Can Incentive-Based Pay Increase the Marginal Value of Public Spending on Energy Efficiency? Experimental Evidence from the Weatherization Assistance Program

Peter Christensen*, Paul Francisco*, and Erica Myers †

July 18, 2022

Abstract

Aligning compensation with recipient outcomes has the potential to improve the efficiency of government programs. We perform a field experiment to evaluate the impact of performance bonuses on the returns to spending in a large low-income energy efficiency assistance program. The performance bonuses increase natural gas savings by 24%, increase the social net benefits of incentivized retrofits from 0.72 to 1.2, and increase returns from total program spending by 10-20%. We find evidence of outsized impacts among higher quality contractors but do not find evidence of deficiencies in non-incentivized tasks when performance incentives are combined with reasonably well-enforced quality control inspections.

---

*University of Illinois at Urbana-Champaign; †University of Calgary; Emails: Christensen (pchrist@illinois.edu), Francisco (pwf@illinois.edu), Myers (erica.myers@ucalgary.ca). We are indebted Mick Prince, Chad Wolfe, the PRISM Climate Group for assistance with data and experimental implementation. We thank Bryan Parthum, Tengjio Chen, and Hansen Shao for excellent research assistance. We received excellent feedback and comments from Meredith Fowlie, Rob Metcalfe, and seminar participants at the CU Environmental and Resource Economics Workshop, UC Berkeley's Electricity Camp, Georgia State University, Michigan State University, and the Online Agricultural and Resource Economics Seminar (OARES). We acknowledge generous support from the Alfred P. Sloan foundation and the Illinois Department of Commerce and Economic Opportunity's Illinois Home Weatherization Assistance Program.

1 Introduction

When effort is unobservable, contracts that more closely align workers' pay with employers targeted outcomes have the potential to improve effort and productivity relative to fixed pay structures (e.g., Lazear, 2000; Shearer, 2004). Thus, piece rate and other incentive contracts are widely used in many private sector applications including agriculture, retail, manufacturing, and logistics, among others (Gittleman and Pierce, 2013). However, incentive contracts are less common and potentially under-utilized in the public sector (Burgess et al., 2017; Bloom and Van Reenen, 2011). Indeed, a growing body of empirical evidence has demonstrated opportunities for performance pay to increase the effort and effectiveness of public sector workers such as teachers, tax collectors, and civil servants.[1] However, it has been difficult to evaluate the social net benefits of performance pay programs, making it difficult to evaluate their ultimate impacts on the efficiency of public spending in large government programs.

The present study addresses this gap in the literature with findings from a 2-year field experiment that examines the impact of pay-for-performance incentives in the Illinois implementation of the Weatherization Assistance Program (WAP), the largest and most ambitious energy efficiency program in the United States. If done cost-effectively, building energy retrofits can be one of the lowest cost climate mitigation strategies.[2] However, energy efficiency program savings frequently fall short of expectations, making them less cost-effective than anticipated.[3] Research into underlying mechanisms has identified contractor performance as a primary contributor and has revealed that moral hazard and incentive problems affect the quality of energy efficiency retrofits done by contractors (Christensen et al., 2020; Giraudet, Houde, and Maher, 2018; Blonz, 2018). Aligning contractor incentives with social benefits has the potential to increase the marginal value

---

[1] For teachers, see (e.g., Biasi, 2021; Lavy, 2020; Lavy, 2009; Aucejo, Romano, and Taylor, 2019), for tax collectors see (Khan, Khwaja, and Olken, 2019; Khan, Khwaja, and Olken, 2016), and for civil servants see (Bandiera et al., 2021; Bertrand et al., 2020; Kim, Kim, and Kim, 2020; Burgess et al., 2017).

[2] Reductions of greenhouse gases can effectively be achieved at negative cost if the benefits from the stream of energy savings, which reduce emissions, outweigh the upfront costs.

[3] Realized savings have been shown to be lower than projection across settings including home retrofit programs (Fowlie, Greenstone, and Wolfram, 2018; Allcott and Greenstone, 2017; Zivin and Novan, 2016; Berry and Gettings, 1998; Dalhoff, 1997; Sharp, 1994), appliance rebate programs (Houde and Aldy, 2014; Davis, Fuchs, and Gertler, 2014), and efficient new construction (Levinson, 2016; Bruegge, Deryugina, and Myers, 2019; Davis, Martinez, and Taboada, 2020).

of spending in energy efficiency initiatives, which is of urgent concern given that these programs are a central component of global climate policy with hundreds of billions of dollars invested annually (European Parliament, 2012; EEA, 2018; ARB, 2017; IEA, 2020).[4]

We partnered with the Illinois Home Weatherization Assistance Program (IHWAP) to study the effects of a piece rate incentive for air sealing retrofits, which are one of the "big four" energy-saving retrofits in the program, accounting for just over 12% of total expenditures in the average home.[5] Air sealing retrofits provide a promising starting point for incentive-based pay in the program because they are straightforward to evaluate using a blower door test that measures the "leakiness" of a home. The blower boor test is routinely conducted as part of a pre-weatherization home energy audit and again upon completion of the work contract. As illustrated in Figure 1, the CFM50 levels measured by a blower door test are highly correlated with energy consumption.[6] Air sealing is considered to be cost-effective in all homes, but opportunities for it can vary widely from house to house, creating potential for the attention and skill of the contractor to have an impact on identifying and properly sealing the leaks. Heterogeneity in the housing stock and contractor skills interact to produce substantial variation in the the marginal cost of additional air sealing improvements.

Our intervention randomized contracts into one of three pay-for-performance treatments. Treatment consisted of two different bonus regimes for reductions beyond the minimum air sealing standard: a "high" payment ($1.00) and a "low" payment ($.40) for each unit of air sealing achieved beyond target, as measured by CFM50 reductions in pre/post blower door tests on all homes. A third group of control jobs were not be eligible

---

[4] In the United States, The Biden Administration's American Jobs Plan proposes investments of $213 billion to "produce, preserve, and retrofit more than two million affordable and sustainable places to live" (The White House, 2021). Investment in the WAP has expanded dramatically as part of federal climate and stimulus initiatives over the past two decades. Under ARRA, funding was temporarily increased from $450 million annually to almost $5 billion for program years 2011-2012 and, more recently, the Bipartisan Infrastructure Investment and Jobs Act has set aside a historic $3.5 billion to expand the program.

[5] Wall insulation, attic insulation and furnace replacements represent the remaining 3 major retrofits types. The term "air sealing" describes the methodical identification and sealing of air leakage sites, including those in the attic, walls, basement, and/or crawlspace.

[6] The blower door test assesses the volumetric flow rate of air from the home when the home is depressurized by a set amount (50 Pascals [Pa]). The leakier the home, the more air needs to be moved to cause that amount of depressurization. The unit of measurement from the blower door test is cubic feet per minute at 50 Pa (CFM50).

for bonus payments. All other components of the program continued as before, including the enforcement of minimum quality standards on all retrofits. A certified agency quality control inspector conducts an inspection after the work has been completed and can require contractors to rectify any deficiencies. Whereas low-quality workmanship can be partially addressed through the enforcement of minimum quality standards, piece rate bonuses have the potential to allocate contractor effort more efficiently. Quality control inspections were maintained to guarantee a minimum standard of quality in all treatment groups and to prevent the re-allocation of effort from non-compensated to compensated retrofit tasks in treated contracts.

One concern with introducing incentives for one program dimension is contractors may re-allocate effort away from non-compensated to compensated retrofit tasks, even with quality control inspections. We designed the study to assess two types of effort reallocation. First, we collected inspection data to determine whether the incentives induce contractors to divert attention away from non-compensated tasks *within* treatment jobs, as evidenced by an increase in deficiencies in those tasks. Second, our job-level randomization allows us to estimate the effects of effort reallocation *between* jobs. Our intervention creates random variation in the number of simultaneous treatment or control jobs that firms work on at any given time, allowing us to quantify the effects of effort reallocation on the reported deficiencies and energy outcomes in contemporaneous jobs.

We document three key experimental findings on the effects of piece rate air sealing bonus payments for our sample from the IHWAP. Performance incentives lead to: (1) increases in the air-tightness of homes treated by the IHWAP, (2) reductions in the likelihood that contractors are called back by inspectors to fix deficiencies in air sealing retrofits, and (3) additional reductions in household energy use. The magnitudes are substantial – increasing the mean effect of weatherization by an additional 24% and resulting in \$2.79-3.99 in social benefits for every dollar invested in bonus payments. This contrasts with \$0.50-0.72 in social benefits for air sealing retrofits in baseline WAP contracts. We find that contractor skill plays an important role – higher-quality contractors identified from the year prior to the intervention have larger impacts on air sealing.

This study makes contributions to multiple literatures in economics. First, we contribute to a growing body of empirical evidence identifying opportunities for incentive pay to improve effort and effectiveness in the public sector (e.g. Biasi, 2021; Lavy, 2020; Lavy, 2009; Khan, Khwaja, and Olken, 2019; Khan, Khwaja, and Olken, 2016; Bandiera et al., 2021; Bertrand et al., 2020; Kim, Kim, and Kim, 2020). We extend the public sector literature on worker incentives to the growing demand for skill in "green jobs" that are focused on the energy transition and the abatement of greenhouse gas emissions. Our results have implications not just for energy efficiency, but other government-sponsored retrofits including building electrification and rooftop solar installation, where workmanship has sizeable performance effects and similar incentive problems are present (Domanski, Henderson, Payne, et al., 2014).

We also contribute to a literature that, beginning with the seminal work of Holmstrom and Milgrom (1991), demonstrates that performance incentives can induce workers to re-allocate effort from non-compensated to compensated tasks in multi-dimensional contracts and may affect output quality (Muralidharan and Sundararaman, 2011; Lavy, 2009; Barlevy and Neal, 2012; Lavy, 2020; Aucejo, Romano, and Taylor, 2019; Gaduh et al., 2021; Hong et al., 2018). In public sector contexts, maintaining quality across all program aspects can be particularly important.[7] Therefore, our intervention is specifically designed to prevent such unintended consequences by introducing pay-for-performance incentives (for CFM50 reductions) to an existing regime where standards regulate the quality of all retrofit tasks on treated and untreated jobs. Interestingly, we do not find any evidence that incentive treatments increase deficiencies in non-compensated retrofits that are part of a treated contract (i.e. furnace repairs, which do not affect CFM50 reductions). Further we do not find evidence that incentive contracts on some jobs negatively affect performance on contemporaneous jobs. The present study therefore provides evidence that more complete hybrid contracts can improve worker outcomes on incentivized dimensions without deleterious impacts on others.

---

[7] Governments contract with social service providers for a wide range of public goods that have a critical bearing on societal outcomes, including child welfare, workforce development, affordable housing and energy security, criminal justice, and education.

While the general effects of performance pay are fairly well-understood, in many contexts it has been difficult to disentangle the effects of performance contracts on productivity sorting (Dohmen and Falk, 2011; Lemieux, MacLeod, and Parent, 2009) versus productivity increases within workers (Friebel et al., 2017; Lazear and Shaw, 2007; Lazear, 2000). In the IHWAP setting, the assignment of jobs to contractors is independent of treatment, allowing us to isolate the productivity effects of performance pay incentives on a constant pool of human capital and reducing the potential for selection-driven inequality in WAP recipient outcomes (Fioretti and Wang, 2019). We find that performance incentives result in improvements in air sealing and energy efficiency outcomes for high- and low-skill contractors. However, consistent with results from canonical models of incentive-based pay (Lazear, 2018) and emerging evidence in other empirical settings (Frederiksen, Hansen, and Manchester, 2022; Franceschelli, Galiani, and Gulmez, 2010), we find that contractors who performed better at baseline (a measure of ability) respond to incentives with disproportionately larger improvements in their air sealing work. This suggests that the longer-run impacts of performance pay could be even larger if higher ability workers increasingly select into an incentive-based WAP program on the basis of their higher expected payoffs.

Finally, our work contributes to the literature on the cost-effectiveness of energy efficiency programs and its role for climate policy. Projected savings typically overestimate the energy savings realized in a wide range of home retrofit, appliance rebate, and new efficient housing programs, many of which rely on contractors for key labor inputs (e.g., Nadel and Keating, 1991; Davis, Martinez, and Taboada, 2018; Allcott and Greenstone, 2017; Zivin and Novan, 2016; Levinson, 2016; Davis, Fuchs, and Gertler, 2014). Economists have begun to advocate that climate policy investments should be redirected to other types of programs due to this evidence of performance shortfalls, while the International Energy Agency estimates that 44% of all global emissions reductions by 2040 could come from improvements in building energy efficiency (IEA, 2018). The present study provides the first empirical evidence of the impact of performance-based pay on the marginal value of public funds in a major residential retrofit program, suggesting that a shift to incentive-based contracts could increase the return on investment in residential energy efficiency in the coming decades. More evidence is needed to assess the comprehensive impacts of incentive-based pay on energy efficiency improvements that lie outside the scope of this study, but the initial results reported here suggest a promising path forward.

2 Setting and Experimental Design

2.1 Institutional Background

The Weatherization Assistance Program (WAP) is the largest U.S. residential weatherization program, having provided weatherization assistance to over 8 million households since it began in 1976. It aims to reduce energy bills for low income households while maintaining health and safety. The Bipartisan Infrastructure Investment and Jobs Act, which was passed in November 2021, set aside \$3.5 billion to significantly expand the program over the following years. Therefore, it is timely to consider ways to improve the efficiency of delivering services in the program.

Our study was carried out in cooperation with Illinois' implementation of the federal program. Implementing agencies throughout the state are responsible for administering the program. There are approximately 33 agencies serving Illinois' 106 counties, with some agencies serving a single county when that county has a larger urban center and some agencies serving up to nine counties in less-populated portions of the state. Which retrofits are performed in each house are determined during a pre-weatherization energy audit. An agency energy auditor collects detailed measurements on the structure of the house, characteristics of mechanical systems, and health and safety information. DOE funding for a home is allocated using these measurements as inputs to an optimization strategy, which is intended to direct program funding to the most cost-effective among the feasible retrofits. Cost-effectiveness is determined by the savings-to-investment ratio (SIR) and retrofits are selected from highest to lowest SIR until either (1) the available funding is exhausted or (2) there are no more remaining retrofits with SIRs of 1.0 or greater.

