Khanna, Tarun M. et al.

Working Paper

Meta-analysis on the efficacy of behavioral, information and economic interventions in increasing energy-efficiency adoption among households

Ruhr Economic Papers, No. 1141

Provided in Cooperation with:
RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Khanna, Tarun M. et al. (2024) : Meta-analysis on the efficacy of behavioral, information and economic interventions in increasing energy-efficiency adoption among households, Ruhr Economic Papers, No. 1141, ISBN 978-3-96973-324-0, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen, https://doi.org/10.4419/96973323

This Version is available at:
https://hdl.handle.net/10419/315481

RUHR
ECONOMIC PAPERS

Tarun M. Khanna
Diana Danilenko
Lukas Tomberg
Sven Hansteen
Mark A. Andor
Paul M. Lohmann
Jan C. Minx

Meta-Analysis on the Efficacy of Behavioral, Information and Economic Interventions in Increasing Energy-Efficiency Adoption Among Households

RUHR UNIVERSITÄT BOCHUM RUB RWI #1141

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Ludger Linnemann
Technische Universität Dortmund, Department of Business and Economics
Economics – Applied Economics
Phone: +49 (0) 231/755-3102, e-mail: Ludger.Linnemann@tu-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83 - 3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Almut Balleer, Prof. Dr. Manuel Frondel,
Prof. Dr. Ansgar Wübker
RWI, Phone: +49 (0) 201/81 49 - 213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #1141

Responsible Editor: Manuel Frondel

ISSN 1864-4872 (online) – ISBN 978-3-96973-324-0

Ruhr Economic Papers #1141

Tarun M. Khanna, Diana Danilenko, Lukas Tomberg, Sven Hansteen,
Mark A. Andor, Paul M. Lohmann, and Jan C. Minx

Meta-Analysis on the Efficacy of Behavioral, Information and Economic Interventions in Increasing Energy-Efficiency Adoption Among Households

RUHR UNIVERSITÄT BOCHUM RUB RWI

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

https://dx.doi.org/10.4419/96973324
ISSN 1864-4872 (online)
ISBN 978-3-96973-323-3

Tarun M. Khanna, Diana Danilenko, Lukas Tomberg, Sven Hansteen,
Mark A. Andor, Paul M. Lohmann, and Jan C. Minx*

Meta-Analysis on the Efficacy of Behavioral, Information and Economic Interventions in Increasing Energy-Efficiency Adoption Among Households

Abstract

Energy efficiency is often cited as a critical component of mitigation pathways that avoid the worst effects of climate change but suffers from chronic underinvestment. This paper evaluates the efficacy of a range of voluntary approaches—monetary incentives, information and behavioral interventions—on the willingness of households to pay for energy efficient appliances, the market share of efficient appliances, and the subsequent savings in energy consumption to understand which interventions work, under what conditions, and why. We find that information provision, labeling, rebates and subsidies increase willingness to pay for efficient appliances moderately, while loans are ineffective. The effects of such interventions on market shares and associated rebound effects on energy consumption of purchase of energy-efficient appliances remain unclear given the limited evidence. Closing this should be a priority to facilitate better understanding of the role of such interventions in climate and energy policy. Real-world effects are also likely to be smaller than those reported due to study design limitations and potential reporting biases. Overall, the existing evidence does not strongly support the effectiveness of these interventions in achieving large-scale energy efficiency improvements required for decarbonization.

JEL-Codes: Q58, Q48, D11, D83

Keywords: Energy efficiency; meta-analysis; energy conservation; behavioral interventions; economic incentives; policy instruments

December 2024

* *Tarun M. Khanna, University of British Columbia; Diana Danilenko, Potsdam Institute for Climate Impact Research; Lukas Tomberg, RWI; Sven Hansteen, RWI; Mark A. Andor, RWI and RUB; Paul Lohmann, University of Cambridge; Jan C. Minx, Potsdam Institute for Climate Impact Research. – All correspondence to: Tarun M. Khanna, e-mail: tarun.khanna@ubc.ca*

Introduction

Energy efficiency is often cited as a critical component of mitigation pathways to keep global warming to well-below 2 or even 1.5 degrees [1], [2], [3], [4]. In fact, lower levels of energy consumption are a key feature of scenarios that tend to perform better in achieving other SDGs and limit transitional risks [5], [6]. For example, the International Energy Agency's Net Zero Roadmap for 2050 requires doubling the annual global primary energy intensity improvement rate from just over 2% in 2022, to slightly above 4% until 2030. At least a third of the improvements required to reach this key figure is expected to come from improvements to technical efficiency [7]. In addition to structural changes to buildings and the adoption of electric vehicles, the IEA's scenario highlights the use of more efficient appliances such as refrigerators and air conditioners as a key action.

While energy efficiency is widely regarded as a least cost mitigation option, investments in efficiency are plagued by the "energy efficiency gap" [8]: many households and firms do not sufficiently invest in energy efficiency even though such investments promise substantial (monetary) benefits (see Gerarden et al., 2017 [9] for a review). The reasons underlying the energy efficiency gap are diverse, such as a potentially long-time horizon until such investments pay off, risk aversion, hassle costs, information asymmetries, or the lack of adequate pricing of externalities. Unsurprisingly, many governments adopt a diverse set of policies aimed at increasing the willingness to invest in energy efficiency including monetary incentives, informational strategies (e.g., information provision, label, audits) and behavioral interventions (e.g., choice architecture, feedback) [8], as such policies hold the potential of simultaneously achieving private and public benefits.

Understanding which of these different policy interventions work, under what conditions, and why is critical for climate and energy policy design as well as advancing research and the underlying evidence base in the field [10]. This requires assessing the available science and providing the best evidence to decision-makers – evidence that is rigorous, transparent, timely and fit-for-purpose [11]. Evidence synthesis methods such as systematic reviews and meta-analyses have been designed for the job by minimizing the different sources of bias common in traditional literature reviews [12].

There have been relatively few attempts to assess the evidence in the existing scientific literature around the energy efficiency gap and the potential of different interventions to address it. Traditional literature reviews in economics have described the economics of energy efficiency and used results from selected studies to estimate the size of the energy efficiency gap [13], [14], [15]. A range of systematic reviews and meta-analyses have evaluated the role of monetary and non-monetary interventions for fostering energy conservation or pro-environment behavior [16], [17], [18], [19], [20]. Andor and Fels, 2018 [21] qualitatively reviewed studies that evaluated the effect of labels on willingness to pay for energy efficient appliances, energy saving potential, and hypothetical purchase decisions regarding energy efficient appliances or actual energy consumption. Nisa et al., 2019 [22] conducted a meta-analysis of the impact of behavioral interventions such as appeals and nudges on various energy-related outcomes, with only four studies on the outcome of energy efficiency adoption. Other reviews have focused on the effect of choice architecture on a range of behaviors [23]. With existing literature focusing primarily on actions with short-term consequences, implications of behavioral interventions for fostering decisions with long-term outlook like energy efficiency investments are not very well understood [8].

Here, we provide a comprehensive, rigorous, machine-learning enhanced systematic review and meta-analysis on the impact of interventions in uptake and use of energy efficient appliances. No previous reviews have focused specifically on appliances even though air-conditioners, refrigerators and freezers, lighting, water heaters, ovens, and washing machines account for up to 55% and 65% of the residential electricity consumption in the European Union [24] and in the US [25] respectively. Governments

