



Energy Institute WP 297R

**The Welfare Costs of Misaligned Incentives:
Energy Inefficiency and the Principal-Agent Problem**

Joshua Blonz

Revised March 2021

Energy Institute at Haas working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to review by any editorial board. The Energy Institute acknowledges the generous support it has received from the organizations and individuals listed at <https://haas.berkeley.edu/energy-institute/about/funders/>.

© 2021 by Joshua Blonz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit is given to the source.

The Welfare Costs of Misaligned Incentives: Energy Inefficiency and the Principal-Agent Problem

Joshua A. Blonz*

March 2021

Abstract

In many settings, misaligned incentives and inadequate monitoring lead employees to take self-interested actions. This paper identifies and quantifies the costs of this principal-agent problem in the context of an energy efficiency appliance replacement program. I show that contractors (agents) hired by the electric utility (the principal) increase their compensation by intentionally misreporting program data to authorize replacement of non-qualified refrigerators. I estimate that unqualified replacements reduce welfare by an average of \$106 and save only half as much electricity as replacements that follow program guidelines. The same program without a principal-agent distortion would increase welfare by \$60 per replacement.

*Federal Reserve Board, 20th & C Street, NW, Washington, D.C. 20551, Email: joshua.blonz@frb.gov. The views in this paper are those of the author and do not necessarily represent the views or policies of the Board of Governors of the Federal Reserve System or its staff. I am deeply grateful to my advisor, Severin Borenstein, for his invaluable advice. I also thank Maximilian Auffhammer, Becka Brolinson, Fiona Burlig, Peter Christensen, Lucas Davis, Aluma Dembo, Meredith Fowle, Sylvan Herskowitz, Sandile Hlatshwayo, Koichiro Ito, Erica Myers, Karen Palmer, Louis Preonas, Catherine Wolfram, and seminar participants at the NBER Summer Institute, AERE 2019, and the Energy Institute at Haas Energy Camp for their thoughtful and helpful comments. I thank the California Public Utilities Commission Public Advocates Office for providing data for this study.

1 Introduction

The notion that economic agents respond to incentives is central to economics. The principal-agent problem is a canonical example of the power of incentives, in which self-interested agents may act against the interests of the principals. A standard formulation of the principal-agent problem involves a principal employing an agent to conduct a task where the principal cannot observe the agent’s effort. In this setting, performance-based payments (rather than a fixed wage) offer the potential to align agents’ and principals’ incentives, by compensating agents proportionately to their output. Performance pay has been shown to increase productivity in cases where agents’ output is easily observable (Lazear 2000; Shearer 2004), but there are many cases where the principal cannot observe output directly. When performance pay is used in these settings, agents have the incentive to undertake unobservable actions to maximize their compensation at the expense of the principal (Baker 1992).

The challenge of imperfectly observed output is exacerbated if workers can exert effort on multiple dimensions or across many tasks (Kerr 1975; Holmstrom and Milgrom 1991), which applies to a wide range of settings including teacher pay based on test scores (Jacob and Levitt 2003), worker retraining programs (Courty and Marschke 1997, 2003), and commission-based sales behavior (Tayan 2016). While compensation structures with misaligned incentives can result in welfare losses, it has been difficult to empirically evaluate their consequences, because agent actions are typically unobservable.

In this paper, I exploit a unique opportunity to observe agents taking self-interested actions at the expense of a principal, which enables me to estimate the welfare costs of the principal-agent problem. In this empirical setting—an appliance upgrade program for low-income households—the principal does not directly monitor the agent, but I can observe how agents adjust their behavior in response to the level of incentive on one of their outputs. I develop a principal-agent framework to model how agents increase their compensation by disobeying guidelines set by the principal. Combining this modeling framework with empirical estimates of agents’ misreporting behavior and outcome data, I find that the principal-agent problem turns an otherwise welfare-increasing program into a welfare-reducing program.

I analyze data from 180,000 home energy efficiency retrofits conducted between 2009 and 2012 as part of the California Energy Savings Assistance (ESA) program, which provided free appliance replacements to low-income households. The electric utility Southern California Edison (SCE; the principal) hired contractors (the agents) to enroll households in the program, assess which upgrades each home was eligible for, and install upgrades.¹ By exploiting changes in program eligibility rules, I show that contractors intentionally misreported the age of refrigerators to increase the number of replacements they were able to provide, thereby increasing their compensation. I estimate that 13.8 percent of ESA enrollees were qualified for a refrigerator replacement, while 27.3 percent were

1. SCE runs the ESA program at the direction of the California Public Utilities Commission (CPUC). For simplicity, I refer to SCE as the principal because it directly administers the program, but the CPUC shares responsibility as the principal.

determined to be eligible by the contractors. This implies that contractors misreported data for 13.5 percent of the households that participated in the program, resulting in about half of the refrigerator replacements conducted in the ESA program not qualifying based on program rules.

Next, I focus on how contract structure—which created variation in the financial incentive to misreport—affected the misreporting rate. Eleven of the 22 contractors in the ESA program conducted all the steps in the upgrade process, including enrolling households, determining refrigerator replacement eligibility, and (on a separate visit) installing the replacement refrigerators. I estimate that these “integrated-task” contractors had an average incentive to misreport of \$123 per household, which resulted in these contractors misreporting data for 19 percent of the households they enrolled. The other 11 contractors enrolled households and determined refrigerator eligibility, but were *not* responsible for refrigerator replacement. I estimate that these “separated-task” contractors had an average incentive to misreport of only \$25 per household, which resulted in these contractors misreporting data for 7.8 percent of the households they enrolled. In short, I find that the incentives created by the contractor compensation structure directly influenced misreporting and the size of the principal-agent distortion.

To evaluate how intentional misreporting affects program outcomes, I use data on monthly household-level electricity consumption before and after an upgrade. I combine this with model number data collected during assessments to classify refrigerator replacements conducted in the program as qualified or unqualified. I find that qualified replacements reduce household electricity consumption by 73 kilowatt hours (kWh) per month, whereas unqualified replacements save only 38 kWh per month. I then calculate the key components of the changes in welfare resulting from qualified and unqualified refrigerator replacements. On average, qualified replacements increase welfare by \$60, while unqualified replacements *decrease* welfare by \$106. Contractor misreporting therefore substantially undermined the benefits of the ESA program.

Finally, I combine my theoretical framework and empirical estimates to evaluate trade-offs between integrated-task contracts and separated-task contracts. Integrated-task contracts could be preferable if there were significant coordination benefits from having a single contractor conduct both initial assessments and physical replacements. However, the empirical results imply that the costs of misaligned incentives outweigh any potential coordination benefits. I find that integrated-task contracts reduce welfare by an average of \$12.32 more per assessment than separated-task contracts. The relative welfare benefits of integrated-task contracts would need to be \$95 per refrigerator replacement, or 135 percent above the average payment to contractors to conduct an assessment, in order to rationalize integrated-task contracts over separated-task contracts. The reduction in benefits from a separated-task contract is equal to 16 percent of the total electricity reduction benefits from the 25,000 refrigerator replacements conducted by integrated-task contractors, and demonstrates it would have been preferable to use the separated-task contract to implement the ESA program.

This paper makes four contributions. For one, to the best of my knowledge, this is the first paper to estimate the welfare costs of a principal-agent distortion using detailed outcome

data. The existing empirical principal-agent literature primarily identifies distorted agent behavior but has been unable to evaluate the welfare effects of these distortions (e.g., Jensen and Murphy 1990; Oyer 1998; Courty and Marschke 2004, 2008; Marion and West 2019).² This is because of insufficient outcome data and difficulty in identifying a valid counterfactual. In this paper, I am able to overcome these obstacles using policy-driven variation in agent incentives and electricity consumption data to estimate the welfare costs of misaligned incentives between principals and agents. Of course, the estimates do not necessarily generalize to other contexts. However, my unique setting provides general insights on the role of contract structure in the principal-agent problem and specific insights on energy efficiency programs.

Second, I leverage variation in contract types to identify how contract structure influences both agent behavior and overall welfare outcomes. Most of the previous work in this area of the contracts literature has focused on the impact of performance incentives compared with fixed wages (e.g., Asch 1990; Lazear 2000; Shearer 2004; Bandiera, Barankay, and Rasul 2007; Copeland and Monnet 2009; Lewis and Bajari 2011, 2014).³ In contrast, my results highlight how even within a given payment scheme, contract structure can distort incentives (all contractors in the study are paid a piece-rate wage) and reduce welfare.

Third, I develop a model that extends the existing principal-agent framework to consider trade-offs between the coordination benefits of integrated-task contracts and the costs of increasing the incentive to misreport. Using this model, I calculate under which circumstances the principal (as the social planner) prefers an integrated-task contract. The model links the contracts literature, which focuses on how performance-based pay affects efficiency, and the empirical principal-agent literature, which studies how agent incentives can distort behavior. I then use the model to provide novel evidence of how contract structure can interact with agent incentives to reduce welfare.

Finally, this is the first paper to identify and quantify the mechanism and incentives behind a principal-agent problem and demonstrate their impact on an energy efficiency program. The closest paper is Giraudet, Houde, and Maher (2018), which shows lower energy efficiency savings from home retrofits completed on Fridays. The authors attribute this to moral hazard and negative productivity shocks to workers at the end of the week. In my paper, I am able to directly observe agents misreporting and measure how that affects program outcomes. My findings add to a growing number of papers that have identified mechanisms behind the underperformance of energy efficiency programs and standards, including car manufacturers gaming of fuel economy standards (Reynaert and Sallee 2021), the landlord-tenant problem (Myers 2020), heterogenous workmanship, and biased engineering models (Christensen et al. 2021).

Energy efficiency programs have become increasingly popular in the United States, with energy efficiency subsidies totaling \$8.2 billion in 2016 (CEE 2018). However, rigorous empirical analysis has consistently demonstrated that energy efficiency subsidies deliver lower energy savings than expected (Metcalf and Hassett 1999; Davis, Fuchs, and Gertler 2014; Graff Zivin and Novan

2. See Prendergast (1999) and Holmstrom (2017) for a survey of the empirical principal-agent distortion literature.

3. See Lazear (2018) for a survey of the performance-based compensation literature.

2016; Allcott and Greenstone 2017; Burlig et al. 2020; Fowlie, Greenstone, and Wolfram 2018). In this paper, I provide the first empirical evidence for a cause of the non-performance of energy efficiency programs: the principal-agent problem. I am also able to quantify the relationship between realized energy savings and contract structure. The findings are of direct policy relevance. A CPUC decision on the ESA program cited this paper and stressed the goal of reducing or eliminating the principal-agent problem I identify (CPUC 2019). More broadly, energy efficiency is a critical component of most policies to mitigate carbon emissions and combat climate change (Loftus et al. 2015; Fowlie, Greenstone, and Wolfram 2018). While the ESA program’s contract structure may exacerbate the principal-agent problem, many energy efficiency retrofit programs use contract structures similar to that of the ESA program. This suggests that the principal-agent problem likely contributes to lower-than-expected savings in many energy efficiency settings.

The principal-agent problem I study is not unique to energy efficiency. Most doctors in the United States receive compensation to diagnose patients and provide treatment. Research suggests that this could lead to the over-provision of medical services, potentially contributing to higher healthcare costs (Hillman and Goldsmith 2010; Larkin and Loewenstein 2017). A similar incentive structure exists for automobile mechanics. Schneider (2012) shows that mechanics frequently recommend expensive repairs even when there is an alternative low-cost solution.⁴ The ubiquity of the principal-agent problem highlights the importance of considering contract structure and agents’ incentives when designing public policy.

The rest of the paper is organized as follows: Section 2 describes energy efficiency policy and the ESA program. Section 3 introduces the model. Section 4 provides evidence of contractor misreporting. Section 5 estimates the changes in electricity consumption from qualified and unqualified refrigerator replacements. Section 6 conducts a welfare analysis, calculates the benefits of an integrated-task contract using an expression derived in Section 3, and calculates a benefit-cost ratio. Section 7 concludes.

2 Empirical Setting

2.1 Energy Efficiency Policy

Energy efficiency subsidy programs are typically considered a win-win policy intervention because energy efficiency investments are projected to reduce energy expenditures enough to more than offset their up-front costs, with the added bonus of reducing greenhouse gas emissions and local air pollution created by energy use. The majority of energy efficiency programs are mandated by state regulators, administered by the local electricity or gas utility, and funded by a surcharge on utility bills. Most of the programs take the form of subsidized energy-efficient appliance replacements or building retrofits.⁵

4. The California smog check program attempts to address this particular principal-agent problem by using a separated-task framework that utilizes some “test-only” stations that cannot conduct repairs.

5. See Allcott and Greenstone (2012), Gerarden, Newell, and Stavins (2017), and Gillingham, Keyes, and Palmer (2018) for reviews of energy efficiency programs and the academic research on their effectiveness.

Another policy tool used to increase energy efficiency is federal minimum efficiency standards. In the United States, most major appliances—including refrigerators—must satisfy a minimum level of efficiency. Before 1990, the average refrigerator consumed 127 kWh/month and represented one-quarter of total household electricity consumption (EIA 1987). A similar refrigerator consumed only 65 kWh/month after the standards were tightened starting in 1993.⁶ These binding standards have been responsible for progressively increasing energy efficiency, as the fleet of refrigerators is replaced over time.

2.2 ESA Program

The ESA Program provides free home energy efficiency upgrades to 400,000 low-income California households annually. I focus on the ESA program administered by SCE, which is one of the largest utilities in the United States and has about 100,000 ESA participants each year (SCE 2012b).⁷ The program is mandated by the CPUC, which sets guidelines and funds the policy with a per-kWh charge on all electricity sold in California.⁸

Between 2009 and 2012, SCE hired 22 geographically dispersed contractors to assess who was eligible for the program and provide upgrades. The energy efficiency retrofits proceeded in the following steps:⁹

Step 1) SCE reached out to eligible households through mail or phone calls. Interested households signed up to start the enrollment process.

Step 2) SCE gave household contact information to an ESA contractor. The contractor scheduled a visit with the household to conduct a home assessment.

Step 3) A contractor visited the household to verify that it was low income to complete the program enrollment. The contractor then assessed eligibility for refrigerator replacement and any other major upgrades.

Step 4a) If a household was not eligible for any major upgrade, no action was taken and the contractor left.

Step 4b) If a household was eligible for a major upgrade, the contractor provided up to five energy-efficient light bulbs and scheduled a second appointment to replace the refrigerator.¹⁰

Step 5) The contractor provided the refrigerator replacement and any other major upgrades.

6. In Southern California, the average household uses 450 kWh/month, which means a refrigerator manufactured before 1993 consumed up to 20 percent of a household's electricity. See Appendix A for a description of federal minimum efficiency standards.

7. See Appendix C.1 for a description of program eligibility criteria, a full list of upgrades the ESA program provides.

8. The utility does not earn a rate of return on the investments made through the ESA program.

9. To supplement my understanding of the ESA program, I participated in a number of ride-alongs. Appendix C.2 discusses some of the insights I gained from directly observing the program.

10. ESA program rules dictated that the contractor could install light bulbs only if the household qualified for a major upgrade.

In this paper, I focus on the contractor behavior around the assessment and replacement of refrigerators, which were the most common major upgrade in the program. Of households that enrolled in the ESA program, 23.1 percent received a refrigerator replacement, which totaled one-fourth of program expenditures and almost one-half of the program’s projected energy savings (SCE 2012a). The contractor assessed refrigerator eligibility in Step 3 by determining the year the refrigerator was manufactured.¹¹ A household was eligible for a free refrigerator replacement if its oldest working refrigerator was manufactured in 1992 or earlier. This eligibility cutoff was chosen because federal minimum efficiency standards tightened in 1993. As a result, replacing a refrigerator manufactured in 1992 or earlier is projected to save more electricity than replacing a refrigerator manufactured in 1993 or later.

There were two types of contractors that implemented the ESA program between 2009 and 2012. Half of the 22 contractor firms were able to conduct all five steps in the ESA program, from assessment to installing a major upgrade. These contractors, which I call “integrated-task” firms, were responsible for 50 percent of the assessments and they conducted 100 percent of the refrigerator replacements in my data. The other 11 contractors, which I call “separated-task” firms, were able to conduct only steps 1 through 4 of the ESA program—importantly, they did not replace refrigerators—and conducted the other 50 percent of the assessments. If a separated-task contractor determined a household was eligible for a refrigerator replacement, one of the 11 integrated-task firms provided that replacement.