Prior to 2016, all funding – including non-DOE funding – followed DOE rules to select retrofits. However, starting in some pilot locations in 2016 and then program-wide in 2017, additional funding from non-DOE sources, including LIHEAP (HHS) funds, state funds, and utility funds has been used to do additional retrofits in homes that may not always meet DOE rules. This "braided funding" is often used to do measures that are ineligible for DOE funding, such as replacing old water heaters or old air conditioners or doing health and safety issues that are beyond WAP guidelines. Many times the goal is to use additional funds to do measures that treat the whole house and help the family independent of DOE eligibility and SIR. Therefore measures with SIR<1 are often selected. The extent to which braided funds can be used depends on the household – not all funding sources have the same eligibility requirements and so some families are not eligible for some funds. Once the complete list of retrofits have selected across all funding sources, the administrative software directly converts this list into a work order, which is provided to the contractor who will implement the work.

Air sealing, the focus of our experiment, is performed in all homes. It is always deemed cost-effective given the low cost and long-lived energy-saving effects of finding and sealing leakage sites in attic, walls, basements or crawl spaces. This type of work affects the tightness of the building envelope, which is measured by a blower door test. To implement the test, an experienced energy specialist will temporarily put a large fan into the frame of a home's outside doorway. Then, after calibrating the device, the fan draws air out of the house and reduces the air pressure inside. The test assesses the volumetric flow rate of air from the home when the home is depressurized by a set amount (50 Pascals [Pa]). In a leakier home, more air needs to be moved to cause the same amount of depressurization. The unit of measurement from the blower door test is cubic feet per minute at 50 Pa (CFM50). Lower CFM50 values indicate that the home is better air-sealed, and thus well suited to retain heated or cooled air. The IHWAP program uses a pre-defined formula to estimate the amount of air sealing expected based on a pre-retrofit blower door test performed during the initial energy audit. The leakier the home, the greater the expected air leakage reductions. This formula is used to determine a quantity target for each individual home.

Which contractors serve the program and how much they are paid for each retrofit is determined through a competitive bidding process at the start of each program year. Contractors submit itemized bids for their labor and materials costs for each of the suite of retrofits performed by the program. For example they will submit their costs per cubic foot of wall or attic insulation or furnace installation. For the selected contractors, their bids determine the compensation that they receive for work performed in that program year. Thus, payments are pre-determined but vary job-to-job for each firm according to which measures appear on the work order and may vary across firms for the same work order due to differences in their bids.

A house enters a job queue once it been approved and the auditor has selected the retrofits to be carried out. Once in the queue, most jobs are assigned in sequential order independent of any characteristics about the home, measures assigned, or the contractors themselves.[8] Contractors often receive a bundle of work orders at once.

The WAP enforces quality standards through inspections performed when the work is completed.[9] The agency inspector documents findings and deficiencies for any measures that did not hit targeted performance. Certain findings will warrant a call-back for the contractor to return to the site to correct a deficiency, such as poor wall insulation or a furnace installation issue. Others do not, if for instance, air sealing does not quite hit target or duct work was not ideally performed.[10] Call-backs are costly for firms since they do not receive any additional compensation for having to perform the extra work.

---

[8] There are some instances, especially in smaller agencies, where jobs may be assigned to a particular contractor on the basis of something like equipment availability.

[9] The quality control inspectors receive accredited training and must be certified by the Building Performance Institute (DOE, 2013).

[10] In practice, the state allows contractors to achieve slightly less than the target on any given job. The guidance is to allow a contractor to achieve a CFM50 reduction that is no more than 10% above the "target" on at least 90% of jobs. The 10% allowance by the state acts as a safeguard against imposing excessive penalties against contractors by accounting for the fact that some homes have characteristics that make achieving target especially difficult.

2.2 Experimental Design

In this study, we evaluate the impact of piece rate bonus incentives to contractors for air sealing improvements based on blower door readings. The study was implemented over the course of two program years 2018 and 2019.[11] The sampling frame included all Illinois jobs completed outside of Cook County (which includes Chicago), which represents approximately half of the state program. The agency serving Chicago is the one agency that we are aware of that already compensates their contractors for air sealing based on measured air-tightness outcomes, whereas the rest of the state compensates contractors using pre-set sums that are based on the degree of expected air leakage reductions. As a result, Cook County-based jobs were excluded from the intervention. The sampling frame was further restricted to single-family homes served by one of the 3 utilities that with whom we had a data sharing agreement.

The intervention consisted of two different piece rate bonus regimes: a "high" payment ($1.00) and a "low" payment ($.40) per CFM50 reduction. Bonus payments were made on top of pre-set compensation assigned to a specific project, such that contractors who achieved CFM50 readings below the minimum target required by the state received a bonus paid on the number of additional CFM50 reduction achieved.[12] A third set of control jobs were not eligible for bonus payments, instead receiving nothing beyond the normal pre-set compensation.

At the annual IHWAP meeting before the start of our intervention in the program year 2018, we alerted the implementing agencies that contractors need to sign up as vendors with the University of Illinois in order to receive bonus payments. We also made them aware that information about the bonus payments would appear at the top of the work orders for treatment jobs. As of the beginning of the program year, eligible jobs were randomized into the three treatment groups. As new work orders were initiated in IHWAP's administrative software, they were automatically assigned to the "high"

---

[11] A program year begins in September and ends in May of the following year, such that program year 2018 included jobs from September, 2017-May 2018.

[12] At the start of the experiment, bonuses were paid based on reductions achieved beyond 10% above target, DOE's guidance for the minimum allowable reduction without having to perform a call back. However, the adjustment was made due to budgetary considerations.

payment treatment, "low" payment treatment, or control regimes. Randomization was implemented through custom application that was embedded in the software that generated work orders for the IHWAP program. The magnitude of incentive payments under example scenarios were clearly printed on the work order that each contractor received at the outset of a job:

> *This job is eligible for a bonus of $1 per CFM50 below target. The target for this job is 1200 CFM50 reduction. A reduction of 1400 CFM50 will receive a $200 bonus payment. A reduction of 1600 CFM50 will receive a $400 bonus payment. A reduction of 1800 CFM50 will receive a $600 bonus payment.*

Several program features described above are worth highlighting for interpreting the results our experimental intervention. First, treatment assignment happens at the point at which the work order is printed, i.e. after the pre-weatherization audit is completed. Therefore it cannot affect blower door targets or retrofits performed because the energy auditor has no way of knowing what treatment status will be when inputting initial measurements. Second, payment per measure is fixed for each firm at the start of the program year. As a result, the compensation for non-air sealing components of the job cannot be affected by the treatment. Third, because jobs are assigned through a queue system, it limits the potential for productivity sorting of firms, allowing us to isolate the impact of piece rate incentives on worker output with a constant pool of human capital (Dohmen and Falk, 2011). Fourth, because contractors must perform call backs at their own expense, the existing quality control regime will reduce effort reallocation away from uncompensated tasks within and across jobs. Finally, the use of braided funding during the program years studied here means that many measures were performed simply as a means to offer aid to low income households and were not intended to be cost-effective in terms of energy savings. Under this regime spending per home almost doubled, and the overall expected SIR of the suite of retrofits was much lower compared to previous years (see Appendix Table B1). This has important implications for interpreting cost-effectiveness of the WAP, since it is not possible to disentangle the impacts of DOE versus other funding sources in this period.

3 Model

In what follows, we present a principal-agent model of contractor effort allocation in a multi-task setting that builds off of insights from both Holmstrom and Milgrom (1991) and Lazear (2000). Contractors have heterogenous ability, which determines the effort required for a given level of output. We first consider minimum quality standards and then introduce a piece rate bonus for performance above a certain minimum level for a task. Finally, we consider the impacts of the introduction of a piece rate bonus on one task on effort on other tasks.

3.1 Set Up

The principal has m different tasks for the agent (a contractor), to perform. The contractor chooses levels of effort for the vector of tasks $\boldsymbol{t} = (t_1, ..., t_m)$ at a personal, strictly convex cost, $C(t_1, ...t_m)$. Let a denote ability, where output, $\mu(t, a)$, is an increasing function of both ability and effort for each task (i).

$$\mu_i = f(t_i(a), a) \tag{1}$$

Differentiating equation 1, we can see that higher ability reduces the effort that a contractor requires to accomplish a given level of output – subscripts on f denote derivatives with respect to ability or effort:

$$\frac{\partial t_i}{\partial a} = -\frac{f_a}{f_{t_i}} < 0$$

The effort expended by a contractor on the set of tasks ($\boldsymbol{t}$) generates a vector of information signals, $\boldsymbol{x(t)} = \boldsymbol{\mu(t)} + \boldsymbol{\varepsilon}$. Assume μ is concave in a and t and the error terms (ε) are normally distributed mean vector zero and stochastically independent across tasks.

3.2 Minimum Quality Standards

The principal pays the contractor a fixed wage for each task, w. She would like to maintain a minimum standard of quality in each dimension with a high degree of likelihood. Quality is determined on the basis of the information signals received, $\boldsymbol{x(t)}$. The principal will "call back" contractors to rectify any problems if they do not meet a minimum standard of $\boldsymbol{x^0}$ that she sets.[13] When minimum quality standards are met for all tasks, the agent receives w.

Without loss of generality, assume the callback has a fixed cost $\boldsymbol{\lambda} = \lambda_1, ...\lambda_m$. The probability of failing the minimum quality standard, $Pr(\boldsymbol{x} < \boldsymbol{x^0}) = \boldsymbol{\phi(t, a)}$, is decreasing and convex in effort and ability, such that the expected cost of a callback, $\boldsymbol{k(t)} = E[\boldsymbol{\lambda'\phi(t, a)}]$, is also decreasing and convex in effort $(k_1 < 0; k_{11} > 0)$ and ability.

Assume that the agent is risk neutral.[14] The agent chooses the vector $\boldsymbol{t^0}$ that minimizes total costs of effort and callbacks. The first order conditions equate the marginal cost of effort with the expected cost of a callback for each task i as follows, where subscripts i indicate derivatives with respect to t_i.

$$C_i(t) = -k_i(t, a) \tag{2}$$

For any given set of outputs and wages $\{\boldsymbol{x^0}, \boldsymbol{w}\}$ there is a group of contractors who will accept the job. Let $\pi(\boldsymbol{x^0(\underline{a}), w}) = \pi(\boldsymbol{0, 0})$ denote profit of the minimum ability agent that would accept the job in lieu of not working. All contractors with ability levels higher than $\underline{a}$ earn rents from the program. Those willing to work in the program must not have preferred work alternatives. Let the profit that an agent of ability a can get at the best alternative be given by $\pi(\boldsymbol{\hat{x}(a), \hat{w}(a)})$ with associated wage and output levels $\boldsymbol{\hat{x}(a), \hat{w}(a)}$. Given that higher-ability contractors may benefit most from outside options that demand more but pay more, there may exist an upper cutoff in ability $\bar{a}$ such that $\pi(\boldsymbol{x^0(\bar{a}), w}) = \pi(\boldsymbol{\hat{x}(\bar{a}), \hat{w}(\bar{a})})$, and only those contractors with abilities $[\underline{a}, \bar{a}]$ participate

---

[13] Given that there is uncertainty in the signal, this could be something like within 10% of the targeted CFM reductions or caulking of gaps.

[14] If the agent were risk averse, payments for effort would be higher, reflecting the risk premium, but the qualitative comparative statics of the model would remain the same.

in the program.

3.3 Piece Rate Bonuses

Suppose the principal introduces a piece rate bonus, which pays b_i for each unit of output above a minimum level $\bar{x}_i$, where $\bar{x}_i \geq x_i^0$, such that compensation for firms is as follows.

$$\pi(x,w) = \begin{cases} w - k(a,t^0) - C(t^0,a) & \forall x_i \leq \bar{x}_i, \\ w - k(a,t^*) - C(t^0,a) + b_i(x_i^* - \bar{x}_i) & \forall x_i > \bar{x}_i \end{cases}$$

Contractors choose the maximum of $[\pi(x^0,w), \pi(x^*,w)]$, where x^* is the output associated with the optimal amount of effort t^*, which solves the following first order condition where subscripts i indicate derivatives with respect to t_i.

$$C_i(t,a) + k_i(t,a) = b_i \cdot x_i^*(t,a) \tag{3}$$

With a piece rate, contractors choose t^* such that total private marginal costs are equal to the bonus payment. Totally differentiating 3 with respect to ability and rearranging we can see that high ability contractors will be more responsive to the bonus.

$$\frac{\partial t^*}{\partial a} = \frac{\frac{\partial x_i^*}{\partial a}}{C_{ii} + k_{ii} - x_{ii}^* \cdot b} > 0 \tag{4}$$

The level of optimal effort is increasing in ability. Given that higher ability contractors produce more output at any given level of effort, the change in output of higher ability contractors in response to the bonus will increase more than proportional to their increase in effort. The presence of the piece rate bonus makes the program more attractive relative to outside options, which may in turn draw in higher ability workers (i.e. increase the level of $\bar{a}$).

