Incentive-Based Pay and Building Decarbonization: Experimental Evidence from the Weatherization Assistance Program

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

August 29, 2025

Abstract

Building energy efficiency has become a cornerstone of greenhouse gas mitigation strategies, with billions of dollars set aside for extensive upgrades in the coming years. However, impact evaluations have revealed actual energy savings from home upgrade programs often fall short of projections, in part due to contractor underperformance. Using field experiment results, we show refining one program design element—offering performance bonuses to contractors–increased natural gas savings by 24% and generated \$5.39-\$14.53 in social benefits per dollar invested. Hence, changes to worker incentives can have sizable impacts on the cost effectiveness of GhG abatement in energy efficiency programs.

Key words: Climate Policy, Returns to Government Spending, Energy Efficiency, Performance Based Pay, Effort Allocation
JEL Classification: J33, Q48, Q54, O93

---

*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, Hansen Shao, and Mateus Souza for excellent research assistance. We received excellent feedback and comments from Hunt Alcott, Judson Boomhower, Meredith Fowlie, Andy Garin, Nathan Hendren, Matt Kahn, Kelsey Jack, Jean-William Laliberte, Rob Metcalfe, Dave Rapson, Rich Sweeney and seminar participants at the CU Environmental and Resource Economics Workshop, the Dallas Fed, UC Berkeley's Electricity Camp, Georgia Institute of Technology, the Kelley School of Business, Michigan State University, the Online Agricultural and Resource Economics Seminar (OARES), UC Irvine, UC Santa Barbara, UC San Diego, UT-Austin, the Heartland Workshop, the University of Pennsylvania, and the Urban Economics Association meeting. 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. This research has been approved by the University of Illinois at Urbana-Champaign Institutional Review Board (IRB), under protocol number 17672.

1 Introduction

Energy use in buildings accounts for 26% of global greenhouse gas emissions (IEA, 2023), making energy efficiency and electrification key climate mitigation strategies. In response, the U.S. has recently committed $13 billion to residential retrofits alone[1]. Beyond climate benefits, home retrofit programs often have other goals, such as lowering energy costs for low-income families. While retrofitting can deliver cost-effective improvements, realized savings from home energy efficiency upgrades have frequently fallen short of projections due to modeling bias and contractor underperformance (Christensen et al., 2023; Blonz, 2023; Fowlie, Greenstone, and Wolfram, 2018).[2] Improving program design to address these shortfalls could significantly enhance the effectiveness of energy efficiency programs as a climate mitigation strategy.

Using results from a randomized controlled trial within Illinois's implementation of the U.S.'s largest energy efficiency program, this paper shows that refining contractor incentives, a crucial element of program design, can significantly improve the cost-effectiveness of whole-home energy efficiency upgrades. We find that performance incentives for air sealing, a critical component of these upgrades, increase natural gas savings by 24%, generating $5.39–14.53 in social net benefits per dollar of investment. These findings demonstrate a low-cost, easily adoptable strategy for improving the cost-effectiveness of energy efficiency programs and offer broader insights for public program design. In particular, they show that adjusting within-program funding allocations can have first-order effects on public spending efficiency. Thus, while evaluating existing public programs remains crucial for understanding budgetary trade-offs, experimenting with mechanisms to correct misallocation within programs may be equally important.[3]

We partnered with the Illinois Home Weatherization Assistance Program (IHWAP) to offer novel evidence on the effects of piece-rate incentives for air sealing retrofits. WAP funding is allocated to states based on their low-income population and climate. As a result, Illinois, along with four other states (New York, Michigan, Pennsylvania, Ohio) account for

[1] The IRA allocated nearly $9 billion to states and Tribes for consumer home energy rebate programs and an additional $837.5 million to the Department of Housing and Urban Development for energy efficiency and building electrification. The Infrastructure Investment and Jobs Act set aside $3.5 billion to expand the Weatherization Assistance Program (*Infrastructure Investment and Jobs Act* 2021).

[2] 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, 2017; Davis, Fuchs, and Gertler, 2014), and efficient new construction (Levinson, 2016; Bruegge, Deryugina, and Myers, 2019; Davis, Martinez, and Taboada, 2020).

[3] For example, Gillingham and Stock (2018) list the costs of delivering a ton of carbon abatement across various government programs; changes of even 10% could substantially alter the ranking of which programs to prioritize.

roughly a third of program funding.[4] The primary goal of the program is to reduce costs for low-income families. However, since the aim is to target cost-effective upgrades, the results are highly relevant for climate change considerations. Air sealing upgrades are one of the "big four" energy-saving retrofits performed in whole-house upgrade programs. In IHWAP, they represent just over 12% of total expenditures in the average home.[5]

Air sealing retrofits provide a promising starting point for introducing incentive-based pay into energy efficiency programs because they are straightforward to evaluate using a blower door test measuring the "leakiness" of a home. The blower door test, a crucial component of energy auditing, is widely utilized across federal, state, and utility programs, as well as by private contractors. The CFM50 levels measured by a blower door test are highly correlated with energy consumption.[6] Air sealing opportunities 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 marginal cost of additional air sealing improvements. While we study low-income housing, air sealing opportunities exist in homes across the income spectrum, particularly those older than 20 years (Papineau, Rivers, and Yassin, 2023). Therefore, piece rate payments for blower door readings could readily be adopted in other government and utility programs or as part of a private contract.

In this study, we randomly assigned contracts (jobs) into two treatment groups that correspond to "high" ($1.00) or "low" ($0.40) payments for each unit of air sealing achieved beyond target. Contracts assigned to the control group were not eligible for bonus incentives.[7] We use the high and low treatment levels to examine how effort responds to different incentive levels. Aside from the presence or absence of bonus payments for air sealing, all other components of the program continued as before, including the enforcement of minimum quality standards on all retrofits. A certified quality control inspector in each agency 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.

[4] See `https://www.energy.gov/sites/default/files/2023-06/WPN%2018-2%20-%20Archived.pdf`, accessed May 27, 2024.

[5] Wall insulation, attic insulation and furnace replacements are the other 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]) and is measured in cubic feet per minute at 50 Pa (CFM50). The leakier the home, the more air needs to be moved to cause that amount of depressurization.

[7] Job-level randomization offered substantial gains in statistical power relative to contractor-level.

We report three key findings on the overall effectiveness of air sealing performance incentives in the IHWAP. Performance bonuses lead to: (1) increases in the homes' air-tightness, (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. Because randomization occurs at the job level and jobs are arbitrarily assigned to contractors through a queue, our estimates isolate within-worker productivity effects of performance incentives (Shearer, 2004; Friebel et al., 2017; Guiteras and Jack, 2018) and exclude potential sorting/selection processes that can also shape the impacts of performance pay interventions (Lemieux, MacLeod, and Parent, 2009; Dohmen and Falk, 2011; Rothstein, 2015; Ashraf et al., 2020).

We also examine whether performance incentives cause workers to shift effort from non-compensated to compensated tasks in multi-dimensional contracts (e.g., Paarsch and Shearer, 2000; Dumont et al., 2008; Mullen, Frank, and Rosenthal, 2010; Feng Lu, 2012; Johnson, Reiley, and Muñoz, 2015; Hong et al., 2018; Aucejo, Romano, and Taylor, 2022). In their seminal work on incentive contracts in multitask settings, Holmstrom and Milgrom (1991) show that the effect of piece-rate bonuses on effort reallocation is theoretically ambiguous, depending on whether tasks are complements, substitutes, or independent in the contractor's private cost function. Our research design and data from inspection reports, blower door readings, and energy bills allows us to directly measure changes in effort on non-compensated tasks and to provide the following evidence on reallocation in an important policy context. First, we find no evidence that incentive treatments increase deficiencies in non-compensated retrofits (e.g., furnace repairs, which do not affect CFM50 reductions) within treated contracts. Second, we control for and quantify the effects of effort reallocation between treatment and control jobs by exploiting random variation in the number of simultaneous treatment or control jobs assigned to contractors. We find that incentive-based contracts on some jobs do not reduce performance on contemporaneous jobs. These findings have important implications for WAP and other energy efficiency programs aiming to improve cost-effectiveness, suggesting that, with minimum quality standards in place, incentivizing air sealing can yield substantial benefits without compromising performance on other aspects of the job.

Finally, we find that contractors who initially perform at a high level respond more strongly to the incentives, implying that the program's cost-effectiveness improves through more efficient allocation of effort among workers with varying marginal effort costs. While performance pay is often thought of as a correction for moral hazard, it can also improve effort allocation. In many settings, estimating the value added by individual workers is complicated by negative assortative matching (e.g. between managers and stores or workers in hospitals) (Metcalfe, Sollaci, and Syverson, 2023; Kosfeld and Von Siemens, 2011). However, leveraging the queue-assignment feature of the IHWAP, we are able to identify high-quality contractors at baseline by evaluating their individual performance during the program year *prior to* the performance-pay intervention. Consistent with results from canonical models of performance pay (Lazear, 2018) and emerging evidence in other empirical settings (Frederiksen, Hansen, and Manchester, 2022; Franceschelli, Galiani, and Gulmez, 2010), we find that those who performed better at baseline respond to incentives with disproportionately larger improvements in their air sealing work. We also find a higher variance of CFM50 and gas use outcomes with the piece rate, which is consistent with the prediction that contractors will invest additional effort on lower marginal cost homes under that regime. These findings demonstrate that cost-effectiveness increases under the bonus result from more efficient allocation of effort within the program, thus indicating that minimum air sealing standards are a significant source of effort misallocation. They also suggest the potential for larger social returns from interventions that target incentives to high performing contractors.

This study also contributes to a growing body of empirical evidence highlighting the potential for incentive pay to improve effort and effectiveness of public sector workers, including teachers (e.g., Lavy, 2009; Muralidharan and Sundararaman, 2011; Duflo, Hanna, and Ryan, 2012; Lavy, 2020; Leaver et al., 2021), tax collectors (Khan, Khwaja, and Olken, 2019; Khan, Khwaja, and Olken, 2016), and civil servants (Burgess et al., 2017; Bandiera et al., 2021; Bertrand et al., 2020; Kim, Kim, and Kim, 2020; Luo et al., 2020). The current paper expands the public sector literature by examining the role of performance pay in the context of government contractors. Even modest improvements in the returns to social spending from government service contracts could have sizeable budgetary impacts in places like the U.S., where $188 billion was spent contracting services through its civilian agencies in 2020 (U.S. Government Accountability Office, 2021). Unlike teachers and other bureaucrats whose counterfactual compensation is a salary or wage with a fixed number of hours, contractors already have incentives to complete jobs efficiently to minimize costs. However, government contractors are often incentivized based on outputs—such as tasks completed or recipients reached—that may not translate directly into welfare gains for clients (Hawkins, Bieretz, and Brown, 2019).

In addition, we contribute to a burgeoning literature that seeks to quantify returns to spending from public programs to inform policymakers' budgetary decisions (Finkelstein and Hendren, 2020; Hendren and Sprung-Keyser, 2020). Being able to rank programs according to their returns to social welfare is particularly relevant for identifying the most potent and scalable strategies for climate policy (Gillingham and Stock, 2018; Hahn et al., 2024). The WAP setting is well suited for evaluating the social net benefits of performance pay because we can directly observe and quantify impacts on the program's welfare-relevant outcome: energy consumption. Further, we are able to develop a framework that quantifies the impacts of incentive-based contracts on both the net benefits and the marginal value of spending in a major government program. Using the high and low treatment levels, we construct a measure of producer surplus to identify two points on the supply curve of air sealing (as a function of the piece rate) and identify the optimal level of bonus payment. Our results indicate that the optimal bonus in the IHWAP program is likely in the neighborhood of and is no higher than \$0.40 per CFM50 reduction.

Finally, we contribute to a literature that examines the effects of "plumbing"— experimenting with the mechanics of policies implemented in the real-world to improve their outcomes—as a path to enhancing government programs with economics research (Duflo, 2017; Duflo, 2020). We show that the marginal value of public funds (MVPF) or the cost per ton of carbon abated are not a fixed attribute of residential energy efficiency programs, as is often implicitly assumed in the literature. On the contrary, design interventions can result in very large changes relative to the baseline MVPF of decarbonization programs. The pay-for-performance intervention that we examine increased the MVPF spent on incentivized retrofits (those that directly affect air sealing) by more than 20%.[8] This suggests that evidence-based refinements to the design of decarbonization programs could play an important role in achieving meaningful reductions in the cost of GhG abatement in the coming decades. More broadly, our findings illustrate how incentive-based contracts can be used as a design tool for addressing the misallocation of effort in government programs, allowing agencies to target contractor effort to welfare-relevant outcomes and to calibrate payments based on observed responsiveness to incentives.

2 Setting and Experimental Design

2.1 Institutional Background

The Weatherization Assistance Program (WAP) is the largest U.S. residential weatherization program, having provided 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, 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, the IHWAP. Implementing agencies throughout the state are responsible for ad-

[8] These include air sealing, foundation, duct repair, and doors and windows, which taken together, comprise around a fifth of total program spending.

ministering 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. Department of Energy (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 IHWAP 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 (Department of Health and Human Services) funds, state funds, and utility funds has been used to do additional retrofits 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 rectifying health and safety concerns that fall outside WAP guidelines. 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 been 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. Figure 2 shows the the CFM50 levels measured by a blower door test are highly correlated with energy consumption. Air sealing is always deemed cost-effective given the low-cost and long-lived energy savings achieved when finding and sealing leakage sites in attic, walls, basements or crawl spaces. This type of work directly affects the tightness of the building envelope. Other retrofit categories can impact building envelope tightness as well, including some types of attic insulation, wall insulation, crawl space insulation, basement insulation, rim insulation and windows and door upgrades. The tightness of the building envelope is measured by a blower door test. To implement the test, an experienced energy specialist will temporarily install a large fan in the frame of a home's outside doorway. After calibrating the device, the fan draws air out of the house and reduces the air pressure inside. The test then 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 (CFM50 reductions) for each individual home.