Contractors were paid a fixed piece rate for every service they provided. If a household was eligible for a major replacement, contractors were paid \$25 to install five compact fluorescent light bulbs (CFLs) at the end of the assessment step. Contractors were paid \$224 to remove the old refrigerator for disposal and replace it with a new energy-efficient unit. Separated-task contractors, who never installed replacement refrigerators, received only the \$25 payment if CFLs were installed. Integrated-task contractors received both payments.¹² In most cases, a household only received CFLs when the contractor conducting the assessment determined that the dwelling was eligible for a refrigerator replacement. This program design gave separated-task contractors the incentive to find a household eligible for a refrigerator replacement so that they could install the CFLs and receive the \$25 payment.

While the ESA program did conduct audits during the period studied, these audits were not able to prevent unqualified refrigerator replacements resulting from contractor misreporting. In particular, the ESA program used audits to verify that contractors provided the replacements they billed to the program. The audits were completed after refrigerators were replaced, too late to check the refrigerator’s manufacture year. Program guidelines set by the regulator did not require verification of the reported manufacture year, and SCE did not implement its own monitoring

11. The manufacture year can be determined by looking at the serial number, which is usually on a plaque inside the main compartment. Contractors recorded information for all refrigerators in the dwelling.

12. All households were given the same energy-efficient refrigerator. The new unit was meant to be approximately the same size as the one that was removed. Of refrigerator eligible households, 15.8 percent did not receive the replacement because of not wanting the replacement or logistical challenges. See Appendix D.7 for more details.

system. As a result, contractors could intentionally misreport that an ineligible household was eligible for a refrigerator replacement without much chance of being caught.

Contractors in the ESA program were paid a piece rate to incentivize them to conduct as many upgrades as possible (Research into Action 2011). This compensation structure, combined with the lack of monitoring, created the principal-agent problem where the principal wanted the contractors to replace only qualifying refrigerators, but contractors had the incentive to misreport and provide ineligible replacements. Misreporting was particularly costly in this context, because the 1992 eligibility cutoff was chosen to maximize the savings from replacing the least efficient refrigerators. It is possible that adjusting the contract structure or monitoring could have reduced the incentive to misreport; however, it is outside the scope of this paper to make this determination.¹³

3 Model

I consider a principal-agent model where the principal is an electric utility working under the supervision of the regulator, and the agent is a contractor hired by the utility. The electric utility may not directly care about the effectiveness of the ESA program. However, I assume the utility wants the program to function without misreporting to avoid reprimand or punishment from the regulator.¹⁴ The principal hires an agent to provide services to a large number of households. Not all households qualify for the ESA program, and the principal cannot directly observe eligibility. Therefore, an agent must assess eligibility of customers on behalf of the principal before services can be provided. The contract proceeds in two steps. In the first step, “assessment,” the principal hires an agent to assess n potential recipients’ eligibility. In the second step, “service,” the principal hires an agent to provide the service to participants who were deemed eligible in the assessment step.

Agents receive a piece-rate wage based on the number of assessments and services they provide. In the assessment step, the agent receives a fixed fee, A , for each assessment she completes. If an agent reports that a potential recipient does not qualify, no services are provided. In addition to the flat fee, the agent receives a “referral payment,” d , for each recipient she decides qualifies for the service.¹⁵ In the service step, an agent receives g for each service she provides. Both d and g are net of the marginal cost of providing the referral and service. The model explicitly considers that

13. It would have been possible, for example, to require contractors to take a picture of the plaque that had the refrigerator’s serial and model number on it during the assessment step. Duflo et al. (2013) used an experiment that combined changing the contract structure with monitoring. The authors find that the intervention improved the effectiveness of pollution regulation in India.

14. I assume the regulator, which in this case is the CPUC, wants to maximize the welfare benefits of the ESA program.

15. In the ESA program, the referral payment is the \$25 payment a contractor receives for installing five CFLs. The referral payment can take different forms in different settings. In a medical context, the referral payment captures that when an orthopedist determines that a patient needs a service (e.g., an MRI), the patient will end up returning for a follow-up appointment.

in the assessment step, the agents may increase their compensation by misreporting unqualified potential recipients as qualified.¹⁶

I consider two different contracts that a principal could offer to an agent in this setting. In the first, the principal uses an integrated-task contract, where the same agent conducts the assessment and service steps. In the integrated-task setup, the agent is the residual claimant on any rents from service provision, giving her an increased incentive for misreporting. The second contract arrangement is a separated-task contract, where the assessment and service steps are done by two different agents. By separating the tasks between two agents, the assessment agent has a lower incentive to intentionally misreport that a potential recipient qualifies. However, the assessment agent still has a small incentive to misreport on a separated-task contract, because the referral payment for qualified recipients occurs at the assessment stage.¹⁷ Separating the contract between different agents is costly. The agent must pay an “information acquisition cost,” F , to complete her task during both the assessment and service steps.¹⁸ In the integrated-task contract, the information acquisition cost is paid only once during the assessment visit because the same agent is responsible for both steps. The two agents cannot pass information in the separated-task contract, which requires the service provision agent to also pay the information acquisition cost for every recipient of the service. This extra information cost is the main disadvantage of a separated-task contract relative to an integrated-task contract.¹⁹ The principal’s objective is to pick a contract structure that maximizes welfare knowing that the agent may intentionally misreport. There are many homogeneous agents that compete for a contract. The principal sets g , d , and A , which are driven by market rates and can vary across the two contract types. Because these payments are transfers between the principal and the agent, they do not factor into the final welfare calculation.²⁰

The principal can observe only whether a contractor provided a service. To match the empirical setting I assume monitoring is not feasible or too costly for the principal to distinguish between a service provided to a qualified recipient and one resulting from false reporting in the assessment step. The principal is aware that the agent may misreport, but the only action the principal can take to punish an agent for suspected intentional misreporting is to cancel the contract with that agent for future assessments and services. For both types of contracts, I assume that agents are risk neutral and maximize expected contract value. By comparing the total expected contract

16. In the model, agents know when they are misreporting, and they do not make mistakes in their reporting. I rule out any other types of agent misreporting, such as reporting that an assessment or service was provided to a potential recipient when it was not. Potential recipients do not influence the misreporting decision or take actions in the model.

17. See Hirao (1993), Lewis and Sappington (1997), and Khalil, Kim, and Shin (2006) for alternate formulations of this optimal task design problem in different contexts.

18. In the ESA program, the information acquisition cost includes the costs associated with gaining the occupant’s trust to schedule an appointment and enter the dwelling and learning about the occupants’ appliances and their electrical system. During a ride-along, I observed a qualified refrigerator replacement that could not occur because the assessment contractor did not correctly note that the occupant did not have a grounded electrical socket.

19. There may be other, smaller advantages of an integrated-task contract. For example, if the principal hires fewer agents in the integrated-task contract, this could reduce transaction costs.

20. Under this assumption, payments for misreported services are not considered welfare costs. If the principal assigns a cost to compensating agents for misreported services, then the current formulation is a lower bound on the cost of misreporting.

surplus under each contract type, I derive the conditions under which the principal prefers the integrated-task contract.²¹

3.1 Integrated-Task Contract

Under an integrated-task contract, the principal offers the agent a contract to both assess potential recipients and provide the service to those that qualify. When a potential recipient is deemed qualified during the assessment visit, the same agent schedules a second visit to provide the service. For all qualified potential recipients, even if the agent intentionally misreported them as eligible, the integrated-task agent provides the service and receives g .

There are two types of potential recipients. A portion α of potential recipients qualify for the service under program rules.²² The principal and the agent have a shared but noisy expectation about the value of α . The expected number of qualified recipients is $n \cdot \alpha$, and the expected number of unqualified participants is $n \cdot (1 - \alpha)$. Subject to participation, the agent chooses a level of misreporting $Z_I \in [0, 1]$ to maximize her expected payout, which is a static decision the agent makes at the beginning of the contract.

$$E(\max_{Z_I} \pi_I) = \underbrace{n\alpha(g + d) + n(1 - \alpha)(g + d)Z_I}_{\text{Referral and service provision payments}} - \underbrace{n(1 - \alpha)C(Z_I)}_{\text{Cost to agent of misreporting}} + \underbrace{n(A_I - F)}_{\text{Assessment payment minus information acquisition cost}}. \quad (1)$$

The I subscript indicates the integrated-task contract. The cost to the agent of misreporting is $C(Z)$, which captures the disutility of misreporting and the effort spent doing that misreporting. $C(Z)$ is a function of the difference between the proportion of recipients that received a new refrigerator and the regulator's expected proportion of eligible households, α .²³ The misreporting cost function is what deters agents from misreporting on every job they conduct, even though the principal is not able to monitor agent actions. Both $C'(Z_I)$ and $C''(Z_I)$ are positive, which reflects the idea that as misreporting rates rise, the chance of the principal catching the agent increases at an increasing rate. A higher chance of being caught by the principal increases the agent disutility of misreporting and the amount of effort the agent exerts to misreport. The principal engages in a simple form of monitoring where it compares the expected population eligibility rate, α , with the proportion of assessments that the agent deems eligible. There is no specific threshold at which the principal will detect misreporting and punish the agent, but the agent perceives the risk of the principal catching her as increasing with her misreporting rate.²⁴ Following Baker (2002), I use a

21. See Appendix B.2 for a discussion of the risk-neutrality assumption and other assumptions in the model.

22. In the ESA program, α reflects the proportion of ESA-qualified households with a refrigerator manufactured in 1992 or earlier.

23. The misreporting cost includes agent efforts spent on misreporting and avoiding detection. The agent misreporting cost is a welfare cost. See Appendix B.3 for a discussion of this assumption and an alternate version of the model where the agent misreporting cost is an expected fine the agent pays to the principal if she is caught misreporting.

24. For example, if the principal has an expectation that 13 percent of the population qualifies and observes 16 percent of the assessed population receiving the service, it likely would not suspect agent misreporting. If, however, 55 percent of the assessed population receives the service, then the principal is much more likely to terminate the

functional form of $C(Z_I) = \frac{1}{2}\gamma_I Z_I^2$ for the misreporting cost function.²⁵ The parameter γ_I converts the misreporting rate into dollars for the integrated-task contract.

Plugging the misreporting cost function into equation (1), I derive the optimal level of agent misreporting Z_I^* using the first-order condition as follows:

$$\frac{\partial \pi_I}{\partial Z_I} = n(1 - \alpha)(g + d) - n(1 - \alpha)\gamma_I Z_I^* \quad (2)$$

$$Z_I^* = \frac{g + d}{\gamma_I}. \quad (3)$$

Equation (3) shows that the optimal misreporting rate is equal to the benefits of misreporting ($g + d$), scaled by γ_I . This equation can be used to derive an estimate for γ_I when the values of Z_I^* , g , and d are known.

I next consider the expected value of this contract to the principal while taking into account the optimal level of agent misreporting, Z_I^* . The value to the principal of the integrated-task contract is as follows:

$$E(V_I^{\text{principal}}) = \underbrace{w_1 n \alpha}_{\text{Welfare benefits of qualified service provision}} + \underbrace{w_2 n (1 - \alpha) Z_I^*}_{\text{Welfare costs of unqualified service provision}} - \underbrace{n(1 - \alpha)(g + d) Z_I^* - n \alpha (g + d) - n A_I}_{\text{Payments to agent for assessment, referral, and service provision}}. \quad (4)$$

w_1 and w_2 are the welfare impacts of a qualified and an unqualified service provision, respectively. Both w_1 and w_2 include the marginal cost to the agent of the referral and providing the service.²⁶ For simplicity, I assume $w_1 \geq 0 > w_2$, which means that a qualified service is weakly welfare improving and an unqualified one is welfare reducing.²⁷

To calculate the welfare impact of the contract, I add together the value to the principal (equation (4)) and the profit of the agent (equation (1)):

$$E(\text{Surplus}_I) = \underbrace{w_1 n \alpha}_{\text{(1) Welfare benefits of qualified service provision}} + \underbrace{w_2 n (1 - \alpha) Z_I^*}_{\text{(2) Welfare costs of unqualified service provision}} - \underbrace{\frac{1}{2} n (1 - \alpha) \gamma_I Z_I^{*2}}_{\text{(3) Cost to agent of misreporting}} - \underbrace{n F}_{\text{(4) Agent information acquisition cost}}. \quad (5)$$

Equation (5) is the main formulation of the surplus from the integrated-task contract divided into four components. Part (1) captures the welfare benefits from qualified service provision. Part (2) is the welfare loss from unqualified service provision. It is always negative, since $w_2 < 0$. Part (3) is

contract because of suspected misreporting. The agent is aware of this, and therefore has increasing disutility and cost at higher levels of misreporting.

25. Baker (2002) uses the functional form to model the cost of effort for the agent. I adapt it to this setting to model the cost of misreporting.

26. I choose to have w_1 and w_2 include the marginal cost of service provision because it yields a simpler expression of contract surplus. The results are unchanged if the marginal cost of providing the service is a separate term.

27. A more general formulation could allow $w_1 > w_2 > 0$ where w_2 includes the opportunity cost of providing additional higher benefit qualified services.

the cost of agent misreporting. The last term is the information acquisition cost paid by the agent at every assessment.

3.2 Separated-Task Contract

Under the separated-task contract, the “assessment agent” is responsible for the assessment, and a different “service agent” conducts the service step. The assessment agent’s problem is similar to the integrated-task problem in equation (1). The assessment agent can still intentionally misreport that a potential recipient household qualifies when it does not, but the assessment agent no longer receives g for providing services. The referral payment d continues to provide an incentive for the assessment agent to misreport.²⁸ The optimal level of misreporting for the assessment agent in the separated contract is

$$Z_S^* = \frac{d}{\gamma_S}. \quad (6)$$

The S subscript indicates the separated-task contract. I assume that the misreporting rate is lower in the separated-task contract than in the integrated-task contract. The service agent provides the service to all the potential recipients that are deemed qualified by the assessment agent and receives g .²⁹

The remainder of the separated-task problem is similar to the integrated-task case, and its derivation is shown in Appendix B.1. The surplus from the separated-task contract is

$$\begin{aligned}
 E(\text{Surplus}_S) = & \underbrace{w_1 n \alpha}_{(1) \text{ Welfare benefits of qualified service provision}} + \underbrace{w_2 n (1 - \alpha) Z_S^*}_{(2) \text{ Welfare costs of unqualified service provision}} - \underbrace{\frac{1}{2} n (1 - \alpha) \gamma_S Z_S^{*2}}_{(3) \text{ Cost to agent of misreporting}} \\
 & - \underbrace{nF}_{(4) \text{ Assessment agent information acquisition cost}} - \underbrace{F[n(1 - \alpha) Z_S^* + n\alpha]}_{(5) \text{ Service agent information acquisition cost}}.
 \end{aligned} \quad (7)$$

The main difference between the separated-task surplus (Surplus_S) and the integrated-task surplus (Surplus_I) is term (5), which captures that in the separated-task contract, the information acquisition cost must be paid a second time during the service provision. In the next section, I calculate the relative benefit of an integrated-task contract by comparing the surplus from each contract type.

28. I rule out any side deals between the assessment and service agents. I do not find any evidence of side deals between contractors in the ESA program.

29. The ESA program has the same contract structure, except there is no dedicated service agent. Instead, the integrated-task contractor is responsible for providing all the upgrades, even if she did not conduct the assessment. The incentive to misreport in the assessment step is the same in the model and in the ESA program.