3.4 Effort Reallocation

Now consider the effect of introducing a piece rate bonus on one task (i) on effort in another (j), while all tasks remain subject to minimum quality standards. Importantly, with minimum quality standards in place, even with some reallocation, effort will still be exerted on non-bonus tasks. We can quantify reallocation by totally differentiating the first order condition in (2) for task j with respect to b_i and solving for, $\frac{\partial t_j}{\partial b_i}$, as follows.

$$C_{ji}\frac{\partial t_i}{\partial b_i} + C_{jj}\frac{\partial t_j}{\partial b_i} + k_{ji}\frac{\partial t_i}{\partial b_i} + k_{jj}\frac{\partial t_j}{\partial b_i} = 0$$

$$\frac{\partial t_j}{\partial b_i} = -\frac{(C_{ji} + k_{ji})}{(C_{jj} + k_{jj})}\frac{\partial t_i}{\partial b_i} \tag{5}$$

The effect of the bonus on task i on effort in task j depends on whether i and j are complements or substitutes in the contractor's private cost function. To see this, note that a bonus on one task will increase effort in that task $\frac{\partial t_i}{\partial b_i} > 0$ and $C(t, a)$ and $k(t, a)$ are convex such that their second derivatives are positive, thus $(C_{jj} + k_{jj}) > 0$. If i and j are complementary in a firm's private cost function, such that $(C_{ji} + k_{ji}) < 0$, the expression is positive and a bonus on i leads to an increase in effort on j. Whereas if they are substitutes, $(C_{ji} + k_{ji}) > 0$, the expression is negative and a bonus on i leads to a decrease in effort on j.

Weatherization may be one setting where complementaries exist among certain tasks. For example, better wall and attic insulation can increase returns to effort on air sealing. To the extent these complementarities exist, there may be some potential for a piece rate bonus on the blower door reading to increase effort not only on air sealing, but other tasks as well. However to the extent some tasks are substitutes to air sealing, minimum quality standards will help reduce incentives to pull effort away from those tasks. We can see from equation 5, the more convex the effort cost and callback function, the less responsive effort in one dimension will be to bonuses on another dimension.

4 Data

We make use of three types of data to estimate the effects of pay-for-performance intervention on program outcomes and cost-effectiveness. We obtain comprehensive administrative data on the homes and households served by the IHWAP program. This includes all home characteristics and building measurements recorded during the pre-weatherization energy audit, pre/post blower door test, homeowner demographics and household information, contractor information, the expected labor and material costs for all retrofits completed as part of a project, and project audit and completion dates.

We additionally obtained information on all inspection reports completed by quality control inspectors serving each of the agencies serving the program. This information includes measurements of deficiencies in contractor work (blower door test, insulation thickness, installations), which are classified as one of two categories: (1) callbacks identify a deficiency that is sufficiently problematic to warrant the return of a contractor prior and re-inspection prior to finalizing a project and issuing payment; (2) findings identify a deficiency irrespective of whether it is sufficiently problematic to warrant a callback. Finally, we obtained pre-treatment and post-treatment monthly billing data from the three major utilities (gas and electric) that serve households participating in the IHWAP. We convert energy use measurements to monthly MMBTUs for a consistent metric for measuring effects on gas and electricity consumption.

Table 1 provides descriptive statistics for the variables examined in the study and reports the results of tests for balance across each of the treatment groups. The average home in the sample has 2.8 bedrooms, 1.3 stories, and a total living space of 1,450 square feet. The average household has 2.4 occupants. The average CFM50 level recorded in a pre-weatherization blower door test is 3,800-4,000. The baseline CFM50 levels can range up to 10,000 in some homes. We exclude homes with baseline CFM50 levels that exceed 10,000 due to concern about measurement error in the tests for this sample of homes. Homes can legitimately have CFM50 levels over 10,000, but those homes also have extremely large discrete leaks and so are unlike the rest of the sample.

As a stratified randomization process was not possible in our context, there are a few characteristics of the jobs that are not completely balanced across treatments. In particular, the low treatment group has somewhat leakier homes (higher CFM50) than the control and high treatment groups at baseline. Since the pre-weatherization blower door reading determines the target CFM50 reductions for a home, this imbalance could affect the comparison of raw means across treatments. Therefore, all of our estimates of treatment effects include controls for the pre-weatherization blower door reading. We also provide estimates of treatment effects controlling for home characteristics and expenditures to control for any other imbalance in the samples of homes.

5 Results

5.1 The Contractor Response to Performance Incentives

We begin by investigating the effects of treatment on the contracted outcome. A first set of estimates tests the effects of treatment on CFM50 reductions (post - pre) and a second set of estimates identifies the effects on deficiencies in air sealing identified as part of a post-weatherization quality control control inspection.

We estimate treatment effects using the following model:

$$Y_i = \beta_0 + \beta_1 T_{Hi} + \beta_2 T_{Li} + \mathbf{X_i'}\beta + \varepsilon_i, \tag{6}$$

where Y_i is an experimental outcome (callback indicator or CFM50 reduction), T_{Hi} is an indicator for the high treatment group with signed contractor, T_{Li} is an indicator for the low treatment group with signed contractor, The matrix $\mathbf{X}_i$ contains vectors of controls. We include month-year fixed effects for the month the job was completed, and we flexibly control for any unobserved characteristics caused by the imbalance in pre-treatment covariates using indicators for binned values of a home's pre-treatment blower door measurement and for the retrofit spending and home characteristics variables in Table 1.

A small number of contractors did not sign up for the performance pay program and contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois. Given the possibility that take-up was correlated with unobservable skills that affect success in CFM50 outcomes, we report all main results obtained from a 2-stage least squares estimator that uses the randomized treatment assignment as an instrument for treatment jobs with eligible contractors.

Effects on Air-Sealing (CFM50 Reductions)

Table 2 reports estimates of average treatment effects on CFM50 reductions. Panel A reports estimates for the pooled treatments groups. All estimates include controls for the pre-weatherization blower door reading to control for imbalance in that characteristic among the low treatment group and the high treatment and control groups. Columns 2-4 progressively add more inclusive sets of controls for expenditures across the retrofit categories, month of completion and home characteristics. The magnitudes of the estimates are consistent across all specifications. The most precisely estimated coefficient in column 4 indicates that, on average, the bonus payment regime results in an additional reduction of 89 CFM50 beyond the 1563 control mean reduction achieved using the pre-set compensation at the level of the target (standard). Panel B reports estimates for each treatment separately. Point estimates suggest that the magnitude of CFM50 reductions was higher among homes randomized into the high bonus treatment than the low bonus treatment, though the difference is not not statistically distinguishable in our sample.

We then examine differences in the effects of the performance pay treatment on higher- versus lower-quality contractors. We identify higher-quality contractors on the basis of their performance during program year 2017, the year prior to the performance pay intervention. We do this by estimating each contractor's mean affect on gas reductions on jobs in 2017 conditional on observable characteristics about the home and household and expenditures on retrofits performed.[15] We then group the contractors into quintiles based on their mean reductions. We define the top two quintiles as "high quality" contractors.

Table 3 reports estimates of treatment effects from a model adding: 1) an indicator for the job being performed by a high quality contractor and 2) an interaction between

---

[15] Mechanically, these are the estimated contractor fixed effects from a model regressing house-specific gas savings on contractor fixed effects along with flexible controls for home and household characteristics, service utility, and expenditures on retrofits performed. We calculate house-specific savings in two steps. First, we estimate counterfactual consumption based on county, month, and year fixed effects along with flexible controls indicating bins of home and household characteristics. Then, we subtract observed consumption from this counterfactual to get house-month treatment effects.

that indicator and treatment to equation 6. Consistent with a model where higher ability contractors can achieve output at lower costs, the results reveal significantly and substantially stronger responses to the bonus from the high quality group. In our sample, CFM50 reductions in homes with the performance bonus were more than twice as large when assigned to contractors in the high quality group as they were when assigned to their counterparts in the lower quality group.

Effects on Air-Sealing Callbacks

Table 4 reports estimates of average treatment effects on deficiencies in air sealing work documented in each project's post-weatherization quality control inspection. Callback data were collected from government agencies administering the program and were available for nearly all (1670/1698) of the jobs in our sample.[16] Panel A reports estimates for the pooled treatment groups. The estimate in Column 4 indicates that performance incentives reduced the probability of a deficiency by 2.95 percentage points, a reduction of just under half of the 8.4% control mean callback rate. Point estimates in Panel B suggest that effects on callback rates may primarily come from responses to the higher bonus payment, though the differences between the groups are not statistically significant in our sample.

5.2 Do Performance Incentives Reduce Energy Use?

The welfare relevant outcome of interest is energy savings. The random assignment of performance bonuses allows us to disentangle the effect of these incentives from the base effect of weatherization under minimum quality standards using the following model:

$$Y_{it} = \beta_0 + \beta_1 W * T_{Hit} + \beta_2 W * T_{Lit} + \beta_3 W_{it} + W_{it} \cdot \mathbf{X_i'}\beta + \delta_t + \gamma_i + \varepsilon_{it} \tag{7}$$

where Y_{it} is the energy use (MMBtu) for household i in month t, $W * T_{Hit}$ is an indicator for the post-weatherization condition in the high treatment group, $W * T_{Lit}$ is an indicator for the post-weatherization condition in the low treatment group. We include month-of-sample (δ_t) fixed effects to control for monthly consumption patterns common to all households and home fixed effects (γ_i) to control for any time-invariant unobserv-

---

[16] In Appendix Table D1 we show that the blower reduction results with this subsample are consistent with the full sample presented in Table 2

able factors about a home that affect consumption. To control for any imbalances in covariates among the groups, we allow the baseline weatherization effect to vary by observable characteristics of the home and household. The matrix $W_{it} \cdot \mathbf{X_i}$ is an interaction between an indicator for post-weatherization and a vector of controls, including home's pre-treatment blower door measurement, spending for each retrofit category, indicators for the month of sample the project was completed, and home characteristics variables. As before, we estimate the model using 2 stage least squares estimator where the randomized treatment assignment instruments for treatment jobs with eligible contractors. Because utility consumption data is only available for a subset of the homes in our experiment (1216/1670), we also report results on our primary blower door outcome for this subsample in Appendix Table D2, which are consistent in magnitude with those presented in Table 2.

Table 5 reports estimates of average treatment effects on monthly gas consumption. Panel A provides pooled estimates of treatment effects. All models reported include house and month-of-sample fixed effects as well as a control for the interaction between weatherization and pre-treatment CFM50 levels. Columns 2-4 progressively add controls for weatherization interacted with: retrofit-specific expenditures (Column 2), fixed effects for the month-year of project completion (Column 3), and property characteristics (Column 4). Estimates are consistent across the specifications and indicate a pooled treatment effect of 0.29-0.38 MMBtu from pay-for-performance incentives. Estimates with the most comprehensive controls indicate that the bonus increases the 1.6 MMbtu mean weatherization effect by 24%. Point estimates in Panel B indicate a small additional effect from high bonus treatment, though these estimates cannot rule out a potentially large additional effect.

We then examine differences in the effects on gas reductions for higher versus lower quality contractors at baseline. Table 6 reports estimates of the differential treatment effects for high quality contractors relative to lower quality contractors. Point estimates suggest stronger responses from pay for performance incentives in the high quality group, though these differences are not statistically significant in our sample.

We report estimates of average treatment effects on monthly electricity consumption in Appendix Table C1. The estimates in this table suggest no evidence of an effect on electricity consumption, which is considered a comparatively minor component of the Weatherization Assistance Program. This can be seen from the comparative mean effect of Weatherization -0.47 MMbtu in control homes, which is approximately 30% the magnitude of the effect of Weatherization on gas consumption. Since the majority of households in the sample heat their homes with gas and use electricity for lighting and households appliances, the primary channel for effects on electricity consumption come through the subset of homes that intensively use air conditioning in summer.

5.3 Do Workers Re-allocate Effort?

Between-Job Reallocation Effects

If contractors were reallocating effort from control jobs that are completed simultaneously with a treated contract, it could bias our treatment effect estimates. One contribution of our study is that our experimental design allows us to estimate the impact of reallocation between jobs. To do so, we include controls for the numbers of simultaneously treated and control projects completed by the same contractor in our main regression specifications. We define a job as being simultaneous to this job if its completion window, defined as the time between the pre-weatherization audit and the completion date overlapped with this jobs completion window. Random assignment guarantees that for CFM50 reduction for a given project i in month t the ratio of work orders in treatment vs in control assigned simultaneously will be exogenous. We estimate effects using the following model:

$$\begin{aligned} Y_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 T_{it} \times \#SimulTreatJobs + \beta_3 T_{it} \times \#SimulControlJobs + \\ \beta_4 \#SimulTreatJobs + \beta_5 \#SimulControlJobs + \mathbf{X_i'}\beta + \epsilon_{it}, \quad (8) \end{aligned}$$

where Y_{it} is the CFM50 reduction for project i in month t, $\#SimulTreatJobs$ is the number of treated jobs (demeaned) completed simultaneously with project i, and $\#SimulControlJobs$ is the number control jobs (demeaned) completed simultaneously with project i. As in our main treatment effects estimation, the matrix $\mathbf{X}_i$ contains month-year fixed effects for the month the job was completed, and indicators for binned values of a home's pre-treatment blower door measurement and binned values for the retrofit spending and home characteristics variables in Table 1. We use the pooled version of treatment in the main text for ease of interpretation, given the number of interaction terms. However, we provide results including interactions of high and low treatment separately in Appendix Table G1.

Given that we used the demeaned number of treatment and control jobs, β_1 represents the effect of the incentive treatment for a job where the contractor has the mean number of simultaneous treatment and control jobs. The coefficients β_4 and β_5 are estimates of the effect of an additional simultaneous treatment or control job on CFM50 reductions for all jobs. Whereas β_2 and β_3 estimate the differential effect of simultaneous treatment or control jobs on CFM50 reductions for treated jobs. If the coefficient on treatment changes little with the inclusion of these controls, it indicates that spillovers to simultaneous jobs are not a significant biasing factor in our main estimation.

Table 7 reports estimates of the impacts of simultaneous jobs on CFM50 reductions. We find no evidence of statistical differences in CFM50 reductions from additional control or treatment homes on either treatment or control jobs. While we cannot rule out modest spillover effects given the point estimates and standard errors, the point estimate of the treatment coefficient changes little with the inclusion of these controls. This suggests that reallocation is not driving the magnitude of our treatment effects in Table 2.

