumption is negative with high inclusion probability. Cooling devices seem to provide households with a degree of flexibility that results in a greater reduction in peak demand. This result is consistent with previous meta-analyses (Srivastava et al., 2018). Interestingly, we do not find any such strong relationship with heating demand. In fact, the coefficient for heating

is positive in the subsample with total consumption, and we only find a weak negative relationship between heating and peak consumption.

The results reported by the studies vary with the context and methodology of the study. Studies with a higher treatment duration report lower effect sizes. The effect is particularly strong for the subsample with total consumption, but weak for the sample with peak consumption. This implies that the treatment effects on peak consumption might erode somewhat over time, and the reduction in total consumption may only be seen in experiments with a shorter duration. Studies where households opted into the pilot report higher effect size though the variable is not picked up in other BMA specifications with different priors. We are therefore not able to strongly support previous findings in the literature (Faruqui et al., 2017; Khanna et al., 2021). However, the sign of the coefficient is negative for all models for the subsample with total consumption and largely negative in the peak consumption subsample. Studies that control for weather in their analysis consistently report smaller effect sizes—a finding that is consistent with previous meta-analyses (Delmas et al., 2013; Khanna et al., 2021). On the other hand, the duration of the peak period does not seem to affect the estimated reduction in energy consumption. Studies that control for the demographics of the household seem to report a higher effect size though this variable is not picked up in other BMA specifications with different priors (see Appendix). Controlling for other variables such as the season and type of house does not seem to be significant.

We do not detect any bias on account of other methodological choices made in the study. Further, studies that we tagged as having a potentially low out-of-sample bias and reporting biases do not seem to systematically differ in their outcomes as compared to studies that do exhibit higher biases.

Figure 3: Models inclusion in Bayesian Model Averaging



Notes: The variables are sorted according to their posterior inclusion probabilities from the highest at the top to the lowest at the bottom. The horizontal axis measures cumulative posterior model probability. Red = the estimated parameter for the variable is negative (higher effectiveness of dynamic pricing). Blue = the estimated parameter for the variable is positive (lower effectiveness of dynamic pricing). No color = the variable is not included in the model. Numerical results are reported in Table 3.

Table 3: Explaining differences in the estimates of the effectiveness of dynamic pricing, numerical results from the Bayesian Meta-Analysis (BMA) for the whole sample, and only estimates of peak consumption.

	All observations				Observations with Peak Consumption = 1				Observations with Total Consumption = 1				Frequentist Check		
	PIP	Post Mean	Post SD	Cond. positive sign	PIP	Post Mean	Post SD	Cond. positive sign	PIP	Post Mean	Post SD	Cond. positive sign	Coefficient	SE	p-value
Peak	1.00	-8.94	0.56	0.00									-9.49	1.34	0.000
Total	0.07	-0.07	0.34	0.00									-0.85	1.12	0.455
CPP	0.08	-0.08	0.33	0.01	0.29	0.74	1.37	0.97	0.13	-0.28	1.10	0.04			
TOU	0.25	0.39	0.80	1.00	0.73	2.84	2.30	1.00	0.13	-0.25	1.02	0.14			
'Standard error'	0.11	-1.24	4.33	0.00	0.51	-13.62	15.37	0.00	0.08	-0.15	2.29	0.43	-11.70	6.43	0.075
Utility Report	0.05	-0.03	0.20	0.11	0.04	0.00	0.18	0.57	0.10	0.00	0.42	0.59			
Any Enabling Technology	1.00	-4.29	0.87	0.00	0.99	-3.89	1.07	0.00	0.87	-2.37	1.24	0.00	-3.94	1.11	0.001
Any Energy Savings Feedback	0.31	0.55	0.95	1.00	0.05	0.03	0.29	0.83	0.15	-0.19	0.69	0.19			
Panel Effects	0.18	0.25	0.62	1.00	0.52	1.26	1.39	1.00	0.25	0.37	0.81	1.00			
DID	0.10	0.10	0.35	1.00	0.29	0.57	1.02	1.00	0.11	0.10	0.52	0.84			
Asia	0.22	-0.63	1.36	0.00	0.89	-5.43	2.59	0.00	0.75	4.02	2.99	1.00			
Europe	0.07	0.07	0.41	0.90	0.06	-0.02	0.40	0.32	0.15	-0.18	0.65	0.12			
Canada	0.04	0.01	0.19	0.73	0.05	-0.00	0.29	0.49	0.89	-3.65	1.83	0.00			
Intervention Duration	1.00	0.05	0.01	1.00	0.10	0.00	0.01	1.00	0.92	0.04	0.02	1.00	0.04	0.01	0.004

	All observations				Observations with Peak Consumption = 1				Observations with Total Consumption = 1				Frequentist Check		
	PIP	Post Mean	Post SD	Cond. positive sign	PIP	Post Mean	Post SD	Cond. positive sign	PIP	Post Mean	Post SD	Cond. positive sign	Coefficient	SE	p-value
Peak Duration	0.03	-0.00	0.02	0.42	0.10	0.02	0.07	1.00	0.17	0.03	0.10	0.94			
Opted-In	0.37	-0.58	0.86	0.00	0.05	-0.02	0.22	0.31	0.47	-1.04	1.31	0.00			
Household Type	0.60	-1.32	1.25	0.00	0.19	-0.44	1.08	0.00	0.88	4.75	2.43	1.00	-2.24	0.86	0.013
Residence Type	0.08	0.11	0.51	0.97	0.24	0.85	1.78	1.00	0.17	0.32	1.09	0.85			
Season	0.04	0.01	0.25	0.58	0.18	0.30	0.77	1.00	0.28	-0.54	1.10	0.02			
Weather	1.00	2.66	0.65	1.00	0.14	0.19	0.56	1.00	0.12	-0.09	0.49	0.23	3.04	0.54	0.000
Heating	0.21	0.34	0.77	1.00	0.05	-0.02	0.28	0.23	0.17	0.24	0.68	1.00			
Cooling	0.45	-0.73	0.92	0.00	0.89	-2.71	1.34	0.00	0.11	-0.08	0.36	0.06			
Heating-Cooling	0.08	-0.08	0.39	0.08	0.06	0.05	0.37	0.92	0.08	-0.02	0.39	0.31			
No-Out-of-Sample-Bias	0.26	-0.39	0.75	0.00	0.09	0.10	0.43	0.99	0.50	-1.18	1.40	0.00			
No-Reporting-Bias	0.13	-0.25	0.83	0.02	0.69	-3.20	2.65	0.00	0.22	-0.71	1.72	0.00			
Ln(Peak price:baseline price)	1.00	-3.04	0.55	0.00	1.00	-4.66	0.75	0.00	0.09	0.01	0.27	0.60	-3.09	0.57	0.000
(Intercept)	1.00	1.64	NA	NA	1.00	-0.83	NA	NA	1.00	-0.80	NA	NA	2.84	1.50	

Notes: SD = standard deviation. SE = standard error. PIP = posterior inclusion probability. Cond. positive sign = percentage of models in which the variable is included with a positive coefficient. Response variable = percentage change in electricity consumption. Bayesian model averaging (BMA) employs the unit information prior. The frequentist check (OLS) includes the variables recognized by BMA to comprise the best model and is estimated using standard errors clustered at the study level.