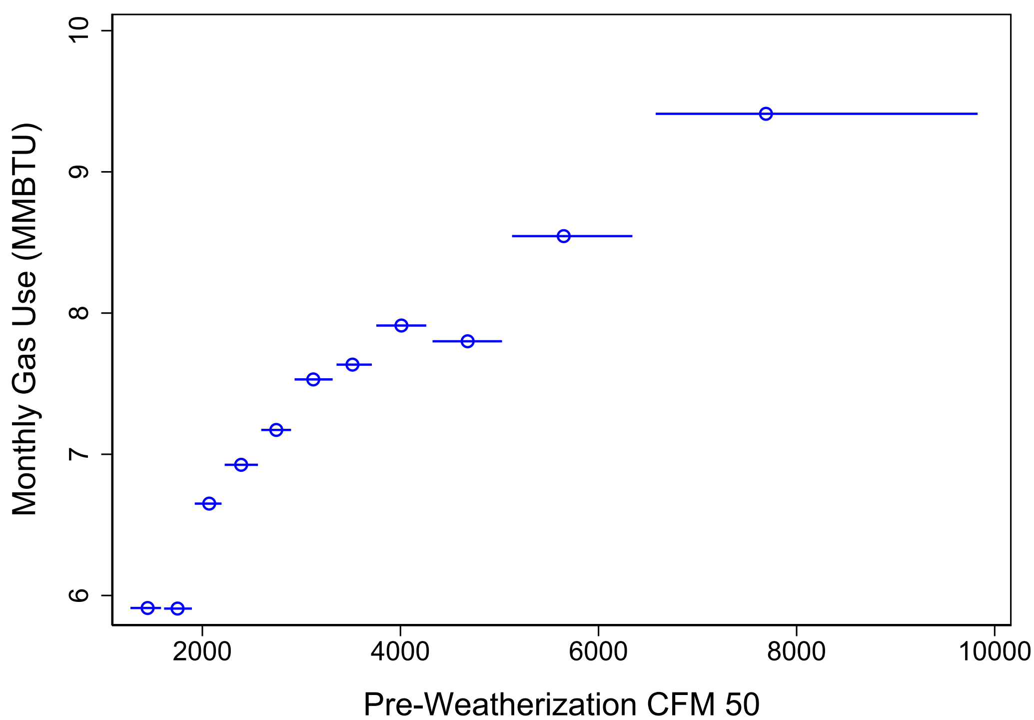


Figure 1: Correlation Between Energy Consumption and CFM Pre-Weatherization

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 (Cattaneo et al., 2019a; Cattaneo et al., 2019b). For the purposes of scaling, the mean of each variable is added back to the residuals.

The set of firms, or "contractors," that 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. These bids determine the compensation that the selected contractors receive for work performed in that program year. As a result, payments are pre-determined in each contract 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 has been approved and the auditor has selected the retrofits to be implemented. Once in the queue, most jobs are assigned in sequential order independent of any characteristics about the home, measures assigned, or the contractors themselves.[9] Contractors frequently receive a bundle of work orders at once.

The WAP enforces quality standards through inspections performed when all work is completed.[10] The agency inspector documents findings and deficiencies for any measures that did not meet the quality standard associated with a given type of retrofit. Certain findings will warrant a callback for the contractor to return to the site to correct a deficiency, such as poor wall insulation or a furnace installation issue. Others will not, if for instance, air sealing does not quite hit target or duct work was not ideally performed.[11] Callbacks are costly since firms do not receive additional compensation for the extra work.

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.[12] 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 in the U.S. that we are aware of that already compensates their contractors for air sealing based on measured air-tightness outcomes (performance pay contracts). 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 major 3 utilities that serve the study region, with whom we had a data sharing agreement.

The intervention consisted of two different piece-rate bonus regimes: a "high" payment

[9] 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.

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

[11] 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.

[12] A program year begins in July and ends in June of the following year.

($1.00) and a "low" payment ($.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.[13] The average bonus paid during our intervention is $114 for the low payment and $283 for the high payment, which account for a respective 10% and 25% of the cost of air sealing retrofits or 1% and 3% of the total costs (labor plus materials) for the average home in our sample. 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 can be certain that all contractors who registered as vendors were aware of the bonus program. We also made contractors 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. After new work orders were initiated in IHWAP's administrative software, they were automatically assigned to the "high" payment treatment, "low" payment treatment, or control regimes. Randomization was implemented through a 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:

"Architectural work on 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 occurs when the work order is printed, which happens after the pre-weatherization audit is completed. As a result, treatment status cannot influence blower door targets or retrofits, as the energy auditor does not know the treatment status when recording initial measurements. Second, contract pay for each retrofit task is specified and 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, jobs are assigned through a queue system, which limits the potential

[13] 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 callback. However, the adjustment was made due to budgetary considerations.

for productivity sorting of firms and allows us to isolate the impact of piece-rate incentives on worker output with a constant pool of human capital (Dohmen and Falk, 2011).

Finally, the use of braided funding during the program years studied here allowed many measures to be 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 Online Appendix Table A1). This has implications for interpreting cost-effectiveness of the WAP. Because it is not possible to disentangle the impacts of DOE versus other funding sources in this period, the cost effectiveness we report for the overall program (Panel B of Table 10) is substantially lower than those reported in recent years without braided funding (e.g., Christensen et al., 2023).

3 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 administering 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 observable job characteristics for each of the experimental 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 3,600-3,900. The baseline CFM50 levels can range up to 10,000 in some homes. For the study, we exclude homes with baseline CFM50 levels that exceed 10,000 due to concern about measurement error. While homes may legitimately have CFM50 levels above this threshold, they would have to have large discrete leaks such as broken windows, which make them unlike the rest of the sample.

A stratified randomization process was not feasible in our context and a few characteristics are statistically imbalanced across treatment arms due to finite-sample variation. Since treatment assignment was random and attrition was negligible (less than 1% of jobs with work orders were not completed), these imbalances do not threaten the internal validity of the study (Bruhn and McKenzie, 2009). By design, any unobservable job characteristics uncorrelated with observables are, in expectation, orthogonal to treatment assignment. As discussed in the next section, we include all of the observable job characteristics in Table 1 in our analysis to improve the precision of the estimated treatment effects.[14]

4 Results

4.1 Contractor Response to Performance Incentives

We begin by investigating the effects of the bonus treatment on the contracted outcome, building envelope tightness. Given that there is a minimum quality standard for air sealing with a minimum guaranteed payment, the introduction of a piece rate will not reduce effort, and, to the extent that some workers respond to the incentive, average effort and output will increase (Lazear, 2000). In our setting, workers should respond with increases in CFM50 reductions (beyond the minimum standard) whenever the marginal costs of effort and materials required are lower than the piece-rate bonus.[15]

We test whether the bonus treatment increases output (CFM50 reductions) and effort, as proxied by fewer deficiencies on incentived measures identified as part of a post-weatherization quality control control inspection. Air sealing is performed with the sole goal of improving building envelope tightness, and thus is directly incentivized by the bonus, though other retrofit categories, including windows, doors, and insulation of all types (e.g. attic, wall, foundation) can also improve blower door readings. We estimate treatment effects using the following model:

[14] Whether baseline differences between treatment groups matter for estimation is determined by the predictive power of each covariate as well as any potential correlation with treatment assignment. Even covariates with small, statistically insignificant differences across groups can meaningfully improve precision if they are strongly predictive of outcomes, while covariates with large, statistically significant imbalances may have little effect if they are weak predictors (Altman, 1985).

[15] See Appendix B, for a derivation of this result using principal-agent model of effort allocation in a multi-task setting that integrates insights from both Holmstrom and Milgrom (1991) and Lazear (2000).

Table 1: Summary Statistics and Balance

	Control ITT (1)	Low ITT (2)	High ITT (3)	T-test Difference (1)-(2)	(1)-(3)	(2)-(3)
Stories	1.328	1.372	1.324	-0.044	0.004	0.048
	(0.015)	(0.024)	(0.023)			
Square Feet	1462.272	1451.030	1431.970	11.242	30.302	19.060
	(22.075)	(27.997)	(32.515)			
Occupants	2.346	2.330	2.403	0.016	-0.057	-0.073
	(0.057)	(0.079)	(0.087)			
Bedrooms	2.817	2.818	2.847	-0.000	-0.029	-0.029
	(0.033)	(0.043)	(0.044)			
Year Built	1951.645	1948.329	1951.470	3.316**	0.175	-3.142*
	(0.952)	(1.325)	(1.341)			
Pre Blower Door (CFM)	3573.425	3856.638	3608.707	-283.213**	-35.282	247.931*
	(61.401)	(95.340)	(92.371)			
General Expenditure	173.836	186.380	150.474	-12.544	23.363	35.907
	(18.122)	(25.696)	(23.693)			
Furnace Expenditure	2366.780	2331.202	2291.773	35.578	75.007	39.429
	(54.580)	(88.130)	(78.090)			
Foundation Expenditure	745.119	847.921	756.534	-102.803	-11.416	91.387
	(31.359)	(54.548)	(47.592)			
Door Expenditure	104.689	108.946	115.340	-4.256	-10.651	-6.395
	(10.322)	(15.865)	(15.996)			
Baseload Expenditure	528.509	572.825	512.925	-44.316	15.584	59.900*
	(17.398)	(24.434)	(23.745)			
Attic Expenditure	1335.578	1320.355	1314.904	15.223	20.674	5.451
	(33.564)	(48.147)	(48.848)			
Air Conditioning Expenditure	1976.039	1747.748	1845.919	228.291***	130.120	-98.171
	(46.419)	(67.644)	(68.504)			
Water Heater Expenditure	929.627	943.229	952.844	-13.603	-23.217	-9.615
	(34.489)	(50.647)	(52.124)			
Wall Insulation Expenditure	413.227	427.287	314.356	-14.060	98.871**	112.931*
	(29.534)	(46.996)	(36.524)			
Window Expenditure	72.752	57.380	62.584	15.372	10.169	-5.203
	(11.307)	(16.820)	(19.491)			
Number of observations	821	406	372	1227	1193	778

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.

$$Y_i = \beta_0 + \beta_1 T_{Hi} + \beta_2 T_{Li} + \mathbf{X_i'}\beta + \varepsilon_i, \quad (1)$$

where Y_i is an experimental outcome (CFM50 reduction or callback indicator), 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 $\mathbf{X}_i$ contains a vector of controls and ε_i is the idiosyncratic error term. We include month-year fixed effects for job completion month, and flexibly control for pre-treatment covariates using indicators for binned values of retrofit spending and home characteristics variables in Table 1 to improve the precision of the estimated treatment effects. Following Abadie et al. (2023), we use heteroskedasticity-robust standard errors.[16]

Contractors were only eligible to receive payments for treatment jobs if they were signed up as vendors with the University of Illinois and a small number of contractors did not sign up for the bonus program. Given the possibility that take-up was correlated with unobservable skills that affect 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. This is also the policy-relevant result because if this bonus were rolled out program-wide, all contractors would receive payments from the implementing agencies directly rather than through a third party vendor. See Online Appendix C for tables of corresponding reduced form estimates.[17]

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 baseline outcome, pre-weatherization blower door reading. 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 specifications. The most precisely estimated coefficient in column 4 indicates that the bonus payment regime results in an average reduction of 89 CFM50 beyond the 1563 control mean reduction achieved using the pre-set compensation at the level of the target (standard). The magnitude of this effect is equivalent to 5.7% of the CFM50 reduction achieved in the control group (dependent variable mean) and approximately 2.5% of the pre-weatherization CFM50 level. 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.

[16] Because treatment is randomly assigned at the job level, it is independent of contractor-level unobservables. As a result, any within-contractor correlation in outcomes does not compromise the validity of inference. Online Appendix Tables D4 and D3 show that clustering at the contractor level yields the same level of statistical significance for the treatment effects on key outcomes: CFM50 reduction and gas use.

[17] These intent-to-treat estimates are qualitatively quite similar–usually within 90% in magnitude–to those reported in the main text, because almost all contractors signed up as vendors.

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
Control Pre-Weatherization Blower Mean	3609.7	3609.7	3609.7	3585.2
Control Group Dep. Variable Mean	-1569.306	-1569.306	-1569.306	-1562.708

Notes: The dependent variable is the change in building envelope tightness (CFM 50) from WAP upgrades (Post-Pre). 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. Models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Effects on Air-Sealing Callbacks

Table 3 reports estimates of average treatment effects on deficiencies in air sealing work documented in post-weatherization quality control inspections. Callback data collected from agencies administering the program were available for nearly all (1670/1698) jobs in our sample.[18] Panel A reports estimates for the pooled treatment groups. The Column 4 estimate indicates that performance incentives reduced the probability of a deficiency by 2.95 percentage points, a reduction of just over a third of the 7.8% control mean callback rate. Point estimates in Panel B suggest that effects on callback rates may result primarily from responses to the higher bonus payment, though the differences between the groups are not statistically significant in our sample. These results are consistent with increased effort on air sealing in response to the bonus payments.

We performed a similar analysis for the effect of treatment on callbacks for deficiencies among all incentivized tasks combined (Appendix Table E1) and the point estimates are similar, though not as precise, to those for air sealing alone.[19]

4.2 Effects on Energy Consumption

Energy savings is the welfare-relevant outcome of interest in the program. 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:

$$Y_{it} = \beta_0 + \beta_1 W_{it} * T_{Hit} + \beta_2 W_{it} * T_{Lit} + \beta_3 W_{it} + W_{it} \cdot \mathbf{X_i'}\beta + \delta_t + \gamma_i + \varepsilon_{it} \tag{2}$$

where Y_{it} is the energy use (MMBtu) for household i in month t, $W_{it} * T_{Hit}$ is an indicator for the post-weatherization condition in the high treatment group, $W_{it} * T_{Lit}$ 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 unobservable factors about a home that affect consumption. To improve precision, 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

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

[19] Incentivized tasks include any retrofit categories that are done with the express purpose of reducing air leakage and include: thermal boundary, air sealing, windows, and doors. We designate retrofit categories that do not affect the blower door reading as non-incentivized, including non-insulation, mechanical system standards such as "Combustion Efficiency Venting," "Heating System Replacement," "Water Heater Retrofits," or "Gas Ovens".

Table 3: Effects of Bonus Treatments on Callback Rate: Air Leakage

Air Leakage Callback	(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
Control Group Dep. Variable Mean	0.078	0.078	0.078	0.078

Notes: The dependent variable indicates an air sealing callback. 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. Models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

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.[20] An idiosyncratic error term is represented by ε_{it} As before, we estimate the model using 2 stage least squares estimator where the randomized treatment assignment instruments for treatment jobs with eligible contractors.[21] We cluster standard errors at the household level

[20]Given that there is continuous random assignment to treatment throughout the sample period, our estimates will not suffer from the near term bias that can be present with staggered rollout (e.g., Goodman-Bacon, 2021). There are also unlikely to be significant differences in treatment effects across cohorts in the sampling period biasing our effects. We demonstrate in Online Appendix F that the treatment effects are quite similar when we reduce our sample size to ensure each WAP project is weighted equally.

[21]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 Online Appendix Table D2, which are consistent in magnitude with those presented in Table 2.

(Abadie et al., 2023).[22]

Table 4 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 and 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), indicators for month-year of project completion (Column 3), and household and property characteristics (Column 4). Estimates are consistent across specifications and indicate a pooled treatment effect of 0.29-0.38 MMBtu. Estimates with the most comprehensive controls indicate that the bonus reduces monthly gas consumption by 5.5% relative to the control mean of 6.88 MMBtu. Compared to the baseline weatherization effect observed in the control group (-1.597 MMBtu), the effects of the bonus are substantial–increasing the overall impact by 20-25%.[23] There are several plausible explanations for why the effect of the bonus treatment is proportionally larger compared to the baseline weatherization effect for energy than for CFM50. For example, some retrofits, such as insulation, reduce air leakage, but reduce energy consumption proportionally more. Or, contractors may have responded to the incentive with increased supervision of employees' air sealing work, which could have made it easier to increase supervision on other architectural aspects of the job, such as insulation.

