Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program

Meredith Fowlie, Michael Greenstone, and Catherine Wolfram*

June 14, 2017

Abstract

A growing number of policies and programs aim to increase investment in energy efficiency, because conventional wisdom suggests that people fail to take-up these investments even though they have positive private returns and generate environmental benefits. Many explanations for this energy efficiency gap have been put forward but there has been surprisingly little field testing of whether the conventional wisdom is correct. This paper reports on the results of an experimental evaluation of the nation's largest residential energy efficiency program conducted on a sample of approximately 30,000 households. The findings suggest that the upfront investment costs are about twice the actual energy savings. Further, the model-projected savings are more than three times the actual savings. While this might be attributed to the "rebound" effect – when demand for energy end uses increases as a result of greater efficiency – the paper fails to find evidence of significantly higher indoor temperatures at weatherized homes. Even when accounting for the broader societal benefits derived from emissions reductions, the costs still substantially outweigh the benefits; the average rate of return is approximately -7.8% annually.

*We received many helpful comments from seminar participants at Carnegie Mellon, Columbia University, ETH Zurich, National University of Singapore, NBER Summer Institute, Resources for the Future, the University of Basel, the University of Maryland, the University of Michigan, and the University of Wisconsin. The authors gratefully acknowledge the financial support of the Alfred P. Sloan Foundation, the MacArthur Foundation, the Rockefeller Foundation and the UC Berkeley Energy and Climate Institute, and institutional support from the Poverty Action Lab (JPAL) at MIT, the Center for Local, State, and Urban Policy (CLOSUP) at the University of Michigan, and the Energy Institute at Haas. We thank James Gillan, Brian Goggin, Walter Graf, Erica Myers, Daniel Stuart, and Matthew Woerman for excellent research assistance. We are indebted to Jesse Worker for outstanding management of a challenging project. Finally, we thank our contacts at both our partner utility and the community action agencies, without whom this project would not have been possible.

1 Introduction

Energy efficiency investments are widely believed to offer the rare win-win opportunity. Detailed engineering projections, such as those summarized by the well-known McKinsey curves (McKinsey & Company, 2009), routinely project that investments pay for themselves through the energy saved alone (win #1). Moreover, by reducing the energy necessary to achieve a given level of energy services (e.g., indoor heating), these investments promise to decrease the greenhouse gas emissions causing climate change and other pollutants that compromise human health (win #2).

Despite these apparent opportunities, there is a large and persistent difference between the levels of investment in energy efficiency that are projected to save consumers money and the investments that individuals actually pursue. This has become known as the "efficiency gap." Over the last three decades, a wide variety of explanations have been offered for this apparent failure of consumers to avail themselves of profitable investment opportunities. The most popular explanations have emphasized the possibility of market failures, such as imperfect information, capital market failures, split incentive problems, and behavioral explanations, including myopia, inattentiveness, and prospect theory and reference-point phenomena (see, for example, Allcott and Greenstone, 2012; Gillingham and Palmer, 2014; Gerarden et al., 2015). In contrast, relatively little attention has been paid to the more pedestrian possibility that the real world returns on energy efficiency investments are lower than the engineering models indicate.[1]

Mounting concern about climate change has increased the urgency of understanding this phenomenon. Governments around the world are pursuing a wide range of policies designed to narrow or close the energy efficiency gap. For example, the International Energy Administration has outlined a suite of policies that do not harm economic growth and limit warming to the 2 degrees C recommended by climate scientists; in this scenario, end-use energy efficiency improvements account for 49% of the greenhouse gas emissions abatement in 2020 (IEA, 2013).[2] U.S. electric utilities are rapidly expanding their energy efficiency programs (Barbose et al., 2013), and federal and state regulators routinely tighten energy efficiency building codes, appliance standards, and

[1]Early work by Joskow and Marron (1992) raised concerns about overstated efficiency potential and underscored the importance of using ex post measures of consumer behavior to estimate energy savings.

[2]Indeed, energy efficiency is a central plank for virtually all serious climate mitigation plans (Loftus et al., 2015).

fuel economy standards for automobiles and trucks.

This paper provides the first large-scale field evidence on the returns to energy efficiency investments from a randomized controlled trial. Specifically, we use experimental, as well as quasi-experimental, variation in participation in the federal Weatherization Assistance Program (WAP), to identify the returns to these investments. WAP is the nation's largest residential energy efficiency program and has provided over 7 million low-income households with weatherization assistance since its inception in 1976. Recipient households in our study received approximately $5,150 worth of home improvements on average, at zero out-of-pocket costs. The most common measures included furnace replacement, attic and wall insulation, and infiltration reduction. Importantly, WAP only pays for energy efficiency measures that pass a cost-benefit test, based on ex ante engineering projections, with the aim of ensuring that only beneficial investments are undertaken.

The randomized encouragement design experiment was conducted on a sample of approximately 30,000 Michigan households that were presumptively eligible for participation in WAP. Approximately one quarter of these households were randomly assigned to a treatment group that was encouraged to apply for the program and received significant application assistance. The control households were free to apply for WAP but were not contacted or assisted in any way by our team.

There are three primary findings. First, an aggressive encouragement intervention increased WAP participation from less than 1% in the control group to about 6% in the encouraged group. The encouragement was implemented by a firm with extensive experience managing outreach campaigns and neighborhood canvassing operations, including among low-income populations. The field activities included almost 7,000 home visits, more than 32,000 phone calls, and 2,700 follow-up appointments. Ultimately, these extensive efforts only managed to increase the participation rate by 5 percentage points at a cost of more than $1,000 per weatherized household (Fowlie et al., 2015), revealing low demand in the eligible population for a program with considerable potential benefits.

Second, the findings suggest that the benefits of these investments are substantially less than the upfront costs. We estimate that the WAP energy efficiency investments reduce monthly energy consumption by 10-20% on average. Although this surely provides a substantial assist to participating low-income households in the form of reduced energy bills, the upfront investment costs are about twice the realized energy savings. Further, the ex post estimated savings are roughly 30% of the model-projected savings.[3]

Third, while the modest energy savings might be attributed to the "rebound" effect, when demand for energy end uses increases as a result of greater efficiency, the paper fails to find evidence of significantly higher indoor temperatures at weatherized homes. This finding comes from a novel survey of measured indoor temperatures and thermostat set points that we conducted in the study population. Across a variety of metrics, the WAP energy efficiency investments appear to be poor performers on average. While these investments were free to participating households, we can nevertheless estimate the private returns if households had been responsible for the upfront costs, which is the case for households that do not qualify for WAP. Counting private returns - a household's reductions in energy bills and willingness to pay for any change in indoor temperatures – the annual internal rate of return that would rationalize these efficiency investments is -2.3%. This finding of low, indeed negative, returns suggests that, at least for residential home retrofits, there may not be much of an efficiency gap to explain. Rather, just like in all other sectors of the economy, investments with low returns are not taken-up frequently. Importantly, the engineering model used by WAP is similar, and in some cases identical, to those used to develop recommended residential efficiency investments in the broader population (i.e., not just low-income households).

In principle these investments could still be beneficial socially, but, on average, this is not the case with the measures we evaluate. In contrast to the private calculation, the social one accounts for the benefits of reduced greenhouse gas and local pollutant emissions and the fact that part of households' energy savings is a transfer from other energy consumers, rather than genuine social savings. When we account for these factors, the annual social internal rate of return that would justify these investments is -7.8%. Finally, we also calculate the average cost per ton of avoided CO_2 under a range of assumptions. The most plausible estimates are $200/ton, which is significantly larger than the U.S. government's estimate of the social cost of carbon of roughly $38 (Greenstone

[3] Allcott and Greenstone (2017) evaluate the returns to residential efficiency investments in a different setting using a non-experimental approach to measure energy savings (although they exploit experimental variation in energy efficiency audit take up). They find that the returns to investments are negative socially. Measured natural gas savings amount to only 29% of the savings predicted by the engineering model.

et al., 2013).

This paper makes three primary contributions to the literature on energy efficiency. First and most importantly, it is one of the first studies to provide causal evidence on the returns to residential energy efficiency investments from a large-scale field test. The closest paper methodologically is Dubin et al. (1986), which reports results from an experiment conducted more than three decades ago on fewer than 400 households. There have been a series of other important academic papers that have taken on issues related to residential energy efficiency using various sources of non-experimental variation including Metcalf and Hassett (1999) on returns to insulation improvements, Davis et al. (2014) about appliance replacement programs in Mexico, Jacobsen and Kotchen (2013), Kotchen (2017), and Levinson (2016) on the energy savings associated with building code standards, and Allcott and Greenstone (2017) who use experimental and non-experimental variation to estimate the welfare consequences of residential energy efficiency programs that are not means tested. There is also a large (measured in the many hundreds) "grey" literature that is often commissioned by agents with a stake in the outcome (e.g., utilities that could face regulatory sanctions if their energy efficiency programs fail to deliver promised savings). These studies generally do not meet modern standards of evidence, often relying on simulations, rather than real world data, failing to specify clear counterfactuals/control groups, and/or not reporting standard statistical tests.[4]

Second, this paper is the first effort to directly measure the rebound effect and to develop a framework or method that allows for rebound behavior to count as a benefit or part of the returns to efficiency investments. The previous literature has generally had to rely on methods of detecting rebound that place great faith in the validity of the engineering predictions of savings. For example, Dubin et al. (1986) assume that the engineering estimates are correct and use them to model demand for energy services. Because they find elastic demand for energy services, the paper concludes that there is a rebound effect. Though the existence of the rebound effect has been

[4] A survey by the American Council for an Energy Efficient Economy revealed that in 81% of US states with significant energy efficiency activities, regulators use ex-ante engineering estimates to measure the savings from the programs (Kushler et al., 2012). Reports that do endeavor to develop ex post estimates of savings vary in terms of analytical rigor, rarely meeting standards applied in refereed journals (e.g., key details necessary to follow the approach to evaluation are often missing). Further, many do not use modern approaches to statistical inference; it is not uncommon, for example, for these studies to simply report differences without an accompanying statistical test of significance. See, for example, CPUC (2015) for a summary of grey literature evaluations of California efficiency programs implemented from 2010-2012.

the subject of much debate (Gillingham et al., 2013), our study is the first to provide a direct field test of this phenomenon across a broad spectrum of residential energy efficiency investments.

Third, the paper provides a demanding and credible test of the validity of engineering models' predictions about the returns to residential energy efficiency investments. This test is important because the case for the energy efficiency gap – the motivation for almost all energy efficiency policy – is derived from the promise of abnormally high returns to these investments; yet, the promises are derived from engineering models that have largely been untested in the field. Further, the paper's comparison of actual and modeled savings is conducted among the most frequent residential efficiency investments. Finally, WAP's engineering model is used extensively in other contexts and there is considerable similarity between its underlying assumptions that generate ex ante predictions in the returns to efficiency investments and the assumptions in competing engineering models. Thus, it seems that the pedestrian explanation for the energy efficiency gap that low take-up is due to low returns deserves greater consideration.

The paper proceeds as follows. Section 2 introduces a conceptual framework useful for deriving the private value of energy efficiency investments, including any behavioral adjustments they may cause. Section 3 outlines key details on the Weatherization Assistance Program and describes our study design. Section 4 describes the data sources and provides summary statistics. Section 5 reports the main results on actual savings and on observed rebound effects. Section 6 develops measures of both the private and social returns to energy efficiency investments and discusses the results' external validity to other settings. Section 7 concludes.

2 Conceptual Framework for Measuring Private Returns

Private gains from an investment in energy efficiency are realized through two main channels: reduced energy consumption and increased consumption of energy services (e.g. lighting, space heating, air conditioning) due to reductions in the price of energy services. With respect to the first channel, any reduction in dollars spent on energy can be allocated to other forms of welfare-enhancing consumption. The second channel becomes important when an efficiency-induced reduction in energy end-use costs leads to an increase or "rebound" in the demand for the energy service.