## 4. Publication bias

Previous reviews in the literature on dynamic pricing have not assessed the presence of publication bias<sup>4</sup>, even though publication bias has been detected previously in the literature on experiments involving household energy consumption more generally (Buckley, 2020). Publication bias has been also diagnosed in economics (e.g., Bruns et al., 2019; Ioannidis et al., 2017)), management (e.g., Goldfarb & King, 2016), and the sciences more generally (Fanelli et al., 2017). Economic theory suggests that peak pricing and time-of-use tariffs decreases the consumption of energy by households both during peak periods and overall (S. Borenstein, 2009). Therefore, researchers, reviewers, or editors may suppress effect sizes that are counter-intuitive, i.e., are positive. Authors may also be reluctant to submit effect sizes that are small or statistically insignificant to peer-reviewed journals and editors may not be keen to publish such papers either. Such results would then go unpublished in peer-reviewed journals even if they are reported as a part of a working paper series.

Publication bias is natural, and inevitable, and does not require any ulterior motives on the side of authors, editors, or referees (e.g., Bruns & Kalthaus, 2020). On the other hand, the reports submitted by utilities to regulators providing an assessment of their pilots on dynamic pricing should not suffer from such a problem. Such pilots are usually designed to be revenue neutral, i.e., they are designed in a fashion such that the expected revenue for the utilities should remain the same. The utilities, therefore, do not have a clear motivation to inflate the effects. Further, since utilities are required to report their results annually, no matter what the

---

<sup>4</sup> Selection of statistically significant results may appear at the level of analyses within studies (commonly known as *p*-hacking (e.g., Bruns and Ioannidis, 2016) or at the level of studies (often referred to as the "file-drawer problem" or "publication bias" (Rosenthal, 1979). We use publication bias in the widest sense encompassing both selective reporting at the level of analyses and at the level of studies, following, for example, Havranek et al. (2017).

results, it is less likely that some of the results go completely unreported even by chance. About half of the estimates in our dataset come from such utility reports. The reports are submitted by more than a dozen utilities across the world to their local regulators at various points in time. Given the lack of inherent motivation and this diversity, we do not expect the results from these utility pilots to be subject to publication bias.

Publication bias results in a correlation between standard error and effect size. The correlation can result due to two reasons. First, researchers (or editors or referees) may prefer statistically significant results. Researchers with small sample sizes and large standard errors may try, for example, different combinations of control variables until they obtain an estimate large enough to offset the standard error. Second, researchers may prefer an intuitive sign of the estimates and discard those with the opposite sign. The association between standard error and effect size can be used to detect and correct for publication bias.

It is helpful to evaluate the relationship visually using the so-called funnel plot: a scatter plot with the effect sizes on the horizontal and their standard errors on the vertical axis. Based on the intuition described in the previous paragraph, an asymmetry of the funnel plot suggests publication bias (Egger et al., 1997). Figure 3 and Figure 4 show funnel plots for the academic literature (including those published in peer-reviewed journals and those that are derived from unpublished sources) and the utility reports. The x-axis shows the reported percentage change in energy use and the y-axis shows the inverse of the standard error, which we proxy with the inverse square root of the effective sample size (section 2.1. ). There are a few noticeable differences between the plots for the two types of publications. First, the precision of the effect sizes for the academic reports is generally lower as compared to the utility reports. This is primarily because the sample size in utility pilots is on average larger. Second, the distribution of effect sizes for the utility reports is almost perfectly symmetric around the estimated average effect. There are also several estimates with the counterintuitive (positive) sign for total consumption. This is not the case for peak consumption, perhaps due to the relatively large

average effect. Third, the funnel plots for academic literature in case of peak consumption show some bias for larger effect sizes. Noticeably, the effect sizes from the unpublished academic sources are in general smaller than the average effect. The funnel plot for total consumption also shows a clear asymmetry towards larger, more negative effect sizes, especially in the more imprecise studies.

We use statistical tests more formal than the funnel plot to confirm the presence of publication bias and to estimate the ‘true’ average effect size adjusted for publication bias. We use PET (precision-effect test) and PEESE (precision-effect estimate with SE) for estimating/testing the ‘true’ effect in the presence of publication bias (Stanley & Doucouliagos, 2014). PET and PEESE use the following regression models:

$$\hat{y}_{ij} = \beta_0 + \beta_1 SE_i + s_i + m_j + u_i \quad \text{PET}$$

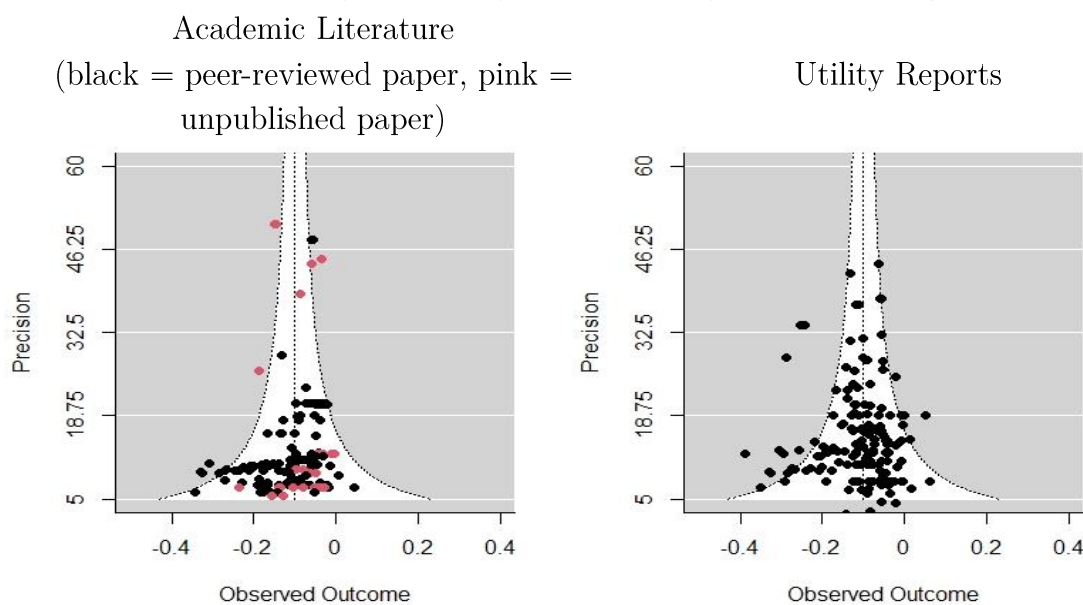
$$\hat{y}_{ij} = \gamma_0 + \gamma_1 SE_i^2 + s_i + m_j + u_i \quad \text{PEESE}$$