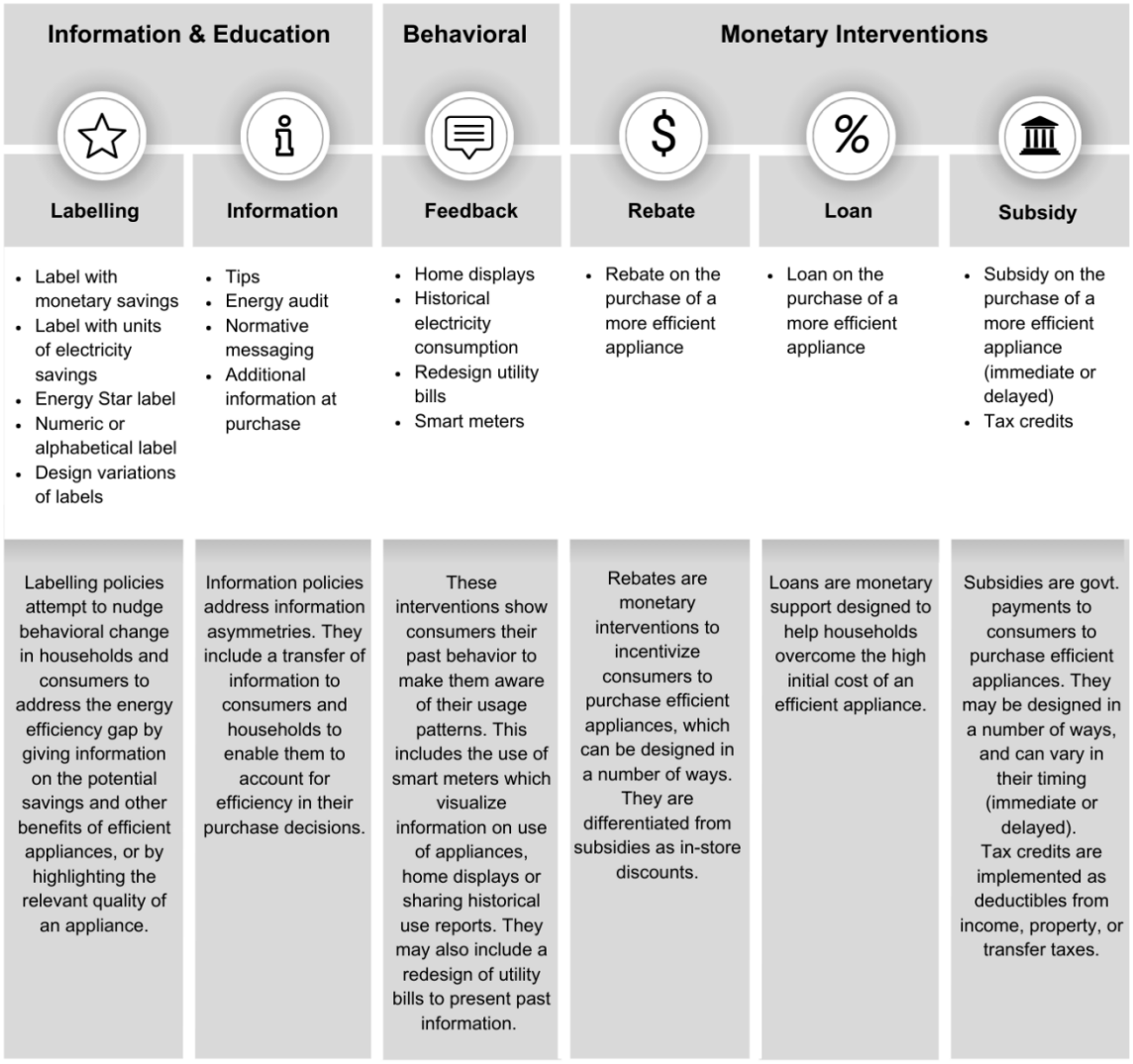
allocate substantial financial resources, besides organizational capacity and time, towards promoting high efficiency home appliances and equipment in households. For example, the Inflation Reduction Act in the US alone has set aside USD 4.5 billion in funding for this[26]. Across the world, governments have allocated USD 264 billion since 2020 towards energy efficient buildings, appliances and net-zero buildings[27]. Incentivizing sales of efficient appliances has been a focus area and has led to considerable experimentation with intervention designs. The knowledge accumulated from such experiments and implemented policies can help to gain a deeper understanding of household investments in energy efficiency, which can also be applied to other areas of residential energy demand beyond appliances.

The academic literature is divided into three strands: studies that evaluate the impact of behavioral, information and economic interventions on willingness to pay for efficient appliances, impact of these interventions on overall sales share of such appliances, and the energy savings following the purchase of efficient appliances. We systematically review this literature together, using a mix of quantitative meta-analysis and narrative review, to create a comprehensive understanding of the field. The results of our meta-analysis indicate that, on average, interventions targeting individual purchase decisions have a moderate treatment effect of 0.06 to 0.14, depending on the respective intervention. But the analysis of heterogeneity in reported effects using novel Bayesian Model Averaging methods suggests that the effect in the real world is likely to be smaller as studies with self-reported behavior and stated preferences exaggerate the impact of interventions. Studies that consider the impact of interventions on observed market share of efficient appliances report very heterogeneous effects. Lastly, studies that measure households' energy use post adoption of efficient appliances suggest reduction in energy consumption by 2%-12% but three studies that consider endogeneity in household behavior report an actual increase in consumption. Overall, these findings together suggest that information, monetary and behavioral interventions are only partially able to motivate the purchase of more efficient household appliances and can only be seen as one component of a broader policy mix aimed at achieving more ambitious energy efficiency and climate change targets.

A diverse and clustered evidence base

This review follows state-of-the-art systematic review guidelines and open-synthesis principles [28], [29] (see Methods and the pre-registered review protocol[30]) to answer the following review questions: are the interventions targeting the adoption of efficient appliances by households generally effective? Which category of interventions is most effective? How does energy consumption change after adoption of efficient appliances through such interventions? The interventions used in the studies were classified into information & education (information provision, label, audits); behavioral interventions (choice architecture, feedback); and monetary incentives (tax credits, rebates, subsidies) following previous studies in energy use among households [16], [17], [18] (Figure 1). This classification has been adopted across a series of reviews and meta-analyses that form an "ecosystem of reviews" that evaluate interventions in households' consumption choices in food, transport and buildings (see Supplementary Note 1 for papers included in the "ecosystem").

Figure 1: Overview of the primary studies classified by the key dependent variable, appliance and intervention. Figure in parentheses is the corresponding number of effect sizes.



We searched for relevant studies in the five major bibliographic databases (Web of Science, Scopus, JSTOR, RePec and Google Scholar) returning over 30,000 unique records after deduplication. We then applied a machine-learning-enabled prioritized screening approach[31], [32] to search for relevant experimental and quasi-experimental studies (see Methods) that is able to find considerably more relevant evidence than conventional systematic as well as traditional reviews[16]. Overall, we identified and coded a total of 58 relevant studies after full-text screening and extracted 278 effect sizes from them, corresponding to an average of about five effects per study.

No previous reviews have focused specifically on appliance purchase and related energy savings, apart from Nisa et al. 2019 [22] which included four studies on this outcome variable (Supplementary Table 3). The identified studies differ in three important dimensions: the dependent variable studied, the intervention used, and the appliance considered. As shown in Table 1, 44 studies examine a household's or an individual's decision to purchase energy efficient appliances, while only five look at the aggregate effects on the market or sales share of efficient appliances. Nine studies report on the energy consumption of households after purchase of efficient appliances following such interventions. Evidence is also scattered unevenly across interventions. Though studies often use a combination of interventions, the most widely used intervention in the literature is labeling followed by other

information strategies and monetary incentives. Refrigerators are the most frequently studied appliance in our sample, along with other high energy use appliances like water heaters and air conditioners. Clothes dryers, dishwashers, and TVs noticeably remain under-investigated, and some evidence is available on newer appliances like heat pumps. Our approach to reviewing the evidence across the three outcomes was informed by the number of studies that we identified. We conducted a quantitative meta-analysis of the studies on purchase decisions given the relatively large sample size and a narrative review of the small sample of market-share and energy-use studies.

Table 1: Overview of the primary studies classified by the key dependent variable, appliance, and intervention. Figure in parentheses is the corresponding number of effect sizes; it is sorted according to the number of effects sizes for each dependent variable.

	Probability of purchase, Willingness-to-pay	Market share/ aggregate effects	Energy use
Total	44 studies (215)	5 studies (27)	9 studies (36)
Approach to synthesis	Meta-analysis	Descriptive review	Descriptive review
Interventions	Label (77) Information (68) Rebate (31) Subsidy (27) Loan (12)	Rebate (15) Label (12)	Label (12) Rebate (11) Standards (7) Subsidy (4) Information (2)
Appliances	Refrigerator (61) Water Heater (43) Air-Conditioner (35) Lighting (23) Heat Pump (18) Washing Machine (13) TV (12) Clothes Dryer (6) Dishwasher (4)	Refrigerator (16) Washing Machine (4) Dishwasher (4) Air-Conditioner (3)	Air-Conditioner (21) Heat Pump (5) Lighting (4) Dishwasher (2) Refrigerator (2) TV (2)

The sample of studies included in the quantitative analysis represents research on a total of 162,892 households/individuals and it encompasses experiments and programs in 25 countries. The summary statistics for the dataset included in the meta-analysis are presented in Supplementary Table 5. The earliest studies date back to the early 1980s, but around half of the sample is from studies conducted after 2016. Studies are spread around the globe: about 34% of the effect sizes come from studies in Asia, 30% from continental Europe and UK, 24% from North America, and the remaining 12% from Africa, Oceania and cross-country studies. Our focus on studies published in English only is an important limitation of this review. Included studies rely on multiple statistical techniques, with variations of logit being the most common, and multiple study designs, where about half of the observations come from quasi-experimental studies, a third come from randomized control trials, and the rest from pre- and post-studies. An important aspect of the study design is whether the results are based on revealed behavior of households and individuals, for example the actual purchase of an appliance, or whether the study asks the participants to self-report actions that they took, or if the study only reports the stated preference of the participants. About half the observations in our sample are based on stated preferences, 32% on revealed preferences and the remaining 20% on self-reported behavior. Reliance on stated preferences is a limitation of the literature, and we investigate its consequences in our analysis.

