Free Riders and the High Cost of Energy-Efficiency Subsidies

Judson Boomhower Lucas W. Davis*

March 2013

Abstract

Economists have long argued that many recipients of energy-efficiency subsidies may be "free riders," getting paid to do what they would have done anyway. Demonstrating this empirically has been difficult, however, because of endogeneity concerns and other challenges. In this paper we use a regression discontinuity analysis to examine participation in a large-scale appliance replacement program in Mexico. Comparing behavior just on either side of several eligibility thresholds, we find that program participation increases with larger subsidy amounts, but that the magnitude of the increase is small. For example, when an air-conditioner subsidy increases from \$110 to \$170, the number of participants increases by only 21%. The large fraction of inframarginal households means that larger subsidy amounts are almost certainly not cost-effective. Overall, we find that accounting for free riders decreases the cost-effectiveness of the program by about 50%.

Key Words: Energy Efficiency, Regression Discontinuity, Free Riders

JEL: D12, H23, Q40, Q54

*(Boomhower) Agricultural and Resource Economics, University of California, Berkeley and Energy Institute at Haas. Email: boomhower@berkeley.edu. (Davis) Haas School of Business, University of California, Berkeley; Energy Institute at Haas; and National Bureau of Economic Research. Email: ldavis@haas.berkeley.edu. We are thankful to seminar participants at UC Berkeley for helpful comments. This research was supported in part under a general research contract from the California Energy Commission to the Energy Institute at Haas and by a National Science Foundation Graduate Research Fellowship to Boomhower. The authors have not received any financial compensation for this project nor do they have any financial relationships that relate to this research.

1 Introduction

Total energy consumption worldwide is forecast to increase 54% by 2030. While the energy mix is becoming somewhat less carbon-intensive, carbon dioxide emissions are still forecast to increase by 33% over the same period.[1] There is wide agreement among economists that the best approach for reducing carbon dioxide emissions and other negative externalities from energy use would be to use a carbon tax or cap-and-trade program. Although there has been some recent progress in this area, the vast majority of carbon dioxide emissions worldwide remain untaxed. And there are many countries, including the United States, where it seems unlikely that there will be large-scale carbon policy in the near term.

Instead what is receiving much attention is energy efficiency.[2] Proponents of energy-efficiency policies argue that they are a "win-win," reducing private energy expenditures while also helping to reduce the externalities associated with energy production and consumption. Many believe that there is scope for energy-efficiency investments to dramatically reduce energy consumption. McKinsey and Company (2009), for example, argues that energy-efficiency investments are a "vast, low-cost energy resource" that could reduce energy expenditures by billions of dollars per year. Subsidies and other energy-efficiency policies are presented as a way to motivate agents to tap into this "efficiency resource."

Despite all of the resources aimed at energy-efficiency programs, there is a surprisingly small amount of direct evidence evaluating their effectiveness.[3] Little is known, in particular, about what determines program participation. Many energy-efficiency programs work by subsidizing households and firms to adopt energy-efficient technologies. A fundamental question in evaluating the cost-effectiveness of these programs is how many of the participants would also have adopted these technologies with a lower subsidy, or even with no subsidy at all. Economists have long argued that many participants in energy-efficiency programs may be "free riders" (Joskow and Marron, 1992), but demonstrating this empirically has been difficult.[4]

[1]These statistics come from U.S. DOE, EIA, "International Energy Outlook", released September 2011, Figures 1 and 10. Total global energy consumption increased from 350 quadrillion Btu in 1990, to 500 quadrillion Btu in 2010, and is forecast to increase to 770 quadrillion Btu by 2030. Total energy-related carbon dioxide emissions increased from 20 billion metric tons in 1990, to 30 billion metric tons in 2010, and is forecast to increase to 40 billion metric tons by 2030.

[2]In the United States, for example, $17 billion dollars were targeted to energy-efficiency programs through the American Recovery and Reinvestment Act of 2009. See http://www1.eere.energy.gov/recovery. Customer-funded utility-administered energy-efficiency program spending in the United States totaled $4.8 billion dollars in 2010, and under current state and federal policies is projected to increase to $8.1 billion dollars per year by 2025 (Barbose et al., 2013).

[3]This lack of evidence was recently noted in Allcott and Greenstone (2012). The authors go on to argue that there is, "great potential for a new body of credible empirical work in this area, both because the questions are so important and because there are significant unexploited opportunities for randomized control trials and quasi-experimental designs that have advanced knowledge in other domains."

[4]The term "free rider" has long been used in the context of energy-efficiency to describe program

Determining the causal relationship between subsidies and technology adoption is challenging because one must construct a credible counterfactual for how much adoption would have occurred in the absence of the policy. Cross-sectional comparisons are misleading because places with generous subsidies are different from places with less generous subsidies. For example, "green" communities like Berkeley, California have more generous subsidy programs but also more eager adopters. Similarly, although programs change over time, it is difficult to separate the causal effect of these changes from other time-varying factors. Changes over time in energy-efficiency subsidies are endogenous, and thus correlated with changes in technology, pricing, and consumer preferences.

In this paper we address these challenges using a regression discontinuity (RD) analysis. Using evidence from a national appliance replacement program in Mexico, we compare behavior just on either side of several eligibility thresholds that result in observationally-equivalent households being offered very different subsidy amounts. The program is particularly conducive to an empirical analysis because of its large scale. The analysis is performed using household-level electric billing records for the universe of 25+ million Mexican residential customers, including data from 1 million households who participated in the program. This large sample size makes the RD estimates highly credible because we can focus attention on narrow bandwidths around the thresholds.

We find that program participation increases with larger subsidy amounts, but not by very much. For example, when an air-conditioner subsidy increases from \$110 to \$170 (both in U.S. 2010 dollars), the number of participants increases by only 21%. That is, our estimates indicate that more than 75% of households at this threshold would have participated in the program even with the lower subsidy amount. For the four main thresholds in our analysis we find that 70%+ of households are inframarginal. This large fraction of inframarginal households means that the larger subsidy amounts are almost certainly not cost-effective.

We then examine the implications of these results for cost-effectiveness and welfare. Under our preferred assumptions, the estimates imply that about one-third of participants would have replaced their appliances even with no subsidy whatsoever. These free riders add cost to the program without yielding reductions in energy consumption. Overall, we find that the cost per kilowatt hour abated is about 50% higher than a naive estimate that ignores free riders. Based on available estimates in the literature for the social cost of carbon and other key parameters, it appears that the total benefits of the program are almost certainly less than the costs.

Our paper is the first that we are aware of to use RD to study participation in an energy-efficiency program. We see broad potential for applying this approach in similar contexts. This distinction between marginal and inframarginal participants is a key part

participants who receive a subsidy for doing something they would have done anyway. This is distinct from the use of the term in public finance, which most economists are more familiar with. The well-known "free rider problem" in economics is the primary rationale for government investment in public goods. In this context individuals are "free riders" because, not internalizing the benefits to others, they underinvest in public goods.

of program design not just with energy-efficiency programs, but with many different types of government subsidies.[5] Although eligibility requirements vary widely across programs, the desire to simplify program design often results in the kind of discrete changes that we exploit here, making RD a natural approach to causal inference.

The paper is organized as follows. Section 2 sketches a conceptual framework and defines free riders. Section 3 discusses the relevant literature. Section 4 provides background about the program and describes the construction of the dataset. Section 5 presents the main results. Section 6 calculates the implied cost-effectiveness and discusses implications for welfare. Section 7 concludes.

2 Conceptual Framework

2.1 Technology Adoption With Externalities

Energy-efficiency subsidies are typically motivated as a way to reduce negative externalities from energy use when the optimal policy, a Pigouvian tax on emissions, is unavailable. In this section we propose a simple framework for thinking about the costs and benefits of energy-efficiency subsidies. We illustrate the welfare loss introduced by transfers to inframarginal participants and show how the optimal subsidy amount depends on the relative shares of marginal and inframarginal participants.

We begin with a simple graphical partial equilibrium analysis. Figure 1 describes the market for an energy-efficient technology. Along the x-axis is the number of adopters. Private demand is given by the downward-sloping private marginal benefits curve. The benefits of adoption vary across potential adopters due to differences in expected utilization and other factors. Supply is described by the private marginal cost curve. Here we have assumed constant marginal costs.

The privately optimal level of adoption is labeled in the figure as Q_0. These consumers adopt the technology purely on the basis of private benefits, even with no subsidies or other form of government intervention. If there are no externalities, then Q_0 is socially optimal. Once externalities are introduced, however, this is no longer the case. The figure illustrates the case in which there is a positive marginal external benefit from adoption, so the social marginal benefit exceeds private marginal benefit. The socially optimal level of adoption is labeled in the figure as Q^*. This optimum is defined as the intersection of the social marginal benefit and private marginal cost curves. The optimal subsidy is s^*.

The total amount paid in subsidies is indicated by the rectangle (A/B/C). Rectangle (A) is a transfer to free riders, i.e. adopters who would have adopted the energy-efficient technology even with no subsidy whatsoever. Rectangle (B) is excess payment to

[5] Gruber (2013) makes this distinction, for example, when discussing tax subsidies for charitable giving and home ownership.

adopters who are induced to participate because of the subsidy. Most adopters receive a subsidy that is more than the minimum amount necessary to induce them to replace. And rectangle (C) is the payment required to make adopters between Q_0 and Q^* indifferent between adopting and not adopting.

Before proceeding it is worth highlighting a couple of important assumptions. First, we have assumed that the external benefits from adoption are the same for all potential adopters. This is not necessarily going to be true because of differences in utilization levels and other factors. Recent studies have emphasized the potential gains from targeting energy conservation policies toward high value participants (Allcott and Mullainathan, 2012; Allcott et al., 2012). We have also assumed constant marginal costs. With increasing marginal costs the analysis is similar, but the incidence of the subsidy is partly on sellers. As a result there are "free riders" on both sides of the market. Subsidies increase the equilibrium price of the good, leading to higher revenues for sellers even for transactions which would have occurred anyway.

2.2 The High Cost of Energy-Efficiency Subsidies: Incorporating the Marginal Cost of Public Funds

The simple graphical analysis above ignores the distortionary effects of taxes and subsidies. Consider the following welfare function,

$$W = U(Q(s)) - C(Q(s)) + \tau Q(s) + Q(s)s - \eta Q(s)s. \tag{1}$$

Here $Q(s)$ is the quantity of technology adoption, which is a weakly increasing function of the subsidy s. $U(\cdot)$ and $C(\cdot)$ are private benefits and costs from the energy-efficient technology. In our graphical analysis, these correspond to the areas under the private marginal benefit and private marginal cost curves to the left of Q. τ is the constant external benefit of technology adoption derived from, for example, reduction in carbon dioxide emissions.

The final two terms reflect that while the subsidies represent a gain to adopters, they represent a loss to taxpayers. Moreover, collecting tax revenue may introduce distortions in labor and other markets. Summing the subsidy payments over all adopters gives the total amount of subsidies paid out by the program. The marginal cost of public funds (MCPF) is denoted η. If the MCPF is one, then there is no deadweight loss from taxation, and the transfers to adopters exactly offset the costs to taxpayers. If the MCPF exceeds 1, financing the transfers imposes efficiency costs.[6]

[6]This discussion is related to a substantial literature that examines general equilibrium interactions for environmental taxes. See, e.g., Bovenberg and Goulder (2002) and references therein. In this literature, environmental taxes create "tax interaction" and "revenue recycling" effects. Environmental taxes exacerbate pre-existing distortions in the economy, for example, by further reducing incentives to work (the tax interaction effect). On the other hand, environmental taxes generate revenues that can be used to finance cuts in other distortionary taxes (the revenue recycling effect). Most numerical studies have found that tax interaction is more important than revenue recycling, so there is no "double dividend",

The welfare change from a marginal increase in the subsidy is given by,

$$\frac{dQ}{ds}\Big[U'(Q(s)) - C'(Q(s)) + \tau - (\eta - 1)s\Big] - (\eta - 1)Q(s). \tag{2}$$

This complicated-looking expression is actually quite intuitive. The additional adoption induced by the subsidy increase is $\frac{dQ}{ds}$. The left-hand term gives the welfare effect of bringing these marginal participants into the program: private marginal benefits minus private marginal costs, plus external benefits, minus the distortions required to finance the subsidy payments to new participants. The right-hand term gives the welfare cost of increased payments to inframarginal households: The $Q(s)$ households already adopting the technology each receive an infinitesimal increase in subsidy payment, financed at the MCPF. If $\frac{dQ}{ds}$ is large relative to $Q(s)$, then the left-hand term matters more than the right-hand term. In other words, the welfare effects of increasing the subsidy depend on the relative numbers of marginal and inframarginal participants. As the subsidy level increases, Q becomes larger and payments to inframarginal households become more and more important.

If the MCPF is equal to one then equation (2) simplifies considerably and it is optimal to set the subsidy equal to marginal external damages (τ). This is exactly what we described in Figure 1 with s^* and Q^*. If the MCPF is greater than one, however, it is not optimal to set subsidies equal to marginal external damages. The optimal subsidy amount balances the benefits of increased adoption with the efficiency costs of larger transfers.

This analysis assumes that the government must pay all adopters the same subsidy. If it were possible to price discriminate, the government could pay each adopter the minimum amount they required to adopt. In this case, there would be no payments to inframarginal households and the right-hand term in equation 2 disappears. Perfect price discrimination moves the optimal subsidy amount closer to the first-best optimal subsidy, Q^*. This is because the total amount of transfers is reduced to the minimum level needed to induce participation by each adopter. Of course in practice, equity concerns, imperfect information, and other factors severely limit the government's ability to price discriminate. So the government must typically offer the same subsidy to all potential adopters, bearing the additional efficiency costs of transfers to inframarginal households.

The optimal uniform subsidy problem is similar to the standard monopsonist problem. When procuring inputs a monopsonist stops short of equating marginal revenue with

and optimal tax rates on externalities are generally below marginal damages (Bovenberg and Goulder, 1996). The subsidy case has received less attention. An important exception is Parry (1998), which points out that subsidies have a negative revenue recycling effect and a positive tax interaction effect. In other words, additional taxation is required to finance subsidies, but they may alleviate pre-existing distortions. In particular, subsidies that reduce the price of consumption goods increase the real household wage, thus encouraging labor and reducing the pre-existing distortion from labor taxation. Parry finds that, "In general the revenue-financing effect exceeds the tax-interaction effect, and the optimal level of subsidy is still positive, but somewhat below that implied by a first-best analysis." In the interest of simplicity, we let η represent the MCPF net of the tax interaction effect for subsidies.

input price, because it takes into account that each additional unit of input requires paying a larger input price to both marginal and inframarginal suppliers. The same goes for a government "procuring" the adoption of energy-efficient technologies. The government increases the subsidy amount only until the social marginal cost of increasing participation is equal to the social value of the benefits derived from the decrease in externalities. Not only is the optimal subsidy reduced by the need to finance costly transfers to new participants; it is further reduced because of the transfers to inframarginal households.