where  $\hat{y}_{ij}$  is the  $i$ -th estimate from the  $j$ -th study, and  $SE_i$  is the corresponding standard error,  $m_j$  is the between-study effect for the  $j$ -th study, normally distributed with variance  $\sigma_m^2$ ,  $s_i$  is the within-study effect, for the  $i$ -th effect size, distributed with a mean of zero and the variance  $\sigma_s^2$ ; and  $u_i$  is the sampling error. The coefficient of the standard error measures publication bias and can be thought of as a test of asymmetry in the funnel plot. The intercept of the models ( $\beta_0$  and  $\gamma_0$ ) estimates the ‘true’ effect beyond publication bias. PET is a powerful test for the presence of an authentic effect beyond publication bias (Stanley, 2008), but  $\beta_0$  tends to underestimate the true mean effect when there is a nonzero treatment effect. If an indication is given for the presence of a genuine effect,  $\gamma_0$  is the preferred estimate (Stanley & Doucouliagos, 2014). We use the inverse square root of effective sample size as a substitute for the standard error and both PET and PEESE are perfectly suited for this (Nakagawa et al., 2022).



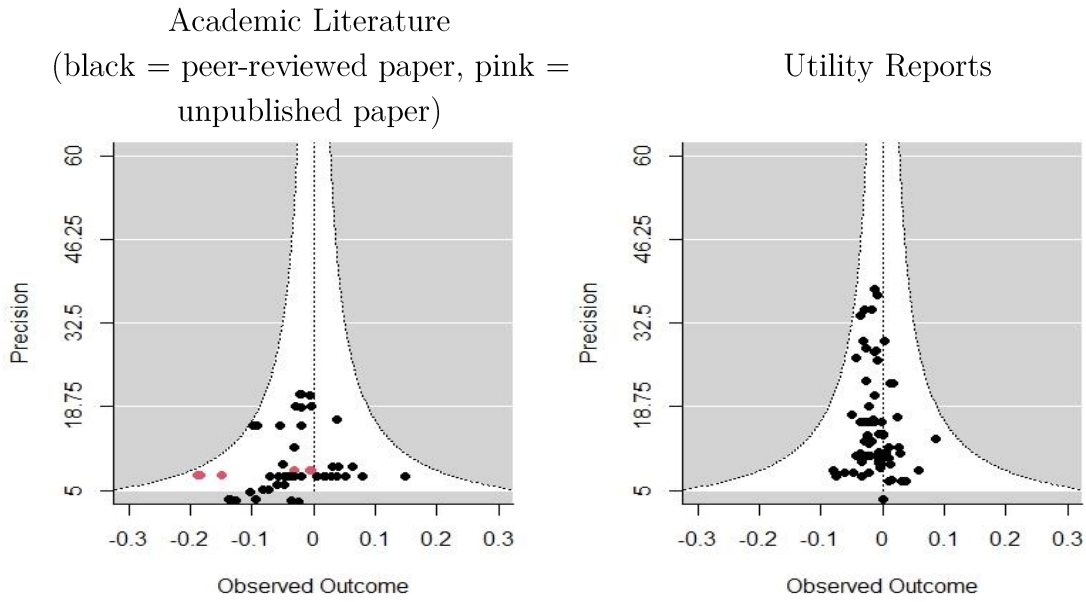
Other commonly used methods to correct for publication bias require the actual standard errors, such as the WAAP (weighted average of the adequately powered studies, Ioannidis et al., 2017) or the stem-based approach (Furukawa, 2019). These methods, however, have in common that they estimate the average treatment effect for the most precise studies but they differ in how they select these studies. Therefore, we employ as a further approach to adjust for publication bias the top 10% estimator that uses only the top 10% of the most precise estimates available in our dataset (Stanley et al., 2010). This approach can be easily adapted to the use of effective sample size as a measure of precision and follows the same rationale as other commonly used approaches to correct for publication bias.

Figure 4: Funnel plots showing the percentage change in energy use and the inverse of standard error as a measure of precision for the subset of data measuring peak consumption



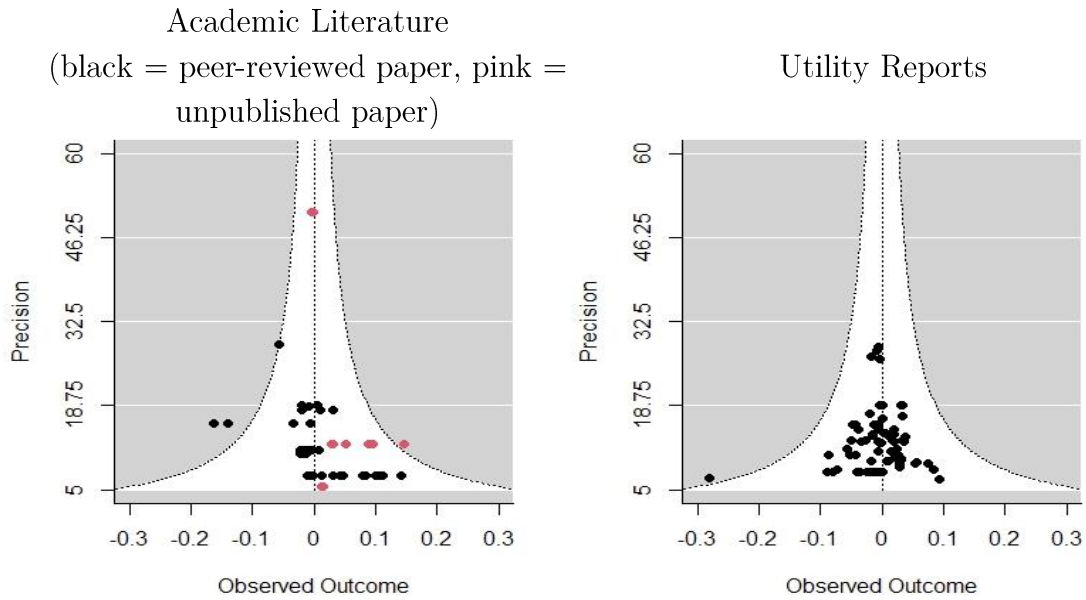
Notes: In the absence of publication bias the funnel plot should be symmetrical. Outliers are excluded from the figure for ease of exposition. The reference line is the estimated average effect.

Figure 5: Funnel plots showing the percentage change in energy use and the inverse of standard error as a measure of precision for the subset of data measuring total consumption



Notes: In the absence of publication bias the funnel plot should be symmetrical. Outliers are excluded from the figure for ease of exposition. The reference line is the estimated average effect.