Results

Impact of interventions on purchase decisions

Quantitative synthesis of effect sizes

Forty-four primary studies in our sample investigate the probability of purchasing, or willingness-to-pay for, an efficient appliance over standard appliances, with the exact specification varying across studies. We first standardized the effects by converting the estimates reported for each study to Fisher's Z and then used meta-analysis models to calculate the average effect across studies and the causes for heterogeneity in results across studies (see Methods). We apply a multi-level random effects model that allows us to deal with dependent effect sizes within and between studies flexibly.

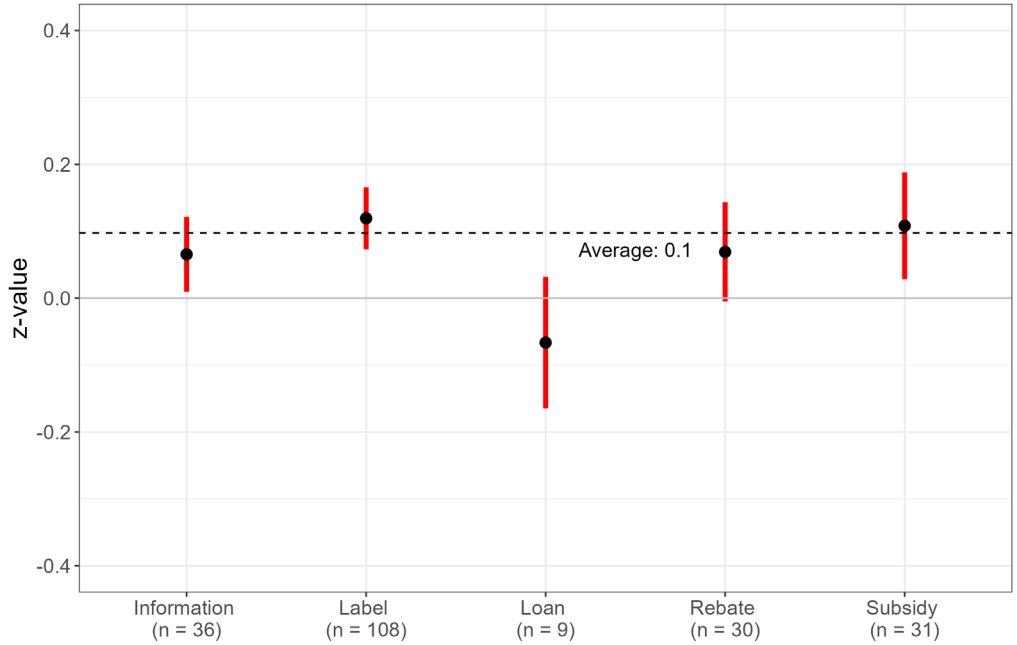
We find an overall moderately sized, average treatment effect size of 0.10 [95% CI = (0.06; 0.14)] across interventions. The average effect sizes for information-based interventions (0.07 [95% CI = (0.01; 0.12)] for information and 0.12 [95% CI = (0.07; 0.17)] for labeling) are broadly comparable to monetary incentives (0.07 [95% CI = (0.00; 0.14)] for rebates and 0.11 [95% CI = (0.03; 0.19)] for subsidies) with overlapping confidence intervals (Figure 2). Loan is the only intervention category for which the average effect size is negative and not statistically significant. Only three studies provide estimates for this category and all report no significant effect of loan on stimulating purchase of efficient appliances (ref. Revelt & Train (1998), Train & Atherton (1995), Liu & Jin (2019)). The confidence intervals are in general larger for rebates and subsidies, due to the lower number of effect sizes compared to information-based interventions and a greater variation in results with several studies reporting negative treatment effects. The average treatment effects were estimated after controlling for small-study bias and potential outliers (see Methods). Not adjusting for publication bias and outliers does not alter the results in a substantive manner, except that the average effect sizes are larger for all categories of interventions and the average effects for rebates and subsidies have larger confidence intervals (see Supplementary Table 6).

Previous meta-analyses on the impact of interventions on other areas of household energy consumption have reported similar average effects but higher effects for monetary incentives as compared to behavioral and informational strategies[16], [17] but we do not find any statistical or substantive differences between these types of interventions here. A Wald test is not able to reject the null hypothesis that the effects are equivalent. On the one hand, it could be that this difference is attributable to the different contexts of these reviews. For example, in the energy sector, specific price-based approaches such as time-of-use and critical peak pricing can be used, which could potentially be more effective than the rewards, rebates and other monetary incentives that have been studied in the context of household appliance purchasing decisions.

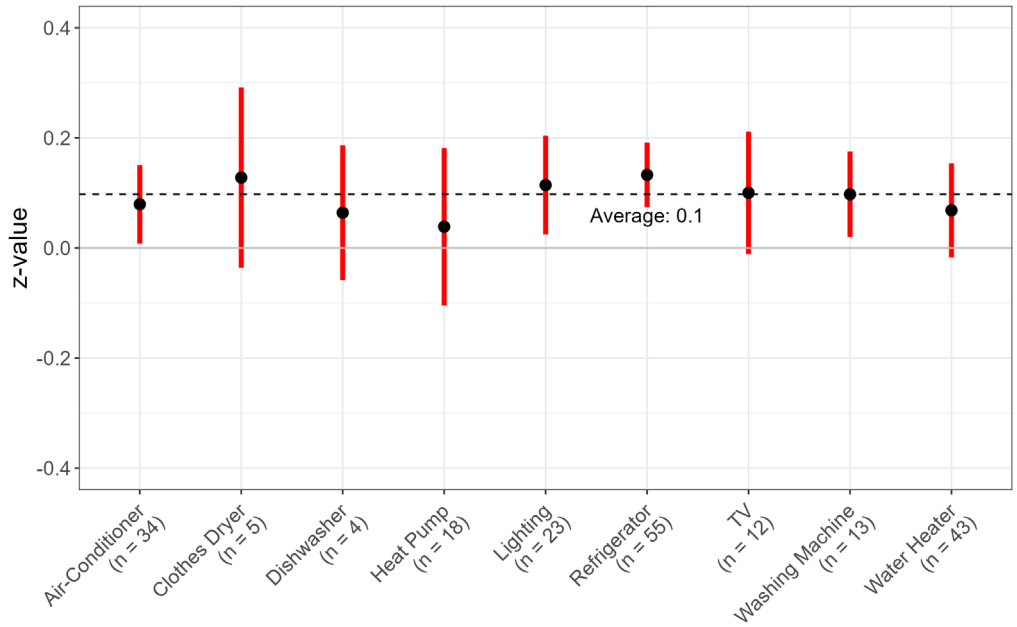
Amongst the appliance categories, refrigerators, lighting and washing machines have the strongest treatment effects that are also statistically significant (Figure 2). The average effect size for other high electricity consumption appliances, namely water heaters, heat pumps and air-conditioners, are weak and not statistically significant after controlling for small-study bias. The differences across appliance categories, irrespective of energy intensity or cost of the appliance, are however not substantive and not statistically significant.

Figure 2: Average effect size along with the 95% confidence interval across interventions and appliances. Z > 0 implies an increase in WTP or probability of purchase. The results are from a multilevel meta-regression model that accounts for publication bias and influential observations.

(a) Average effect size across intervention categories



(b) Average effect size across different categories of appliances studies



Heterogeneity

There is considerable unexplained heterogeneity in effect sizes that we extract from primary studies (I^2 = 97-98% for random effects model with intervention/appliance as moderator variable). Estimates of the effect size may vary because of variables that affect the willingness to pay for appliances, e.g. type of intervention and appliance, location of the experiment, type of participant (homeowners vs. tenants),

etc. or variables that represent the design elements or biases contained in primary studies (e.g., setting of the experiment, statistical techniques used for analysis, reporting biases, etc.) and included those also in the meta-analysis. We employed the state-of-the-art hybrid-frequentist Bayesian approach to identify the most important variables that could explain the heterogeneity in results out of a set of 16 potentially relevant variables and their 2^{16} possible combinations. We employ variable selection based on Bayesian Model averaging (specifically, the variables with posterior inclusion probability above 50% are selected; see Methods) and then use these identified variables in a frequentist model with clustered standard errors to estimate the effect of variables on variation in effect sizes (results in Table 2).

The heterogeneity analysis confirms that there is no statistical difference between the average effect of interventions except "Loans" that have lower effect. There is some variation in the impact of these interventions by the appliance type: refrigerator and lighting have the highest effects even after controlling for other parameters. Interestingly, we find that homeowners are more likely to purchase energy efficient appliances as compared to only "residents" (Table 2). None of the other variables that could drive heterogeneity in results including location, energy prices, demographics of the household, environmental attitudes, or the year of publication were identified as being statistically significant.