In most contexts it will also be important to take into account additional non-private costs. In particular, there are substantial indirect costs from administering energy-efficiency programs. These costs include marginal costs like the administrative and transaction costs of remitting subsidies, and fixed costs like program design and advertising. Incorporating marginal indirect costs causes the optimal subsidy amount to be even lower. And if the fixed costs are larger than the difference between total benefits and total costs, then it is not worth having the program at all.

The main takeaway from this section is that, in general, the optimal subsidy amount is lower than marginal external benefits. The welfare effects of a subsidy depend critically on the effect of the subsidy on program participation. If $\frac{dQ}{ds}$ is small relative to the number of existing participants, then the benefits from increased adoption will be small relative to the efficiency costs of the payments to inframarginal participants. Accordingly, this is where we focus our attention in the empirical analyses which follow.

3 Related Literature

We have not seen previous studies that conceptualize the free rider problem quite like we have done in Section 2. But this general intuition has been around for a long time. Economists have long argued that many participants in energy-efficiency programs may be free riders, and that this is important to take into account when measuring cost-effectiveness. In one of the first papers to seriously consider this issue, Joskow and Marron (1992) review data from a representative sample of U.S. energy-efficiency programs, finding that most ignore free riders when calculating the cost per kilowatt hour saved and other measures of cost-effectiveness. They conclude that accounting for free riders and other issues likely leads utilities to overestimate the cost-effectiveness of these programs by a factor of two or more.

In the extreme, if all households are inframarginal, then an energy-efficiency subsidy has no environmental benefit at all. In our conceptual framework this could occur, for example, if supply of the good was constrained at a level below Q_0. An example of this occurred in 2006 and 2007 with federal tax credits for the Toyota Prius. These credits were imposed during a period in which there were long waitlists for the Prius, so the credits had little to no impact on the total number of vehicles sold (Sallee, 2011).

In most cases, however, supply is not constrained. This brief period of shortages with the Toyota Prius was unusual and in most cases manufacturers are able to adjust quantities to subsidy-induced increases in demand. A couple of studies have used panel data to measure the effect of hybrid vehicle subsidies on the adoption of hybrids. Chandra et al. (2010) examine subsidies offered by Canadian provinces, and Gallagher and Muehlegger (2011) focus on subsidies offered by U.S. states. Both studies find statistically significant effects of subsidies on adoption, thus ruling out the null hypothesis that all participants are free riders.

In related work, Mian and Sufi (2012) find that the U.S. vehicle replacement subsidy program *Cash for Clunkers* induced the purchase of 370,000 cars during July and August 2009. Under the program there were almost 680,000 total purchases, so their estimate implies that about 50% of participants were free riders. Mian and Sufi (2012) also examine vehicle purchase behavior in the months following the program, finding that the initial impact was considerably offset by fewer replacements over the following 10 months. The *Cash for Clunkers* program is a bit different than most energy-efficiency programs in that it was always viewed primarily as a form of fiscal stimulus.

With regard to residential energy efficiency, there is surprisingly little evidence on free riders. A number of papers have used U.S. utility-level data to estimate the effect of total spending on energy-efficiency programs on electricity consumption (Loughran and Kulick, 2004; Auffhammer et al., 2008; Arimura et al., 2012). These analyses address free riders indirectly by comparing realized aggregate savings to engineering estimates. Loughran and Kulick (2004), for example, postulate that the relatively small savings they observe could be due to free riders. While certainly plausible, there are many reasons why actual realized savings differ from engineering estimates, and it is impossible with aggregate data to determine which explanations are most important. Moreover, because utilities are typically running many energy-efficiency programs at the same time, it is hard to learn about which types of energy-efficiency programs are most effective, or about how differences in program design impact effectiveness.

More broadly, it is worth pointing out that this panel-data approach to evaluating energy-efficiency programs faces several challenges. First, there are important concerns about the quality of utility-level data on energy-efficiency spending. There are a large number of missing values, even for large utilities, and a lack of consistency both across utilities and over time in what gets reported. Second, and perhaps most importantly, spending is endogenous. This has been typically ignored in these analyses with the exception of Arimura et al. (2012) which instruments for current spending using lagged spending and other variables. Third, the estimates in these studies tend to be estimated without much precision. For example, Arimura et al. (2012) find that energy savings come at an average cost of 5 cents per kilowatt hour, but the estimate has a 90th percentile confidence interval that goes from 0.3 cents to nearly 10 cents.

Similar challenges arise in the vehicle adoption studies described above. The cross-sectional comparisons in Mian and Sufi (2012) rely on states with few "clunkers" being a good counterfactual for what would have happened in states with more "clunkers". And studies like Chandra et al. (2010) and Gallagher and Muehlegger (2011) must address the endogeneity of subsidy levels. Decisions by state and provincial governments to adopt these subsidies are not random, and choices may be correlated with both time-invariant factors such as state and province demographics, as well as with time-variant factors such as gasoline prices, macroeconomic conditions, and tastes for green products.

Our analysis sharply differs from these previous studies in that we use RD rather than difference-in-differences. Perhaps the closest existing study is Ito (2012) which uses an RD analysis to examine a recent California policy which paid households to reduce their electricity consumption. Exploiting eligibility requirements for the program which required households to have moved into their homes before a particular date, Ito finds that participants in hot areas in California reduce their consumption 5-10% compared to observationally-equivalent ineligible households. Ito (2012) is a study of an energy *conservation* program, so while this response may in part reflect adoption of energy-efficiency technologies, it also likely reflects changes in the utilization of air conditioners and other energy-using durables.

An important advantage of RD is that it requires a considerably weaker identifying assumption than difference-in-differences (Hahn et al., 2001; Lee and Lemieux, 2010). With panel studies the counterfactual is constructed using cross-sectional and time-series comparisons and one must assume that energy-efficiency subsidies are not correlated with unobservables. In contrast, the counterfactual in our analysis is constructed by comparing behavior on either side of eligibility thresholds, with households who just narrowly missed qualifying for a more generous subsidy as a comparison group for those who did qualify. As we will show in our analysis, households on both sides of the threshold are observationally-equivalent, so treatment is as good as randomly assigned and these highly-localized comparisons yield unbiased estimates of the causal effect of the larger subsidy.

4 Program Background and Construction of Dataset

4.1 Background

Our empirical analysis focuses on a large-scale energy-efficiency program in Mexico. The program was launched in March 2009 and was ended in December 2012. During this period, the program helped 1.5 million households replace their refrigerators or air conditioners with energy-efficient models. To participate in the program a household had to have a refrigerator or air conditioner that was at least 10 years old and agree to purchase a new appliance of the same type. The new appliance had to exceed Mexican energy-efficiency standards by at least 5%.

Table 1 describes the subsidies available under the program. The cash grants came in three different amounts, approximately corresponding to \$170, \$110, and \$30 (all in U.S. 2010 dollars). For reference, the mean prices of appliances purchased through the program were \$406 for air conditioners and \$427 for refrigerators. Participation was limited to households with relatively high electricity consumption. Within this group of eligible households, households with lower average historical electricity consumption received larger cash grants. This structure was designed to target the larger subsidies to lower-income households.[7]

In addition to direct cash payments, the program offered on-bill financing at subsidized interest rates. These loans had an annual interest rate of 13.8% and were repaid over four years. If households would have otherwise financed these purchases using credit cards, the economic value of each \$1 of loans would be about 18 cents.[8] That is, each dollar in direct cash payment would have been worth about five times as much to participants as a dollar in loans. From the government's perspective, providing direct cash payments was much more costly than providing subsidized loans. In fact, the government's cost of borrowing may have even been lower than the 13.8% charged to participants. In addition, default risk was low because these loans were repaid directly through electricity bills. If a household defaulted it would lose electricity service.

As illustrated in Table 1, the maximum loan amounts increase across consumption categories. At the middle two thresholds, the increase in maximum loan amount is exactly equal to the decrease in cash subsidies. For example, as the cash subsidy decreases from \$170 to \$110, the maximum loan amount increases by \$60. For a household who would have used this entire additional loan amount and who otherwise would have financed the purchases using a credit card over four years at typical credit card rates the economic value of the change is about \$49 instead of \$60.

The maximum loan amount increases sharply in the highest consumption category. For both appliances in this last category the increase in maximum loan amount at this threshold (\$280) greatly exceeds the decrease in direct cash payments (\$30). For households with a high cost of borrowing and who wish to buy relatively expensive appliances, the economic value of the program actually decreases at this threshold. The large offsetting change in the maximum loan amount makes it more difficult to interpret any observed changes in behavior at this last threshold. Accordingly, in the empirical analysis which follows we emphasize the other thresholds where there is a clear and unambiguous change

[7] We are not aware of any studies that measure the correlation between electricity consumption and household income in Mexico. Recent studies from the United States indicate that this correlation is surprisingly small, in part because higher-income households tend to live in more energy-efficient homes. See Borenstein (2012) and Borenstein and Davis (2012).

[8] According to Banco de Mexico, "Indicadores Básicos de Tarjeta de Crédito," October 2012, Table 1, credit cards in Mexico charged an average interest rate of 25.3% in 2011. This is not a perfect measure of the private cost of borrowing. On the one hand, not all households in Mexico have access to credit cards, and the interest rates on alternative forms of borrowing will vary. On the other hand, collateralized loans for the purchase of durable goods (for example, department store credit programs) typically can be made at lower rates. More than half of participants took out the maximum loan amount, which suggests that the market cost of borrowing exceeds 13.8% for most consumers.

in the value of the program.

4.2 Construction of the Dataset

A key feature of our analysis is the use of high-quality, household-level microdata, both about program participants and about the entire pool of potential participants. Often data from energy-efficiency programs includes information about participants, but little or no information about non-participants. The fact that we observe non-participants is important because, as usual, the objective in the empirical analysis is to construct a credible counterfactual, and this is hard to do without information about the broader pool.

The first component of this database is a two-year panel dataset of household-level electric billing records describing bimonthly electricity consumption and expenditure for the universe of Mexican residential customers from May 2009 through April 2011. The complete set of billing records includes data from 26,278,397 households. This represents the entire pool of potential participants in the program. From this complete set of records we drop 15,262 households (<0.001%) for which the records are improperly formatted and 1,113 households for whom no state was indicated. We also drop 491,788 observations (1.9%) with zero reported usage in every month of the panel.

The second component of this database is a record of all households who participated in the program between March 2009 and June 2011. In the complete dataset there are a total of 1,162,775 participants. We drop 51,823 participants (4.5%) for whom no installation date for the new appliance was recorded. We then merged the remaining list with the electric billing records using customer account numbers. For 86% of program participants, there was a household with the identical account number in the electric billing records. We exclude the remaining 14% for whom there is no perfect match rather than attempting to perform some kind of a"fuzzy" match for account numbers that were presumably entered incorrectly in one of the two datasets, leaving us with 957,080 total program participants. For each participant, we know the exact dates of purchase and replacement, whether the appliance replaced was a refrigerator or an air-conditioner, and the amount of direct cash subsidy and credit received.

Participating retailers determined which subsidy amount a household was eligible for by entering the household's account number into a website designed for this purpose. This centralization is important from an empirical design perspective because it means that there was no scope for retailer discretion in assigning subsidy amounts. The formula used to calculate average historical electricity consumption was complicated. In particular, exactly which billing cycles were used for the calculation differed between the refrigerator and air-conditioner programs, and differed across locations in Mexico. For example, in hot parts of the country, only non-summer bills were used in assessing eligibility for the refrigerator program. Moreover, "summer" is defined differently in different parts of the country. For details see Appendix 1.

We used our database to calculate average historical electricity consumption for each household. We then applied the program rules to determine which subsidy each household should have been eligible for. For program participants we know which subsidy they actually received, so we can use this information to validate our construction of the running variable. And as we show in the following subsection, we are able to correctly predict subsidy levels for 97%+ of all households.

5 Empirical Strategy and Results

5.1 Estimating Equation

Our empirical strategy exploits the discrete eligibility thresholds that determined whether a household was eligible for zero subsidy, \$30, \$110, or \$170. There are six total thresholds; three for air-conditioners and three for refrigerators. At each of these thresholds, we use a standard RD estimating equation (Lee and Lemieux, 2010):

$$1[Participate]_i = \alpha + f(X_i) + \rho 1[Below\ Threshold]_i + \eta_i \tag{3}$$

where $1[Participate]_i$ is an indicator variable equal to one if a household participated in the program and zero otherwise. We include in the regression $f(X_i)$, a polynomial in average historical electricity consumption, and $1[Below\ Threshold]_i$ an indicator variable equal to one if the household's average historical electricity consumption was below the given threshold. The coefficient of interest is ρ, which measures the discontinuous change in program participation at the threshold. Moreover, we normalize X_i to be equal to zero at the threshold so the coefficient α corresponds to the predicted probability of participating just below the threshold, and $\alpha + \rho$ corresponds to the predicted probability just above the threshold.

The error term η_i captures unobserved determinants of the participation decision. Hahn et al. (2001) show that identification with RD requires that the conditional mean function $E[\eta_i | X_i]$ is continuous at the discontinuity. Participation in energy-efficiency programs is driven by a large number of factors so, in general, one would expect η_i to be correlated with X_i and thus with $1[Below\ Threshold]_i$. This correlation does not invalidate RD estimates. In the limit, one is comparing outcomes with an arbitrarily small neighborhood around each threshold and the identifying assumption requires only that there not be a discontinuous change in these other factors that occurs exactly at the eligibility thresholds.

The idea behind RD is that within an arbitrarily narrow window around these thresholds households are observationally-equivalent. For example, the average characteristics of households who consume 499 kilowatt hours per month are likely to be very similar to the average characteristics of households who consume 500 kilowatt hours per month. Of course, in practice there are few observations within an arbitrarily small neighborhood around these thresholds, and so there is a tradeoff between bias and efficiency. Flexibly parameterizing the polynomial $f(X_i)$, allows us to expand the sample to include households farther away from the threshold. We selected a cubic polynomial based on graphical analysis because it seems to adequately describe the underlying relationship between program participation and average historical electricity consumption. Higher-order polynomials tended to perform poorly, behaving erratically at the thresholds and increasing and decreasing dramatically to fit individual observations.

We report results using a variety of different bandwidths. In our preferred specification, we include all households within 100 kilowatt hours of the thresholds for air conditioners, and within 50 kilowatt hours of the thresholds for refrigerators. The wider bandwidth for air conditioners reflects that these thresholds were much higher (500, 750, and 1000 kilowatt hours compared to 175, 200, and 250) and the density of households in that part of the distribution is considerably lower. With refrigerators, the thresholds are close enough together that, in some cases, the bandwidth includes more than one threshold. One possibility would be to use a single RD estimating equation for all thresholds. This would mean, however, that the bandwidths would not be symmetric. Instead, in the results which follow we use one estimating equation per threshold, but we include intercept terms for any additional thresholds.