Air sealing and other retrofits also increase efficiency of electricity usage for space cooling. We report estimates of average treatment effects on monthly electricity consumption in Online Appendix Table G1. The estimated treatment effects of our intervention are statistically non-significant, largely due to the relatively smaller mean effect of weatherization on electricity consumption in control homes, which is -0.47 MMbtu—about 30% of the effect observed for gas. However, we cannot rule out an effect size similar to that observed for gas, which shows a 23% increase in the effect of the Weatherization Assistance Program (WAP). To be conservative in our estimates of social benefits, we assume that the intervention did not affect electricity usage and limit our benefits estimates to those from gas reductions.

[22] This accounts for correlation in the error term within a household due to unobserved household-level characteristics and serial dependence in consumption. As with the blower door and callback outcomes, any within-contractor correlation does not affect the validity of inference because random assignment breaks the link between treatment and contractor-level unobservables or shocks. Online Appendix Table D5 demonstrates that inference regarding treatment effects on natural gas consumption is robust to clustering at the contractor level, with no change in statistical significance.

[23] The treatment effect size for both the blower door reading and gas consumption appear to have remained consistent over the trial period and do not change as a function of experience with the bonus (see Appendix Tables I1 and I2).

Table 4: Effects of Bonus Treatments 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**	-0.353***	-0.345**	-0.353***
	(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
Baseline Weatherization Reduction	-1.582***	-1.582***	-1.582***	-1.623***
	(0.101)	(0.101)	(0.101)	(0.0982)
Control Mean Pre-Weatherization Consumption	7.283	7.283	7.283	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. Standard errors are clustered at the house level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

4.3 Do Workers Re-allocate Effort?

In our setting, contractors complete multiple retrofit tasks for a contract, such that increasing effort on one type of retrofit task may impact effort allocated to other retrofits. Changes in effort and output in non-incentivized tasks in response to the introduction of an incentive depends on whether inputs to those tasks are complements or substitutes in the worker's private cost function (Holmstrom and Milgrom, 1991). In Appendix B, we discuss a basic model that derives these predictions for our setting. If non-incentivized retrofits are complementary to air sealing, the introduction of the bonus could increase effort and output in the non-incentivized dimension. If they are substitutes, then the bonus could reduce effort and output in the non-incentivized dimension.

It is not obvious ex ante whether tasks in a weatherization contract are characterized by complementarity or substitution. On the one hand, improvements in wall and attic insulation can increase returns to effort on air sealing. Improved supervision of a crew's air sealing work may also reduce the cost of supervising other aspects of the job. To the extent these kinds of complementarities exist, a piece-rate bonus on the blower door reading could increase effort not only on air sealing, but on other tasks. However, to the extent that contractors face constraints in their capacity to adjust the size of their crew, or time spent on each job, an air sealing incentive may result in re-allocation from substitute tasks. Or it could be that certain retrofit tasks may not act as complements or substitutes in a contractor's cost function, such that an incentive payment on one type of retrofit would not affect effort or output on other retrofits. The evidence of proportionally larger net reductions in gas consumption than blower door readings reported in the prior section are consistent with complementarities and suggest a limited role for substitution from non-incentivized to incentivized tasks in the WAP. However, our experimental design and data collection allows for a more comprehensive analysis of contractor substitution than is often possible in empirical work on incentive pay in a multi-task setting. We are able to directly test for the impacts of reallocation from: 1) control jobs and 2) treatment job tasks that do not affect the blower door reading.

Between-Job Reallocation Effects

We begin by considering reallocation of effort from control jobs that are completed simultaneously with a treated contract. We take advantage of the fact that batches of several jobs are typically assigned to a contractor at a time through a queue. Random assignment guarantees that for a given project i in month t the ratio of work orders in treatment versus in control assigned simultaneously will be exogenous. We define jobs as occurring simultaneously if there is any overlap between their completion windows–the period between the pre-weatherization audit and completion dates.

To test for reallocation between jobs, we include controls for the numbers of simultaneously treated and control projects completed by the same contractor in our main regression specifications:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 T_i \times Simultaneous\,Treat\,Jobs_i + \beta_3 T_i \times Simultaneous\,Control\,Jobs_i + \beta_4 Simultaneous\,Treat\,Jobs_i + \beta_5 Simultaneous\,Control\,Jobs_i + \mathbf{X_i'}\beta + \epsilon_i, \quad (3)$$

where Y_i is the CFM50 reduction for project i, $Simultaneous\,Treat\,Jobs$ is the number of treated jobs (demeaned) simultaneously assigned with project i, *Simulataneous Control Jobs* is the number control jobs (demeaned) simultaneously assigned with project i and ε_i is the idiosyncratic error term. As in our main estimation, the vector $\mathbf{X}_i$ contains indicators for: 1) month-year of completion, 2) indicators for binned values of pre-treatment blower door measurement, 3) indicators for binned values of retrofit spending levels, and 4) indicators for binned home characteristics described in Table 1. We use the pooled version of treatment in the main text for ease of interpretation, given the large number of interaction terms. Online Appendix Table H1 shows qualitatively similar results with interactions with high and low treatments separately.

Since we define simultaneous jobs using 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 in the sample. The coefficients β_2 and β_3 estimate the differential effect of simultaneous treatment or control jobs on CFM50 reductions for treated jobs. The coefficients β_4 and β_5 estimate the effect of an additional simultaneous treatment or control job on CFM50 reductions for all jobs. If the coefficient on treatment (β_1) does not change across specifications that include or omit these controls, it indicates that spillovers to simultaneous jobs are not a significant biasing factor in the primary estimates reported in Table 2.

Table 5 reports estimates of the impacts of simultaneous jobs on CFM50 reductions. While we cannot rule out small spillovers in either direction given the point estimates and standard errors, we find no evidence that a contractor working on a treatment or a control job achieves a statistically different CFM50 reduction when the job is completed simultaneously with another control or treatment home. The point estimate of the treatment coefficient (β_1) with these controls is also consistent with estimates in Table 2, indicating that reallocation does not affect the magnitude of our treatment effects.

Table 6 reports analogous estimates of the impacts of simultaneous jobs on gas consumption. We estimate equation 2 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. Rather than additional treatment jobs inducing a substitution of effort from simultaneous jobs, we find some evidence of *larger* pre/post weatherization differences in control jobs when completed alongside additional simultaneous control jobs and *smaller* pre/post weatherization differences in control jobs when completed alongside additional simultaneous treatment jobs. However, these differences are small in magnitude (less than 10% of the pooled treatment effect) and given the random assignment of contract incentives, the expected impact of these equal and opposite effects on simultanously completed jobs equates to a statistical zero. We do not find any evidence of simultaneous jobs deferentially affecting

Table 5: Effects on Building Envelope Tightness (CFM 50): Simultaneous Jobs Within Contractor

CFM50 (Post - Pre)	(1)	(2)	(3)	(4)
Treat	-64.41*	-80.14**	-79.34**	-88.72***
	(33.38)	(32.39)	(33.01)	(32.14)
Treat × Simultaneous Treat Jobs	4.242	6.925	6.278	3.515
	(4.713)	(4.618)	(4.620)	(4.520)
Treat × Simultaneous Control Jobs	-6.822	-7.450	-7.230	-4.196
	(5.111)	(4.872)	(4.850)	(4.819)
Simultaneous Treat Jobs	-4.356	-5.187	-3.440	0.0841
	(3.765)	(3.864)	(4.210)	(3.951)
Simultaneous Control Jobs	3.973	3.704	1.812	-1.229
	(4.326)	(4.165)	(4.576)	(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	1696	1600
Control Pre-Weatherization Blower Mean	3609.7	3609.7	3609.7	3585.2
Control Group Dep. Variable Mean	-1569.306	-1569.306	-1569.306	-1562.708

Notes: The dependent variable is the change in building envelope tightness (CFM 50) from WAP upgrades (Post-Pre). 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. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

treatment jobs. As with the blower door outcome, the estimated effects of treatment on gas consumption (β_1) are consistent with those in Table 4, indicating that spillover effects are not a significant biasing factor in our primary analysis.[24]. Taken together, these results suggest that the presence of treatment jobs does not appear to create complementarities with or substitution from control jobs.

Further, these patterns do not change as a function of experience with the bonus. Online Appendix Tables I1 and I2, show that neither treatment nor control outcomes change in a statistically or economically significant way with the number of bonus jobs assigned,

[24] Similarly, we find no evidence of simultaneous treatment or control jobs affecting energy consumption for high and low treatments separately nor affecting air sealing callbacks (Online Appendix H)

Table 6: Effects on Gas Usage (MMBtu): Simultaneous Jobs Within Contractor

Gas MMBtu	(1)	(2)	(3)	(4)
Weatherization × Treatment	-0.283**	-0.357***	-0.388***	-0.415***
	(0.119)	(0.116)	(0.119)	(0.120)
Weatherization × Treat × Simultaneous Treat Jobs	0.0171	0.0194	0.0189	0.0216
	(0.0172)	(0.0162)	(0.0164)	(0.0164)
Weatherization × Treat × Simultaneous Control Jobs	-0.0208	-0.0202	-0.0204	-0.0238
	(0.0188)	(0.0174)	(0.0178)	(0.0178)
Weatherization × Simultaneous Treat Jobs	-0.0252**	-0.0243**	-0.0342**	-0.0337**
	(0.0114)	(0.0110)	(0.0133)	(0.0133)
Weatherization × Simultaneous Control Jobs	0.0271**	0.0259**	0.0366**	0.0388**
	(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
r2_a	0	0	0	0
Baseline Weatherization Reduction	-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 post-retrofit. 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. Standard errors are clustered at the house level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

suggesting a limited role for learning-by-doing within the experiment.

Within-Job Reallocation Effects

Rather than shift effort between jobs, contractors may respond to incentives by allocating their effort differently *within* a project when assigned to treatment. We consider callbacks for deficiencies associated with "non-incentivized" tasks–any retrofit categories that are mechanical and do not impact the blower door reading, such as furnace replacements. Table 7 reports the results. The estimates do not suggest any evidence of an effect of performance incentives on the quality of the work done on non-incentivized tasks. This indicates that substitution of effort from non-incentivized to incentivized tasks within jobs does not appear to be a significant concern in our setting. Tasks that do not affect the blower door reading such as furnace replacement are likely independent from air sealing in a contractor's private cost function. Taken together, the results on within and between job reallocation have important implications for the WAP and other energy efficiency programs considering ways to improve cost-effectiveness because it suggests that, in the presence of minimum quality standards, incentivizing airsealing can have large benefits without deleterious impacts on other components of the job.

Table 7: Effects on Callback Rate: Non-Incentivized Retrofits

Non-Building Envelope Callback	(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
Control Group Dep. Variable Mean	0.051	0.051	0.051	0.053

Notes: The dependent variable indicates an air sealing callback. 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. Models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

4.4 Misallocation Under Minimum Quality Standards

Many public programs adopt minimum quality standards or quotas using metrics such as "tasks performed" or "clients served" to incentivize worker effort (Hawkins, Bieretz, and

Brown, 2019). To the extent there is variation in marginal costs across cases due to heterogeneity across recipients' circumstances or heterogeneity in worker ability, piece-rate bonuses have the potential to allocate effort more efficiently. We examine this mechanism in our setting with two testable hypotheses from the model in Online Appendix B: 1) higher-quality contractors will be more responsive to the intervention, given their lower marginal effort costs and 2) outcome variance will be higher for treatment jobs than control jobs. Housing stock heterogeneity creates variation in the marginal cost of additional air sealing improvements such that under the piece-rate bonus, contractors–particularly those who are high quality–will go further on lower marginal cost jobs relative to higher marginal cost jobs. This is something minimum quality standards do not encourage, demonstrating the efficiency advantage of piece-rate contracts. [25]

Our setting creates a unique opportunity to identify higher-quality contractors at baseline by measuring their performance during the program year *prior to* the performance-pay intervention (program year 2017). Unlike many settings where estimating individual value added is complicated by negative assortative matching (e.g. between managers and stores or workers in hospitals), WAP jobs are arbitrarily assigned to contractors through a queue (Metcalfe, Sollaci, and Syverson, 2023). As demonstrated in Christensen et al. (2023), this allows us to separately identify worker skill from selection into jobs. We do this by estimating each contractor's mean effect on gas reductions on jobs in 2017, conditional on observable characteristics about the home and household and expenditures on retrofits performed. Mechanically, these are the estimated contractor fixed effects from a model that regresses 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. We then subtract observed consumption from this counterfactual to get house-month treatment effects (See Online Appendix J). Finally, we group the contractors into quintiles based on their mean reductions and define the top two quintiles as "high quality" contractors.

Table 8 reports estimates of treatment effects from a model that adds: 1) an indicator for the job being performed by a high-quality contractor and 2) an interaction between that indicator and treatment to equation 1. 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

[25] This is in line with Lazear's (2000) canonical model, predicting workers choose type-specific levels of output under piece rates, thereby increasing the variance in productivity.

to contractors in the high quality group as they were when assigned to their counterparts in the lower-quality group. These results are consistent with a model where higher ability contractors can achieve output at lower costs. Table 9 reports estimates of the differential treatment effects on gas consumption for high-quality contractors relative to lower-quality contractors. While these differences are not statistically significant in our sample, point estimates suggest stronger responses from performance incentives in the high-quality group, which is consistent with the findings on CFM50 reductions.

We test whether the variance in both CFM50 and gas reductions are higher under piece-rate performance incentives (treatment) relative to minimum quality standards (control). Using a ratio test for homogeneity of variance, we find the standard deviation for treatment CFM50 reductions (1235) and gas reductions (1.99 MMBtu) is statistically significantly higher than for control (1151 and 1.83 MMBtu respectively). The one-sided p-value for the alternative hypothesis that the ratio of the standard deviation of control to treatment is less than one is $p = 0.023$ for CFM50 reductions and $p = 0.017$ for gas reduction (See Online Appendix H). This is consistent with the hypothesis that piece-rate performance incentives induce a more efficient allocation of effort than minimum quality standards because contractors invest additional effort on lower marginal cost homes in the former regime, but not in the latter. Both sets of hypotheses tested in this section indicate that piece-rate performance incentives can help address the misallocation of effort resulting from compensation for tasks performed and clients served subject to minimum quality standards, which are a feature of contract design in many public programs.

Table 8: Effects of Bonus Treatments on Building Envelope Tightness (CFM 50) by Contractor Quality

CFM50 (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
Control Pre-Weatherization Blower Mean	3609.7	3609.7	3609.7	3585.2
Control Group Dep. Variable Mean	-1557.014	-1557.014	-1557.014	-1552.336

Notes: The dependent variable is the change in building envelope tightness (CFM 50) from WAP upgrades (Post-Pre). 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. Models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table 9: Effects of Bonus Treatments 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	-0.117	-0.221	-0.232	-0.178
	(0.294)	(0.284)	(0.283)	(0.292)
Weatherization × High Quality	-0.0329	0.126	0.125	0.0489
	(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	-0.0694	-0.172	-0.172	-0.143
	(0.320)	(0.315)	(0.313)	(0.319)
Weatherization × High Quality	-0.0326	0.126	0.126	0.0501
	(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
Baseline Weatherization Reduction	-1.582***	-1.582***	-1.582***	-1.623***
	(0.101)	(0.101)	(0.101)	(0.0982)
Control Mean Pre-Weatherization Consumption	7.283	7.283	7.283	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 post-retrofit. 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. Standard errors are clustered at the house level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

4.5 Multiple Hypothesis Testing

In Appendix K, we evaluate the sensitivity of our inferences to multiple hypothesis testing adjustments, focusing on the three primary outcomes specified in our pre-registration: blower door readings, air sealing callbacks, and gas consumption. In Table K1, we report sharpened q-values for pooled treatment effects using the procedures defined in Benjamini, Krieger, and Yekutieli (2006) (BKY) and Benjamini and Hochberg (1995) (BH). Sharpened q-values using the BH procedure increase slightly but do not change the interpretation of significance for any of the primary outcomes in the pooled treatment. Sharpened q-values using the BKY procedure indicate that the pooled estimate of -0.04 for the effect of treatment on air sealing callbacks may be significant at the 0.05 level versus the unadjusted p-value that is only significant at the 0.1 level.

