

Motivating and Evaluating Energy Efficiency Policy

Kenneth Gillingham, Amelia Keyes, and Karen Palmer

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org



Motivating and Evaluating Energy Efficiency Policy

Kenneth Gillingham, Amelia Keyes, and Karen Palmer

Abstract

This paper reviews the recent evidence on the effectiveness and cost-effectiveness of energy efficiency interventions. After a brief review of explanations for the energy efficiency gap, we explore key issues in the evaluation of energy efficiency, including challenges and benefits to randomized controlled trials and incentives faced by those performing evaluations. We provide a summary table of savings results by type of efficiency intervention. We also develop an updated estimate of the aggregate cost-effectiveness of utility energy efficiency programs of 2.86 cents per kilowatt hour (kWh) net savings, but note that this estimate is based on utility-reported aggregate energy savings. Our review of the economics literature provides a mixed picture of the cost-effectiveness of specific interventions, with only some appearing to be highly cost-effective.

Key Words: energy efficiency, evaluation, energy savings, cost-effectiveness

© 2017 Resources for the Future. All rights reserved. No portion of this paper may be reproduced without permission of the authors.

Resources for the Future (RFF) is an independent, nonpartisan organization that conducts rigorous economic research and analysis to help leaders make better decisions and craft smarter policies about natural resources and the environment.

Working papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review. Unless otherwise stated, interpretations and conclusions in RFF publications are those of the authors. RFF does not take institutional positions.

Contents

1. Introduction.....	1
2. Brief Review of Key Concepts in Energy Efficiency Economics	3
3. From Theory to Evaluating Policy and Programs.....	5
4. Recent Evidence from the Academic Community	9
4.1. Empirical Evidence Testing Economic Theory	10
4.2. Empirical Evaluations of Energy Efficiency Programs	11
4.3. Behavioral and Information Programs.....	13
4.4. Product Standards	15
4.5. Financial Incentives	17
5. The Cost-Effectiveness of Energy Efficiency in Aggregate.....	19
6. Conclusions.....	22
References	25
Appendix: Dollars Per Ton of CO₂ Avoided	32

Motivating and Evaluating Energy Efficiency Policy

Kenneth Gillingham, Amelia Keyes, and Karen Palmer*

1. Introduction

As global interest in mitigating climate change continues to increase, there is an ever-growing demand for policies aimed at improving the energy efficiency of the economy. Policymakers have largely been reluctant to impose caps or prices directly on CO₂ emissions, and with the adoption of the Paris Agreement in 2015, international climate negotiations under the United Nations Framework Convention on Climate Change (UNFCCC) have shifted from a focus on an international emissions cap-and-trade approach—as was envisioned under the Kyoto Protocol—to more decentralized national commitments to emissions reductions with a collection of strategies or mechanisms to achieve those reductions. Energy efficiency not only typically has more political support than policies to limit emissions directly, but also is often perceived by policymakers as a low-cost approach to reducing carbon emissions that can also save consumers money through lower energy costs, even after accounting for the costs of the necessary investments in more efficient technology and the policies to bring those investments about. This deviation between the level of energy efficiency that appears to make economic sense and what we observe is known as the energy efficiency gap. Indeed, if the market does not elicit efficient levels of investment in energy efficiency, there may be economic justification for policies to promote efficiency investments even if externalities from carbon emissions are internalized.

To help policymakers identify the energy savings, emissions reduction potential, and costs associated with energy efficiency policies, analysts from federal and state governments, industry, and nonprofits have conducted hundreds of studies over the past several decades (e.g., Wilson et al. 2017; Holmes and Mullen-Trento 2017). There is a large and growing body of energy efficiency impact evaluations, often performed by professional evaluators at the request of policymakers, utilities, and energy and utility regulators. More recently, there has been an explosion of economic literature on energy efficiency policy, aiming both to improve our theoretical understanding of when, where, and why the market fails to deliver economically

* Gillingham: School of Forestry & Environmental Studies, School of Management, Department of Economics, Yale University, New Haven, CT 06511, email: kenneth.gillingham@yale.edu; Keyes: research assistant, Resources for the Future, Washington, DC 20036, email: keyes@rff.org; Palmer: senior fellow, Resources for the Future, Washington, DC 20036, email: palmer@rff.org. Palmer and Keyes thank the Alfred P. Sloan Foundation for financial support of this research.

efficient levels of energy efficiency investment and to evaluate selected energy efficiency programs to assess their cost-effectiveness. Despite this plethora of studies, there is still much we do not know about energy efficiency, in particular about the role of policy and market interventions in delivering energy savings and emissions reductions. However, the evidence is growing, and this literature can offer insights for policy design and future policy evaluations.

In this paper, we review the academic literature on energy efficiency, with a focus on the recent academic empirical evidence that seeks to identify the roles of various theoretical explanations for the energy efficiency gap and that looks at the effectiveness and cost-effectiveness of different types of policies. Our review differs from the existing set of energy efficiency reviews in four primary ways.¹

First, we focus on the estimates of effectiveness and cost-effectiveness of recent academic studies that use experimental and quasi-experimental approaches to evaluate specific energy efficiency policies to draw out useful estimates for policymakers. Second, this review of estimates highlights a major thread of the recent literature that draws on behavioral economics to assess the relevance of potential behavioral failures that could justify energy efficiency policy and interventions to address these failures. Third, we use utility-reported data through 2015 to update the aggregate estimates of the cost-effectiveness of utility energy efficiency programs from Gillingham et al. (2006), finding an aggregate estimate of 2.86 cents per kilowatt hour (kWh) net savings, or 1.94 cents per kWh gross savings. Fourth, we discuss the benefits and challenges of using quasi-experimental and experimental approaches in energy efficiency evaluation, as well as incentive problems that currently afflict evaluations.

The empirical studies by economists that seek to evaluate energy efficiency interventions include behavioral programs, subsidies for efficient appliances and energy savings, building codes, and weatherization programs. These studies typically find that the empirically estimated savings fall short of ex ante engineering estimates of savings, when available, although the size of the difference varies across studies and programs. The heterogeneity in experimental results, due to various factors including the time period of evaluation and consumer characteristics, provides insights into the specific mechanisms behind energy savings (e.g., changes in energy conservation habits versus investments in more efficient durable goods) and lessons for improving the cost-effectiveness of policies, such as targeting the most responsive consumers.

¹ Previous reviews include Gillingham et al. (2006, 2009); Gillingham and Palmer (2014); Allcott and Greenstone (2012); Allcott (2016); and Gerarden et al. (2017).

The rest of the paper is organized as follows. In Section 2, we briefly review the key concepts in energy efficiency economics, including the recent insights from behavioral economics. In Section 3, we discuss methods for evaluation of energy efficiency policy and programs. In Section 4, we review recent empirical evidence testing economic theory and evaluating the results of energy efficiency programs. In Section 5, we discuss the results of our analysis quantifying the cost-effectiveness of energy efficiency in aggregate. Section 6 concludes.

2. Brief Review of Key Concepts in Energy Efficiency Economics

Much of the discussion around energy efficiency policy cites the “energy efficiency gap.” This refers to a significant difference between observed levels of energy efficiency and some notion of optimal energy use (Jaffe et al. 2004). In the context of climate change mitigation, this notion has been used to identify potential upgrades to energy-using capital that might yield low- or negative-cost CO₂ emissions reductions when the upfront cost of the upgrades appears lower than the value of the anticipated energy savings (Creys et al. 2007; Granade et al. 2009; McKinsey & Company 2009). Gillingham et al. (2009), Gillingham and Palmer (2014), and Gerarden et al. (2017) outline the range of explanations that economists have posited for this phenomenon. These potential explanations include a range of classic market failures, behavioral failures rooted in behavioral economics, and analyst miscalculations of energy savings or costs of achieving those savings or both, with the final explanation suggesting that there is no energy efficiency gap at all.

An economic rationale for energy efficiency policy exists when market barriers that slow the diffusion of energy-efficient technologies can be demonstrated to stem from market or behavioral failures. Market failures assume that consumers are rational and can be defined as a deviation between socially optimal and privately optimal decisions. Market failures are sometimes described as contributing to the “social energy efficiency gap” because the private market does not provide the socially optimal level of energy efficiency, even if consumers are fully rational (see Gerarden et al. 2017). For example, negative externalities from pollution and landlord-tenant principal-agent issues (e.g., the tenant chooses how much energy to use, while the landlord pays for it) are classic market failures relevant to energy efficiency. Innovation market failures leading to an underinvestment in research and development (R&D) and capital market liquidity constraints leading to an underinvestment in more expensive energy-efficient technologies are additional market failures that may apply to energy efficiency (Palmer et al. 2012). These market failures, and others not mentioned here, are discussed at much greater length in Gillingham et al. (2009) and Gillingham and Palmer (2014).