3.3 Comparing the Two Contracting Approaches

The principal's objective is to maximize welfare by choosing the preferred contract structure for the setting. To determine which contract the principal prefers, I compare the surplus in the two contract types by taking the difference between Surplus_I and Surplus_S to calculate the relative benefits of an integrated-task contract:

$$\begin{aligned}
 E(\text{Surplus}_I - \text{Surplus}_S) = & \underbrace{w_2[n(1 - \alpha)(Z_I^* - Z_S^*)]}_{\substack{(1) \\ \text{Welfare costs from} \\ \text{increased unqualified service} \\ \text{provision with integrated contract}}} + \underbrace{-n(1 - \alpha)\frac{1}{2}(\gamma_I Z_I^{*2} - \gamma_S Z_S^{*2})}_{\substack{(2) \\ \text{Higher misreporting cost to agent} \\ \text{in integrated contract}}} \\
 & + \underbrace{F[n(1 - \alpha)Z_S^*]}_{\substack{(3) \\ \text{Information acquisition} \\ \text{cost in separated contract} \\ \text{for unqualified service provision}}} + \underbrace{Fn\alpha}_{\substack{(4) \\ \text{Information acquisition} \\ \text{cost in separated contract} \\ \text{for qualified service provision}}} \tag{8}
 \end{aligned}$$

Equation (8) decomposes the factors that determine the welfare consequences of an integrated-task contract relative to a separated-task contract. I break it down into four main parts. Term (1) captures the welfare costs of higher misreporting rates in the integrated-task contract. Term (1) always reduces the relative value of an integrated-task contract because $w_2 < 0$ and $Z_I^* > Z_S^*$. Term (2) is the difference in the misreporting cost to the agent between the integrated- and separated-task contracts. It always reduces the relative value of an integrated-task contract, because agents misreport more under the integrated-task contract, leading to a higher misreporting cost. Terms (3) and (4) capture the extra information acquisition cost that contractors pay in the separated-task contract. They show how information acquisition raises the cost of the separated-task contract for qualified *and* unqualified replacements, which increases the relative value of the integrated-task contract.

In summary, the first two terms in equation (8) reduce the relative value of an integrated-task contract, and the last two terms increase the relative value of an integrated-task contract. The model does not predict which set of terms dominates or whether the integrated- or separated-task contract is preferred by the principal. The model is applicable to many real-world situations where agents are responsible for both determining eligibility and providing the service. For example, many doctors use an integrated-task contract, whereby they are paid to both diagnose and treat a patient's illness (Berenson and Rich 2010). Similarly, auto mechanics both assess what is wrong with a car and receive payment to conduct the repair (Hubbard 1998). The model provides a general framework to calculate the welfare benefits of an integrated-task contract relative to a separated-task contract. In this paper, I apply the model to the ESA program to determine which contract type is welfare maximizing using empirical values derived in the following sections.

4 Intentional Misreporting

In this section, I identify and quantify the principal-agent distortion in the ESA program. I first look at how contractors intentionally misreport ineligible households as qualified for a refrigerator replacement. I use changes in program rules to develop a counterfactual against which I compare contractor behavior to identify intentional misreporting. I next use variation in contract type to examine how the incentive to misreport affects the size of the principal-agent distortion. Finally, I use the results to calculate the integrated-task and separated-task misreporting rates (Z_I and Z_S) used in the model.

4.1 ESA Program Data

I use two main datasets in this paper. First, I use confidential data from the SCE ESA program on assessments and installations. My main sample consists of 271,126 households that enrolled in the ESA program through SCE between January 2009 and December 2014.³⁰ The data include the reported refrigerator manufacture year and model number, along with information on the date of each visit, what upgrades or measures were installed, and which contractor performed each service.

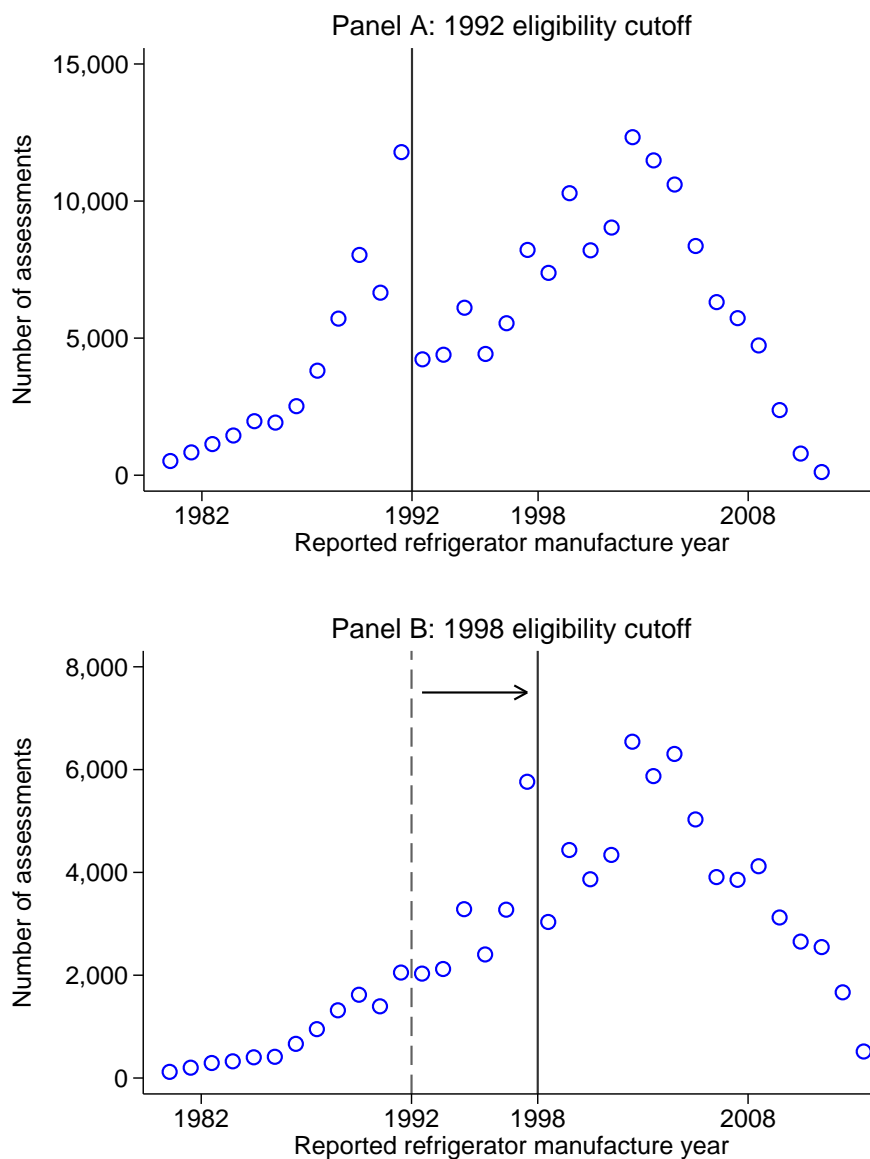
4.2 Overall Misreporting

In this section, I investigate whether contractors intentionally misreport refrigerator manufacture years. Panel A of Figure 1 shows the distribution of refrigerator ages *reported* by contractors on the 180,105 assessments they conducted between January 2009 and the end of August 2012. Households to the left of the vertical line—those with refrigerators reported to be manufactured in 1992 or earlier—qualified for a free refrigerator replacement in the ESA program, while households to the right, with reported refrigerator manufacture year of 1993 or later, did not. Overall, contractors reported finding a qualified refrigerator in 27.3 percent of assessments.

Panel A shows substantial bunching to the left of the 1992 threshold. About 12,000 assessments had a reported refrigerator manufacture year of 1992 compared with just over 4,000 reported for 1993. Next, I test whether the bunching is due to contractor manipulation or reflects the true distribution of assessed refrigerators. It could be the case, for example, that households with pre-1993 refrigerators disproportionately enrolled in the program. To examine alternate explanations, I leverage a change in the ESA refrigerator replacement program. Starting in September 2012, the refrigerator eligibility cutoff year moved from 1992 to 1998. Panel B of Figure 1 shows the distribution of contractor-reported refrigerator ages for the 91,021 assessments conducted between September 2012 and December 2014. The solid vertical line indicates the new cutoff, and the dashed line represents the old 1992 cutoff. In Panel B, the bunching to the left of the 1992 cutoff is no longer present. Instead, bunching appears to the left of the 1998 threshold, with almost 6,000 assessments having a reported refrigerator manufacture year of 1998, compared with just 3,000

30. See Appendix D.1 for details on the data cleaning process.

Figure 1: Reported Refrigerator Manufacture Years with 1992 and 1998 Eligibility Cutoffs



Note: Panel A of this figure shows the *reported* refrigerator age from 180,105 assessments conducted between January 2009 and August 2012. Households are given free replacement refrigerators if their existing units are reported to be manufactured in 1992 or earlier. Of these assessments, 27.3 percent have reported refrigerator ages to the left of the 1992 cutoff, indicated by the solid vertical line. Panel B shows the reported refrigerator ages from 91,021 assessments conducted between September 2012 and December 2014. These households were eligible for replacements if their existing units were manufactured in 1998 or earlier, indicated by the solid vertical line. The dashed line indicates the previous 1992 refrigerator replacement eligibility cutoff. Both panels show bunching of reported refrigerator manufacturer years to the left of the eligibility cutoff.

manufactured in 1999. The bunching is smaller around the 1998 cutoff in Panel B than the 1992 cutoff in Panel A. The reduction in misreporting is likely because moving the cutoff caused a higher proportion of assessed households to be qualified for a refrigerator replacement, which could change a contractor’s misreporting decisions.

One possible explanation for the bunching in Panel A of Figure 1 is that low-income households with refrigerators manufactured in 1992 or earlier enrolled in the early phase of the ESA program at high rates knowing they could get free replacements. There are four reasons this is unlikely. First, the 1992 eligibility cutoff was not advertised. A potential enrollee could have determined the refrigerator eligibility cutoff with some searching, but a large portion of the eligible population would have to have been informed about the cutoff to have generated this bunching pattern.

Second, the change in bunching from the 1992 cutoff to the 1998 cutoff happened discontinuously in September 2012, the month the eligibility cutoff shifted. Figure 2 shows the timing of this changeover. The vertical axis reports the ratio of 1992 assessments to 1993 assessments conducted each month, which is a measure of bunching at the 1992 cutoff. The month the eligibility cutoff shifted to 1998, the assessments no longer exhibited this bunching to the left of the 1992 eligibility cutoff. It would have been difficult for enrolling households to learn about the changing program guidelines and adjust their behavior so rapidly, suggesting that it is contractors—who were aware of the eligibility shift in advance—adjusting their reporting behavior to the new guidelines.³¹

Third, if newly eligible households did respond to the eligibility shift shown in Panel B, one would expect to see a jump in the number of 1993–1997 refrigerators. Instead, the only visible discontinuous increase is in refrigerators reported to be manufactured in 1998.

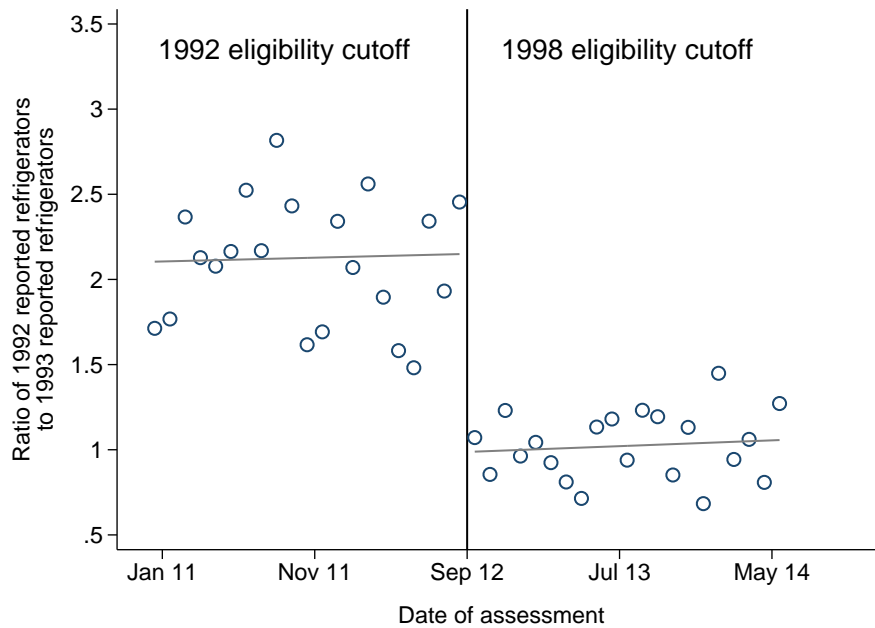
Fourth, the neighboring utility San Diego Gas & Electric (SDG&E) does not exhibit the same bunching in assessments to the left of the 1992 eligibility cutoff. The only difference between the two programs is that SDG&E used a different contract structure that does not have the same incentives to misreport during the assessment step.³² Both utilities were subject to the same rules from the regulator, and the eligible households live in similar regions. Figure 3 displays data from 106,179 assessments conducted during the same time period by contractors hired by SDG&E. The lack of bunching to the left of the 1992 eligibility cutoff suggests that it is the incentives of the contractors in the SCE program, not the behavior of the enrolling households, that are responsible for the bunching in Figure 1.³³

31. Appendix D.2 discusses this changeover in more detail and formally estimates a regression discontinuity in time at the changeover date.

32. SDG&E had one contractor firm that was responsible for administering the program and conducting all the assessments. Appendix D.3 provides a further discussion of assessment contractor incentives in the SDG&E ESA program.

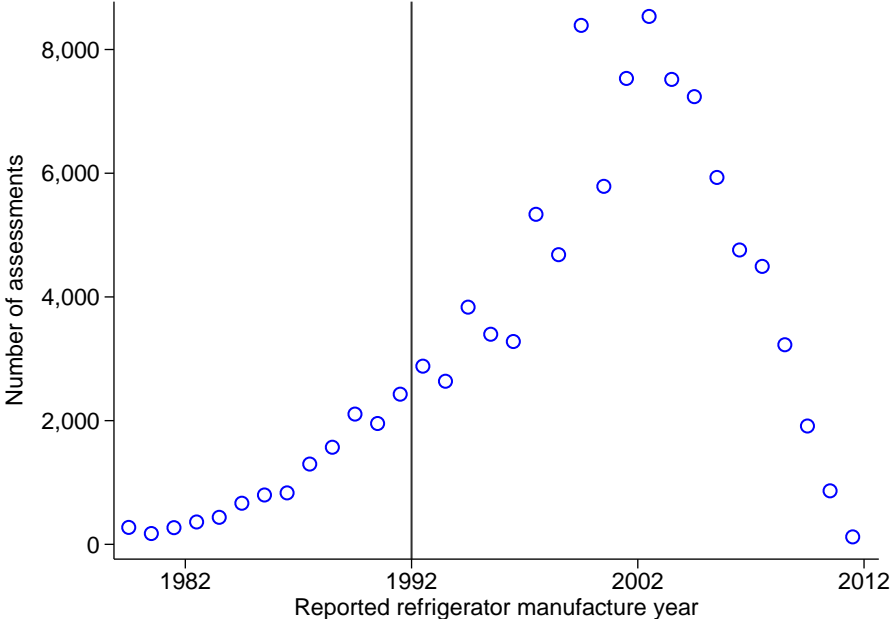
33. I do not use the SDG&E data to estimate misreporting because the two programs were run differently and serve different populations.

Figure 2: Contractor Reporting Behavior and the Change in Refrigerator Eligibility



Note: This figure shows how contractors adjust their behavior in response to the eligibility cutoff shifting from 1992 to 1998. The vertical axis reports the ratio of 1992 assessments to 1993 assessments each month—a measure of bunching around the 1992 cutoff. Each dot represents about 200 assessments, and the horizontal lines are from a local linear regression described in Appendix D.2. The figure shows that immediately after the eligibility cutoff shifted to 1998, contractors began reporting an equal number of 1992 and 1993 refrigerators. The reporting behavior changed the month after the policy shift, which suggests that it was contractor intentional misreporting that caused the bunching to the left of the 1992 threshold. The discontinuity is significant with a coefficient of -1.16 and a standard error of .19.

Figure 3: ESA Program Administered by Neighboring Utility SDG&E



Note: This figure shows ESA assessment data from 106,179 assessments in the neighboring utility (SDG&E), where contractors did not have an incentive to intentionally misreport during the assessment step. The data were collected during the same time period as in Panel A of Figure 1, but there is no bunching to the left of the 1992 eligibility threshold.

4.3 Estimating Misreporting

The ESA program’s change in eligibility requirements in September 2012 provides an opportunity to estimate the proportion of the ESA population with refrigerators eligible for upgrades at the time of the January 2009 to August 2012 assessments. Establishing a credible estimate for the proportion of eligible refrigerators is central to quantifying contractor misreporting. To generate this counterfactual, I use assessment data from 2013, after the refrigerator eligibility cutoff was moved to 1998 and contractors no longer had the incentive to misreport newer refrigerators as being manufactured in 1992 or earlier.³⁴ Figure 2 shows that starting in September 2012, the reported manufacture years no longer bunch around the 1992 cutoff, as was observed in the assessments collected between January 2009 and August 2012. Instead, contractors reported 1992 and 1993 refrigerators in nearly equal proportions.