Table 8 reports analogous estimates of the impacts of simultaneous jobs on gas consumption. We estimate equation 7 with additional terms for post-weatherization-by-number of simultaneous treatment or control jobs and post-weatherization-by-treatment-by number of simultaneous treatment or control jobs. We again use demeaned counts of simultaneous jobs. Contrary to the re-allocation hypothesis, we find some evidence of *larger* weatherization effects with additional simultaneous control jobs and *smaller* weatherization effects with additional simultaneous treatment jobs. However, the point estimates are relatively small in magnitude: less than 10% of the pooled treatment effect. Given the random assignment of contract incentives, the expected impact of these equal and opposite effects on simultanously completed jobs euqates to a statistical zero. We do not find any evidence of simultaneous jobs deferentially affecting treatment jobs. As with the blower door outcome, the estimated effects of treatment on gas consumption are quite consistent with those in Table 5, suggesting that spillover effects are not a significant biasing factor in our primary analysis.

Callback Rates within Treated Jobs

Rather than shift effort between jobs, contractors may respond to incentives by allocating their effort away from non-incentivized tasks *within* a given project when it is assigned to the performance pay treatment. We use data on callbacks for deficiencies associated with non-incentivized retrofit tasks to test for evidence of the effects of the performance pay treatment on the probability of a documented deficiency on non-incentivized tasks: furnace repairs and replacements, water heater repairs and replacements, lighting, and setback thermostats. Table 9 reports the results of this test using the same set of specifications used to examine effects on deficiencies in air sealing (Table 4). The estimates in Panels A and B do not suggest any evidence of an effect of performance incentives on the quality of work done on non-incentivized tasks. This suggests that quality control standards, which are an important feature of contracts in the IHWAP, may limit the re-allocation of contractor effort in the context of performance incentives.

Complementarities Among Tasks

Our evidence suggests that substitution among tasks between and within jobs does not appear to be a strong driver of our treatment effects. However, as described in our model, to the extent that there are complementarities among tasks, the bonus may induce effort in non-bonus tasks. Figure 2 plots the relationship between residual variation in CFM50 reductions and reductions in monthly gas usage from our preferred specification of equation 6. The plot illustrates a strong positive correlation between increases in building tightness and delivered gas savings in the control and treatment groups. Interestingly the relationship appears stronger in the treatment group. This indicates that, conditional on delivering the same blower door reduction and on all of the observable things about the home, work performed, and household, energy reductions were larger in treatment homes than in control. This suggests that contractors responded to the bonus by improving quality in dimensions of the job that do not directly affect the blower door reading. One possible explanation is that there were complementaries in contractors' costs, which lead to out-sized effects of the bonus payments on other dimensions of the job.

6 Weatherization: Marginal Benefit of Public Funds

We now use the experimental estimates above to estimate the marginal value of public funds of the performance pay bonuses in the IHWAP using the framework proposed by Finkelstein and Hendren (2020). The marginal value of public funds (MVPF) is defined as the ratio of the marginal benefit of public expenditure to the net marginal cost to the government as follows.

$$\text{MVPF} = \frac{\sum_{t=1}^{T}\left[\frac{\hat{\beta}^{e}\times \text{cost}_{\text{elec}}}{(1+r)^{t}} + \frac{\hat{\beta}^{g}\times \text{cost}_{\text{gas}}}{(1+r)^{t}}\right] + PS}{\text{Bonus Cost}} \tag{9}$$

The numerator reflects the marginal benefit of the incentives where $\hat{\beta}^e$ and $\hat{\beta}^g$ are the estimated annual electricity and natural gas savings. We convert predicted savings into monetary benefits using the social benefits of avoided energy consumption, including avoided generation, transmission and distribution costs, as well as benefits from reduced GHG and local air pollution, indicated by $\text{cost}_{\text{elec}}$ and cost_{gas} (Davis and Muehlegger, 2010; Borenstein and Bushnell, 2018).[17] T represents the cost-weighted average expected lifetime of the retrofits and r is the DOE-recommended discount rate of 2%. PS is the producer surplus associated with the incentives. We assume that there are no non-energy benefits to the household from weatherization, such as health impacts, which are likely small in magnitude. In the denominator, Bonus Cost, reflects the mean bonus payment.

---

[17] For electricity, we estimate the difference between retail prices and social marginal costs for the areas of the state which we analyze using data provided by Borenstein and Bushnell (2018). We then apply that difference to the month-of-year averages of residential electricity prices for our study period. The resulting social marginal benefits of reductions from all retrofits are: \$8.51 per MMBtu for natural gas and \$37.95 per MMBtu for electricity. We calculate marginal private costs of natural gas for each month-of-year, based on month-of-year average citygate prices in Illinois for the sample years. We assume a social cost of carbon of \$40 per ton, with emissions factors from EPA (1998).

We assume that this reflects the full effects of the performance pay intervention on the government budget as it would not meaningfully affect any other fiscal outlays.[18]

Returns from Performance Incentives

Panel A of Table 10 reports estimates of the MVPF for the performance incentive. The first column displays the average payment for the high and low payment intent-to-treat jobs. Column 2 displays the computed producer surplus for the contractors using the estimated effects of the high ($1.00/CFM50) and low ($0.40/CFM50) performance incentives on CFM50 reductions from Table 2 as two points on the supply curve. We assume that supply is piece wise approximately linear from the CFM50 target to the mean reduction at $0.40 and between the mean reductions at $0.40 and $1.00. The resulting estimates are $13/home in producer surplus for contractors in the low treatment and $99/home in the high treatment.

Columns 3 and 5 display the net present value of the energy benefits assuming two different retrofit lifespans. We consider 19.25, which is the cost-weighted average lifetime of the retrofits performed in our sample, using the expected lifespans for individual upgrades from internal WAP documentation. The individual lifespans range from from 5 years for fluorescent lamps to 25 years for insulation. However, recent engineering literature suggests that insulation measures should have substantially longer lifespans, such as 50 years for cellulose fiber (ISOCELL GmbH, 2014), 35-50 years for expanded polystyrene (EPS) (EUMEPS, 2017; IVH, 2015), or the full building lifetime for extruded polystyrene (XPS) (50-150 years) (EXIBA, 2019). Allowing for a longer lifespan on some of these measures yields a cost-weighted average lifespan of 34.5 years. Given that Table C1 indicates that the average effect of the bonus payments on electricity consumption is not different from 0, we assume $\hat{\beta}^e = 0$. For annual gas savings $\hat{\beta}^g$, we multiply the predicted monthly savings in column 5 in Table 5 by 12.

We calculate the MVPF according to equation 9. Under either lifespan assumption, the incentive treatment was remarkably cost effective, ranging from $1.91 to $6.98 in

---

[18] Unlike the EITC and other programs that affect tax revenue as discussed in (Finkelstein and Hendren, 2020), any fiscal externalities from performance pay in the WAP would likely be small in magnitude.

benefits created for $1 of government expenditure. Notably, the benefits from the low bonus treatment, $4.26 (19.25 year lifespan) to $6.98 (34.5 year lifespan), are substantially higher than those from the high bonus treatment, which range from $1.91 (19.25 year lifespan) to $3.15 (34.5 year lifespan).

Effects of Incentives on Returns to Weatherization Tasks

Panel B of Table 10 reports estimates of the costs, net present energy benefits, and MVPF for the average control home that received weatherization assistance through the IHWAP program during the study period. Annual energy savings are calculated using estimated reductions from the last rows of Table 5 and Table C1. As in Panel A, we monetize the value of these savings using the social costs of natural gas and electricity respectively.

Given the constraints of the empirical exercise, we cannot estimate producer surplus from the IHWAP as we were able to for the incentive payments. Therefore, the value in the MVPF column reflects just the ratio of the net present energy benefits to the program costs. For the sample program years in IHWAP, the social energy benefits are $0.5 per dollar spent on the home under the conservative lifespan and $0.81 per dollar spent under the longer lifespan.[19] In row 2 of Panel B, we use these same returns to spending to estimate the benefits of the air sealing retrofits targeted by the incentive.[20]

What is the effect of performance pay on the MVPF of air sealing measures completed in a home? To answer this question, we compare the estimates of the MVPF from air sealing retrofits in control to the social net benefits from air sealing retrofits with performance bonuses, adding the benefits and costs from baseline air sealing (Panel B) to the benefits and costs from performance bonuses on air sealing (Panel A). The results of this exercise are in Panel C. We find that the bonus treatment has an enormous effect, increasing the MVPF from baseline air sealing retrofits by 65-75%. The MVPF increases from 0.5 to over 0.80-0.86 using the 19.25 year lifespan or from 0.81 to 1.3-1.4 using the

---

[19] In interpreting this result, note that returns per dollar spent were affected by the introduction of braided funding wherein some measures were performed to help households irrespective of SIR. See Appendix Table B2 for cost effectiveness in the years preceding the braided funding regime.

[20] Since we estimate the net benefits of weatherization for each home, we cannot disentangle the net benefits of air sealing from other retrofits done in each home.

34.5 year lifespan. Under the more realistic, higher lifespan assumptions, the performance bonuses turn the social net benefits from air sealing retrofits from not cost-effective to cost-effective.

A final set of calculations in Panel D compares the estimates of the MVPF from the broader set of tasks that may have been affected through complementarity with air sealing retrofits in achieving CFM50 reductions including, foundation, duct repair, and doors and windows. We compare the social net benefits from this set of "All Incentivized" tasks in control (Panel B) to their benefits in a setting with performance bonuses. We find that the bonus treatment has an important effect, increasing the MVPF from this broader set of incentivized tasks from 0.5 to over 0.72-0.76 using the 19.25 year lifespan or from 0.81 to 1.15-1.16 using the 34.5 year lifespan.

Effects of Incentives on Returns to WAP Investments

An additional policy-relevant question from the perspective of the Weatherization Assistance Program specifically concerns the effect of performance incentives on the social net benefits from investments in the WAP program as a whole. We note two caveats regarding our intervention. First, the incentives introduced in the current experiment affect a fairly narrow set of tasks that represent a modest fraction of total average home expenditures – 12% for air sealing tasks and 22% for air sealing tasks plus those that could be affected through compelementarity. Nevertheless, this results in an increase of 10-20% across the full set of IHWAP expenditures. The experimental evidence above indicates that they have an outsized impact on contractor work without compromising quality on non-incentivized tasks. This suggests that the application of performance incentives to all tasks in the program could be substantial. Second, we cannot disentangle the effects of our treatment on returns to DOE funds versus other sources that do not solely target cost-effective reductions in household energy consumption. For instance, non-DOE funds may be used for retrofits that are not intended to be cost-effective in terms of energy savings, but simply as a means to offer aid to low income households. Therefore, the fact that the intervention does not "flip the sign" on overall cost-effectiveness, does not discount its impact. In the program years directly preceding the introduction of these other funding sources, IHWAP investments were cost-effective on average (Appendix Table B2). Thus, in the DOE-only funding regime, the 10-20% increase in cost-effectiveness found here would lead to MPVF that is well above one.

7 Conclusion

As governments increase their reliance on social service providers and contract work to provide key public goods, some have called for the use of performance incentives to better align contracts with measurable outcomes that directly relate to program/policy objectives. This paper presents findings form a 2-year field experiment on the impacts of performance incentives in the Weatherization Assistance Program. We find that performance incentives generate significant increases in gains in from incentivized retrofits (CFM50 reductions), reduce the probability of a reported deficiency in a contractor's air sealing work, and increase the energy savings from weatherization by 24%. Expenditures on performance incentives are highly cost-effective, even when they take the form of bonuses paid on top of standard payments for air sealing tasks in the program.

We test several hypotheses regarding the behavioral mechanisms underlying these effects. We find evidence of disproportionate increases in air sealing outcomes among contractors who were performing at a high level at baseline, suggesting that producers may respond according to (lower) expected marginal costs. We do not find any evidence that performance incentives on air sealing (CFM50) outcomes lead to increases in reported deficiencies on non-incentivized tasks or compromise the energy efficiency outcomes on non-incentivized projects that are completed the same time as incentivized projects. Minimum quality standards on retrofit tasks that are not directly incentivized through the performance pay intervention in the IHWAP may limit the expected returns from the re-allocation of effort in the context of a performance incentive.

These results shed new light on the mechanisms underlying performance pay and their potential impacts in a nation-wide social welfare program. They also have implications for the Weatherization Assistance Program and other energy efficiency programs, where increases in cost-effectiveness could have a critical impact on public investments in climate policy over the next 2 decades. We note that there is a strong precedent for the use of performance incentives in the WAP – the Cook County (CEDA) program in the IHWAP adopted performance pay contracts for air sealing in 2016. While there has been no formal evaluation of the success of that intervention in the CEDA program, our experimental results provide evidence to suggest that it may have important impacts on the cost-effectiveness of CEDA projects and may worth considering at scale.