Information problems represent a type of market failure that is an often-cited contributor to the energy efficiency gap. Information problems take a variety of forms, and in addition to causing systematic underinvestment in energy efficiency, they can also undermine the effectiveness of certain policies. One important form of information problem is simply a lack of information. Palmer et al. (2012) argue that a lack of credible information about actual energy savings from investment in efficient equipment may limit demand for efficiency improvements. Another information problem is asymmetric information, in which one agent in an economic transaction possesses information that another agent does not. This can lead to principal-agent problems, also commonly described as split incentives. These problems are particularly relevant for energy efficiency outcomes when consumers do not pay for their own energy consumption or make energy efficiency investments that they cannot perfectly observe.

Much of the recent energy efficiency literature has become increasingly focused on the possibility of behavioral failures, and these phenomena are often cited as major drivers behind the energy efficiency gap. Behavioral failures can be defined as any feature of decisionmaking that leads the consumer to exhibit a deviation between the utility made at the time of the decision—decision utility—and the utility at the time that the consequences of the decision occur. In contrast to market failures, which create externalities and contribute to a social energy efficiency gap, behavioral failures create internalities and can contribute to a private energy efficiency gap (Gerarden et al. 2017). These anomalies in decision-making may stem from cognitive biases such as status quo bias, loss and risk aversion, sunk-cost effects, temporal and spatial discounting, and availability bias (see Frederiks et al. 2015). Energy efficiency decisions have also been linked to prospect theory, in which consumers have reference-dependent preferences, exhibit loss aversion, and subjectively weight probabilities when faced with uncertainty (see Heutel 2017). Much of the energy efficiency literature has focused on how consumers implicitly discount future fuel savings from energy-efficient investments and whether they appear to undervalue future fuel savings relative to what would be expected based on how consumers discount in other contexts.

Whether cognitive biases are actual behavioral failures, however, has not gone unchallenged. Brennan (2013) questions the assumption that consumers fail to make utility-maximizing choices when investing in energy efficiency. Smith and Moore (2010) suggest a consumer choice model that incorporates additional choice constraints such as cognitive needs, providing an alternative theory for consumer behavior that does not require irrationality. Sallee (2014) furthers this point and argues that observed behavioral failures may be explained at least in part by rational inattention, in which consumers choose to act based on incomplete

information if information is costly to acquire. In these cases, what may appear to the analyst to be an undervaluation of future fuel savings may simply be unobserved costs.

If behavioral failures cause systematic bias in energy efficiency decision-making, then they clearly contribute to the energy efficiency gap and motivate a role for policy to correct the bias and improve economic efficiency. Additionally, policies designed to address market failures would not sufficiently correct for the biases created by behavioral failures. Tsvetanov and Segerson (2013) use a theoretical model that incorporates temptation to show that product standards can have better welfare outcomes than price-based policies like externality taxes because they reduce consumers' opportunities to give in to the temptation of purchasing low-cost products that are less energy efficient. Additionally, their findings suggest that a policy combining product standards with externality taxes can yield higher welfare outcomes.

In a similar vein, Allcott and Mullainathan (2010) suggest that policies designed specifically to address behavioral failures can have as much of an effect on behavior as price-based policies. Behavioral-based policies, often referred to as "nudges," are low-cost interventions designed to persuade consumers to change their behavior. Nudges generally incorporate strategies such as invoking social approval, providing energy consumption feedback, and asking for goal setting and commitments on energy conservation. A commonly used nudge, for example, is the provision of feedback to homeowners or renters on their energy usage and their peers' usage. Allcott et al. (2014) argue that in the presence of market and behavioral failures, an optimal policy would involve an energy tax set below marginal damages in conjunction with policies that target behavioral failures. This conclusion follows from their finding that consumers who have the greatest undervaluation of energy costs are also the least sensitive to energy taxes, so the optimal policy would target more biased customers while limiting distortions to less biased customers.

3. From Theory to Evaluating Policy and Programs

There are many possible explanations for the energy efficiency gap, but what really matters from a policy perspective is the effectiveness and cost-effectiveness of energy efficiency policies and programs. From a theoretical perspective, if the apparent energy efficiency gap can be entirely attributed to hidden costs and mismeasurement by the analyst, then there is no actual gap. In this case, energy efficiency programs are not likely to yield greater benefits than costs, and evaluations that appropriately account for hidden costs and correct for mismeasurement would tend to bear this out.

However, if there are market failures or behavioral failures underpinning the apparent energy efficiency gap, then it is possible for an evaluation that accounts for all hidden costs and corrects for any mismeasurement to reveal that a program not only is cost-effective relative to other approaches but also yields savings that exceed costs. The presence of such failures is a necessary but not sufficient condition for evaluations to reveal a program with positive net benefits. Programs to promote energy savings in such a setting could be poorly designed, poorly implemented, or not targeted appropriately to the relevant market or behavioral failure, thereby leading to non-cost-effective outcomes. The real challenge for policy design and evaluation lies in developing accurate evaluations that appropriately identify changes in energy use attributable to the intervention as well as all costs of the intervention. An accurate accounting of all the costs and benefits of energy efficiency interventions is often a challenging task given the real-world constraints under which evaluators and academics work.

To understand the challenges inherent in evaluations, it is instructive to first consider an evaluation of an energy efficiency policy or program in an ideal world. In this ideal world, unlimited resources could be brought to bear, and the policy or program could be rolled out as a large-scale, well-designed randomized controlled trial (RCT). An RCT will randomly pre-assign entities who could potentially be affected by the policy or program to a treatment or control group, allowing for a comparison between these two groups on completion of the experimental treatment (List and Metcalfe 2014). Depending on the scale and scope of the policy or program, the randomization may be possible at different levels—household, neighborhood, or even city—as long as the sample size is sufficiently large. This ideal vision is well understood by many, if not most, in the economics and energy efficiency evaluation profession (e.g., see Vine et al. 2014 for a perspective from the energy efficiency evaluation community).

Unfortunately, this ideal is often far from reality. While energy efficiency RCTs are being run by both economists and evaluators, the number is small relative to the very large number of evaluations being performed every year. For example, in the California Measurement Advisory Council database of energy efficiency impact evaluations, only 13 of the 467 evaluations of California energy efficiency programs in the past 10 years used an RCT methodology. One of the major challenges is simply the resource intensity of RCTs, which are typically much more expensive and time-consuming to develop than other types of evaluations. Policymakers are often unwilling to earmark a high percentage of the cost of a program to evaluation, limiting the possibility of RCTs. In some cases, the entire budget for an efficiency program may not be large enough to fund a single RCT with a sufficiently large sample size to

permit an analysis with adequate statistical significance (the sufficient sample size can be determined based on the expected treatment effect using power calculations).

A related challenge may be the effort required prior to implementing the RCT. Implementing an RCT requires substantial upfront investment in design, which may not be possible for programs with tight statutory or other deadlines. A further challenge is the possibility of spillovers and the difficulty of ensuring that the control group is not contaminated by the treatment.² This issue may be particularly problematic for large-scale marketing campaigns, such as radio or other media campaigns, which inherently cover large geographic areas. It may also be problematic for “market transformation” energy efficiency programs (defined by the American Council for an Energy-Efficient Economy [ACEEE] as programs that “remove barriers or exploit opportunities”), which usually include a bundle of measures targeted market-wide.³ Finally, some seemingly promising treatments may simply be difficult to implement in an RCT setting. For instance, Allcott and Sweeney (2017) found that a randomized information treatment implemented by home appliance sales agents had little effect, a result attributed to the limited incentive agents had under the experiment to deviate from their standard sales pitch.

Thus, while RCTs may be the “gold standard” for credible evaluation, most studies by economists and evaluators are not RCTs. Other modern empirical methods are available to evaluate policies and programs, however. A common theme of these approaches is that they attempt to isolate sources of variation in the data that can be considered plausibly random (Angrist and Pischke 2008). For example, many studies exploit a boundary between two regions that is arguably arbitrarily defined, and thus entities on either side of the boundary are plausibly the same except for the treatment (i.e., as good as randomly assigned). These are often simply difference-in-difference studies that assume common trends on either side of the border. Other studies exploit an unexpected dramatic change that occurs at a certain time, through a regression discontinuity design. The assumption in this research design is that the pre-period can act as a sufficient control for the treatment period. Regression discontinuity designs can also be used when there are other thresholds that lead to a random assignment of the treatment to one side of the threshold, such as program eligibility cutoffs. A third common approach is matching, whereby treated entities are matched with similar control entities based on observables to attempt

² In some cases, spillovers or “treatment externalities” are the object of interest of an RCT (e.g., Miguel and Kremer 2004). These RCTs typically require very careful research design.

³ See <http://aceee.org/portal/market-transformation> for more on how energy efficiency advocates think about market transformation activities.

to replicate randomization as closely as possible. All of these approaches are often called “quasi-experimental” (Greenstone and Gayer 2009) and are increasingly being used in energy efficiency evaluations (Stewart and Todd 2015).