However, the 2013 data is likely to include fewer refrigerators from 1992 or earlier due to naturally occurring replacements that take place over time independent of the ESA program. Using the distribution of refrigerator manufacture years from 2013 as a counterfactual for the 2009 to 2012 assessments may lead to an underestimate of the number of eligible refrigerators present at that time. To account for refrigerator retirements, I adjust the 2013 data by subtracting one from each reported manufacture year in the 2013 data, which shifts the full distribution down by a year.³⁵ I use this approach because it is simple and transparent. A single year adjustment to the 2013 data may not account for the full difference between the two samples, but it is a conservative assumption, and a larger adjustment might overstate the misreporting rate.³⁶ The result is a counterfactual distribution of refrigerator ages in the ESA population. I can then calculate rates of misreporting by comparing this counterfactual to the distribution of ages recorded in the 2009 to 2012 assessments.³⁷

Figure 4 shows the observed distribution of the 2009 to 2012 data with blue hollow dots and the counterfactual distribution with green solid dots. The lower area, shaded light green, is the implied distribution of qualified refrigerators among ESA eligible households based on the 2013 assessment data. It shows that only 13.8 percent of households in the ESA program should have qualified for a replacement. This is in contrast to the 27.3 percent of households that were reported by contractors to be eligible. The gray shaded area represents the implied misreported refrigerators, which total 13.5 of the assessments conducted between 2009 and 2012.

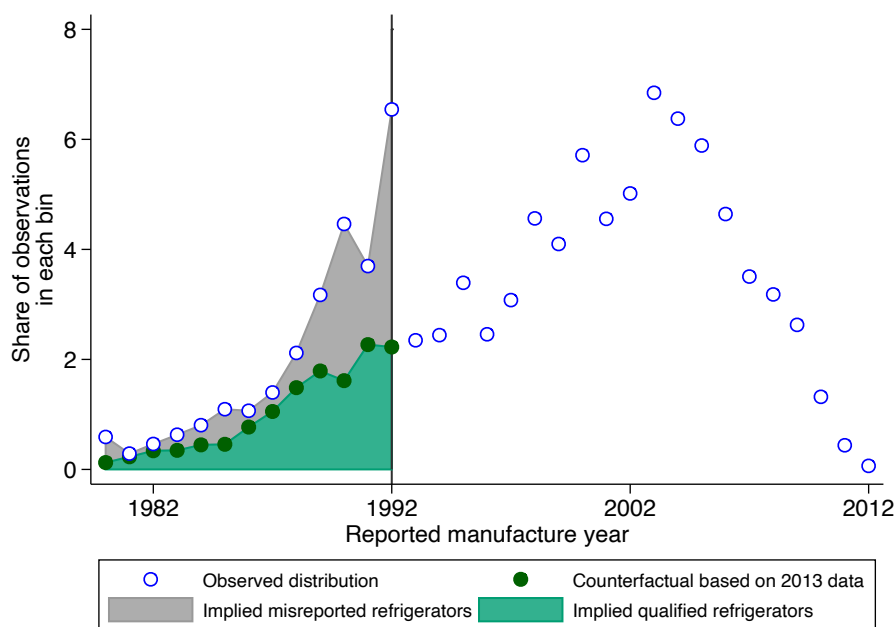
34. After the eligibility change, contractors have the incentive to report a refrigerator was manufactured in 1998 or earlier, which removes the incentive to misreport around the old 1992 cutoff.

35. I considered other more complex approaches to calculating a counterfactual; however, each requires additional assumptions without clear benefits to the quality of the resulting estimates.

36. The one-year adjustment increases the proportion of eligible refrigerators by 2.2 percentage points. A two-year adjustment would increase the proportion of predicted eligible refrigerators by an additional 2.4 percent. See Appendix D.4 for a detailed discussion of the adjustment.

37. I find no evidence in Section 4.2 of selection into the program for people with refrigerators manufactured in 1992 or earlier. If people with qualifying refrigerators were early adopters of the ESA program, then the counterfactual could underpredict the amount of 1992 or earlier refrigerators in the 2009 to 2012 study window, which would make the misreporting estimate a lower bound.

Figure 4: Reported Refrigerator Manufacture Years and Counterfactual Based on 2013 Data



Note: This figure shows the reported refrigerator ages with blue hollow dots and the counterfactual based on 2013 data with green solid dots. The observed distribution is the same as Panel A of Figure 1 with the vertical axis rescaled to the share of observations in each bin. The light green area below the green dots is the implied distribution of qualified refrigerators among ESA eligible households and is equal to 13.8 percent of assessments conducted between 2009 and 2012. The gray area represents the implied misreported refrigerators. It is the difference between the observed distribution and the counterfactual distribution and totals 13.5 percent of the assessments conducted. The counterfactual is calculated using data collected in 2013 when contractors did not have the incentive to misreport around the old 1992 eligibility threshold.

I test the statistical significance of the 13.5 percent misreporting rate using a t-test to estimate whether the proportion of assessments reporting a manufacture year of 1992 or earlier is the same in the observed and counterfactual distributions. I account for correlation within each firm by clustering at the firm level. Because there are only 22 firms in my sample, I use the wild cluster bootstrap, finding that the misreporting rate is significant at the 1 percent level (Cameron, Gelbach, and Miller 2008).³⁸

The shape of the distributions in Figure 4 provide two additional insights. First, the majority of the misreporting was done with assessments indicating a manufacture year of 1989 to 1992. This could reflect contractors preferring to misreport within a few years of the 1992 threshold.³⁹ Second, the calculated counterfactual appears to line up with the observed distribution on the right side

38. I implement the t-test using a regression of a dummy variable equal to one if the assessment reported the refrigerator was manufactured in 1992 or earlier on a dummy of the observation being from the 2013 counterfactual period.

39. If a contractor was caught, a smaller difference between the true age and misreported manufacture year may make it easier to claim that the misreporting was accidental. I find anecdotal evidence of contractors preferring smaller misreporting jumps in a subsample where I can precisely identify misreporting behavior.

of the 1992 cutoff. This is not a necessary condition for a reasonable counterfactual because the misreported refrigerators are drawn from the years above 1992, which could lead to a lower share of reported manufacture years in 1993 than the 1992 counterfactual. The smoothness through the 1992 cutoff suggests that contractors misreported refrigerators from many years above the 1992 cutoff.

This approach also allows for quantification of the misreporting without relying on statistical manipulation tests (e.g. McCrary (2008) and Frandsen (2017)).⁴⁰ I use the 2013 data to construct a counterfactual with few functional form assumptions or expectations about the shape of the misreporting. Figure 4 suggests contractors are misreporting refrigerators to be manufactured as far back as 1989. Statistical manipulation tests may not effectively capture this type of misreporting so far from the threshold and would likely undercount the misreporting. My counterfactual also captures contractors taking shortcuts and misreporting earlier refrigerators (e.g. pre 1990) as being manufactured in 1992, while statistical manipulation tests could count this behavior as additional misreporting.

Another advantage of using the 2013 data to calculate the counterfactual is that it accounts for unintentional misreporting that occurs due to error. If these errors are random and unrelated to the 1992 eligibility cutoff, they should, on average, occur at the same rate in the 2013 data. The error rate in the predicted distribution will be the same as in the reported distribution, allowing estimates of unintentional misreporting to be netted out. This allows me to attribute the estimated misreporting to the principal-agent problem and contractor incentives to maximize their compensation, not to errors in the assessment process.

4.4 Model Numbers and Strategic Contractor Behavior

SCE instructed contractors to record the refrigerator model number during an assessment. The model number data, when accurately reported, could be used to verify the reported refrigerator manufacture year.⁴¹ However, contractors omit the model number 10.1 percent of the time. I examine the extent to which omitted model numbers are correlated with misreporting.

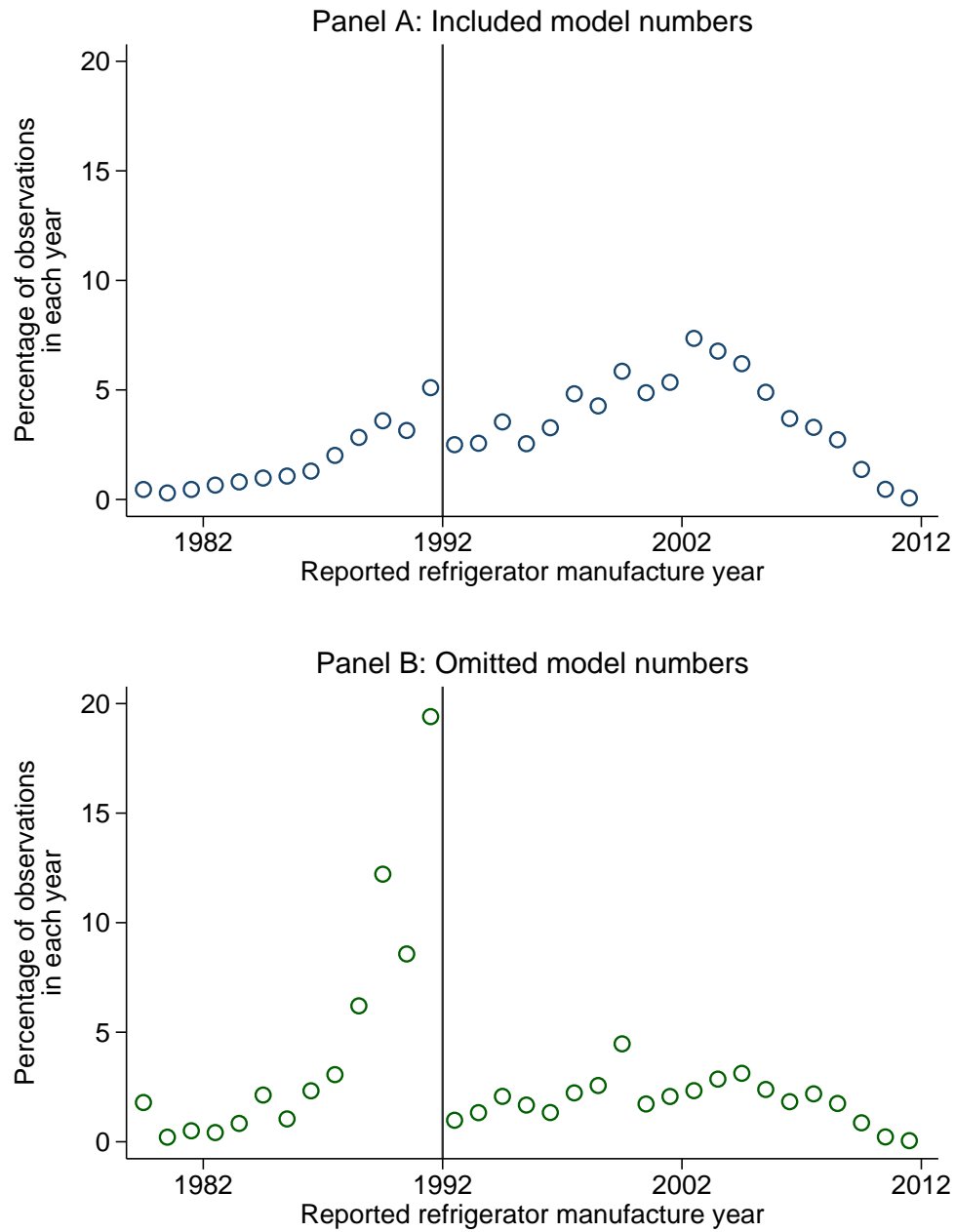
Panel A of Figure 5 shows the 161,955 assessments where contractors reported model numbers. There is less bunching to the left of the 1992 cutoff, with only 23.6 percent of households eligible for a replacement. The smaller distortion implies a misreporting rate of 9.7 percent, which is significant at the 1 percent level, and is 3.8 percentage points lower than the overall sample. This shows that when contractors report a model number, they are less likely to intentionally misreport on an assessment.

Panel B shows the 18,150 assessments where contractors did not record model numbers. I find more bunching of assessments to the left of 1992 cutoff than in the full sample, with 60.9 percent

40. These tests use smooth approximations of the variable of interest near a threshold to test for manipulation. They are commonly used to test the validity of regression discontinuity designs, but Frandsen (2017) can also be used to estimate the fraction of missing observations at the threshold.

41. SCE has yet to use the model numbers for this purpose, but the potential for retrospective auditing could affect contractor behavior.

Figure 5: Assessments with Included versus Omitted Model Numbers



Note: This figure shows the distribution of reported refrigerator manufacture years based on whether the contractor reported a model number on the assessment. Panel A shows the 161,955 assessments where contractors reported model numbers, which has an implied misreporting rate of 9.7 percent. Panel B shows the 18,150 assessments where contractors omitted the model numbers, which has an implied misreporting rate of 47.1 percent. The vertical line in both panels indicates the 1992 refrigerator replacement eligibility cutoff. The vertical axis reports the percentage of observations in each year to facilitate comparison across the panels.

of these assessments having a reported refrigerator manufacturer year before this date. The large distortion implies a misreporting rate of 47.1 percent, which is significant at the 1 percent level, and suggests that when contractors intentionally misreport, they are also more likely to omit the model numbers.⁴² About 22 percent of refrigerator replacements in the ESA program were conducted based on assessments that did not report a model number, even though they represented only one-tenth of the assessments.

One explanation for the high level of implied misreporting in Panel B is that the model numbers are missing from the refrigerators. Over time, model number labels could become hard to read, which would lead to more pre-1993 refrigerators having model numbers that are not visible. However, once the eligibility threshold shifted from 1992 to 1998, the percent of assessments with omitted model numbers that reported manufacture years before 1993 dropped from 60.9 to 18.3. This discontinuous 42.6 percentage point drop suggests that the change in the eligibility cutoff caused the contractors to adjust their model number omission behavior. The fact that contractors choose to disproportionately omit model numbers when misreporting the refrigerator manufacture years provides further evidence that they are trying to increase their compensation while minimizing the chance of being detected.

4.5 The Occupant and Misreporting

The occupants of ESA upgraded dwellings were required to be home during ESA assessments, but they were not told the criteria for a refrigerator replacement.⁴³ However, occupants had a large incentive to receive a free refrigerator and they might have encouraged contractors to misreport. Contractors could also have solicited bribes to misreport. If occupants were playing a role in the misreporting decisions, it would not change my analysis of misreporting. However, it is useful to understand how occupants could affect misreporting.

To test for occupant influence, I compare the misreporting rates for homeowners and renters. Homeowners had a larger incentive to encourage misreporting because they own the replacement unit. Renters, in contrast, only benefit from using the refrigerator while they live in the rental unit. Landlords do not play a role in the assessment process that would give them the opportunity to bribe contractors.⁴⁴ After accounting for different implied qualified refrigerator rates for renters and owners, I find that contractors misreported assessments for 13.7 percent of renters and 13.5 percent of homeowners, which are not statistically different from each other, and suggests that homeowners were not influencing the contractors' misreporting decisions.⁴⁵

42. All the contractors engaged in this behavior; the results are not driven by a small number of contractors.

43. On the ride-alongs I participated in, the contractors worked quickly and did not explain what the dwelling was eligible for during the assessment.

44. Landlords were only notified that an ESA upgrade was going to occur after the assessment was conducted, which makes it unlikely they would be able to change the outcome with a bribe.

45. I calculate separate implied qualified refrigerator rates using the same approach used in Section 4.3 broken down by renters and owners. I find 14.9 percent of renters are qualified and 11.3 percent of homeowners are qualified. The higher rate of replacement qualification for renters suggests that a landlord-tenant problem might be leading to less efficient refrigerators in rental dwellings (Myers 2020).

Renters and homeowners also appear to behave similarly during other parts of the ESA process. Logistical constraints, such as scheduling a follow-up appointment or technical issues, results in only 84.2 percent of households found eligible for a refrigerator replacement receiving one.⁴⁶ Renters and owners have similar completion rates of 84.4 percent and 83.6 percent, respectively. While not definitive, the comparable behavior and outcomes for owners and renters in the ESA program suggests that the occupant does not play a significant role in the misreporting decision.