Figures

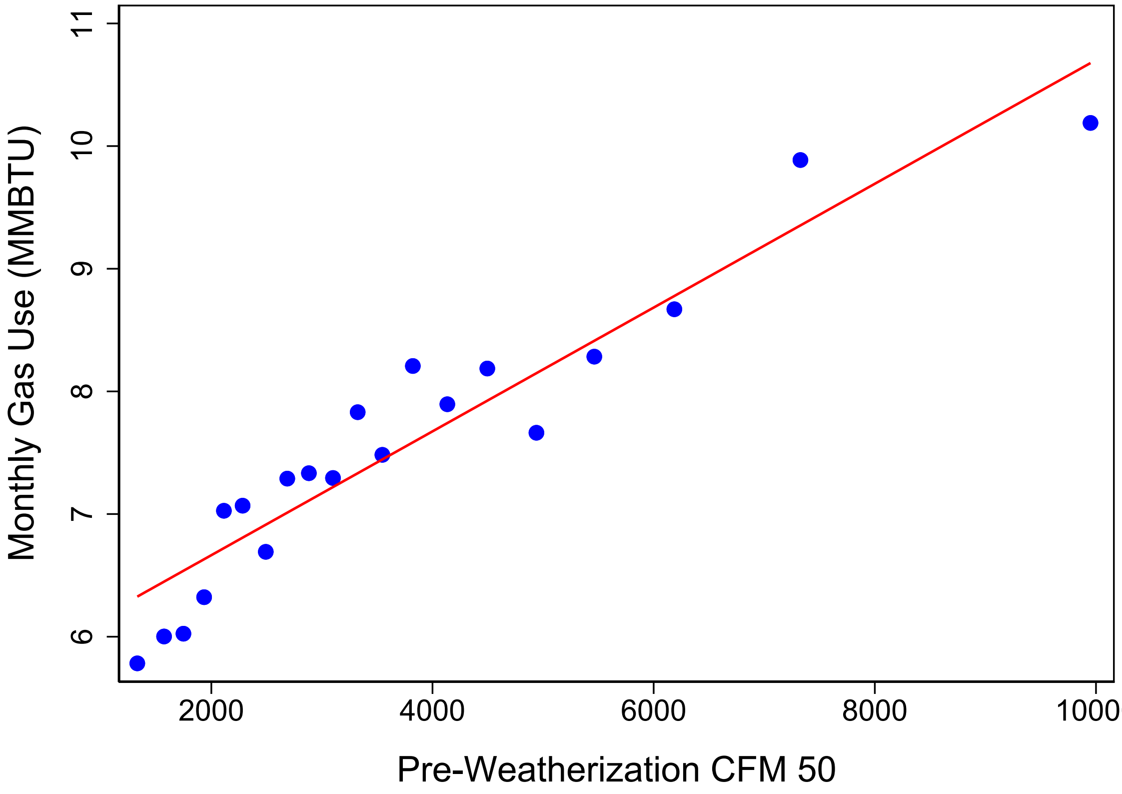

Figure 1: Correlation Between Energy Consumption and CFM Pre-Weatherization

Notes: This figure is a binned scatter plot of the relationship between residual variation in gas consumption and the pre-weatherization blower door reading after controlling for year-month fixed effects. For the purposes of scaling, the mean of each variable is added back to the residuals.

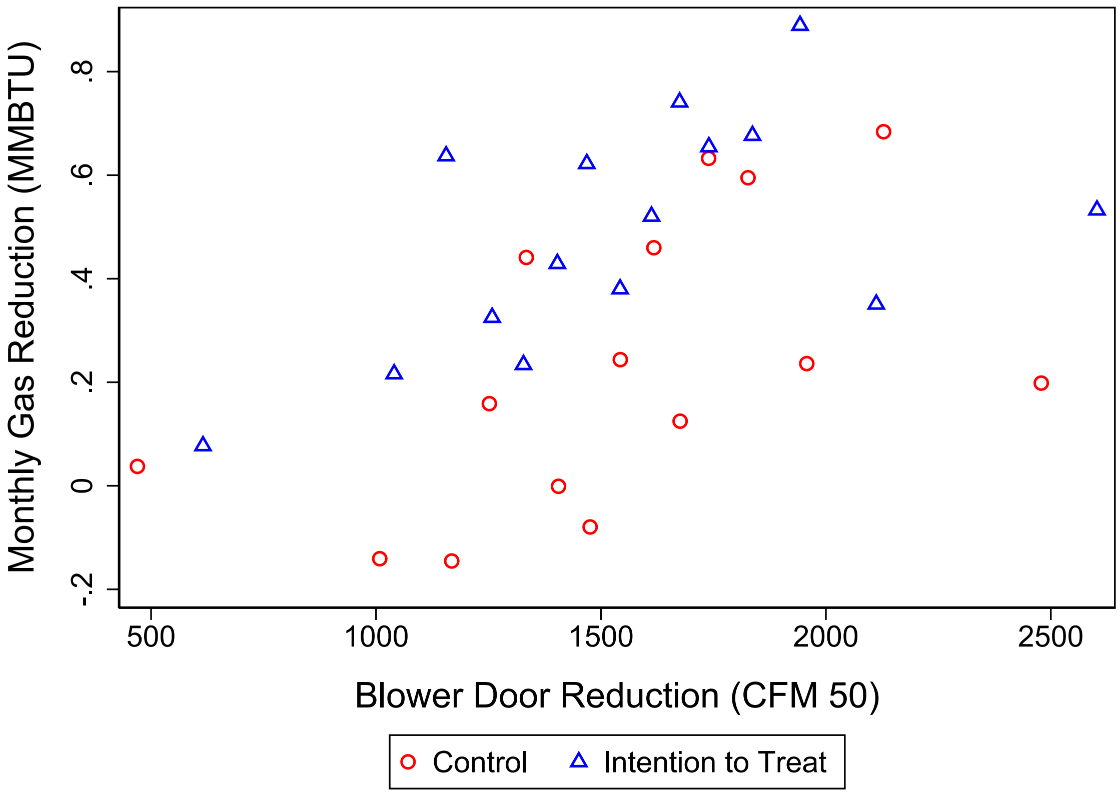

Figure 2: Correlation Between Gas and Blower Door Reductions

Notes: This figure is a binned scatter plot of the relationship between the residual variation in monthly gas reduction and blower door reduction in the preferred specification. The preferred specification includes controls for the pre-weatherization blower door reading, binned expenditure categories for each retrofit, year-month of completion, and binned categories of household characteristics. For the purposes of scaling, the mean of each variable is added back to the residuals.

Tables

Table 1: Balance Tests

	Control ITT	Low ITT	High ITT	T-test Difference		
	(1)	(2)	(3)	(1)-(2)	(1)-(3)	(2)-(3)
Pre Blower Door (CFM50)	3835.990	4090.085	3826.255	-254.095*	9.735	263.830*
	(74.824)	(111.648)	(108.161)			
Stories	1.336	1.376	1.335	-0.039	0.002	0.041
	(0.015)	(0.024)	(0.023)			
Sq. Feet	1469.056	1459.442	1443.180	9.615	25.876	16.261
	(21.846)	(28.668)	(31.959)			
Occupants	2.374	2.368	2.403	0.006	-0.029	-0.036
	(0.057)	(0.079)	(0.084)			
Bedrooms	2.835	2.831	2.855	0.004	-0.021	-0.025
	(0.033)	(0.044)	(0.043)			
General Exp	161.984	172.955	140.070	-10.971	21.914	32.885
	(16.570)	(23.564)	(21.485)			
Furnace Exp	2226.074	2193.126	2166.609	32.948	59.465	26.517
	(53.987)	(85.524)	(77.832)			
Foundation Exp	727.090	808.886	745.107	-81.797	-18.017	63.780
	(30.498)	(51.662)	(47.938)			
Door Exp	96.495	99.175	120.000	-2.680	-23.505	-20.825
	(9.457)	(14.515)	(15.858)			
Baseload Exp	499.631	539.256	484.439	-39.625	15.192	54.817*
	(16.631)	(23.507)	(23.045)			
Attic Exp	1249.658	1229.956	1229.635	19.703	20.024	0.321
	(32.744)	(46.797)	(47.703)			
Air Sealing Exp	977.799	1013.656	967.594	-35.857	10.205	46.062
	(21.801)	(33.399)	(31.675)			
Air Conditioning Exp	1870.994	1634.712	1724.629	236.282***	146.365*	-89.917
	(46.023)	(65.397)	(66.607)			
Water Heater Exp	876.408	892.281	883.056	-15.873	-6.648	9.225
	(32.870)	(47.931)	(49.070)			
Wall Insulation Exp	408.388	408.804	290.798	-0.416	117.590**	118.006**
	(29.040)	(44.335)	(33.620)			
Window Exp	69.454	52.234	56.654	17.220	12.800	-4.420
	(10.616)	(15.330)	(17.499)			
Total Observations	908	446	415			
Observations matched with site details	853	419	387			
Observations matched with callback data	646	316	309			

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 2: Effects of Bonus Treatments on Building Envelope Tightness (CFM50)

CFM50 (Post - Pre)	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Treat	-64.09*	-73.97**	-75.96**	-88.67***
	(32.88)	(31.96)	(32.36)	(31.48)
Panel B: Effect by Treatment Group				
Low Treat	-57.25	-58.58	-54.25	-66.17*
	(40.23)	(39.41)	(39.67)	(39.08)
High Treat	-71.49*	-90.26**	-99.91**	-113.8***
	(40.49)	(39.43)	(39.98)	(39.06)
Pre Blower (CFM)	Yes	Yes	Yes	Yes
Expenditures	No	Yes	Yes	Yes
Month of Completion FE	No	No	Yes	Yes
Characteristics	No	No	No	Yes
Observations	1698	1697	1697	1601
Adjusted R^2	-0.002	-0.034	-0.051	-0.082
Control Group Depvar Mean	-1569.306	-1569.306	-1569.306	-1562.708

Notes: The dependent variable is the change in building envelope tightness as a result of Weatherization Assistance Program upgrades (Post-Pre). Building envelope tightness is measured in units of 50 cubic feet per minute (CFM50), where higher values indicate a leakier home. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Panel A reports results from regressions pooling High and Low payment treatments into one single treatment indicator. Panel B reports results from regressions with separate indicators for High and Low Treatment. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table 3: Effects of Bonus Treatments on Building Envelope Tightness by Contractor Quality

CFM60 (Post - Pre)	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Treat	-41.63	-47.78	-47.26	-61.83*
	(36.73)	(35.68)	(36.22)	(35.12)
Treat × High Quality	-140.9*	-141.0*	-146.3*	-136.5*
	(79.46)	(79.66)	(79.03)	(79.40)
High Quality	-66.74	-30.28	-28.14	-49.12
	(44.01)	(45.53)	(45.67)	(44.12)
Panel B: Effect by Treatment Group				
Low Treat	-45.24	-44.91	-35.68	-45.47
	(45.83)	(44.94)	(45.21)	(43.92)
High Treat	-37.87	-51.26	-60.63	-80.68*
	(43.08)	(41.95)	(42.78)	(42.06)
Low Treat × High Quality	-68.47	-62.73	-85.88	-97.41
	(88.20)	(88.17)	(88.57)	(92.52)
High Treat × High Quality	-233.6**	-240.4**	-225.3**	-188.9*
	(117.8)	(115.2)	(112.4)	(110.9)
High Quality	-66.87	-30.99	-28.64	-49.01
	(44.04)	(45.60)	(45.73)	(44.23)
Pre Blower (CFM)	Yes	Yes	Yes	Yes
Expenditures	No	Yes	Yes	Yes
Month of Completion FE	No	No	Yes	Yes
Characteristics	No	No	No	Yes
Observations	1670	1669	1669	1579
Adjusted R^2	0.003	-0.032	-0.049	-0.078
Control Group Depvar Mean	-1557.014	-1557.014	-1557.014	-1552.336

Notes: The dependent variable is the change in building envelope tightness as a result of Weatherization Assistance Program upgrades (Post-Pre). Building envelope tightness is measured in units of 50 cubic feet per minute (CFM50), where higher values indicate a leakier home. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Panel A reports results from regressions pooling High and Low payment treatments into one single treatment indicator. Panel B reports results from regressions with separate indicators for High and Low Treatment. High Quality indicates that the contractor that performed the work was in the upper 2 quintiles of performance in the program year that preceded the intervention (2017). Performance was measured as mean gas reductions associated with each contractor, conditional on the measures performed and home and household characteristics. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table 4: Effects on Callback Rate: Air Leakage

Reducing Air Leakage	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Treat	-0.0325*	-0.0358**	-0.0265	-0.0295*
	(0.0169)	(0.0167)	(0.0173)	(0.0176)
Panel B: Effect by Treatment Group				
Low Treat	-0.0201	-0.0222	-0.0161	-0.0195
	(0.0214)	(0.0212)	(0.0214)	(0.0216)
High Treat	-0.0455**	-0.0496***	-0.0374*	-0.0402**
	(0.0192)	(0.0189)	(0.0196)	(0.0203)
Pre Blower (CFM)	Yes	Yes	Yes	Yes
Expenditures	No	Yes	Yes	Yes
Month of Completion FE	No	No	Yes	Yes
Characteristics	No	No	No	Yes
Observations	1226	1225	1225	1164
Adjusted R^2	-0.006	-0.050	-0.072	-0.117
Control Group Depvar Mean	0.085	0.085	0.085	0.084

Notes: The dependent variable is an indicator variable for whether the contractor performing the job was "called back" by the quality control inspector (QCI) to perform additional work to rectify an issue related to air sealing. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Panel A reports results from regressions pooling High and Low payment treatments into one single treatment indicator. Panel B reports results from regressions with separate indicators for High and Low Treatment. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table 5: Effect on Gas Usage (MMBtu)

Gas MMBtu	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Weatherization × Treatment	-0.285**	-0.347***	-0.359***	-0.376***
	(0.117)	(0.114)	(0.114)	(0.115)
Panel B: Effect by Treatment Group				
Weatherization × Low Treat	-0.327**	-0.352**	-0.345**	-0.352***
	(0.141)	(0.137)	(0.136)	(0.136)
Weatherization × High Treat	-0.239	-0.340**	-0.375**	-0.402***
	(0.149)	(0.146)	(0.147)	(0.151)
House FE	Yes	Yes	Yes	Yes
Month of Sample FE	Yes	Yes	Yes	Yes
Weatherization × Demeaned Pre Blower (CFM)	Yes	Yes	Yes	Yes
Weatherization × Expenditures	No	Yes	Yes	Yes
Weatherization × Month of Completion FE	No	No	Yes	Yes
Weatherization × Characteristics	No	No	No	Yes
No. of Homes	1216	1216	1216	1164
Observations	66423	66423	66423	63676
Adjusted R^2	0.001	0.007	0.006	0.006
Control Mean Weatherization Effect	-1.607***	-1.607***	-1.607***	-1.623***
	(0.0974)	(0.0974)	(0.0974)	(0.0982)
Control Mean Pre-Weatherization Consumption	7.264	7.264	7.264	7.257

Notes: The dependent variable is monthly gas consumption (MMBtu). Homes with 12 months or more of both pre and post weatherization gas consumption data were included. Weatherization indicates consumption observations post-retrofits. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Panel A reports results from regressions pooling High and Low payment treatments into one single treatment indicator. Panel B reports results from regressions with separate indicators for High and Low Treatment. Standard errors are clustered at the house level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table 6: Effect on Gas Usage (MMBtu) by Contractor Quality