Economists and many evaluators consider these approaches generally preferable to the “deemed savings” approach, which was the standard approach to evaluate energy efficiency programs for many years and is still common.⁴ This approach is usually substantially less expensive. While there are many variations in practice, the deemed savings approach uses engineering estimates of the energy savings from actions taken as part of the program and sums them up to get the total energy savings from the program.⁵ This estimate may also include some forecast of total savings over the program’s lifetime. The total savings estimate is then compared with the cost of the program to calculate cost-effectiveness. One common concern with the deemed savings approach is that it often (but not always) ignores behavioral responses and implementation challenges. For instance, there may be a rebound effect, whereby households increase their use of a more efficient appliance or their overall energy use or both because they are spending less on energy for the more efficient product. Alternatively, it may be that in real-world conditions, the energy efficiency improvement is not installed correctly, reducing the energy savings. Challenges may also occur if the deemed savings estimates were developed in a different setting (e.g., different climate zone) than where the program takes place.

Using revealed preference estimates from RCTs or quasi-experimental approaches may be more resource-intensive, but it can provide a benchmark against which to compare deemed savings estimates. As shown in the next section, it appears that the deemed savings estimates tend to be overestimates of the energy savings from energy efficiency programs. These types of findings may be further fueling the trend toward increased use of RCTs and quasi-experimental approaches in evaluations. However, it is important to note that not all empirical approaches that use household-level consumption data and compare changes in consumption after the program with a control group do a good job of approximating true randomization. For example, suppose a city decides to offer a rebate for energy-efficient light bulbs, and only households that are undertaking major renovations purchase these light bulbs with the rebate. These households may

⁴ While this approach is common for evaluating particular efficiency measures, such as incentives for efficient light bulbs or other equipment upgrades, it is less common for policy evaluations that have a broader form of impact, such as new building codes or energy efficiency resource standards. Note that we are using “deemed savings” to describe any approach that does not use a comparison group to derive savings estimates (Palmer 2016).

⁵ See Jayaweera et al. (2013) for more details on standard practices for calculating energy savings associated with a host of specific energy efficiency measures.

also be improving their efficiency in other ways through their renovations. Thus, unless the renovations are observed, any attempt at a matching approach based on observed household characteristics would lead to an upwardly biased estimate of the energy savings attributable to the rebate. This issue is referred to as a “selection bias” because the households that opted into the program were selected from a population that was different from the control group. In the example given, the selected population would have experienced some energy savings anyway, even without the more efficient light bulbs. The beauty of the RCT is that it eliminates this selection bias, because the control and treatment groups are randomly selected initially. Finding ways to minimize selection bias in evaluations is an area of ongoing effort and aspiration for both economists and professional evaluators.

Incentives for evaluators and academic economists also matter in the conduct and findings of an evaluation as well as the types of questions that each seeks to address. Evaluators face incentives both to be rigorous and to make their clients—often utilities—happy. Sometimes these incentives align to yield accurate evaluations, but other times they do not. Keeping clients satisfied may involve confirming expectations (e.g., making sure the savings are not too different from previously estimated savings) or simply showing at least some savings so that the clients save face in their discussions with regulators. Given that it is a repeated game between utilities and professional evaluators, there is a clear incentive for evaluators to make sure that their results will not preclude their being hired for future evaluation work. One practice that could help keep these incentives in check is for evaluators not to contract with the utility that operates the efficiency program, but to work for a third-party entity (Kaufman and Palmer 2012). Combined with the added challenges of RCTs, this incentive can translate into a resistance to RCTs and a preference for deemed savings approaches. In contrast, academic economists are often rewarded in the publication process for results that are counterintuitive on the surface and findings that buck trends, reflecting a desire to change the way people perceive an issue. Thus academics may face the opposite incentive: an incentive to develop rigorous studies to find that what others have been doing in the past is incorrect.

4. Recent Evidence from the Academic Community

The recent empirical literature on energy efficiency economics has made two major contributions. The first includes studies that use empirical methods to evaluate the validity of economic theory explaining the energy efficiency gap. The second includes policy evaluation studies that assess energy efficiency programs by estimating their energy savings and cost-effectiveness. In both, we focus on only the most recent academic evidence.

4.1. Empirical Evidence Testing Economic Theory

The first set of literature uses empirical methods to examine evidence on market or behavioral failures that could contribute to the energy efficiency gap.

Of those focused on market failures, several studies find evidence of asymmetric information causing principal-agent or split incentive problems and leading to higher energy consumption and less investment in energy efficiency. Davis (2012) finds evidence of a split incentive problem in the tenant-landlord relationship, concluding that renters paying their own utility bills are significantly less likely to have energy-efficient appliances such as refrigerators, clothes washers, and dishwashers. Gillingham et al. (2012) also investigate split incentives between tenants and landlords and find that tenants paying for their own heating and cooling are 16 percent more likely to change the temperature setting at night. Additionally, owner-occupied homes in which the residents pay for heating and cooling are 20 percent more likely to have attic and ceiling insulation and 13 percent more likely to have exterior wall insulation. Following this work, Myers (2015) demonstrates that asymmetric information in the tenant-landlord relationship creates split incentives and reduces the effects of changes in the relative prices of natural gas and oil on the demand for both products. Jessoe et al. (2017) investigate the split incentive problem in the commercial sector and find that it leads to higher electricity usage for the top consuming customers. Finally, Giraudet and Houde (2014) and Giraudet et al. (2016) show that asymmetric information between energy users and the suppliers of energy efficiency enhancing products can lead to moral hazard in the provision of quality for energy efficiency investments. For home energy retrofits, a hard-to-observe measure, they find evidence that contractors take advantage of homeowner lack of expertise by providing lower-quality retrofit installations, which leads to lower realized energy savings. These information problems therefore have the potential to reduce the effectiveness of energy efficiency policy.

Many studies focus on determining whether behavioral failures cause systematic bias in consumer decisions. Sallee (2014) finds empirical evidence of rational inattention, in which what appear to be cognitive biases are indeed consumers behaving rationally. He finds that this explanation is plausible in markets for durable goods such as vehicles and home appliances, where the variation in future energy costs is minor relative to the variation in up-front product prices.

A major thread of recent studies uses empirical methods to determine whether consumers appear to have high implicit discount rates (consistent with myopia) or undervalue future fuel savings when purchasing durable goods. Allcott and Wozny (2014) and Grigolon et al. (2017) find that consumers purchasing vehicles modestly undervalue future fuel costs. However, Busse et al. (2013) and Sallee et al. (2009) do not find evidence of undervaluation of future fuel costs in

vehicle purchase decisions. Allcott (2016) concludes that consumers make systematic mistakes in the purchase of energy-consuming durable goods but that the magnitude of these mistakes is small in comparison with the impact of the regulatory standards in place.

Another aspect of possible myopia, described at length in Ungemach et al. (forthcoming) stems from bounded rationality, under which consumers' decisions are affected by the framing of their choices. The authors find that fuel attribute descriptions affect car choices and that consumers with strong pro-environmental values are more likely to choose fuel-efficient cars when their fuel economy is described in terms of greenhouse gas emissions. Such a response to framing could lead to under- or overvaluation of future fuel savings.

Further studies focus on other types of behavioral failures that bias energy efficiency decision making. Sexton (2015) focuses on the issue of energy price salience. He studies the effects of automatic bill payment (ABP) programs on energy consumption and finds that ABP enrollment increases energy usage for both firms and households. Such programs have the potential to reduce price elasticity of electricity demand and challenge efforts to increase the price salience of energy with implications for optimal policy design. Heutel (2017) finds empirical evidence that consumers behave consistently with prospect theory when it comes to energy efficiency by observing that consumers with higher loss aversion are less likely to invest in energy-efficient goods such as alternative fuel vehicles and efficient light bulbs. This result suggests that a high degree of loss aversion should be met with greater energy efficiency subsidies and that the standard Pigouvian externality tax should be modified to account for behavioral failures. Similarly, Tsvetanov and Segerson (2014) study energy efficiency decisions in the purchase of refrigerators and find evidence that when individuals do not act rationally, product standards can result in higher welfare than price-based policies. The greatest gains tend to accrue to lower-income households, which are more likely to select less efficient products.

4.2. Empirical Evaluations of Energy Efficiency Programs

The second set of literature we review covers studies that aim to estimate causal relationships to evaluate specific energy efficiency interventions. The ex post studies we focus on hold distinct advantages over the bounty of ex ante studies and have led, in many cases, to substantially different results. Many studies surveyed in this section employ well-designed RCTs, and others follow quasi-experimental methods, such as those outlined in Section 3. Additionally, many studies have attempted to capture the nuances of program results by examining factors such as heterogeneity of effects across consumers and changes in outcomes over time. This section reviews studies of three major types of energy efficiency programs: behavioral and information programs, building codes, and financial incentives. Energy savings and cost-effectiveness estimates are displayed in Table 1.