Table K2 shows parallel results from models that control for simulatenous jobs; conclusions are substantively unchanged. Table K3 expands the analysis to all nine pairwise comparisons (each treatment level vs. control and vs. the other treatment) for the three outcomes. Sharpened q-values indicate that the significance of most results are qualitatively unchanged, with the exception of (1) the estimate of -0.04 for callbacks for the high treatment group, which is significant at the 0.1 level using the unadjusted p-value and the BKY procedure but no longer significant at the 0.1 level using the BH procedure, (2) the estimate of -66.17 for increases in air sealing for the low treatment group, which is significant at the 0.1 level using the unadjusted p-value, but no longer significant at the 0.1 level using either the BKY procedure or the BH procedure.

Overall, these results indicate the pooled treatment effects are robust to conservative corrections for multiple hypothesis testing. Importantly, all treatment effects on the welfare-relevant outcome (gas consumption) remain statistically significant. We do find that some marginal findings for callbacks and air sealing are sensitive to the choice of correction method.

5 Returns from Energy Efficiency with Incentive Pay

We use the experimental results above to provide estimates of two distinct measures of the welfare impacts of the bonus payments. Equation 4 defines the more traditional net present value of social benefits calculation and Equation 5 defines the marginal value of public funds (MVPF) of the performance pay bonuses in the IHWAP using the framework proposed in Hendren and Sprung-Keyser (2020) and Finkelstein and Hendren (2020). As shown in recent work, these two measures can lead to different conclusions about optimal public spending on a given program (Garcia and Heckman, 2022; Hendren and Sprung-Keyser, 2022).[26] We find that their comparison yields an informative set of insights regarding spending on our intervention, particularly when considering the optimal level of bonus incentives.

Equation 4 defines the net present value of social benefits from the bonus payment, defined as the difference between the marginal benefit of public expenditure to the net marginal cost to the government:

$$\text{NSB} = \sum_{t=1}^{T_i} \left[\frac{\hat{\beta}^e \times \text{social benefit}_{\text{elec},t}}{(1+r)^t} + \frac{\hat{\beta}^g \times \text{social benefit}_{\text{gas},t}}{(1+r)^t} \right] + PS - \text{Bonus Cost} \tag{4}$$

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:

$$\text{MVPF} = \frac{\sum_{t=1}^{T} \left[\frac{\hat{\beta}^e \times \text{social benefit}_{\text{elec}}}{(1+r)^t} + \frac{\hat{\beta}^g \times \text{social benefit}_{\text{gas}}}{(1+r)^t} \right] + PS}{\text{Bonus Cost}} \tag{5}$$

where $\hat{\beta}^e$ and $\hat{\beta}^g$ are the estimated annual electricity and natural gas savings. Given that Online Appendix Table G1 indicates that the average effect of the bonus payments on electricity consumption is not different from 0, we assume $\hat{\beta}^e = 0$. We compute annual gas savings $\hat{\beta}^g$ by multiplying the monthly treatment effect estimates in column 4 in Table 4 by 12.

We convert estimated energy savings ($\hat{\beta}^g$) 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(Davis and Muehlegger, 2010; Borenstein and Bushnell, 2022). We calculate these social marginal benefits, indicated by $\text{cost}_{\text{elec}}$ and cost_{gas} to be \$8.51 per MMBtu for natural gas and \$37.95

[26] A key limitation of the NSB approach is the typical assumption that a government has a fixed quantity of funds, making it difficult to compare the return to on an additional dollar of spending across different programs or alternative instruments such as subsidy levels (within a program) that may vary in size (Hendren and Sprung-Keyser, 2022). A key limitation of the MVPF approach is the inability to draw conclusions about the optimal size of the government budget or the welfare effects of expanding the budget to fund new programs (Garcia and Heckman, 2022). The approaches also differ in how they address the distortionary cost of generating revenue by raising linear income taxes. We follow current practice and ignore that issue in the analysis, allowing for direct comparability between the measures.

per MMBtu for electricity.[27] We use a range of estimates of the social cost of carbon (SCC) to account for the benefits from avoided GhG emissions, including recent estimates of $121 and $185 per ton (Carleton and Greenstone, 2022; Rennert et al., 2022). Since the EPA has not yet adopted their recently proposed value of $190 per ton and most prior work on the cost-effectiveness of WAP and other energy efficiency programs has relied upon older estimates of $40-60 per ton, we include the previously EPA-approved estimate of $51 per ton to illustrate the effects of updated SCC estimates on the cost-effectiveness of energy efficiency investments. T represents the expected lifetime of the retrofits. Baseline estimates are computed using the cost-weighted average lifetime of the 34.5 years for the retrofits performed in the IHWAP program (Christensen et al., 2023).[28] Benefits are discounted using the 2018 DOE-recommended discount rate during the study period (r) of 2%. We examine sensitivity of estimates to alternate lifespan assumptions in Online Appendix M.

We calculate the producer surplus associated with the incentives (PS) as the difference between the bonus payment and contractors' supply curve for CFM50 reductions. We approximate these values using the treatment effect estimates from Table 2, assuming that supply curves are piece-wise linear from the CFM50 target to the mean reduction at $0.40 and between the mean reductions at $0.40 and $1.00 (Online Appendix L).

We assume that there are no non-energy benefits to the household from weatherization, such as health impacts, which could increase true benefits relative to our estimates.[29] $Bonus\,Cost$ reflects the incentive payment for the average treatment contract. We provide estimates using the high and low incentive treatments. 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.[30] We also assume that the effect of the transfer

[27] For electricity, we estimate the mean difference between retail prices and social marginal social cost estimates by region and pollutant from table A1 from Borenstein and Bushnell (2022), focusing on the costs for Ameren (the dominant service territory in our sample). For CO2, the estimates by load tercile ($27.12, $28.41, and $28.67) are all close to $28. A marginal CO2 emissions cost of $28 at $50/ton from Borenstein and Bushnell (2018) yields a marginal emissions rate of 0.56 and an increase in the social marginal cost of cost of 0.56 for the $51/ton SCC, $39.76 for the $151/ton SCC, and $75.6 for the $185/ton SCC. We then apply that difference to the month-of-year residential electricity price averages for our study period. We account for expected fuel price escalation using the DOE's energy price indices and projections used by OMB for Census Region 2.

[28] We use the same retrofit lifespans as the WAP except for the 25 year air sealing and insulation lifespans as recent engineering literature finds they are much longer: 50 years for cellulose fiber (ISOCELL GmbH, 2014), 35-50 years for expanded polystyrene (EUMEPS, 2017; IVH, 2015), and 50-150 years for extruded polystyrene (EXIBA, 2019). Accounting for longer lifespans results in a 34.5 year average.

[29] It is also possible that increasing the tightness of a building envelope results in negative health impacts from indoor air pollution, though each IHWAP project also requires a ventilation system to avoid adverse health impacts, making this unlikely.

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

to program recipients will not result in meaningful increases in tax revenue through consumption on other goods. Either of these effects would make our estimates a lower bound on the true MVPF from the intervention.

Returns from Performance Incentives

Panel A of Table 10 reports estimates of returns from the performance incentive using the NSB from equation. 4 and the MVPF from equation. 5. Column 1 reports the average payment for the high and low treatment jobs. Column 2 reports estimates of producer surplus (for IHWAP contractors) for high ($1.00/CFM50) and low ($0.40/CFM50) performance incentives as described above. The resulting estimates are $13/home in producer surplus for contractors in the low treatment and $99/home in the high treatment.

Columns 3-8 report estimates of (a) the social net benefits and (b) the marginal value of public funds from expenditures on performance incentives and the resulting energy reductions across the three levels of SCC considered. The incentive treatment was remarkably cost effective under all SCC estimates considered, ranging from $773 to $1,644 in social net benefits for the low treatment and from $880 to $1,872 for the high treatment. The equivalent MVPF estimates range from $6.89 to $14.53 for the low bonus level and from $3.46 to $6.96 for the high bonus level. Comparing the two welfare measures offers important policy insights. The social net benefits from the high bonus incentive is slightly higher than the low bonus incentive on a per home basis, indicating that program expansion by using larger bonus payments ($1.00 per CFM50) would yield additional surplus relative to a program with the equivalent number of contracts at the lower bonus level. However, the MVPF is much higher at the lower bonus level ($0.40 per CFM50), suggesting that a smaller incentive level may be advantageous when considering trade-offs between allocating larger incentive payments to a smaller pool of contracts versus smaller incentive payments to a larger pool of contracts.[31]

Experimental variation in incentive levels provides guidance on the optimal payment. The results indicate that increasing payments has a very inelastic effect on social benefits. Under the middle scenario–using a social cost of carbon (SCC) of $121–the estimated social benefit is nearly flat, rising from $1,114 at the low payment to just $1,115 at the high payment. In contrast, program costs increase almost linearly with the incentive. The marginal cost per dollar of incentive is roughly constant: $114/0.4 = $285 when increasing from no payment to the low payment, and $169/0.6 = $282 from the low to the high payment.

[31] These measures do not account for the distortionary cost associated with raising revenue from a linear income tax, which under standard assumptions would reduce the optimal level of total expenditure in the program. Feldstein (1999) estimates a greater than $2 deadweight loss associated with increases existing tax rates, though estimates used in empirical welfare analysis vary widely.

Figure 2 plots the experimentally estimated social benefits (green) and costs (red) at each payment level, connected with piecewise linear segments. If benefits were sufficiently smooth, the optimal payment would occur where the marginal cost of incentivizing air sealing equals its marginal social benefit. Over the range we experimentally examined, however, the two are not equal for nearly any smoothly increasing functional form between $0 and $0.40: before $0.40, marginal benefits are higher than marginal costs; after $0.40, they are lower, implying that net benefits are maximized in the neighborhood of the $0.40 bonus and that returns to higher payments are negligible.[32]

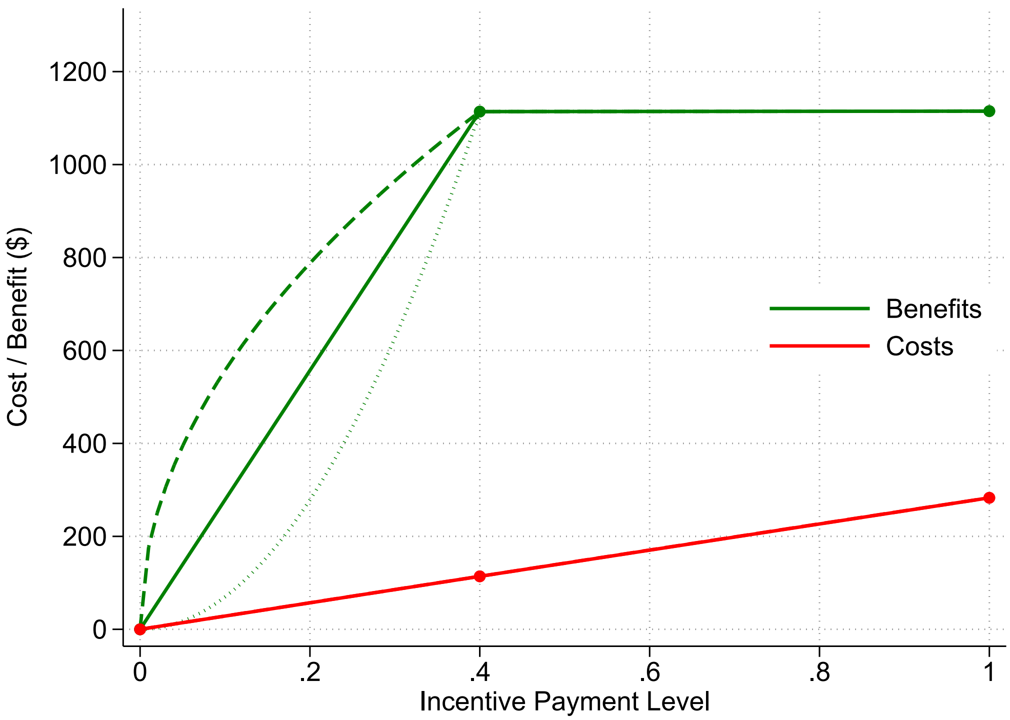


Figure 2: Social Benefits and Costs by Incentive Payment

Notes: Green circles represent estimated social benefits and red circles estimate program costs using the $121 per ton social cost of carbon scenario. Solid lines connect values using piecewise linear segments.

Effects of Incentives on Returns to Weatherization Tasks

Panel B Table 10 reports estimates of the net benefits and MVPF for the baseline IWHAP program, in the absence of any incentives. We follow a methodology similar to that of the performance incentive outlined above, except the benefits include estimated savings

[32] It remains possible that benefits increase sharply at very low payments and then flatten below $0.40, in which case the optimal payment could be lower. Additional targeted experiments could provide evidence on the shape of the net benefits function just below the $0.40 level.