Gas MMBtu	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Weatherization × Treatment	-0.266**	-0.310**	-0.322**	-0.351***
	(0.129)	(0.125)	(0.126)	(0.127)
Weatherization × Treat × High Quality	-0.116	-0.220	-0.231	-0.177
	(0.294)	(0.284)	(0.283)	(0.292)
Weatherization × High Quality	-0.0329	0.126	0.125	0.0488
	(0.198)	(0.193)	(0.191)	(0.192)
Panel B: Effect by Treatment Group				
Weatherization × Low Treat	-0.317**	-0.324**	-0.317**	-0.333**
	(0.159)	(0.153)	(0.152)	(0.152)
Weatherization × High Treat	-0.212	-0.294*	-0.328**	-0.372**
	(0.161)	(0.159)	(0.159)	(0.164)
Weatherization × High Treat × High Quality	-0.169	-0.277	-0.300	-0.220
	(0.412)	(0.395)	(0.399)	(0.417)
Weatherization × Low Treat × High Quality	-0.0676	-0.171	-0.170	-0.141
	(0.320)	(0.315)	(0.313)	(0.319)
Weatherization × High Quality	-0.0326	0.126	0.126	0.0500
	(0.198)	(0.193)	(0.191)	(0.192)
House FE	Yes	Yes	Yes	Yes
Month of Sample FE	Yes	Yes	Yes	Yes
Weatherization × Demeaned Pre Blower (CFM)	Yes	Yes	Yes	Yes
Weatherization × Expenditures	No	Yes	Yes	Yes
Weatherization × Month of Completion FE	No	No	Yes	Yes
Weatherization × Characteristics	No	No	No	Yes
No. of Homes	1204	1204	1204	1154
Observations	65905	65905	65905	63254
Adjusted R^2	0.001	0.007	0.005	0.006
Control Mean Weatherization Effect	-1.607***	-1.607***	-1.607***	-1.623***
	(0.0974)	(0.0974)	(0.0974)	(0.0982)
Control Mean Pre-Weatherization Consumption	7.264	7.264	7.264	7.257

Notes: The dependent variable is monthly gas consumption (MMBtu). Homes with 12 months or more of both pre and post weatherization gas consumption data were included. Weatherization indicates consumption observations post-retrofits. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Panel A reports results from regressions pooling High and Low payment treatments into one single treatment indicator. Panel B reports results from regressions with separate indicators for High and Low Treatment. High Quality indicates that the contractor that performed the work was in the upper 2 quintiles of performance in the program year that preceded the intervention (2017). Performance was measured as mean gas reductions associated with each contractor, conditional on the measures performed and home and household characteristics. Standard errors are clustered at the house level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table 7: Effects on Blower Door: Simultaneous Jobs Within Contractor

CFM50 (Post - Pre)	(1)	(2)	(3)	(4)
Treat	-49.58	-75.65**	-74.08**	-84.22**
	(36.17)	(35.64)	(34.37)	(35.03)
Treat by Simul. Treat Jobs	4.242	6.925	3.391	3.515
	(4.713)	(4.618)	(4.342)	(4.520)
Treat by Simul. Control Jobs	-6.822	-7.450	-2.949	-4.196
	(5.111)	(4.872)	(4.559)	(4.819)
No. of Simul. Treat Jobs	-4.356	-5.187	-1.520	0.0841
	(3.765)	(3.864)	(4.370)	(3.951)
No. of Simul. Control Jobs	3.973	3.704	0.713	-1.229
	(4.326)	(4.165)	(4.287)	(4.389)
Pre Blower (CFM)	Yes	Yes	Yes	Yes
Expenditures	No	Yes	Yes	Yes
Month of Completion FE	No	No	Yes	Yes
Characteristics	No	No	No	Yes
Observations	1697	1696	1695	1600
Adjusted R^2	-0.002	-0.036	-0.080	-0.082
Control Group Depvar Mean	-1569.306	-1569.306	-1569.306	-1562.708

Notes: The dependent variable is the change in building envelope tightness as a result of Weatherization Assistance Program upgrades (Post-Pre). Building envelope tightness is measured in units of 50 cubic feet per minute (CFM50), where higher values indicate a leakier home. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Results are from regressions pooling High and Low payment treatments into one single treatment indicator. Controls for the demeaned number of simultaneous treatment and control jobs that the contractor worked on are included along with the interaction of each of these controls with the treatment indicator. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table 8: Effects on Gas Use: Simultaneous Jobs Within Contractor

Gas MMBtu	(1)	(2)	(3)	(4)
Weatherization × Treatment	-0.266**	-0.353***	-0.381***	-0.405***
	(0.130)	(0.127)	(0.129)	(0.131)
Weatherization × Treat × No. of Siml. Treat Jobs	0.0171	0.0194	0.0189	0.0216
	(0.0172)	(0.0162)	(0.0164)	(0.0164)
Weatherization × Treat × No. of Siml. Control Jobs	-0.0209	-0.0202	-0.0204	-0.0238
	(0.0188)	(0.0174)	(0.0178)	(0.0178)
Weatherization × No. of Siml. Treat Jobs	-0.0252**	-0.0243**	-0.0342**	-0.0338**
	(0.0114)	(0.0110)	(0.0133)	(0.0133)
Weatherization × No. of Siml. Control Jobs	0.0271**	0.0259**	0.0366**	0.0389**
	(0.0133)	(0.0125)	(0.0150)	(0.0151)
House FE	Yes	Yes	Yes	Yes
Month of Sample FE	Yes	Yes	Yes	Yes
Weatherization × Demeaned Pre Blower (CFM)	Yes	Yes	Yes	Yes
Weatherization × Expenditures	No	Yes	Yes	Yes
Weatherization × Month of Completion FE	No	No	Yes	Yes
Weatherization × Characteristics	No	No	No	Yes
No. of Homes	1215	1215	1215	1163
Observations	66372	66372	66372	63625
Adjusted R^2	0.001	0.007	0.005	0.006
Control Mean Weatherization Effect	-1.607***	-1.607***	-1.607***	-1.623***
	(0.0974)	(0.0974)	(0.0974)	(0.0982)
Control Mean Pre-Weatherization Consumption	7.264	7.264	7.264	7.257

Notes: The dependent variable is monthly gas consumption (MMBtu). Homes with 12 months or more of both pre and post weatherization gas consumption data were included. Weatherization indicates consumption observations post-retrofits. Weatherization and Weatherization×Treat are each interacted with the demeaned number of simultaneous treatment or control jobs the contractor worked on. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Results are from regressions pooling High and Low payment treatments into one single treatment indicator. Standard errors are clustered at the house level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table 9: Effects on Callback Rate: Non-Incentivized Retrofits

Non-Building Envelope	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Treat	0.0219	0.0221	0.0151	0.0174
	(0.0141)	(0.0144)	(0.0149)	(0.0152)
Panel B: Effect by Treatment Group				
Low Treat	0.0227	0.0199	0.0131	0.0150
	(0.0178)	(0.0178)	(0.0186)	(0.0195)
High Treat	0.0212	0.0245	0.0172	0.0199
	(0.0179)	(0.0185)	(0.0186)	(0.0186)
Pre Blower (CFM)	Yes	Yes	Yes	Yes
Expenditures	No	Yes	Yes	Yes
Month of Completion FE	No	No	Yes	Yes
Characteristics	No	No	No	Yes
Observations	1226	1225	1225	1164
Adjusted R^2	-0.002	-0.049	-0.072	-0.118
Control Group Depvar Mean	0.040	0.040	0.040	0.040

Notes: The dependent variable is an indicator variable for whether the contractor performing the job was "called back" by the quality control inspector (QCI) to perform additional work to rectify an issue related to retrofits that are not incentivized by the bonus payments. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Panel A reports results from regressions pooling High and Low payment treatments into one single treatment indicator. Panel B reports results from regressions with separate indicators for High and Low Treatment. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table 10: Marginal Value of Public Funds

			19.25 Year Lifespan		34.5 Year Lifespan	
	Cost	Producer Surplus	Net Present Energy Benefits	MVPF	Net Present Energy Benefits	MVPF
Panel A. Performance Incentive						
Low Treat	$115	$13	$477	$4.26	$790	$6.98
High Treat	$290	$99	$543	$1.91	$899	$3.15
Panel B: Baseline WAP						
All Baseline Retrofits	$8,744	.	$4,356	$0.50	$7,062	$0.81
Baseline Air Sealing	$1,085	.	$541	$0.50	$876	$0.81
Baseline All Incentivized	$1,926	.	$960	$0.50	$1556	$0.81
Panel C: Baseline Air Sealing + Incentive						
Baseline Air Sealing + Low Treat	$1,200	.	$1,018	$0.86	$1,666	$1.40
Baseline Air Sealing + High Treat	$1,375	.	$1,083	$0.80	$1,776	$1.30
Panel D: Baseline All Incentivized + Incentive						
Baseline Air Sealing + Low Treat	$2,041	.	$1,542	$0.76	$2,345	$1.16
Baseline Air Sealing + High Treat	$2,216	.	$1,503	$0.72	$2,455	$1.15

Notes: Table reports the marginal value of public funds (MVPF) for: (1) performance incentives for air sealing retrofits, (2) baseline investments in air sealing retrofits completed as part of weatherization in all homes, and (3) air sealing retrofits with performance incentives. Panel A reports estimates of the social net benefits and MVPF for expenditures on performance incentives on air sealing retrofits using estimates of treatment effects from Table 5. Panel B reports estimates of social net benefits and MVPF for all baseline retrofits and baseline air sealing retrofits. Estimates of benefits from baseline air sealing retrofits are assumed to be proportional to expenditures on air sealing given control mean weatherization effect. Panel C reports estimates of social net benefits and MVPF of combining the baseline air sealing investments with performance incentives. Net present energy benefits use gas and electricity prices per MMBTU for 2017. Emissions factors were obtained from EPA (1998). Data provided by Borenstein and Bushnell (2018) is used to estimate the difference between retail residential electricity prices and social marginal costs for the study region. The resulting social marginal benefits of reductions from all retrofits are: $8.51 per MMBtu for natural gas and $37.95 per MMBtu for electricity. Retrofit lifespans are based on the weighted average for of retrofit-specific lifespans in the average home in the sample: 19.25 years when assuming a 20-year lifespan for long-lived insulation materials vs. 34.5 years when assuming a 150-year lifespan for long-lived air-sealing materials.

References

Allcott, Hunt and Michael Greenstone (2017). “Measuring the Welfare Effects of Residential Energy Efficiency Programs”. *NBER Working Paper* 23386.

Aucejo, Esteban, Teresa Romano, and Eric Taylor (2019). “Does evaluation distort teacher effort and decisions? Quasi-experimental evidence from a policy of retesting students”.

Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat (2021). “The allocation of authority in organizations: A field experiment with bureaucrats”. *The Quarterly Journal of Economics* 136(4), pp. 2195–2242.

Barlevy, Gadi and Derek Neal (2012). “Pay for percentile”. *American Economic Review* 102(5), pp. 1805–31.

Berry, Linda G. and Michael B. Gettings (1998). “Realization Rates of the National Energy Audit”. *Thermal Performance of the Exterior Envelopes of Buildings VI Conference.* URL: https://weatherization.ornl.gov/wp-content/uploads/pdf/1996_2000/RealizationRatesoftheNationalEnergyAudit.pdf.

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu (2020). “The glittering prizes: Career incentives and bureaucrat performance”. *The Review of Economic Studies* 87(2), pp. 626–655.

Biasi, Barbara (2021). “The labor market for teachers under different pay schemes”. *American Economic Journal: Economic Policy* 13(3), pp. 63–102.

Blonz, Joshua (2018). “The Welfare Costs of Misaligned Incentives: Energy Inefficiency and the Principal-Agent Problem”. *University of California, Berkeley. Energy Institute. Working Paper* 297.

Bloom, Nicholas and John Van Reenen (2011). “Human Resource Management and Productivity”. *Handbook of labor economics.* Vol. 4. Elsevier, pp. 1697–1767.

Borenstein, Severin and James B Bushnell (2018). “Do Two Electricity Pricing Wrongs Make a Right? Cost Recovery, Externalities, and Efficiency”. *NBER Working Paper* 24756. URL: http://www.nber.org/papers/w24756.

Bruegge, Chris, Tatyana Deryugina, and Erica Myers (2019). “The distributional effects of building energy codes”. *Journal of the Association of Environmental and Resource Economists* 6(S1), S95–S127.

Burgess, Simon, Carol Propper, Marisa Ratto, Emma Tominey, et al. (2017). “Incentives in the Public Sector: Evidence from a Government Agency”. *Economic Journal* 127(605), pp. 117–141.

California Air Resources Board (2017). *California’s 2017 Climate Change Scoping Plan.* [Online; accessed in 2020]. URL: https://ww3.arb.ca.gov/cc/scopingplan/scoping_plan_2017.pdf.

Christensen, Peter, Paul Francisco, Erica Myers, and Mateus Souza (2020). *Decomposing the Wedge Between Projected and Realized Returns in Energy Efficiency Programs.* Working Paper.

Dalhoff, Gregory K. (1997). “An Evaluation of the Performance of the NEAT Audit for the Iowa Low-Income Weatherization Program”. *1997 Energy Evaluation Conference.* URL: https://weatherization.ornl.gov/wp-content/uploads/pdf/1996_2000/IEPECDalhoffIowaNEATEvaluation.pdf.

Davis, Lucas W, Sebastian Martinez, and Bibiana Taboada (2018). *How Effective is Energy-Efficient Housing? Evidence from a Field Experiment in Mexico.* Tech. rep. National Bureau of Economic Research.

Davis, Lucas W, Sebastian Martinez, and Bibiana Taboada (2020). “How effective is energy-efficient housing? Evidence from a field trial in Mexico”. *Journal of Development Economics* 143, p. 102390.

Davis, Lucas W., Alan Fuchs, and Paul Gertler (2014). “Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico”. *American Economic Journal: Economic Policy* 6(4), pp. 207–38.