Figure 6: Funnel plots showing the percentage change in energy use and the inverse of standard error as a measure of precision for the subset of data measuring off-peak consumption



Notes: In the absence of publication bias the funnel plot should be symmetrical. Outliers are excluded from the figure for ease of exposition. The reference line is the estimated average effect.

We estimate PET-PEESE and the top10% approach using both the multilevel model (Figure 6) and the fixed effects model (see Appendix) with clustered standard errors. The results from the regressions confirm the presence of publication bias in the peer-reviewed publications for both the estimates reported for total and peak consumption. We do not find any evidence of publication bias for estimates stemming from utility reports.

This finding is also consistent with utility reports frequently reporting estimates for peak, off-peak and total consumption while academic articles often focus on peak consumption where the genuine effect is large and statistically significant estimates can be easily obtained. The share of estimates for off-peak and total consumption in our data from utility reports is much higher (64% and 65% respectively) than share of estimates for peak consumption (55%). When total and off-peak consumption is reported, the effect size is higher in peer-reviewed literature. The average percentage change in total consumption reported in peer-reviewed papers is -2.61% vs. -1.35% in utility reports. Differences are starker for off-peak consumption. Peer-reviewed papers report an average increase of 1.62% in off-peak consumption (consistent with theory) while utility reports have an average decrease of 0.44%. Additionally, the proportion of documents that report any statistically insignificant effects at all is much larger for utility reports (57%) than peer-reviewed publications (35%). We conclude that estimates for total consumption or off-peak that have a smaller genuine effect (if any) are frequently not reported in academic articles even though they may have been estimated. This observation is consistent with publication bias towards theory-consistent and statistically significant estimates and the suppression of non-significant estimates.

We estimate that accounting for publication bias, the estimated effect of dynamic pricing on peak consumption is -8.7% (95% CI = [-10.5; -6.8]) using PEESE and -10.6% (95% CI = [-13.2; -8.1]) using the top10% estimator. The confidence intervals are overlapping, and the estimates are statistically indistinguishable from the uncorrected average effect of -10.4% (95% CI = [-12.1; -8.7]). The presence of many highly precise estimates may explain why the correction of

the average treatment effect is virtually absent, even though publication bias is present. We estimate a pooled average treatment effect to increase precision but allow for publication bias with regard to the three different types of publications (peer-reviewed academic articles, unpublished academic articles, and utility reports). Separate models for the different types of publication support both the estimation of a pooled average treatment effect and the absence of publication bias for utility reports (see appendix).

The effect of dynamic pricing on total consumption is estimated to be -1.1% (95% CI = [-2.5; 0.4]) using PEESE and -1.5% (95% CI = [-3.0; -0.1]) using the top10% estimator. Confidence intervals are again overlapping, and the estimates are statistically indistinguishable from the uncorrected average treatment effect of -2.6% (95% CI = -4.2; -1.0]). Models separately estimated for the different types of publication tend to support again the pooled estimation of the average treatment effect and the absence of publication bias for utility reports. However, the sample size is smaller, particularly for academic articles and the uncorrected average treatment effect is somehow larger than the estimates corrected for publication bias and the estimates for the utility reports (see appendix). Estimates from the fixed effects models largely confirm the results from the multilevel models (see appendix).

*Table 4: Estimates of publication bias and the average treatment effect by type of demand*

	Total Consumption				Peak Consumption			
		PET	PEESE	Top10%		PET	PEESE	Top10%
Average Effect	-2.6*				-10.4*			
	[-4.2; -1.0]				[-12.1; -8.7]			
Average beyond publication bias		-0.7	-1.2#	-1.5*		-6.8*	-8.7*	-10.6*
		[-2.5; 1.1]	[-2.5; 0.4]	[-3.0; -0.1]		[-9.7; -3.8]	[-10.5; -6.8]	[-13.2; -8.1]
SE: Peer-Reviewed Paper		-23.5*				-68.1*		
		[-42.0; -4.9]				[-105.1; -31.]		
SE: Unpublished Academic Paper		-89.5				-22.4		
		[-284.8; 105.9]				[-54.3; 9.6]		
SE: Utility Report		-1.3				-24.8		
		[-19.7; 17.2]				[-50.9; 1.3]		
			-93.3*				-426.3*	

	Total Consumption				Peak Consumption			
		PET	PEESE	Top10%		PET	PEESE	Top10%
SE <sup>2</sup> : Peer-Reviewed Paper			[-161.0; -25.7]				[-573.2; -279.4]	
SE <sup>2</sup> : Unpublished Academic Paper			-840.3				-75.9	
SE <sup>2</sup> : Utility Report			17.1				-46.0	
			[-43.2; 77.4]				[-99.1; 7.2]	
DF Resid.	128	125	125	113	331	328	328	32
nobs	129	129	129	114	332	332	332	33

Notes: PET = Precision effect test and PEESE = precision-effect estimate with SE. Models estimated using the multilevel model using the *metafor* package in R with robust standard errors clustered at the study level. The square brackets report 95% confidence intervals obtained from bootstrap clustering. The star (\*) denotes that the null hypothesis value (zero) is outside the confidence interval. (#) denotes that the effect has a p-value of 0.1157, and is just outside the significance bounds.

**Reporting and out-of-sample bias.** The results presented above are robust to potential reporting biases and out-of-sample validity of the different pilots. As a part of the critical appraisal, each study was evaluated by the coding team for potential reporting or out-of-sample biases. Following the Cochrane Risk Tool, we categorized each study as ‘Probably Biased’, ‘Probably Not Biased’, or ‘Unclear’ if this information was not available to make this judgment. We calculated the average treatment effect after controlling for such biases. However, the coefficient on both the reporting bias and out-of-sample bias was not statistically significant in any model specification.

## 5. Discussion

In this meta-analysis, we synthesized over 600 observations representing results from trials across the world that use time-varying electricity prices to influence the electricity consumption in households. We differentiate between the impact of such schemes on the peak, total, and off-peak electricity consumption. A primary motivation of the dynamic pricing schemes is to lower the consumption of electricity by households during peak hours. Our meta-analysis suggests that dynamic pricing in all its forms is successful in achieving this aim: the average effect size is -8.7% (95% CI = [-10.5; -6.8]) after correcting for publication bias. This is smaller than the

effect size reported by previous analyses (Faruqui et al., 2017; Faruqui & Sergici, 2011). CPP schemes that entail providing monetary incentives to cut down consumption for short durations achieve the highest reductions in demand. The average effect size is -13.4% (95% CI = [-16.2; -10.6]), which is somewhat smaller than sometimes expected. Rebates, including peak time rebates, are similarly effective but the variation in outcomes is larger. TOU schemes, that are used for shaping the demand curve on a daily basis have an average effect size of -8.0% (95% CI = [-18.6; -5.2]), which is lower than rebates or CPP but can be used for long term management of peak load.