Table 1. Energy Savings and Cost-Effectiveness of Energy Efficiency Interventions for Energy Use in Buildings

Authors and year	Evaluation type	Where multiple results available, relevant subset	Percent change in energy usage	Cost-effectiveness (cents per kWh saved, 2015\$) ¹	Engineering estimates of percentage change in energy usage
Behavioral Programs					
Allcott (2011)	RCT		-2%	3.6	
Allcott and Rogers (2014)	RCT	One-shot intervention		4.4	
		Two-year intervention		1.1 to 1.8	
		Four-year intervention		1.2 to 1.8	
Ayres et al. (2012)	RCT	Sacramento	-2%	5.5	
		Puget Sound	-1.2%	2.0	
Costa and Kahn (2013)	RCT	Political liberals	-3.6%		
		Political conservatives	-1.1%		
Graff Zivin and Novan (2016) ²	RCT	Households with AC units	-24%		
		Households without AC units	0%		
LaRiviere et al. (2014)	RCT		2%		
Building Codes					
Aroonruengsawat et al. (2012)	PD regression, fixed effects	Nationwide	-5.0%		
		California	-4% to -7.7%		-6%
Jacobsen and Kotchen (2013)	RD analysis	Electricity	-4.3%		-2%
		Natural gas	-6.7%		
Kotchen (2017)	RD analysis	Electricity	0%		
		Natural gas	-13.5%		
		Electricity and natural gas	-2.9%		-2%
Levinson (2016)	Multiple methodologies	Electricity	-10% to -15%		-80%
		Natural gas	-25%		
Novan et al. (2017) ³	RD analysis		-13%	24.4 ⁴	-20%
Efficient Equipment or Energy Savings Subsidy					
Alberini and Towe (2015)	Matching		-5.3%	3.9 ⁴	
Alberini et al. (2016)	DID regression	Rebate of \$1,000 or more	0%		
		Rebate of \$450	-5.5%	47.9 ⁴	
		Rebate of \$300	-6.2%	28.2 ⁴	
		No rebate	-16%		
Burlig et al. (2017) ⁵	Machine learning		-2.9 to -4.5%		-11.6 to -18% ⁴
Costa and Gerard (2015)	DID regression	Short-term effects (9 months)	-23%		-20%
		Long-term effects (10 years)	-11%		
Davis et al. (2014)	DID regression	Refrigerators	-8%	27.2	
		Air conditioners	1.7%		
Ito (2015)	RD analysis	Inland	-4%	3.0	
		Coastal	0%	114.7	
Information Provision					
Alberini and Towe (2015)	Matching		-5.5%	4.5 ⁴	
Asensio and Delmas (2015)	RCT	Information on social costs	-8.2%		
		Information on private costs	Insignificant		
Byrne et al. (2014)	Multiple methodologies	Immediate treatment effect	0%		
		Treatment effect after 7 months	-4.6%		
		Households who overestimate their usage	7%		
		Households who underestimate their usage	-5%		
Savings Subsidy					
Allcott and Greenstone (2017)	RCT	Natural gas	-5.3%		-18.5%
		Electricity	-5.2%		-1.9%
Weatherization					
Fowle et al. (forthcoming)	RCT		-19%		
	DID regression		-9.5%		
Graff Zivin and Novan (2016)	FE regression	Households with AC units	-7%		-14%
		Households without AC units	0%		

¹When dollar years are not specified we use the dollar year from the final year of the program(s).²Evidence of energy savings is conditional on participation in WAP program.³Authors suggest that projected savings should be closer to actual savings because energy use in pre-retrofit houses was lower than projected.⁴Calculated using estimates from study.⁵Study focuses on energy efficiency in schools.

4.3. Behavioral and Information Programs

A major area of growth in the academic literature has been studies focusing on energy efficiency programs that address behavioral failures and information problems. These programs generally administer interventions that “nudge” consumers to conserve energy. Examples of such nudges include giving consumers social comparisons of their energy usage and asking consumers for nonbinding commitments to reduce energy usage. Several recent studies provide estimates of program treatment effects by employing randomized controlled natural field experiments. For example, several run RCTs using the company Opower, which sends home energy reports to residential utility customers about their energy usage relative to that of neighbors, along with suggestions regarding ways to conserve energy. Many studies estimate both energy savings results and overall program cost-effectiveness, and some examine either persistence of the effects over time or the heterogeneity in results due to different household characteristics.

Allcott (2011) studies the effects of the Opower treatment to find energy savings of 2 percent and treatment effects that decay over time after home energy reports are received but increase again whenever the next report is received. A subsequent Opower study by Allcott and Rogers (2014) investigates treatment effect decay over a longer time period and observes a similar initial pattern of usage reductions followed by backsliding. However, this cycle lessens over time, and average treatment effects are greater for customers that received reports for at least two years than for those with shorter exposure to the program. These results suggest that consumers may primarily make behavior changes in early treatment periods and either adopt lasting habits in the longer term or invest in energy-efficient durable goods.

Habituation of energy-efficient behavior and investment in durable goods can improve the cost-effectiveness of longer-lasting programs, and in fact, Allcott and Rogers (2014) find that a single Opower report has a cost-effectiveness of 4.31 cents per kWh saved, while sending reports to customers over two years has a cost-effectiveness of 1.08–1.77 cents per kWh saved. Ayres et al. (2012) also evaluate Opower programs and do not observe any short-term pattern of action and backsliding until the next report is received, concluding that quarterly treatments generate energy reduction effects similar to those of monthly treatments but are more cost-effective. The authors also attempt to draw conclusions about the mechanisms behind the treatment effects by observing that treatment effects were highest on Sundays, a result suggesting that the impacts may have been driven by behavioral changes rather than changes in the household stock of durable goods.

Studies that examine the heterogeneity of results across customers can also provide lessons for designing more cost-effective interventions. For example, the effects of interventions may differ depending on what consumers learn about their baseline level of consumption. Allcott (2011) finds that consumers with higher baseline consumption conserve more electricity, and those with lower baseline consumption display a “boomerang effect” by actually increasing their electricity usage. Byrne et al. (2014) study the effects of information provision in Australia using an RCT and find evidence of the boomerang effect for consumers that underestimated their baseline energy consumption. They observe that underestimating households decrease consumption, while households that had overestimated their relative usage increase consumption. In contrast, Ayres et al. (2012) find no evidence of the boomerang effect for low baseline consumption customers.

Other work provides suggestive evidence of the mechanisms through which consumers are incentivized to exhibit more energy-efficient behavior. This evidence comes primarily from experiments that exploit variation in consumer characteristics and those that vary the framing of interventions (i.e., framing information around private cost savings versus social cost savings). A study of Opower by Costa and Kahn (2013) finds that liberal and environmentalist households have a more energy-efficient baseline than conservative households, and they are also more responsive to the home energy reports (although both liberal and conservative households reduce energy usage), suggesting that private costs may not be the most salient factor in consumer decision making. LaRiviere et al. (2014) conduct an RCT intervention similar to that of Opower, in which home energy reports are sent to households, and also find that energy savings results differ based on political affiliation; however, they find that when nudges are framed to emphasize the public good of energy savings, Republican and mixed-affiliation households are most responsive. In a related but non-Opower study, Asensio and Delmas (2015) find that environmental- and health-based messages are better drivers of behavioral change than monetary savings information and are particularly effective for families with children.

Heterogeneity in treatment effects can also stem from differences in households’ durable energy-consuming goods. Graff Zivin and Novan (2016) study an RCT that administered behavioral interventions—including information provision and requests for nonbinding energy conservation commitments—to recipients of free home retrofits through the federal Weatherization Assistance Program (WAP) and find that these behavioral interventions led to an additional 23 percent decrease in electricity usage but only in households with air-conditioning units. This result suggests that changes in air-conditioning usage are a substantial source of energy use reductions under such behavioral programs.

Another set of studies examines whether information and behavioral interventions to encourage home energy audits can lead to audit uptake and energy efficiency investments. Gillingham and Tsvetanov (2017) use an RCT to investigate the effects of information provision on home energy audit uptake and find that a message to consumers combining the effects of social norms and salience improves audit uptake by 20 percent. The effects are strongest in rural communities with strong social networks. LaRiviere et al. (2014) find that in an information provision program, the framing of information affected results, as privately framed signals (emphasizing electricity expenditures and kWh usage) tended to affect mainly audit uptake, and public good framed signals (emphasizing CO₂ emissions from electricity usage) tended to affect mainly electricity consumption. This finding supports the theoretical argument that different types of framing can have varying levels of effectiveness (Darby 2006).

Furthermore, privately framed signals that are effective at inducing audit uptake are not effective at inducing extensive margin adjustments for energy efficiency installations. Allcott and Greenstone (2017) examine two similar programs that encourage and subsidize home energy audits. They find that the information and behavioral interventions had impacts that were statistically and economically insignificant, and only audit subsidies increased audit uptake. Consistent with LaRiviere et al. (2014), they also find that consumers who were marginal to audit subsidies were less likely to ultimately make energy efficiency investments than those that were inframarginal. Alberini and Towe (2015), however, study the effects of a free home energy audit program in Maryland and find that it led to reduced electricity usage of 5.5 percent.

4.4. Product Standards

Product standards may apply to energy-using goods ranging from household appliances to entire buildings. They are designed to require goods to achieve a minimum threshold of energy efficiency. Standards are less economically efficient than pricing policies that directly address the externality, and their effectiveness may be reduced by factors such as the rebound effect; however, they are popular policies around the world. Empirical estimates of the effectiveness of such policies are mixed.