For utility-maximizing agents, any re-optimization of consumption that occurs in response to an efficiency improvement will be (weakly) welfare improving.

These basic ideas are illustrated in the upper quadrant of Figure 1, which plots consumption of a particular energy service, home heating, on the horizontal axis and consumption of the numeraire (i.e., all other goods), X, on the vertical axis. This framework focuses on home heating because it is a particularly important end-use in our empirical setting; over 90% of projected energy savings from the weatherization investments we analyze are heating related.

The two downward sloping lines in the upper quadrant of the figure reflect budget constraints, the lower before the efficiency improvement (i.e., the status quo) and the higher after weatherization. The budget constraint pivots post-weatherization because the price of heating services (e.g., the price of keeping the house at a certain indoor temperature in winter) has fallen; energy efficiency improvements reduce the cost of purchasing any given level of thermal comfort.

The figure also illustrates a family of indifference curves for a representative consumer, each of which trace out the bundles of the numeraire, X, and heating services, H, that deliver the same level of utility. The U-shape of the indifference curves reflects that households do not like to be too hot or too cold. In the status quo (i.e., absent an efficiency improvement), the representative agent will maximize utility through the choice of H_0 and X_0. The weatherization-induced expansion of the budget constraint allows the agent to move to a higher level of utility associated with the bundle of H_1 and X_1. In the figure, status quo consumption occurs below the satiation point for thermal comfort. Thus, when the price falls, demand for heating services increases by $H_1 - H_0$. The positive income effect also increases consumption of the numeraire by $X_1 - X_0$.

The paper's empirical challenge is to measure the welfare gains conferred by weatherization investments. Our empirical setting allows us to develop a measure of willingness to pay (WTP) for weatherization that accounts for both reductions in energy expenditures and increased consumption of heating services. The effect of energy efficiency improvements on other consumption can be measured using data on monthly energy expenditures which vary one-for-one with the consumption of the numeraire. Put another way, a \$1 decrease in energy expenditures allows for a \$1 increase in consumption of all other goods. Measuring willingness to pay for the increase in heating services (i.e., the direct rebound effect) is more challenging because demand for energy services, such as heating, is not readily observable in household energy consumption or expenditure data.

To obtain an estimate of the efficiency-induced increase in demand for heating ($H_1 - H_0$ in the figure), we conduct a survey of indoor temperatures in weatherized and non-weatherized homes. With this estimate of the treatment-induced change in indoor temperatures in hand, we can construct bounds for the welfare consequences of any observed increase in warmth by imposing some structure on the relationship between energy demand and heating services.[5] The bottom quadrant of Figure 1 plots a representative building-specific relationship between heating services and the energy required to achieve that temperature, E, holding constant outdoor temperatures and building characteristics, Z.[6] Efficiency improvements to the building envelope (e.g., insulation improvements, window sealing, a furnace upgrade) reduce the energy required to deliver any given level of heating services. This implies that the slope of the relationship between heating services and energy consumption becomes less steep following an efficiency improvement, depicted as the pivot in the function that determine energy consumption from $E(H; SQ, Z)$ to $E(H; W, Z)$ in the figure. Section 5.3 details our approach to empirically estimating the relationship between energy consumption and heating demand, which, in practice, is measured as indoor temperature.

We use the empirical relationship between indoor temperatures and energy consumption to construct bounds on the utility gains from any efficiency-induced increase in the demand for indoor temperature using revealed preference logic. Since the agent chooses to increase heating services from H_0 to H_1 following the efficiency improvement, it follows that a lower bound for the associated increased utility is the increase in heating costs incurred after weatherization. This value is represented by P_E*($E_{1,W}$-$E_{0,W}$) in the figure, where P_E is the exogenous price of energy. Note that

[5]Our measure of the returns to weatherization investments is predicated on a measure of increased indoor temperature, but we do not account for any increase in comfort conditional on indoor air temperature. Researchers have noted that improvements in insulation enhance comfort by reducing drafts and increasing humidity (Schwarz and Taylor, 1995). If a home is less drafty and more humid, a consumer may be able to achieve the same comfort at a lower indoor temperature. Consequently, what we measure here is the net effect of efficiency improvements on heating demand. We return to this point below.

[6]We assume a linear relationship between air temperature and energy demand over the relevant range of temperatures; we provide empirical support for this assumption in Section 5. Also, for ease of exposition, this figure depicts the limited range of indoor air temperatures over which energy consumption is increasing in indoor temperatures. This range is bounded from below by the outdoor air temperature (if the thermostat is set below the outdoor air temperature, no heating services are required). The lines stop where the capacity of the home heating system binds. Beyond this point, increased energy consumption ceases to generate heating services.

the agent chose less heating than H_1 prior to the efficiency improvement. Thus, the cost of the change in heating services prior to weatherization, measured by $P_E{*}(E_{1,SQ} - E_{0,SQ})$, provides an upper bound on the welfare gain. Preferences revealed prior to the efficiency improvement suggest that the agent values the increase in heating services less than this incremental cost.

With this conceptual framework as a guide, the paper will estimate the causal effect of WAP participation on annual energy consumption and willingness to pay for changes in heating services.

3 Background and Study Design

3.1 Weatherization Assistance Program

The Weatherization Assistance Program (WAP) is the nation's largest residential energy-efficiency program. WAP supports improvements in the energy efficiency of dwellings occupied by low-income families. Since its inception in 1976, over 7 million low-income households have received weatherization assistance through the program. The American Recovery and Reinvestment Act PL111-5 (ARRA) dramatically increased the scale and scope of WAP.[7] Our analysis seeks to estimate the impacts of weatherization assistance over the ARRA-funded time period.

WAP funds are distributed to states based on a formula tied to a state's climate, the number of low-income residents, and their typical energy bills. The states distribute WAP money to over 1,000 local sub-grantees, which are typically community action agencies (CAAs) or similar nonprofit groups. These sub-grantees are then tasked with identifying and serving eligible households. The average participating household in our data received an average of \$4,130 of energy efficiency investments and over \$1,000 worth of additional house improvements at zero out-of-pocket costs.[8]

Before implementing a weatherization retrofit, CAA program staff conduct an energy audit of the home. The purpose of the audit is to recommend specific efficiency improvements for the home. During the visit, program auditors collect detailed information about the building structure,

[7] Funding increased from \$450 million annually in 2009 to almost \$5 billion for the 2011-2012 program years. Under the ARRA-funded program, all owner-occupied households at or below 200% of the poverty line were eligible to apply for assistance.

[8] During the course of the retrofit, additional costs are incurred to ensure the safe and effective installation of the weatherization measures. For example, electric wiring updates or asbestos removal may be required to ensure a safe working environment. Once these safety measures are accounted for, the average cost per household is \$5,155.

heating and cooling systems, appliances, insulation, ventilation, etc. This information is combined with local climate conditions and retrofit measure costs, then fed into a computer-based audit tool: the National Energy Audit Tool (NEAT). This tool is comprised of a set of easy-to-use but advanced energy audit computer programs that identify the cost-effective energy-efficiency retrofit measures for a home after taking into account local weather conditions, retrofit measure costs, fuel costs, and specific construction details of the home. NEAT was designed specifically to help states and local weatherization agencies implement WAP, although it is also a widely used tool for private-sector weatherization audits (EERE, 2010).[9]

NEAT produces an estimate of the energy savings and costs associated with different combinations of efficiency measures. The present value of projected energy savings are calculated using a discount rate of 3% and an engineering estimate of the lifespan of the measures. The 3% discount rate is consistent with OMB guidance on how to evaluate benefits of federal spending but is substantially lower than the cost of borrowing for most households, especially low income ones. The WAP program requires that all recommended measures return a minimum of $1.00 in incremental savings for every $1.00 expended in labor and material costs.

The process of applying for weatherization is highly onerous and time intensive, at least partially to prevent fraud. Applicants must submit extensive paperwork documenting their eligibility, including utility bills, earnings documentation, social security cards for all residents of the home and deeds to the home. Local agencies often identify potential applicants from the pool of households that are receiving other social services, although walk-in clients are routinely admitted. Agencies screen potential applicants for eligibility. Eligible applicants are then prioritized following guidelines that recommend CAAs assign the household a high rank if it has an elderly resident, a person with disabilities or children, or where the occupants typically face a high energy burden (energy as a share of income) or have high residential energy use (see 10 CFR 440.16(b) (1-5)).[10] Given the non-random nature of the process by which households end up in the program, any comparisons of energy consumption across weatherized and un-weatherized households risks confounding the effect

[9] The NEAT audit tool is documented extensively at http://weatherization.ornl.gov/assistant.shtml.

[10] Given the high ARRA funding levels during our study period, the prioritization scheme was less binding as compared to lower funding periods.

of the program with pre-existing differences in determinants of energy consumption.

3.2 Research Design

Our analysis focuses on a sample of low-income Michigan households. Michigan is one of the largest recipients of WAP program funding on account of its cold winters and large low-income population. Further, we were able to secure collaborative agreements with a major Michigan utility and five CAAs working in this utility's service territory. This allowed us access to detailed, household-level energy consumption and weatherization program data. A close collaboration with two of these agencies allowed us to implement a large-scale field experiment.

Michigan received over $200 million in ARRA funding for weatherization assistance. A series of bureaucratic delays - for example, ensuring that contractors were paid prevailing wages - delayed spending until early 2011 (Radnofsky, 2010). Around March 2011, weatherization activities in Michigan increased markedly. All stimulus funds had to be spent by March 2012. After that point, the pace of weatherization activity dropped precipitously. We stopped collecting data in April 2014. A number of households that applied for WAP had not yet received services by the end of our study period.

The paper's empirical challenge is to obtain causal estimates of the effect of participation in the WAP program on energy consumption and indoor heating demand. We estimate versions of the following equation:

$$ln\left(y_{imt}\right) = \beta\mathbf{1}\{WAP\}_{imt} + \alpha_{im} + \alpha_{mt} + \epsilon_{imt}, \tag{1}$$

where $ln\left(y_{imt}\right)$ measures the natural log of energy consumption (natural gas, electricity, or a combined MMBtu measure) at household i in month m and year t. The equation also includes household-by-month-of-year fixed effects, α_{im}, to account for permanent differences in a household's energy consumption across months. It is possible to include such a rich set of fixed effects because we observe households across multiple years. The model also includes month-by-year fixed effects, α_{mt}, to adjust for the average effects of time-varying factors (e.g., winter temperature) that generate variation in average consumption across all households.[11]

The parameter of interest is β, which measures the mean difference in energy consumption subsequent to the completion of WAP energy efficiency investments, after adjustment for the fixed effects. It is a difference-in-differences estimator that compares the change in energy consumption after weatherization to before, relative to consumption among households that have either not yet weatherized through WAP or never did during our sample period. We take two approaches to developing unbiased estimates of β, described in the next two sub-sections.

3.2.1 Randomized Encouragement Design

The paper's primary empirical estimates are derived from an experimental research design. The basis of the experiment is a randomly assigned encouragement intervention that aims to increase the probability of treatment households' participation in WAP through recruitment and significant application assistance. This randomized encouragement design is particularly useful in contexts such as this one where program participation cannot be directly denied - we cannot prohibit eligible households from participating in this federal program.

The recruitment and assistance was conducted by FieldWorks LLC, a private company that specializes in running neighborhood canvassing operations and managing outreach campaigns. We chose them because they had substantial experience working with low-income populations and their staff had generated millions of phone calls and knocked on millions of doors in previous engagements.

The experimental sample comprised 34,161 households that were both presumptively eligible for WAP and located within the counties served by the two CAAs that partnered with us for the field experiment. Approximately one quarter of the sample households were randomly assigned to our encouragement "treatment." For the remaining households assigned to the control group, we simply observe energy consumption and program participation decisions.

The encouragement campaign got underway as ARRA funds began flowing to the implementing agencies. Encouragement activities ran from March to May 2011. Table 1 summarizes our encouragement and enrollment activities in detail. During the encouragement phase, field staff

[11] We also estimate specifications that include month-by-year-by-county and month-by-year-by-billing segment fixed effects.

made almost 7,000 initial, in-person house visits.[12] The ground operations were complemented with 23,500 targeted "robo-calls" to raise awareness of both the weatherization program and our encouragement campaign.