Our analysis also hints at the potential mechanism behind such reductions. Households responding to dynamic pricing can reduce peak consumption by shifting the demand to off-peak hours (peak-shifting) or by reducing the demand as such (peak-shaving). While previous meta-analyses focus on peak demand, we also collect data and study the effect on total and off-peak consumption. We find no statistically significant average effect of dynamic pricing on off-peak consumption for either CPP or rebates and only a small, statistically insignificant increase for TOU. We also detect a small, but statistically significant reduction in total consumption for rebates and TOU. Peak pricing is usually applied for a duration of 4-6 hours in a day. Considering that these hours constitute on average about 20% of the overall consumption in a day, an average reduction of 8.7% in peak consumption implies a 1.7% reduction in total electricity consumption which is very close to our estimate of the average treatment effect. Our analysis, therefore, supports the idea that dynamic pricing leads to peak shaving rather than peak shifting by households.

Our heterogeneity analysis highlights that an important source of peak reductions could be that households reduce their demand from energy-intensive appliances like air-conditioners. The coefficient of cooling is statistically significant and negative for the subsample of peak consumption. Previous meta-analyses have also found that TOU schemes are more successful in areas with cooling demand (Srivastava et al., 2018). But the reduction in peak demand could

have welfare effects for consumers in terms of reducing comfort and health impacts because the reduction in peak seems to be achieved by peak-shaving rather than shifting.

Another important finding of our meta-analysis is that peak consumption is responsive to the size of the monetary incentive, while total consumption is not. We estimate that a 10% increase in the peak-to-baseline price ratio, *ceteres paribus*, reduces peak consumption by 0.47%. This is lower than the estimates provided by Faruqui et al. (2017), potentially because they do not control neither for publication bias nor confounding variables in their analysis. As per our estimated response, an average price ratio of 4:1 is associated with a 6.5% reduction in peak consumption, with diminishing marginal reductions. A relatively high price ratio of 10:1 is associated with a 10% reduction in peak consumption. This reflects to an extent “scope neglect” among households i.e., people pay attention to the fact that there’s a change in price, but not a lot to the size of the change. Given the welfare implications of exposing households to ever higher peak prices, it appears that relatively moderate peak-to-baseline price ratios could be sufficient. Instead of using high monetary incentives, the focus should be on providing smart meters with in-home displays that seem to increase the effectiveness of dynamic pricing.

Our meta-analysis also indicates some concerns about dynamic pricing are probably overemphasized. Implementing dynamic pricing on a large scale does not seem to reduce effectiveness. Our heterogeneity analysis shows that results from studies where households were required to opt-in versus those which are applied to a more general population are quite similar. In fact, several of the studies, especially at the utility level, have been performed at the utility level with large sample and the reductions in peak and total consumption are significant even for these studies. We also do not find that regional context seems to be an important predictor of heterogeneity in results, once other factors are controlled for. One caveat to this finding is that only a few studies in our dataset were from non-OECD countries. The results from non-OECD countries could be systematically different. For example, Capitán et al. (2021) find that in certain country contexts, the absence of energy intensive devices that allow greater control

over demand, behavioral changes to reduce consumption during peak hours may not be enough to offset the increased consumption during off-peak hours, leading to higher overall energy consumption.

Our analysis of the relationship between the higher precision of the studies and their reported results, the so-called publication bias, shows that peer-reviewed publications tend to publish higher effect sizes. We also see that researchers tend to cherry-pick results of peak and total consumption that are of more interest than off-peak consumption results that are reported without bias. The differences in the funnel plots of peak and off-peak consumption are instructive. We do not find such biases in the results reported by the utility reports. The presence of publication bias in the peer-reviewed literature does not reverse the broader conclusions of the meta-analysis, though it could have led to an overestimation of price elasticity in previous analyses.

Our heterogeneity analysis, however, points to two potential factors that limit the effectiveness of dynamic pricing. Firstly, the BMA results weakly support the idea that there could be potential non-complementarities in using moral suasion and monetary incentives for reducing energy consumption in households. More research is required in this area as policy programs tend to take a “more-is-better” approach to interventions in household energy consumption that could be counterproductive. Secondly, studies with a higher treatment duration report lower effect sizes. The effect is particularly strong for the subsample with total consumption, but weak for the sample with peak consumption. This implies that the treatment effects on peak consumption might erode somewhat over time, and the reduction in total consumption may only be seen in experiments with a shorter duration. Long-term evaluations of experiments, therefore, continue to be of value.

We are not able to provide evidence that dynamic pricing can facilitate the decarbonization of the electricity grid. The electrification of heating and transport is expected to add significantly



to electricity demand, especially peak consumption. Dynamic pricing is often thought of as a potential solution to prevent sharp increases in peak demand. We find only weak support for this. The coefficient for heating in the BMA is negative for peak consumption but it is not a strong predictor of heterogeneity. The size of the effect could be small due to the relatively small sample (mostly a few states in the US and Canada) that uses electricity for heating. Direct heating rather than heat pumps coupled with poor insulation could also mean that people are not able to reduce or shift heating demand at present. This may change with heat pumps and better building insulation, but further research is required. This meta-analysis is also limited in its ability to understand the impact of dynamic pricing of electric vehicles on demand as the studies in our data do not specifically target or provide information about households with EVs. Some jurisdictions have introduced specific dynamic tariffs for EVs and the impact of these needs to be studied.

## 6. Conclusions

The literature on dynamic pricing of electricity for households has focused on two distinct advantages of time-varying prices: the gains in economic efficiency and the potential benefits for the integration of renewable energy.

Our meta-analysis shows that dynamic pricing can be an important tool for managing peak consumption of electricity, through reductions in both peak and total consumption of energy. We estimate the average effect size of 8.7%-10.6% reduction in electricity consumption during peak periods and 1.1%-1.5% reduction in total consumption. Our estimates are lower than those in the previous meta-analyses as we account for potential publication bias in the literature. We do not find statistically significant average reductions in off-peak consumption, lending support to the idea of peak shaving rather than demand shifting by households at an aggregate level.

The average effect size for peak consumption is highest for CPP and rebates. While TOU pricing is associated with a more moderate reduction in peak consumption, it can be employed on a more sustained basis. TOU and rebates are also associated with a reduction in total energy consumption.

Our heterogeneity analysis shows that peak consumption is responsive to the peak-to-baseline price ratio, though with diminishing marginal effect. Though, we control for other factors that lead to heterogeneity in results and show that the response is lower than that reported by previous meta-analyses <sup>s</sup>. Total consumption is not responsive to the peak-to-baseline price ratio. We also show that the effectiveness of dynamic pricing does not vary significantly across geographies after controlling for other design aspects though further research is required to understand the effect in non-OECD countries. We do not find evidence that the effectiveness of dynamic pricing decreases with scale.