Building codes are an especially popular form of product standard. These codes are the primary policy to influence the energy efficiency of newly constructed or renovated buildings, and a strain of research has estimated their effects on energy consumption. Aroonruengsawat et al. (2012) study the effects of US state-level building codes and find that on average, building codes reduce energy consumption by 2–5 percent.

Another set of studies examines the effects of specific building codes in more detail. Jacobsen and Kotchen (2013) exploit a tightening of Florida's state energy code and, using a regression discontinuity approach, find that the more stringent code caused a 4 percent decrease in annual electricity consumption and a 6 percent decrease in annual natural gas consumption. They also find larger effects on days when the demand for heating and cooling was highest, suggesting that reduced energy consumption was consistent with reduced consumption of air-conditioning and heating. Levinson (2016) raises concerns that these findings do not account for the possibility that newer and older homes vary in ways that are correlated with energy consumption but not caused by the energy code changes. Accounting for this possibility, he finds that houses constructed under the California residential building energy code consumed 10–15 percent less electricity than those built before the codes were instituted; however, they did not increase their electricity usage less on high-temperature days, and the overall electricity consumption decrease was no different from that in other states with less stringent building codes. For natural gas, usage declined by 25 percent, and the newer houses did increase gas usage less on low-temperature days, but the difference was no greater for houses in California than in the other states. These findings imply that building codes are less effective than Jacobsen and Kotchen (2013)'s results suggest.

Novan et al. (2017) enter this debate by looking at the effects of California's energy code and find, in contrast, that new codes led to sizable electricity savings. Specifically, houses in Sacramento County built after the code was implemented increased their cooling energy usage less on high-temperature days, corresponding to a 3 percent reduction in total energy usage. Finally, in a follow-up work, Kotchen (2017) addresses the issues raised in Levinson (2016) and concludes that over a longer time period, electricity consumption in the Florida post-energy code houses converged to the consumption levels of the pre-code houses. However, natural gas consumption did not converge, implying that building codes saved energy in the long run. This debate leads to a more nuanced view of building code effectiveness, in which the outcomes for electricity consumption are mixed, but there is some consensus that effects on natural gas consumption are more substantial. Savings estimates for building codes in general tend to be close to deemed savings estimates, although this is mainly driven by natural gas savings.

Product standards are also applied to vehicles, most notably in the case of the federal CAFE standards. Knittel (2012) studies the trade-offs between fuel economy and other vehicle attributes and finds that the greatest progress in energy efficiency vehicle technology occurred when CAFE standards were most stringent and fuel costs were high, consistent with previous work showing that the rate of energy efficiency innovation depends on both energy prices and

regulatory standards (Newell et al. 1999). Jacobsen (2013) conducts an analysis that incorporates consideration of the effects of CAFE in used vehicle markets and estimates that a 1 mile-per-gallon incremental addition to the stringency of CAFE standards reduces long-term gasoline use by 3 percent. Gillingham (2012) and Jacobsen and van Benthem (2015) investigate two effects that reduce the effectiveness of fuel economy standards like CAFE. Gillingham finds evidence of the rebound effect, in which consumers drive more as a result of lower per-mile fuel costs, and estimates that it lowers total fuel savings by around 15 percent. Jacobsen and van Benthem (2015) find evidence of used car leakage, in which used vehicle owners postpone their decision to scrap vehicles, particularly poor fuel economy vehicles, estimating that this reduces energy savings by 13–16 percent. Klier and Linn (2011) study the trade-offs among vehicle fuel economy, power, and weight and find that the ability to choose among these characteristics lowers the overall cost of achieving fuel economy standard goals. Additionally, Ito and Sallee (2014) find that the common practice of employing attribute-based fuel economy standards, such as vehicle weight or size, as in the recent CAFE standards, distorts incentives to focus on the specified attribute rather than the true target of the standard.

Another set of product standards applies to energy-using appliances such as clothes washer and refrigerators. Houde and Spurlock (2015) conduct an ex post analysis to determine the impacts of past revisions of US energy efficiency appliance standards and find that they resulted in increased quality of products with only modest changes—and in some instances, even decreases—in price.

4.5. Financial Incentives

Another category of energy efficiency programs involves subsidizing energy efficiency investments. Subsidies often focus on replacing specific appliances or investing in building retrofits. Accurately measuring the effectiveness of these programs requires addressing potential free riding, adverse selection, and the rebound effect. Several studies empirically estimate the energy savings and cost-effectiveness of subsidy programs, accounting for these potential issues. Again, studies have generally found mixed results, as shown in Table 1.

Alberini and Towe (2015) use a matching design to assess the impacts of subsidized heat pump replacement in Maryland, finding that the replacements led to lower electricity usage. Alberini et al. (2016) expand on this research, evaluating electricity usage of households that had both incentivized and un-incentivized recent heat pump replacements. They find that those who did not receive incentives reduced their electricity usage by 16 percent, while incentive recipients did not reduce their electricity usage. Furthermore, the larger the rebate a household received, the

less the household reduced energy usage. The authors posit that these results may be due to a large rebound effect. Davis et al. (2014) also use a matching methodology to evaluate a program in Mexico that subsidized replacement of refrigerators and air-conditioning units and find that while refrigerator replacements reduced electricity consumption by 8 percent on average annually, air conditioner replacements actually increased electricity consumption, again consistent with a rebound effect. Rivers and Shiell (2016) examine Canadian subsidies for natural gas furnace retrofits and find strong evidence of free riding, estimating that in the long run, over 80 percent of subsidy recipients would have eventually purchased identical furnaces without a subsidy.

Other studies examine the effects of building retrofits. Fowlie et al. (Forthcoming) use both experimental and quasi-experimental designs to evaluate how receiving free WAP investments, such as insulation, affected energy usage in Michigan. They find that these investments reduced energy consumption on average by 10–19 percent, with the majority of those reductions coming from natural gas. The authors also analyze the indoor temperatures of homes with and without WAP investments and find no evidence of a rebound effect. Graff Zivin and Novan (2016) find that WAP investments decreased consumption by an average of 7 percent in households with air-conditioning units but had no effect on households without air conditioners. Burlig et al. (2017) investigate results from a school retrofitting program. Using a machine learning methodology to generate counterfactual outcomes, they find that the retrofits can lead to energy savings of 2.9 to 4.5 percent. On average, these realized savings represent only about 25 percent of the expected savings, but for some categories of retrofits, such as lighting and heating, ventilation, and air-conditioning, the realized savings are similar to the expected savings.

As an alternative to subsidizing specific efficient investments, other programs directly subsidize (or penalize) consumers for their energy consumption. Ito (2015) studies a financial incentive program in which consumers are offered rebates if they reduce their electricity usage by a specified percentage. The author finds that consumers in inland areas (generally characterized by higher temperatures and lower average income) reduced electricity consumption, while those in coastal areas (lower temperatures and higher average income) did not reduce consumption. Effects were higher for households with air conditioners, suggesting some portion of the effects was due to behavioral changes related to cooling. Another study by Costa and Gerard (2015) examines a temporary program in Brazil aimed to mitigate an energy supply crisis. Residential consumers were given electricity quotas and various financial incentives to meet those quotas. The authors examine whether energy savings continued in the

long term after the termination of the program and find that energy savings persisted up to 10 years after program termination. Surveys of households involved in the study indicate that the main mechanisms in continued energy savings were electricity utilization habits rather than efficient appliance investments.

5. The Cost-Effectiveness of Energy Efficiency in Aggregate

The empirical literature reviewed above focuses on particular efficiency interventions and provides estimates of energy savings associated with specific market interventions and the cost of those savings.⁶ With the exception of building codes, most of the interventions examined are the types of information, behavioral or incentive programs that utilities routinely pursue. A larger question is how effective and cost-effective these utility programs are in the aggregate. Performing this type of assessment is a challenging task. Aggregating findings across different empirical studies provides some insight but gives an insufficient picture of aggregate savings, because of the lack of empirical evidence for many popular types of programs such as residential and commercial lighting upgrades, which account for 60 percent of energy savings reported by utilities for their residential efficiency programs (Hoffman et al. 2015). There is also no single database for assessing aggregate savings outcomes.

One source of information that can be used to assess aggregate savings is the data collected by the Energy Information Administration (EIA) on survey form 861 from utilities and energy efficiency program agencies funded by utility ratepayer dollars on energy efficiency expenditures and associated energy savings.⁷ EIA collects data annually from these entities on efficiency expenditures, and up until 2012, it also collected information on both incremental energy savings in a particular year and annual net savings resulting from efficiency expenditures during that year and all relevant prior years. After 2012, based on consultations with those who design, operate, and evaluate energy efficiency programs, EIA changed its survey questions to focus on annual efficiency spending by program administrators and lifetime energy savings associated with those programs. It also changed the survey to focus specifically on gross savings and thus does not exclude savings by program participants who would have altered their energy use without the program (often referred to as free riders).

⁶ Some studies also provide (or provide instead) estimates of the cost per ton of CO₂ avoided.