5.2 Validity of Research Design

Before presenting our main results, we perform several tests aimed at assessing the validity of our research design. In section 5.2.1 we test whether there is a discontinuity at our six thresholds. These discontinuities are the basis for our research design so it is critical that we verify them empirically. Then in sections 5.2.2 and 5.2.3 we test for manipulation of the running variable. If households are able to change their behavior to qualify for more generous subsidies this would potentially violate the identifying assumption for RD.

5.2.1 The Discontinuity

Figure 2 plots the fraction of participants who received the larger subsidy as a function of average historical electricity consumption. We show separate plots for each of the six thresholds. The dots represent mean values for three kilowatt hour, non-overlapping bins. The sample used to construct these figures includes all households who replaced appliances through the program in 2011. We focus on 2011 because calculating eligibility for some earlier participants would require data before May 2009, the first month in our dataset.

In all six cases there is a clear discontinuity at the threshold. Almost all households with average historical consumption below the threshold receive the higher subsidy amount and almost all households with average historical consumption above the threshold receive the lower subsidy amount. Figures 2A, 2B, and 2C describe the thresholds for air-conditioners. The share of participants receiving the larger subsidy falls from near one to near zero at all three eligibility thresholds.

The discontinuities are slightly less sharp for refrigerators (Figures 2D, 2E, and 2F). Near the threshold, a small number of participants with average historical electricity consumption above the threshold receive the higher subsidy, and vice-versa below the threshold. This is due to measurement error in our reconstruction of average historical electricity consumption. As we explain in more detail in Appendix 1, the program rules for refrigerators were especially complicated. Eligibility for the air conditioner subsidies was calculated using summer consumption during the previous calendar year, regardless of when during the year the appliance was purchased. Eligibility for refrigerators, in contrast, was based on the most recent 12 months of billing history.

We observe the dates for each billing cycle, but not the dates when meter readings were entered into the central system. Consequently, with refrigerators it is impossible for us to be sure which bills were in the system at the moment a household participated in the program. In other words, we cannot be sure for each household which billing cycles were used when calculating eligibility. This means that for some observations we measure average historical consumption with error because we include months that were not actually included when that household's eligibility was calculated. The direction and magnitude of this error will vary across households. Importantly, however, for most households we measure historical consumption with no error because we choose the same billing cycles as were actually used to calculate eligibility.

For the set of observations where average historical electricity consumption is measured with error, the subsidy received will not change discontinuously at the observed threshold. Random noise in the running variable instead leads to a smooth "S"-shaped relationship between subsidy amount and historical average electricity consumption for these observations. Thus, for the full sample, the observed discontinuity in subsidy amounts at the threshold is reduced by the fraction of households for whom the running variable includes measurement error. In Figures 2D, 2E, and 2F, the observed change in the probability of receiving the higher subsidy is about 0.8 - 0.9.

Battistin et al. (2009) show that this type of measurement error biases sharp RD estimates downward in proportion to the fraction of observations measured with error. However, as long as the measurement error is uncorrelated with the subsidy amount offered conditional on the value of the running variable, the fuzzy RD estimator will correctly estimate the effect by "canceling out" the measurement error (See Appendix 2). In our case, this is a reasonable assumption given that it seems unlikely that uncertainty about bill delivery dates would be correlated with a household's electricity consumption. We return to this issue below. In practice, because a small share of observations are measured with error, sharp RD and fuzzy RD produce very similar estimates.

Table 2 shows the number of total households and number of participants within our preferred bandwidths above or below each threshold. It also shows the number of participants within that range for whom we incorrectly predict historical usage. We correctly predict subsidies for at least 97% of participants at all thresholds. Moreover, the incorrect predictions are approximately symmetric, with near-equal numbers of participants receiving higher and lower subsidy amounts than we predict.

5.2.2 Smoothness of the Running Variable

A standard concern with RD analyses is manipulation of the running variable. When agents can completely or partially manipulate their treatment status this represents a substantial threat to the identifying assumption. Even a modest amount of selection can mean that agents on one side of the threshold may not be a good comparison group for agents on the other. Understanding any strategic behavior in response to eligibility thresholds is also of significant independent interest because it may introduce additional inefficiencies, as agents alter their behavior to qualify for more generous subsidies.[9]

Figure 3 plots the frequency distribution of average historical electricity consumption for all households. We use three kilowatt hour bins and include separate plots for air conditioners and refrigerators because the measure of average historical electricity consumption used to determine eligibility was different for the two appliance types. Examining the smoothness of the running variable is a valuable first test for manipulation (McCrary, 2008). If households are changing their behavior to qualify for the more generous subsidy, we would expect to see bunching to the left of the thresholds, with more households than expected just barely qualifying for the higher subsidies, and with fewer households than expected just failing to qualify.

For both appliance types, the frequency distributions appear smooth across the eligibility thresholds. This lack of evidence of manipulation is perhaps not surprising given that it is difficult for a household to control its average historical electricity consumption. Perhaps most importantly, this is *historical* consumption, so at the time of participating in the program, there is no scope for the household to go back and change its electricity consumption patterns in the past.

Pushing this a bit farther, one could ask whether a forward-looking household could have changed their electricity consumption patterns (e.g. turning off lights, etc.), in the months leading up to participation. Although possible in principle, these changes would require a household to be unusually well-informed about the program and about

[9] In related work, Sallee and Slemrod (2012) examine the strategic response of automakers to discrete changes in the *Gas Guzzler Tax* which penalizes vehicles with poor fuel economy. At one so-called "notch", for example, a car with a 14.5 miles-per-gallon (mpg) rating is subject to a $4500 tax, while a car with 14.4 mpg is subject to a $5400 tax. Their results indicate that vehicle fuel economy can be relatively easily manipulated, for example, by making minor modifications to vehicle weight, engine tuning, and tires. They find strong evidence of bunching on the low-tax side of notches and then describe the welfare consequences of this behavior, highlighting the inefficiencies introduced by this local manipulation.

their own electricity consumption patterns. The formula for calculating average historical consumption was available online, but was complicated. A household would have had to put considerable effort into understanding and then performing the calculation. Moreover, the description of the formula was confusing enough so that even a careful household would have been uncertain about whether or not they performed the calculations correctly.

5.2.3 Testing for Strategic Behavior

In this section we perform a final test aimed at a more subtle form of strategic behavior. In particular, we test for whether households in the refrigerator program strategically *delayed* participation when they were close to a threshold. Suppose a household applies for the program and somehow learns that they just narrowly missed eligibility for one of the more generous subsidies.[10] At least in theory, this household could wait for a billing cycle (or more), perhaps while simultaneously taking steps to reduce electricity consumption, and then reapply. This kind of strategic delay would lead us to find more participants just on the generous side of these thresholds relative to a program for which eligibility was assigned only once.

This type of strategic delay is much less feasible for air conditioners. Subsidy amounts for air conditioner replacement were based on summer consumption in the *previous* calendar year, so a household would need to wait six or more months before changes in electricity consumption could affect their subsidy eligibility (see Section 4.2 and Appendix 1).

To test for strategic delay in the refrigerator program, we examine the subsidies that households would have received if they had purchased their new appliance one billing cycle earlier. If there is no strategic delay, the probability of a household's average historical electricity consumption falling enough to increase their subsidy eligibility should be the same for participants and non-participants. In contrast, if participants are more likely than non-participants to be eligible for larger subsidies at purchase than sixty days earlier, this may indicate strategic delay.

We implement this test using households within 15 kWh below the 175 kWh and 200 kWh thresholds (i.e., households who would have received the larger subsidy at each threshold). To minimize the influence of time-varying shocks like weather, we examine participants from a relatively narrow time window: February 20 to March 10, 2011. We calculate average historical electricity consumption for participants on the day that they

[10]As described earlier, participating retailers determined whether a household was eligible for the program by entering the household's account number into a website designed for this purpose. Households could not access this site without a retailer's login and password. The website reported the subsidy level for which a household qualified, but did not describe the intermediate calculations which determined eligibility or let a household know when it was close to a more generous subsidy level. Thus, in general, there was no obvious mechanism for households to learn that they had just narrowly missed an eligibility threshold.

replaced, and for non-participants on March 1, 2011. Then, for both groups, we also calculate their average historical electricity consumption sixty days earlier.

Table 3 reports the results. Overall, historical average electricity consumption changed in a way that increased subsidy eligibility for about 25% of households. The similarity across participants and non-participants suggests these changes were driven more by time-varying shocks like weather than by strategic delay. At the 175 kWh threshold, eligibility increased for 25.4% of participants and 23.8% of non-participants. Taken literally, this result suggests that about 1.5 out of every 100 refrigerator participants might have strategically timed their refrigerator purchase to maximize their subsidy. Using a χ^2 test, we test the null hypothesis that the probability of being eligible for a smaller subsidy two months ago is equal across participants and non-participants. We find that the difference is weakly statistically significant (at the 10% level). At the 200 kWh threshold, eligibility increased for 26.6% of participants and 26.1% of non-participants. This difference is not statistically significant.

To summarize, we find weak evidence that there may have been a very small amount of strategic purchase timing for refrigerators at one threshold and no evidence of strategic purchase timing at the other. Changes in subsidy eligibility over time appear to be driven much more by time-varying shocks like weather than by strategic behavior.

5.3 Graphical Evidence

We now turn to our main results, first presenting graphical evidence and then reporting regression estimates in Section 5.4 and alternative specifications in Section 5.5. Figures 4A and 4B plot program participation against average historical electricity consumption for air conditioners and refrigerators, respectively. The dots represent the percentage of households within a three kilowatt-hour historical usage bin who participated in the program.

It is first worth noting that there is essentially no participation by households who used less than the minimum levels of electricity required for participation. This is not surprising given the way the program was administered. Retailers had to document eligibility for all participants using the program website, and thus had no discretion in determining subsidy amounts. The small number of participating households to the left of the minimum eligibility thresholds for refrigerators reflects, instead, a small amount of measurement error in average historical electricity consumption.

For air conditioners, participation increases steadily between 250 and 500 kilowatt hours, levels off between 500 and 750, and then declines slowly after 750. Our main interest is in behavior at the 500, 750, and 1000 kilowatt hour thresholds. In the first two cases, there appears to be a discontinuous decrease in participation at the threshold. The second decrease is particularly visible and appears to occur exactly at the threshold in which the subsidy amount decreases from $110 to $30. It is difficult to make strong statements based on this graphical evidence because the participation rate moves around across bins, but at this threshold the participation rate appears to drop from about 1.5% to about 1%. At the final threshold there does not appear to be any discontinuous change in participation.

For refrigerators, participation follows a similar inverted "U" pattern, peaking at about 1.75% participation near 150 kilowatt hours and then decreasing steadily between 150 and 300 kilowatt hours. Again, our primary interest is behavior at the thresholds. At both the 175 and 200 kilowatt hour thresholds there are visible discontinuous decreases in participation. At the 250 kilowatt hour threshold there is no apparent decrease. This overall pattern is similar to what is observed for air conditioners, with decreases at the first two thresholds and no visible decrease at the third threshold.

Figure 5 describes these six thresholds with more detail by substantially narrowing the window and switching to one kilowatt hour bins. In addition to plotting the percentage of households who participated in the program for each bin, the figure plots cubic polynomial functions of average historical electricity consumption with intercepts at the thresholds. For air conditioners, there again are discontinuous decreases in participation at the 500 and 750 kilowatt hour thresholds. Participation falls by about 0.3 percentage points in Figure 5A and by about 0.5 percentage points in Figure 5B. For refrigerators, there are clear decreases at the 175 and 200 kilowatt-hours thresholds (Figures 5D and 5E), both of about 0.2 percentage points.

Figures 5C and 5F show that there is no observed change in participation for either appliance when the subsidy falls from \$30 to \$0. As we discussed in Section 4.1, this threshold was different from the others in that there was a large offsetting increase in the maximum loan amount. We were expecting to see a much smaller change in participation at this threshold, and the data appears to bear this out. The near zero change in participation implies that, on average, the increase in maximum loan amount (\$280) had about the same value to households as the \$30 decrease in cash. We find this very interesting, but in order to focus primarily on the cash subsidies we limit the empirical analysis which follows to the other thresholds where there is a clear and unambiguous change in the value of the program.

5.4 Regression Estimates

Table 4 reports RD estimates and standard errors from four separate regressions. For each threshold we report estimated program participation at each side of the threshold as well as the percent change in participation. Because we have normalized the running variable to be equal to zero at the threshold, these statistics come right out of our estimating equation. From each regression, column (2) reports the estimated intercept, column (3) reports our estimate of the intercept plus our estimate of the discontinuous change at the threshold, and column (4) reports the percent change between the two. Columns (5) and (6) report the implied price elasticity and cost per additional program participant (both defined below).

Consistent with the graphical evidence, participation increases at all four thresholds. The magnitudes, however, are reasonably small. With air conditioners, increases are 21% and 44%. The larger increase in participation corresponds to the subsidy increase from \$30 to \$110. This is the largest subsidy increase and was the most obvious change in the graphical evidence for air conditioners. Both changes are statistically significant. For refrigerators the estimated changes in participation are similar, 20% and 35%. As with air conditioners, the larger increase corresponds to the subsidy increase from \$30 to \$110. The estimates for refrigerators are more precisely estimated because of the large number of households with average historical electricity consumption near these thresholds.

These relatively small changes in participation imply that most participants at these thresholds are inframarginal. That is, most households who just barely qualified for a \$170 subsidy, for example, would have participated even if they had only received \$110. The percent inframarginal can be calculated by dividing column (2) by column (3). For example, when the air conditioner subsidy increases from \$110 to \$170, our estimates imply that 83% ($\frac{1.34}{1.62} = 0.83$) of households are inframarginal. Across thresholds the percentage inframarginal ranges from 70% to 84%.

In column (5) we report the implied price elasticity of demand for appliance replacement. These are calculated for each threshold by dividing the percent change in participation by the percent change in the average price of appliance replacement net of the subsidy. In calculating this net price we use the average price paid by program participants at the threshold.[11] For the change in net price, we incorporate both the direct cash payments and the implied cash value of the subsidized loans (assuming a 25.1% cost of borrowing; see Section 4.1). Demand is relatively elastic. For both appliances, the elasticities are near one at the first threshold, and near two at the second.

These estimated price elasticities are of significant independent interest. Wolfram. et al. (2012) argues that demand for residential appliances will have an enormous influence on future energy consumption growth in low- and middle-income countries. Appliance prices have been falling for decades and our estimates imply that continued decreases will accelerate the rate at which appliances are replaced. If households are more quickly replacing appliances this means that improvements in energy-efficiency will more quickly be reflected in the appliance stock.