4.6 Contract Structure and Misreporting

Next, I leverage the two types of contractors that SCE used to implement the ESA program to identify the intentional misreporting rates for the integrated- and separated-task assessments (Z_I and Z_S). Figure 6 shows all the assessments, split by whether the assessment was conducted by a separated-task contractor (i.e., one of the 11 firms that performs *only* assessments) or an integrated-task contractor (i.e., one of the 11 firms that *also* performs replacements). Panel A shows the distribution for 90,084 integrated-task contractor assessments, and Panel B shows the distribution for 90,021 separated-task contractor assessments. The bunching of reported refrigerator manufacturer years is larger for the integrated assessments, with 32.9 percent of all assessments having a recorded refrigerator age of 1992 or earlier compared with 21.8 percent for the separated assessments. Both of these are higher than the predicted eligibility rate of 13.8 percent, suggesting that both types of contractors misreport. I find a misreporting rate of 19.0 percent for the integrated contracts, which corresponds to Z_I in the model.⁴⁷ The misreporting rate for separated contracts (Z_S) is less than half the size, at 7.8 percent, which confirms the assumption in the model that $Z_S < Z_I$ and reflects that separated-task contractors only receive \$25 for misreporting.⁴⁸ Both of the misreporting estimates are significant at the 1 percent level and I can reject that integrated- and separated-task contractors misreport at the same rate.

Households are assigned to contractors in a quasi-random manner. When SCE passes along the contact information of households that are enrolling in the SCE program to contractors (Step 2 in Section 2.2), it does so using a round-robin algorithm based on where the household is located and which contractors have available capacity.⁴⁹ This assignment mechanism prevents contractors from targeting specific types of households for ESA upgrades.⁵⁰

46. See Appendix D.7 for a detailed discussion of the refrigerator replacement completion percentage.

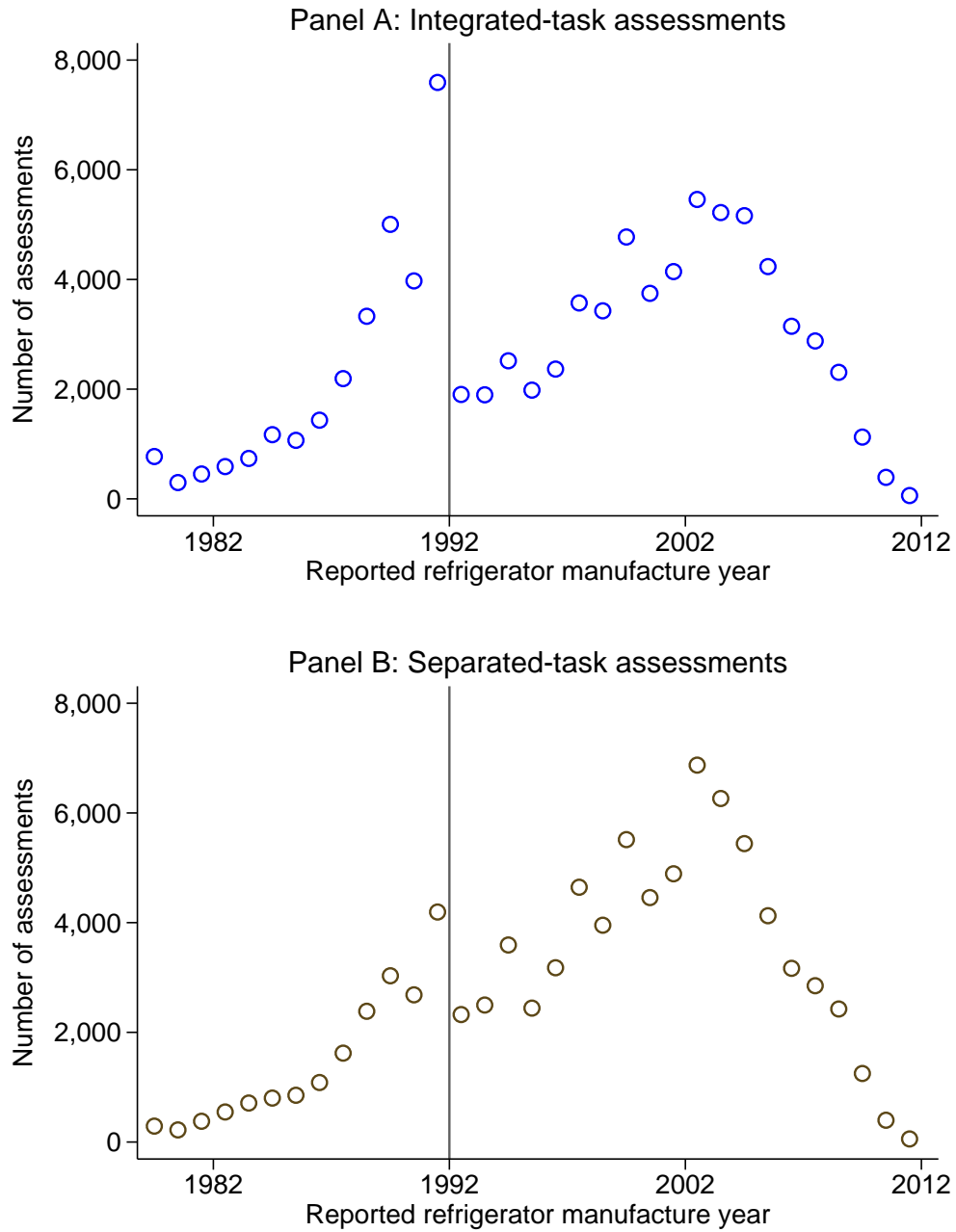
47. There is a range in misreporting rates across contractors. Some separated-task contractors misreport on over 15 percent of their assessments, while some integrated-task contractors misreport on less than 12 percent. I report the implied misreporting rate for each of the anonymized 22 contractors in Appendix D.5.

48. Separated-task contractors did not know who would conduct the upgrade, making it difficult to negotiate side payments. I verify the absence of side payments by examining the relationship between separated-task contractor misreporting and who conducts the upgrade, finding no pattern.

49. The round-robin contract-assignment process is described in more detail in Appendix D.6.

50. I find that, conditional on being declared eligible during the assessment step, refrigerator replacements are completed at the same rate for integrated- and separated-task contractors. I also find that there are no major differences across contract types or whether the dwelling is rented versus owned. See Appendix D.7 for a detailed breakdown.

Figure 6: The Effect of Contract Type on Misreporting



Note: This figure shows the reported refrigerator manufacture year by which type of contractor conducted the assessment. Panel A shows the 90,084 assessments that were conducted by integrated-task assessments. Integrated-task contractors have an implied misreporting rate of 19.0 percent. Panel B shows the 90,021 assessments that were conducted by separated-task contractors. Separated-task contractors have an implied misreporting rate of 7.8 percent. The vertical line in both panels indicates the 1992 refrigerator replacement eligibility cutoff.

However, contract type in the ESA program is not randomly assigned. SCE signs contracts with existing firms in their service territory based on their capacity. Some firms may have the capital and workforce to provide refrigerator replacements, while others are able only to conduct assessments. Without random assignment of contracts, there could be unobserved factors at the contractor level that might be driving misreporting rates. Firms with the ability to replace refrigerators may be more willing to misreport on average. There are limited firm-level observable characteristics in the ESA data to compare the two contractor types. I am able to show that integrated- and separated-task contractors look similar with regard to the geographic area they service and the number of assessments conducted per month.⁵¹

Most of the contractors do not work exclusively for the ESA program. To better understand the types of contractors and their misreporting behavior, I group them into three broad classes based on the services they offer *outside* of the ESA program. The three types of contractors are termed “general low-income services,” “utility subsidy contractors,” and “general contractors.”⁵²

Breaking down contractors by the type of work they do provides two main insights. First, the misreporting rate varies across different types of contractors. Among general low-income service providers, separated- and integrated-task providers misreport at 17.5 percent and 36.2 percent, respectively, which is higher than the overall misreporting rates. The different behavior could reflect a different misreporting cost parameter (γ) to the firm or that these firms are more inclined to misreport to benefit low-income households. Second, within all three contractor classifications, integrated-task contractors misreport at higher rates. Even if contractor altruism toward low-income households is responsible for some of the misreporting, the level of misreporting is still influenced by the level of incentive. These findings provide evidence that there does not appear to be selection of specific types of contractors into either the separated- or integrated-task category and that contract structure influences the misreporting rate even within contractor type.

The findings in this section show that contractors are intentionally misreporting assessment data to increase their own compensation. Contractors quickly readjust their misreporting behavior in response to shifts in program rules. They also strategically omit model number data to avoid the potential for detection. Contractor misreporting rates increase with the returns to misreporting, showing that contractors directly respond to financial incentives. Taken together, these results show that misaligned incentives between the principal and agent are significantly affecting how the program is implemented.

5 Estimating Changes in Electricity Consumption

To quantify the cost of the principal-agent problem in the ESA program, I estimate the effect of refrigerator replacements on electricity consumption. I use model number data to separately

51. See Appendix D.8 for a comparison of the available observable characteristics.

52. I make these classifications using the contractor firm’s name to search for other services they provide. Appendix D.9 shows these classifications by contract type along with the number of assessments conducted and the implied misreporting rates by group.

identify the causal effect of a qualified and an unqualified refrigerator replacement on household electricity consumption. I use these results to show how intentional contractor misreporting reduces the electricity savings from refrigerator replacements in the ESA program.

5.1 Electricity Consumption Data

To estimate how a refrigerator replacement reduces electricity consumption, I combine ESA program data with confidential monthly household-level electricity consumption data. I merge model number data, reported by contractors during the assessment step, with external data on refrigerator characteristics to determine whether a refrigerator replacement was qualified or unqualified. From this model number data, I verify the refrigerator manufacture year for 57,847 households using exact matches.⁵³ I combine the ESA data from these households with monthly electricity consumption data from 2007 to 2014. I restrict the sample to households for which I have at least six months of consumption data before and after the ESA assessment was conducted, which removes households that moved shortly before or after enrolling in the ESA program and limits the sample to 38,008 households.⁵⁴ My primary specification examines electricity usage in the two years before and after enrolling in the ESA program, leaving me with 1,648,915 observations.⁵⁵

5.2 Empirical Strategy

This section describes the equation used to estimate the causal impact of a refrigerator replacement on household electricity consumption. The main specification is a panel fixed effects regression with a rich set of weather controls, household fixed effects, and time fixed effects that vary by region. Causal identification comes from comparing electricity use before and after a refrigerator replacement with electricity use by the control group that participates in the ESA program but does not receive a refrigerator replacement. In particular, I estimate

$$Q_{icmy} = \beta_1 \mathbf{1}[\text{refrigerator} = 1]_{im y} + \sum_u \beta_u \mathbf{1}[\text{upgrade} = u]_{im y} + CDD_{im y} + HDD_{im y} + \zeta_{myc} + \delta_{im} + \epsilon_{icmy}, \quad (9)$$

53. I was unable to match the model numbers to a large portion of the assessments because of incomplete or imprecisely recorded model numbers. I could not match refrigerators that did not have a reported model number. See Appendix E.1 for a description of how I match model numbers and the effects of unmatched model numbers. I hand verified a subset of the refrigerators to check how often the matching process produced errors and found an error rate of less than 1 percent. I also conduct a robustness check using a less conservative model matching approach in Appendix F.5. I do not use the matched model numbers to calculate misreporting rates because, as shown in Section 4.4, contractors are more likely to omit the model number when they are misreporting.

54. I use only households whose current residents participated in the ESA program. If the residents moved, I observe the dwelling with a new electricity account number. Requiring six months of pre-period data also allows me to examine pre-replacement trends in electricity consumption before the replacement.

55. The average household in my sample has between 20 and 21 months of data before it enrolled in the program and the same amount after enrollment. The results are robust to including all available pre- and post-enrollment consumption data. See Appendix E.2 for more details on the data cleaning process.

where Q_{icmy} is the electricity consumption in kWh for customer i in climate zone c in month m and year y . Climate zones are designated by the California Energy Commission and represent geographic regions with similar energy usage and climates.⁵⁶ The variable $\mathbf{1}[\text{refrigerator} = 1]_{icmy}$ is an indicator equal to 1 once a household’s refrigerator was replaced by the ESA program. β_1 , the coefficient of interest, represents the average change in monthly household electricity consumption in kWh caused by a refrigerator replacement. The variable $\mathbf{1}[\text{upgrade} = u]_{icmy}$ is a separate indicator for the installation of each non-refrigerator upgrade that is conducted by the ESA program.⁵⁷ CDD_{icmy} and HDD_{icmy} are controls for monthly cooling degree days and heating degree days.⁵⁸ ζ_{mcy} is a climate zone by month-of-sample fixed effect to control for contemporaneous regional shocks to electricity consumption. δ_{im} is a set of household-month-of-year fixed effects that flexibly controls for time-invariant household characteristics. Each household has a separate fixed effect for each month of the year (m) because household consumption varies seasonally throughout the year.

ϵ_{icmy} is the error term. The panel nature of this analysis makes the errors potentially correlated both within household over time and across households. I two-way cluster at the household and month-of-sample levels (Cameron, Gelbach, and Miller 2011). As a result, the standard errors are robust to both within-household and within-month correlation.

One potential concern in identifying β_1 is that households may have experienced other changes at the time of their refrigerator replacements that also could have affected electricity consumption, and could have resulted in changes in electricity consumption being misattributed to the refrigerator replacement.

The identifying assumption is that a refrigerator replacement is uncorrelated with electricity consumption conditional on the other ESA upgrades and fixed effects. Formally, this is written as $\text{cov}(\mathbf{1}[\text{refrigerator} = 1]_{icmy}, \epsilon_{icmy} | X_{icmy}) = 0$, where X_{icmy} represents the other upgrades a household received, weather controls, and fixed effects. In other words, I assume households that received a refrigerator replacement had parallel trends in consumption to those that did not, and that these trends would have continued in the absence of an upgrade. To support this parallel-trends assumption, I use an event study in Section 5.4 and conduct robustness checks with household-specific time trends. I examine the pre-upgrade trends in electricity consumption by group in Appendix F.1, finding similar patterns in pre-upgrade electricity usage across the treatment and control groups.⁵⁹

56. Climate zones are used to set building standards and other regulations. There are nine climate zones in SCE’s service territory.

57. See Appendix Table F.11 for the non-refrigerator upgrades that were conducted.

58. Weather data were obtained from MesoWest. See Appendix E.3 for more details.

59. A potential source of bias would be households with different unobserved time-varying trends in electricity consumption systematically selecting into the refrigerator replacement or control group. However, this is unlikely given the limited influence households have in the implementation of the ESA program and the analysis of pre-upgrade electricity consumption trends in Section 5.4.

5.3 Sample Selection

I report regression results for two different treatment groups: qualified replacements and unqualified replacements. I use model number data to classify replacements as qualified or unqualified based on the true manufacture year of the existing refrigerator at the time of the assessment. I classify a replacement as qualified or unqualified if I can verify its model number was manufactured before or after the 1993 eligibility cutoff. Table 1 shows average pre-period summary statistics. Group 1 consists of the 3,715 households that received qualified replacements because their existing units were manufactured in 1992 or earlier.⁶⁰ Group 2 consists of the 1,261 unqualified households that received refrigerators because the contractors misreported during the assessment step. Group 3 is the control group, which consists of 33,032 households that went through the ESA program but did not receive refrigerators because the contractors correctly reported that their existing units were manufactured in 1993 or later.⁶¹

Table 1 shows statistically significant differences in observable characteristics between each of the treatment groups (group 1 and group 2) and the control group. In particular, households that receive qualified upgrades (group 1) consumed significantly more electricity in the pre-ESA upgrade period than the control group. This difference in consumption is partially attributable to the fact that group 1 households had older refrigerators that consumed more electricity. Group 2 also consumed more electricity than the control group, which is likely because these households live in warmer regions than the control group.⁶² The difference in electricity usage, however, will not bias the estimate of β_1 , because differences in levels are absorbed by the household-month-of-year fixed effects. I also directly control for cooling and heating degree days to account for weather. All other differences in time-invariant characteristics, such as proportion renters and English speakers, will also be absorbed by the household-month-of-year fixed effect.⁶³

5.4 Event Study: Support for Parallel Trends

This section presents graphical results from an event study specification, which supports the parallel trends assumption. I estimate the event study design with the following equation:

$$Q_{icmy} = \sum_{q=-8}^8 \beta_q \mathbf{1}[\text{quarter to enrollment} = q]_{im y} + \sum_u \beta_u \mathbf{1}[\text{upgrade} = u]_{im y} + CDD_{im y} + HDD_{im y} + \zeta_{myc} + \delta_{im} + \epsilon_{icmy}, \quad (10)$$

60. It is possible that a contractor could misreport both the refrigerator manufacture year and model number. This would lead me to classify an unqualified replacement in group 1, which would bias downward the savings estimates from qualified replacements. I do not find any evidence of this behavior, but I cannot rule it out.