After the encouragement phase, we transitioned to an enrollment phase, which lasted through February 2012. Our staff made over 9,000 personal phone calls to provide assistance with the onerous application process and to coordinate in-person meetings. Over the course of 2,720 home visits, our field staff helped individuals assemble documentation and complete paperwork. In some cases, our field staff provided transportation to and from the program agency offices.

The final row of Table 1 reports that we spent around $475,000 on the encouragement or a little more than $55 per household in the treatment group. It is noteworthy that we did not initially intend to devote such extensive efforts and resources to the encourage and enrollment phases. However, the early results suggested that we were failing to have a substantial impact on applications and concerns about the ultimate precision of our estimated treatment effects motivated us to raise additional funds to expand the share of treated households.[13]

We use the random assignment to encouragement as an instrumental variable for weatherization status. In a first stage, we use OLS to estimate:

$$\mathbf{1}\{WAP\}_{imt} = \theta\mathbf{1}\{Encouraged\}_{imt} + \delta_{im} + \delta_{mt} + \eta_{imt}, \tag{2}$$

where the dependent variable is an indicator variable that switches from zero to one in the month after a household's weatherization retrofit is complete. The indicator variable, $\mathbf{1}\{Encouraged\}_{imt}$ is set to zero for all households prior to the encouragement intervention. After March 2011, this indicator switches to 1 for the 25% of the households randomly assigned to the treatment group.

We substitute $\mathbf{1}\{W\hat{A}P\}_{imt}$ from the estimation of equation (2) to fit equation (1) and obtain

[12]Most- but not all- houses assigned to the treated group were contacted. A small fraction were deemed inaccessible (e.g., because of a locked gate).

[13]On the one hand, our encouragement costs may have been higher than necessary. To our knowledge, ours was the first encouragement program for WAP, and we learned from our initial experiences how to refine our intervention. On the other hand, the costs in Table 1 do not reflect the time that the research team devoted to overseeing the encouragement effort. The Online Appendix 1 and Fowlie et al. (2015) provide more details on the encouragement and application assistance programs.

$\hat{\beta_{IV}}$. In this instrumental variables (IV) framework, $\hat{\beta_{IV}}$ is identified using the exogenous variation in program participation that is generated via the random assignment of encouragement.

3.2.2 Quasi-Experimental Design

The quasi-experimental research design uses data collected from households that applied for WAP after March 2011 with one of the five CAAs that shared data with us to estimate a version of equation (1). In this sample of households, we compare patterns in energy consumption among weatherized households and households that applied for WAP but had not been weatherized by mid-2014, when our data ends. Forty percent of applicants in our data were weatherized through WAP during this period.

A critical issue for the validity of the quasi-experimental estimates is the manner in which households in this sample were chosen for weatherization. The road from application to energy efficiency investments is long and there are many potential off-ramps. Applicant households may fail to complete the necessary – and involved – paperwork or may be deemed ineligible based on the information they provide. Once paperwork is completed successfully, households are put on a list, and the waiting times can exceed one year. Once at the top of the list, homeowners must schedule an energy audit. Households may fail to receive weatherization if they miss an audit appointment, or if the auditors discover risks to WAP contractors (e.g., asbestos in the home). Because of significant delays in ramping up weatherization activities under ARRA, agencies were unable to complete the weatherizations they anticipated prior to the March 2012 ARRA deadline, which helps to explain why fewer than half of the applicants in our sample were weatherized by mid-2014. Some of the explanations for variation in treatment status among applicants are orthogonal to household characteristics that determine energy consumption patterns, while others clearly are not.

The primary threat to consistent estimation of β using the quasi-experimental approach is the possibility that time-varying factors that affect household demand for energy also influence WAP participation. For instance, households may push forward their WAP application more aggressively when they anticipate an increase in their demand for energy as would be the case when the number of people in their household increases or they lose a job and expect to spend more time at home. While the quasi-experimental approach does not have a direct solution to this threat to identification, we take steps to balance observable characteristics and trends across weatherized and unweatherized households. To evaluate the robustness of our findings, we re-estimate equation (1), using alternative sets of controls and regression weights. Our preferred quasi-experimental specification re-weights control observations in order to achieve covariate balance across weatherized and un-weatherized controls, explained in more detail below.

Of course, it is possible that the effect of a WAP weatherization varies across households (i.e., β_i) either based on observables or unobservables. Indeed, it is possible that households sort into WAP participation based on knowledge of their idiosyncratic returns; this has been labeled "essential heterogeneity" Heckman et al. (2006). In the presence of heterogeneity in how weatherization retrofits impact residential energy consumption, the expectation of unbiased quasi-experimental and experimental estimates of β_i need not be equivalent. The quasi-experimental approach is designed to provide an estimate of the average treatment effect on all treated households (ATET). In contrast, with some additional assumptions (Angrist et al., 1996), the randomized encouragement design estimates the so-called local average treatment effect (LATE), or the average effect for the subset of the population who must be encouraged to participate in the program (i.e., the compliers). Therefore, significant differences in the quasi-experimental and experimental estimates could be due to bias in the former or differences in the LATE and ATET. Section 5.2 and Section 6 of the online appendix explore the sources of the differences in our quasi-experimental and experimental estimates of the effect of WAP weatherization.

4 Data Sources and Summary Statistics

4.1 Data Sources

The data collected to support this analysis correspond to two overlapping groups of households. The first group comprises the 34,161 households in our experimental sample drawn from the counties served by our two partner CAAs. The second group of households corresponds to our quasi-experimental research design. As this design did not require the same degree of coordination with our agency partners, we were able to expand the scope of this sample by collecting detailed data from three additional implementing agencies. The quasi-experimental sample includes the 7,304 households that applied for weatherization assistance at these five agencies. The quasi-experimental sample is smaller overall but has a larger number of applicants and weatherized households, relative to the experimental sample. The two groups overlap as 1,773 applicant households are also part of our experimental sample.

4.1.1 Energy Consumption Data

We obtained monthly natural gas and electricity consumption data over the period June 2008 to May 2014. This period includes at least two years of pre-retrofit data for all weatherized households in our sample. The utility data track monthly kilowatt-hours (kWh) of electricity and thousand cubic feet (Mcf) of gas used at the dwelling. Some specifications convert both natural gas and electricity consumption to million British thermal units (MMBtu) using the conversion factor employed by the NEAT audit tool.

Energy consumption records obtained from the utility are merged with households in our experimental sample and the applicant data obtained from the five implementing agencies. Data are merged using detailed name and address information. Not all households find an exact match in consumption records. Match rates are 88% and 81% in our experimental and quasi-experimental samples, respectively. The higher match rate in our experimental sample is to be expected; when selecting this sample we focused exclusively on zip codes within the territory of our partner utility. Online Appendix 2 provides more detail on missing data and attrition. After dropping observations with no match, our experimental sample is 28,888.

4.1.2 Application Data

We obtained the data about households collected through the application process including information used to determine program eligibility (e.g., income, household size, number of children) and information that can be used to prioritize successful applicants (e.g., elderly residents). Application materials also collect other demographic characteristics including race and education. These data are used to assess balance across the quasi-experimental treatment and control groups.

4.1.3 Efficiency Audit Data

Detailed information about the dwelling is collected during the household-level energy efficiency audit, which is a critical part of the WAP implementation as described above. An output of these audits is the list of energy efficiency measures for which projected energy savings exceed projected costs. These are the measures that comprise the weatherization retrofit. We also acquired the job reports that are filed after the weatherization retrofit is completed. These allow us to confirm that the recommended measures were installed and to compare realized costs with projections.

4.1.4 Indoor Temperature Data

Two years after the encouragement effort was initiated, we randomly selected a subset of weatherized and unweatherized households for a field survey. These households were selected from the quasi-experimental sample. The primary purpose of the survey was to measure thermostat set points and indoor temperatures to test for a direct "rebound effect."

Michigan field staff attempted to contact 6,400 households on cold days (projected maximum temperature below 45 degrees Fahrenheit) in March and early April 2013. With the homeowner's permission, surveyors entered the home, closed the door, moved to the center of the room, and recorded multiple indoor air temperature measurements using two different thermometers. Of our initially targeted sample, surveyors spoke with 1,658 homeowners. Of these, 899 allowed us to enter their homes and record their thermostat set point and 688 allowed us to close the door, and collect two or more indoor thermometer readings.

4.1.5 Demographic Data

Detailed data on occupant and dwelling characteristics are not available for the households in our study that did not apply for weatherization assistance. To construct a more complete picture, we map households in the study to census-block level data on variables of interest (such as home age, poverty level of residents, etc.). We use these data to assess balance in observable characteristics across our experimental encouraged and control groups.

4.2 Summary Statistics

Table 2 summarizes pre-treatment information on the households in both the experimental and quasi-experimental samples. The top panel summarizes monthly energy consumption during the two years immediately preceding the treatment period. The first two columns summarize means for the randomized encouragement and experimental control groups, respectively. The third column reports p-values for the differences between the treatment and control groups.

There are seasonal patterns in energy consumption that differ across natural gas and electricity. Winter natural gas consumption (which is dominated by space heating) is significantly higher than summer gas use (comprised primarily of hot water heating and cooking). Electricity usage is fairly consistent across seasons. Because households in the experimental sample were randomly assigned to the encouraged and control groups, it is unsurprising that differences in energy consumption across these two groups are all small and statistically indistinguishable from zero.

The table also provides an opportunity to judge the credibility of the comparisons that underlie the quasi-experimental estimates. The fourth column reports on all weatherized households in the territory covered by the five CAAs. The fifth column summarizes households in these territories that applied for weatherization but had not received weatherization assistance as of April 2014. In practice, the variation in the weatherization dates means that the identification of the quasi-experimental estimator is not just based on comparisons between the samples summarized in columns (4) and (5), but also relies on within household comparisons. It is nevertheless informative to compare these two sets of households. P-values for the mean differences in column (6) show that weatherized households have historically consumed significantly less natural gas than the unweatherized applicants during both winter and summer months.

To assess the balance in characteristics other than energy consumption across the experimental control and encouraged groups, we locate households in the experimental sample within census blocks and summarize dwelling and household information at the census-block level. Panel B summarizes these data. P-values in column (3) account for clustering within census blocks. None of these differences in group means are statistically distinguishable from zero.

Panel C of Table 2 summarizes the detailed demographic information and dwelling characteristics that are collected for most clients as part of the application process. Our summary of these variables focuses exclusively on program applicants because these data are not available for the majority of households in the experimental sample that did not apply to the program. Column (6) documents important differences between the weatherized and unweatherized applicant subsamples. Note that differences across Panels B and C should be interpreted carefully because variables are measured differently across the two data sources. For example, income reported in Panel B summarizes incomes at the census-block level and includes all income received on a regular basis. Income reported for the purpose of determining weatherization program eligibility (measured at the household level and summarized in Panel C) includes cash receipts (e.g., wages, salaries, pensions) but excludes other potential income sources (such as capital gains, tax refunds, or housing allowances). Weatherized households have higher incomes as compared to unsuccessful applicants. In addition, applicants who ultimately receive weatherization are farther above the poverty line, which adjusts for household size, than the unsuccessful applicants. Weatherized households also report having more children, are more likely to report an elderly or disabled resident, and are more likely to use natural gas as their primary heat source.

The significant differences between the weatherized and unweatherized applicants motivate us to re-weight observations in the control group so that observable factors are distributed similarly in the weatherized and unweatherized applicant groups. Although our preferred specifications includes household-by-month-of-sample fixed effects that control for all time invariant differences between households, we are concerned that observable fixed differences might be correlated with time-varying differences. We use an estimated propensity score to balance covariates that presumably play a role in determining program take-up. Appendix 3 provides a detailed discussion of the participation equation we estimate. Column (7) reports p-values for the mean differences between weatherized households and matched controls. Reassuringly, the differences in average covariate values across weatherized households and propensity-score weighted controls are statistically indistinguishable.