⁷ These energy efficiency agencies include Efficiency Vermont, Efficiency Maine, and DC Sustainable Energy Utility.

We used the information on efficiency expenditures collected by EIA and a perpetual inventory methodology used by Gillingham et al. (2006) to develop an estimate of the energy efficiency capital stock from ratepayer-funded programs in each year beginning in 1997 and extended through 2015.⁸ We then applied a rental price of capital to the efficiency capital stock value for each year to estimate annual efficiency costs and divide that by the annual energy savings estimates reported to EIA to determine cost-effectiveness in each year.⁹ Because of the change in the definition of the energy savings information requested by EIA after 2012, we developed a methodology to extend the time series of net annual energy savings through 2015. To do this, we used information on average program lifetimes for utility efficiency investments (between 10 and 11 years) collected by EIA to translate lifetime savings reported to EIA into annual incremental savings to be added to energy savings carried forward from prior years.¹⁰ To create consistent time series of both net and gross energy savings, we used a net-to-gross ratio of 68 percent estimated by Navigant through a study of 42 US and Canadian jurisdictions (Brannan et al. 2013; Violette and Rathburn 2014) to translate gross annual savings post 2013 into net annual savings and, analogously, to translate net savings prior to 2013 into gross savings.¹¹

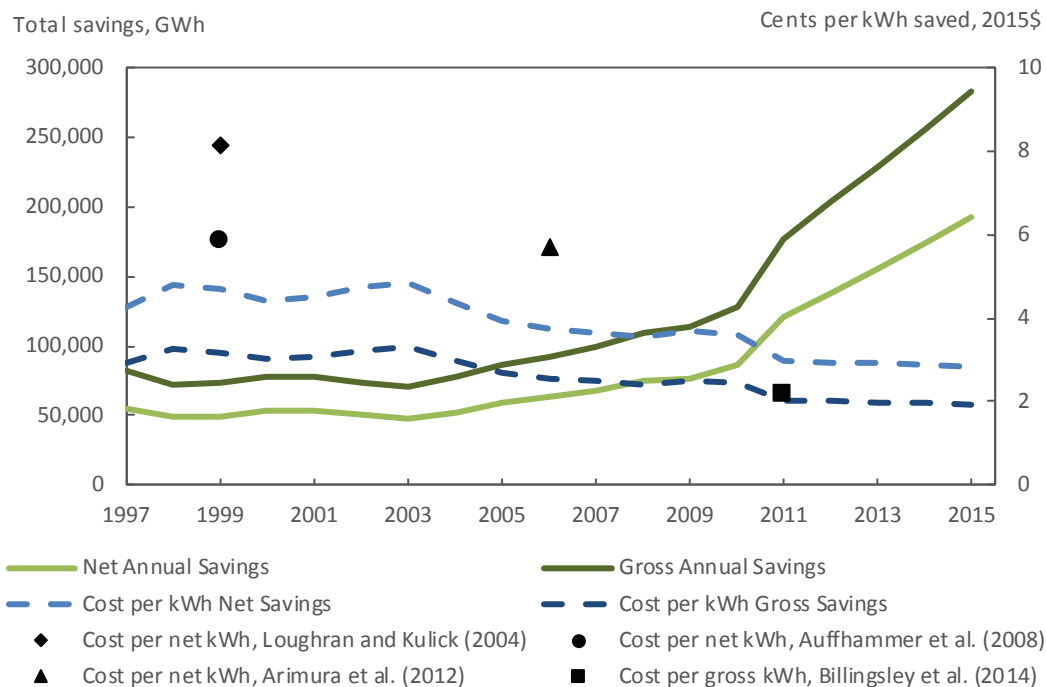
The resulting time series estimates of energy savings and cost-effectiveness (in 2015\$) are presented in Figure 1. Aggregate annual energy savings are measured against the left-hand axis and presented in both gross and net terms, and cost-effectiveness (the dotted lines) is measured against the right-hand axis. The aggregate cost-effectiveness in 2015 was 2.86 cents per kWh net savings, or 1.94 cents per kWh gross savings. Savings have increased over time, particularly since the burst of spending associated with the American Recovery and Reinvestment Act in 2009, which directed roughly \$17 billion dollars to support energy efficiency programs (ACEEE 2009).

⁸ This methodology involves cumulating an annual measure of aggregate energy efficiency capital where expenditures are treated as investment and that investment cumulates and depreciates over time.

⁹ The rental price of capital is the sum of the 11 percent annual depreciation rate for general private industrial equipment used in the national income accounts (Fraumeni 1997) and the real discount rate estimated at 7 percent, for a total rental price of capital of 18 percent.

¹⁰ To estimate annual energy savings for 2013 and beyond, we assumed based on the 10- to 11-year average lifetime for efficiency savings that annual savings reported in 2012 persisted on average for 5 years after 2012 and then added incremental savings (calculated by dividing lifetime savings from expenditures in each subsequent year by average lifetime) in each successive year after 2012.

¹¹ The 68 percent net-to-gross ratio is net of free ridership. When spillovers are added to the mix, the net-to-gross ratio rises to 86 percent for participants and 87 percent for nonparticipants.

Figure 1. Annual Energy Efficiency Savings and Cost-Effectiveness

Notes: Cost-effectiveness point estimates for Loughran and Kulick (2004) and Auffhammer et al. (2008) are the lower bound of estimates. The upper bound estimates are 29.5 and 16.7 cents per kWh, respectively. The point estimates for Arimura et al. (2012) and Billingsley et al. (2014) are the average estimates.

Several important caveats apply to these estimates. First, the savings estimates are reported by utilities and make use of a variety of methodologies, including deemed savings and estimates from engineering studies where installations are measured, but many assumptions about actual use apply. For the most part, the estimates of savings are not based on RCTs or on quasi-experimental studies, and in recent years, they are based on predictions of lifetime savings, so the methodologies are different between this aggregate study and the analyses of specific programs and policies presented above. Second, the cost estimates are program administrator costs and do not capture the full costs to consumers associated with many efficiency interventions. Third, the conversion from gross savings to net savings is based on an aggregate estimate that is applied in all years, but that ratio may be changing over time. For example, if recent programs have a larger component of nudge programs run as RCTs, then the savings estimates for those programs are all net savings, so the ratio of net to gross may be changing over time as the mix of program types evolves.

Other papers have used EIA data to estimate the cost-effectiveness of energy savings programs. The findings of these studies (all point estimates) are included in Figure 1 at the point in time that approximates the end of the time-series component of the panel data used for the

analysis. Three of these papers (Loughran and Kulick 2004; Auffhammer et al. 2008; Arimura et al. 2012) study the relationship between energy efficiency spending and energy sales to conduct an ex post analysis of how energy efficiency programs affected electricity sales at the utility level, which by virtue of the approach provides an estimate of net impacts. In contrast, Billingsley et al. (2014) aggregates information on gross energy savings from state energy efficiency program evaluations similar in spirit to those reported to EIA and information on costs to develop a cost-effectiveness estimate based on gross energy savings.

The papers focused on net impacts on electricity sales provide further evidence on the distinction between ex post evaluation estimates and deemed savings estimates. Loughran and Kulick (2004) find that realized savings were significantly lower than deemed savings, leading to realized cost-effectiveness estimates of 14–22 cents per kWh saved compared with projections of 2–3 cents per kWh saved. However, Auffhammer et al. (2008) revisit the Loughran and Kulick (2004) paper and find instead that the evidence does not allow them to reject the deemed savings estimates. Arimura et al. (2012) also calculate savings estimates within a range of deemed savings and find their cost-effectiveness estimates of 5 cents per kWh saved within the range of cost-effectiveness estimates based on reported expenditures and deemed savings. Following from these findings, the information in the graph shows that ex post evaluations tend to produce higher estimates of cost per kWh saved; however, they are not statistically significantly different from the estimates derived directly from savings data reported to EIA or collected by states.

6. Conclusions

The literature on the motivation for and evaluation of energy efficiency interventions is extensive and rapidly growing. Our review finds that recent academic studies of the topic have produced mixed results. It is clear that many energy efficiency policies lead to energy savings, and some appear to be cost-effective. The evidence suggests that behavioral and information programs, often characterized as “nudge” programs, tend to be the most cost-effective; however, the magnitude of their savings potential is relatively small. More traditional policies, including product standards and financial incentives, may have higher energy savings potential but appear to be more likely subject to misvaluation, and several recent academic papers using “gold standard” evaluation approaches estimate relatively high costs per kWh saved. Our analysis of program cost-effectiveness in aggregate uses the latest EIA data to update prior estimates and finds that as a whole, these programs appear to be relatively cost-effective. In 2015, the surveyed energy efficiency programs had an aggregate cost-effectiveness of 2.86 cents per kWh net savings, or 1.94 cents per kWh gross savings.

We also raise challenging questions for evaluation, the solutions to which generally involve trade-offs. While in an ideal world, all programs would be evaluated using RCTs, their cost and time requirements render this infeasible, and other approaches need to be used in many cases; otherwise, program evaluation may not happen at all. Increasing the use of RCTs will certainly help in improving estimates of program results, although we acknowledge that they are not the sensible approach in every case. Expanded use of quasi-experimental methods exploiting program features such as eligibility criteria or a staggered phase-in of programs over time could help improve the accuracy of program evaluations.