Davis, Lucas W. and Erich Muehlegger (2010). “Do Americans consume too little natural gas? An empirical test of marginal cost pricing”. *The RAND Journal of Economics* 41(4), pp. 791–810. URL: http://www.jstor.org/stable/25746054.

Dohmen, Thomas and Armin Falk (2011). "Performance pay and multidimensional sorting: Productivity, preferences, and gender". *American economic review* 101(2), pp. 556–90.

Domanski, Piotr A, Hugh I Henderson, W Vance Payne, et al. (2014). *Sensitivity analysis of installation faults on heat pump performance*. US Department of Commerce, National Institute of Standards and Technology.

European Extruded Polystyrene Insulation Board Association (2019). "Extruded Polystyrene (XPS) Foam Insulation with halogen free blowing agent". *Institut Bauen und Umwelt e.V. (IBU)*. URL: https://ibu-epd.com/en/published-epds/.

European Manufacturers of Expanded Polystyrene (2017). "Expanded Polystyrene (EPS) Foam Insulation (shape moulded, density 25 kg/m^3)". *Institut Bauen und Umwelt e.V. (IBU)*. URL: https://eumeps.construction/content/8-downloads/2-epds-for-eps/expanded-polystyrene-eps-foam-insulation-shape-moulded-density-25-kgm.pdf.

European Parliament (2012). "Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency". URL: http://data.europa.eu/eli/dir/2012/27/oj.

Executive Office of Energy and Environmental Affairs (2018). *Massachusetts Global Warming Solutions Act: 10-Year Progress Report*. [Online; accessed in 2020]. URL: https://www.mass.gov/doc/gwsa-10-year-progress-report/download.

Finkelstein, Amy and Nathaniel Hendren (2020). "Welfare analysis meets causal inference". *Journal of Economic Perspectives* 34(4), pp. 146–67.

Fioretti, Michele and Hongming Wang (2019). *Subsidizing inequality: Performance pay and risk selection in medicare*. HAL Working Paper.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram (2018). "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program". *The Quarterly Journal of Economics* 133(3), pp. 1597–1644.

Franceschelli, Ignacio, Sebastian Galiani, and Eduardo Gulmez (2010). “Performance pay and productivity of low-and high-ability workers”. *Labour Economics* 17(2), pp. 317–322.

Frederiksen, Anders, Daniel Baltzer Schjødt Hansen, and Colleen Flaherty Manchester (2022). “Does Group-Based Incentive Pay Lead To Higher Productivity? Evidence from a Complex and Interdependent Industrial Production Process”.

Friebel, Guido, Matthias Heinz, Miriam Krueger, and Nikolay Zubanov (2017). “Team incentives and performance: Evidence from a retail chain”. *American Economic Review* 107(8), pp. 2168–2203.

Gaduh, Arya, Menno Pradhan, Jan Priebe, and Dewi Susanti (2021). “Scores, Camera, Action”.

Giraudet, Louis-Gaëtan, Sébastien Houde, and Joseph Maher (2018). “Moral hazard and the energy efficiency gap: Theory and evidence”. *Journal of the Association of Environmental and Resource Economists* 5(4), pp. 755–790.

Gittleman, Maury and Brooks Pierce (2013). “How prevalent is performance-related pay in the United States? Current incidence and recent trends”. *National Institute Economic Review* 226, R4–R16.

Holmstrom, Bengt and Paul Milgrom (1991). “Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design”. *JL Econ. & Org.* 7, p. 24.

Hong, Fuhai, Tanjim Hossain, John A List, and Migiwa Tanaka (2018). “Testing the theory of multitasking: evidence from a natural field experiment in Chinese factories”. *International Economic Review* 59(2), pp. 511–536.

Houde, Sébastien and Joseph E. Aldy (2014). “Belt and Suspenders and More: The Incremental Impact of Energy Efficiency Subsidies in the Presence of Existing Policy Instruments”. *NBER Working Paper* 20541.

IEA (2020). “World energy investment 2020”. IEA Paris.

Industrieverband Hartschaum e.V. (2015). “EPS rigid foam (Styropor®) for ceilings/floors and as perimetric insulation (in German)”. *Institut Bauen und Umwelt e.V. (IBU).* URL: https://ibu-epd.com/en/published-epds/.

International Energy Agency (2018). "Energy Efficiency 2018: Analysis and outlooks to 2040".

ISOCELL GmbH (2014). "Blown insulation made of cellulose fibre". *Bau EPD GmbH.* URL: http://www.bau-epd.at/wp-content/uploads/2014/11/EPD-_ISOCELL_Ecoinvent_20140825-English.pdf.

Khan, Adnan Q, Asim I Khwaja, and Benjamin A Olken (2016). "Tax farming redux: Experimental evidence on performance pay for tax collectors". *The Quarterly Journal of Economics* 131(1), pp. 219–271.

Khan, Adnan Q, Asim Ijaz Khwaja, and Benjamin A Olken (2019). "Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings". *American Economic Review* 109(1), pp. 237–70.

Kim, Hyuncheol Bryant, Seonghoon Kim, and Thomas T Kim (2020). "The role of career and wage incentives in labor productivity: Evidence from a two-stage field experiment in malawi". *Review of Economics and Statistics* 102(5), pp. 839–851.

Lavy, Victor (2009). "Performance pay and teachers' effort, productivity, and grading ethics". *American Economic Review* 99(5), pp. 1979–2011.

Lavy, Victor (2020). "Teachers' Pay for Performance in the Long-Run: The Dynamic Pattern of Treatment Effects on Students' Educational and Labour Market Outcomes in Adulthood". *The Review of Economic Studies* 87(5), pp. 2322–2355.

Lazear, Edward P (2000). "Performance pay and productivity". *American Economic Review* 90(5), pp. 1346–1361.

Lazear, Edward P (2018). "Compensation and Incentives in the Workplace". *Journal of Economic Perspectives* 32(3), pp. 195–214.

Lazear, Edward P and Kathryn L Shaw (2007). "Personnel economics: The economist's view of human resources". *Journal of economic perspectives* 21(4), pp. 91–114.

Lemieux, Thomas, W Bentley MacLeod, and Daniel Parent (2009). "Performance pay and wage inequality". *The Quarterly Journal of Economics* 124(1), pp. 1–49.

Levinson, Arik (2016). "How Much Energy Do Building Energy Codes Save? Evidence from California Houses". *American Economic Review* 106(10), pp. 2867–94.

Muralidharan, Karthik and Venkatesh Sundararaman (2011). “Teacher Performance Pay: Experimental evidence from India”. *Journal of political Economy* 119(1), pp. 39–77.

Nadel, Steven and Kenneth Keating (1991). “Engineering Estimates vs. Impact Evaluation Results: How do they compare and why?” *American Council for an Energy-Efficient Economy, Research Report.* URL: https://www.aceee.org/sites/default/files/publications/researchreports/U915.pdf.

Sharp, T.R. (1994). “The North Carolina Field Test: Field Performance of the Preliminary Version of an Advanced Weatherization Audit for the Department of Energy’s Weatherization Assistance Program”. *ORNL/CON-362, Oak Ridge National Laboratory, Oak Ridge, Tennessee.* URL: https://weatherization.ornl.gov/wp-content/uploads/pdf/1991_1995/ORNL_CON-362.pdf.

Shearer, Bruce (2004). “Piece rates, fixed wages and incentives: Evidence from a field experiment”. *The Review of Economic Studies* 71(2), pp. 513–534.

The White House (2021). “FACT SHEET: The American Jobs Plan”. URL: https://www.whitehouse.gov/briefing-room/statements-releases/2021/03/31/fact-sheet-the-american-jobs-plan/.

US Department of Energy (2013). “Weatherization Program Notice 14-4”. *Published by the National Association for State Community Services Programs.* URL: https://nascsp.org/wp-content/uploads/2018/02/WPN-14-4.pdf.

US Environmental Protection Agency (1998). “AP 42, Fifth Edition Compilation of Air Pollutant Emissions Factors, Volume 1: Stationary Point and Area Sources”. *Technical Report.* URL: https://www3.epa.gov/ttn/chief/ap42/ch01/final/c01s04.pdf.

Zivin, Joshua G. and Kevin Novan (2016). “Upgrading Efficiency and Behavior: Electricity Savings from Residential Weatherization Programs”. *The Energy Journal* 37(4).

Appendices

A Experimental Design

Performance Incentive: Contract Language

High Bonus Contract:

Architectural work on this job is eligible for a bonus of $1.00 for each CFM50 reduced below 10% above the target number. For example, if the target is 2,500 CFM50, every CFM50 below 2,750 is paid a bonus of $1.00. If the final blower door reading is:

- 2,650 then you will be paid a bonus of $100.00

- 2,550 then you will be paid a bonus of $200.00

- 2,450 then you will be paid a bonus of $300.00

Readings will be based on the QCI's final blower door reading. All discrepancies will be determined by the agency. Any purposely inflated numbers to receive unearned bonuses will be cause for disqualification from the bonus program. In order to receive bonus payments, contractors performing architectural work must be registered as a vendor with the University of Illinois.

Low Bonus Contract:

Architectural work on this job is eligible for a bonus of 40 cents for each CFM50 reduced below 10% above the target number. For example, if the target is 2,500 CFM50, every CFM50 below 2,750 is paid a bonus of 40 cents If the final blower door reading is:

- 2,650 then you will be paid a bonus of $40.00

- 2,550 then you will be paid a bonus of $80.00

- 2,450 then you will be paid a bonus of $120.00

Readings will be based on the QCI's final blower door reading. All discrepancies will be determined by the agency. Any purposely inflated numbers to receive unearned bonuses will be cause for disqualification from the bonus program. In order to receive bonus pay-

ments, contractors performing architectural work must be registered as a vendor with the University of Illinois.

B IWHAP Trends in Spending and Cost Effectiveness by Program Year

Table B1: IHWAP Program Trends

Program Year	DOE Predicted SIR	Homes Served	Average Spending (2021 USD)
2013	3.59	5687	$6,078
2014	2.27	3638	$6,370
2015	2.22	4683	$6,195
2016	3.02	2477	$7,129
2017	1.67	1763	$12,104
2018	2.12	1532	$12,535
2019	2.09	1891	$11,089
2020	1.78	1611	$10,739

Notes: Table reports the DOE predicted overall SIR for a home, number of homes served and average spending per home for each program year in the full IHWAP program.

Table B2: Average Net Present Benefits by Program Years 2009-2016

Panel A: Evaluated at Social Marginal Costs of Energy

Program Years	Average NPB (US$)	Std. Dev.	Number of Homes
PY 2009	-434.85	1910.31	497
PY 2010	-1021.81	1821.02	1015
PY 2011	-1145.38	1749.61	990
PY 2012	-173.81	1904.09	570
PY 2013	726.95	2111.72	489
PY 2014	736.25	1806.88	438
PY 2015	615.62	1816.55	554
PY 2016	-388.92	1851.94	96
PYs 2009-2012	-809.33	1868.39	3072
PYs 2013-2016	622.50	1928.74	1577

Panel B: Evaluated at Retail Energy Prices

Program Years	Average NPB (US$)	Std. Dev.	Number of Homes
PY 2009	87.08	2281.54	497
PY 2010	-568.88	2204.09	1015
PY 2011	-748.27	2148.88	990
PY 2012	324.80	2291.56	570
PY 2013	1516.03	2475.11	489
PY 2014	1493.54	2181.53	438
PY 2015	1371.05	2154.17	554
PY 2016	228.95	2212.31	96
PYs 2009-2012	-354.75	2255.49	3072
PYs 2013-2016	1380.50	2286.90	1577

Notes: This table presents average home-specific net present benefits by program year. Those were obtained by first estimating home-specific net benefits, as in section ??, and then taking simple averages of those net benefits based on which homes were served in each program year.

C Impacts on Electricity Usage

Table C1: Impacts on on Electricity Usage (MMBtu)

Elec MMBtu	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Weatherization × Treatment	0.00292	-0.00805	-0.0107	-0.0120
	(0.0518)	(0.0519)	(0.0519)	(0.0517)
Panel B: Effect by Treatment Group				
Weatherization × Low Treat	0.0380	0.0271	0.0257	0.0338
	(0.0587)	(0.0585)	(0.0585)	(0.0581)
Weatherization × High Treat	-0.0365	-0.0474	-0.0529	-0.0656
	(0.0703)	(0.0711)	(0.0702)	(0.0713)
House FE	Yes	Yes	Yes	Yes
Month of Sample FE	Yes	Yes	Yes	Yes
Weatherization × Demeaned Pre Blower (CFM)	Yes	Yes	Yes	Yes
Weatherization × Expenditures	No	Yes	Yes	Yes
Weatherization × Month of Completion FE	No	No	Yes	Yes
Weatherization × Characteristics	No	No	No	Yes
No. of Homes	1452	1452	1452	1386
Observations	69334	69334	69334	66245
Adjusted R^2	-0.001	0.002	0.001	0.002
Control Mean Weatherization Effect	-0.471***	-0.471***	-0.471***	-0.471***
	(0.0308)	(0.0308)	(0.0308)	(0.0313)
Control Mean Pre-Weatherization Consumption	2.767	2.767	2.767	2.766

Notes: The dependent variable is monthly electricity consumption (MMBtu). Weatherization indicates consumption observations post-retrofits. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Panel A reports results from regressions pooling High and Low payment treatments into one single treatment indicator. Panel B reports results from regressions with separate indicators for High and Low Treatment. Standard errors are clustered at the house level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table C2: Impacts on on Electricity Usage (MMBtu):Above Average Spending

Elec MMBtu	(1)	(2)	(3)
Weatherization × Treatment	0.0598	0.0380	0.0407
	(0.0731)	(0.0713)	(0.0718)
Weatherization X Treat X Above Average Cost	-0.106	-0.0846	-0.0994
	(0.110)	(0.111)	(0.112)
Weatherization	-0.410***	0	0
	(0.0581)	(.)	(.)
Weatherization X Above Average Cost	0.00956	-0.0101	0.000574
	(0.0691)	(0.0672)	(0.0675)
House FE	Yes	Yes	Yes
Month of Sample FE	Yes	Yes	Yes
Weatherization × Demeaned Pre Blower (CFM)	Yes	Yes	Yes
Weatherization × Month of Completion FE	No	Yes	Yes
Weatherization × Characteristics	No	No	Yes
No. of Homes	1169	1169	1169
Observations	59551	59551	59551
Adjusted R^2	0.003	-0.002	-0.001
Control Mean Weatherization Effect	-0.405***	-0.405***	-0.405***
	(0.0323)	(0.0323)	(0.0323)
Control Mean Pre-Weatherization Consumption	2.759	2.759	2.759

D Robustness Across Samples

The primary estimates of the effects of performance incentives on building envelope tightness (Table 2) use data from all treated homes in the sample. In this Appendix, we test the robustness of our preferred estimates to subsamples for which we also have data on: (1) gas consumption and (2) contractor callback data.