In the Online Appendix, Table 1 summarizes some of the detailed information collected during the household energy efficiency audits. A typical weatherization retrofit involved several measures such as furnace replacement (34% of retrofits), attic insulation (85%), wall insulation (44%), and infiltration reduction (76%). The NEAT model predicts that on average these investments will reduce natural gas consumption by 46% and electricity consumption by 15%. The average project involved over $5,150 in total expenditures. This includes materials, labor, and construction costs, but does not include any program overhead. Using a 3% discount rate, the projected net present value of energy bill savings average $10,611. The average projected savings:investment ratio (across measures) exceeds 2:1.

5 Results

5.1 Experimental Estimates of Energy Savings

5.1.1 First-Stage: Program Take-Up

It may seem straightforward to encourage households to participate in a program that provides free efficiency retrofits worth an average of approximately $5,000 that are designed to significantly reduce energy expenditures. In our experience, that was hardly the case. The impact of reducing barriers to participation (e.g., information and process costs) on program uptake is of independent interest both to policymakers and researchers. This section evaluates the impact of our intervention on a multi-part participation process.

Table 3 summarizes program take-up at three separate junctures. In the first stage, our goal was to increase the share of households filing applications. The coefficient estimate in column (1) indicates that the encouragement intervention increased the rate of application to the program by 13 percentage points from the control group mean of 2%. Column (2) reveals that the fraction of households who received an energy audit was 5 percentage points higher in the encouraged group (off a base of about 1%). As we discussed above, several factors can explain why so many households that submit an application are not audited, including if the household fails to follow through on requests for more information or if the submitted information indicates that the household does not meet the program's eligibility requirements.

Column (3) of Table 3 documents that the treatment increased the fraction of households that were successfully weatherized by about 5 percentage points, against a 1% rate in the control group.[14] The encouragement treatment is a statistically significant predictor of weatherization, which we will use to instrument for program participation.

The low take-up rates in the encouraged group are quite striking. Program participants receive substantive home improvements, yet incur no out-of-pocket expenses. All households in the encouraged group received some information about the program via a phone call or door hanger. A majority of households (i.e., those who spoke with our canvassers in person or by phone) received further information about our offer of application assistance. Given that households had detailed, specific information about the program, it seems reasonable to surmise that some combination of high perceived costs of applying for the program, low expectation of an application leading to a weatherization, high unmeasured process costs, and low expected benefits of participating in the program are impediments to WAP participation. In the end, the average cost of encouragement per completed weatherization was about $1,050, which is more than 20% of the average costs of measures installed. See Fowlie et al. (2015) for further details.

5.1.2 Instrumental Variables Estimates of Energy Savings

Figure 2 provides a graphical overview of the experimental estimates. The broken line shows the cumulative effect of the randomly assigned encouragement intervention on the monthly rate of weatherization, relative to the control group which received no encouragement. This effect increases over time as the treatment households submit applications and receive weatherization assistance. The figure also plots month-by-month estimates of intent to treat (ITT) effects on energy consumption. Conceptually, monthly estimates of the local average effect of weatherization on energy consumption can be constructed as a ratio of the monthly ITT estimates and the corresponding effect of encouragement on program participation.

Panel A of Table 4 summarizes results that relate measures of energy consumption to the

[14] A small fraction of households get audited but not weatherized, primarily because the auditors deem the home a possible danger to weatherization contractors (e.g., due to the presence of asbestos).

WAP participation indicator. In the first two columns, the dependent variable is the log of total energy consumption (MMBtu/month). In the third and fourth columns, the dependent variables are the log of natural gas and electricity consumption, respectively. Standard errors are clustered by household in all specifications.

For comparison, the first column reports the coefficient associated with the WAP indicator from an OLS regression with month-of-sample and household-month fixed effects that is fit to the experimental sample. This specification indicates that WAP participation is associated with a 10% decline in energy consumption.

The IV estimates based on our randomized encouragement design are reported in columns (2) - (4). The estimate in column (2) indicates that WAP participation causes a reduction in monthly energy consumption of approximately 20% among households that were encouraged into the WAP program. This is the first indication that the realized savings from the WAP-induced energy efficiency investments are substantially smaller than the projections from the engineering model. Columns (3) and (4) report local average treatment effects for natural gas and electricity, respectively. Natural gas accounts for 94% of projected savings (measured in MMBtu), so it is not surprising that natural gas consumption is more significantly impacted by weatherization.[15]

Our IV strategy is predicated on an exclusion restriction: we assume that our encouragement activities affected energy consumption only through its effect on participation in the WAP program. To informally test whether the treatment's encouragement activities had a direct effect on energy consumption, we test for an effect of our encouragement activities on the households in the encouraged group that did not receive weatherization assistance. We fail to reject the null of no effect on energy consumption among these households.[16]

[15]The coefficient in column (2) is not a weighted average of the gas and electricity coefficients in columns (3) and (4) for two reasons. First, the samples differ slightly across the columns, and second the coefficients on the fixed effects and other covariates are not constrained to be equal in columns (3) and (4). It is also worth noting that the effect of weatherization on gas consumption is more precisely estimated. This is not surprising, because natural gas consumption is driven primarily by the end uses targeted by weatherization (space and water heating), whereas electricity consumption is derived from many end uses that are unaffected by weatherization (e.g., lighting and appliances).

[16]To conduct this test, we drop all households receiving weatherization assistance from the experimental sample. Using the remaining households, we regress the log of monthly energy consumption on an indicator that equals zero before encouragement activities were initiated and one after they began in March 2011 for all households assigned to the encouraged group. We estimate a precise zero effect which suggests that our encouragement intervention had no effect on energy use in these households and supports the validity of the exclusion restriction. Of course, our

Panel B of Table 4 computes the present value of the estimated energy savings under alternative assumptions about investment time horizons and discount rates. To express the estimates of monthly energy savings in dollar terms, we first impute the average consumption among unweatherized compliers.[17] Estimates of monthly natural gas savings (in percentage terms, based on the coefficient in column (3)) are multiplied by the product of this imputed average monthly gas consumption among control group compliers and the residential retail price of natural gas in Michigan in 2013 (EIA, 2015).[18] Similarly, average 2013 electricity savings (in percentage terms, based on the coefficient in column (4)) are multiplied by the product of average monthly electricity consumption among control group compliers and retail electricity price ($0.11/kWh).[19] Taken together, our estimate of average energy savings is approximately $233 per year.

To compute the net present value of energy savings over the useful life of the improvements, we invoke some additional assumptions. First, we rely on the NEAT simulation program's assumptions about measure-specific lifespans. These projected lifespans range from 3 years (for a furnace tune-up) to 20 years (for attic insulation). The energy savings-weighted average lifespan for installed measures in our dataset is 16 years. In the table, we report discounted benefits with assumed lifespans of 10, 16, and 20 years. We also assume that the effect of weatherization on energy consumption - and real energy prices - do not vary over the life of the measure. Discounted benefits are calculated at discount rates of 3%, 6%, and 10% (see columns).

The estimated energy savings are small relative to the projected savings and the upfront costs.

exclusion restriction also implies that the encouragement intervention does not directly effect energy consumption among households taking up the weatherization treatment but we cannot test this assumption directly.

[17]In Table 3 we estimate the share of always takers and compliers in the population to be 1% and 5%, respectively. We directly observe the average energy consumption among unweatherized households in the encouraged group (i.e., the never takers) during the post-assignment period. We also observe the average consumption among unweatherized households during the post-assignment period in the control group. As this represents a weighted average of consumption across never takers and compliers, we can impute the average post-assignment consumption among unweatherized compliers. This is the imputed counterfactual consumption reported in Table 4.

[18]The average retail price of natural gas in Michigan in 2013 was $10.46/MMbtu (expressed in $2013). This is higher than the average price charged by this utility over the entire treatment period ($7.98/MMBtu). Natural gas prices were at historic lows over this period, so using the observed average price would likely underestimate the real prices that will prevail over the life of these investments. Prices in 2013 are somewhat lower than the average real prices over the period 2000-2013. The shale gas boom has arguably ushered in a new domestic price regime, such that a longer average real price will overestimate future prices.

[19]The NEAT program audits assume an electricity price of $0.11/kWh and a natural gas price of $11.46/MMBtu. The higher gas price is presumably based on 2006 prices which averaged around $11.50/MMBtu in this service territory in 2006.

Our estimates imply energy savings of 17 MMBtu per year, whereas the average projected savings among compliers is 56 MMBtu (aggregating electricity and gas measured in terms of site energy content). Our central estimate of the realized average savings per household is roughly $2,349. This is significantly lower than the average net present value of savings predicted by the ex ante engineering analysis which is $9,810 among compliers (and $10,611 averaged across all weatherized households in our data).[20] The nine estimates of the present value of the savings range from approximately $1,428 (high discount rate and short time horizon) to about $3,500 (low discount rate and long time horizon). These estimates are just 31% to 76% of the average upfront cost of the energy efficiency measures among compliers.[21]

5.2 Quasi-Experimental Estimates of Energy Savings

Table 5 presents the quasi-experimental estimates based on the estimation of equation (1). The dependent variable in all regressions is the log of monthly energy consumption (i.e., the sum of electricity and natural gas both measured in MMBtu). The first two columns use data from all weatherization applicants, with the second specification allowing time period effects to vary across counties. Columns (3) and (4) trim the sample to obtain estimates that are more directly comparable to the experimental estimates. Specifically, this sample is limited to the implementing agencies that participated in the experiment and to applicants that applied after the encouragement intervention was initiated. Columns (5) and (6) report estimates comparable to columns (1) and (2) reweighted by the propensity score.[22]

The first row in Table 5 reports the estimated average treatment effect. Across the columns, the estimates suggest that WAP participation reduces energy consumption by roughly 8-10%.

[20]The average present discounted savings among weatherized households in the control group was projected to be $10,989. The average projected discounted savings among weatherized households in the encouraged group is $9,982. We estimate that compliers account for 85% of weatherized households in the encouraged group. We thus impute that the average projected savings among complier households is $9,810.

[21]To compute these percentages, we use an average cost of $4,585 among compliers. To construct this average, we use our estimate that 85% of weatherized households in the encouraged group are compliers. We observe the average cost among always takers in the control group to be $5,363. We interpret the average cost among weatherized households in the encouraged group, $4,698, to be a weighted average across compliers and always takers. Note that the range of the percentages could be different if we had household-level estimates of savings and treatment costs.

[22]Because some of the covariates included in the propensity score estimating equation (e.g., reported disability and number of children) are not reported by all households, this sample is somewhat smaller.

Figure 3 provides another perspective on the results in Table 5. It reports on the estimation of a version of the column (1) specification, except the weatherization indicator is interacted with indicator variables for each of the potential quarters before and after weatherization was completed (the time zero effect captures energy consumption in the month of weatherization). We also interact the quarter indicators with heating degree days (HDD) and HDD squared, to account for the fact that weatherization retrofits may lead to higher energy savings during colder quarters.[23] The figure then plots the coefficients on the quarter dummy variables plus the coefficients on the HDD and HDD squared variables multiplied by the average HDD and HDD squared for the corresponding month. We also plot 95% confidence intervals associated with these coefficient sums. The decline in energy consumption is apparent and seems roughly constant throughout the period of our study.

Panel B of Table 5 computes the present value of the estimated energy savings using essentially the same approach described in section 5.1.2. To estimate monetary savings, we need to disaggregate our energy savings estimates by fuel. We estimate equation (1) separately for gas and electricity, respectively (see Appendix 4). Again, the energy savings are small relative to the upfront costs. The nine estimates of the present value of the savings range from approximately $1,000 (high discount rate and short time horizon) to about $2,400 (low discount rate and long time horizon). These estimates are between 22% and 53% of the average upfront cost of the energy efficiency retrofits.