Table 10: Effects of Performance Incentives on Social Welfare

	Cost	Producer Surplus	*SCC* = \$51 Net Benefits	*SCC* = \$51 MVPF	*SCC* = \$121 Net Benefits	*SCC* = \$121 MVPF	*SCC* = \$185 Net Benefits	*SCC* = \$185 MVPF
Panel A. Performance Incentive								
Low Treat	\$114	\$13	\$773	\$6.89	\$1,228	\$10.88	\$1,644	\$14.53
High Treat	\$283	\$99	\$880	\$3.46	\$1,398	\$5.29	\$1,872	\$6.96
Panel B: Baseline WAP								
Baseline: All Retrofits	\$9,655	.	\$-2,584	\$0.73	\$10,621	\$2.10	\$22,655	\$3.36
Baseline: Air Sealing	\$1,128	.	\$-303	\$0.73	\$1,246	\$2.10	\$2,658	\$3.36
Baseline: All Incentivized	\$2,037	.	\$-547	\$0.73	\$2,250	\$2.10	\$4,800	\$3.36
Panel C: Baseline Air Sealing + Incentive								
Baseline Air Sealing + Low Treat	\$1,242	\$13	\$369	\$1.3	\$2,373	\$2.91	\$4,201	\$4.38
Baseline Air Sealing + High Treat	\$1,411	\$99	\$393	\$1.28	\$2,460	\$2.74	\$4,346	\$4.08
Panel D: Baseline All Incentivized + Incentive								
Baseline All Incentivized + Low Treat	\$2,151	\$13	\$124	\$1.06	\$3,376	\$2.57	\$6,342	\$3.95
Baseline All Incentivized + High Treat	\$2,320	\$99	\$149	\$1.06	\$3,464	\$2.49	\$6,487	\$3.80
Panel E: Baseline All Retrofits + Incentive								
Baseline All Retrofits + Low Treat	\$9,769	\$13	\$-1,925	\$0.80	\$11,735	\$2.21	\$24,185	\$3.49
Baseline All Retrofits + High Treat	\$9,938	\$99	\$-1,987	\$0.80	\$11,736	\$2.19	\$24,244	\$3.45

Notes: Panel A reports estimates of the NSB and MVPF for expenditures on performance incentives on air sealing retrofits using estimates of treatment effects from Table 4. Panel B reports estimates of social net benefits and MVPF for all baseline retrofits. Estimates of benefits from baseline air sealing retrofits are assumed to be proportional to expenditures on air sealing given control mean weatherization effect. Panels C-D report estimates of social net benefits and MVPF of combining the baseline air sealing investments (Panel C) or all incentivized investments (Panel D) with performance incentives. Panel E reports estimates of social net benefits and MVPF for all baseline retrofits combined with performance incentives. Net present energy benefits use gas and electricity prices per MMBTU for 2017. Emissions factors for natural were obtained from EPA (1998). Data provided by Borenstein and Bushnell (2022) 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: 34.5 years when assuming a 150-year lifespan for long-lived air-sealing materials.

for both electricity and natural gas and we do not include a value of producer surplus, as it could not be separately estimated for the baseline program. For the sample program years in IHWAP, the estimates of social returns from investments in IHWAP retrofits are highly sensitive to assumptions about the social cost of carbon. With the lowest SCC estimate of \$51, which has generally been used to estimate the cost-effectiveness of investments in energy efficiency programs, the baseline program yields \$-2,584 in net benefits per home and a MVPF of \$0.73. This increases to \$10,621 per home and a MVPF of \$2.10 using the SCC estimate of \$121 and to \$22,655 and a MVPF of \$3.36 using the SCC estimate of \$185. In interpreting this result, note that returns per dollar spent in the baseline IHWAP are lower due to the introduction of braided funding during this period wherein some measures were performed to help households irrespective of SIR. For example, in the three years before braided funding, NPBs in IHWAP ranged between $616 and $736 2021 USD (Online Appendix Table A2). In rows 2 and 3 of Panel B, we assume these same returns to spending for air sealing retrofits and the broader category of incentivized retrofits targeted by the intervention since we cannot disentangle returns to spending on individual retrofits.

What is the effect of performance pay on the returns from air sealing measures completed in a home? To answer this question, we compare the estimates of benefits from air sealing retrofits in control (Panel B) to the benefits from air sealing retrofits with performance bonuses. To recover the latter estimates, we combine the benefits and costs from baseline air sealing (Panel B) with the benefits and costs from performance bonuses on air sealing (Panel A). We report the results of this exercise in Panel C. We find that the bonus treatment has a large effect on both the low and high incentive levels, increasing the net benefits to positive irrespective of the assumed SCC and increasing the MVPF from baseline air sealing retrofits by 78-130%. The MVPF increases from 2.10 to 2.91 for the low treatment (2.74 for the high treatment) using the more conservative of the two recent SCC estimates, $125, or from 3.36 to 4.38 (4.08 for the high treatment) using the higher estimate of SCC, $185.

The 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 also has an important effect on this larger category of expenditures, increasing the MVPF from 2.10 to 2.57 using the low bonus treatments (2.49 for high treatment) and the more conservative of the two recent estimates of SCC, $121, or from 3.36 to 3.95 (3.80 for high treatment) using the higher estimate of SCC, $185.

Effects of Incentives on Overall Returns to WAP Investments

While the incentive-based pay intervention studied in the current paper applies the incentive treatment to a fraction of the tasks in the average IHWAP contract, an additional policy-relevant question concerns the effect of the bonus payments for air sealing on the social net benefits from investments in the IHWAP program as a whole (See Panel E). Similar to the approach used to recover estimates in Panel C and D, we combine the benefits and costs from baseline retrofit tasks (Panel B) with the benefits and costs from performance bonuses on air sealing (Panel A).

The impact of air sealing bonus on overall program cost-effectiveness is also substantial.

Using the most conservative of the two recent SCC estimates ($51) the incentive payments increase the social net benefits by 34%, from $-2,584 to $-1,925. Using the highest of the SCC estimates ($185), the incentive payments in both low and high bonus regimes also result in an economically substantial increase from $22,655 to $24,185. Across both payment levels and irrespective of assumptions about SCC, we find that the incentive pay intervention increases the overall MVPF of the IHWAP program by 4-10%. Any differences in the returns from low versus high incentive payments becomes less economically meaningful when considering the social benefits from spending on the program as a whole.

As illustrated in Panel B, treated tasks represent a modest fraction of total average home expenditures – 12% for air sealing tasks and 21% for all incentivized tasks. Even when applied to this relatively small fraction of contracted expenditures, performance incentives have a meaningful impact on the overall returns to spending in the IHWAP program. In the absence of braided funding, the intervention would have an even larger effect on NSB and MPVF.

6 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 welfare-relevant program/policy objectives. This paper presents findings form a 2-year field experiment on the impacts of piece-rate performance incentives in the Weatherization Assistance Program. We find that these performance incentives increased natural gas savings by 24% and generated $5.39-$14.53 in social benefits per dollar invested.

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 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.

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 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 are worth considering at scale. Blower door tests are a standard component of energy audits and are widely used in federal, state, and utility programs, as well as by private contractors. Most homes—especially those more than 20 years old—have opportunities for air sealing. As a result, piece-rate payments for conducting blower door tests could be readily incorporated into other government or utility initiatives, or into private contracts.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge (2023). "When should you adjust standard errors for clustering?" *The Quarterly Journal of Economics* 138(1), pp. 1–35.

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

Altman, Douglas G (1985). "Comparability of randomised groups". *Journal of the Royal Statistical Society Series D: The Statistician* 34(1), pp. 125–136.

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S Lee (2020). "Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services". *American Economic Review* 110(5), pp. 1355–94.

Athey, Susan and Guido W Imbens (2022). "Design-based analysis in difference-in-differences settings with staggered adoption". *Journal of Econometrics* 226(1), pp. 62–79.

Aucejo, Esteban, Teresa Romano, and Eric S Taylor (2022). "Does evaluation change teacher effort and performance? Quasi-experimental evidence from a policy of retesting students". *Review of Economics and Statistics* 104(3), pp. 417–430.

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.

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing". *Journal of the Royal statistical society: series B (Methodological)* 57(1), pp. 289–300.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli (2006). "Adaptive linear step-up procedures that control the false discovery rate". *Biometrika* 93(3), pp. 491–507.

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.

Blonz, Joshua A (2023). "The costs of misaligned incentives: Energy inefficiency and the principal-agent problem". *American Economic Journal: Economic Policy* 15(3), pp. 286–321.

Borenstein, Severin and James B Bushnell (2022). "Do Two Electricity Pricing Wrongs Make a Right? Cost Recovery, Externalities, and Efficiency." *American Economic Journal: Economic Policy* 14(4), pp. 80–110.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2021). "Revisiting Event Study Designs: Robust and Efficient Estimation". *Working Paper*.

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.

Bruhn, Miriam and David McKenzie (2009). "In pursuit of balance: Randomization in practice in development field experiments". *American economic journal: applied economics* 1(4), pp. 200–232.

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.

Callaway, Brantly and Pedro HC Sant'Anna (2021). "Difference-in-differences with multiple time periods". *Journal of Econometrics* 225(2), pp. 200–230.

Carleton, Tamma and Michael Greenstone (2022). "A Guide to Updating the US Government's Social Cost of Carbon". *Review of Environmental Economics and Policy* 16(2), pp. 196–218.

Cattaneo, Matias D, Richard K Crump, Max H Farrell, and Yingjie Feng (2019a). "Binscatter regressions". *arXiv preprint arXiv:1902.09615*.

Cattaneo, Matias D, Richard K Crump, Max H Farrell, and Yingjie Feng (2019b). "On binscatter". *arXiv preprint arXiv:1902.09608*.

Christensen, Peter, Paul Francisco, Erica Myers, and Mateus Souza (2023). "Decomposing the wedge between projected and realized returns in energy efficiency programs". *Review of Economics and Statistics* 105(4), pp. 798–817.

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 (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.

de Chaisemartin, Clément and Xavier D'Haultfœuille (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". *American Economic Review* 110(9), pp. 2964–96. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20181169.

Dohmen, Thomas and Armin Falk (2011). "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101(2), pp. 556–90.

Duflo, Esther (2017). "The Economist as Plumber". *American Economic Review* 107(5), pp. 1–26.

Duflo, Esther (2020). "Field Experiments and the Practice of Policy". *American Economic Review* 110(7), pp. 1952–1973.

Duflo, Esther, Rema Hanna, and Stephen P Ryan (2012). "Incentives work: Getting teachers to come to school". *American economic review* 102(4), pp. 1241–1278.

Dumont, Etienne, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer (2008). "Physicians' multitasking and incentives: Empirical evidence from a natural experiment". *Journal of health economics* 27(6), pp. 1436–1450.

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.

Feldstein, Martin (1999). "Tax Avoidance and the Deadweight Loss of the Income Tax". *Review of Economics and Statistics* 81(4), pp. 674–680.

Feng Lu, Susan (2012). "Multitasking, information disclosure, and product quality: Evidence from nursing homes". *Journal of Economics & Management Strategy* 21(3), pp. 673–705.

Finkelstein, Amy and Nathaniel Hendren (2020). "Welfare Analysis Meets Causal Inference." *Journal of Economic Perspectives* 34(4), pp. 146–67.

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.

Garcia, Jorge Luis and James J Heckman (2022). *On Criteria for Evaluating Social Programs.* NBER Working Paper 30005.

Gillingham, Kenneth and James H Stock (2018). "The Cost of Reducing Greenhouse Gas Emissions." *Journal of Economic Perspectives* 32(4), pp. 53–72.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". *Journal of Econometrics* 225(2), pp. 254–277.

Guiteras, Raymond P and B Kelsey Jack (2018). "Productivity in piece-rate labor markets: Evidence from rural Malawi". *Journal of Development Economics* 131, pp. 42–61.

Hahn, Robert W, Nathaniel Hendren, Robert D Metcalfe, and Ben Sprung-Keyser (2024). *A welfare analysis of policies impacting climate change*. National Bureau of Economic Research Working Paper 32728.

Hawkins, Rayanne, Brian Bieretz, and Madeline Brown (2019). "Incentivizing Results: Contracting for Outcomes in Social Service Delivery".

Hendren, Nathaniel and Ben Sprung-Keyser (2020). "A Unified Welfare Analysis of Government Policies". *The Quarterly Journal of Economics* 135(3), pp. 1209–1318.

Hendren, Nathaniel and Ben Sprung-Keyser (2022). *The Case for Using the MVPF in Empirical Welfare Analysis*. NBER Working Paper 30029.

Holmstrom, Bengt and Paul Milgrom (1991). "Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *The Journal of Law, Economics, and Organization* 7, pp. 24–52.

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, Sebastien and Joseph E Aldy (2017). "Belt and Suspenders and More: The Incremental Impact of Energy Efficiency Subsidies in the Presence of Existing Policy Instruments." *American Economic Journal: Economic Policy* 9(4), pp. 227–255.

IEA (2023). *The Breakthrough Agenda Report 2023: Accelerating Sector Transitions Through Stronger International Collaboration.* https://www.iea.org/reports/breakthrough-agenda-report-2023. Accessed 2025-02-26.

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/.

Infrastructure Investment and Jobs Act (2021). United States Code. Public Law 117-58. U.S. Congress. URL: https://www.congress.gov/bill/117th-congress/house-bill/3684.

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.

Johnson, Ryan M, David H Reiley, and Juan Carlos Muñoz (2015). ““The war for the fare”: how driver compensation affects bus system performance”. *Economic Inquiry* 53(3), pp. 1401–1419.

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.

Kosfeld, Michael and Ferdinand A Von Siemens (2011). “Competition, cooperation, and corporate culture”. *The RAND Journal of Economics* 42(1), pp. 23–43.

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.

Leaver, Clare, Owen Ozier, Pieter Serneels, and Andrew Zeitlin (2021). "Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools". *American economic review* 111(7), pp. 2213–2246.

Lemieux, Thomas, 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.

Luo, Renfu, Grant Miller, Scott Rozelle, Sean Sylvia, and Marcos Vera-Hernández (2020). "Can bureaucrats really be paid like ceos? Substitution between incentives and resources among school administrators in China". *Journal of the European Economic Association* 18(1), pp. 165–201.

Metcalfe, Robert D, Alexandre B Sollaci, and Chad Syverson (2023). *Managers and Productivity in Retail*. Tech. rep. National Bureau of Economic Research.

Mullen, Kathleen J, Richard G Frank, and Meredith B Rosenthal (2010). "Can you get what you pay for? Pay-for-performance and the quality of healthcare providers". *The Rand journal of economics* 41(1), pp. 64–91.

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

Paarsch, Harry J and Bruce Shearer (2000). "Piece rates, fixed wages, and incentive effects: Statistical evidence from payroll records". *International Economic Review* 41(1), pp. 59–92.

Papineau, Maya, Nicholas Rivers, and Kareman Yassin (2023). "Evaluating a major home energy efficiency retrofit program: long-run savings, realization rates and the distribution of benefits". *Available at SSRN 4348993*.

Rennert, Kevin, Frank Errickson, Brian C Prest, Lisa Rennels, Richard G Newell, William Pizer, Cora Kingdon, Jordan Wingenroth, Roger Cooke, Bryan Parthum, et al. (2022). "Comprehensive Evidence Implies a Higher Social Cost of CO2." *Nature*, pp. 1–3.

Rothstein, Jesse (2015). "Teacher quality policy when supply matters". *American Economic Review* 105(1), pp. 100–130.

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.

Strezhnev, Anton (2018). "Semiparametric weighting estimators for multi-period difference-in-differences designs". *Working Paper*. URL: https://www.antonstrezhnev.com/research/.

Sun, Liyang and Sarah Abraham (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". *Journal of Econometrics* 225(2), pp. 175–199.

U.S. Government Accountability Office (June 2021). *A Snapshot of Government-Wide Contracting For FY 2020 (infographic)*. URL: https://www.gao.gov/blog/snapshot-government-wide-contracting-fy-2020-infographic.

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.

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

Online Appendix

A Trends in Total Household-Level Spending and Cost-Effectiveness by Program Year

Table A1: 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 in 2021 USD.

Table A2: Average Net Present Benefits by Program Years 2009-2016

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

Notes: This table presents average home-specific net present benefits by program year as estimated in (Christensen et al., 2023). 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. This table is reproduced from the online appendix in (Christensen et al., 2023).