In column (6) we report the implied cost of increasing program participation at each threshold. An example is helpful. When the air conditioner subsidy increases from \$110 to \$170, the cost of increasing participation is the \$170 paid to each marginal household (17% of all households at the threshold), plus an additional \$60 (\$170-\$110) paid to each inframarginal household. This is then divided by the fraction of households who

[11] Specifically, we use the average price paid by participants at the low-subsidy side of each threshold. For air conditioners, these prices were \$402 at the 175 kWh threshold and \$402 at the 200 kWh threshold. For refrigerators the prices were \$425 at the 500 kWh threshold and \$427 at the 750 kWh threshold.

are marginal,

$$\frac{(.17)(\$170) + (.83)(\$60)}{(.17)} = \$459. \tag{4}$$

Across thresholds, the cost per additional participant for air conditioners ranges from \$290 to \$472, and for refrigerators ranges from \$350 to \$500. These are important statistics from a program design perspective. Larger subsidies increase program participation. But as we illustrated in Section 2, encouraging adoption becomes increasingly expensive because a higher and higher fraction of participants are inframarginal.

5.5 Alternative Specifications

Table 5 reports regression estimates from alternative specifications. For each specification we report the estimated percent change in participation at each threshold. Column (3) reports our baseline estimates, identical to the estimates reported in Table 4. Columns (2) and (4) assess the sensitivity of our estimates to alternative bandwidths. Overall, the results are very similar across bandwidths. Moreover, there is no consistent pattern. As we move to different bandwidths, some point estimates increase while others decrease.

Column (5) reports estimates from a fuzzy RD specification. In this specification, we scale the estimates by the size of the discontinuity at the threshold following Hahn et al. (2001) and Battistin et al. (2009). Specifically, we run a first stage regression of an indicator for the larger subsidy ($1[Larger\ Subsidy]$) on ($1[BelowThreshold]$) and a cubic polynomial of average historical consumption, $g(X)$,

$$1[Larger\ Subsidy]_i = \alpha + g(X_i) + \gamma 1[Below\ Threshold]_i + \epsilon_i. \tag{5}$$

We then divide our baseline estimates by γ to remove any bias caused by measurement error (see section 5.2.1 and Appendix 2). The estimates are very similar with the fuzzy RD specification. The air conditioner estimates are essentially identical to the sharp RD estimates, consistent with the near perfect discontinuity observed in Figures 2A, 2B, and 2C. For refrigerators, the scaling increases the point estimates by between four and seven percentage points, consistent with the graphical evidence in Figures 2D, 2E, and 2F, which exhibit a small amount of measurement error in average historical electricity consumption.

6 Cost-Benefit Analysis

6.1 Accounting for Free Riders in Measures of Cost-Effectiveness

Table 6 reports several different measures of cost-effectiveness. Row (1) reports naive estimates that ignore free riders. These numbers come directly from Davis et al. (2012), which estimates the electricity savings per replacement from the program using difference-in-difference and matching estimators. Using a 5% discount rate and other assumptions, they find that the program reduces electricity consumption at a direct program cost of $0.30 per kilowatt hour and reduces carbon dioxide emissions at $521 per ton.

Row (2) reports estimates of cost-effectiveness which account explicitly for free riders. In particular, these calculations exclude "savings" from households who, according to our estimates, would have replaced their appliances even without the program. For this calculation we use the elasticity estimates in Table 4. For participants who received the $170 cash payment, we use the elasticity corresponding to the threshold between $110 and $170, and for participants who received $30 or $110, we use the elasticity corresponding to the threshold between $30 and $110.

In using these elasticities, we are assuming that behavior at the thresholds is representative of households in a given subsidy tier. This is a weak assumption for a short tier like the $110 tier for refrigerators, but is a considerably stronger assumption for longer tiers like the $170 tier for both appliances. The need for some kind of assumption here reflects the inherent tradeoff with RD. The advantage of RD is that it yields highly credible estimates. The disadvantage is that these estimates are valid only at the discontinuities, and some type of assumption is necessary if we are to predict behavior for households outside the immediate neighborhood of these thresholds or to make broader statements about cost-effectiveness.

Under these assumptions, our estimates imply that 31% of participants are free riders, and thus would have replaced their appliances even with no subsidy whatsoever. Excluding the "savings" from these replacements makes the program much less cost-effective. The program cost per kilowatt hour increases from $.30 to $.45, and the program cost per ton of carbon dioxide increases from $512 to $756.

These measures of cost-effectiveness are high compared to most available estimates in the literature. For example, Arimura et al. (2012) find that U.S. utility-administered energy efficiency programs yield savings at an average cost of 5 cents per kilowatt hour. After accounting for free riders, our estimate is 9 times larger. In part, this large difference reflects the fact that free riders are expensive; leading in this case to substantial increases in expenditures without corresponding decreases in electricity consumption.

Panel B reports cost-effectiveness under two alternative program designs. When we restrict the maximum subsidy amount to $110 (compared to $170), this decreases the total number of participants, but actually makes the program more cost-effective because the average subsidy level is lower. The last row considers an even more modest program which caps the level of the subsidy at only $30. Under this program design the total number of induced replacements is considerably smaller and 69% of participants are free riders. Nonetheless, this design is even more cost-effective than the $110 cap because of the even lower average subsidy level.

6.2 Implications for Welfare

Cost-effectiveness measures are a valuable first step in evaluating energy-efficiency programs, but by themselves are insufficient for determining whether or not a program is welfare-improving. In this section we combine our cost-effectiveness measures with estimates of the MCPF, the private cost of appliance replacement, and the social value of reduced externalities. Our objective is to make a preliminary assessment about the overall welfare implications of the program and to assess how this is affected by free riders. These calculation should be interpreted with caution because they require us to make several strong assumptions.

After accounting for free riders, we are finding that it costs \$245 in direct program cost per appliance replacement. If the MCPF is, for example, 1.34, then the economic cost of raising these funds would be \$83 per replacement.[12] Estimates of the MCPF vary widely, both because studies use different methodologies and because revenue sources vary in the amount of distortion that they impose. But this back-of-the-envelope calculation suggests that the economic costs of these distortions could be substantial.

Another important component is the private cost of replacement. As we emphasized in Section 2, increasing adoption yields both costs and benefits, and what matters for welfare is the difference between the two. This was represented in in Figure 1 as triangle C. The subsidy changes who bears this cost, but does not make it disappear. The average subsidy amount received by program participants was \$149. Assuming private marginal benefits are linear, the average private cost of replacement (net of benefits) is half of this, about \$75. Of course, private benefits need not be linear, but this back-of-the-envelope calculation shows that this component of the economic costs of adoption can indeed be substantial.

Finally, it is important to incorporate the indirect costs from designing, advertising, and administering the program. Data is not available on these costs, but we can make a rough estimate using evidence from the United States. Joskow and Marron (1992) examines detailed energy-efficiency program data from seven U.S. utilities, finding that administrative costs as a fraction of direct costs averaged 36%.[13] This ratio implies, given the \$245 per replacement in direct costs and using our baseline MCPF (1.34), that indirect costs yield an additional economic cost of \$118 per replacement. Of course, indirect costs could well be lower than 36%, particularly given the fact that this is an unusually large program and, thus, can benefit from economies-of-scale in administration. But this rough calculation highlights the potential for indirect costs to be large.

[12] This estimate comes from Bovenberg and Goulder (2002) who report estimates from the United States ranging from 1.11 to 1.56. That is, raising each dollar of revenue imposes a distortion ranging from \$.11 to \$.56. The midpoint of this range is \$.34. We are not aware of any estimates from Mexico.

[13] Joskow and Marron (1992) examined these seven utilities because they were the only ones for which detailed data were available. The ratio of administrative cost to direct measured costs ranges dramatically across utilities from 7% to 70%. Interestingly, the two utilities with the most detailed cost data had the highest overall ratios, while the authors argue that some of the lower ratios may reflect incomplete records of administrative costs for those utilities.

These costs must be compared to the benefits in the form of reduced externalities. In Mexico 0.59 tons of carbon dioxide are emitted per megawatt hour of electricity generation[14], so the estimates in Davis et al. (2012) imply that appliance replacement reduces lifetime carbon dioxide emissions by an average of 0.32 tons. Greenstone et al. (2011) find a central value of the social cost of carbon dioxide emissions of \$21 per ton, which would imply that the social benefits from carbon dioxide abatement are about \$7 per replacement. At the 95th percentile estimate in Greenstone et al. (2011) the social cost is \$65 per ton per ton of carbon dioxide, which would imply \$21 in benefits per replacement.

It is also important to incorporate external benefits from reduced emissions of criteria air pollutants like sulfur dioxide. We are not aware of any estimates of these benefits for Mexico, but estimates from the United States provide some sense of the potential magnitudes. Muller et al. (2011) estimates the external damages from sulfur dioxide, nitrogen oxide, and particulates for different forms of U.S. power generation. Coal-fired power plants are the most damaging (2.8 cents per kilowatt hour), while oil (2.0 cents) and, in particular, natural gas (0.2 cents) are less damaging. Mexican plants emit higher levels of criteria pollutants than U.S. plants. According to CEC (2011), Mexican plants emit per kilowatt hour on average 2.4 times as much sulfur dioxide, 1.7 times as much nitrogen oxide, and 2.2 times as much particulates (PM_{10}).[15] Taking this into account, and using the mix of electricity generation in Mexico[16], the estimates from Muller et al. (2011) imply that the implied lifetime external benefit from reduced emissions of criteria pollutants is \$16 per replacement.

Thus overall it appears that the total benefits per replacement are almost certainly less than the costs. These calculations require many strong assumptions. Perhaps most importantly, all of these calculations continue to assume that our estimated elasticities are representative of all households, including households not close to the eligibility thresholds. Nevertheless, the fact that the gap between benefits and costs is so large under these baseline assumptions implies that one would have to make considerably

[14] According to Mexico Secretaría de Energía, "Balance Nacional de Energía", 2010, pages 53-54, electricity generation in Mexico in 2009 produced 113.4 million tons of carbon dioxide. According to Mexico Instituto Nacional de Estadística y Geografía, "El Sector Energético en Mexico 2009", Table 2.4.1, total electricity generation in 2009 was 193 million megawatt hours. Thus each megawatt hour of electricity generation implies an average of (113.4) / (193) = 0.59 tons of carbon dioxide emissions. The equivalent emissions factor for the United States is 0.58. From U.S. Department of Energy, Energy Information Administration, Annual Energy Review 2011, released September 2012, Table 11.5a, "Emissions from Energy Consumption for Electricity Generation," total carbon dioxide emissions in 2010 for electricity generation were 2.39 billion metric tons. From Table 8.1 "Electricity Overview," total electricity generation in 2010 was 4.13 billion megawatt hours.

[15] CEC (2011) reports plant-level emissions from all 3000+ major power plants in North American. We calculated average emissions factors for Mexico and the United States using total electricity generation and emissions levels from their Tables 2.2 and Table 2.3. The most damaging criteria pollutant is sulfur dioxide, so in these calculations we scaled damages by 2.4 to reflect the higher level of emissions from Mexican plants.

[16] According to SENER, "Prospectiva del Sector Eléctrico, 2010-2025", Figure 37, net generation in Mexico in 2010 was 52% natural gas, 17% oil, 12% coal, and 11% hydroelectric.

different assumptions in order to conclude that the program is welfare improving. For example, a much lower MCPF, in itself, would not be enough to reduce the costs below the estimated benefits per replacement. These somewhat disappointing results come in large part from the relatively small amount of electricity savings per replacement. When combined with the fact that it appears that a large fraction of participants are free riders; it becomes difficult to make an economic argument for the program.

6.3 Internalities

This discussion of the welfare implications of the program assumes that households know their private benefits (and costs) from replacing. This was the starting point in the conceptual framework outlined in Section 2, and is implicit above when we estimate that private costs (net of private benefits) are about $75 per replacement. If, however, there is some other market failure that leads potential adopters to systematically undervalue these private benefits, or overvalue the costs of adoption, then this gap between private costs and benefits could be much smaller, or even negative.

The question of what energy policy should look when there are these so-called "internalities" is considered in detail by Allcott et al. (2012). Internalities provide an additional potential rationale for subsidizing technology adoption. The idea is that, absent any public intervention, agents are making mistakes – adopting fewer energy-efficient durable goods than is privately optimal. It is possible in this case for a subsidy to lead agents to make decisions which they realize *ex post* were better than they expected.

The existing empirical evidence on internalities is mixed. In a recent survey, Allcott and Greenstone (2012) conclude that, "when one tallies up the available evidence from different contexts, it is difficult to substantiate claims of a pervasive energy efficiency gap." Some of the most credible recent empirical evidence comes from automobile purchase decisions. Busse et al. (2013) calculate implied discount rates by estimating the effect of gasoline prices on the prices of cars with different fuel economies, finding little evidence of consumer myopia. On the other hand, Allcott and Wozny (2012) find that consumers value future gasoline expenditures at only $.76 per dollar in vehicle purchase price. It seems plausible, moreover, that decisions about electricity-using durable goods like air conditioners and refrigerators might be particularly susceptible to systematic undervaluing of energy costs, given that operating costs are less salient than, for example, gasoline expenditures.

Even if there are internalities, however, whether or not a subsidy is corrective depends on the exact form the undervaluation takes. Allcott et al. (2012) show that if all households undervalue future energy costs by exactly the same amount, then a subsidy can be implemented that exactly induces households into making the decisions that are correct *ex post*. They also point out, however, that if households undervalue by different amounts, or if some agents follow "rules-of-thumb" like always buying the cheapest alternative, then subsidies are much less successful, in many cases actually distorting decisions away from privately optimal choices.

Moreover, it is worth emphasizing there are important alternative policy mechanisms which may be more appropriate than subsidies for addressing particular forms of internalities. For example, if consumers are simply not aware of particular alternatives, then information provision may be the best approach. Energy-efficiency standards have also long been a key component of policy in this area, and are appropriate when there are alternatives in the market that few agents would adopt if they were perfectly informed about all the relevant costs and benefits.

7 Conclusion

It is hard to provide incentives for socially-valuable behavior without substantial transfers to those who would have done these behaviors anyway. Subsidies for energy efficiency are a key example, both because the potential external benefits are large and because first-best policies seem, for the moment, to be impossible politically. Credible empirical estimates of free-riding are critical, because if a large enough fraction of participants are free riders then a program will not be welfare improving.

Economists have long argued that many recipients of energy-efficiency subsidies may be free riders, but endogeneity concerns and other empirical challenges have made this difficult to demonstrate empirically. Our RD analysis avoids many of the problems in previous studies by focusing on behavior within narrow windows around eligibility thresholds. Although RD is a natural approach to causal inference in this context, we are not aware of any previous RD analyses of free riders. Our estimating equations, tests of strategic behavior, and cost-benefit analysis could be widely applied elsewhere. Although the exact eligibility requirements vary across programs, it is typical to see discontinuous thresholds of the type observed here.

The results are striking. Across thresholds, we find that most households would have participated even for much lower subsidy amounts. Under our preferred assumptions, the estimates imply that about one-third of participants would have replaced their appliances even with no subsidy whatsoever. These free riders add substantial cost to the program because raising the funds necessary for these transfers distorts labor and other markets. Even for modest values of the marginal cost of public funds, we show that the economic costs of the program almost certainly exceeded the benefits from reduced externalities.