We find that the design elements of original studies are statistically significant in explaining differences in outcomes reported. Studies that measured the impact of the interventions in terms of changes in self-reported behaviors or stated preferences of the subjects of the study reported higher impact as compared to studies that measured the change in actual purchase behavior in response to the interventions. The differences resulting from studies that employ the different incentive-compatibility designs are statistically significant (Table 2). This result is expected for two primary reasons. First, customers might have a tendency to say that they would purchase a high-efficiency appliance more readily than they actually do. Second, any time or effort that the customer must expend to receive a rebate is not reflected in the stated- preference data[33].

Critical appraisal of primary studies

We also carried out critical appraisal of the evidence by: (1) a qualitative assessment of the studies for reporting and sampling biases, and (2) assessment of publication bias and test whether study quality has a relationship with heterogeneity in the reported treatment effects.

Table 2: Results from meta-regression based on moderator variables identified through Bayesian Model Averaging.

Variable	Beta coefficient	Standard Error	p-value
Participant type (Base: Residents)			
Homeowners	0.09*	0.05	0.10
Unspecified	0.18***	0.06	0.00
Intervention type (Base: Information)			
Label	- 0.00	0.06	0.93
Loan	- 0.09	0.06	0.14
Rebate	0.13	0.09	0.17
Subsidy	0.08	0.07	0.25
Appliance (Base: Refrigerator)			
Clothes Dryer	- 0.18*	0.11	0.10
Dishwasher	- 0.11	0.06	0.11
Heat Pump	- 0.27***	0.08	0.00

Variable	Beta coefficient	Standard Error	p-value
Lighting	0.03**	0.06	0.59
Air-Conditioner	-0.14**	0.07	0.04
TV	- 0.25**	0.08	0.00
Washing Machine	- 0.07	0.08	0.35
Water Heater	- 0.12**	0.06	0.04
Incentive compatibility (base: Actual behavior)			
Self-reported Behavior	0.09	0.07	0.24
Stated preferences	0.19*	0.11	0.10
Risk of Bias (Base: Probably No)			
Reporting Bias: "Probably Yes"	0.10**	0.05	0.05
Out of Sample Bias: "Probably Yes"	0.02	0.04	0.62
Square of Standard Error of Z	2.53***	0.30	0.00

*Note: The dependent variable is the standardized Fisher's Z calculated for each effect size reported by studies. Z > 0 indicates a higher willingness to pay for energy efficient appliances. The model estimated is an Ordinary Least Squares with clustered standard errors (see Methods for details). ***p <= 0.01; **p <= 0.05; *p <= 0.10.*

We test the presence of publication bias and estimate the 'true' average effect size adjusted for publication bias using a graphical analysis (funnel plot) and statistical tests in the form of Egger's regression (see Supplementary Note 2 for details). We confirm the presence of publication bias in the literature, with more studies than expected reporting a positive effect. Adjusting the estimates of the average treatment effect for publication bias, the "true" mean effect is estimated to be 0.10 which is slightly lower than the unadjusted mean effect of 0.11 for all interventions taken together reported in Figure 2. These results are also confirmed by the heterogeneity analysis: the square of standard error of Z is statistically significant even after controlling for other possible sources of heterogeneity. We, therefore, account for publication bias in all the estimates provided in the main text, including those in the previous sub-sections.

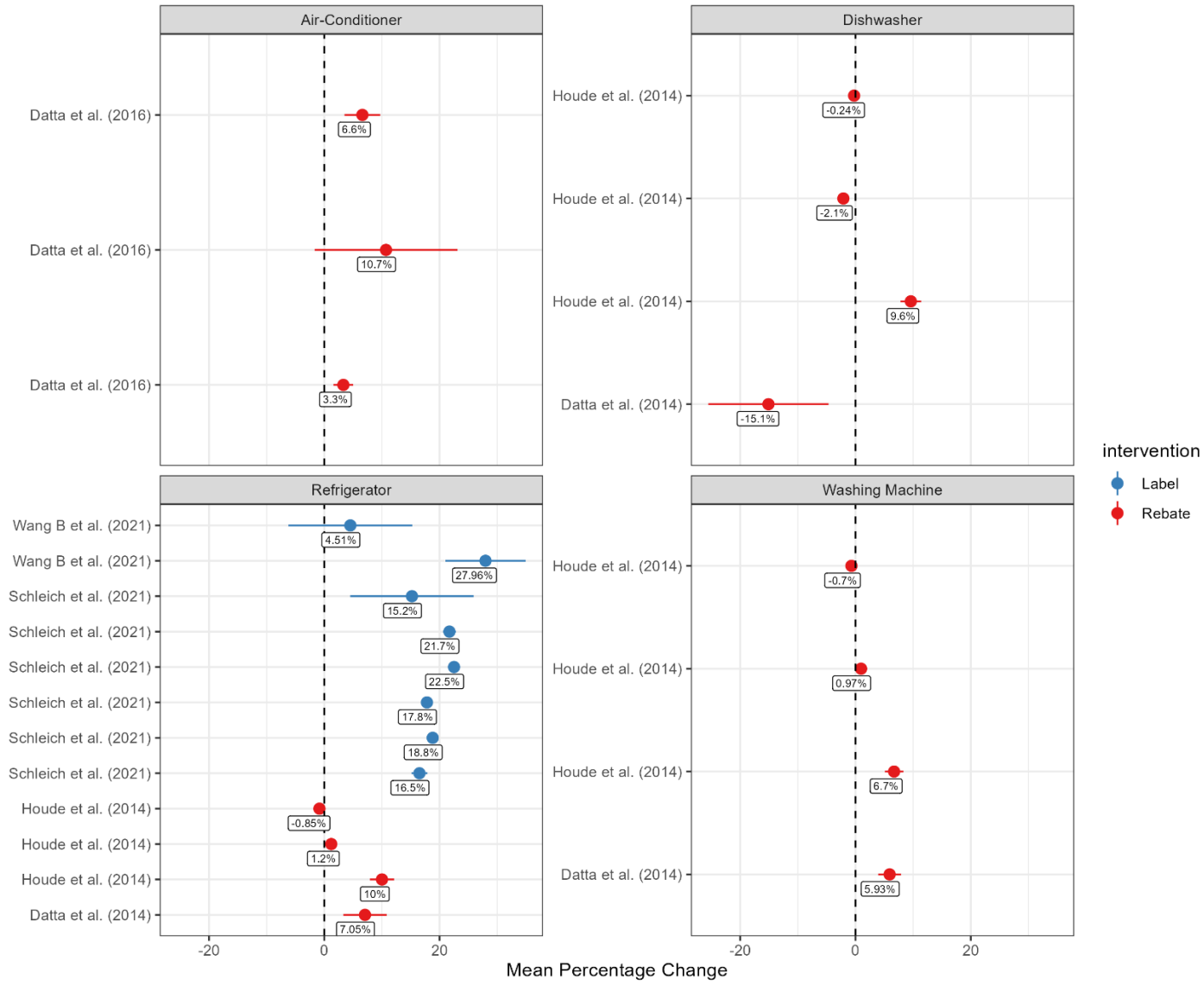
Impact of interventions on market share of efficient appliances

Beyond the literature on purchase probability and willingness to pay, five studies also investigate the impact of information and economic interventions from different energy conservation programs on the overall market share of efficient appliances in the US and Europe using econometric analysis. These studies provide valuable insight into the impact of real-world policies that scale up interventions in an effort to promote energy efficiency. Two studies focus on the ENERGY STAR program that was introduced in 1992 by the United States Environmental Protection Agency (EPA) as a voluntary labeling program designed to promote the use of energy-efficient products and thus help to reduce the emissions of greenhouse gases. Federal and local governments and utility companies across the US promote the adoption of ENERGY STAR-labeled products by offering financial incentives, usually in the form of mail-in rebates. Dutta and Filippini[34] estimate that the rebates increase the sales share of ENERGY STAR appliances between 9%-18% based on average market share. The results are higher than the findings in Datta and Gulati[35] who find that, on average, rebates increase the sales of ENERGY STAR washing machines by 6% but do not have statistically significant impact on sales share of dishwashers and refrigerators. Dutta and Filippini use an instrument variable approach to account for the endogeneity of the rebate policy itself, so the results may be considered more robust than Dutta and Gulati.

Houde and Aldy[36] evaluate another program in the US—the State Energy Efficient Appliance Rebate Program (SEEARP)—under which rebates are given to consumers for purchasing residential appliances that met or exceeded the ENERGY STAR certification requirement. They evaluate the impact of the rebate program on sales of efficient star rated refrigerators, dishwashers and washing machines and the subsequent reduction in household energy consumption. They find that rebate programs increased efficient appliance sales by 7%-10% during rebate period, which translates into a market share increase of energy star appliances of 1-2 percentage points. But they find that about 70% of the consumers who claimed the rebate would have bought the efficient appliance anyway, so rebates did not have a strong impact on attracting new consumers. An additional 15%-20% of consumers changed the timing of their purchase of an efficient appliance by a few weeks.