The experimentally estimated local average treatment effect of weatherization on energy consumption is approximately twice as large as the quasi-experimental estimate of the average effect among all households receiving weatherization assistance. Appendix 6 documents that there are significant observable differences between the group of complier households in the experimental sample and the weatherized households in the quasi-experimental sample, but they only explain about 30% of the difference between the experimental and quasi-experimental treatment effect estimates.

"Essential heterogeneity" (Heckman et al., 2006), where households sort into participation based

[23]Heating degree days reflect the difference between the outdoor temperature and a base temperature of 65° F. A day's HDD equal zero for days when the hourly temperature exceeds 65° F for all hours of the day. It is equal to a weighted difference between 65° F and hourly temperatures when the temperature dips below the base. Weights are determined by the share of hours at each temperature. Quarterly HDDs sum daily HDDs over all days in the quarter.

on their gains, is a potential explanation but the data fail to support standard sorting stories because the experimentally estimated LATE exceeds the estimated ATET from the quasi-experimental where household choose to participate in WAP without our encouragement. Measurement error is perhaps a more likely explanation; specifically, the results are consistent with the possibility that the quasi-experimental estimates are attenuated due to measurement error and that the experimentally derived instrumental variable removes this bias. Indeed, the finding that the OLS estimate in column (1) of Table 4 is comparable to the quasi-experimental estimator is striking and consistent with measurement error playing an important role here.[24] Finally, we are using different (albeit overlapping) data sets to implement the experimental and quasi-experimental estimation, so the differences in the estimated parameters could be due to sampling variation.

5.3 Household Reoptimizing Behavior, Building Thermal Properties, and the Welfare Implications of Rebound

This section implements the utility-maximization framework outlined in Section 2 to bound the average willingness to pay for any efficiency-induced increase in energy services.

5.3.1 Does Weatherization Lead to Temperature "Take Back"?

We first test for an effect of weatherization on household demand for space heating. Table 6 summarizes the results from our survey of indoor air temperatures and thermostat set points collected during the winter of 2013 at a subset of our quasi-experimental households. All weatherized households surveyed had received efficiency improvements at least one year before the survey was administered, allowing plenty of time for residents to observe the extent to which the weatherization retrofit affected winter heating costs. A total of 899 households allowed our survey team to enter their home to record the thermostat set point, and 688 allowed our surveyors to linger long enough to record the actual indoor temperature. Approximately half of the households for which we have thermostat or temperature data had received weatherization assistance (453/899 and 349/688).

[24]Measurement error could exist, for instance, if there are typos in the CAAs' paperwork that cause some household to be incorrectly labeled as program participants. It is noteworthy that the results are robust to requiring more and less documentation from the CAAs to define a household as "weatherized."

The remaining households in the control group had applied for- but not received- weatherization assistance. Anticipating some measurement error, we used two different devices to measure indoor temperatures at each home.

Table 6 reports results from regressing thermometer readings and thermostat set points on a binary variable indicating whether the household had been weatherized.[25] Columns (1) and (3) report results from our base specification. Survey respondents comprise a non-random subset of our sample as only a fraction of the targeted household were home and/or willing to open the door to receive our surveyors. Further only 41% of households who opened the door to our surveyors allowed us to come in and collect temperature measurements. We observe differences between survey respondents and the larger sample along potentially important dimensions (such as past energy use and demographics). We estimate specifications where survey household observations are weighted to match the covariate means in the larger quasi-experimental sample of weatherized households.[26] These results are reported in columns (2) and (4). All specifications control for the outdoor temperature on the day of the survey, measured by heating degree days (HDD).

Columns (1) and (2) show that, after controlling for outdoor temperatures, indoor temperatures may be slightly warmer at weatherized households. The point estimate suggests an increase of 0.65 degrees. This effect is imprecisely estimated; we fail to reject the null of zero. The coefficient associated with the HDD variable is statistically significantly negative in the temperature specifications, suggesting that indoor temperatures measure lower on colder days. Although we asked surveyors to wait several minutes before recording temperatures, this finding suggests that cold air brought with the surveyor could be affecting the measurements. Because the standard errors of our estimates do not allow us to reject a small increase in indoor temperatures, we will estimate an upper bound on possible welfare gains using an estimated increase in temperature of 0.65 degrees

[25]If surveyors recorded temperatures from both devices, the temperature specifications include two measurements per household. Standard errors are clustered at the household level.

[26]For both the thermostat set point and indoor temperature samples, we developed pairwise comparisons of observable household and dwelling characteristics across weatherized and unweatherized surveyed households and the larger quasi-experimental sample. The surveyed sub-sample is observationally similar among most - but not all - dimensions. For example, for both dependent variables, survey respondents are significantly more likely to report having children or elderly family members, less likely to be unemployed, and are more likely to use gas as their primary source of heat, as compared to the larger quasi-experimental sample. Among unweatherized households, surveyed households tend to be larger.

based on the point estimate in Column (2).[27]

Columns (3) and (4) suggest that weatherized households set their thermostats lower, on average, by approximately 0.6 degrees F. This is inconsistent with a rebound in demand for indoor heat. A decrease in thermostat settings may seem inconsistent with the suggestive evidence of an increase in indoor air temperatures. One possible explanation is that retrofits, by reducing cold air infiltration, allow households to maintain the same (or slightly higher) levels of indoor comfort at lower thermostat set points. It is also potentially relevant that the study period covers the aftermath of the Great Recession when the marginal utility of income was elevated.

5.3.2 Building Energy Performance

To estimate how a given increase in demand for indoor space heating translates to an increase in monthly gas consumption, we need to estimate the thermal properties of a representative building. We use the so-called "degree day method" to model the energy required to increase temperatures in an average home in our data (Thorpe, 2013). The literature that analyzes energy use in buildings commonly assumes a linear relationship between energy consumption and the difference between outdoor temperatures and a temperature that people find comfortable indoors (e.g., Friedman, 1987; Dyson et al., 2014). We find empirical support for this assumption in our setting.

In residential buildings, it is standard to summarize the technical relationship between energy consumption and heating demand by regressing energy consumption on HDDs. Specifically, we estimate the following equation:

$$C_{imt} = \alpha_i + \beta_1 \mathbf{1}\{WAP\}_{imt} + \beta_2 HDD_{mt} + \beta_3 HDD_{mt} * 1\{WAP\}_{imt} \quad (3)$$

$$+\beta_4 HDD^2_{mt} + \beta_5 HDD^2_{mt} * 1\{WAP\}_{imt},$$

where C_{imt} measures the natural gas consumption at household i in month m and year t. We estimate this equation using data collected from program applicants during winter months

[27]The positive coefficient on the weatherization dummy in columns (1) and (2) appears sensitive to outliers in our data, which may reflect coding errors by our survey team. For example, when we trim the sample by excluding the top and bottom 5% of the observations, the coefficient on a specification equivalent to column (2) is 0.01 (standard error = 0.27). The results on the thermostat set point are not sensitive to data trimming.

(September-March) over the entire sample period. Panel data allow us to include household-level fixed effects in this regression. We include both a linear and a quadratic HDD term, allowing each coefficient to vary with weatherization status, as well as a separate intercept for WAP participants.

All of the estimated coefficients in equation (3) are highly statistically significant and very precisely estimated. The R-squared value is 0.69. The estimated slope of the relationship is less steep among weatherized homes. That is, weatherizations effectively reduce the marginal cost of indoor space heating during the winter. Moreover, it is noteworthy that this relationship is approximately linear though we estimated quadratic terms. Appendix Figure A2 summarizes the estimated relationship, which is analogous to the bottom panel of Figure 1 (flipped into the top quadrant).

We can use this estimated relationship between HDD and monthly energy consumption to estimate the effect of increased demand for indoor temperatures on energy consumption and associated expenditures. This approach assumes that a household's choice of the indoor temperature is independent of outdoor temperatures, thus outdoor temperatures are a valid proxy for the desired level of heating services.[28] The product of the slope of the relationship between natural gas consumption and HDD (estimated from equation 3)and the average natural gas price in the post-encouragement period provides a measure of the the marginal cost of gas heating. Among unweatherized households, this product is equal to approximately $0.072 per heating degree day (or $2.17 per heating degree month). The analogous calculation for weatherized households lead to an estimated marginal cost of $0.056 per heating degree day (or $1.67 per heating degree month). This implies that weatherization led to a reduction in the marginal cost of approximately 20%.[29]

[28] Note that this assumption is consistent with the results in Table 6 which fail to reject that there is no relationship between HDD and thermostat set point.

[29] These estimates of incremental heating costs are comparable to a "rule of thumb" popularized by the American Council for an Energy-Efficient Economy. This rule states that a household will pay approximately 3% on their gas bill for a degree increase in winter thermostat settings (see, for example, http://www.improvement.com/a/5-easy-ways-to-lower-your-gas-bill-during-the-winter). Average natural gas bills during winter months are $85.95 and $57.96 at unweatherized and weatherized homes, respectively.

5.3.3 Bounding the Average Valuation of Increased Indoor Heat

Estimates of the marginal costs of heating among weatherized and non-weatherized households can be combined with the estimated 0.65 degree increase in indoor temperatures (see Table 6, column (2)) to bound the average welfare gain from weatherized households' reoptimization.[30] The lower bound of households' valuation of the higher temperatures is given by 0.65°F * $1.67/degree-month or $1.09 per winter month. At this lower bound, the utility gains from increased warmth are exactly offset by the increase in the energy expenditures incurred to achieve the temperature increase, implying a zero gain in welfare. To define the upper bound, we note that by revealed preference, unweatherized households chose not to pay 0.65°F * $2.17/degree-month or $1.41 per winter month to achieve this incremental increase in temperature. It follows that average marginal benefits from this temperature increase cannot exceed $1.41 per winter month. Since increasing indoor temperatures by 0.65 °F costs the average weatherized household $1.09 per winter month, the average net gain from the weatherization-induced increase in warmth does not exceed $0.32 per winter month. Assuming 6 full winter months in Michigan, this implies an upper bound on the annual welfare gain of roughly $1.92 from higher indoor temperature. In sum, this bounding exercise suggests that the welfare gains from any efficiency-induced rebound in heating demand are very small, likely less than 1% of the energy expenditure savings. Put another way, the efficiency induced rebound in demand for indoor heating appears to be inconsequential in this setting. Our methodology also allows us to calculate the potential welfare gains from much larger increases in indoor temperatures. Note that even a 10 degree F increase in indoor temperatures would lead to a welfare gain of less than $30 per year.

6 Interpretation

6.1 Returns on Residential Energy Efficiency Investments

Panel A of Table 7 evaluates the internal rate of return (IRR) on energy-efficiency investment from a private perspective. More precisely, we report the discount rate at which the discounted value of

[30]Because the relationship between heating services and energy consumption is approximately linear, the average temperature increase yields the average willingness to pay for heating services.

average avoided energy expenditures exactly equals the average upfront investment. While these investments were free to participating households, it is nevertheless informative to estimate the private returns if households had been responsible for the upfront costs because most households that consider the exact same investments do not qualify for WAP. That is, in the broader population, the IRR is a critical factor in determining take-up of energy efficiency investments.

Column (1) computes the internal rate of return using the average upfront investment costs[31] and the average reductions in annual energy expenditures projected by the WAP program audit. These savings are valued at the average retail residential natural gas and electricity prices in 2013 as described above. Over a range of time horizons, the estimated IRR is quite high, as they should be given the investment rule that required projected savings exceed costs. The rate of return associated with the savings-weighted average lifespan (i.e., 16 years) is approximately 12%. By this measure, efficiency investments supported under WAP appear to be very attractive investments that greatly exceed typical returns available in equity, real estate, and bond markets.