References

Allcott, H. and M. Greenstone, "Is There an Energy Efficiency Gap?," *Journal of Economic Perspectives*, September 2012, *26* (1), 3–28.

_ **and N. Wozny**, "Gasoline Prices, Fuel Economy, and the Energy Paradox," 2012.

_ **and S. Mullainathan**, "External Validity and Partner Selection Bias," *National Bureau of Economic Research Working Paper*, 2012.

_ **, _ , and D. Taubinsky**, "Energy Policy with Externalities and Internalities," *National Bureau of Economic Research Working Paper*, 2012.

Arimura, T., S. Li, R. Newell, and K. Palmer, "Cost-Effectiveness of Electricity Energy Efficiency Programs," *Energy Journal*, 2012, *33* (2), 63–99.

Auffhammer, M., C. Blumstein, and M. Fowlie, "Demand-Side Management and Energy Efficiency Revisited," *Energy Journal*, 2008, *29* (3), 91–104.

Barbose, G., C. Goldman, I. Hoffman, and M. Billingsley, "The Future of Utility Customer-Funded Energy Efficiency Programs in the United States: Projected Spending and Savings to 2025," *Lawrence Berkeley Laboratory Working Paper, LBNL-5803E*, 2013.

Battistin, Erich, Agar Brugiavini, Enrico Rettore, and Guglielmo Weber, "The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach," *American Economic Review*, September 2009, *99* (5), 2209–26.

Borenstein, S., "The Redistributional Impact of Nonlinear Electricity Pricing," *American Economic Journal: Economic Policy*, September 2012, *4* (3), 56–90.

_ **and L. Davis**, "The Equity and Efficiency of Two-Part Tariffs in U.S. Natural Gas Markets," *Journal of Law and Economics*, 2012, *55* (1), pp. 75–128.

Bovenberg, A. and L. Goulder, "Optimal Environmental Taxation in the Presence of Other Taxes: General-Equilibrium Analyses," *American Economic Review*, 1996, *86* (4), 985–1000.

_ **and** _ , "Environmental Taxation and Regulation," *Handbook of Public Economics*, 2002, *3*, 1471–1545.

Busse, M., C. Knittel, and F. Zettelmeyer, "Are Consumers Myopic? Evidence from New and Used Car Purchases," *American Economic Review*, 2013, *103*, 220–256.

Chandra, A., S. Gulati, and M. Kandlikar, "Green Drivers or Free Riders? An Analysis of Tax Rebates for Hybrid Vehicles," *Journal of Environmental Economics and Management*, 2010, *60* (2), 78–93.

Davis, L., A. Fuchs, and P. Gertler, "Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico," *EI @ Haas Working Paper*, 2012.

Gallagher, K. and E. Muehlegger, "Giving Green to Get Green? Incentives and Consumer Adoption of Hybrid Vehicle Technology," *Journal of Environmental Economics and Management*, 2011, *61* (1), 1–15.

Greenstone, M., E. Kopits, and A. Wolverton, "Estimating the Social Cost of Carbon for Use in U.S. Federal Rulemakings: A Summary and Interpretation," *NBER Working Paper 16913*, 2011.

Gruber, Jonathan, *Public Finance and Public Policy,* 2013.

Hahn, J., P. Todd, and W. Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69* (1), 201–209.

Ito, K., "Does Conservation Targeting Work? Evidence from a California Statewide Electricity Rebate Program," *Stanford University Working Paper*, 2012.

Joskow, P. and D. Marron, "What Does a Negawatt Really Cost? Evidence from Utility Conservation Programs," *Energy Journal*, 1992, pp. 41–74.

Lee, D. and T. Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48*, 281–355.

Loughran, D. and J. Kulick, "Demand-Side Management and Energy Efficiency in the United States," *Energy Journal*, 2004, pp. 19–43.

McCrary, J., "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 2008, *142* (2), 698 – 714.

Mian, A. and A. Sufi, "The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program," *Quarterly Journal of Economics*, 2012, *127* (3), 1107–1142.

Muller, N., R. Mendelsohn, and W. Nordhaus, "Environmental Accounting for Pollution in the United States Economy," *American Economic Review*, 2011, *101* (5), 1649–1675.

Parry, I., "A Second-best Analysis of Environmental Subsidies," *International Tax and Public Finance*, 1998, *5* (2), 153–170.

Sallee, J., "The Surprising Incidence of Tax Credits for the Toyota Prius," *American Economic Journal: Economic Policy*, 2011, *3* (2), 189–219.

_ **and J. Slemrod**, "Car Notches: Strategic Automaker Responses to Fuel Economy Policy," *Journal of Public Economics*, 2012, *96*, 981–999.

Wolfram., C, O. Shelef, and P. Gertler, "How Will Energy Demand Develop in the Developing World?," *Journal of Economic Perspectives*, 2012, *26*, 119–138.

Figure 1: The Market for an Energy-Efficient Technology

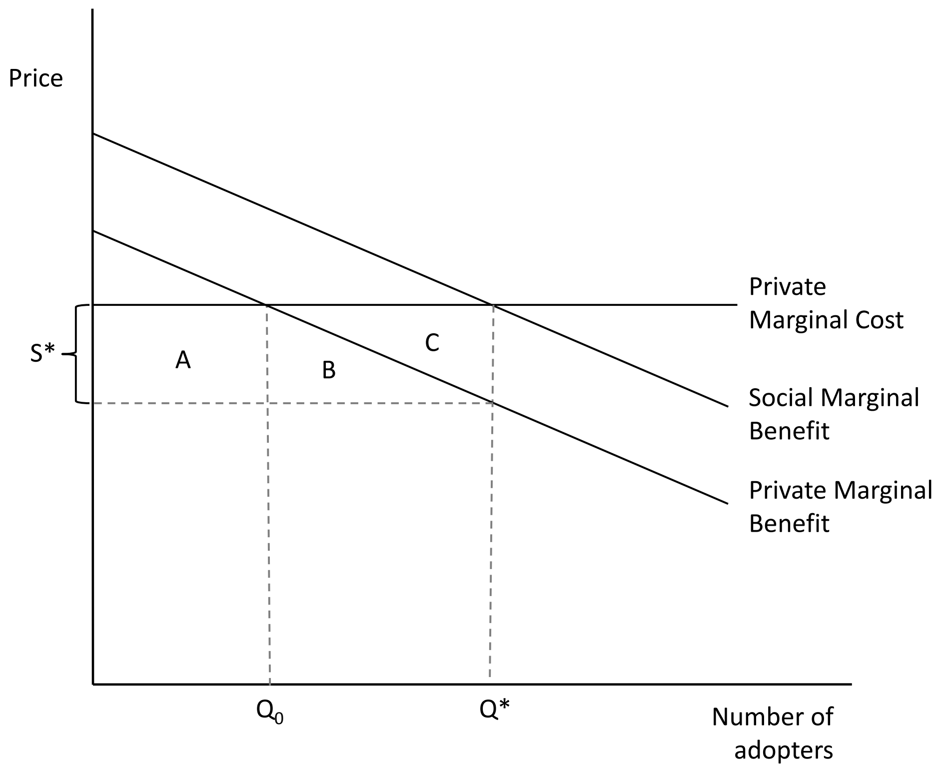
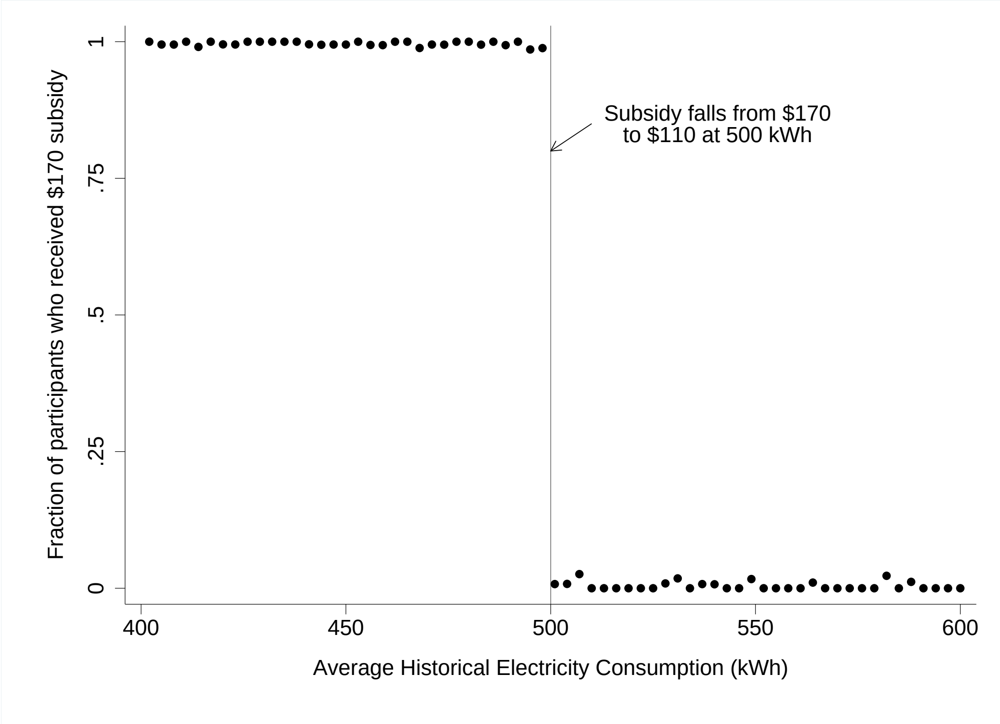
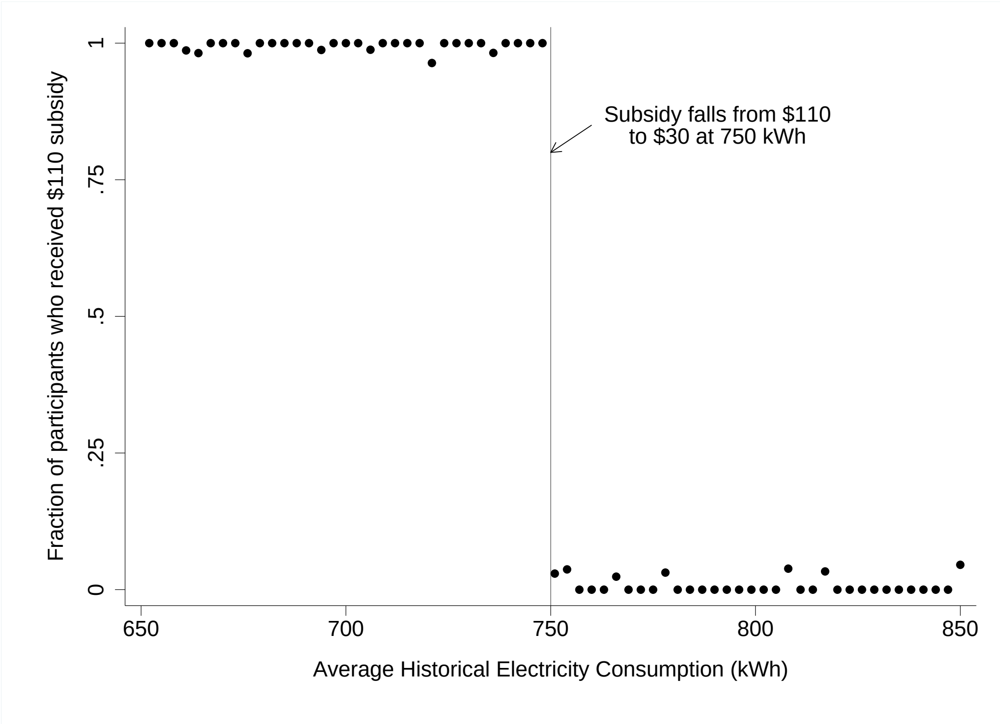