Schleich, Durand and Brugger[37] investigate the effectiveness of minimum energy performance standards (MEPS) and energy labels, which member states of the EU and several other countries rely on. While MEPS eliminate the worst performing products from the market, labeling schemes help consumers to make more informed choices. MEPS and energy labels are therefore considered pillars of the EU's strategy to achieve its energy efficiency and climate policy targets and studies that model energy demand pathways consider substantial reductions of 40%-60% in energy demand of appliances[38]. They find that MEPS and the energy labels together increased the sales share of cold appliances with an energy label of A+ and better by about 15-38 percentage points. However, they are unable to disentangle the contribution of the standards and energy labels. They also attempt to estimate the increase in market share of efficient appliances that would have taken place without any policy intervention, for example because of technological progress or societal trends. They estimate these autonomous effects to range between 5-10 percentage points per year for the change in market shares, which would imply significant overestimation of the effects of policy interventions. This result could be suggestive for more substantial effects when the various interventions studied here are combined with other regulatory instruments, but we caution not to jump to conclusions due to study limitations and a lack of additional evidence. As such, this could be an important area for future research that could be key to understanding the potential of behavioral interventions in their specific institutional and regulatory context.

Figure 3: Forest plot of studies with the dependent variable as change in sales share of efficient appliances.



Changes in energy consumption after purchase of efficient appliances

Increasing the sale of energy efficient appliances is only a means to an end. Energy-efficient appliance adoption needs to translate into actual reductions in energy consumption to reduce energy related household carbon emissions. Literature on achieved reductions in energy consumption is therefore a second complementary strand of research. Theoretical and empirical literature[8], [39] has highlighted the possibility of so-called (direct) rebound effects, i.e., efficiency gains reduce the effective price seen by users by increasing the energy services provided by a unit of energy, so they tend to increase physical energy use above what simple engineering calculations would predict. If the rebound effect is strong enough, it could "backfire" and increase, rather than decrease, energy consumption. We identified nine studies that evaluate the energy use of households after replacement of standard appliances with efficient appliances. These studies employ a diverse set of methodologies and data (see Supplementary Table 4) and include both high income countries (US, Japan) and low-middle income countries (China, Mexico, Pakistan).

Despite some differences in effect magnitude, all studies report some level of direct rebound effect associated with the use of newly purchased energy-efficient appliances. Hence, the available evidence strongly suggests that savings in energy consumption are likely to be lower than engineering estimates

that do not account for behavioral change. The net savings in energy consumption, inclusive of the rebound effect, are between 8-10% in response to information interventions when combined with mandatory technology upgrades. Mizobuchi and Takeuchi[40] compare the changes in energy consumption of households that owned an efficient vs. inefficient air conditioner in response to an awareness campaign encouraging households to reduce their consumption. They find that households that owned an efficient air conditioner reduced their consumption by about 9% though the effect was statistically significant only for households that had recently purchased an air conditioner rather than replacing the existing air conditioner with a more efficient one. Furthermore, when information intervention is applied alone, as in case of Sun[41] the intervention resulted in no statistical reduction in energy consumption.

Studies further report that monetary incentives usually increase the purchase of efficient appliances, but they lead to larger rebound effects - possibly even backfire effects - in developing countries. Alberini, Gans, Towe[42] find that while households achieve an overall decrease in energy consumption through mandatory efficiency standards coupled with monetary incentives, an increase in monetary incentives corresponds to a smaller reduction in electricity use. The authors hypothesize that this is because the incentives reduce the lifetime cost per unit of energy services to a point sufficient to trigger a large increase in the demand for these energy services. Additionally, rebates allow households to upsize and defray the cost of bigger units, or of units that end up being used more. Houde and Aldy[36] find that State Energy Efficient Appliance Rebate Program (SEEARP) in the US did not translate into any meaningful reduction in energy consumption as rebates led consumers to upgrade to appliances with more features but that are less efficient. These findings support that of Davis, Fuchs, Gertler[43] who use a large dataset and comprehensively investigate the effect of financial incentives for adoption of efficient air-conditioners in Mexico. They find that households that use these incentives to buy new air conditioners are likely to have increased their consumption because of upsizing and increase in usage of air conditioning during the summer months by a relatively modest 20%. The authors hypothesize that rebound effects are expected to be larger in developing countries, where the demand for energy services is suppressed by the ability to pay for these services. This is confirmed by Yao, Liu, Yan[44] who find that a subsidy for energy efficient appliances in China led to an increase in the energy consumption of households though the study does not employ a robust control treatment design. Hence, to the extent the different interventions studied here incentivize the uptake of energy-efficient household appliances, their impact on energy consumption of households is limited by the presence of rebound effects, but more rigorous primary evidence is needed to study the magnitude of this effect or different appliances in different parts of the world.

Discussion & conclusions

This systematic review and meta-analysis aggregates evidence from experimental and quasi-experimental studies that evaluate the impact of behavioral, information and economic interventions on purchase of energy efficient appliances, market share of efficient appliances, and the subsequent savings in energy consumption after the purchase of efficient appliances. We argue that it is critical to discuss the effect of the interventions on these different outcomes jointly as they provide complementary pieces of information to our larger effectiveness question. We do this despite stark differences in the availability of evidence across the three outcomes, which affects the choice of our synthesis methodologies. We conduct a quantitative meta-analysis of the studies on purchase decisions given the relatively large sample size and provide a narrative review of the literature on the market-share and energy-use studies.

Our meta-analysis shows a small to medium average effect size (average Z score: 0.06-0.14) of monetary, information and behavioral interventions on the purchase of energy-efficient appliances. In light of recent interdisciplinary discussions on the interpretation of effect sizes, however, such effect sizes seem to be relatively common when interpreting the results of experiments[45], [46]. Even though estimated average effect sizes are statistically significant and robust to outliers and model selection, the results should be interpreted with caution. About half of the effect sizes in our dataset come from studies that rely on stated preferences of consumers rather than actual observed behavior, and our heterogeneity analysis suggests that the studies based on stated preferences are likely to report higher effect size as compared to those that use actual purchase data. The main benefit of using stated preferences approach is full and complete control over the experimental setting; however, the choices are based on hypothetical situations, decisions are often made removed from any context and respondents are not financially committed to their choices, which may bias results upward[47]. In addition to this, studies that were tagged as possibly having a higher risk of bias tend to report larger effect sizes. These findings, along with the presence of publication bias in the literature, suggest that the effects reported in the literature reviewed tend to overstate the efficacy of monetary, information and behavioral interventions.

On the question whether increasing use of efficient appliances in households can translate into reduced energy consumption at household level, the scarce evidence suggests the presence of rebound effects limiting the effectiveness in conserving energy, i.e. the modest increase in sales of efficient appliances through voluntary policy instruments translate only into very moderate energy savings. This is also supported by recent literature on energy efficiency improvements and renovation, such as Fowlie, Greenstone, & Wolfram[48]. Encouragingly though, the size of the rebound effect also does not seem to be large enough to lead to a backfire effect in most studies.

Ultimately, the practical consequences of the average effect sizes reported in this review, however, depend on at least three factors. Firstly, consideration should be given to how often a person makes decisions that could be influenced by the interventions under investigation[46]. For example, a widespread introduction of energy labels would not only influence one purchase decision but could potentially have a cumulative effect on many different purchase decisions, which would significantly increase the actual expected effectiveness of an intervention. Secondly, the scalability and costs of interventions are important[49]. A relatively cheap and easy-to-implement intervention that can be rolled out in such a way that it can influence the decisions of millions of people, such as energy labels, is to be evaluated differently than an intervention that is rather costly and resource-intensive to implement, such as a personal energy audit. Therefore, it may be advisable to deploy a scalable and cost-effective intervention on a large scale, even if the effect size is small, while more demanding interventions require a high effect to be cost-effective and therefore advisable. Thirdly, interventions

may differ strongly in their welfare effects[50], [51], which for example incorporate the actual reduction in externalities, such as greenhouse gas emissions, achieved by the intervention, the cost that those who are affected by the intervention incur when changing their behavior, i.e., the potentially higher cost or lower comfort of an energy efficient appliance, potential effect heterogeneity and any other intervention effects that affect individual utility. Hence, the estimated meta-analytic average effects may serve as a good starting point for assessing the potential of interventions to promote the adoption of efficient appliances, but to increase the informative value of meta-analyses for such questions, as also recommended in Andor & Fels (2018)[21], studies should be more transparent about the costs of interventions, scalability, and possibly welfare effects, so that these aspects can also be included in the quantitative research synthesis or the marginal value of public funds can then be estimated and compared (see for example Hahn et al. (2024)[52]).