61. In the primary specifications, I do not include the households that were deemed eligible (either qualified or unqualified) but ultimately did not receive an upgrade. I include these households in a robustness check in Appendix Table F.20.

62. Electricity use during the winter months is the same for group 2 and the control group.

63. Most of the differences in observable characteristics, while statistically different because of the large sample, are not economically meaningful.

Table 1: Summary Statistics for ESA Participating Households in Regression Sample

Variable	Group 1 Qualified replacements	Group 2 Unqualified replacements	Group 3 No replacements (control)	P-value of difference	
				Group 1 vs. Group 3	Group 2 vs. Group 3
Pre-period use (kWh/month)	476 (229)	455 (233)	437 (233)	<.01	.01
Average monthly CDD	95 (55)	98 (53)	91 (53)	<.01	<.01
Average monthly HDD	148 (64)	172 (70)	147 (63)	.2	<.01
Proportion renter	.64 (.48)	.67 (.47)	.66 (.47)	.01	.43
Proportion single-family home	.68 (.47)	.69 (.46)	.66 (.47)	<.01	.02
Proportion disabled resident	.09 (.29)	.10 (.30)	.07 (.26)	<.01	<.01
Proportion English speaker	.56 (.50)	.50 (.50)	.48 (.50)	<.01	.19
Households	3,715	1,261	33,032		
Observations	161,818	54,557	1,432,540		
Refrigerator manufacture year (model number verified)	pre-1993	1993-2012	1993-2012		

Note: This table shows pre-period summary statistics broken down by the three groups used in the empirical analysis. Standard deviations are shown in parentheses. Group 1 consists of households that qualified for and received refrigerator replacements. Group 2 consists of households that were not qualified under ESA program rules to receive refrigerators but received replacement refrigerators because of intentional contractor misreporting. I differentiate between these two groups using the model numbers reported during the assessment step. Group 3 consists of control households that went through the ESA program and did not receive replacements because they did not qualify. The right two columns report p-values of the difference between each treatment group and the control group. CDD and HDD signify cooling and heating degree days, respectively. Pre-period use is the average monthly electricity consumption for all households in each group before they enroll in the ESA program.

where $1[\text{quarter to enrollment} = q]_{imyt}$ is an indicator in event time.⁶⁴ The household enrolls in the ESA program in quarter $q = 0$. Treatment effects are measured relative to the quarter before enrollment $q = -1$. The specification includes the eight quarters before and after enrollment. All other variables are the same as the panel fixed effects design shown in equation (9).

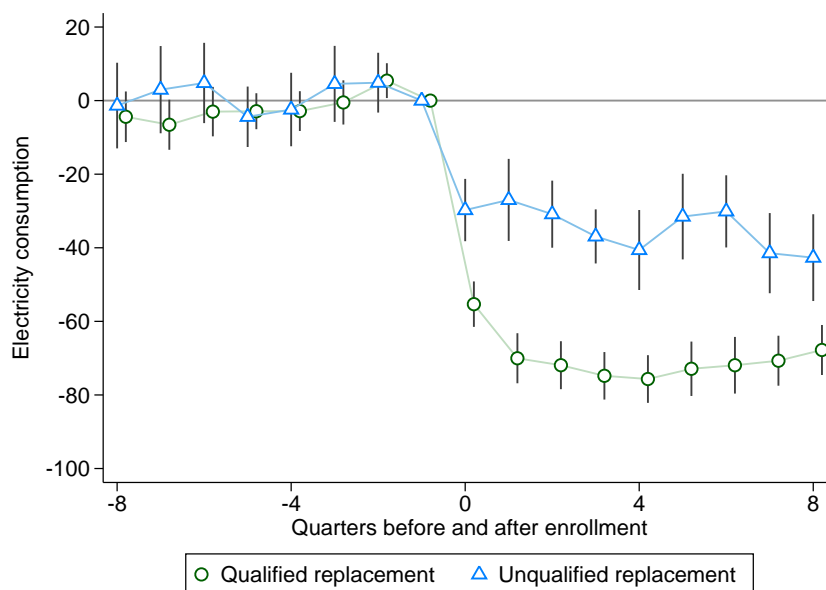
Figure 7 shows graphically—in two separate event study regressions—the causal effect of a qualified and an unqualified refrigerator replacement. The qualified replacement results compare households that received qualified replacements (Group 1) with those that did not qualify and did not receive replacements (Group 3). The unqualified replacement results compare households that received unqualified replacements (Group 2) with the same control group (Group 3). The horizontal axis is time in quarters before and after enrollment. Each dot represents the average change in

64. I use quarter of enrollment instead of quarter of refrigerator replacement because the quarter of enrollment better lines up the treatment group and the control group in event time.

monthly household kWh consumption compared with the consumption by the control group, and the vertical gray lines show 95 percent confidence intervals.

The treatment effect for both regressions in the months leading up to enrollment is flat and statistically indistinguishable from zero. This flat pre-treatment result suggests that households in the treatment group are similar to households in the control group, conditional on controls and fixed effects. It provides additional supporting evidence that this specification satisfies the parallel trends assumption, and that the estimates reflect the causal estimate of a refrigerator replacement.⁶⁵

Figure 7: Electricity Savings from Qualified and Unqualified Refrigerator Replacements



Note: This figure shows two event studies for refrigerator replacements, estimated using equation (10). The qualified replacement event study, indicated with circles, shows the electricity savings caused by qualified refrigerator replacements. The unqualified replacement event study, indicated with triangles, shows the electricity savings caused by unqualified refrigerator replacements. The results show that qualified replacements save about twice as much electricity as unqualified replacements. All results are in kWh per month and are relative to the quarter before enrollment. Both regressions control for non-refrigerator upgrades conducted in the ESA program and include weather controls, household-month-of-year fixed effects, and month-of-sample by climate-zone fixed effects. Standard errors are clustered at the household and month-of-sample levels.

The results show that a qualified refrigerator replacement provides statistically significant electricity reductions between 70 and 75 kWh/month, or around 16 percent of average monthly consumption. The quarter of enrollment in the ESA program ($q = 0$) has a smaller savings estimate than the months that follow. Not all refrigerator replacements are completed in the first three months after enrollment, so the full electricity savings are not realized until the next quarter. The

⁶⁵ I use a placebo event study in Appendix F.2 to show that enrolling in the program does not cause a change in electricity use for the control group.

savings estimates for the quarters after enrollment are flat, suggesting a stable long-run treatment effect for at least two years.⁶⁶

The unqualified refrigerator replacement results (indicated with triangles) show that unqualified refrigerator replacements save about half as much electricity as a qualified replacement. Starting in 1993, all new refrigerators were subject to more stringent federal minimum efficiency standards, making them more efficient than similar refrigerators sold in 1992. In other words, the ESA program specifically excluded post-1992 refrigerators because of the timing of the standards. My results estimate that replacing a pre-1993 refrigerator saves substantially more electricity than replacing a post-1992 refrigerator. This means that contractor misreporting led to significantly smaller reductions in electricity consumption.

5.5 Panel Fixed Effects Results

Table 2 shows the results from estimating the effects of a refrigerator replacement using equation (9). Column 1 reports the changes in monthly electricity consumption from a qualified refrigerator replacement. The first row shows that a qualified refrigerator replacement causes a 73 kWh/month reduction in electricity consumption, or 16.4 percent of the average household's pre-upgrade electricity consumption.⁶⁷ These estimates are consistent with engineering predictions of the impacts of replacing a refrigerator manufactured in 1992 or earlier with a new energy-efficient model.⁶⁸

Column 2 of Table 2 estimates the same regression for unqualified refrigerator replacements using sample groups 2 and 3. I find that replacing a refrigerator manufactured in 1993 or later reduces consumption by 38 kWh/month.⁶⁹ The estimated savings are consistent with the engineering estimate prediction of replacing refrigerators manufactured after 1992 with new energy-efficient units. The savings from an unqualified replacement are half as large as and statistically different from the savings from a qualified replacement.⁷⁰ This finding shows how intentional contractor misreporting can reduce the electricity savings from a refrigerator replacement.

I conduct a number of robustness checks in Appendix F, all of which support the above results. Appendix F.6 shows the results estimated with household-specific time trends, which yields quantitatively similar savings estimates. Appendix F.7 runs the same regression as Table 2 with

66. I estimate the event study with a longer time horizon and find that the results are stable and persist for five years, which is the maximum I can test with the available data. See results in Appendix F.3.

67. Appendix F.4 reports the results with individual coefficients for each of the upgrades conducted in the ESA program.

68. There are no official ex-ante energy savings engineering estimates for the ESA program because the projected savings are based on an ex-post evaluation (Evergreen Economics 2013). As a result, there is no energy efficiency gap because the projected savings are from a retrospective analysis that likely includes qualified and unqualified refrigerator replacements. I calculate the engineering predictions using the average increase in minimum efficiency standards over time. See Appendix A for more details. The similarity of the engineering projections to the estimated effects of a qualified replacement suggests that model numbers in Group 1 are accurately reported and that contractors are not misreporting both the manufacture year and the model number.

69. With a less conservative model matching approach, I include about 14,000 more households and find similar savings estimates for both qualified and unqualified replacements. See Appendix F.5 for these results.

70. See Appendix Table F.21 for the two effects estimated jointly.

Table 2: The Effect of Refrigerator Replacements on Monthly Electricity Consumption

	(1) Qualified replacements	(2) Unqualified replacements
Refrigerator replacement	-73.45 (2.68)	-38.02 (3.20)
Cooling degree days	0.12 (0.02)	0.11 (0.02)
Heating degree days	0.03 (0.01)	0.03 (0.01)
Controls for other upgrades	Yes	Yes
Pre-period consumption (kWh/month)	448	445
Refrigerator replacements	3,715	1,261
Households	36,747	34,293
Observations	1,581,024	1,474,841

Note: This table reports regression coefficients from two regressions using equation (9). The dependent variable in both regressions is monthly household electricity consumption. Refrigerator replacement is an indicator for a household’s refrigerator being replaced. The refrigerator replacement coefficient in Column 1 shows that a qualified replacement causes a 73 kWh/month reduction in monthly consumption. Column 2 shows that an unqualified replacement causes a 38 kWh/month reduction in monthly consumption. Both Column 1 and Column 2 use the same control group of customers that went through the ESA program but did not qualify for refrigerator replacements. Both regressions control for non-refrigerator upgrades conducted in the ESA program and include weather controls, household-month-of-year fixed effects, and month-of-sample by climate-zone fixed effects. Standard errors are in parentheses and are clustered at the household and month-of-sample levels.

the control group limited to households with refrigerators manufactured between 1993 and 1996. I find almost identical energy savings, suggesting that households that have newer refrigerators and did not receive replacements are not driving the results. Appendix F.8 shows that qualified and unqualified refrigerator replacements reduce electricity consumption by the same amount for renters and homeowners. This result suggests that unobserved non-ESA program upgrades are not biasing the refrigerator savings estimates because renters have a smaller incentive to purchase non-ESA energy efficiency upgrades. Appendix F.9 shows the main specification results for the subset of households that only received CFLs or refrigerators to verify that potential contractor misreporting on other major upgrades is not biasing the refrigerator savings estimates. The estimates in Appendix F.9 are similar to the main results in Table 2. Appendix F.10 breaks down the results in Columns 1 and 2 by the minimum efficiency standard of the refrigerator that was replaced. The results show that replacing older refrigerators saves more electricity. Replacing a refrigerator manufactured before 1990 saves 85 kWh/month. In contrast, replacing a refrigerator manufactured in 2001 or later reduces consumption by only 20 kWh/month.

The results in this section show that qualified replacements save almost twice as much electricity as unqualified replacements. This suggests that there is a large cost to ESA program outcomes from intentional contractor misreporting.

6 Welfare Costs of Misreporting

6.1 Welfare Analysis

In this section, I use the above electricity savings estimates to calculate the social welfare impacts of a qualified and unqualified refrigerator replacement, which are w_1 and w_2 in Section 3. I capture the key components of welfare in equation (11), consisting of the private benefits, which include net consumer and producer surplus impacts of reducing electricity consumption; the carbon and local air pollution externality benefits of reduced electricity consumption; the costs associated with replacing refrigerators (including the distortion from raising the revenue to fund the program); and the contractor labor cost. The welfare measure is intended to capture the sum of all consumer and producer surplus and externality reduction benefits while subtracting the capital replacement costs and efficiency costs of funding the program through higher electricity bills. I assume that there are no income effects from a refrigerator replacement and that any remaining unmeasured welfare impacts are small.⁷¹

$$\begin{aligned} \Delta \text{Welfare} = & \text{Private benefit of electricity savings} + \text{Reduction in externalities} \\ & - \text{Capital replacement costs} - \text{Distortion from raising funds} - \text{Labor cost.} \end{aligned} \quad (11)$$

I first construct the counterfactual electricity consumption in the absence of an ESA refrigerator replacement. Using assessment data, I model the proportion of old refrigerators that would break or otherwise be replaced with new refrigerators each year. The assessment data allows for a more nuanced characterization of when refrigerators would have been scrapped and replaced compared with the standard assumption in the literature that the program accelerated appliance replacement by five years.⁷² There are three costs associated with refrigerator replacements. The first is the cost of scrapping an old, but working, refrigerator and moving the purchase of a replacement refrigerator up to the present.⁷³ For simplicity, I assume the ESA program provides the same refrigerator that households would eventually purchase when replacing their own refrigerators in the absence of the program.⁷⁴ The second is the labor cost associated with delivering and installing

71. Appendix F.8 shows that homeowners and renters have the same energy savings from a refrigerator replacement, which suggests owners are not experiencing a substantial income effect from receiving a new refrigerator that could lead to increased energy consumption. See Appendix G for a detailed discussion of the assumptions behind the welfare calculation including the potential for contractor market power. I assume the market for refrigerators is perfectly competitive and that there are no producer surplus benefits for refrigerator manufacturers from the ESA program.

72. The assessment data captures ESA-eligible households purchasing used refrigerators or keeping older refrigerators. See Appendix G.2 for a discussion of household behavior absent the ESA program and how I calculate the counterfactual refrigerator scrapping and replacement.

73. If old refrigerators are not scrapped then the welfare calculation would depend on how the old refrigerator is used. Not removing the existing refrigerator is risky for a contractor because they could be caught by a post-upgrade audit conducted by the ESA program. I test for non-scrapping by examining individual households that received refrigerators through the program to see if their electricity consumption increased discontinuously after an installation. I find no evidence of non-scrapping. See Appendix G.1 for a discussion.

74. An alternate formation would have households in the counterfactual purchasing refrigerators that met the minimum standards. In practice, the simplifying assumption I use does not have a large effect on energy savings because minimum efficiency standards increased in 2014 to a similar level as the ESA provided energy-efficient refrigerator. See Appendix G.3 for a detailed discussion and calculation.

a refrigerator earlier than would have happened in the absence of the ESA program.⁷⁵ The third is the welfare loss from increasing electricity prices a small amount to raise money to pay for the refrigerators. This welfare cost, while important to account for, is only 2 to 3 cents per refrigerator, because the demand for electricity is inelastic.⁷⁶ The benefits of refrigerator replacement come from reducing electricity consumption between when the refrigerator was replaced in the ESA program and when it would have been replaced in the absence of the program.⁷⁷ The calculation in equation (11) does not account for the non-energy benefits to a household from having a newer refrigerator.⁷⁸ I calculate these reductions using the regression estimates in Table 2 separately for qualified and unqualified replacements. I value the total reduction in electricity consumption at the social marginal cost of electricity of \$0.10/kWh and use a 3 percent annual real discount rate (Borenstein 2012).⁷⁹

Table 3 shows the results for three scenarios. Column 1 shows the welfare impacts for the base case. A qualified replacement *increases* welfare by \$60 per replacement, and an unqualified replacement *decreases* welfare by \$106, which correspond to w_1 and w_2 , respectively, in the model in Section 3.⁸⁰ Much of this difference is because a qualified ESA replacement saves on average 3,677 kWh over its lifespan, while an unqualified replacement saves only 2,598 kWh. The last two rows in the table show a summary statistic for the percentage of refrigerators that would have remained in the absence of the ESA program. A larger portion of post-1992 refrigerators (55 percent) are still being used after five years than their pre-1993 counterparts (42 percent). This difference reflects post-1992 refrigerators being newer and therefore being replaced more slowly in the absence of the ESA program.