Figure 2: The Discontinuity

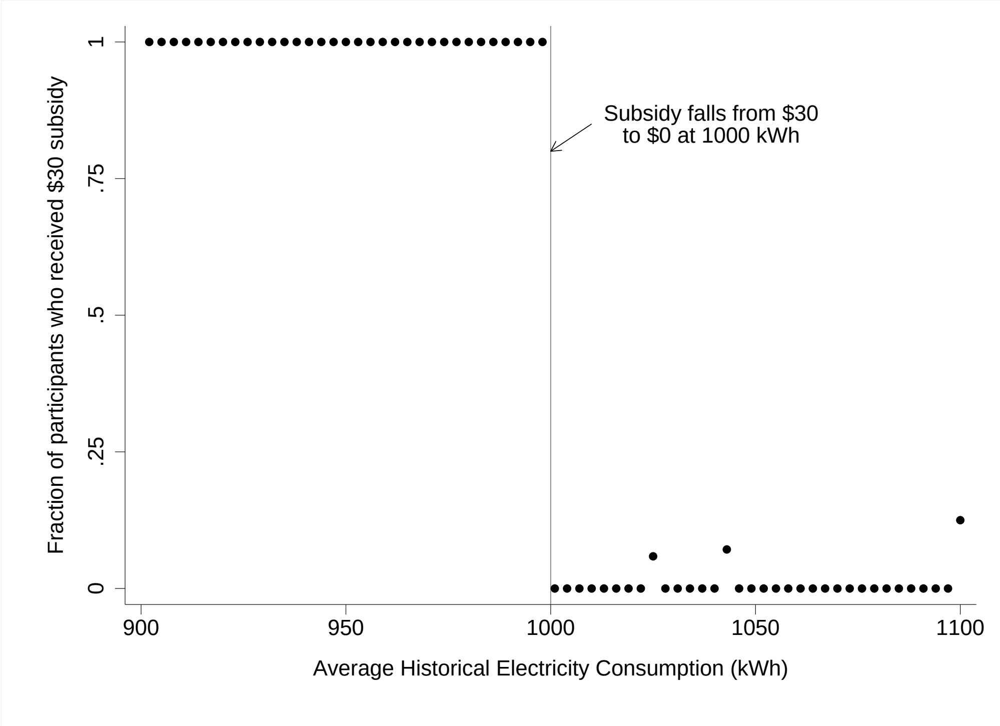
A. Air Conditioners, 500 kWh Threshold



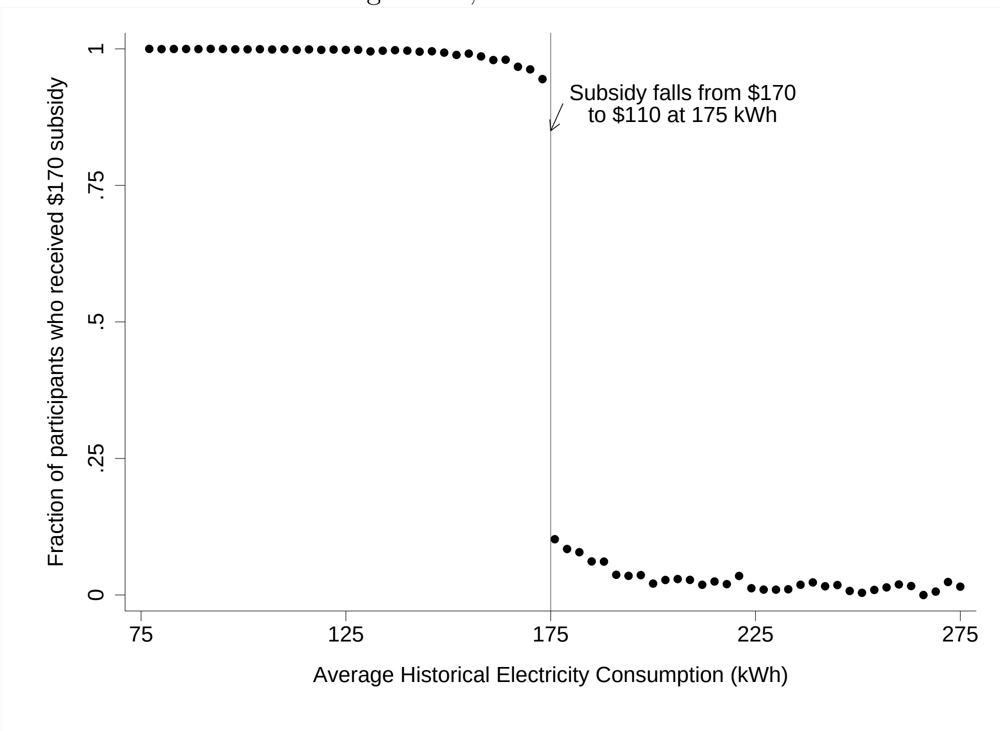
B. Air Conditioners, 750 kWh Threshold



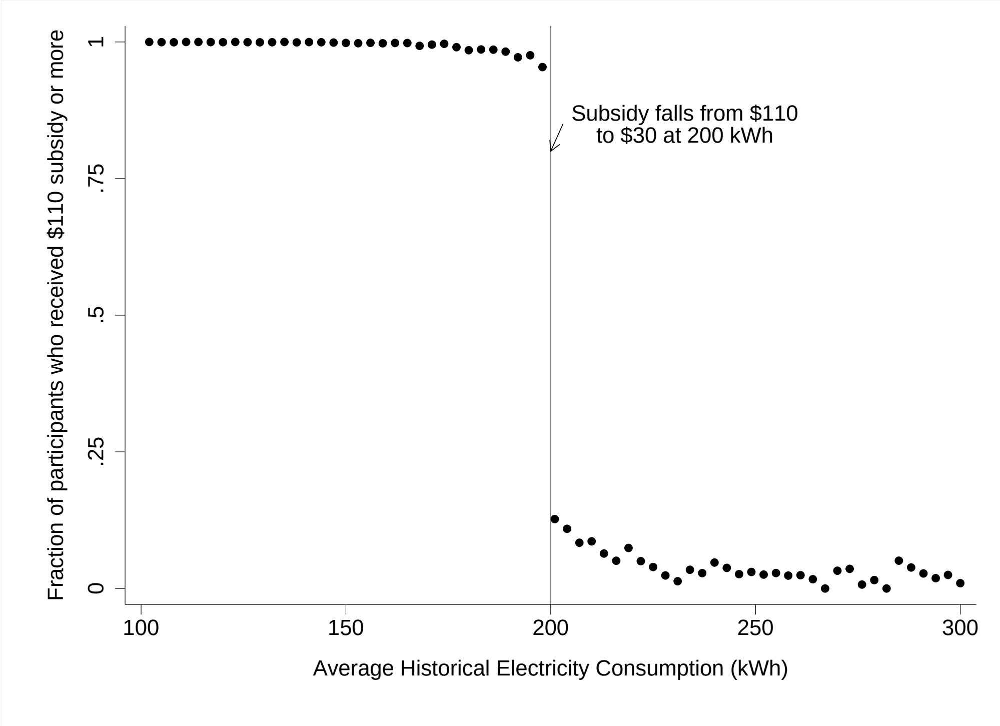
C. Air Conditioners, 1000 kWh Threshold



D. Refrigerators, 175 kWh Threshold



E. Refrigerators, 200 kWh Threshold



F. Refrigerators, 250 kWh Threshold

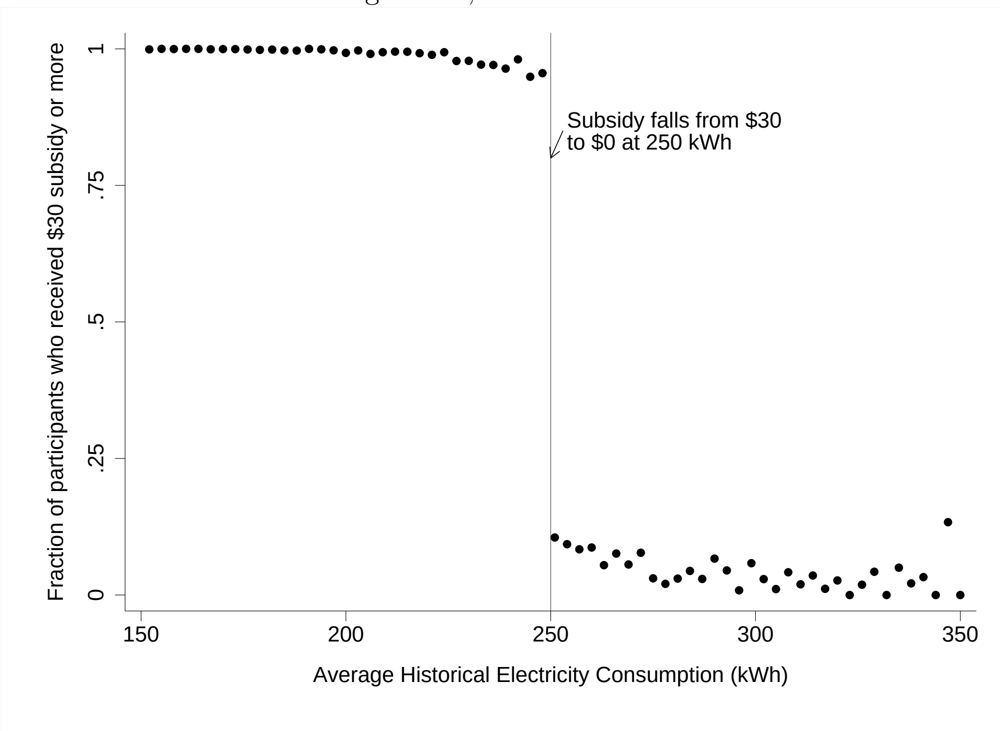
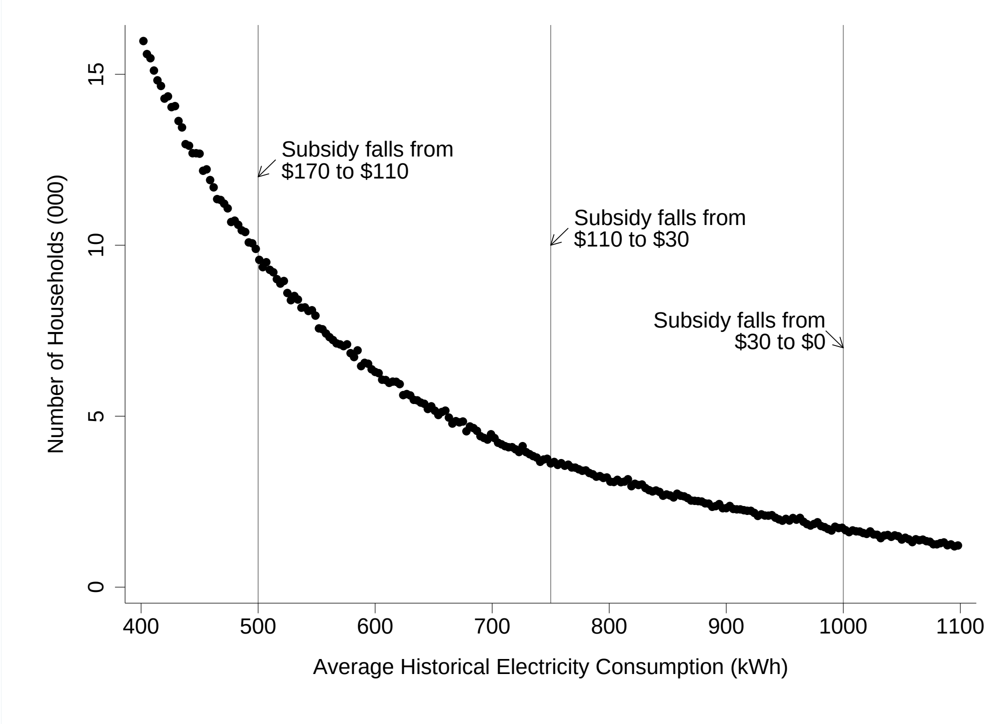


Figure 3: Smoothness of Running Variable Across Subsidy Thresholds

A. Air Conditioners



B. Refrigerators

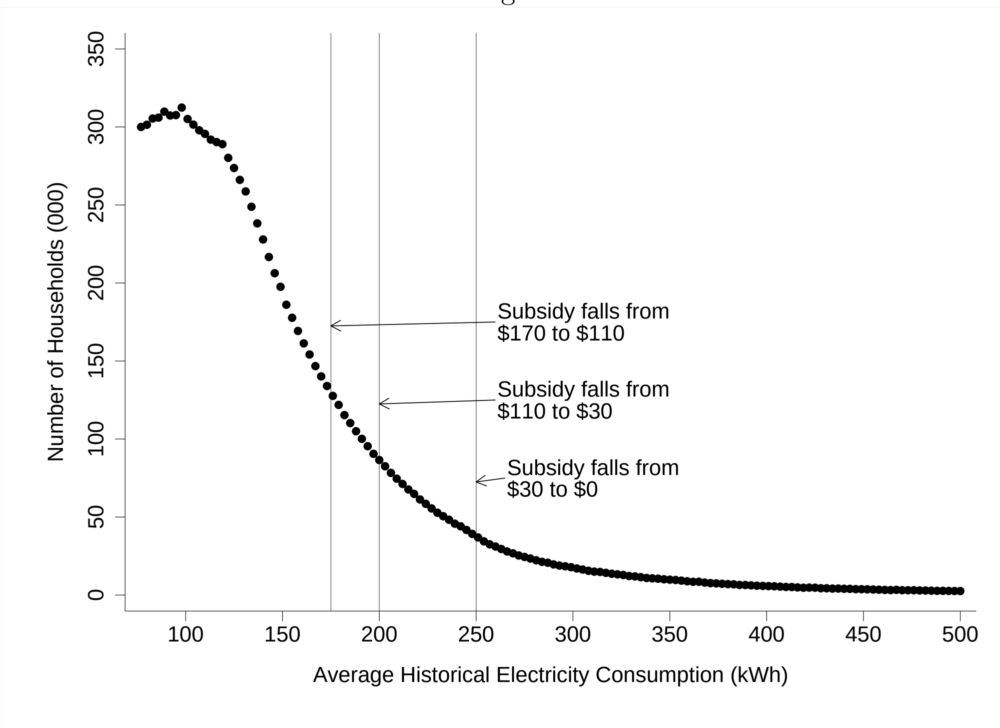
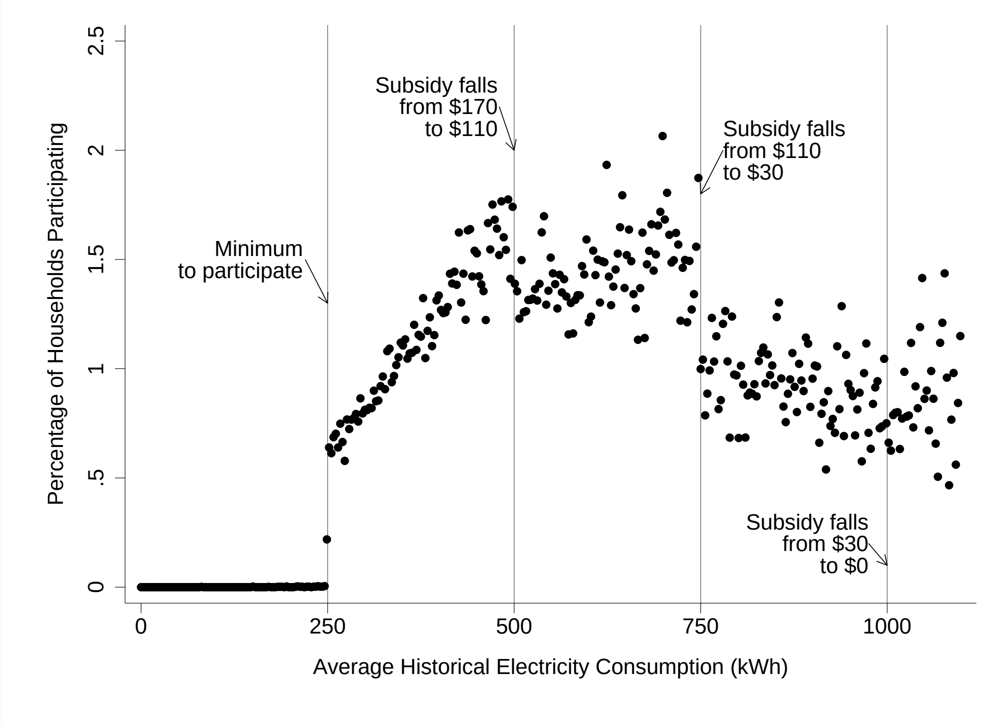