Hence, our systematic review reveals important evidence gaps. Evidence is limited on some key interventions such as loans or subsidies. Despite the larger number of labeling and information-based measures deployed by governments, ex-post policy evaluations of these programs that identify a causal effect are still scarce. The present studies are not always able to distinguish between the effect of such voluntary approaches and mandatory phasing out of inefficient appliances. There is a more structural gap in primary evidence on the effects on market shares and energy consumption. Closing those should be a priority as exploring how our set of interventions might be more effective as part of smart policy packages e.g. around other regulatory instruments or standards. Particularly, answering this latter question could be fundamental for better understanding their role in climate and energy policy. Substantive understanding on the overall impacts of incentive programs is critical to ensure that governments and institutions that are devoting substantial time, energy, and resources in championing a behavioral approach to increase energy adoption do so in a manner truly capable of bringing about transformative change in energy consumption.

Methods

Literature search and data extraction

For data collection we relied primarily on (1) string-based searches of bibliographic databases; and (2) a rolling review of existing literature reviews on the same topic. In compliance with recommendations for systematic evidence syntheses, we included several bibliographic databases in our search (Web of Science Core Collections Citation Indexes, Scopus, JSTOR, and RePEc), as well as the web-based academic search engine Google Scholar. The Web of Science Core Collection Citation Indexes included in our search were: Science Citation Index Expanded (SCI-EXPANDED) --1900-present, Social Sciences Citation Index (SSCI) --1900-present, Arts & Humanities Citation Index (A&HCI) --1975-present, Conference Proceedings Citation Index- Science (CPCI-S) --1990-present, Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH) --1990-present, Emerging Sources Citation Index (ESCI) --2015-present. The records pulled from JSTOR were limited to only publications after 1960 due to an inconsequential number of publications available from years before.

We applied a comprehensive search string, which was developed based on the PICOS (population, intervention, comparator, outcome, and study design) approach recommended by the Campbell Collaboration (Supplementary Table 1). The search string was developed reiteratively by checking the results of the search against a set of relevant studies identified using previous literature reviews. The search was aimed at articles that dealt with household energy (or electricity) consumption and/or the purchase of energy-efficient appliances along with one or more interventions of interest (see Supplementary Table 2 for inclusion-exclusion criteria). We only tagged as relevant those studies that dealt with appliance purchases by households or individuals and/or contained information about specific appliance energy consumption in a household or dormitory. Further, studies tagged as relevant contained a quantitative estimate for appliance sales or energy saved through the effect of the relevant intervention. Studies that only provided an effect size but not the associated variance were not included in the final synthesis. In addition, studies where no obvious comparator group was available (untreated control group or pre-intervention data) or where the sample size was too small to extract meaningful estimates were excluded from the analysis. Studies where behavioral interventions targeted overall energy consumption rather than certain appliances were not considered. Overall, studies focusing on energy consumption were only included if they regarded energy usage in relation to appliance upgrade (i.e. rebound effect studies). The decision was made due to sufficient evidence on the effects of behavioral interventions on energy consumption of households in general from existing reviews as described in the introduction to this paper.

Since no exclusions were made based on the year of the study, methodology, or the field of publication, the search returned a large number of studies (30,902 entries after removing duplicates and empty records). To make manual screening of relevant papers possible, a machine learning algorithm was applied that used support vector machines to rate the studies based on the predicted relevance at the abstract level (Supplementary Figure 1, 2). The ROSES flowchart for the screening and coding process is available in the supplementary materials (Supplementary Figure 1), as well as the complete list of studies included in the analysis (Supplementary Table 9). After training the model with the previously known relevant documents (32) and a random sample of 500 documents, several reiterations were performed, where a team of five reviewers screened the predicted most relevant abstracts manually. The abstracts were single screened, however, to ensure consistency, two samples of 50 studies were screened by the five reviewers independently at the abstract level (Kappa = 0.68). We applied ML-assisted screening process with implementation of a statistical stopping criteria for screening these documents at abstract level, which resulted in around 85% work savings and a recall of 0.9 achieved at $p = 0.018$. We used the

NACSOS software for managing search results, removing duplicates, and screening records. In total, 4,788 studies were screened manually at the abstract level, with 576 rated as relevant. The studies were tagged with one of the categories: a) uses behavioral interventions for purchase of energy efficient appliances, b) studies usage of energy efficient appliances after purchase for estimating the rebound effect, or c) uses behavioral interventions for reducing energy consumption of specific appliances (not necessarily energy efficient). This review focuses on the efficacy of behavioral interventions in promoting purchase of energy-efficient appliances (categories a and b). Accordingly, 170 studies were screened at the full-text level, including 32 studies identified from previous reviews. The final sample included 58 studies, whereby all the exclusions have been recorded and categorized. As can be seen in Supplementary Figure 1, 13 studies could not be retrieved at full text due to either access restrictions (N = 9) or general unavailability of full text due to incomplete records (N = 4). Among further exclusion reasons were a) dependent variable outside of the scope of this study (N= 50, these studies mostly focused on building retrofits, or appliances that were not considered in this review, or dependent variables that were not in any way related to direct purchase decisions / observed energy usage); b) independent variable something other than an intervention of interest (N = 10, a common thread among these studies was to analyze the effect of socio-demographic characteristics on appliance purchase decisions which lies outside the scope of this analysis); c) irrelevant study design or incomplete uncertainty information (N = 48, most of these studies applied modelling as their main methodology; there were 10 theoretical papers, 3 literature reviews, and 3 case studies; 10 of these studies provided no quantitative estimates, and in 7 of them no uncertainty measurements could be calculated for the effect sizes) and d) language of the text was not English (N = 1).

Four reviewers extracted the relevant data from the included studies following the guideline described in detail in a codebook (Supplementary Table 10) and holding regular discussions of the coded fields to see what disagreements occurred, while making suitable adjustments to the codebook. A single reviewer double checked the final data collected for all the included studies. The data was extracted manually using MS Excel.

Standardizing effect sizes

As the studies included in this review covered a range of dependent variables, they were categorized into a) purchase probability of energy efficient appliance, b) willingness-to-pay (WTP) for an efficient appliance, c) energy usage after efficient appliance adoption, and d) market share of an efficient appliance. The effect sizes for purchase probability and WTP reflected different dependent variables (e.g. choice of a high-efficiency class appliance, WTP for a high-efficiency class appliance, etc.), which were harmonized and analyzed together, whereas usage and market share effect sizes were studied separately. We harmonized the effect sizes on purchase probability and WTP following a standard procedure described by Ringquist (2013)[53] of first converting the recorded regression coefficients into correlation coefficients (*r*) using the total sample size, which we then converted to Fisher's Z.

If a study applied the difference-of-means design, the standardized mean differences (or Cohen's d) was calculated first and then converted to Fisher's Z. The formulas for the described conversions are specified by Ringquist (2013)[53]. These values were then used for meta-regressions. For the analysis on energy usage and market share impact, there were fewer studies available, and the dependent variables were homogeneous. Thus, the effect sizes on usage and market share were harmonized as percentage change estimates so that a descriptive analysis could be performed. Most of the studies within these two categories had reported percentage change explicitly, while others reported coefficients from models that depending on the model specifications were easily converted to percentage change values.

Synthesis

After the harmonization of individual effect sizes for the dependent variables "purchase probability" and "WTP" through conversion to Fisher's Z [53], we implemented a random effects model with intervention and appliance category as predictors. The *metafor* package in R was used to get the estimates from FE and REML model specifications. While the estimates from the FE model were lower compared to REML, they were still positive and statistically significant. We present the estimates from a REML model in the main text, as the underlying effects in our sample are likely to be heterogenous, in which case a random effects model is considered more appropriate than a fixed effect model [53], [54]. Furthermore, as some of the studies in the sample reported multiple effect sizes were extracted from each study, we applied the multi-level random effects model to account for possible interdependence of the effect sizes.

We also checked that no single study or specific effect size disproportionately affected the model results. We calculated Cook's distance values for coefficients in the meta regression. Observations with the highest values of this parameter were dropped which resulted in 7 and 7 influential values for intervention-specific and appliance-specific models respectively. The results presented in the main text are after exclusion of the influential observations. Not excluding these observations does not materially change the results except that the confidence intervals for the estimated effect sizes become larger (see Supplementary Table 6).

We analyzed the subset of appliances that had the best coverage in our sample (air-conditioner, lighting, refrigerator and water heater), and economic incentives seemed to have a much larger effect for refrigerators, and a smaller effect compared to information incentives for air-conditioners and water heaters. However, when we excluded outliers based on Cook's difference, the interaction effect between refrigerator and economic incentives was completely diminished.

Furthermore, we analyzed the subset of effect sizes where more than one intervention was implemented at the same time, comprising of 40 observations. We found no statistically significant effects when interventions are applied in combination with each other. The most powerful combinations of interventions appeared to be information coupled with a subsidy, while combining information with rebate or labels with rebates has a slightly lower average effect.