Even when programs are accurately evaluated, another question arises regarding the evaluation's external validity, which is the legitimacy of applying its results to infer the results from other programs in other contexts. If high-quality evaluations like RCTs are externally valid, then one could imagine something akin to benefits transfer to estimate the results of similar programs. Allcott and Mullainathan (2012) provide some evidence suggesting that we should be careful of this option; they find evidence that RCT programs tend to be selected on program-level characteristics that make them favorable for conducting rigorous evaluations and might render extrapolation to other programs less accurate. Finally, the incentives faced by evaluators and academics may influence results, and we believe that efforts to address the issue would help further improve the accuracy of evaluations. One possible solution on the evaluator side could involve allocating evaluation money into a third party-administered fund.

The issue of external validity raises a final question for policymakers: What is the threshold of evidence sufficient to justify repeating or expanding a program? Should results from less-than-ideal studies be used? This is a particularly tricky question in light of calls to decrease the default p -value threshold for statistical significance of a new finding from 0.05 to 0.005 because of a desire to ensure reproducibility (Benjamin et al. 2017). Such a change would require even larger and more expensive RCTs to provide sufficient statistical power, which might mean that even fewer RCTs are run because of their expense. This trade-off between ensuring the credibility of findings and having any findings at all leaves policymakers in a difficult position. It would seem unreasonable to spend most of a program's budget on evaluating the program. Perhaps a compromise could be that every type of program should have at least one large-scale RCT examined at the highest level of rigor, while subsequent program evaluations can use other, less expensive approaches. A major challenge with this compromise proposal is that heterogeneity in program results across time and space may mean that it is inappropriate to use the results of the single RCT to inform decisions made about the same type of program in other contexts.

While we recognize the difficulties in addressing these challenges going forward, we are impressed by the many new developments in this field of inquiry over the past decade and are optimistic that continuing advances in the study of energy efficiency policy coupled with collection of more detailed data on energy use and associated factors will lead to new insights and improved policymaking in the future.

References

- ACEEE. 2009. American Recovery and Reinvestment Act of 2009. Washington, DC: American Council for an Energy-Efficient Economy.
- Alberini A, Gans W, Towe C. 2016. Energy Efficiency Incentives: Do They Work for Heating and Cooling Equipment? Evidence from Maryland Homeowners. *Energy Journal* 37.
- Alberini A, Towe C. 2015. Information v. Energy Efficiency Incentives: Evidence from Residential Electricity Consumption in Maryland. *Energy Economics* 52: S30–S40.
- Allcott H. 2011. Social Norms and Energy Conservation. *Journal of Public Economics* 95: 1082–95.
- Allcott H. 2016. Paternalism and Energy Efficiency: An Overview. *Annual Review of Economics* 8: 145–76.
- Allcott H, Greenstone M. 2012. Is There an Energy Efficiency Gap? *Journal of Economic Perspectives* 26: 3–28.
- Allcott H, Greenstone M. 2017. Measuring the Welfare Effects of Residential Energy Efficiency Programs. E2e Working Paper 028. Berkeley: University of California.
- Allcott H, Mullainathan S. 2010. Behavior and Energy Policy. *Science* 327: 1204–05.
- Allcott H, Mullainathan S. 2012. External Validity and Partner Selection Bias. NBER Working Paper 18373. Cambridge, MA: National Bureau of Economic Research.
- Allcott H, Mullainathan S, Taubinsky D. 2014. Energy Policy with Externalities and Internalities. *Journal of Public Economics* 112: 72–88.
- Allcott H, Rogers T. 2014. The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review* 104: 3003–37.
- Allcott H, Sweeney R. 2017. The Role of Sales Agents in Information Disclosure: Evidence from a Field Experiment. *Management Science* 63: 587–624.
- Allcott H, Wozny N. 2014. Gasoline Prices, Fuel Economy, and the Energy Paradox. *Review of Economics and Statistics* 96: 779–95.
- Angrist JD, Pischke J-S. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

- Arimura TH, Li S, Newell RG, Palmer K. 2012. Cost-Effectiveness of Electricity Energy Efficiency Programs. *Energy Journal* 33: 63–99.
- Aroonruengsawat A, Auffhammer M, Sanstad A. 2012. The Impact of State-Level Building Codes on Residential Energy Consumption. *Energy Journal* 33: 31–52.
- Asensio OI, Delmas MA. 2015. Nonprice Incentives and Energy Conservation. *Proceedings of the National Academy of Sciences* 112: E510–E15.
- Auffhammer M, Blumstein C, Fowlie M. 2008. Demand-Side Management and Energy Efficiency Revisited. *Energy Journal* 29: 91–104.
- Ayres I, Raseman S, Shih A. 2012. Evidence from Two Large Field Experiments That Peer Comparison Feedback Can Reduce Residential Energy Usage. *Journal of Law, Economics, and Organization* 29: 992–1022.
- Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, et al. 2017. Redefine Statistical Significance. *Nature Human Behaviour*. doi:10.1038/s41562-017-0189-z.
- Billingsley MA, Hoffman I, Stuart E, Schiller SR, Goldman CA, LaCommare K. 2014. *The Program Administrator Cost of Saved Energy for Utility Customer-Funded Energy Efficiency Programs*. Berkeley, CA: Lawrence Berkeley National Laboratory.
- Brannan D, Violette DM, Seiden K, Hummer J, Erickson J. 2013. *Custom Free Ridership and Participant Spillover Jurisdictional Review*. Boulder, CO: Navigant Consulting.
- Brennan T. 2013. Energy Efficiency Policy Puzzles. *Energy Journal* 34: 1–25.
- Burlig F, Knittel CR, Rapson D, Reguant M, Wolfram C. 2017. Machine Learning from Schools about Energy Efficiency. E2e Working Paper 032. Berkeley: University of California.
- Busse MR, Knittel CR, Zettelmeyer F. 2013. Are Consumers Myopic? Evidence from New and Used Car Purchases. *American Economic Review* 103: 220–56.
- Byrne DP, La Nauze A, Martin LA. 2014. Tell Me Something I Don't Already Know: Informedness and External Validity in Information Programs. Working Paper 1180. Victoria, Australia: University of Melbourne.
- Costa DL, Kahn ME. 2013. Energy Conservation “Nudges” and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. *Journal of the European Economic Association* 11: 680–702.

- Costa F, Gerard F. 2015. Hysteresis and the Social Cost of Corrective Policies: Evidence from a Temporary Energy Savings Program. Working Paper. Available at SSRN: <https://ssrn.com/abstract=2666704>.
- Creys J, Derkach A, Nyquist S, Ostrowski K, Stephenson J. 2007. *Reducing U.S. Greenhouse Gas Emissions: How Much at What Cost?* New York: McKinsey and Company.
- Darby S. 2006. The Effectiveness of Feedback on Energy Consumption. A Review for DEFRA of the Literature on Metering, Billing and Direct Displays. Oxford: Environmental Change Institute.
- Davis L. 2012. Evaluating the Slow Adoption of Energy Efficient Investments: Are Renters Less Likely to Have Energy Efficient Appliances? In *The Design and Implementation of U.S. Climate Policy*, edited by D Fullerton and C Wolfram, 301–16. Chicago: University of Chicago Press.
- Davis LW, Fuchs A, Gertler P. 2014. Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico. *American Economic Journal: Economic Policy* 6: 207–38.
- Fowlie M, Greenstone M, Wolfram C. Forthcoming. Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program. *Quarterly Journal of Economics*.
- Fraumeni B. 1997. The Measurement of Depreciation in the U.S. National Income and Product Accounts. *Survey of Current Business*, 7–23.
- Frederiks ER, Stenner K, Hobman EV. 2015. Household Energy Use: Applying Behavioural Economics to Understand Consumer Decision-making and Behaviour. *Renewable and Sustainable Energy Reviews* 41: 1385–94.
- Gerarden T, Newell R, Stavins RN. 2017. Assessing the Energy-Efficiency Gap. *Journal of Economic Literature* 55: 1–40.
- Gillingham K. 2012. Selection on Anticipated Driving and the Consumer Response to Changing Gasoline Prices. Working Paper. New Haven, CT: Yale University.
- Gillingham K, Harding M, Rapson D. 2012. Split Incentives in Residential Energy Consumption. *Energy Journal* 33: 37–62.
- Gillingham K, Newell R, Palmer K. 2006. Energy Efficiency Policies: A Retrospective Examination. *Annual Review of Environment and Resources* 31: 161–92.