Figure 4: Program Participation

A. Air Conditioners



B. Refrigerators

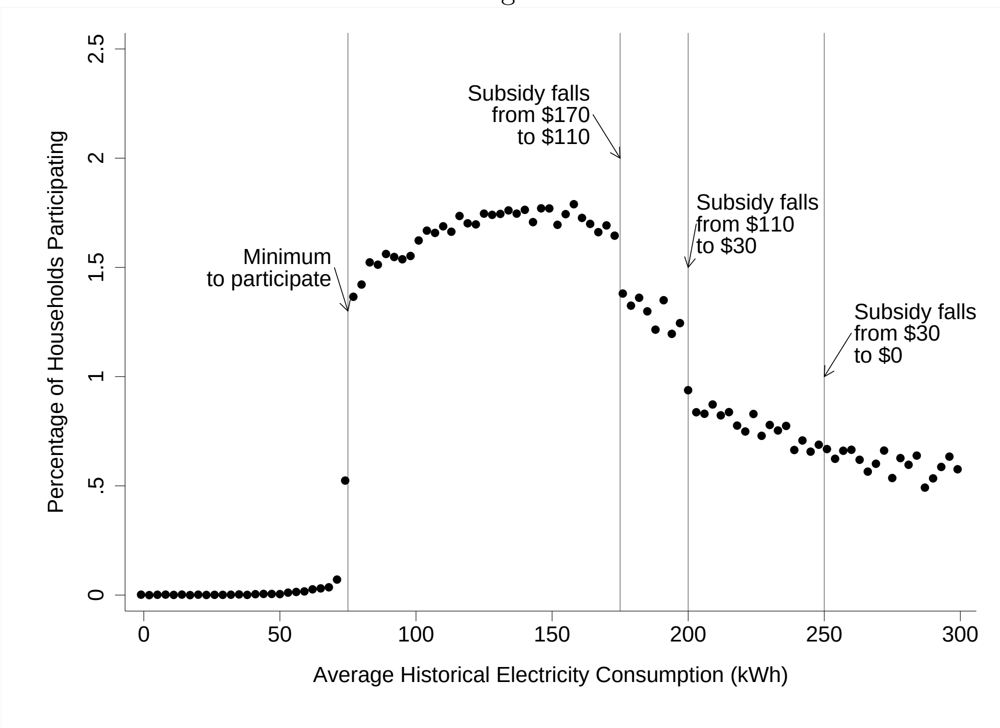
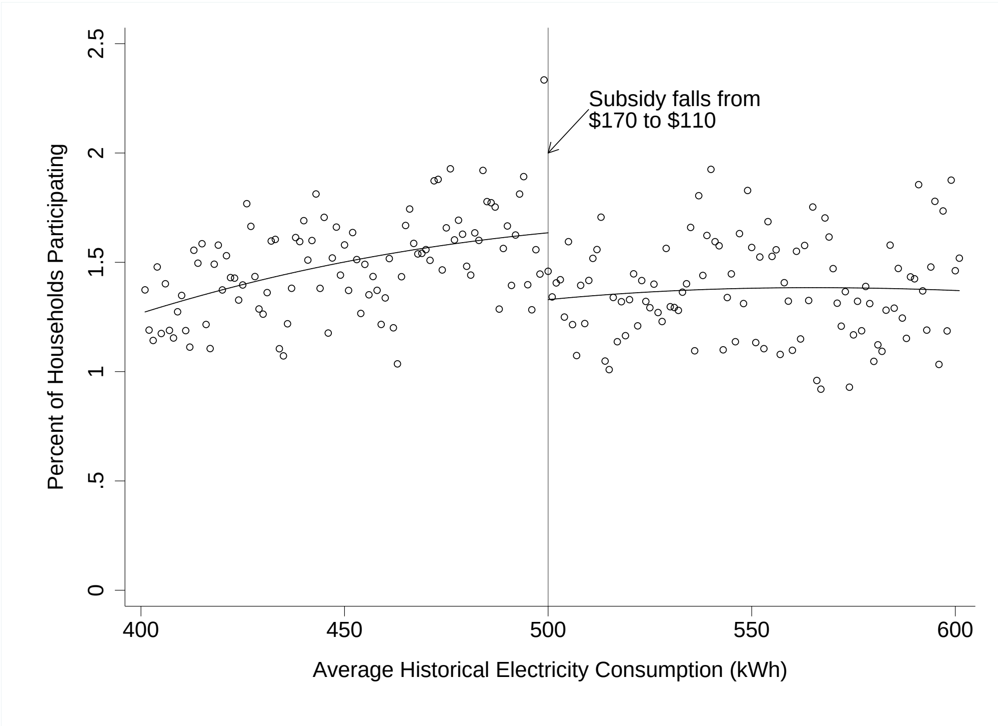
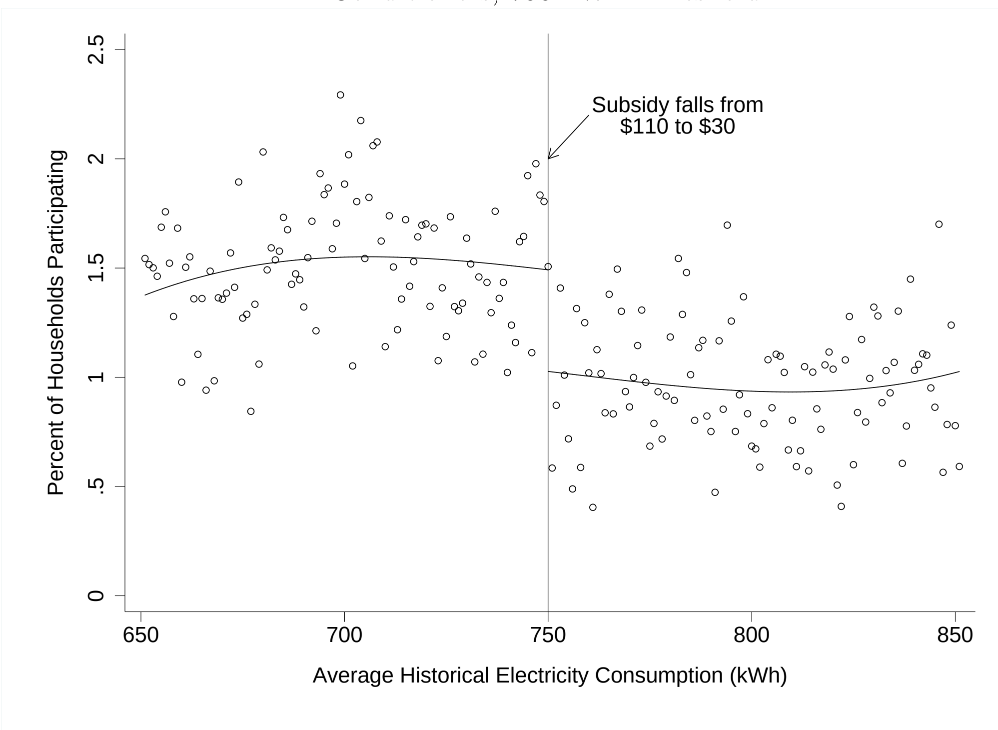


Figure 5: The Effect of Subsidy Size on Program Participation, RD Estimates

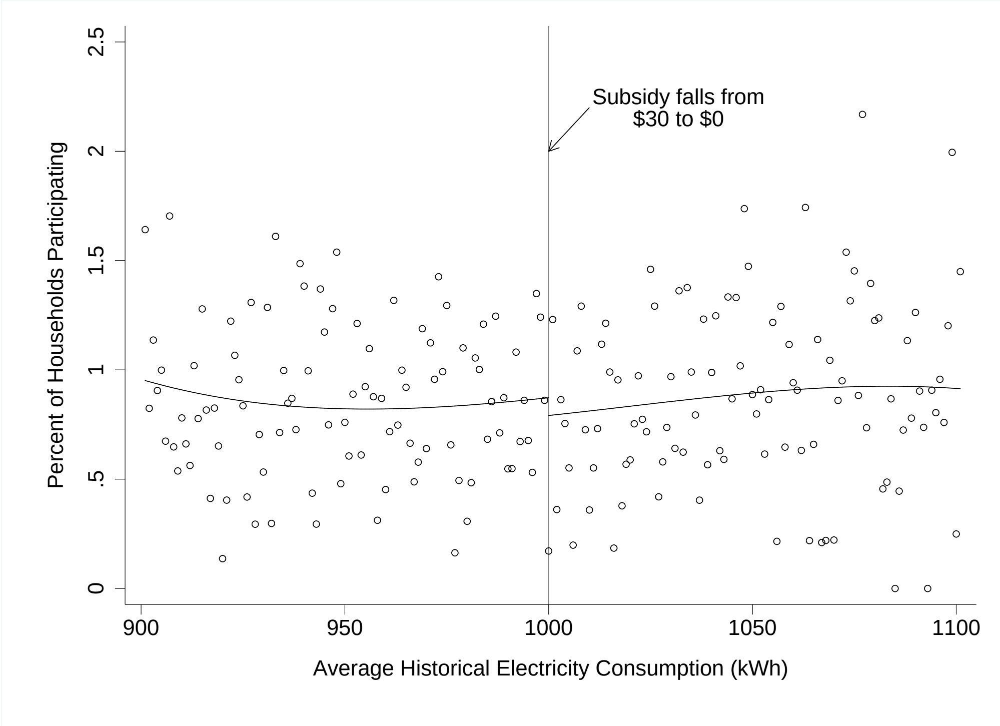
A. Air Conditioners, 500 kWh Threshold



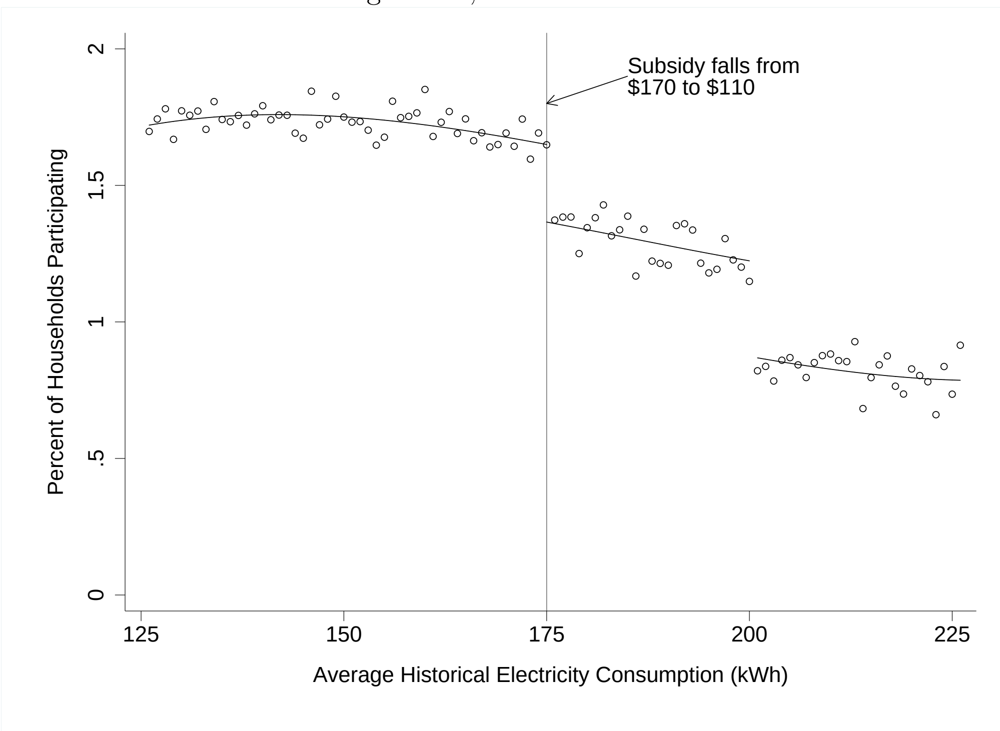
B. Air Conditioners, 750 kWh Threshold



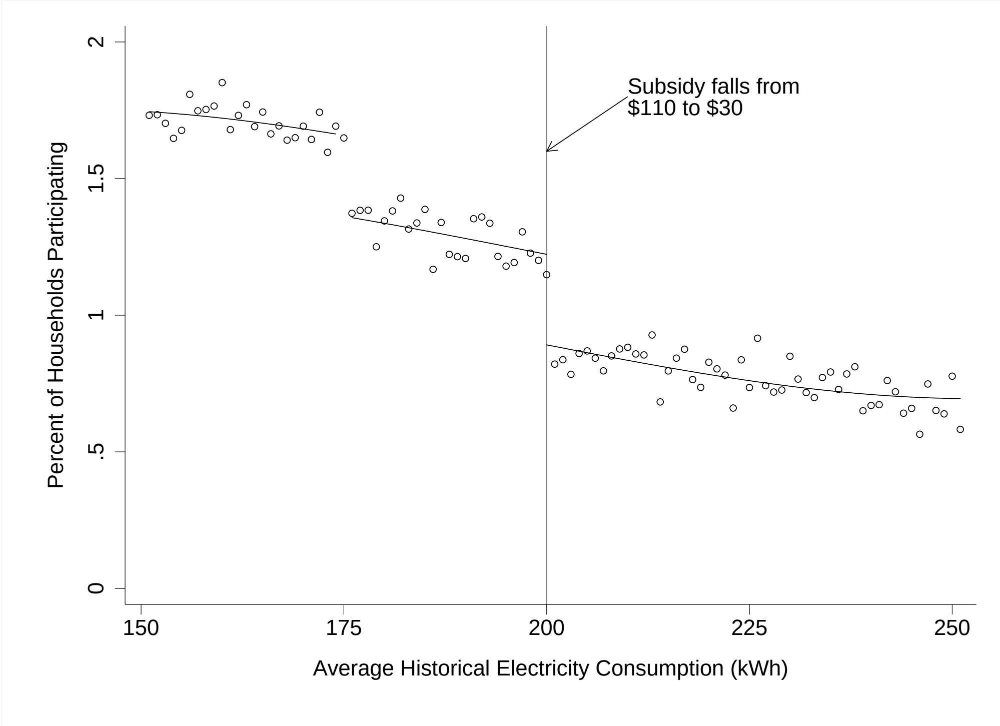
C. Air Conditioners, 1000 kWh Threshold



D. Refrigerators, 175 kWh Threshold



E. Refrigerators, 200 kWh Threshold



F. Refrigerators, 250 kWh Threshold

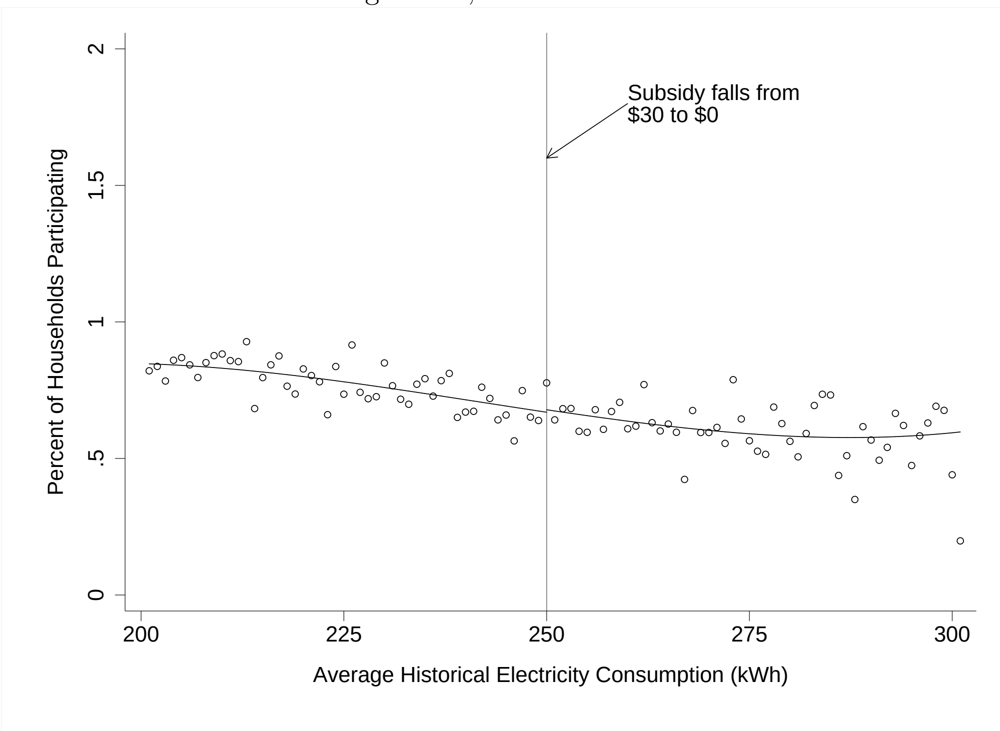


Table 1: Subsidy Amounts

Consumption Categories	Direct Cash Payments	Maximum Credit Amount
	Panel A. Air Conditioners	
< 251	Ineligible for cash subsidy	Ineligible for subsidized credit
251 - 500	2,200 Pesos ($170 dollars)	3,400 Pesos ($270 dollars)
501 - 750	1,400 Pesos ($110 dollars)	4,200 Pesos ($330 dollars)
751 - 1,000	400 Pesos ($30 dollars)	5,200 Pesos ($410 dollars)
1,000+	Ineligible for cash subsidy	8,700 Pesos ($690 dollars)
	Panel B. Refrigerators	
< 76	Ineligible for cash subsidy	Ineligible for subsidized credit
76 - 175	2,200 Pesos ($170 dollars)	3,400 Pesos ($270 dollars)
176 - 200	1,400 Pesos ($110 dollars)	4,200 Pesos ($330 dollars)
201-250	400 Pesos ($30 dollars)	5,200 Pesos ($410 dollars)
250+	Ineligible for cash subsidy	8,700 Pesos ($690 dollars)

Notes: This table describes the direct cash payments and maximum credit amounts available to households with different levels of average historical electricity consumption (in kilowatt hours per month). For further details, see Appendix 1. Dollar amounts are reported in U.S. 2010 dollars using the annual average exchange rate for 2010 (12.645 Pesos per dollar). For expositional clarity, we rounded all dollar amounts to the nearest $10.