Heterogeneity in effect sizes

There is considerable unexplained heterogeneity in effect sizes that we extract from primary studies (I^2 = 97-98% for random effects model with intervention/appliance as moderator variable). Estimates of the effect size may vary both because of the variation in the underlying response to interventions (e.g., due to the type of intervention and appliance, regional variation, homeowners vs. tenants, etc.) or because of differences in characteristics of the studies (e.g., setting of the experiment, statistical techniques used for analysis, reporting biases, etc.). We coded 16 variables, which can take 35 separate values, to capture several characteristics of the studies. The complete list of the study characteristics that we code as potential explanatory variables, and their summary statistics are presented in the supplementary information (Supplementary Tables 4-5). We use the following regression to estimate the relationship between the potential explanatory factors discussed above, and the willingness to pay reported by the studies, standardized as Z value, in our dataset:

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

where $\hat{Z}_{ij}$ is the *i*-th estimate from the *j*-th study, and $X_{k,ij}$ is the corresponding value of the k-th explanatory variable. α is the intercept and θ is a vector of k parameters, ε 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 possible combinations (2^{16} in our case) of the categories of the moderator 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, 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. [55], [56], [57]). 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[59], [60].

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[59]. 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[56], [57]. The results from the BMA are provided in supplementary information (Supplementary Table 7). The tables show the posterior inclusion probability (PIP) that indicates the relevance of each variable. Commonly, variables with a PIP above 0.5 are interpreted to be relevant explanatory factors, while variables with lower PIPs are unable to capture the observed heterogeneity. The table furthermore provides the posterior mean and standard deviation of the estimated effect averaged across all meta-regressions that include the respective variable. As a robustness check of our baseline BMA specification and to provide more readily interpretable results we run a hybrid frequentist-Bayesian model. We employ variable selection based on BMA, i.e., we only include the variables with posterior inclusion probability (PIP) above 50% and estimate the resulting model using an Ordinary Least Squares (OLS) model with clustered standard errors. The results for this exercise are presented in the main text in Table 2.

Acknowledgements

Funding from the Kopernikus-Projekt Ariadne (grant reference: 03SFK5J0-2) by the German Federal Ministry of Education and Research is gratefully acknowledged. The authors acknowledge the contribution of Sangeeth Selvaraju whose master thesis initially explored this literature and contributed towards development of the search query. Feedback from participants of the What Works Global Summit 2023, Ottawa and BIG Difference BC 2024, Vancouver is also appreciated.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information and will be made available on GitHub on publication. All the information collected in this project is publicly available in line with the systematic reviews reporting

protocol, providing the transparency and reproducibility required to conform with Open Synthesis principles (see the ROSES checklist).

Code availability

We used the NACSOS software to manage search results, remove duplicates, screen records and extract data, and the *metafor* package in R for the meta-regressions. All the software packages used are open source and freely accessible. The code developed for the paper is available in the Supplementary Information and will be made available on GitHub on publication.

References

[1] IEA, "World Energy Outlook 2023," IEA, Paris, 2023. [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2023, License: CC BY 4.0 (report); CC BY NC SA 4.0 (Annex A)

[2] IPCC, "Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)].," IPCC, Geneva, Switzerland, 2023. [Online]. Available: 10.59327/IPCC/AR6-9789291691647

[3] J. Barrett *et al.*, "Energy demand reduction options for meeting national zero-emission targets in the United Kingdom," *Nat Energy*, vol. 7, no. 8, pp. 726–735, Aug. 2022, doi: 10.1038/s41560-022-01057-y.

[4] N. Zhou, N. Khanna, W. Feng, J. Ke, and M. Levine, "Scenarios of energy efficiency and CO2 emissions reduction potential in the buildings sector in China to year 2050," *Nat Energy*, vol. 3, no. 11, pp. 978–984, Nov. 2018, doi: 10.1038/s41560-018-0253-6.

[5] C. von Stechow *et al.*, "2 °C and SDGs: united they stand, divided they fall?," *Environmental Research Letters*, vol. 11, no. 3, p. 34022, 2016, doi: 10.1088/1748-9326/11/3/034022.

[6] A. Grubler *et al.*, "A low energy demand scenario for meeting the 1.5 °c target and sustainable development goals without negative emission technologies," *Nature Energy*, 2018, doi: 10.1038/s41560-018-0172-6.

[7] IEA, "Energy Efficiency 2023," IEA, Paris, 2023. [Online]. Available: https://www.iea.org/reports/energy-efficiency-2023, License: CC BY 4.0

[8] H. D. Saunders *et al.*, "Energy Efficiency: What Has Research Delivered in the Last 40 Years?," *Annual Review of Environment and Resources*, vol. 46, no. 1, pp. 135–165, 2021, doi: 10.1146/annurev-environ-012320-084937.

[9] T. D. Gerarden, R. G. Newell, and R. N. Stavins, "Assessing the Energy-Efficiency Gap," *Journal of Economic Literature*, vol. 55, no. 4, pp. 1486–1525, Dec. 2017, doi: 10.1257/jel.20161360.

[10] L. Berrang-Ford *et al.*, "Editorial: Evidence synthesis for accelerated learning on climate solutions," *Campbell Systematic Reviews*, vol. 16, no. 4, p. e1128, 2020, doi: 10.1002/cl2.1128.

[11] J. Elliott *et al.*, "Decision makers need constantly updated evidence synthesis," 2021.

[12] N. r. Haddaway, P. Woodcock, B. Macura, and A. Collins, "Making literature reviews more reliable through application of lessons from systematic reviews," *Conservation Biology*, vol. 29, no. 6, pp. 1596–1605, 2015, doi: 10.1111/cobi.12541.

[13] H. Allcott and M. Greenstone, "Is There an Energy Efficiency Gap?," *Journal of Economic Perspectives*, vol. 26, no. 1, pp. 3–28, Feb. 2012, doi: 10.1257/jep.26.1.3.

[14] K. Gillingham, R. G. Newell, and K. Palmer, "Energy Efficiency Economics and Policy," *Annual Review of Resource Economics*, vol. 1, no. 1, pp. 597–620, 2009, doi: 10.1146/annurev.resource.102308.124234.

[15] K. Gillingham and K. Palmer, "Bridging the Energy Efficiency Gap: Policy Insights from Economic Theory and Empirical Evidence," *Review of Environmental Economics and Policy*, vol. 8, no. 1, pp. 18–38, Jan. 2014, doi: 10.1093/reep/ret021.

[16] T. M. Khanna *et al.*, "A multi-country meta-analysis on the role of behavioural change in reducing energy consumption and CO2 emissions in residential buildings," *Nature Energy*, vol. 6, no. 9, pp. 925–932, 2021, doi: 10.1038/s41560-021-00866-x.
[17] P. Buckley, "Prices, information and nudges for residential electricity conservation: A meta-analysis," *Ecological Economics*, vol. 172, no. January, p. 106635, 2020, doi: 10.1016/j.ecolecon.2020.106635.
[18] M. A. Delmas, M. Fischlein, and O. I. Asensio, "Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012," *Energy Policy*, vol. 61, pp. 729–739, Oct. 2013, doi: 10.1016/j.enpol.2013.05.109.
[19] W. Abrahamse, L. Steg, C. Vlek, and T. Rothengatter, "A review of intervention studies aimed at household energy conservation," *Journal of Environmental Psychology*, vol. 25, no. 3, pp. 273–291, 2005, doi: 10.1016/j.jenvp.2005.08.002.
[20] M. Bergquist, M. Thiel, M. H. Goldberg, and S. van der Linden, "Field interventions for climate change mitigation behaviors: A second-order meta-analysis," *Proceedings of the National Academy of Sciences*, vol. 120, no. 13, p. e2214851120, Mar. 2023, doi: 10.1073/pnas.2214851120.
[21] M. A. Andor and K. M. Fels, "Behavioral Economics and Energy Conservation – A Systematic Review of Non-price Interventions and Their Causal Effects," *Ecological Economics*, vol. 148, no. January, pp. 178–210, 2018, doi: 10.1016/j.ecolecon.2018.01.018.
[22] C. F. Nisa, J. J. Bélanger, B. M. Schumpe, and D. G. Faller, "Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change," *Nature communications*, vol. 10, no. 1, pp. 1–13, 2019.
[23] S. Mertens, M. Herberz, U. J. J. Hahnel, and T. Brosch, "The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains," *Proceedings of the National Academy of Sciences*, vol. 119, no. 1, p. e2107346118, Jan. 2022, doi: 10.1073/pnas.2107346118.
[24] T. S. Tsemekidi, P. Bertoldi, L. Castellazzi, T. M. Gonzalez, E. Clementi, and D. Paci, "Energy Consumption and Energy Efficiency trends in the EU, 2000-2020," Nov. 2022. doi: 10.2760/727548.
[25] "Electricity use in homes - U.S. Energy Information Administration (EIA)." Accessed: Oct. 22, 2024. [Online]. Available: https://www.eia.gov/energyexplained/use-of-energy/electricity-use-in-homes.php
[26] "Home Electrification and Appliance Rebates," Energy.gov. Accessed: Oct. 31, 2024. [Online]. Available: https://www.energy.gov/scep/home-electrification-and-appliance-rebates
[27] "Government Energy Spending Tracker – Analysis," IEA. Accessed: Oct. 31, 2024. [Online]. Available: https://www.iea.org/reports/government-energy-spending-tracker-2
[28] N. R. Haddaway, "Open Synthesis: On the need for evidence synthesis to embrace Open Science," *Environmental Evidence*, vol. 7, no. 1, pp. 4–8, 2018, doi: 10.1186/s13750-018-0140-4.
[29] N. R. Haddaway, B. Macura, P. Whaley, and A. S. Pullin, "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*, 2018, doi: 10.1186/s13750-018-0121-7.
[30] T. M. Khanna *et al.*, "PROTOCOL: Emissions savings from energy efficient appliances: a systematic review and meta-analysis of experimental evidence," Jul. 2022, doi: https://doi.org/10.17605/OSF.IO/AMNHR.
[31] A. O'Mara-Eves, J. Thomas, J. McNaught, M. Miwa, and S. Ananiadou, "Using text mining for study identification in systematic reviews: a systematic review of current approaches," *Systematic Reviews*, vol. 4, no. 1, p. 5, Jan. 2015, doi: 10.1186/2046-4053-4-5.
[32] M. Callagahan, F. Müller-Hansen, J. Hilaire, and Y. T. Lee, *NACSOS: NLP Assisted Classification, Synthesis and Online Screening*. (v0.1.0) [Computer software]. Zenodo, 2020. doi: 10.5281/ZENODO.4121526.