The other columns in Table 3 use alternate refrigerator replacement rates to understand the impact of this assumption on the estimates of w_1 and w_2 . Column 2 assumes an accelerated

75. The capital replacement and labor costs make up the marginal cost of providing the service in Section 3.

76. In California, the price of electricity is higher than the social marginal cost (Borenstein and Bushnell 2018). Any price increase will inefficiently reduce electricity consumption. See Appendix G.3 for a detailed discussion of the three costs associated with refrigerator replacement.

77. There are many reasons why a household might not have replaced their old existing refrigerator despite potential financial benefits, including liquidity constraints and the landlord-tenant problem.

78. Non-energy benefits could include quieter operation, more consistent cooling, and better internal organization. Non-energy benefits are on average larger for qualified replacements, which are older and likely less desirable refrigerators. There are also some non-energy costs. The replacement refrigerators are the lowest-cost base model without extra features, while the existing refrigerator that was replaced may have had features such as an ice maker. In a small survey, 24 percent of respondents said their replacement was lower quality than their previous refrigerator (Research into Action 2011). See Appendix G.1 for a discussion of how this could affect the welfare calculations. It would be possible to estimate the demand system with data on refrigerator prices (e.g. Houde (2018)), but I do not have the relevant data for that exercise.

79. This is equivalent to a welfare calculation that adds up the private and externality benefits of a refrigerator replacement. See Appendix G.4 for a detailed discussion of this accounting. I find similar results using a range of discount rates, which can be seen in Appendix G.5. I do not use the estimate of social marginal cost from Borenstein and Bushnell (2018) because it is based on a short run marginal cost that would be less appropriate for the analysis of a refrigerator replacement that lasts many years.

80. The welfare costs for unqualified replacements do not include any costs to the agents of misreporting. The welfare cost of agent misreporting is captured in term 3 of equations (5) and (7) for integrated- and separated-task contractors, respectively. Including the welfare cost of misreporting increases the cost of an unqualified replacement to \$168 for integrated-task contractors and \$119 for separated-task contractors. See Appendix G.6 for details.

Table 3: Welfare Impacts of Qualified and Unqualified Refrigerator Replacements

	(1) Base case	(2) High refrigerator failure rate	(3) Low refrigerator failure rate
Welfare impact of qualified replacement (w_1)	\$60	\$25	\$94
Welfare impact unqualified replacement (w_2)	-\$106	-\$91	-\$119
Lifetime kWh savings from qualified replacement	3,677	2,574	4,759
Lifetime kWh savings from unqualified replacement	2,598	1,819	3,244
Pre-1993 refrigerators in year 5 without ESA program	42%	29%	54%
Post-1992 refrigerators in year 5 without ESA program	55%	38%	71%

Note: This table shows the welfare effects of qualified and unqualified replacements for four scenarios using the electricity savings estimates from Table 2. Each scenario varies the rate at which refrigerators would break or otherwise be replaced in the absence of the ESA program. The table shows that qualified replacements increase welfare and unqualified replacements decrease welfare under all the scenarios. Electricity savings are valued at the social marginal cost of \$0.10/kWh.

refrigerator replacement rate, which reduces the welfare impacts for both qualified and unqualified replacements. Column 3 assumes that old refrigerators are replaced at a slower rate in the absence of the ESA program and finds that unqualified replacements have larger negative welfare consequences than in the base case. In this scenario, the lifetime kWh reductions from an unqualified replacement are larger than in the base case, but the costs of replacement also go up, as the existing refrigerators would have lasted longer in the absence of the program.

The results in this section show that under a wide range of assumptions, qualified replacements increase welfare and unqualified replacements reduce welfare. This discrepancy highlights the costs of the principal-agent problem: agent misreporting leads to welfare losses.

6.2 Alternative Cost-Benefits

There are shortcomings in using a single welfare measure to quantify the effect of qualified and unqualified refrigerator replacements. The previous section’s welfare calculations assume that SCE, working under the CPUC’s direction, wants to maximize total surplus. That welfare measure also assumes no welfare costs, aside from the distortion caused by raising funds for the program, associated with paying contractors to provide unqualified replacements because those payments are only transfers. If, for example, the CPUC or SCE’s ratepayers assign a cost to the payments to contractors for unqualified replacements, then the estimated welfare costs of unqualified replacements could be an underestimate.

An alternative approach is to consider the costs and benefits of refrigerator replacements from five different perspectives: the households participating in the ESA program, the reduction in externality damages, the contractors making the installations, SCE electricity customers, and SCE as a company. Households saved around \$414 in private electricity expenditure from a qualified replacement and \$288 from an unqualified replacement relative to not receiving a replacement through the program. The dwelling owner also receives the \$650 replacement refrigerator at no

charge.⁸¹ The electricity savings also reduced CO₂ emissions and local air pollution. Using a \$50 social cost of carbon and local air pollution costs in California from Muller, Mendelsohn, and Nordhaus (2011), I calculate that qualified and unqualified replacements reduced externality damages by \$78 and \$54 over the lifetime of the replacements, with 84 percent of those benefits coming from CO₂ reductions.

The ESA program is administered by SCE and paid for by their electricity customers through higher electricity prices. SCE spent \$35.2 million over four years on 41,500 refrigerator replacements, \$17.5 of which was on unqualified replacements. Of those unqualified replacements, SCE spent \$12.9 million purchasing the refrigerators and paid contractors a total of \$4.6 million to conduct the upgrades. The ESA program is financed through higher electricity bills, which I estimate increased by \$0.17/month for the average SCE residential electricity customer.⁸²

Breaking down the ESA program costs and benefits into categories shows some clear winners and losers. Households that received a new refrigerator are better off, along with the contractors that provided replacements. Unqualified replacements make SCE worse off, although its objective function is harder to specify. If SCE's goal is to sell as much electricity as possible, then the unqualified replacements reduced consumption beyond a hypothetical ESA program that only provided replacements to the 13.8% of qualified households.⁸³ Raising electricity rates to pay for the unqualified replacements also will reduce long-run electricity sales by a small amount. SCE may also wish to maintain a good relationship with the PUC to receive favorable treatment in other matters. The discovery of large numbers of unqualified replacements in the ESA program could negatively affect that relationship. This paper has documented the misreporting in the ESA program, resulting in a PUC decision asking SCE to address this issue (CPUC 2019).

Calculating the impacts of refrigerator upgrades on various groups is helpful to understand the ESA program's costs and benefits. However, it can be harder to judge the overall impacts of ESA refrigerator replacements with this approach. Excluding program participants and contractors, SCE ratepayers are on net worse off from unqualified replacements. The total externality benefit of \$1.1 million from unqualified replacements, most of which comes from CO₂ reductions, is meaningfully smaller than the \$17.5 million in utility expenditure required to finance them. While the single welfare measure calculated in Section 6.1 has drawbacks, it is useful when considering the ESA program's full impacts.⁸⁴ In the next section, I use the welfare measure of refrigerator replacements to investigate the benefits of contract integration.

81. In rental units, the free refrigerator's value goes to the landlord, who may also be able to offer higher rent due to lower energy bills for their tenants.

82. SCE has almost 4.3 million residential customers that consume on average 6,894 kWh/year (EIA 2019). I estimate that prices would go up by \$0.0003 per kWh for all residential customers in SCE. The excess burden caused by further increasing prices above the social marginal cost is small at only \$0.02 per customer per year using a demand elasticity from (Ito 2014).

83. I assume that unqualified replacements did not crowd out qualified replacements. Contractors have no incentive to pass on qualified replacements in favor of unqualified replacements, and I find no evidence that this happened.

84. Reporting the costs and benefits from different perspectives does not solve the problem of the ESA program's hard-to-measure impacts, such as the non-energy benefits of a new refrigerator.

6.3 Calculating Benefits of Integration

This section combines the empirical analysis done in Sections 4 and 5, the welfare analysis in Section 6.1, and the theoretical predictions in Section 3 to estimate how the welfare consequences of the principal-agent problem interacts with contract type. I derive most of the key parameters empirically. I take those that I cannot observe from program data on payments made to contractors and first-order conditions from the model.

My estimates of the payment to a contractor for a refrigerator upgrade (g) and five CFLs (d) are based on ESA payments to contractors.⁸⁵ SCE makes bulk purchases of refrigerators and CFLs, which it gives directly to the contractors to install. I assume that there is no labor cost to install a light bulb.⁸⁶ Contractors are paid \$5 per CFL they install, making the referral payment d \$25.⁸⁷

Contractors are paid \$224 per refrigerator replacement, which covers the costs of installing the new unit along with removing and recycling the old unit in an environmentally appropriate manner (SCE 2012a). However, the contractor's expected value of g is lower than \$224 for four reasons. First, the contractor must pay a disposal fee for an old refrigerator that is removed, which in the SCE service territory is \$27 (WM 2018).⁸⁸ Second, there is a labor cost for providing the replacement, which I estimate at \$30 per replacement.⁸⁹ Third, only 84.2 percent of households that are eligible based on the assessments receive replacements.⁹⁰ Fourth, logistical constraints may prevent the integrated-task contractor who conducts the assessment from providing the replacement. The same integrated-task contractor does both steps 69 percent of the time, and contractors do not know which replacements they will be completing during the assessment step.⁹¹ Accounting for all these factors, integrated-task contractors receive in expectation \$123 per refrigerator they report as eligible for replacement, while separated-task contractors receive only \$25.

I calculate the values for γ_I and γ_S , which convert agent misreporting rates into a misreporting cost, using first-order conditions (equations (3) and (6)) and estimates of d , g , Z_I , and Z_S . I leverage

85. I omit the integrated and separated subscripts on d and g because SCE pays all contractors the same amount for each service they provide.

86. On the ride-alongs I participated in, I observed that it did not take contractors much time or effort to replace the CFLs once they had already conducted the home assessment.

87. Of households that receive a refrigerator, 82 percent also receive at least one CFL during the assessment step. In most cases contractors install all five CFLs, with 73 percent of households that receive at least one CFL receiving all five.

88. Many electricity utilities or local governments offer energy efficiency programs that pay households between \$25 and \$50 to remove old refrigerators and dispose of them. I assume contractors are not able to take advantage of these programs and must pay for the refrigerator disposal at a local facility.

89. The labor cost estimate is based on the appliance delivery fee charged by major appliance retailers. In Appendix G.7 I conduct robustness checks using labor costs of zero and \$60.

90. See Appendix D.7 for a discussion of why all eligible households do not receive refrigerator replacements. Integrated-task and separated-task contractors complete 84.9 and 83.0 of replacements in eligible households. The small difference suggests that contract structure is not affecting refrigerator installation completions.

91. Much of this discrepancy is due to scheduling issues. Appendix G.8 discusses this own-firm upgrade rate in more detail.

the functional form assumption that the cost of misreporting is $C(Z) = \frac{1}{2}\gamma Z^2$.⁹² Using equations (3) and (6), I calculate a value of 646 and 319 for γ_I and γ_S , respectively.

Table 4 shows the parameters I use for the calculation of equation (8) and how I estimate them. I cannot measure the information acquisition cost, which captures the benefits from having the same contractor responsible for both assessment and refrigerator replacement, with the available data. Instead, I first use the model to calculate how large the information acquisition cost needs to be to justify an integrated-task contract. In the base case using the parameters displayed in Table 4, I estimate that the benefits would have to exceed \$95 for an integrated-task contractor to be preferred. Because contractors are paid only \$70 to conduct an assessment (Steps 3-4 in Section 2.2), it is extremely unlikely that the information acquisition cost is this large.

Table 4: Estimated and Derived Parameters to Calculate Benefits of Integrated-Task Contract

Parameter	Value	Origin of estimate
Predicted qualified eligibility rate (α)	.138	Calculated from ESA data in Section 4.3
Separated misreporting rate (Z_S)	.078	Calculated from ESA data in Section 4.6
Integrated misreporting rate (Z_I)	.19	Calculated from ESA data in Section 4.6
Welfare costs of unqualified replacement (w_2)	\$103	Calculated in Section 6.1 using regression results
Referral payment (d)	\$25	ESA program documents and author's calculations
Refrigerator installation payment (g)	\$98	ESA program documents and author's calculations
Integrated misreporting cost parameter (γ_I)	646	Equation (3) using values for Z_{int} , d , and g
Separated misreporting cost parameter (γ_S)	319	Equation (6) using values for Z_{sep} and d

Note: This table shows the parameters used to calculate the benefits of an integrated-task contract from equation (8). The parameters are derived empirically from program data and first-order conditions from the model.

I next calculate the benefits of integration using a range of information acquisition cost estimates. I use the \$70 SCE pays contractors to conduct household assessment visits as the basis for my estimates. The assessment visit includes a variety of costs to the contractor including scheduling an appointment, transportation to a household, looking at a household's W2 income statement, walking through every room in the dwelling, and recording data. Because the full \$70 payment includes agent activities that do not involve acquiring information related to the refrigerator replacement, I approximate the information acquisition cost for refrigerator replacements in the primary specification as \$35.

Table 5 shows the calculation of the benefits of integration per assessment using the base case refrigerator replacement scenario. The first column shows the results using an information acquisition cost of \$35 per refrigerator replacement, which results in a benefit of integration of -\$12.32 per assessment. In this case, it is welfare improving to separate the assessment and refrigerator replacement between two different contractors. In total, integrated contract assessments reduce welfare in the ESA program by \$1,109,000 across 90,084 assessments. The assessments resulted in 25,000

⁹². I conduct robustness checks using a different functional form in Appendix G.9.

refrigerator replacements (about 58 percent unqualified), which created \$6,774,000 in benefits from reducing electricity consumption. The distortion introduced by using an integrated-task contract rather than a separated-task contract is equal to 16 percent of the total benefits of the refrigerator replacement program.

Table 5: Benefits of an Integrated-Task Contract Estimates

	(1) Base case	(2) High information acquisition cost	(3) Zero agent misreporting cost
Benefits of integration per assessment	-\$12.32	-\$5.11	-\$3.07
Information acquisition cost (F)	\$35	\$70	\$35
Welfare costs of unqualified replacement (w_2)	-\$106	-\$106	-\$106
Integrated misreporting cost parameter (γ_I)	646	646	0
Separated misreporting cost parameter (γ_S)	319	319	0

Note: This table shows the benefits of an integrated-task contract relative to a separated-task contract, calculated using equation (8). Each column uses a different set of assumptions. Column 1 shows that the costs of integration are high when using the base case parameters from Table 4. Columns 2 and 3 make progressively stronger assumptions to benefit the integrated-task contract, but the separated-task contract is still preferred.

In Column 2, I present a scenario where the information acquisition cost is \$70. Under this conservative assumption, the benefits of integration are -\$5.11 per assessment. Even assuming this high information acquisition cost, the benefits of integration are negative.⁹³

To tilt the balance further in favor of the integrated-task contract, I zero out the cost associated with contractor misreporting while assuming contractors still misreport at the same rate. The results of the zero agent misreporting cost scenario are shown in Column 3 of Table 5. This scenario does not reflect a plausible scenario, but it is a useful lower bound because it enables me to estimate the importance of the misreporting cost for the benefits of integration.⁹⁴ I find that even with no agent misreporting cost, the benefits of integration are -\$3.07 per assessment, and thus a separated-task contract is still preferable.

I next calculate the information acquisition cost that would justify an integrated-task contract in the absence of a contractor misreporting cost. I find that the value of integration would have to be at least \$50 per refrigerator replacement. This estimate serves as a lower bound on how large the information acquisition cost would have to be to justify task integration. It does not require a functional form assumption on the misreporting cost function. Even in this scenario, designed to favor task integration, I find that a high information acquisition cost is necessary to justify the integrated-task contract.

93. See Appendix G.10 for a discussion of the benefits of integration in the ESA program.

94. If there was no misreporting cost, contractors would misreport every time under both contract types. This scenario also reflects an alternate formulation of the model where the agent misreporting cost is not a welfare cost. See Appendix B.3 for a discussion of this alternate formulation of the model.