B Principal-Agent Model of Contractor Effort: Multi-Task Settings with Heterogeneous Ability

In this section, we present a principal-agent model of effort allocation for heterogeneous ability contractors in a multi-task setting that integrates insights from both Holmstrom and Milgrom (1991) and Lazear (2000). The model generates predictions about the effect of a piece-rate bonus on the allocation of effort by contractors, as well as the potential variation for contractors who vary in ability.

B.1 Model Set Up

Suppose the principal has m different tasks for their agent (a contractor), to perform. The contractor chooses levels of effort, $\boldsymbol{t} = (t_1, ..., t_m)$ for a vector of m tasks at a personal, strictly convex cost in each dimension, $C(t_1, ...t_m)$. We begin with the assumption that effort in one dimension i does not affect the marginal cost of effort in any other dimension j, so that the cross-partial is zero, i.e. $C_{ij} = 0$. Let a denote ability, where output, $\mu(t, a)$, is an increasing concave function of both ability and effort for each task (i).

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

Ability determines the amount of effort required for a given level of output. By differentiating equation B.1, we can see that higher ability is associated with lower effort for achieving 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 principal observes a vector of signals $\boldsymbol{x(t) = \mu(t) + \varepsilon}$ about the effort expended by a contractor, which are a function of true output ($\boldsymbol{\mu(t)}$) and error terms ($\boldsymbol{\varepsilon}$). Assume the error terms are normally distributed, have a mean vector zero, and are stochastically independent across tasks.

B.2 Minimum Quality Standards

In a baseline condition, the principal uses minimum quality standards to ensure that the agent's incentives are compatible with allocating sufficient effort to maintain quality in the program. The principal pays the contractor a fixed wage, w_i, for each task i, but she withholds payment until minimum quality has been achieved for all tasks. 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.[33]

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

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, $\boldsymbol{k(t, a)} = E[\boldsymbol{\lambda'\phi(t, a)}]$, is also decreasing and convex in effort and ability ($k_t < 0, k_a; k_{tt} > 0, k_{aa}$).

Assume that the agent is risk neutral, such that the agent chooses the vector $\boldsymbol{t^0}$ that minimizes total costs of effort and callbacks.[34] 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.

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

Participation Constraint

For any given set of minimum quality 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 in the program.

B.3 Impacts of Piece-Rate Bonuses

Suppose the principal introduces a piece-rate bonus for task i, 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 across all contractor tasks is as follows.

$$\pi(\boldsymbol{x, w}) = \begin{cases} \sum_{j=1}^{m} \left[w_j - k(t_j^0, a) - C(t_j^0, a)\right] \quad \forall x_i \leq \bar{x}_i, & \text{(B.3)} \\ \sum_{j=1}^{m} \left[w_j - k(t_j^*, a) - C(t_j^*, a)\right] + b_i(x_i^* - \bar{x}_i) \quad \forall x_i > \bar{x}_i & \text{(B.4)} \end{cases}$$

Contractors choose the maximum of B.3 and B.4, where $\boldsymbol{x^*}$ is the vector of output associated with the optimal amount of effort $\boldsymbol{t^*}$ under a piece rate b_i for task i, which solves the following

[34] 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.

first order condition where subscripts i indicate derivatives with respect to t_i

$$C_i(t,a) + k_i(t,a) = b_i \cdot x_i^*(t,a) \quad i = 1 \tag{B.5}$$

and $t_j^* = t_j^0$ for all other tasks $j \neq i$. Under a piece rate, contractors choose t^* such that total private marginal costs are equal to the bonus payment. Whether B.3 or B.4 maximizes contractor profits will depend on a contractor's ability.

By totally differentiating B.5 with respect to ability and rearranging terms, we can see that the optimal level of contractor effort is increasing in ability.

$$\frac{\partial t^*}{\partial a} = \frac{\frac{\partial x_i^*}{\partial a}}{C_{ii} + k_{ii} - x_{ii}^* \cdot b} > 0 \tag{B.6}$$

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 at a rate that is more than proportional to their increase in effort. This model yields two hypotheses regarding the effect of bonuses on contractor effort:

Hypothesis 1: Effort will not decrease in response to the introduction of a piece-rate bonus.[35]

Hypothesis 2: Higher-ability workers will respond with stronger increases in output than lower-ability workers.[36]

B.4 Effort Reallocation

So far, we have assumed that effort in one dimension does not affect the marginal cost of effort in another, $C_{ij} = 0$. This could be the case if the contractor were not constrained in their capacity to bring in more labor either through hiring or giving existing employees more hours. Introducing a bonus for one task in this case would not affect effort on any other tasks.

Now consider the possibility that effort in one dimension could lower or increase the marginal cost of effort in another, depending on whether the tasks are complements or substitutes in the contractor's private cost function.[37] We can quantify reallocation by totally differentiating the first order condition in (B.2) for task j with respect to b_i and solving for, $\frac{\partial t_j}{\partial b_i}$, as

[35] First, as in Lazear (2000), and as long as there is some ability type for which output rises, effort will increase.

[36] A third hypothesis that comes out of this section that we are not able to test empirically is that 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 $\bar{a}$).

[37] Assume for this exercise that the contractor's optimal output under the bonus regime, x^*, is above above a minimum level, $\bar{x}_i$.

follows.

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

$$\frac{\partial t_j}{\partial b_i} = -\frac{(C_{ji} + k_{ji})}{(C_{jj} + k_{jj})}\frac{\partial t_i}{\partial b_i} \tag{B.7}$$

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.

Importantly, minimum quality standards will help reduce incentives to pull effort away from those tasks. We can see from equation B.7, the more convex the effort cost and callback function, the less responsive effort in one dimension will be to bonuses on another dimension. Therefore, whether the bonus leads to reallocation of effort in other, non-incentivized tasks depends on whether those tasks are complements or substitutes to air sealing in contractors' private costs functions and is an empirical question that we can test given our experimental design and data collection.

C Intent-to-Treat Estimates

This Appendix contains the intent-to-treat (ITT) reduced-form estimates for each of the 2SLS results tables in the text. 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. In our 2SLS regression presented in the main body of the paper random assignment instruments for treatment jobs with contractors who are eligible to receive payments. The ITT results can be interpreted as the effect of a job being assigned to treatment and are somewhat lower than the 2SLS estimates because not all contractor-jobs in the program were eligible for payment. ITT results are qualitatively similar to the 2SLS results with the main ITT treatment effect consistently falling within 88 to 95% of the 2SLS estimate.

Table C1: Effects of Bonus on Building Envelope Tightness (CFM50): Intent-to-Treat

CFM50 (Pre - Post)	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
ITT	-57.64*	-66.74**	-69.09**	-81.44***
	(29.57)	(28.84)	(29.43)	(28.90)
Panel B: Effect by Treatment Group				
Low ITT	-51.61	-53.03	-49.54	-61.18*
	(36.30)	(35.67)	(36.01)	(35.88)
High ITT	-64.14*	-81.11**	-90.41**	-103.7***
	(36.25)	(35.30)	(36.03)	(35.40)
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
Control Pre-Weatherization Blower Mean	3609.7	3609.7	3609.7	3585.2
Control Group Dep. Variable 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. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table C2: Effects of Bonus on Air Leakage Callback Rate: Intent-to-Treat

Air Leakage Callback	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
ITT	-0.0292*	-0.0322**	-0.0241	-0.0270*
	(0.0152)	(0.0150)	(0.0157)	(0.0161)
Panel B: Effect by Treatment Group				
Low ITT	-0.0182	-0.0202	-0.0149	-0.0183
	(0.0193)	(0.0192)	(0.0195)	(0.0199)
High ITT	-0.0406**	-0.0442***	-0.0336*	-0.0360**
	(0.0170)	(0.0168)	(0.0176)	(0.0182)
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
Control Group Dep. Variable 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. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table C3: Effects of Bonus on Gas Usage (MMBtu): Intent-to-Treat

Gas MMBtu	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Weatherization × ITT	-0.266**	-0.324***	-0.336***	-0.353***
	(0.110)	(0.106)	(0.107)	(0.108)
Panel B: Effect by Treatment Group				
Weatherization × Low ITT	-0.308**	-0.331**	-0.325**	-0.335***
	(0.133)	(0.129)	(0.128)	(0.129)
Weatherization × High ITT	-0.222	-0.315**	-0.348**	-0.374***
	(0.139)	(0.135)	(0.137)	(0.141)
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
Baseline Weatherization Reduction	-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. 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 C4: Effects on Blower Door: Simultaneous Jobs Within Contractor (Intent-to-Treat)

CFM50 (Post - Pre)	(1)	(2)	(3)	(4)
ITT	-53.86*	-64.78**	-65.89**	-78.33***
	(29.70)	(29.05)	(29.51)	(29.01)
ITT × Simultaneous ITT Jobs	3.876	6.395	5.849	3.410
	(4.295)	(4.191)	(4.186)	(4.050)
ITT × Simultaneous Control Jobs	-6.634	-7.265*	-7.085	-4.420
	(4.619)	(4.369)	(4.349)	(4.291)
Simultaneous ITT Jobs	-3.800	-4.484	-3.091	0.00836
	(3.237)	(3.264)	(3.646)	(3.377)
Simultaneous Control Jobs	3.188	2.769	1.343	-1.100
	(3.579)	(3.364)	(3.801)	(3.671)
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	1696	1600
Control Pre-Weatherization Blower Mean	3609.7	3609.7	3609.7	3585.2
Control Group Dep. Variable 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). Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table C5: Effects on Gas Usage: Simultaneous Jobs Within Contractor (Intent-to-Treat)

Gas MMBtu	(1)	(2)	(3)	(4)
Weatherization × ITT	-0.247**	-0.311***	-0.340***	-0.364***
	(0.110)	(0.107)	(0.109)	(0.111)
Weatherization × ITT × Simultaneous ITT Jobs	0.0155	0.0173	0.0167	0.0197
	(0.0158)	(0.0149)	(0.0151)	(0.0152)
Weatherization × ITT × Simultaneous Control Jobs	-0.0200	-0.0192	-0.0194	-0.0231
	(0.0172)	(0.0159)	(0.0162)	(0.0164)
Weatherization × Simultaneous ITT Jobs	-0.0220**	-0.0208**	-0.0304***	-0.0301***
	(0.00994)	(0.00938)	(0.0116)	(0.0116)
Weatherization × Simultaneous Control Jobs	0.0230**	0.0212**	0.0315**	0.0339***
	(0.0113)	(0.0104)	(0.0127)	(0.0129)
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
r2_a	1	1	1	1
Baseline Weatherization Reduction	-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. 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 C6: Effects on Callback Rate: Non-Incentivized Retrofits (Intent-to-Treat)

Non-Building Envelope Callback	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
ITT	0.0197	0.0199	0.0137	0.0159
	(0.0127)	(0.0130)	(0.0136)	(0.0139)
Panel B: Effect by Treatment Group				
Low ITT	0.0205	0.0180	0.0120	0.0140
	(0.0162)	(0.0162)	(0.0169)	(0.0180)
High ITT	0.0189	0.0219	0.0155	0.0179
	(0.0160)	(0.0165)	(0.0167)	(0.0167)
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
Control Group Dep. Variable 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. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table C7: Effects on Building Envelope Tightness by Contractor Quality: Intent-to-Treat

CFM50 (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
Control Pre-Weatherization Blower Mean	3609.7	3609.7	3609.7	3585.2
Control Group Dep. Variable 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 OLS. ITT (intent-to-treat) refers to the randomized treatment assignment. 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 C8: Effects on Gas Usage (MMBtu) by Contractor Quality: Intent-to-Treat

Gas MMBtu	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Weatherization × ITT	-0.249**	-0.289**	-0.303**	-0.331***
	(0.121)	(0.117)	(0.118)	(0.120)
Weatherization × ITT × High Quality	-0.126	-0.229	-0.239	-0.186
	(0.285)	(0.275)	(0.275)	(0.283)
Weatherization × High Quality	-0.0333	0.122	0.121	0.0428
	(0.198)	(0.193)	(0.191)	(0.193)
Panel B: Effect by Treatment Group				
Weatherization × Low ITT	-0.298**	-0.304**	-0.299**	-0.316**
	(0.149)	(0.144)	(0.144)	(0.145)
Weatherization × High ITT	-0.198	-0.274*	-0.307**	-0.348**
	(0.150)	(0.148)	(0.149)	(0.153)
Weatherization × Low ITT × High Quality	-0.0882	-0.191	-0.190	-0.159
	(0.315)	(0.311)	(0.309)	(0.315)
Weatherization × High ITT × High Quality	-0.166	-0.271	-0.292	-0.215
	(0.391)	(0.374)	(0.378)	(0.395)
Weatherization × High Quality	-0.0329	0.122	0.122	0.0438
	(0.198)	(0.193)	(0.191)	(0.193)
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
Baseline Weatherization Reduction	-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 OLS. ITT (intent-to-treat) refers to the randomized treatment assignment. 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.

D Robustness Across Samples and Clustering Level

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) contractor callbacks and (2) gas consumption

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 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: 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
Control Pre-Weatherization Blower Mean	3704.9	3704.9	3704.9	3687.9
Control Group Dep. Variable 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
Control Pre-Weatherization Blower Mean	3497.9	3497.9	3497.9	3489.9
Control Group Dep. Variable 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.

Tables D3–D5 report results for the three primary outcomes—blower door readings, air sealing callbacks, and gas consumption—with standard errors clustered at the contractor level. Inference for blower door readings and gas consumption remains unchanged, with results staying statistically significant at the same level for our preferred outcomes. For air-sealing-related callbacks, however, the estimates are less precise and no longer statistically significant.

Table D3: Effects of Bonus Treatments on Building Envelope Tightness (CFM50) Contractor-Level Clustering

CFM50 (Post - Pre)	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Treat	-64.09*	-73.97*	-75.96**	-88.67**
	(36.21)	(36.73)	(35.16)	(35.98)
Panel B: Effect by Treatment Group				
Low Treat	-57.25	-58.58	-54.25	-66.17*
	(42.77)	(41.44)	(39.59)	(35.13)
High Treat	-71.49	-90.26*	-99.91**	-113.8**
	(47.33)	(46.69)	(47.02)	(49.68)
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
Control Pre-Weatherization Blower Mean	3609.7	3609.7	3609.7	3585.2
Control Group Dep. Variable Mean	-1569.306	-1569.306	-1569.306	-1562.708

Notes: The dependent variable is the change in building envelope tightness (CFM 50) from WAP upgrades (Post-Pre). 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. Models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Standard errors clustered at the contractor level are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table D4: Effects on Air Leakage Callback Rate: Contractor-Level Clustering

Air Leakage Callback	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Treat	-0.0325	-0.0358	-0.0265	-0.0295
	(0.0241)	(0.0240)	(0.0252)	(0.0265)
Panel B: Effect by Treatment Group				
Low Treat	-0.0201	-0.0222	-0.0161	-0.0195
	(0.0230)	(0.0240)	(0.0261)	(0.0265)
High Treat	-0.0455	-0.0496*	-0.0374	-0.0402
	(0.0274)	(0.0266)	(0.0269)	(0.0304)
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
Control Group Dep. Variable Mean	0.078	0.078	0.078	0.078

Notes: The dependent variable indicates an air sealing callback. 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. Models are estimated using 2SLS where randomized treatment assignment is an instrument for treatment jobs with eligible contractors. Standard errors clustered at the contractor level are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table D5: Effects of Bonus Treatments on Gas Usage (MMBtu) Contractor-Level Clustering

Gas MMBtu	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Weatherization × Treatment	-0.285**	-0.347***	-0.359***	-0.376***
	(0.126)	(0.124)	(0.124)	(0.126)
Panel B: Effect by Treatment Group				
Weatherization × Low Treat	-0.327**	-0.353**	-0.345**	-0.353**
	(0.155)	(0.152)	(0.152)	(0.155)
Weatherization × High Treat	-0.239*	-0.340**	-0.375***	-0.402***
	(0.128)	(0.129)	(0.130)	(0.130)
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	32	32	32	32
Observations	66423	66423	66423	63676
Baseline Weatherization Reduction	-1.582***	-1.582***	-1.582***	-1.597***
	(0.101)	(0.101)	(0.101)	(0.102)
Control Mean Pre-Weatherization Consumption	7.283	7.283	7.283	7.272

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. Standard errors are clustered at the contractor level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

E Effects on Callbacks for All Incentivized Tasks

Table E1 reports results for the effect of treatment on callbacks for all incentivized retrofits combined. These include the following retrofits, all of which have the potential to improve the blower door measure: thermal boundary, air sealing, rim insulation, windows and doors. While the results have somewhat larger standard errors, the treatment effects are consistent in magnitude with those in Table 3 for air sealing callbacks alone. This suggests contractors may be responding by doing all tasks that might impact blower door readings better.