Dynamic pricing is more effective in the presence of cooling demand, but we are not able to confirm a similar effect for heating demand. Further research is required to understand the response of households with heat pumps and electric vehicles to dynamic pricing. This would be crucial for employing dynamic pricing for renewable energy integration. Additional research on the effects of potential non-complementarities in using moral suasion and monetary incentives for reducing energy consumption in households is also required to ensure the effectiveness of policies.

## 7. Data Availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information and on GitHub. All the information collected in this project is publicly available in line with the systematic reviews reporting protocol (Collaboration

for Environmental Evidence, 2013; Haddaway & Macura, 2018), providing the transparency and reproducibility required to conform with Open Synthesis principles (see the ROSES checklist).

## 8. Code Availability

We used the NACSOS software (Callaghan et al., 2020) for managing search results, removing duplicates, and screening records, the *metafor* package in R (Viechtbauer, 2010) for the meta-analysis and the *bms* package in R for implementing the Bayesian model averaging. All the software packages used are open source and freely accessible. The code developed for the paper and its supplementary information is available on GitHub.

## 9. References

- Aigner, D. J., & Lillard, L. A. (1984a). Measuring peak load pricing response from experimental data: An exploratory analysis. *Journal of Business and Economic Statistics*, 2(1), 21–39. <https://doi.org/10.1080/07350015.1984.10509367>
- Aigner, D. J., & Lillard, L. A. (1984b). Measuring Peak Load Pricing Response from Experimental Data: An Exploratory Analysis. *Journal of Business & Economic Statistics*, 2(1), 21–39. <https://doi.org/10.2307/1391352>
- Allcott, H. (2011). Rethinking real-time electricity pricing. *Resource and Energy Economics*, 33(4), 820–842. <https://doi.org/10.1016/j.reseneeco.2011.06.003>
- Andersen, L. M., Hansen, L. G., Jensen, C. L., & Wolak, F. A. (2019). *Can Incentives to Increase Electricity Use Reduce the Cost of Integrating Renewable Resources* (Working Paper No. 25615). National Bureau of Economic Research. <https://doi.org/10.3386/w25615>
- Bajzik, J., Havranek, T., Irsova, Z., & Schwarz, J. (2020). Estimating the Armington elasticity: The importance of study design and publication bias. *Journal of International Economics*, 127, 103383. <https://doi.org/10.1016/j.jinteco.2020.103383>
- Battalio, R. C., Kagel, J. H., Winkler, R. C., & Winett, R. A. (1979). Residential Electricity Demand: An Experimental Study. *The Review of Economics and Statistics*, 61(2), 180–189. <https://doi.org/10.2307/1924585>
- Belenguer, E., García, N., & Sabater-Grande, G. (2019). Assessment of energy efficiency improvement methods in the residential sector through the development of economic experiments. *SN Applied Sciences*, 1(11), 1409. <https://doi.org/10.1007/s42452-019-1439-7>

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. <https://doi.org/10.1002/9780470743386>
- Borenstein, S. (2009). Time-varying retail electricity prices: Theory and practice. *Electricity Deregulation: Choices and Challenges*, 111–130. <https://doi.org/10.1080/10643389.2012.728825>
- Bruns, S. B., Asanov, I., Bode, R., Dunger, M., Funk, C., Hassan, S., Hauschildt, J., Heinisch, D., Kempa, K., König, J., Lips, J., Verbeck, M., Wolfschütz, E., & Buenstorf, G. (2019). Reporting errors and biases in published empirical findings: Evidence from innovation research. *Research Policy*, 48(9). [https://econpapers.repec.org/article/eeerespol/v\\_3a48\\_3ay\\_3a2019\\_3ai\\_3a9\\_3a25.htm](https://econpapers.repec.org/article/eeerespol/v_3a48_3ay_3a2019_3ai_3a9_3a25.htm)
- Bruns, S. B., & Kalthaus, M. (2020). Flexibility in the selection of patent counts: Implications for p-hacking and evidence-based policymaking. *Research Policy*, 49(1), 103877. <https://doi.org/10.1016/j.respol.2019.103877>
- Buckley, P. (2020). Prices, information and nudges for residential electricity conservation: A meta-analysis. *Ecological Economics*, 172(January), 106635. <https://doi.org/10.1016/j.ecolecon.2020.106635>
- Burkhardt, J., Gillingham, K., & Kopalle, P. K. (2019). Experimental Evidence on the Effect of Information and Pricing on Residential Electricity Consumption. *NBER Working Papers*, Article 25576. <https://ideas.repec.org//p/nbr/nberwo/25576.html>
- Bushnell, J. (2005). Electricity Resource Adequacy: Matching Policies and Goals. *The Electricity Journal*, 18(8), 11–21. <https://doi.org/10.1016/j.tej.2005.08.005>

- Callaghan, M., Müller-Hansen, F., Hilaire, J., & Lee, Y. T. (2020). *NACSOS: NLP Assisted Classification, Synthesis and Online Screening (Version v0.1.0)*.  
<https://doi.org/10.5281/zenodo.4121525>
- Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
- Capitán, T., Alpízar, F., Madrigal-Ballesteros, R., & Pattanayak, S. K. (2021). Time-varying pricing may increase total electricity consumption: Evidence from Costa Rica. *Resource and Energy Economics, 66*, 101264.  
<https://doi.org/10.1016/j.reseneeco.2021.101264>
- Cappers, P. A., & Todd-Blick, A. (2021). Heterogeneity in own-price residential customer demand elasticities for electricity under time-of-use rates: Evidence from a randomized-control trial in the United States. *Utilities Policy, 73*, 101314.  
<https://doi.org/10.1016/j.jup.2021.101314>
- Carroll, J., Lyons, S., & Denny, E. (2014). Reducing household electricity demand through smart metering: The role of improved information about energy saving. *Energy Economics, 45*, 234–243. <https://doi.org/10.1016/j.eneco.2014.07.007>
- Chen, V. L., Delmas, M. A., Locke, S. L., & Singh, A. (2017). Dataset on information strategies for energy conservation: A field experiment in India. *Energy Economics, 16*, 713–716. <https://doi.org/10.1016/j.dib.2017.11.084>
- Collaboration for Environmental Evidence. (2013). Guidelines for systematic review and evidence synthesis in environmental management. *Environmental Evidence*.  
<https://doi.org/www.environmentalevidence.org/Documents/Guidelines/Guidelines4.2.pdf>