6.4 Benefit-Cost Ratio

I calculate a benefit-cost ratio for the replacements conducted by separated- and integrated-task contractors using the same approach as in Section 6.1. I account for the full costs and benefits for each contract type, including the agent misreporting cost and the information acquisition cost (\$35 per replacement) in the separated-task contract.⁹⁵ Using this approach, I find that an integrated-task contract has a benefit-cost ratio of 0.73, while a separated-task contract has a higher ratio of 0.86.⁹⁶ The difference between these two values is driven by differences in the contract structure and the larger incentive for contractors to misreport in the integrated-task contract. Both of the contract structures do not have benefits that exceed the costs.

It is also possible to make an out-of-sample calculation and consider what the refrigerator replacement program would look like if it conducted only qualified replacements. I find that conducting only qualified replacements would yield a benefit-cost ratio of 1.22, which is 49 percentage points higher than the integrated-task contract. This large difference demonstrates the importance of the principal-agent problem and contract structure for cost-effectiveness.

6.5 Policy Fixes

There are two main approaches to reducing the misreporting in the ESA program: monitor agent actions or fully separate the task of assessment and service provision between two agents. After reading an earlier version of this paper, SCE implemented a monitoring system that requires contractors to submit a picture of the serial number plaque for each assessed refrigerator. SCE plans to audit a small proportion of these pictures to verify that contractors are submitting accurate data. Photo-based monitoring should lead to a large drop in misreporting because contractors exhibited strategic reporting behavior in the existing program (e.g., Section 4.4), suggesting that they will respond to the introduction of accountability and potential penalties.

Photo-based monitoring has disadvantages. First, agents can evade the monitoring by substituting the picture from a qualified refrigerator.⁹⁷ Second, contractors may shift their misreporting from refrigerators to other unmonitored measures. In the ESA program, there is suggestive evidence that contractors are more likely to install upgrades that earn them higher profits. For example, only four firms are qualified to replace central air systems, giving them a larger incentive to find a household qualified for a replacement. I find that when one of these central-air-installing firms conducts an assessment, households in eligible regions receive a new central air system 30.3 percent of the time, compared with 23.3 percent of the assessments conducted by firms that are not qualified to install central air systems.⁹⁸ Any reallocation of misreporting may undermine some of the

95. See Appendix G.11 for details on how the benefit-cost ratios are calculated.

96. The benefit-cost ratio of the separated-task contract increases to 0.96 if there is no information acquisition cost.

97. It would take a sophisticated contractor a significant effort for this type of misreporting because they would also have to manipulate the metadata (location, time, etc.) embedded in the picture they submitted.

98. Households were only eligible for central air replacements in the hottest regions of SCE's service territory. The central-air-replacing firms have the ability to install refrigerators but install them at a lower rate than other firms. This behavior suggests that central-air-replacing firms may have shifted misreporting towards central air replacements, which earns them \$2,000 in installation costs per replacement, and away from refrigerators.

benefits of monitoring refrigerator installations. Another approach to reduce overall misreporting is to fully separate the assessment from the service provision. In the ESA program, this could be accomplished with a modified separated-task contract where the assessment contractor does not have the \$25 incentive to misreport. Full task separation would reduce the incentive to misreport on all upgrades, not just the ones that can be directly monitored. It could also introduce task coordination costs associated with having different firms conduct each step. It is not possible to determine if monitoring or task separation is better for the ESA program or other principal-agent problems without further research.

It is important to consider how generalizable the results and the policy fixes in this paper are to other settings. The setup of the ESA program may make the principal-agent problem more noticeable and costly, but there are many settings where agents face similar incentives that can result in suboptimal outcomes. Many energy efficiency programs use a piece-rate wage where third-party contractors determine which upgrades they are paid to conduct. In those cases, full task separation or monitoring could reduce the contractor’s incentive to provide services for their personal benefit. Outside of energy efficiency programs, there is piece-rate compensation for physicians, general contractors, auto mechanics, and other specialists that likely have similar principal-agent problems. The proposed policy fixes of contract separation or expanded monitoring likely apply in settings where piece-rate wages give agents the incentive to provide unqualified services with a low chance of being caught. In some other instances where contract separation is not possible or monitoring is infeasible, the insights on the principal-agent problem in this paper may be less applicable. In Appendix G.12, I use the SANS conditions outlined in List (2020) for a more detailed discussion of the external validity of the findings in this paper.

7 Conclusion

In this paper, I measure the welfare costs of the principal-agent problem. In the empirical setting I study—an energy efficiency appliance replacement program—contractors intentionally misreport assessment data to provide unqualified refrigerator replacements and increase their compensation. This profit-seeking agent behavior has significant welfare costs: I estimate that each unqualified replacement reduces welfare by \$106. In contrast, each replacement that follows program rules saves twice as much electricity and increases welfare by \$60. I develop a principal-agent framework to quantify the costs and benefits of having the same agent conduct multiple tasks versus splitting the tasks between different agents. The empirical estimates derived imply that the integrated-task contract structure causes the ESA refrigerator replacement program to be welfare reducing.

These findings shed light on a key question in the energy efficiency literature: why do energy efficiency programs deliver lower savings than predicted? I provide the first evidence that the principal-agent problem may be an important part of the explanation. These findings are applicable to a wide range of energy efficiency policies, because similar incentives exist for contractors to

maximize their income during program implementation. Other programs may use monitoring to avoid the misreporting problems observed in this paper, but monitoring every decision made by a contractor in an energy efficiency program can be expensive and impractical. Burlig et al. (2020) study a program that subsidizes energy-efficient capital upgrades, finding savings that were only 51 percent of ex-ante expectations. In this type of program, utility-licensed contractors recommend upgrades from a menu of options. It is likely that they focus on installing upgrades that are more profitable for them to install but yield lower savings.

More broadly, I am able to quantify the welfare consequences of misaligned incentives between principals and agents. The principal-agent problem is a well-understood theoretical concept, but it has proved challenging to quantify its effects on welfare. Because asymmetric information is at the core of the principal-agent problem, it is difficult for the researcher to observe agent behavior and quantify the welfare costs of self-interested agent actions. The existing literature primarily identifies how agents distort their behavior in a variety of contexts including commission-based sales behavior, teacher performance pay, healthcare provision, and auto repair. The degree to which the observed distortions cause significant welfare losses has generally remained an unanswered question. My research provides answers to this question in one setting, showing that self-interested agents can substantially reduce welfare. Further research is necessary to learn how task assignment can better align the incentives of principals and agents in different settings. As this paper demonstrates, failing to consider the incentives of agents can have large welfare costs and undermine policy goals.

References

- Allcott, Hunt, and Michael Greenstone. 2012. “Is There an Energy Efficiency Gap?” *Journal of Economic Perspectives* 26 (1): 3–28.
- . 2017. “Measuring the Welfare Effects of Residential Energy Efficiency Programs.” NBER Working Paper No. 23386, National Bureau of Economic Research, Cambridge, MA.
- Asch, Beth J. 1990. “Do Incentives Matter? The Case of Navy Recruiters.” *ILR Review* 43 (3): 89-S-106–S.
- Baker, George. 1992. “Incentive Contracts and Performance Measurement.” *Journal of Political Economy* 100 (3): 598–614.
- . 2002. “Distortion and Risk in Optimal Incentive Contracts.” *The Journal of Human Resources* 37 (4): 728–751.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2007. “Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment.” *The Quarterly Journal of Economics* 122 (2): 729–773.
- Berenson, Robert A., and Eugene C. Rich. 2010. “US Approaches to Physician Payment: The Deconstruction of Primary Care.” *Journal of General Internal Medicine* 25 (6): 613–618.
- Borenstein, Severin. 2012. “The Private and Public Economics of Renewable Electricity Generation.” *Journal of Economic Perspectives* 26 (1): 67–92.
- Borenstein, Severin, and James B Bushnell. 2018. “Do Two Electricity Pricing Wrongs Make a Right? Cost Recovery, Externalities, and Efficiency.” NBER Working Paper No. 24756, National Bureau of Economic Research, Cambridge, MA.
- Burlig, Fiona, Christopher Knittel, David Rapson, Mar Reguant, and Catherine Wolfram. 2020. “Machine Learning from Schools about Energy Efficiency.” *Journal of the Association of Environmental and Resource Economists* 7 (6): 1181–1217.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. “Bootstrap-Based Improvements for Inference with Clustered Errors.” *Review of Economics and Statistics* 90 (3): 414–427.
- . 2011. “Robust Inference with Multiway Clustering.” *Journal of Business & Economic Statistics* 29 (2): 238–249.
- CEE, Consortium for Energy Efficiency. 2018. *State of the Efficiency Program Industry*. Technical report. Boston: CEE.

- Christensen, Peter, Paul Francisco, Erica Myers, and Mateus Souza. 2021. “Decomposing the Wedge Between Projected and Realized Returns in Energy Efficiency Programs.” *E2e Working Paper No. 046*.
- Copeland, Adam, and Cyril Monnet. 2009. “The Welfare Effects of Incentive Schemes.” *The Review of Economic Studies* 76 (1): 93–113.
- Courty, Pascal, and Gerald Marschke. 1997. “Measuring Government Performance: Lessons from a Federal Job-Training Program.” *The American Economic Review* 87 (2): 383–388.
- . 2003. “Dynamics of Performance-Measurement Systems.” *Oxford Review of Economic Policy* 19 (2): 268–284.
- . 2004. “An Empirical Investigation of Gaming Responses to Explicit Performance Incentives.” *Journal of Labor Economics* 22 (1): 23–56.
- . 2008. “A General Test for Distortions in Performance Measures.” *The Review of Economics and Statistics* 90 (3): 428–441.
- CPUC, California Public Utilities Commission. 2019. *Decision 19-06-022*. Technical report. San Francisco: CPUC.
- 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): 207–38.
- Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. “Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India.” *The Quarterly Journal of Economics* 128 (4): 1499–1545.
- EIA. 1987. *Residential Energy Consumption Survey*. Technical report DOE/EIA-0321/1(87). Washington, DC: EIA.
- . 2019. *Form EIA-861*. Technical report. Washington, DC: EIA.
- Evergreen Economics. 2013. *PY2011 Energy Savings Assistance Program Impact Evaluation*. Technical report. Portland, Oregon.
- 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): 1597–1644.
- Frandsen, Brigham R. 2017. “Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design When the Running Variable Is Discrete.” In *Advances in Econometrics*, 38:281–315. Emerald Publishing Ltd.

- Gerarden, Todd D., Richard G. Newell, and Robert N. Stavins. 2017. "Assessing the Energy-Efficiency Gap." *Journal of Economic Literature* 55 (4): 1486–1525.
- Gillingham, Kenneth, Amelia Keyes, and Karen Palmer. 2018. "Advances in Evaluating Energy Efficiency Policies and Programs." *Annual Review of Resource Economics* 10:511–532.
- Giraudet, Louis-Gaëtan, Sébastien Houde, and Joseph Maher. 2018. "Moral Hazard and the Energy Efficiency Gap: Theory and Evidence." *Journal of the Association of Environmental and Resource Economists* 5 (4): 755–790.
- Graff Zivin, Joshua, and Kevin Novan. 2016. "Upgrading Efficiency and Behavior: Electricity Savings from Residential Weatherization Programs." *The Energy Journal* 37 (4).
- Hillman, Bruce J., and Jeff Goldsmith. 2010. "Imaging: The Self-Referral Boom and the Ongoing Search for Effective Policies to Contain It." *Health Affairs* 29 (12): 2231–2236.
- Hirao, Yukiko. 1993. "Task Assignment and Agency Structures." *Journal of Economics and Management Strategy* 2 (2): 299–323.
- Holmstrom, Bengt. 2017. "Pay for Performance and Beyond." *American Economic Review* 107 (7): 1753–1777.
- Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7:24–52.
- Houde, Sébastien. 2018. "How Consumers Respond to Product Certification and the Value of Energy Information." *The RAND Journal of Economics* 49 (2): 453–477.
- Hubbard, Thomas N. 1998. "An Empirical Examination of Moral Hazard in the Vehicle Inspection Market." *The RAND Journal of Economics* 29 (2): 406–426.
- Ito, Koichiro. 2014. "Do Consumers Respond to Marginal or Average Price? Evidence from Non-linear Electricity Pricing." *American Economic Review* 104 (2): 537–63.
- Jacob, Brian A., and Steven D. Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *The Quarterly Journal of Economics* 118 (3): 843–877.
- Jensen, Michael C., and Kevin J. Murphy. 1990. "Performance Pay and Top-Management Incentives." *Journal of Political Economy* 98 (2): 225–264.
- Kerr, Steven. 1975. "On the Folly of Rewarding A, While Hoping for B." *Academy of Management Journal* 18 (4): 769–783.
- Khalil, Fahad, Doyoung Kim, and Dongsoo Shin. 2006. "Optimal Task Design: To Integrate or Separate Planning and Implementation?" *Journal of Economics & Management Strategy* 15 (2): 457–478.

- Larkin, Ian, and George Loewenstein. 2017. "Business Model-Related Conflict of Interests in Medicine: Problems and Potential Solutions." *Journal of the American Medical Association* 317 (17): 1745–1746.
- Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review* 90 (5): 1346–1361.
- . 2018. "Compensation and Incentives in the Workplace." *Journal of Economic Perspectives* 32 (3): 195–214.
- Lewis, Gregory, and Patrick Bajari. 2011. "Procurement Contracting with Time Incentives: Theory and Evidence." *The Quarterly Journal of Economics* 126 (3): 1173–1211.
- . 2014. "Moral Hazard, Incentive Contracts, and Risk: Evidence from Procurement." *The Review of Economic Studies* 81 (3): 1201–1228.
- Lewis, Tracy R., and David E. M. Sappington. 1997. "Information Management in Incentive Problems." *Journal of Political Economy* 105 (4): 796–821.
- List, John A. 2020. *Non Est Disputandum de Generalizability? A Glimpse into The External Validity Trial*. Technical report w27535. National Bureau of Economic Research.
- Loftus, Peter J., Armond M. Cohen, Jane C. S. Long, and Jesse D. Jenkins. 2015. "A Critical Review of Global Decarbonization Scenarios: What Do They Tell Us about Feasibility?" *Wiley Interdisciplinary Reviews: Climate Change* 6 (1): 93–112.
- Marion, Justin, and Jeremy West. 2019. "Dirty Business: Principal-Agent Problems in Hazardous Waste Remediation." https://people.ucsc.edu/~jwest1/articles/Marion_West_HazardSites.pdf. Accessed on 2020-1-15.
- McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, The Regression Discontinuity Design: Theory and Applications, 142 (2): 698–714.
- Metcalf, Gilbert E., and Kevin A. Hassett. 1999. "Measuring the Energy Savings from Home Improvement Investments: Evidence from Monthly Billing Data." *The Review of Economics and Statistics* 81 (3): 516–528.
- Muller, Nicholas Z., Robert Mendelsohn, and William Nordhaus. 2011. "Environmental Accounting for Pollution in the United States Economy." *The American Economic Review* 101 (5): 1649–1675.
- Myers, Erica. 2020. "Asymmetric Information in Residential Rental Markets: Implications for the Energy Efficiency Gap." *Journal of Public Economics* 190:104251.
- Oyer, Paul. 1998. "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality." *The Quarterly Journal of Economics* 113 (1): 149–185.

- Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1): 7–63.
- Research into Action. 2011. *Low Income Energy Efficiency Program 2009-2010 Process Evaluation*. Technical report. Portland, OR: Research into Action.
- Reynaert, Mathias, and James M. Sallee. 2021. "Who Benefits When Firms Game Corrective Policies?" *American Economic Journal: Economic Policy* 13 (1): 372–412.
- SCE, Southern California Edison. 2012a. *Southern California Edison Company's (U 338-E) 2012 Annual Report for 2011 Low Income Programs*. Technical report. Rosemead, CA: SCE.
- . 2012b. *The Joint Utilities' 2011 Amended Report on CARE Eligibility and Penetration Rates*. Technical report. Rosemead, CA: SCE.
- Schneider, Henry S. 2012. "Agency Problems and Reputation in Expert Services: Evidence from Auto Repair." *The Journal of Industrial Economics* 60 (3): 406–433.
- Shearer, Bruce. 2004. "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment." *The Review of Economic Studies* 71 (2): 513–534.
- Tayan, Brian. 2016. "The Wells Fargo Cross-Selling Scandal." Rock Center for Corporate Governance at Stanford University Closer Look Series: Topics, Issues and Controversies in Corporate Governance No. CGRP-62; Stanford University Graduate School of Business Research Paper No. 17-1.
- WM, Waste Management. 2018. *Lancaster Landfill & Recycling Center Service and Fee Schedule*. Technical report. Antelope Valley, CA: WM.