Table E1: Effects on Callbacks: All Incentivized Tasks

Building Envelope	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Treat	-0.0233	-0.0276	-0.0192	-0.0231
	(0.0188)	(0.0191)	(0.0197)	(0.0201)
Panel B: Effect by Treatment Group				
Low Treat	-0.00301	-0.00731	-0.000580	-0.00909
	(0.0241)	(0.0246)	(0.0247)	(0.0248)
High Treat	-0.0444**	-0.0483**	-0.0390*	-0.0381
	(0.0212)	(0.0212)	(0.0221)	(0.0232)
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
Control Group Dep. Variable Mean	0.099	0.099	0.099	0.099

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 any "incentivized" task: i.e. thermal boundary, air sealing, attic insulation, wall insulation, crawl space insulation, basement insulation, rim insulation, windows and doors. 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. Heteroskedasticity-robust standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

F Robustness to Potential Cohort Effects

Recent econometric literature as identified potential biases from two-way fixed effects (TWFE) approaches (see, for example: Borusyak, Jaravel, and Spiess, 2021; Athey and Imbens, 2022; Goodman-Bacon, 2021; Strezhnev, 2018; Sun and Abraham, 2021; de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021). Given that treatment is randomly assigned continuously throughout the study period, our two-way fixed effects (TWFE) approach does not suffer from the near-term bias that can occur in staggered rollout settings where the proportion of the sample that is treated is increasing over time. With staggered rollout, TWFE estimators place more weight on portions of the sample with higher variance of the treatment indicator variable (i.e. typically at the middle of the panel).

Given that our estimates of the effects of treatment on household gas use use sample of billing data for projects for which we obtain a minimum of 12 months of pre/post billing data, sample weights for projects occurring earlier in the study period will be greater than those treated later. To the extent there is significant heterogeneity of treatment effects across time or groups of treated units, it could potentially bias our estimates of treatment on the gas use outcome (de Chaisemartin and D'Haultfœuille, 2020). The application process and the queuing system for timing of upgrades in the WAP, make it unlikely that there are substantial and systematic differences in outcomes across the sample period.

Nevertheless, we compare our estimates to an approach that weights the outcomes of each home equally. In Table F1, we estimate the treatment effects of the bonuses on gas consumption using home-specific monthly gas reductions (pre- minus post- weaterization) as the outcome. As with the estimates in the main text, we include homes with at least 12 months of both pre- and post-weatherization gas consumption data. The dependent variable was calculated as follows: (1) we computed the mean consumption for each home-calendar month in both pre- and post-weatherization periods, 2) we computed the annualized monthly mean for the pre- and post-weatherization periods using the mean across the 12 monthly means, 3) we computed energy savings using the difference in annualized monthly means. Because each home in our sample has just one observation with this approach, the estimates weight observations equally. The estimates are quite similar to those estimated with the TWFE approach. The pooled effect of the bonus treatments on gas consumption is 0.354 MMbtu, as compared to 0.376 from the TWFE approach. Given how close the estimates from the two approaches are, it is unlikely that significant heterogeneity in treatment effects across treated units over time are a biasing factor in our main analysis.

Table F1: Effects of Bonus on Gas Reduction (MMBtu): Annualized Monthly Mean

Gas Reduction	(1)	(2)	(3)	(4)
Panel A: Pooled Treatments				
Treat	0.240**	0.325***	0.388***	0.354***
	(0.117)	(0.114)	(0.116)	(0.116)
Panel B: Effect by Treatment Group				
Low Treat	0.231	0.341**	0.384***	0.370***
	(0.142)	(0.140)	(0.139)	(0.140)
High Treat	0.249*	0.307**	0.393***	0.336**
	(0.149)	(0.146)	(0.148)	(0.151)
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	1216	1203	1202	1146
Control Pre-Weatherization Blower Mean	3495.3	3497.9	3497.9	3489.9
Control Group Dep. Variable Mean	0.814	0.819	0.819	0.832

Notes: The dependent variable is mean monthly gas reduction (MMBtu). Homes with at least 12 months of both pre and post weatherization gas consumption data were included. The dependent variable was calculated as follows: Take the mean consumption for each home-calendar month both pre and post weatherization, 2) take the mean across the 12 monthly means for both pre and post, 3) energy savings is the difference in annualized monthly means. 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. Observations are weighted by the inverse probability of being in their respective treatment or control groups. Standard errors are clustered at the house level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

G Impacts on Electricity Usage

In this Appendix, we report estimates of the effects of treatment on electricity usage. While the energy efficiency retrofits mad in the IHWAP program are primarily designed to reduce energy consumption related to winter temperatures, air sealing and other retrofits could also have important effects on the consumption of electricity for cooling in the summer months.

Table G1 reports estimates of the effect of performance incentives on electricity usage. While point estimates in Panel A suggest evidence of a small reduction [-1.2 MMBtu] in monthly electricity consumption, we do not have sufficient power to detect small effect sizes. We note that the effect of summer cooling is small relative to baseload electricity consumption, such that fluctuations in monthly use result in a small change relative to variance in the monthly electricity data. Treatment effects are also not different from zero in the high [-6.6 MMBtu] or low [+3.4 MMBtu] bonus treatments reported in Panel B.

We compare the relative effects of treatment on gas usage to those for electricity usage by examining the ratios of baseline effects of weatherization (control mean weatherization effect) to the additional effect of the incentive treatments as reported in Tables 4 (gas) and G1 (electricity). Estimates in Table 4 (gas) indicate that the pooled effect of performance incentives (-0.376 MMBtu) is equivalent to 23% of the control mean weatherization effect (-1.6723 MMBtu). For electricity consumption, we estimate a control mean weatherization effect of -0.471 MMBtu. If we expect that the relative effect of the incentive treatment to baseline weatherization effects will be constant across the two fuel types, this would yield an expected effect of -10.9 MMBtu in electricity usage. We cannot rule out an effect of this magnitude on the basis of our -1.2 MMBtu [-11.3, 8.9] estimate.

Table G1: Effects of Bonus Treatment 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
Baseline Weatherization Reduction	-0.470***	-0.470***	-0.470***	-0.471***
	(0.0325)	(0.0325)	(0.0325)	(0.0313)
Control Mean Pre-Weatherization Consumption	2.772	2.772	2.772	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.

H 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 3) and the effects of completing additional treated/control projects contemporaneously with a given project on the reductions in CFM50 (Table 5) and gas use (Table 6) in that given project. In Tables H1 and H2, we report estimates of blower door and gas use effects broken out by high and low bonus treatments. Like the pooled sample test provided in Tables 5 and 6, we find no evidence of effects of additional simultaneous treatment or control jobs. In Table H3, 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. We do not find any statistical effect of additional incentivized contracts on the air leakage callback rate for a given home.

Table H4 reports the results of a ratio test for homogeneity of variances between households assigned to treatment (ITT) and households assigned to control for our two outcomes of interest: CFM50 reductions and gas reductions. To estimate gas reductions at the household-level, we first estimate the mean consumption for each calendar month both pre and post weatherization. We then take the mean across calendar months to get an annualized monthly average consumption for both pre and post weatherization. Finally, we subtract the post-weatherization annualized monthly average from the pre-weatherization annualized monthly average to get an average treatment effect. The first to columns in the table display the standard deviation (SD) of the house-level reductions for ITT and control. We report the F-Statistic and lower one-sided p-value for the alternative hypothesis that the ratio of the standard deviation of control to treatment is less than one. The results indicate that the standard deviation is statistically significantly higher for homes assigned to treatment relative to control. This is consistent with piece rates leading to more efficient allocation of effort than minimum quality standards as contractors will go further on lower marginal cost homes under the former regime, but not the latter.

Table H1: Effects on Building Envelope Tightness: Simultaneous Jobs Within Contractor

CFM50 (Post - Pre)	(1)	(2)	(3)	(4)
Low Treat	-60.34	-69.31*	-64.72	-74.76*
	(39.55)	(38.84)	(39.47)	(39.17)
High Treat	-68.44	-91.48**	-96.88**	-106.6***
	(42.69)	(41.55)	(41.81)	(40.90)
Low Treat × Simultaneous Treat Jobs	7.153	10.22*	10.12*	7.319
	(5.490)	(5.349)	(5.342)	(5.332)
Low Treat × Simultaneous Control Jobs	-10.81*	-11.35**	-11.59**	-8.394
	(5.895)	(5.605)	(5.607)	(5.637)
High Treat × Simultaneous Treat Jobs	0.722	2.962	1.498	-1.328
	(5.609)	(5.517)	(5.551)	(5.309)
High Treat × Simultaneous Control Jobs	-2.265	-3.021	-2.071	0.836
	(6.127)	(5.940)	(5.922)	(5.790)
Simultaneous Treat Jobs	-4.356	-5.211	-3.648	-0.0575
	(3.769)	(3.868)	(4.224)	(3.972)
Simultaneous Control Jobs	3.979	3.786	2.102	-1.023
	(4.330)	(4.170)	(4.593)	(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	1696	1600
Control Pre-Weatherization Blower Mean	3609.7	3609.7	3609.7	3585.2
Control Group Dep. Variable 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 H2: Effects on Gas Use: Simultaneous Jobs Within Contractor

Gas MMBtu	(1)	(2)	(3)	(4)
Weatherization × Low Treat	-0.335**	-0.373***	-0.374***	-0.394***
	(0.148)	(0.144)	(0.143)	(0.143)
Weatherization × High Treat	-0.229	-0.340**	-0.407***	-0.443***
	(0.147)	(0.144)	(0.151)	(0.156)
Weatherization × Low Treat × Simultaneous Treat Jobs	0.0248	0.0254	0.0261	0.0300
	(0.0195)	(0.0185)	(0.0186)	(0.0186)
Weatherization × Low Treat × Simultaneous Control Jobs	-0.0277	-0.0247	-0.0270	-0.0308
	(0.0209)	(0.0196)	(0.0199)	(0.0200)
Weatherization × High Treat × Simultaneous Treat Jobs	0.00785	0.0118	0.00996	0.0108
	(0.0243)	(0.0231)	(0.0233)	(0.0237)
Weatherization × High Treat × Simultaneous Control Jobs	-0.0127	-0.0148	-0.0125	-0.0155
	(0.0264)	(0.0247)	(0.0251)	(0.0254)
Weatherization × Simultaneous Treat Jobs	-0.0252**	-0.0243**	-0.0345**	-0.0341**
	(0.0114)	(0.0110)	(0.0134)	(0.0134)
Weatherization × Simultaneous Control Jobs	0.0271**	0.0259**	0.0369**	0.0393**
	(0.0133)	(0.0125)	(0.0151)	(0.0153)
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
r2_a	0	0	0	0
Baseline Weatherization Reduction	-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 H3: Air Leakage Callback Rate: Simultaneous Jobs Within Contractor

Reducing Air Leakage	(1)	(2)	(3)	(4)
Treat	-0.0371**	-0.0378**	-0.0221	-0.0261
	(0.0167)	(0.0163)	(0.0174)	(0.0175)
Treat × Simultaneous Treat Jobs	0.00136	0.000629	0.000913	0.000404
	(0.00283)	(0.00274)	(0.00273)	(0.00290)
Treat × Simultaneous Control Jobs	-0.00238	-0.00156	-0.00218	-0.00112
	(0.00312)	(0.00296)	(0.00299)	(0.00313)
Simultaneous Treat Jobs	0.000125	0.000941	0.00504*	0.00460
	(0.00232)	(0.00232)	(0.00269)	(0.00282)
Simultaneous Control Jobs	0.00281	0.00164	-0.00297	-0.00306
	(0.00267)	(0.00262)	(0.00302)	(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	-0.056	-0.106
Control Group Dep. Variable 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 H4: Test for Homogeneity of Variance

	SD Control	SD ITT	F-Statistic	P-value	Observations
Blower Door Reduction	1150.58	1235.36	.868	.0234	1567
Gas Reduction	1.83	1.99	.841	.0166	1216

Notes: This table reports the results of a ratio test for homogeneity of variances between households assigned to treatment (ITT) and households assigned to control. We report the F-Statistic and lower one-sided p-value for the alternative hypothesis that the ratio of the standard deviation of control to treatment is less than one.

I Evidence of Learning

In this Appendix, we test for evidence that the bonus treatment results in improvements in the performance of air sealing tasks completed across the study period. This could occur, for example, if the bonus incentives cause high or especially low-performing contractors to improve the way that they perform, supervise or internally inspect air sealing tasks. We construct a measure of the number of treated jobs completed prior to each job in the sample. We test for evidence of learning by examining the interaction between the "number of previously treated jobs" measure with the treatment indicator to examine whether contractors who have received a greater number of bonus contracts perform differently in their ability to induce reductions in CFM50 (Table I1) or gas use (Table I2). These tests do not reveal strong evidence that contractor performance is increasing as they complete larger numbers of treated or control jobs during the intervention.