Table D2 reports estimates from the sub-sample of homes that contain a minimum of 12 months of utility billing data on gas consumption, ensuring balance across months of the year. This is the exact same sample that is used to estimate the effects of treatment on household gas consumption. The pooled estimate in our preferred specification (Column 4) is -97.27. The estimated effect of the low bonus is -93.67 and the effect of the high bonus is -101.3 in this subsample. All estimates are statistically different from zero, but none are different from the main estimates reported in Table 2, which are: -88.67 (pooled estimate); -66.17 (low bonus); -113.8 (high bonus).

Table D1 reports estimates from the sub-sample of projects for which we also have contractor callback data provided by quality control inspectors. The pooled estimate in our preferred specification (Column 4) is -118.0. The estimated effect of the low bonus is -86.36 and the effect of the high bonus is -151.7 in this subsample. All estimates are statistically different from zero, but none are different from the main estimates reported in Table 2, which are: -88.67 (pooled estimate); -66.17 (low bonus); -113.8 (high bonus).

Table D1: Effects of Bonus Treatments on Building Envelope Tightness (CFM50)
(Sub-sample with Callback Data)

CFM50 (Post - Pre)	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Treat	-100.8**	-104.3***	-103.0***	-118.0***
	(39.79)	(38.35)	(39.05)	(38.86)
Panel B: Effect by Treatment Group				
Low Treat	-96.07*	-85.05*	-78.78	-86.36*
	(50.39)	(49.25)	(50.31)	(49.53)
High Treat	-105.6**	-123.7***	-128.6***	-151.7***
	(47.60)	(45.86)	(46.25)	(46.17)
Pre Blower (CFM)	Yes	Yes	Yes	Yes
Expenditures	No	Yes	Yes	Yes
Month of Completion FE	No	No	Yes	Yes
Characteristics	No	No	No	Yes
Observations	1226	1225	1225	1164
Adjusted R^2	0.001	-0.043	-0.064	-0.108
Control Group Depvar Mean	-1611.365	-1611.365	-1611.365	-1611.521

Notes: The dependent variable is the change in building envelope tightness as a result of Weatherization Assistance Program upgrades (Post-Pre). Building envelope tightness is measured in units of 50 cubic feet per minute (CFM50), where higher values indicate a leakier home. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Panel A reports results from regressions pooling High and Low payment treatments into one single treatment indicator. Panel B reports results from regressions with separate indicators for High and Low Treatment. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table D2: Effects of Bonus Treatments on Building Envelope Tightness (CFM50) (Subsample with 12+ Months of Gas Consumption Data)

CFM50 (Post - Pre)	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Treat	-67.44*	-86.18***	-99.30***	-97.27***
	(35.31)	(33.31)	(33.94)	(33.84)
Panel B: Effect by Treatment Group				
Low Treat	-84.70**	-94.81**	-97.88**	-93.67**
	(42.16)	(40.17)	(40.14)	(40.85)
High Treat	-48.73	-76.97*	-100.9**	-101.3**
	(43.29)	(41.40)	(42.22)	(41.39)
Pre Blower (CFM)	Yes	Yes	Yes	Yes
Expenditures	No	Yes	Yes	Yes
Month of Completion FE	No	No	Yes	Yes
Characteristics	No	No	No	Yes
Observations	1205	1203	1202	1146
Adjusted R^2	-0.003	-0.046	-0.066	-0.110
Control Group Depvar Mean	-1500.761	-1500.761	-1500.761	-1507.106

Notes: The dependent variable is the change in building envelope tightness as a result of Weatherization Assistance Program upgrades (Post-Pre). Building envelope tightness is measured in units of 50 cubic feet per minute (CFM50), where higher values indicate a leakier home. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Panel A reports results from regressions pooling High and Low payment treatments into one single treatment indicator. Panel B reports results from regressions with separate indicators for High and Low Treatment. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

E Robustness to Potential Cohort Effects

Our main estimates of the effects of treatment on reductions in household gas use (Table 5, use the sample of billing data to projects for which we obtain a minimum of 12 months of pre/post billing data. As a result, it is possible that the sample weights for projects occurring earlier in the study period may be greater than those treated later. In Table E1, we estimate the same regression while constraining the sample of billing data to exactly 12 months of pre/post billing data to ensure that each WAP project gets equal weight in our estimates. In this sample, we find that the pooled effect of performance incentives is -36.2 MMBTU. The effect of the high and low bonus payments are 36.2 MMBTU and -36.3 MMBTU, respectively. All estimates are statistically different from zero, but none are different from the main estimates reported in Table 5, which are: -37.6 (pooled estimate); -35.2 (low bonus); -40.2 (high bonus).

Table E1: Effect on Gas Usage (MMBtu): Exactly 12 Months Pre/Post

Gas MMBtu	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Weatherization × Treatment	-0.242*	-0.328***	-0.356***	-0.362***
	(0.125)	(0.123)	(0.125)	(0.126)
Panel B: Effect by Treatment Group				
Weatherization × Low Treat	-0.301**	-0.364**	-0.372**	-0.362**
	(0.151)	(0.148)	(0.146)	(0.148)
Weatherization × High Treat	-0.177	-0.289*	-0.338**	-0.363**
	(0.160)	(0.158)	(0.163)	(0.168)
House FE	Yes	Yes	Yes	Yes
Month of Sample FE	Yes	Yes	Yes	Yes
Weatherization × Demeaned Pre Blower (CFM)	Yes	Yes	Yes	Yes
Weatherization × Expenditures	No	Yes	Yes	Yes
Weatherization × Month of Completion FE	No	No	Yes	Yes
Weatherization × Characteristics	No	No	No	Yes
No. of Homes	1216	1216	1216	1164
Observations	28416	28416	28416	27211
Adjusted R^2	0.001	0.006	0.004	0.005
Control Mean Weatherization Effect	-1.893***	-1.893***	-1.893***	-1.868***
	(0.151)	(0.151)	(0.151)	(0.153)
Control Mean Pre-Weatherization Consumption	7.498	7.498	7.498	7.498

Notes: The dependent variable is monthly gas consumption (MMBtu). Homes with at least 12 months of both pre and post weatherization gas consumption data were included. Weatherization indicates consumption observations post-retrofits. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Panel A reports results from regressions pooling High and Low payment treatments into one single treatment indicator. Panel B reports results from regressions with separate indicators for High and Low Treatment. Standard errors are clustered at the house level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

F Do Workers Reallocate Effort: Further Evidence

In the main analysis, we provide estimates of the effects of treatment on the callback rate for deficiencies in air leakage retrofits (Table 4) and the effects of completing additional treated/control projects contemporaneously with a given project on the reductions in CFM50 (Table 7) and gas use (Table 8) in that given project. In Table G1, we report estimates of gas use effects broken out by high and low bonus treatments. Like the pooled sample test provided in Table 8, we find no evidence of effects of additional simultaneous treat vs control jobs.

In Tables G2 and G3, we provide an additional test for effects of completing additional treated/control projects contemporaneously with a given project on the *callback rate* associated with deficiencies in air leakage retrofits in that given project. Table G2 reports results from a pooled sample test, while Table G3 reports results for high/low bonus groups. We do not find any statistical effect of additional incentivized contracts on the air leakage callback rate for a given home.

Table G1: Effects on Blower Door: Simultaneous Jobs Within Contractor

CFM50 (Post - Pre)	(1)	(2)	(3)	(4)
Low Treat	-39.03	-60.80	-53.82	-67.24
	(43.28)	(43.33)	(41.50)	(42.94)
High Treat	-60.00	-90.45**	-97.52**	-104.3**
	(47.16)	(46.09)	(45.24)	(45.53)
Low Treat by Simul. Treat Jobs	7.153	10.22*	7.255	7.319
	(5.490)	(5.349)	(5.040)	(5.332)
Low Treat by Simul. Control Jobs	-10.81*	-11.35**	-6.989	-8.394
	(5.895)	(5.605)	(5.326)	(5.637)
High Treat by Simul. Treat Jobs	0.722	2.962	-1.558	-1.328
	(5.609)	(5.517)	(5.327)	(5.309)
High Treat by Simul. Control Jobs	-2.265	-3.021	1.991	0.836
	(6.127)	(5.940)	(5.594)	(5.790)
No. of Simul. Treat Jobs	-4.356	-5.211	-1.754	-0.0575
	(3.769)	(3.868)	(4.407)	(3.972)
No. of Simul. Control Jobs	3.979	3.786	1.050	-1.023
	(4.330)	(4.170)	(4.319)	(4.422)
Pre Blower (CFM)	Yes	Yes	Yes	Yes
Expenditures	No	Yes	Yes	Yes
Month of Completion FE	No	No	Yes	Yes
Characteristics	No	No	No	Yes
Observations	1697	1696	1695	1600
Adjusted R^2	-0.004	-0.038	-0.082	-0.083
Control Group Depvar Mean	-1569.306	-1569.306	-1569.306	-1562.708

Notes: The dependent variable is the change in building envelope tightness as a result of Weatherization Assistance Program upgrades (Post-Pre). Building envelope tightness is measured in units of 50 cubic feet per minute (CFM50), where higher values indicate a leakier home. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. High Treat and Low Treat indicate jobs assigned to high and low treatment respectively. Controls for the demeaned number of simultaneous treatment and control jobs that the contractor worked on are included along with the interaction of each of these controls with the treatment indicator. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table G2: Air Leakage Callback Rate: Simultaneous Jobs Within Contractor

Reducing Air Leakage	(1)	(2)	(3)	(4)
Treat	-0.0314**	-0.0326**	-0.0309*	-0.0222
	(0.0157)	(0.0159)	(0.0158)	(0.0171)
Treat by Simul. Treat Jobs	0.00136	0.000629	0.000513	0.000404
	(0.00283)	(0.00274)	(0.00256)	(0.00290)
Treat by Simul. Control Jobs	-0.00238	-0.00156	-0.00203	-0.00112
	(0.00312)	(0.00296)	(0.00276)	(0.00313)
No. of Simul. Treat Jobs	0.000125	0.000941	0.00217	0.00460
	(0.00232)	(0.00232)	(0.00268)	(0.00282)
No. of Simul. Control Jobs	0.00281	0.00164	0.000565	-0.00306
	(0.00267)	(0.00262)	(0.00282)	(0.00309)
Pre Blower (CFM)	Yes	Yes	Yes	Yes
Expenditures	No	Yes	Yes	Yes
Month of Completion FE	No	No	Yes	Yes
Characteristics	No	No	No	Yes
Observations	1226	1225	1225	1164
Adjusted R^2	0.007	-0.040	-0.099	-0.106
Control Group Depvar Mean	0.085	0.085	0.085	0.084

Notes: The dependent variable is an indicator variable for whether the contractor performing the job was "called back" by the quality control inspector (QCI) to perform additional work to rectify an issue related to air sealing. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. High and Low payment treatments are pooled into one single treatment indicator. Controls for the demeaned number of simultaneous treatment and control jobs that the contractor worked on are included along with the interaction of each of these controls with the treatment indicator. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table G3: Air Leakage Callback Rate: Simultaneous Jobs Within Contractor

Reducing Air Leakage	(1)	(2)	(3)	(4)
Low Treat	-0.0104	-0.0100	-0.0143	-0.00996
	(0.0205)	(0.0207)	(0.0200)	(0.0206)
High Treat	-0.0530***	-0.0562***	-0.0498***	-0.0372*
	(0.0173)	(0.0175)	(0.0173)	(0.0205)
Low Treat by Simul. Treat Jobs	0.00226	0.00168	0.00125	0.00170
	(0.00346)	(0.00335)	(0.00312)	(0.00359)
Low Treat by Simul. Control Jobs	-0.00468	-0.00397	-0.00372	-0.00360
	(0.00382)	(0.00365)	(0.00339)	(0.00389)
High Treat by Simul. Treat Jobs	0.000686	-0.0000143	0.0000128	-0.000655
	(0.00309)	(0.00311)	(0.00307)	(0.00311)
High Treat by Simul. Control Jobs	-0.000248	0.000484	-0.000487	0.00116
	(0.00341)	(0.00341)	(0.00331)	(0.00343)
No. of Simul. Treat Jobs	0.000119	0.000889	0.00201	0.00446
	(0.00232)	(0.00232)	(0.00269)	(0.00283)
No. of Simul. Control Jobs	0.00283	0.00175	0.000709	-0.00287
	(0.00267)	(0.00262)	(0.00283)	(0.00310)
Pre Blower (CFM)	Yes	Yes	Yes	Yes
Expenditures	No	Yes	Yes	Yes
Month of Completion FE	No	No	Yes	Yes
Characteristics	No	No	No	Yes
Observations	1226	1225	1225	1164
Adjusted R^2	0.008	-0.038	-0.100	-0.105
Control Group Depvar Mean	0.085	0.085	0.085	0.084

Notes: The dependent variable is an indicator variable for whether the contractor performing the job was “called back” by the quality control inspector (QCI) to perform additional work to rectify an issue related to air sealing. Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois at the time the work order was printed. The models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. High Treat and Low Treat indicate jobs assigned to high and low treatment respectively. Controls for the demeaned number of simultaneous treatment and control jobs that the contractor worked on are included along with the interaction of each of these controls with the treatment indicator. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.