Table 2: Number of Participants Close to Thresholds

Subsidy Increase	Number of Households Within Preferred Bandwidth	Number of Participants Within Preferred Bandwidth	Number of Participants Incorrectly Predicted to Receive Low Subsidy	Number of Participants Incorrectly Predicted to Receive High Subsidy	Percent Correctly Predicted
(1)	(2)	(3)	(4)	(5)	(6)
		Panel A. Air Conditioners			
$110 to $170	683,253	9,716	16	24	99.6
$30 to $110	252,358	3,217	7	10	99.5
$0 to $30	114,947	988	3	4	99.3
		Panel B. Refrigerators			
$110 to $170	4,806,857	73,910	853	702	97.9
$30 to $110	3,225,615	42,237	490	309	98.1
$0 to $30	1,421,057	10,401	163	126	97.2

Note: This table describes the households near each of the eligibility thresholds used in the RD analysis. In column (2), we report the number of households with average historical electricity consumption above or below the threshold by less than our preferred bandwidths of 100 kilowatt hours for air conditioners and 50 kilowatt hours for refrigerators. Column (3) gives the number of program participants in the same range. We then report in columns (4) and (5) the number of participants for whom our reconstruction of average historical electricity consumption incorrectly predicts the subsidy amount received. Column (6) reports the percentage of participants where we correctly predict the subsidy amount.

Table 3: Testing for Strategic Delay

(1)	Percent who would have been eligible for a smaller subsidy two months earlier (2)	Percent who would have been eligible for the same or larger subsidy two months earlier (3)
Panel A. 175 kWh threshold		
Participants	25.4%	74.6%
Non-participants	23.8%	76.2%
p-value, χ^2 test	0.07	
Panel B. 200 kWh threshold		
Participants	26.6%	73.4%
Non-participants	26.1%	73.9%
p-value, χ^2 test	0.71	

Note: For households eligible for the larger subsidies at time of replacement, this table reports the subsidies they would have received if they had purchased their appliance one billing cycle earlier. We include all households within 15 kWh below the threshold. "Participants" includes 2,166 households at the 175 kWh threshold and 1,015 households at the 200 kWh threshold who replaced their refrigerator through the program between February 20 and March 11, 2011. "Non-participants" includes 727,850 households at the 175 kWh threshold and 490,130 households at the 200 kWh threshold who did not participate in the program. For non-participants, we take "time of replacement" as March 1, 2011. Column (2) reports the percentage of households who would have been eligible for a smaller subsidy 60 days before the time of replacement. Column (3) reports the percentage of households who would have been eligible for the same or a larger subsidy 60 days before the time of replacement. The indicated p-values come from a χ^2 test of homogeneity.

Table 4: Sharp RD Estimates of the Effect of Increased Subsidies on Participation

(1) Subsidy Increase	(2) Percent of Households Participating: At Lower Subsidy Amount	(3) Percent of Households Participating: At Higher Subsidy Amount	(4) Percent Change in Participation at the Threshold	(5) Implied Price Elasticity	(6) Cost per Additional Participant
			Panel A. Air Conditioners		
\$110 to \$170	1.34	1.62	20.8	1.09	\$459
	(0.22)	(0.27)	(8.7)	(0.46)	(122)
\$30 to \$110	1.03	1.48	44.4	2.36	\$290
	(0.22)	(0.30)	(11.5)	(0.61)	(47)
			Panel B. Refrigerators		
\$110 to \$170	1.37	1.64	19.8	1.12	\$472
	(0.11)	(0.13)	(2.6)	(0.15)	(40)
\$30 to \$110	0.90	1.21	35.1	2.00	\$338
	(0.07)	(0.10)	(4.7)	(0.27)	(31)

Notes: This table reports RD estimates of the effect of increased subsidies on program participation from four separate regressions. In each regression, the sample includes all households within the chosen bandwidth. We use a 100 kWh bandwidth for air conditioners, and a 50 kWh bandwidth for refrigerators. All regressions include a cubic polynomial in average historical electricity consumption, normalized to zero at the threshold. Column 2 reports the estimated intercept. Column 3 reports the estimated intercept plus the estimated coefficient on an indicator variable equal to one for households below the eligibility threshold. Column 4 reports the percent change between the previous two columns. Column 5 reports the implied price elasticities evaluated using the net change in the cost of replacement at each threshold. Column 6 reports the implied cost per additional participant, calculated by dividing the amount of increased subsidies paid to each inframarginal participant by the fraction of participants that were induced to participate, and then adding the subsidy payment to the additional participant. See text for details. Standard errors are clustered at the municipality level.

Table 5: Alternative Bandwidths and Specifications

Panel A: Air Conditioners

(1)	(2)	(3)	(4)	(5)
Subsidy Increase	Sharp RD			Fuzzy RD
	125 kWh	100 kWh	75 kWh	100 kWh
\$110 to \$170	20.7	20.8	29.6	21.1
	(8.3)	(8.7)	(9.8)	(8.9)
\$30 to \$110	56.6	44.4	38.0	45.5
	(13.6)	(11.5)	(11.8)	(11.8)

Panel B: Refrigerators

(1)	(2)	(3)	(4)	(5)
Subsidy Increase	Sharp RD			Fuzzy RD
	75 kWh	50 kWh	25 kWh	50 kWh
\$110 to \$170	22.8	19.8	16.5	23.5
	(2.3)	(2.6)	(3.2)	(3.1)
\$30 to \$110	36.1	35.1	38.2	42.1
	(4.6)	(4.7)	(6.3)	(5.7)

Note: This table reports the estimated percent increase in program participation from 16 separate regressions. The percent increase in participation is calculated as in Table 4. Standard errors are indicated in parentheses. All regressions include a cubic polynomial in average historical electricity consumption. Columns (2) through (4) report results from a regression of program participation on an indicator equal to one for households eligible for the higher subsidy, using all households above or below the threshold by less than the indicated bandwidth. Column (5) reports results from scaling the coefficient in Column (3) by the size of the observed change in subsidy amount at the eligibility threshold. See text for details. The specifications in Column (3) are identical to the specifications in Table 4.

Table 6: Accounting for Free Riders in Measures of Cost-Effectiveness

	Total Replacements (1000s)	Total Induced Replacements (1000s)	Percentage Free Riders	Program Cost (Per kWh)	Program Cost Per Ton of Carbon Dioxide
	(1)	(2)	(3)	(4)	(5)
Panel A. Actual Program Design					
Assuming No Free Riders	958	958	0%	\$0.30	\$512
Accounting for Free Riders	958	659	31%	\$0.45	\$756
Panel B. Alternative Program Designs					
Setting Maximum Subsidy \$110	763	465	39%	\$0.35	\$586
Setting Maximum Subsidy \$30	435	137	69%	\$0.21	\$348

Notes: This table describes the cost-effectiveness of the program under four different scenarios. Column (1) is the total number of participants in the program between May 2009 and April 2011. Column (2) is the number of participants who, according to our estimates, would not have replaced their appliance without the program. The difference between column (1) and (2) is the number of free riders. The total number of free riders is the same (298,000) in rows (2), (3), and (4). Columns (4) and (5) report direct program cost per kWh and per ton of carbon dioxide, respectively.

Appendix 1: Program Rules for Calculating Historical Consumption

This appendix describes the program rules that were used in calculating average historical electricity consumption for each household. As described in the text, participating retailers determined subsidy amounts for households using a specially-designed website. In determining the subsidy amount, the website calculated the average historical electricity consumption for each household as of the date they were applying. These calculations were not saved by the website, but we have used our database of electric billing records to recreate these calculations.

First, using billing cycle codes included in household account numbers, we determined as accurately as possible the exact days corresponding to all 300+ million billing cycles in our data. This was important because we needed to determine which bills would have been in the system as of each potential date of application. In Mexico, 93% of residential electricity customers are billed in two-month cycles. Half of these households have their meters read during odd-numbered months (January, March, etc.) and half have their meters read during even-numbered months. An additional 5% of households have their meters read every month. The remaining 2% have billing periods of 3 months or longer. These irregular periods arise for a variety of reasons. For example, some households in extremely rural areas have their meters read less than six times per year.

Program rules differed between air conditioners and refrigerators. For air conditioner replacements, average historical electricity consumption was calculated using all summer bills during the previous calendar year. "Summer" is defined by the Federal Electricity Commission to mean the six months of the year with the highest average temperature. This differs across locations. The four possibilities are February - July, March - August, April - September, and May - October.

For refrigerator replacements, the average was calculated using the most recent 12 months of bills that were available in the online eligibility system. In practice, it seems to have taken about 60 days on average between the day a meter was read and the day that information became available in the system. So, for example, a household whose meter was read in odd-numbered months and who bought a refrigerator in May 2011 would have had their eligibility calculated using bills from April 2010 through March 2011.

For refrigerators the calculation also depended on the household's electricity rate structure. Residential customers in hot parts of the country have electric rates which vary seasonally, while customers in cool parts of the country have rates which are the same all year. For households whose electric rates do not vary seasonally, all of the bills during the most recent 12 months are included when calculating eligibility. For households whose electric rates vary seasonally, only the subset of non-summer bills during that period were used to calculate baseline consumption.

For households participating in the program near the end of a billing cycle, a few days can change average historical consumption considerably depending on whether or not that last cycle was included in the average or not. In reconstructing average historical consumption for refrigerators we did the best we could, but we do not know exactly the date in which each billing cycle was included in the billing system, and thus cannot reconstruct this perfectly.

Appendix 2: Addressing Measurement Error

This section describes how we address measurement error in historical average electricity consumption. This variable is measured with error for a subset of observations, as described in the text.

7.1 Implications of Measurement Error for the RD Design

First, we note that a model of classical measurement error does not fit the pattern of our data. Random measurement errors in each observation would smooth away any discontinuity at the observed threshold. We do not see a completely sharp discontinuity, but there is a clear change in the probability of receiving the higher subsidy at the threshold. This is consistent with a model where only some observations are measured with error. In other words, for some households we include the same bills as were used in the actual calculation of average historical consumption, but for other households we do not.

Battistin et al. (2009) shows that the fuzzy RD estimator consistently estimates the true treatment effect when the running variable is measured with error for only some observations, as long as the measurement error is independent of treatment and outcomes, conditional on the true value of the running variable. To formalize this, let $X_{obs} = ZX^* + (1 - Z)X$, where Z takes the values 0 or 1 and $Pr[Z = 1] = p$. In other words, p is the probability that we observe the true value of historical average electricity consumption for any observation.

Let Y represent the binary outcome variable, program participation. As in Hahn et al. (2001), let $Y^+ = \lim_{x\to 0^+} E[Y|X^* = x]$ and $Y^- = \lim_{x\to 0^-} E[Y|X^* = x]$. With no measurement error, the sharp RD estimator $Y^+ - Y^-$ yields the true treatment effect ω. However, with the type of measurement error described above, the sharp RD estimator is biased downwards. Let $\tilde{Y}^+ = \lim_{x\to 0^+} E[Y|X_{obs} = x]$ and $\tilde{Y}^- = \lim_{x\to 0^-} E[Y|X_{obs} = x]$. Battistin et al. (2009) shows that

$$\tilde{Y}^+ - \tilde{Y}^- = p\Big[\lim_{x\to 0^+} E[Y|X^* = x] - \lim_{x\to 0^-} E[Y|X^* = x]\Big] = p\omega.$$

The sharp RD estimator underestimates the treatment effect ω by a factor p.

Define $\tilde{S}^+$ and $\tilde{S}^-$ analogously for the treatment, receiving a higher subsidy offer. The true change in the probability of receiving the higher subsidy offer at the threshold is 1. But the estimator $\tilde{S}^+ - \tilde{S}^-$ underestimates the change in probability of treatment by a factor p:

$$\tilde{S}^+ - \tilde{S}^- = p\Big[\lim_{x \to 0^+} E[S|X^*] - \lim_{x \to 0^-} E[S|X^*]\Big] = p.$$

The fuzzy RD estimator $\frac{\tilde{Y}^+ - \tilde{Y}^-}{\tilde{S}^+ - \tilde{S}^-}$ consistently estimates the true treatment effect by canceling out the downward bias p.

7.2 Estimation

We estimate $\tilde{S}^+ - \tilde{S}^-$ using the following first stage regression, as explained in the text:

$$1[Larger\ Subsidy]_i = \alpha + g(X_i) + \gamma 1[Below\ Threshold]_i + \epsilon_i.$$

Because we do not observe what subsidy non-participants were (or would have been) offered, the first stage regression includes only the households within the chosen bandwidth who participated in the program. Because uncertainty about bill delivery dates is unlikely to be correlated with program participation, this first stage relationship is likely to be the same for participants as for the full population.

To calculate the fuzzy RD estimates, we divide the coefficient ρ from section 5.1 in the text by the estimated coefficient γ. We calculate standard errors by bootstrapping.