- Gillingham K, Newell RG, Palmer K. 2009. Energy Efficiency Economics and Policy. *Annual Review of Resource Economics* 1: 597–620.
- Gillingham K, Palmer K. 2014. Bridging the Energy Efficiency Gap: Policy Insights from Economic Theory and Empirical Evidence. *Review of Environmental Economics and Policy* 8: 18–38.
- Gillingham K, Tsvetanov T. 2017. Nudging Energy Efficiency Audits: Evidence from a Field Experiment. USAEE Working Paper 17-320. Cleveland: United States Association for Energy Economics.
- Giraudet L-G, Houde S. 2014. Double Moral Hazard and the Energy Efficiency Gap. E2e Working Paper 009. Berkeley: University of California.
- Giraudet L-G, Houde S, Maher J. 2016. Moral Hazard and the Energy Efficiency Gap: Theory and Evidence. CIRED Working Paper. Centre International de Recherche.
- Graff Zivin J, Novan K. 2016. Upgrading Efficiency and Behavior: Electricity Savings from Residential Weatherization Programs. *Energy Journal* 37: 1–23.
- Granade HC, Creyts J, Derkach A, Farese P, Nyquist S, Ostrowski K. 2009. *Unlocking Energy Efficiency in the U.S. Economy*. New York: McKinsey and Company.
- Greenstone M, Gayer T. 2009. Quasi-Experimental and Experimental Approaches to Environmental Economics. *Journal of Environmental Economics and Management* 57: 21–44.
- Grigolon L, Reynaert M, Verboven F. 2017. Consumer Valuation of Fuel Costs and Tax Policy: Evidence from the European Car Market. TSE Working Paper 17-836. Toulouse, France: Toulouse School of Economics.
- Heutel G. 2017. Prospect Theory and Energy Efficiency. NBER Working Paper 23692. Cambridge, MA: National Bureau of Economic Research.
- Hoffman I, Rybka G, Leventis G, Goldman CA, Schwartz L, et al. 2015. *The Total Cost of Saving Electricity through Utility Customer-Funded Energy Efficiency Programs: Estimates at the National, State, Sector and Program Level*. Berkeley, CA: Lawrence Berkeley National Laboratory.
- Holmes C, Mullen-Trento S. 2017. State Level Electric Energy Efficiency Potential Estimates. Palo Alto, CA: Electric Power Research Institute.

- Houde S, Spurlock CA. 2015. Do Energy Efficiency Standards Improve Quality? Evidence from a Revealed Preference Approach. Working Paper. Berkeley, CA: Lawrence Berkeley National Laboratory.
- Ito K. 2015. Asymmetric Incentives in Subsidies: Evidence from a Large-Scale Electricity Rebate Program. *American Economic Journal: Economic Policy* 7: 209–37.
- Ito K, Sallee JM. 2014. The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel-Economy Standards. NBER Working Paper 20500. Cambridge, MA: National Bureau of Economic Research.
- Jacobsen GD, Kotchen MJ. 2013. Are Building Codes Effective at Saving Energy? Evidence from Residential Billing Data in Florida. *Review of Economics and Statistics* 95: 34–49.
- Jacobsen MR. 2013. Evaluating US Fuel Economy Standards in a Model with Producer and Household Heterogeneity. *American Economic Journal: Economic Policy* 5: 148–87.
- Jacobsen MR, van Benthem AA. 2015. Vehicle Scrappage and Gasoline Policy. *American Economic Review* 105: 1312–38.
- Jaffe AB, Newell RG, Stavins, RN. 2004. Economics of Energy Conservation. In *Encyclopedia of Energy*, ed. C Cleveland. Boston, MA: Elsevier, Inc.
- Jayaweera T, Haeri H, Bruchs D, Keeling J, Khawaja MS, et al. 2013. *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*, Golden, CO: National Renewable Energy Laboratory.
- Jessoe K, Papineau M, Rapson D. 2017. Utilities Included: Split Incentives in Commercial Electricity Contracts. E2e Working Paper 029. Berkeley: University of California.
- Kaufman N, Palmer KL. 2012. Energy Efficiency Program Evaluations: Opportunities for Learning and Inputs to Incentive Mechanisms. *Energy Efficiency* 5: 243–68.
- Klier T, Linn J. 2011. New Vehicle Characteristics and the Cost of the Corporate Average Fuel Economy Standard. *RAND Journal of Economics* 43: 186–213.
- Knittel CR. 2012. Automobiles on Steroids: Product Attribute Trade-Offs and Technological Progress in the Automobile Sector. *American Economic Review* 101: 3368–99.
- Kotchen MJ. 2017. Longer-Run Evidence on Whether Building Energy Codes Reduce Residential Energy Consumption. *Journal of the Association of Environmental and Resource Economists* 4: 135–53.

- LaRiviere J, Holladay S, Novgorodsky D, Price MK. 2014. Prices vs. Nudges: A Large Field Experiment on Energy Efficiency Fixed Cost Investments. NBER Working Paper 22939. Cambridge, MA: National Bureau of Economic Research.
- Levinson A. 2016. How Much Energy Do Building Energy Codes Save? Evidence from California Houses. *American Economic Review* 106: 2867–94.
- List JA, Metcalfe R. 2014. Field Experiments in the Developed World: An Introduction. *Oxford Review of Economic Policy* 30: 585–96.
- Loughran D, Kulick J. 2004. Demand-Side Management and Energy Efficiency in the United States. *Energy Journal* 25: 19–41.
- McKinsey and Company. 2009. *Pathways to a Low-Carbon Economy: Version 2 of the Global Greenhouse Gas Abatement Curve*. New York: McKinsey and Company.
- Miguel E, Kremer M. 2004. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica* 72: 159–217.
- Myers E. 2015. Asymmetric Information in Residential Rental Markets: Implications for the Energy Efficiency Gap. E2e Working Paper 246R. Berkeley: University of California.
- Newell RG, Jaffe AB, Stavins RN. 1999. The Induced Innovation Hypothesis and Energy-Saving Technological Change. *Quarterly Journal of Economics* 114: 941–75.
- Novan K, Smith A, Zhou T. 2017. Residential Building Codes Do Save Energy: Evidence from Hourly Smart-Meter Data. Davis Energy Economics Program Working Paper. Davis: University of California.
- Palmer K. 2016. Comments on EPA’s Draft Evaluation, Measurement and Verification Guidance for Demand-Side Energy Efficiency. Washington, DC: Resources for the Future.
- Palmer K, Walls M, Gerarden T. 2012. An Assessment of Energy-Efficiency Financing Programs. Washington, DC: Resources for the Future.
- Rivers N, Shiell L. 2016. Free-Riding on Energy Efficiency Subsidies: The Case of Natural Gas Furnaces in Canada. *Energy Journal* 37.
- Sallee JM. 2014. Rational Inattention and Energy Efficiency. *Journal of Law and Economics* 57: 781–820.

- Sallee JM, West SE, Fan W. 2009. Do Consumers Recognize the Value of Fuel Economy? Evidence from Used Car Prices and Gasoline Price Fluctuations. *Journal of Public Economics* 135: 61–73.
- Sexton S. 2015. Automatic Bill Payment and Salience Effects: Evidence from Electricity Consumption. *Review of Economics and Statistics* 97: 229–41.
- Smith VK, Moore EM. 2010. Behavioral Economics and Benefit Cost Analysis. *Environmental and Resource Economics* 46: 217–34.
- Stewart J, Todd A. 2015. *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. Golden, CO: National Renewable Energy Laboratory.
- Tsvetanov T, Segerson K. 2013. Re-evaluating the Role of Energy Efficiency Standards: A Behavioral Economics Approach. *Journal of Environmental Economics and Management* 66: 347–63.
- Tsvetanov T, Segerson K. 2014. The Welfare Effects of Energy Efficiency Standards When Choice Sets Matter. *Journal of the Association of Environmental and Resource Economists* 1: 233–71.
- Ungemach C, Camilleri A, Johnson EJ, Larrick RP, Weber EU. Forthcoming. Translated Attributes as Choice Architecture: Aligning Objectives and Choices through Decision Signposts. *Management Science*.
- Vine E, Sullivan M, Lutzenhiser L, Blumstein C, Miller B. 2014. Experimentation and the Evaluation of Energy Efficiency Programs. *Energy Efficiency* 7: 627–40.
- Violette DM, Rathburn P. 2014. *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. Golden, CO: National Renewable Energy Laboratory.
- Wilson E, Christensen C, Horowitz S, Robertson J, Maguire J. 2017. *Electric End-Use Energy Efficiency Potential in the U.S. Single-Family Housing Stock*. Golden, CO: National Renewable Energy Laboratory.

Appendix: Dollars Per Ton of CO₂ Avoided

Authors and year	Evaluation type	Where multiple results available, relevant subset	Dollars per ton of avoided CO₂, 2015\$¹
Efficient Equipment or Energy Savings Subsidy			
Alberini and Towe (2015)	Matching		64
Alberini et al. (2016)	DID regression	Rebate of \$450	62 to 72
Davis et al. (2014)	DID regression	Refrigerators	489
Ito (2015)	RD analysis	Inland	56
		Coastal	2,536
Information Provision			
Alberini and Towe (2015)	Matching		52 to 74
Weatherization			
Fowlie et al. (forthcoming)	RCT		329

¹When dollar years are not specified we use the dollar year from the final year of the program(s).