[33] D. Revelt and K. Train, “Mixed Logit with Repeated Choices: Households’ Choices of Appliance Efficiency Level,” *The Review of Economics and Statistics*, vol. 80, no. 4, pp. 647–657, 1998.
[34] S. Datta and M. Filippini, “Analysing the impact of ENERGY STAR rebate policies in the US,” *Energy Efficiency*, vol. 9, no. 3, pp. 677–698, Jun. 2016, doi: 10.1007/s12053-015-9386-7.
[35] S. Datta and S. Gulati, “Utility rebates for ENERGY STAR appliances: Are they effective?,” *Journal of Environmental Economics and Management*, vol. 68, no. 3, pp. 480–506, Feb. 2014, doi: 10.1016/j.jeem.2014.09.003.
[36] S. Houde and J. E. Aldy, “Consumers’ Response to State Energy Efficient Appliance Rebate Programs,” *American Economic Journal: Economic Policy*, vol. 9, no. 4, pp. 227–255, Nov. 2017, doi: 10.1257/pol.20140383.
[37] J. Schleich, A. Durand, and H. Brugger, “How effective are EU minimum energy performance standards and energy labels for cold appliances?,” *Energy Policy*, vol. 149, p. 112069, Feb. 2021, doi: 10.1016/j.enpol.2020.112069.
[38] T. Mandel, L. Kranzl, E. Popovski, F. Sensfuß, A. Müller, and W. Eichhammer, “Investigating pathways to a net-zero emissions building sector in the European Union: what role for the energy efficiency first principle?,” *Energy Efficiency*, vol. 16, no. 4, p. 22, Mar. 2023, doi: 10.1007/s12053-023-10100-0.
[39] K. Gillingham, D. Rapson, and G. Wagner, “The Rebound Effect and Energy Efficiency Policy,” *Review of Environmental Economics and Policy*, vol. 10, no. 1, pp. 68–88, Jan. 2016, doi: 10.1093/reep/rev017.
[40] K. Mizobuchi and K. Takeuchi, “The influences of financial and non-financial factors on energy-saving behaviour: A field experiment in Japan,” *Energy Policy*, vol. 63, pp. 775–787, 2013, doi: 10.1016/j.enpol.2013.08.064.
[41] B. Sun, “Heterogeneous direct rebound effect: Theory and evidence from the Energy Star program,” *Energy Economics*, vol. 69, pp. 335–349, 2018, doi: https://doi.org/10.1016/j.eneco.2017.11.025.
[42] A. Alberini, W. Gans, and C. Towe, “Free riding, upsizing, and energy efficiency incentives in Maryland homes,” *The Energy Journal*, vol. 37, no. 1, 2016.
[43] L. W. Davis, A. Fuchs, and P. Gertler, “Cash for coolers: evaluating a large-scale appliance replacement program in Mexico,” *American Economic Journal: Economic Policy*, vol. 6, no. 4, pp. 207–38, 2014.
[44] X. L. Yao, Y. Liu, and X. Yan, “A quantile approach to assess the effectiveness of the subsidy policy for energy-efficient home appliances: Evidence from Rizhao, China,” *Energy Policy*, vol. 73, pp. 512–518, 2014, doi: 10.1016/j.enpol.2014.06.010.
[45] “Interpreting Effect Sizes of Education Interventions - Matthew A. Kraft, 2020.” Accessed: Oct. 30, 2024. [Online]. Available: https://journals.sagepub.com/doi/10.3102/0013189X20912798
[46] D. C. Funder and D. J. Ozer, “Evaluating Effect Size in Psychological Research: Sense and Nonsense,” *Advances in Methods and Practices in Psychological Science*, vol. 2, no. 2, pp. 156–168, 2019, doi: 10.1177/2515245919847202.
[47] J. Carroll, E. Denny, and S. Lyons, “The effects of energy cost labelling on appliance purchasing decisions: trial results from Ireland,” *Journal of Consumer Policy*, vol. 39, no. 1, pp. 23–40, 2016.
[48] M. Fowlie, M. Greenstone, and C. Wolfram, “Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program,” Jan. 22, 2018, *Social Science Research Network, Rochester, NY*: 2621817. doi: 10.2139/ssrn.2621817.
[49] J. A. List, *The Voltage Effect*. New York, NY: Crown Currency. Accessed: Oct. 30, 2024. [Online]. Available: https://www.penguinrandomhouse.com/books/672117/the-voltage-effect-by-john-a-list/
[50] H. Allcott and J. B. Kessler, “The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons,” *American Economic Journal: Applied Economics*, vol. 11, no. 1, pp. 236–276, Jan. 2019, doi: 10.1257/app.20170328.

[51] H. Allcott, D. Cohen, W. Morrison, and D. Taubinsky, “When Do ‘Nudges’ Increase Welfare?,” Dec. 01, 2022, *Social Science Research Network, Rochester, NY*: 4299626. Accessed: Oct. 30, 2024. [Online]. Available: https://papers.ssrn.com/abstract=4299626
[52] R. W. Hahn, N. Hendren, R. D. Metcalfe, and B. Sprung-Keyser, “A Welfare Analysis of Policies Impacting Climate Change,” Jul. 2024, *National Bureau of Economic Research*: 32728. doi: 10.3386/w32728.
[53] E. J. Ringquist, *Meta-Analysis for Public Management and Policy*. San Francisco: John Wiley & Sons, 2013.
[54] J. P. Nelson and P. E. Kennedy, “The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment,” *Environmental and Resource Economics*, vol. 42, no. 3, pp. 345–377, 2009, doi: 10.1007/s10640-008-9253-5.
[55] J. Bajzik, T. Havranek, Z. Irsova, and J. Schwarz, “Estimating the Armington elasticity: The importance of study design and publication bias,” *Journal of International Economics*, vol. 127, p. 103383, Nov. 2020, doi: 10.1016/j.jinteco.2020.103383.
[56] T. Havranek, Z. Irsova, and K. Janda, “Demand for gasoline is more price-inelastic than commonly thought,” *Energy Economics*, vol. 34, no. 1, pp. 201–207, Jan. 2012, doi: 10.1016/j.eneco.2011.09.003.
[57] T. Havranek, M. Rusnak, and A. Sokolova, “Habit formation in consumption: A meta-analysis,” *European Economic Review*, vol. 95, pp. 142–167, Jun. 2017, doi: 10.1016/j.euroecorev.2017.03.009.
[58] S. Zeugner and M. Feldkircher, “Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R,” *Journal of Statistical Software*, vol. 68, pp. 1–37, Nov. 2015, doi: 10.18637/jss.v068.i04.
[59] T. S. Eicher, C. Papageorgiou, and A. E. Raftery, “Default priors and predictive performance in Bayesian model averaging, with application to growth determinants,” *Journal of Applied Econometrics*, vol. 26, no. 1, pp. 30–55, 2011, doi: 10.1002/jae.1112.
[60] A. E. Raftery, D. Madigan, and J. A. Hoeting, “Bayesian Model Averaging for Linear Regression Models,” *Journal of the American Statistical Association*, vol. 92, no. 437, pp. 179–191, 1997, doi: 10.2307/2291462.