The second column of Table 7 replaces the projected savings with an estimate of the actual energy savings (in monetary terms), derived from the randomized encouragement design, plus a generous estimate of the welfare gain from higher indoor temperatures (i.e., the estimated upper bound of the monetized value of the net welfare gain).[32] The estimate of realized savings is about \$232, which is only 36% (=\$232/\$653) of the average projected annual monetary savings.[33] When the upper bound of the monetized value of the higher indoor temperatures is added to the energy savings, the annual benefits are approximately \$234. Using this measure of annual benefits, the IRR is -10.6% for the 10 year horizon, -2.3% for 16 years and 0.2% for 20 years. Negative returns suggests that, at least for residential home retrofits, there may not be much of an efficiency gap to explain. Investments with these returns are infrequently taken-up in the broader economy.

Panel B conducts a similar exercise but adds the value of avoided emissions to the benefit side

[31]The upfront investment costs are calculated as the imputed average installation, construction, and materials costs among compliers. The average cost per household is \$4585.

[32]As a basis of comparison, the lower bound is zero. However, the decision to use the upper versus lower bound estimate has little impact on these calculations because the estimated rebound effect is so small.

[33]The experimental estimate of average energy savings (measured in MMBtu) is 31% (17 MMBtu/56 MMBtu) of projected energy savings, lower than our estimate of the ratio of realized to projected annual monetary savings. This is because electricity is more expensive than natural gas on a per MMBtu basis, and electricity, for which realized savings are closer to projected savings, thus comprises a relatively larger share of the monetary savings.

of the ledger. Avoided emissions of CO_2 are valued at \$38 per ton of (Greenstone et al., 2013). Nitrogen oxide and sulfur dioxide emissions from residential gas consumption are valued at \$250 per ton and \$970 per ton, respectively (Muller and Mendelsohn, 2009).[34] The IRRs for the 10, 16, and 20 year horizons are -7.6%, 0.1%, and 2.3%, respectively.

The societal perspective is an especially important one to judge these investments, because a broad range of policies encourage residential energy efficiency investments. Panel C reports estimates of the social internal rate of return that are calculated using estimates of the avoided marginal costs, rather than the retail prices, to monetize the energy savings. These avoided costs are then added to the monetized value of the avoided emissions (as in Panel B). The reason for this adjustment is that retail electricity and natural gas prices used in Panels A and B include a fixed cost component; energy providers recover a significant fraction of the fixed costs of supplying energy (e.g., transmission and distribution investments) in their volumetric rates (Davis and Muehlegger, 2010). These fixed costs are not avoided when efficiency investments reduce residential natural gas and electricity consumption. Most electricity and natural gas distribution utilities are subject to cost-plus regulation, and, in these cases, the fixed costs recovery will be shifted to other customers, meaning that households that reduce energy consumption receive a transfer from other households by reducing their contribution to covering fixed costs. We use the average of the 2013 spot prices (\$3.73/MMBtu) set at the Henry Hub distribution point, a standard reference price for natural gas in North America, to reflect the true marginal cost of natural gas over the lifetime of the measures.[35] To value reductions in electricity consumption, we use the average wholesale price in the midwest electricity market and assume transmission and distribution losses of 5%.

These adjustments are reported in Panel C and lead to a meaningful decrease in the IRR, relative to the IRRs in Panels A and B. For example, the social IRR for the 16 year time horizon is

[34] We assume that burning natural gas emits 117 lbs CO_2 per mmbtu, 0.092 lbs NOx per mmbtu, and 0.000584 lbs SO2 per mmbtu. We assume a marginal operating emissions rate of 1.87 lbs CO_2 per kWh in the Midwest power sector (Callaway et al., 2015). Sulfur dioxide and nitrogen oxide emissions from residential natural gas consumption are monetized using the median marginal damage estimates in (Muller and Mendelsohn, 2009). NOx and SO2 emissions from electricity generation are subject to a (barely) binding cap. To convert energy savings measured at the site to emissions avoided at the power plant site, we assume a 5% transmission and distribution line loss rate, based on EIA's estimated nationwide loss rates of in 2013 (http://www.eia.gov/tools/faqs).

[35] This is lower than the average gas recovery charge reported by the utility in regulatory proceedings over the post-weatherization period (\$5.54), but the utility's recovery charge rolls in contract positions.

-7.8%; it is -17.8% and -4.6% for the 10 and 20 year horizons, respectively. Overall, these residential energy efficiency investments have a negative rate of return across all reasonable time horizons.[36]

An alternative method to summarize the return on WAP energy efficiency investments is to estimate the cost per ton of CO_2 avoided. This is calculated as the ratio of the net cost of the investments (i.e., the annual rental cost of the upfront investment less the value of annual energy savings and avoided damages from regional pollutants) and the tons of CO_2 emissions reduced per year. Panel D uses a 3% discount rate to calculate the annual rental cost of capital, while Panel E uses 7%. Both panels report results based on assumptions of 10, 16, and 20 year lifespans for the investments and use wholesale energy prices as above.

As in Panels A through C, the conclusions differ depending on whether one uses *projected* or *realized* energy savings. Using projected energy savings values generated by the NEAT program audits and the 16 year lifespan, the cost per ton of CO_2 avoided is $38 with a 3% discount rate. With these estimates, the energy efficiency investments would break even, from a social perspective, because the abatement cost per ton is approximately equal to the United State Government's official value of the social cost of carbon of $38.

However, the estimates that are based on actual energy savings again tell a different story. When the experimental estimates of actual natural gas and electricity savings are used in column (2), the analogous costs per ton of CO_2 avoided are $201 (3% discount rate) and $285 (7% discount rate). These costs exceed the United States Government's social cost of carbon by a significant margin. On the basis of the costs and benefits we account for here, these residential energy efficiency investments are not a cost effective approach to mitigating climate change.

6.2 What Explains the Low Rate of Return on These Efficiency Investments?

It is natural to ask why the returns to residential energy efficiency investments are so low. After all, WAP contractors only implement measures with projected savings to costs ratios greater than one. Section 5.3 found that the low returns cannot be explained by a rebound effect and this subsection examines other potential explanations with the aim of shedding light on whether these findings are

[36] An alternative approach would be to calculate the social internal rate of return from the government's perspective. This would require accounting for the social cost of public funds and the administrative costs of the program.

externally valid to other settings.

One possibility is that the costs of the investments are inflated in the government-funded WAP program, especially relative to costs in a private market, where most households make energy efficiency investments. Appendix Tables 8 and 9 compare the costs of four common measures in our data with costs incurred by homeowners participating in two residential audit-based energy efficiency programs administered by the Wisconsin Energy Conservation Corporation (WECC),analyzed by Allcott and Greenstone (2017). The WECC program is a standard residential audit-based energy efficiency program, so it seems reasonable to presume that most of its consumers have incomes that exceed the WAP threshold. The clear conclusion from these tables is that, even after adjustment for observables, costs were statistically lower by economically meaningful amount under WAP for three of the four measures and roughly equivalent for the fourth measure. One can speculate on the reasons, but the data fail to support the possibility that the costs used to calculate the rate of return associated with WAP measures are inflated. The analysis is described in greater detail in the online appendix.

Another potential explanation is that contractor quality is low in the WAP program and that this was particularly the case during the Great Recession when WAP greatly increased its scale, making the paper's results non-representative of returns under more normal conditions. We note that ARRA funding faced tremendous scrutiny for mismanagement and that during this period WAP agencies indicated an increased commitment to training, monitoring, and quality assurance. Further, it is possible that the quality of contractors could either increase or decrease during this period because, while WAP was scaling up, the home construction industry was declining dramatically, potentially increasing the supply of skilled contractors.

This is ultimately an empirical question, so the Appendix investigates the possibility that realized savings vary with contractor experience. More precisely, we test for systematic variation in realization rates (i.e., the share of projected savings that manifest as actual savings). The online appendix describes this exercise in detail, including some important data limitations, but the key finding is that we fail to find any relationship between realization rates and our imperfect measure of contractor experience in the WAP program (see Appendix Figure 1). Based on the available data, it appears that contractor heterogeneity cannot explain the low rates of return to energy efficiency investments through the WAP program.

Given that the data reject large rebound effects, and we find no evidence that costs incurred in this government program were abnormally high, we believe that the most likely explanation is that the NEAT efficiency audit tool systematically overstates the real returns to these investments by a significant margin.[37] It is noteworthy that we found in our data that the NEAT program predicts baseline natural gas consumption that exceeds actual consumption by more than 25% prior to weatherization, suggesting that the auditing tool could be under-estimating homes' efficiency properties prior to weatherization. This might help to explain why the benefits of efficiency measures are overstated. [38] Furthermore, recent research (e.g., Allcott and Greenstone (2017), Levinson (2016), and Davis et al. (2014)) has also found low realization rates for energy efficiency measures in other settings. Ultimately, we do not have a definitive explanation for why the engineering models overstate the savings, but we note that this is not an industry where there is a culture of objective ex post evaluation and that contracts are not written with penalties for a failure to deliver promised returns. Regardless of the cause, it seems clear that significant research into improving the functioning of the NEAT efficiency auditing tool (and similar tools) is warranted as it is used by state and local WAP sub-grantees, utility companies, and home energy audit firms.[39]

More broadly, it is possible that the conventional cost-benefit approach to these programs misses important non-energy benefits and that such benefits help explain the popularity of energy efficiency programs and investments. The most oft-cited example is that weatherization leads to improvements in health, presumably due to decreasing thermal stress. The Department of Energy produced an evaluation of the weatherization program, and claim that the retrofits produce substantial health benefits (ORNL, 2015). Among other things, the analysis claims that the typical

[37]Another possibility is that the winters we analyze were warmer than the average winter that was used in the engineering calculations. In fact, the opposite is true in this period: we observe colder than average temperatures and higher than average degree day measures in our sample that should lead to greater than average savings.

[38]Several studies and utility reports have documented how software-based energy analysis of existing homes tends to over-predict pre-retrofit energy use. For example, a recent report found that modeling software consistently overestimated energy consumption; mean modeled total annual use was 40% greater than billed use (SBW, 2012).

[39]While more sophisticated building simulation models exist, they are also very likely more expensive to use. An appeal of NEAT is that it can be inexpensively used by the thousand of implementers who have a wide range of skills and technical training. In fact, the DOE cites NEAT's accessibility to non-technical users as one of its primary benefits (EERE, 2010).

WAP retrofit leads to an astounding nearly $5,000 of mortality reduction benefits. We investigated this claim and found that it is based on replies to the survey question, "In the past 12 months, has anyone in the household needed medical attention because your home was too cold (hot)?". A series of assumptions are invoked to map these responses to fatality rates. The estimates were also derived from a non-standard statistical technique and included no effort to assess statistical significance.[40] Finally, the DOE's own analysis of non-energy benefits fails to account for any health costs associated with the documented increases in radon and formaldehyde levels (Pigg et al. 2014), both of which are known health risks. Based on the available evidence, particularly the small change in indoor temperature that we document (and confirmed by the DOE study), we find little reason to believe that the non-energy benefits are important and some suggestion that they are negative.

Iit is possible that the weatherization retrofits improve the quality of the house in ways beyond energy savings and the non-energy benefits discussed in the previous paragraph. These quality improvements could be reflected in the property markets through higher house valuations. A prime example is new windows, which are easily visible, can improve a house's aesthetics and are frequently undertaken as part of retrofits by non-low-income households. However, the program's requirement that each investment pass a cost-benefit test rules out windows in most cases, because their energy savings generally fall short of their costs. At the same time, this is an immature area and more research is necessary to understand how housing markets capitalize energy efficiency investments.

7 Conclusion

We conducted a large-scale randomized encouragement design experiment on a sample of 30,000 households presumptively eligible for participation in WAP in the state of Michigan. Approximately one quarter of these households were randomly assigned to a treatment group that was encouraged to apply for the program and received significant application assistance. The control households were free to apply for WAP but were not contacted or assisted in any way by our team. We also

[40]We provide a more detailed explanation of the unconventional approach taken to generate what we believe to be unrealistic estimates of non-energy benefits. This summary can be found at https://energyathaas.wordpress.com/2015/10/06/weatherization-assistance-program/.

analyze corroborating evidence from a quasi-experimental analysis covering over twice as many weatherizations as well as a survey of indoor conditions at weatherized and unweatherized homes.