- Collaboration for Environmental Evidence. (2013). *Guidelines for Systematic Review and Evidence Synthesis in Environmental Management Version 4.2*.
- DeAngelo, J., Azevedo, I., Bistline, J., Clarke, L., Luderer, G., Byers, E., & Davis, S. J. (2021). Energy systems in scenarios at net-zero CO<sub>2</sub> emissions. *Nature Communications*, *12*(1), 6096. <https://doi.org/10.1038/s41467-021-26356-y>
- Delmas, M. A., Fischlein, M., & Asensio, O. I. (2013). Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. *Energy Policy*, *61*, 729–739. <https://doi.org/10.1016/j.enpol.2013.05.109>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Eicher, T. S., Papageorgiou, C., & Raftery, A. E. (2011). Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*, *26*(1), 30–55. <https://doi.org/10.1002/jae.1112>
- Elevate Energy. (2015). *ComEd Residential Real-Time Pricing (RRTP) Program* [Annual report]. Commonwealth Edison Company.
- Fabra, N., Rapson, D., Reguant, M., & Wang, J. (2021). Estimating the Elasticity to Real-Time Pricing: Evidence from the Spanish Electricity Market. *AEA Papers and Proceedings*, *111*, 425–429. <https://doi.org/10.1257/pandp.20211007>
- Fanelli, D., Costas, R., & Ioannidis, J. P. A. (2017). Meta-assessment of bias in science. *Proceedings of the National Academy of Sciences*, *114*(14), 3714–3719. <https://doi.org/10.1073/pnas.1618569114>

- Faruqui, A., & Sergici, S. (2011). Household Response to Dynamic Pricing of Electricity—A Survey of the Empirical Evidence. *SSRN Electronic Journal*, February.  
<https://doi.org/10.2139/ssrn.1134132>
- Faruqui, A., Sergici, S., & Akaba, L. (2013). Dynamic pricing of electricity for residential customers: The evidence from Michigan. *Energy Efficiency*, 6(3), 571–584.  
<https://doi.org/10.1007/s12053-013-9192-z>
- Faruqui, A., Sergici, S., & Warner, C. (2017). Arcturus 2.0: A meta-analysis of time-varying rates for electricity. *Electricity Journal*, 30(10), 64–72.  
<https://doi.org/10.1016/j.tej.2017.11.003>
- Fernandez, P., Aguirreamalloa, J., & Corres, L. (2012). Market Risk Premium used in 82 countries in 2012: A survey with 7 , 192 answers. *IESE Business School*, 1–17.  
<https://doi.org/10.2139/ssrn.2084213>
- Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2021). Detecting Selection Bias in Meta-Analyses with Multiple Outcomes: A Simulation Study. *The Journal of Experimental Education*, 89(1), 125–144. <https://doi.org/10.1080/00220973.2019.1582470>
- Fowle, M., Wolfram, C., Baylis, P., Spurlock, C. A., Todd-Blick, A., & Cappers, P. (2021). Default Effects And Follow-On Behaviour: Evidence From An Electricity Pricing Program. *The Review of Economic Studies*, 88(6), 2886–2934.  
<https://doi.org/10.1093/restud/rdab018>
- Furukawa, C. (2019). *Publication Bias under Aggregation Frictions: Theory, Evidence, and a New Correction Method* (SSRN Scholarly Paper No. 3362053).  
<https://doi.org/10.2139/ssrn.3362053>



- Gagnon, P., & Cole, W. (2022). Planning for the evolution of the electric grid with a long-run marginal emission rate. *IScience*, *25*(3), 103915.  
<https://doi.org/10.1016/j.isci.2022.103915>
- Goldfarb, B., & King, A. A. (2016). Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal*, *37*(1), 167–176.  
<https://doi.org/10.1002/smj.2459>
- Haddaway, N. R. (2018). Open Synthesis: On the need for evidence synthesis to embrace Open Science. *Environmental Evidence*, *7*(1), 4–8. <https://doi.org/10.1186/s13750-018-0140-4>
- Haddaway, N. R., & Macura, B. (2018). The role of reporting standards in producing robust literature reviews. *Nature Climate Change*. <https://doi.org/10.1038/s41558-018-0180-3>
- Haddaway, N. R., Macura, B., Whaley, P., & Pullin, A. S. (2018). ROSES Reporting standards for Systematic Evidence Syntheses: Pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence*. <https://doi.org/10.1186/s13750-018-0121-7>
- Hammerstrom, D. J., Ambrosio, R., Carlon, T. A., DeSteele, J. G., Horst, G. R., Kajfasz, R., Kiesling, L. L., Michie, P., Pratt, R. G., Yao, M., Brous, J., Chassin, D. P., Guttromson, R. T., Katipamula, S., Le, N. T., Oliver, T. V., & Thompson, S. E. (2008). *Pacific Northwest GridWise? Testbed Demonstration Projects; Part I. Olympic Peninsula Project* (PNNL-17167, 926113; p. PNNL-17167, 926113).  
<https://doi.org/10.2172/926113>
- Hartway, R., Price, S., & Woo, C. K. (1999). Smart meter, customer choice and profitable time-of-use rate option. *Energy*, *24*(10), 895–903. [https://doi.org/10.1016/S0360-5442\(99\)00040-7](https://doi.org/10.1016/S0360-5442(99)00040-7)

- Havranek, T., Irsova, Z., & Janda, K. (2012). Demand for gasoline is more price-inelastic than commonly thought. *Energy Economics*, *34*(1), 201–207.  
<https://doi.org/10.1016/j.eneco.2011.09.003>
- Havranek, T., Rusnak, M., & Sokolova, A. (2017). Habit formation in consumption: A meta-analysis. *European Economic Review*, *95*, 142–167.  
<https://doi.org/10.1016/j.euroecorev.2017.03.009>
- Havránek, T., Stanley, T. D., Doucouliagos, H., Bom, P., Geyer-Klingeberg, J., Iwasaki, I., Reed, W. R., Rost, K., & van Aert, R. C. M. (2020). Reporting Guidelines for Meta-Analysis in Economics. *Journal of Economic Surveys*, *34*(3), 469–475.  
<https://doi.org/10.1111/joes.12363>
- Herter, K., Wood, V., & Blozis, S. (2013). The effects of combining dynamic pricing, AC load control, and real-time energy feedback: SMUD'S 2011 Residential Summer Solutions Study. *Energy Efficiency*, *6*(4), 641–653. <https://doi.org/10.1007/s12053-013-9209-7>
- Hirth, L. (2013). The market value of variable renewables—The effect of solar wind power variability on their relative price. *Energy Economics*, *38*, 218–236.  
<https://doi.org/10.1016/j.eneco.2013.02.004>
- Hofmann, M., & Lindberg, K. B. (2021). Do households react to variable power prices? - Results from a Norwegian pricing experiment. *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, 1–6.  
<https://doi.org/10.1109/ISGTEurope52324.2021.9639976>
- Holland, S. P., Kotchen, M. J., Mansur, E. T., & Yates, A. J. (2022). Why marginal CO<sub>2</sub> emissions are not decreasing for US electricity: Estimates and implications for climate policy. *Proceedings of the National Academy of Sciences*, *119*(8), e2116632119.  
<https://doi.org/10.1073/pnas.2116632119>