Table I1: Effects of Bonus on Building Envelope Tightness by Number of Previously Treated Jobs

CFM50 (Post - Pre)	(1)	(2)	(3)
Panel A: Pooled Treatments			
Treat	-48.07	-65.94	-108.4**
	(52.84)	(53.04)	(51.31)
Treat × Number of Treated Jobs	-0.615	-0.354	0.807
	(1.440)	(1.438)	(1.456)
Number of Treated Jobs	-0.638	-1.075	-1.433
	(0.989)	(0.986)	(0.999)
Panel B: Effect by Treatment Group			
Low Treat	-20.03	-23.64	-69.08
	(67.06)	(67.59)	(67.37)
High Treat	-75.82	-107.2*	-147.2**
	(63.30)	(62.43)	(61.39)
Low Treat × Number of Treated Jobs	-1.354	-1.308	-0.133
	(1.783)	(1.809)	(1.876)
High Treat × Number of Treated Kobs	0.128	0.559	1.696
	(1.693)	(1.670)	(1.708)
Number of Treated Jobs	-0.638	-1.066	-1.430
	(0.989)	(0.987)	(0.999)
Pre Blower (CFM)	Yes	Yes	Yes
Expenditures	No	Yes	Yes
Characteristics	No	No	Yes
Observations	1698	1697	1601
Control Pre-Weatherization Blower Mean	3609.7	3609.7	3585.2
Control Group Dep. Variable Mean	-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 previous 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 I2: Effects of Bonus on Gas Usage (MMBtu) by Number of Previously Treated Jobs

Gas MMBtu	(1)	(2)	(3)
Panel A: Pooled Treatments			
Weatherization × Treatment	-0.275**	-0.342***	-0.349***
	(0.117)	(0.115)	(0.116)
Weatherization × Treat × Number of Treated Jobs	-0.00466	-0.00241	-0.00453
	(0.00546)	(0.00498)	(0.00516)
Weatherization × Number of Treated Jobs	0.00307	0.000880	0.00315
	(0.00394)	(0.00366)	(0.00376)
Panel B: Effect by Treatment Group			
Weatherization × Low Treat	-0.312**	-0.345**	-0.351**
	(0.142)	(0.139)	(0.139)
Weatherization × High Treat	-0.232	-0.337**	-0.347**
	(0.149)	(0.146)	(0.150)
Weatherization × Low Treat × Number of Treated Jobs	-0.00655	-0.00317	-0.00490
	(0.00620)	(0.00556)	(0.00582)
Weatherization × High Treat × Number of Treated Jobs	-0.00258	-0.00161	-0.00416
	(0.00692)	(0.00649)	(0.00672)
Weatherization × Number of Treated Jobs	0.00307	0.000886	0.00315
	(0.00394)	(0.00366)	(0.00376)
House FE	Yes	Yes	Yes
Month of Sample FE	Yes	Yes	Yes
Weatherization × Demeaned Pre Blower (CFM)	Yes	Yes	Yes
Weatherization × Expenditures	No	Yes	Yes
Weatherization × Characteristics	No	No	Yes
No. of Homes	1216	1216	1164
Observations	66423	66423	63676
Baseline Weatherization Reduction	-1.582***	-1.582***	-1.597***
	(0.101)	(0.101)	(0.102)
Control Mean Pre-Weatherization Consumption	6.910	6.910	6.879

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 previous treatment jobs 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.

J Measuring Contractor Quality

We construct a measure of contractor performance at baseline using a research design that leverages the queue-based assignment of contractors to jobs in the IHWAP program to estimate contractor-specific performance in energy savings achieved (Christensen et al., 2023). Whereas in Christensen et al. (2023), the measure is used to examine the extent that heterogeneity in contractor performance can explain the wedge between projected and realized energy savings in a quasi-experimental design, in the present study we use this performance measure and the experimental variation in bonus payments to examine heterogeneous responses to the piece-rate incentive.

To isolate the contribution of each contractor $\hat{\eta}_j$, we create a quality measure that measures the mean savings attributable to contractor j from all homes worked on during a baseline year. We deploy the approach using IHWAP data on jobs and energy outcomes for program year 2017, the year prior to the intervention. The estimates come from a procedure that first predicts the post-weatherization average monthly gas consumption for each home using characteristics of the home and household as well as pre-weatherization monthly gas consumption:

$$\hat{Y}_{iym} = \beta_1 X_i + \gamma_y + \mu_m + \epsilon_i \tag{J.8}$$

where Y_i is the energy use (MMBtu) for household i during the year prior to weatherization, X_i is a vector of home and household characteristics including county of residence, square footage, number of stories, number of bedrooms, year built and number of occupants. Calendar month and year fixed effects are denoted by μ_m and γ_y respectively. For each home, we then generate an estimate of mean differences in monthly energy use pre/post weatherization by computing the difference between predicted gas consumption in the post-period ($\hat{Y}_i$) from Eq. J.8 and and observed gas consumption in the post-period ($\bar{Y}_i$).

$$\hat{\delta}_i = \bar{Y}_i - \hat{Y}_i \tag{J.9}$$

In a final step, we estimate the contractor-specific contribution to each home's predicted energy savings ($\hat{\delta}_i$) using the following estimating equation:

$$\hat{\delta}_i = \beta_1 X_i + \beta_2 Z_i + \theta_j + \gamma_y + \mu_m \epsilon_i \tag{J.10}$$

where θ_j is a vector of contractor fixed effects, γ_y and μ_m again denote year and calendar month fixed effects respectively, and X_i is the vector of home and household characteristics from Eq. J.8. Z_i is a vector of indicators for binned ranges of spending for each retrofit category. The coefficient estimates on the contractor fixed effects, $\hat{\theta}_j$, capture the mean performance of each contractor j at baseline. The differences among the $\hat{\eta}_j$'s reflect mean contractor-specific differences in the energy savings.

In the absence of unobserved or uncontrolled for determinants of energy savings, these coefficients can be interpreted as a measure of contractor j's performance in achieving energy savings. Given the strong reliance on quasi-experimental variation in Christensen et al. (2023), we conduct several tests for evidence that differences in the coefficient estimates recovered from prior years could be due to unobservable variation (ϵ_j) across the homes to which contractors were assigned. On the basis of that evidence, we argue that the component of a contractor's outcome that is not attributable to quality is likely idiosyncratic in any given year. In the present experimental setting, randomization of treatment ensures that any unobservable variation in our contractor quality measure (ϵ_j) is not correlated with the bonus incentives.

Table K1: Comparison of Pooled Treatment Effects with Multiple Hypothesis Correction

Outcome	Treatment Coefficient	Unadjusted p-value	BKY sharpened q-value	BH q-value
Air Sealing Call Backs	-0.030	0.094	0.033	0.094
Blower Door Reduction	-88.673	0.005	0.005	0.007
Gas Consumption	-0.376	0.001	0.004	0.003

BKY sharpened q-values use the Benjamini, Krieger, and Yekutieli (2006) procedure implemented by Anderson (2008). BH q-values use the Benjamini and Hochberg (1995) procedure.

Table K2: Pooled Treatment Effects Controlling for Simultaneous Jobs with Multiple Hypothesis Correction

Outcome	Treatment Coefficient	Unadjusted p-value	BKY sharpened q-value	BH q-value
Air Sealing Call Backs	-0.026	0.135	0.048	0.135
Blower Door Reduction	-88.719	0.006	0.006	0.009
Gas Consumption	-0.415	0.001	0.002	0.002

BKY sharpened q-values use the Benjamini, Krieger, and Yekutieli (2006) procedure implemented by Anderson (2008). BH q-values use the Benjamini and Hochberg (1995) procedure.

K Multiple Hypothesis Testing

This appendix presents results from multiple hypothesis testing (MHT) corrections applied to the primary outcomes defined in the pre-registration: blower door readings, air sealing callbacks, and gas consumption. Tables K1–K3 report sharpened q-values using the procedures of Benjamini and Hochberg (1995, BH) and Benjamini, Krieger, and Yekutieli (2006, BKY), which control the false discovery rate (FDR) across sets of related hypotheses. Table K1 shows corrections for pooled treatment effects, Table K2 presents pooled effects controlling for simultaneous jobs, and Table K3 reports pairwise comparisons across the three primary outcomes—blower door readings, air sealing callbacks, and gas consumption—for each of the high and low treatment groups relative to control, as well as to each other.

Table K3: Comparison of High and Low Treatment Effects with Multiple Hypothesis Correction

Comparison	Coefficient	Unadjusted p-value	BKY sharpened q-value	BH q-value
Air Sealing Call Backs High vs Control	-0.04	0.05	0.08	0.11
Air Sealing Call Backs Low vs Control	-0.02	0.37	0.22	0.41
Air Sealing Call Backs High vs Low	-0.02	0.37	0.22	0.41
Blower Door Reduction High vs Control	-113.83	0.00	0.03	0.03
Blower Door Reduction Low vs Control	-66.17	0.09	0.12	0.16
Blower Door Reduction High vs Low	-47.67	0.30	0.22	0.41
Gas Consumption High vs Control	-0.40	0.01	0.03	0.03
Gas Consumption Low vs Control	-0.35	0.01	0.03	0.03
Gas Consumption High vs Low	-0.05	0.77	0.52	0.77

BKY sharpened q-values use the Benjamini, Krieger, and Yekutieli (2006) procedure implemented by Anderson (2008). BH q-values use the Benjamini and Hochberg (1995) procedure.

L Estimating Producer Surplus

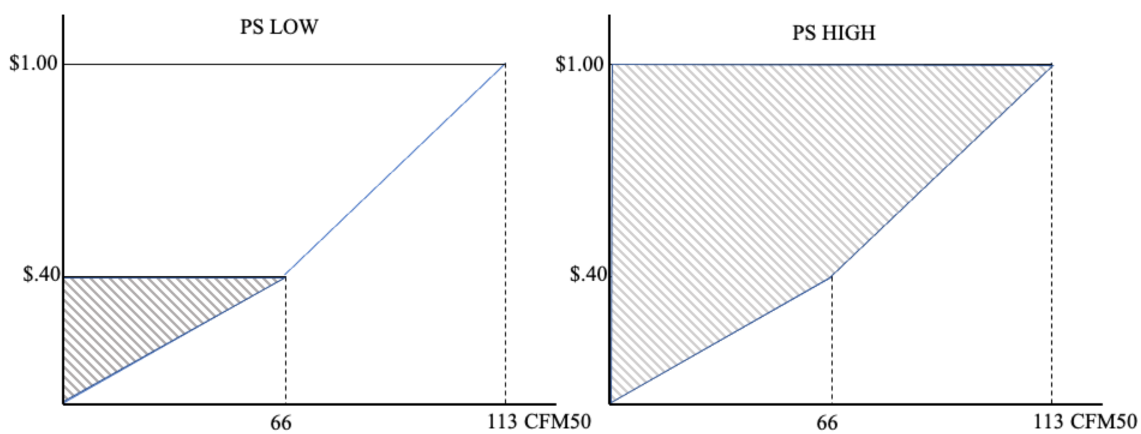
Equations 4 and 5 require estimates of changes in producer surplus that are attributable to the performance incentives. We recover estimates of producer surplus by tracing out the supply curve of air sealing improvements (CFM50 reductions) at the high and low bonus levels extending the model in Appendix B. The supply of CFM50 that is attributable to the piece rate at any given bonus level is:

$$Q_{CFM50} = f(b_{CFM50}(x^*_{CFM50} - \bar{x}_{CFM50})) \tag{L.11}$$

where Q_{CFM50} is the quantity of additional building envelope sealing (CFM50) observed in the program at any given level of bonus payment (b), which is a function of the optimal output of additional sealing by a given contractor beyond baseline target: x^*_i - $\bar{x}_i$. At any given level of bonus payment, b_{CFM50}, we observe Q_{CFM50} and can compute the level of producer surplus using:

$$PS = b_{CFM50}(Q^*_{CFM50}) - \int_{Q_{control}}^{Q^*_{CFM50}} \hat{\beta}_{CFM50}\, dx \tag{L.12}$$

We obtain estimates of $Q_{CFM50} = 66$ for the \$0.4 bonus and $Q_{CFM50} = 113$ for the \$1.00 bonus from Table 2. Assuming a piece-wise linear functional form as depicted below, the magnitude of the producer surplus is \$13 for the average contract at the low bonus level and \$99 for the average contract at the high bonus level.



M Social Welfare: Sensitivity Analysis

Table M1: Marginal Value of Public Funds

	Cost	Producer Surplus	*SCC* = $51		*SCC* = $121		*SCC* = $185	
			19.25 Years	34.5 Years	19.25 Years	34.5 Years	19.25 Years	34.5 Years
Panel A. Performance Incentive								
			Social Net Benefits					
Low Treat	$114	$13	$477	$773	$758	$1,228	$1,015	$1,644
High Treat	$283	$99	$543	$880	$863	$1,398	$1,155	$1,872
			Marginal Value of Public Funds					
Low Treat	.	.	$4.3	$6.89	$6.76	$10.88	$9.01	$14.53
High Treat	.	.	$2.27	$3.46	$3.4	$5.29	$4.43	$6.96
Panel B: Baseline WAP								
			Social Net Benefits					
All Baseline Retrofits	$9,615	.	$-5,184	$-2,584	$3,257	$10,621	$10,949	$22,655
Baseline Air Sealing	$1,128	.	$-608	$-303	$382	$1,246	$1,285	$2,658
Baseline All Incentivized	$2,037	.	$-1,098	$-547	$690	$2,250	$2,320	$4800
			Marginal Value of Public Funds					
All Baseline Retrofits	.	.	$0.46	$0.73	$1.34	$2.10	$2.14	$3.36
Baseline Air Sealing	.	.	$0.46	$0.73	$1.34	$2.10	$2.14	$3.36
Baseline All Incentivized	.	.	$0.46	$0.73	$1.34	$2.10	$2.14	$3.36
Panel C: Baseline Air Sealing + Incentive								
			Social Net Benefits					
Baseline Air Sealing + Low Treat	$1,242	.	$-232	$369	$1,039	$2,373	$2,198	$4,201
Baseline Air Sealing + High Treat	$1,411	.	$-249	$393	$1,061	$2,460	$2,256	$4,346
			Marginal Value of Public Funds					
Baseline Air Sealing + Low Treat	.	.	$0.81	$1.30	$1.84	$2.91	$2.77	$4.38
Baseline Air Sealing + High Treat	.	.	$0.82	$1.28	$1.75	$2.74	$2.60	$4.08
Panel D: Baseline All Incentivized + Incentive								
			Social Net Benefits					
Baseline All Incentivized + Low Treat	$2,151	.	$-722	$124	$1,347	$3,376	$3,233	$6,342
Baseline All Incentivized + High Treat	$2,320	.	$-739	$149	$1,369	$3,464	$3,291	$6,487
			Marginal Value of Public Funds					
Baseline All Incentivized + Low Treat	.	.	$0.66	$1.06	$1.63	$2.57	$2.50	$3.95
Baseline All Incentivized + High Treat	.	.	$0.68	$1.06	$1.59	$2.49	$2.42	$3.80
Panel E: Baseline All Retrofits + Incentive								
			Social Net Benefits					
Baseline All Retrofits + Low Treat	$9,769	.	$-4,821	$-1,925	$3,901	$11,735	$11,850	$24,185
Baseline All Retrofits + High Treat	$9,938	.	$-4,924	$-1,987	$3,837	$11,736	$11,821	$24,244
			Marginal Value of Public Funds					
Baseline All Retrofits + Low Treat	.	.	$0.50	$0.80	$1.40	$2.21	$2.22	$3.49
Baseline All Retrofits + High Treat	.	.	$0.50	$0.80	$1.39	$2.19	$2.19	$3.45

Notes: Table reports sensitivity in estimates reported in Table 10 to assumed lifespans of 19.25 years versus 34.5 years. 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.