We document three primary findings. First, the aggressive encouragement efforts were disappointing. This encouragement increased take-up rates from less than 1% in the control group to about 6% at a cost of over $1,000 per weatherized household. Second, we find that WAP participation reduced energy consumption by 10-20% among participating households. However, the upfront cost of the energy efficiency investments are about twice the cost of the realized energy savings. Further, the projected savings are more thant 3 times the actual savings. Third, while the modest energy savings might be attributed to the rebound effect, when demand for energy end uses increases as a result of greater efficiency, the paper fails to find evidence of economically or statistically significant increases in indoor temperature at weatherized homes.

Overall, the energy efficiency investments we evaluate are poor performers on average across a variety of metrics. From a household's perspective, the annual internal rate of return that would rationalize these efficiency investments is -2.3%. The household's perspective differs from society's because it fails to recognize the benefits of greenhouse gas and local pollutant emissions reductions and because the retail prices for natural gas and electricity exceed their marginal costs of delivery. Accounting for these two factors, the annual social internal rate of return that would justify these investments is -7.8%, which is even less favorable. Finally, we also calculate the average cost per ton of avoided CO_2 under a range of assumptions. The most plausible estimates exceed $200/ton, which is significantly larger than the U.S. government's estimate of the monetized benefits of avoided emissions (i.e., the social cost of carbon) of roughly $38.

This study demonstrates that the returns to common residential energy efficiency investments are negative both privately and socially among low-income households in Michigan. The results are striking because Michigan's cold winters and the likelihood that the weatherized homes were not in perfect condition suggests that it may have been reasonable to expect high returns in this setting. Regardless of one's priors, this paper underscores that it is critical to develop a body of credible evidence on the true, rather than projected, returns to energy efficiency investments in the residential and other sectors. The findings also suggest that the last several decades may have seen too much investigation into the why of the energy efficiency gap and not enough into whether there really was one.

From a policy perspective, WAP does not appear to pass a conventional cost-benefit test, although its full set of goals may not be reflected in such tests. On the broader question of optimal climate change policy, this paper's findings indicate that residential energy efficiency retrofits are unlikely to provide the least expensive carbon reductions. Future research should examine whether the real world returns to energy efficiency investments differ so starkly from engineering projections in other settings.

References

Allcott, H. and Greenstone, M. (2012). Is there an energy efficiency gap? *The Journal of Economic Perspectives*, 6(1):3–28.

Allcott, H. and Greenstone, M. (2017). Measuring the welfare effects of residential energy efficiency programs. *Mimeograph*.

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Barbose, G. L., Goldman, C. A., Hoffman, I. M., and Billingsley, M. A. (2013). The future of utility customer-funded energy efficiency programs in the United States: projected spending and savings to 2025. *Energy Efficiency Journal*, 6(3):475–493.

Callaway, D., Fowlie, M., and McCormick, G. (2015). Location, location, location: the variable value of renewable energy and demand-side efficiency resources. *Working paper*.

CPUC (2015). 2010 to 2012 energy efficiency annual progress evaluation report. Technical report, California Public Utilities Commission.

Davis, L., Fuchs, A., and Gertler, P. (2014). Cash for coolers: evaluating a large-scale appliance replacement program in Mexico. *American Economic Journal: Economic Policy*, 6(4):207–238.

Davis, L. and Muehlegger, E. (2010). Do Americans consume too little natural gas? An empirical test of marginal cost pricing. *The RAND Journal of Economics*, 41(4):791–810.

Dubin, J. A., Miedema, A. K., and Chandran, R. V. (1986). Price effects of energy-efficient technologies: a study of residential demand for heating and cooling. *The RAND Journal of Economics*, 17(3):310–325.

Dyson, M., Borgeson, S., Michaelangelo, T., and Callaway, D. (2014). Using smart meter data to estimate demand response potential, with application to solar energy integration. *Energy Policy*, 73:607–619.

EERE (2010). Review of selected home energy auditing tools: in support of the development of a national building performance assessment and rating program. Technical report, U.S. Department of Energy, Energy Efficiency & Renewable Energy.

EIA (2015). Natural gas monthly. Technical report, Energy Information Administration.

Fowlie, M., Greenstone, M., and Wolfram, C. (2015). Are the non-monetary costs of energy efficiency investments large? Understanding low take-up of a free energy efficiency program. *American Economic Review Papers and Proceedings*, 105(5):201–04.

Friedman, D. (1987). Cold houses in warm climates and vice versa: A paradox of rational heating. *Journal of Political Economy*, 95(5):1089–97.

Gerarden, T. D., Newell, R. G., and Stavins, R. N. (2015). Assessing the energy-efficiency gap. Technical report, Harvard Environmental Economics Program.

Gillingham, K., Kotchen, M. J., Rapson, D. S., and Wagner, G. (2013). Energy policy: The rebound effect is overplayed. *Nature*, 493(7433):475–476.

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

Greenstone, M., Kopits, E., and Wolverton, A. (2013). Developing a social cost of carbon for US regulatory analysis: a methodology and interpretation. *Review of Environmental Economics and Policy*, 7(1):23–46.

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*, 88(3):389–432.

IEA (2013). Redrawing the energy-climate map. Technical report, International Energy Agency.

Jacobsen, G. D. and Kotchen, M. J. (2013). Are building codes effective at saving energy? Evidence from residential billing data in Florida. *Review of Economics and Statistics*, 95(1):34–49.

Joskow, P. L. and Marron, D. B. (1992). What does a negawatt really cost? Evidence from utility conservation programs. *The Energy Journal*, 13(4):41–74.

Kotchen, M. J. (2017). Longer-run evidence on whether building energy codes reduce residential energy consumption. *Journal of the Association of Environmental and Resource Economists*, 4(1):135 – 153.

Kushler, M., Nowak, S., and Witte, P. (2012). A national survey of state policies and practices for the evaluation of ratepayer-funded energy efficiency programs. Technical Report U122, American Council for an Energy-Efficient Economy.

Levinson, A. (2016). How much energy do building energy codes save? evidence from california houses. *American Economic Review*, 106(10):2867–94.

Loftus, P. J., Cohen, A. M., Long, J. C. S., and Jenkins, J. D. (2015). A critical review of global decarbonization scenarios: what do they tell us about feasibility? *Wiley Interdisciplinary Reviews: Climate Change*, 6(1):93–112.

McKinsey & Company (2009). Unlocking energy efficiency in the U.S. economy. Technical report, McKinsey Global Energy and Materials.

Metcalf, G. E. and Hassett, K. A. (1999). Measuring the energy savings from home improvement investments: evidence from monthly billing data. *The Review of Economics and Statistics*, 81(3):516–528.

Muller, N. and Mendelsohn, R. (2009). Efficient pollution regulation: getting the prices right. *The American Economic Review*, 99(5):1714–1739.

Oak Ridge National Laboratory and United States. Department of Energy. Office of Energy Efficiency and Renewable Energy and United States. Department of Energy. Office of Scientific and Technical Information (2015). *Weatherization Works–Summary of Findings from the Retrospective Evaluation of the U.S. DOE's Weatherization Assistance Program*. United States. Department of Energy. Office of Energy Efficiency and Renewable Energy.

Radnofsky, L. (2010). A stimulus project gets all caulked up. *Wall Street Journal*.

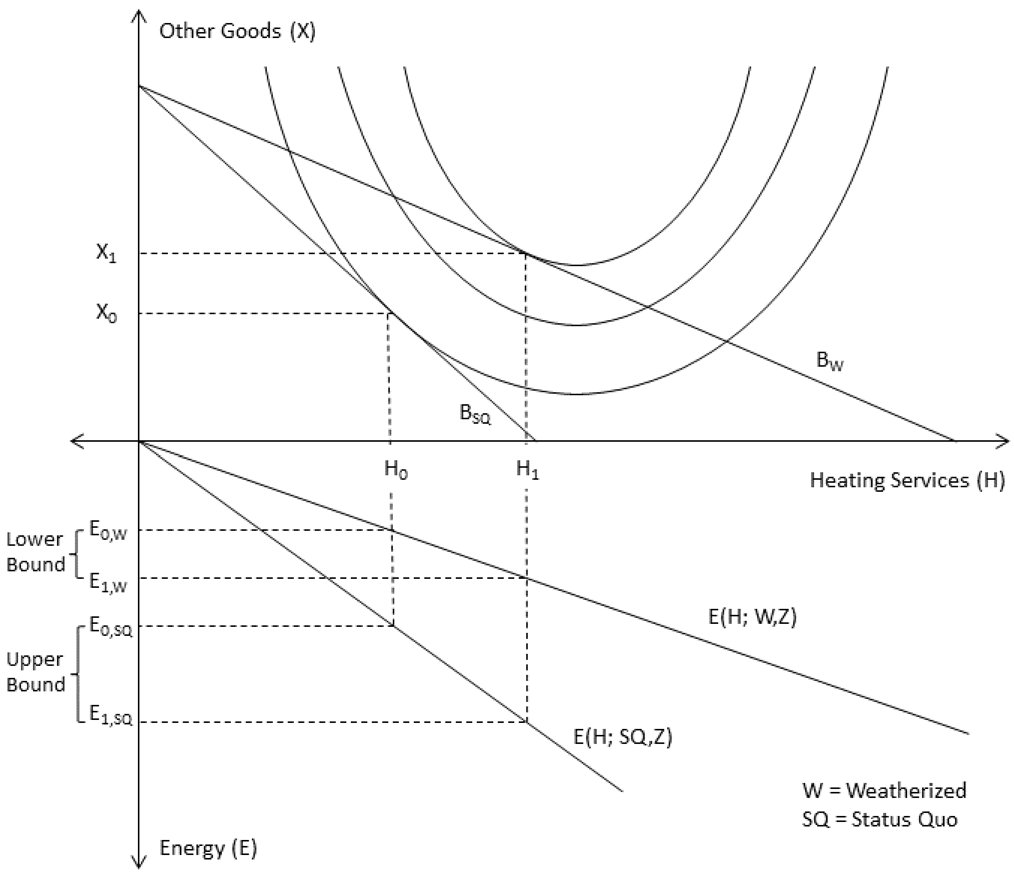
SBW (2012). 2010-2012 PGE and SCE whole house retrofit program process evaluation study. Technical report.

Schwarz, P. M. and Taylor, T. N. (1995). Cold hands, warm hearth? Climate, net takeback, and household comfort. *The Energy Journal*, 16(1):pp. 41–54.

Thorpe, D. (2013). *Energy Management in Buildings: The Earthscan Expert Guide*. Routledge.