- Ioannidis, J. P. A., Stanley, T., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *Economic Journal*, *127*(605), F236–F265.
- Ito, K. (2015). Asymmetric Incentives in Subsidies: Evidence from a Large-Scale Electricity Rebate Program. *American Economic Journal: Economic Policy*, *7*(3), 209–237.  
<https://doi.org/10.1257/pol.20130397>
- Jessoe, K., & Rapson, D. (2014). Knowledge Is (Less) Power: Experimental Evidence from Residential Energy Use. *American Economic Review*, *104*(4), 1417–1438.  
<https://doi.org/10.1257/aer.104.4.1417>
- Joskow, P. L., & Wolfram, C. D. (2012). Dynamic Pricing of Electricity. *American Economic Review*, *102*(3), 381–385. <https://doi.org/10.1257/aer.102.3.381>
- Khanna, T. M. (2022). Using agricultural demand for reducing costs of renewable energy integration in India. *Energy*, *254*, 124385.  
<https://doi.org/10.1016/j.energy.2022.124385>
- Khanna, T. M., Baiocchi, G., Callaghan, M., Creutzig, F., Guías, H., Haddaway, N. R., Hirth, L., Javaid, A., Koch, N., Laukemper, S., Löschel, A., Zamora Dominguez, M. del M., & Minx, J. C. (2021). A multi-country meta-analysis on the role of behavioural change in reducing energy consumption and CO2 emissions in residential buildings. *Nature Energy*, *6*(9), 925–932. <https://doi.org/10.1038/s41560-021-00866-x>
- Lifson, D. P., & Miedema, A. K. (1981). A comparative analysis of time-of-use electricity rate effects: The Arizona experiment. *Energy*, *6*(5), 403–408. [https://doi.org/10.1016/0360-5442\(81\)90003-7](https://doi.org/10.1016/0360-5442(81)90003-7)
- Löschel, A., Rodemeier, M., & Werthschulte, M. (2020). When Nudges Fail to Scale: Field Experimental Evidence from Goal Setting on Mobile Phones. In *CESifo Working Paper Series 8485* (Issue August). CESifo. <https://doi.org/10.2139/ssrn.3693673>

- Matsukawa, I. (2018). Information acquisition and residential electricity consumption: Evidence from a field experiment. *Resource and Energy Economics*, *53*, 1–19.  
<https://doi.org/10.1016/j.reseneeco.2018.02.001>
- Midden, C. J. H., Meter, J. F., Weenig, M. H., & Zieverink, H. J. A. (1983). Using feedback, reinforcement and information to reduce energy consumption in households: A field-experiment. *Journal of Economic Psychology*, *3*(1), 65–86.  
[https://doi.org/10.1016/0167-4870\(83\)90058-2](https://doi.org/10.1016/0167-4870(83)90058-2)
- Mier, M., & Weissbart, C. (2020). Power markets in transition: Decarbonization, energy efficiency, and short-term demand response. *Energy Economics*, *86*, 104644.  
<https://doi.org/10.1016/j.eneco.2019.104644>
- Montigel, B. (2022). *Smart meter deployment and dynamic electricity pricing in Europe: Empirical evidence from Germany, Sweden, Italy, and the Netherlands*.
- Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Sánchez-Tójar, A., Yang, Y., & O’Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods in Ecology and Evolution*, *13*(1), 4–21. <https://doi.org/10.1111/2041-210X.13724>
- Pellerano, J. A., Price, M. K., Puller, S. L., & Sánchez, G. E. (2017). Do Extrinsic Incentives Undermine Social Norms? Evidence from a Field Experiment in Energy Conservation. *Environmental and Resource Economics*, *67*(3), 413–428.  
<https://doi.org/10.1007/s10640-016-0094-3>
- Pon, S. (2017). The Effect of Information on TOU Electricity Use: An Irish residential study. *The Energy Journal*, *38*(01). <https://doi.org/10.5547/01956574.38.6.spon>
- Potter, J. M., George, S. S., & Jimenez, L. R. (2014). *SmartPricing Options Final Evaluation* (SmartPricing Options). Sacramento Municipal Utility District.

- Pustejovsky, J. E., & Tipton, E. (2018). Small sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, *36*(4), 672–683. <https://doi.org/10.1080/07350015.2016.1247004>
- Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association*, *92*(437), 179–191. <https://doi.org/10.2307/2291462>
- Ringquist, E. J. (2013). *Meta-analysis for Public Management and Policy*. John Wiley & Sons.
- Sexton, R. J., & Sexton, T. A. (1987). Theoretical and Methodological Perspectives on Consumer Response to Electricity information. *Journal of Consumer Affairs*, *21*(2), 238–257. <https://doi.org/10.1111/j.1745-6606.1987.tb00201.x>
- Slavin, R. E., Wodarski, J. S., & Blackburn, B. L. (1981). A group contingency for electricity conservation in master-metered apartments. *Journal of Applied Behavior Analysis*, *14*(3), 357–363. <https://doi.org/10.1901/jaba.1981.14-357>
- Srivastava, A., Van Passel, S., & Laes, E. (2018). Assessing the success of electricity demand response programs: A meta-analysis. *Energy Research and Social Science*, *40*(December 2017), 110–117. <https://doi.org/10.1016/j.erss.2017.12.005>
- Stanley, T. D. (2005). Beyond Publication Bias. *Journal of Economic Surveys*, *19*(3), 309–345. <https://doi.org/10.1111/j.0950-0804.2005.00250.x>
- Stanley, T. D. (2008). Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection\*. *Oxford Bulletin of Economics and Statistics*, *70*(1), 103–127. <https://doi.org/10.1111/j.1468-0084.2007.00487.x>

- Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60–78.  
<https://doi.org/10.1002/jrsm.1095>
- Stanley, T. D., & Doucouliagos, H. (2015). Neither fixed nor random: Weighted least squares meta-analysis. *Statistics in Medicine*, 34(13), 2116–2127.  
<https://doi.org/10.1002/sim.6481>
- Stanley, T. D., Jarrell, S. B., & Doucouliagos, H. (2010). Could It Be Better to Discard 90% of the Data? A Statistical Paradox. *The American Statistician*, 64(1), 70–77.  
<https://doi.org/10.1198/tast.2009.08205>
- Viechtbauer, W. (2010). Conducting meta-analysis in R with metafor package. *Journal of Statistical Software*, 36(3), 1–48.
- Wolak, F. A. (2011). Do Residential Customers Respond to Hourly Prices? Evidence from a Dynamic Pricing Experiment. *The American Economic Review*, 101(3), 83–87.
- Zeugner, S., & Feldkircher, M. (2015). Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R. *Journal of Statistical Software*, 68, 1–37.  
<https://doi.org/10.18637/jss.v068.i04>



Mercator Research Institute on

MCC was founded jointly by Stiftung  
Mercator and the Potsdam Institute  
for Climate Impact Research.