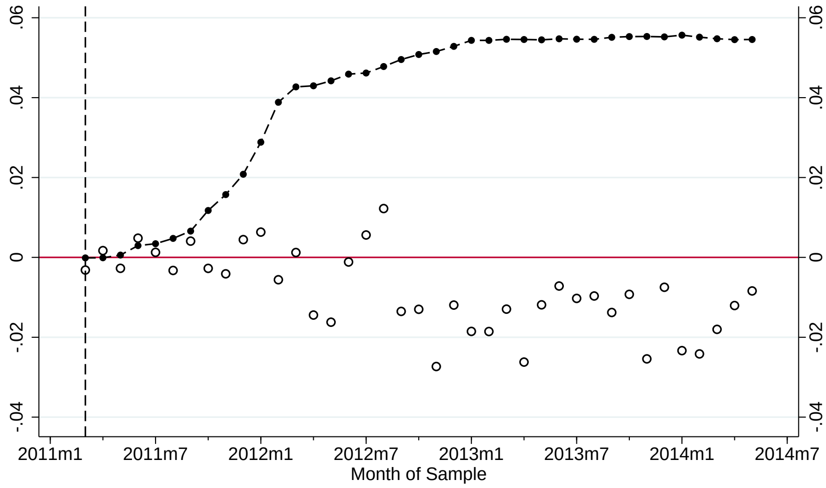
8 Figures

Figure 1: Household-level re-optimization in response to an efficiency improvement



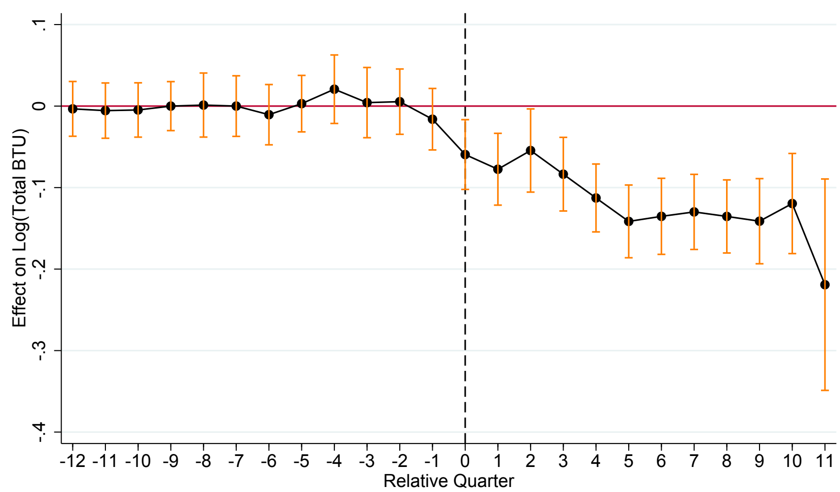
Note: Budget constraints and indifference curves of a representative consumer are plotted in the top quadrant. A linear relationship between heating services and building energy consumption (over a range where heating demand is strictily positive, and heating technology constraints do not bind) are plotted in the bottom quadrant. Please see text for details.

Figure 2: Effect of encouragement on participation and energy consumption



Notes: This figure provides an overview of the local average treatment effect estimates. The broken line tracks the cumulative difference in participation rates across the encouraged and control groups. The circular markers plot the monthly estimates of the intent to treat effects on household energy consumption (in logs).

Figure 3: Event study analysis: Matched quasi-experimental sample



Notes: This figure reports estimated weatherization effects by quarter before and after the weatherization was completed based on the quasi-experimental estimates reported in column (1) of Table 5. The time zero effect captures energy consumption in the month of the weatherization. The effects are also allowed to vary by the realized weather in the quarter. See text for details.

9 Tables

Table 1: Randomized encouragement intervention

Encouragement activity	
Encouraged group (households)	8,648
Initial home visits	6,694
Robo-calls	23,500
Personal calls	9,171
Follow up appointments	2,720
Average cost/encouraged hh	$55.00

Note: The table summarizes efforts to encourage a group of Michigan households to take up weatherization assistance. These households were selected randomly from a sub-population of households who were located in the service territory of our partner utility and presumptively eligible based on ex ante available income information.

Table 2: Differences in sample means between groups of households

	Experimental encouraged (1)	Experimental control (2)	(1) - (2) (3)	All weatherized (4)	Unweatherized applicants (5)	(4) - (5) (6)	(4) - matched applicants (7)
Panel A: Pre-treatment period monthly energy consumption							
Winter gas (MMBtu)	10.40	10.38	0.79	9.88	11.63	0.00**	0.87
Summer gas (MMBtu)	2.84	2.79	0.09	1.80	2.16	0.00**	0.61
Winter electricity (MMBtu)	2.12	2.10	0.27	2.24	2.30	0.15	0.72
Summer electricity (MMBtu)	2.17	2.17	0.84	2.23	2.20	0.37	0.45
Panel B: Census block group demographics and dwelling characteristics							
Mean household income ($)	42530	42870	0.63				
Percent of people below poverty line	27.81	27.12	0.39				
Mean household size (owner-occupied)	2.53	2.53	0.95				
Households with members under 18 (%)	33.64	33.84	0.70				
Households with members over 65 (%)	24.92	25.00	0.89				
Heat with natural gas (% of owner-occupied housing units)	80.49	81.09	0.32				
Median home age (owner-occupied)	52.59	52.94	0.57				
Panel C: Demographics and dwelling characteristics							
Household income				19,617	17,509	0.00**	0.53
Household percent of poverty				115	104	0.00**	0.18
Household size				2.56	2.47	0.13	0.49
Children				0.24	0.15	0.00**	0.76
Disability				0.04	0.03	0.03*	0.91
Elderly				0.23	0.13	0.00**	0.77
Heat with natural gas				0.79	0.57	0.00**	0.73
Age of home				59.15	68.30	0.00**	0.62
Households	7,549	21,339		2,074	2,973		

Note: Columns numbered (1), (2), (4) and (5) report average values. Columns (3), (6), and (7) report the p-values of differences in means. Columns (1) and (2) report sample means for the randomized encouraged and the experimental control groups, respectively. Column (4) reports the sample means for all weatherized households in the quasi-experimental sample while column (5) reports the the sample means for households in the quasi-experimental sample that applied for weatherization but did not receive assistance as of April 2014. Finally, column (7) reports the p-values of the differences in means between all weatherized households in the quasi-experimental sample and propensity-score matched unweatherized households. Household counts summarize energy consumption data. Panel B describes census block-group level data from the 2007-2011 American Community Survey while Panel C focuses exclusively on program applicants. This is because household-level demographic data is not available for the majority of households in the experimental sample that did not apply to the program.

* Significant at the 5 percent level

** Significant at the 1 percent level

Table 3: Randomized encouragement: return on effort

	Application (1)	Efficiency audit (2)	Weatherization complete (3)
Base Rate	0.02**	0.01**	0.01**
	(< 0.01)	(< 0.01)	(< 0.01)
Encouragement	0.13**	0.05**	0.05**
	(< 0.01)	(< 0.01)	(< 0.01)
Households	28,888	28,888	28,888

Note: The table shows the effect of our encouragement on program applications, efficiency audits, and weatherization. Indicators of program participation status are regressed on an encouragement indicator and a constant. The unit of observation is a household.
* Significant at the 5 percent level
** Significant at the 1 percent level

Table 4: Experimental estimates impacts of weatherization on household energy consumption

Panel A: Dependent variable is monthly energy consumption (in logs)

	Total Energy		Gas	Electricity
	(1)	(2)	(3)	(4)
	OLS-FE	IV-FE	IV-FE	IV-FE
WAP	-0.10**	-0.20*	-0.21**	-0.10
	(0.01)	(0.08)	(0.08)	(0.10)
Imputed counterfactual consumption MMbtu/month	7.52		6.39	2.13
F-statistic	.	267.41**	261.06**	266.78**
Households	27,990	27,229	26,054	27,115
Observations	1,662,781	1,653,583	1,528,526	1,638,337

Panel B: Present value of (discounted) savings

Time Horizon		Discount rate	
	3 percent	6 percent	10 percent
10 years	$1,983	$1,711	$1,428
16 years	$2,920	$2,349	$1,819
20 years	$3,459	$2,666	$1,979

Note: Dependent variable measures log of monthly household energy consumption. Panel A reports regression coefficients. With the exception of the first column, all specifications are estimated using 2SLS. Standard errors (in parentheses) are clustered by household. Panel B reports savings projections generated by NEAT audit. All regressions include month-of-sample and household-month fixed effects.

* Significant at the 5 percent level

** Significant at the 1 percent level

Table 5: Quasi-experimental estimated impacts of weatherization on household energy consumption

Panel A: Dependent variable is monthly energy consumption (in logs)

	(1)	(2)	(3)	(4)	(5)	(6)
WAP	-0.08**	-0.09**	-0.08**	-0.09**	-0.10**	-0.10**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
month-of-sample FE	Y	N	Y	N	Y	N
month-of-sample x county FE	N	Y	N	Y	N	Y
P-score matched sample	N	N	N	N	Y	Y
Adjusted R-squared	0.85	0.86	0.83	0.83	0.80	0.81
Households	5,013	5,013	3,334	3,334	3,404	3,404
Observations	282,196	282,196	183,353	183,353	188,287	188,287

Panel B: Present value of (discounted) savings

Time Horizon		Discount rate	
	3 percent	6 percent	10 percent
10 years	$1,393	$1,202	$1,004
16 years	$2,052	$1,651	$1,278
20 years	$2,430	$1,873	$1,391

Note: Panel A reports estimates of the reduction in monthly energy consumption following weatherization (measured in MMBtu). The dependent variable is the log of monthly household energy consumption (electricity and natural gas) measured in MMBtu. All columns include household-by-month-of sample fixed effects. Columns (1) and (2) use data from all weatherization applicants while columns (3) and (4) use a sample limited to implementing agencies that participated in the experiment as well as applicants that applied after the encouragement intervention was initiated. Columns (5) and (6) report estimates comparable to columns (1) and (2) reweighted by the propensity score. Average monthly consumption is 8.48 MMBtu for all applicants and 10.14 MMBtu for the limited sample. The propensity score weighted average is 8.29 MMBtu per month. Standard errors (in parentheses) are clustered at the household level. Panel B reports the net present value of energy savings implied by the preferred estimate reported in column (6). Reductions in energy bills associated with the estimates in column (6) are assumed to accrue over the life of the measure using a range of discount rates and assumed time horizons.

* Significant at the 5 percent level

** Significant at the 1 percent level

Table 6: Indoor temperature survey results

Indoor temperature response to weatherization

	Thermometer		Thermostat	
	(1)	(2)	(3)	(4)
Base temperature	72.36**	72.17**	69.26**	68.91**
	(0.95)	(1.24)	(0.96)	(1.29)
Weatherized home	0.57	0.65	-0.57	-0.56
	(0.41)	(0.44)	(0.29)	(0.33)
Heating Degree Days	-0.16**	-0.15**	0.04	0.05
	(0.03)	(0.04)	(0.03)	(0.04)
Propensity Score Weights?	N	Y	N	Y
R-squared	0.02	0.02	0.01	0.01
Observations	1359	1359	899	899

Note: The table reports measured indoor temperature differentials across weatherized (WAP) and unweatherizedhouseholds. Columns (1) and (2) have the indoor thermometer temperature reading as a dependent variable while columns (3) and (4) use the survey thermostat readings. Columns (2) and (4) are weighted so that surveyed population better represents total quasi-experimental sample. Standard errors clustered at the household level.

* Significant at the 5 percent level

** Significant at the 1 percent level

Table 7: Estimated returns on investments in energy efficiency

Time horizon	Ex ante (NEAT) projections (1)	Empirical estimates (2)
Panel A: Private internal rate of return		
10 years	7.0%	-10.6%
16 years	11.9%	-2.3%
20 years	13.0%	0.21%
Panel B: Private internal rate of return, adding the avoided emissions damages		
10 years	11.4%	-7.6%
16 years	15.4%	0.1%
20 years	16.4%	2.3%
Panel C: Social internal rate of return		
10 years	-3.9%	-17.8%
16 years	3.1%	-7.8%
20 years	5.0%	-4.6%
Panel D: CO_2 **abatement cost - 3 percent discount ($/ton** CO_2**)**		
10 years	$85	$322
16 years	$38	$201
20 years	$22	$161
Panel E: CO_2 **abatement cost - 7 percent discount ($/ton** CO_2**)**		
10 years	$117	$403
16 years	$71	$285
20 years	$56	$248

Note: All calculations use the average retrofit cost of $4,585. This is the imputed average expenditure for compliers, constructed using ex post observed costs per weatherized household as reported by the implementing agencies. Column (1) reflects engineering projections of annual energy savings. In Panels B and C, column (1) also incorporates the value of estimated emissions reductions (valued using a social cost of carbon value of $38 per ton CO_2 and values for avoided local pollutants as described in the text). Column (2) replaces the engineering estimates of energy savings with our experimental estimates of energy savings. Column (2) also incorporates the upper bound on the net welfare gain from increased heating demand using our very small and statistically insignificant point estimate of the upper bound on the efficiency-induced increase in welfare associated with warmer indoor air temperatures. Panels D and E report the implied abatement cost per ton CO_2. These values divide the average levelized investment cost net of fuel savings by the estimated average quantity of emissions avoided (measured in metric tons).