Essays in Environmental and Energy Economics

by

Joshua Aaron Blonz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Severin Borenstein, Co-chair
Professor Maximilian Auffhammer, Co-chair
Professor Lucas Davis
Professor Meredith Fowlie

Spring 2017

Essays in Environmental and Energy Economics

Abstract

Essays in Environmental and Energy Economics

by

Joshua Aaron Blonz

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Severin Borenstein, Co-chair

Professor Maximilian Auffhammer, Co-chair

This dissertation combines research on three topics in applied Energy and Environmental Economics related to the electricity industry. In the first paper, I study the economic welfare impact of an electricity pricing program that increases the price of electricity for small commercial and industrial customers when the cost of generation is high. The second paper explores an energy efficiency retrofit program that provides free upgrades to low-income households in California. Both of these policy interventions were a result of orders from the California Public Utilities Commission, the energy regulator in California. The final paper examines the cost of air quality regulations on employment in the coal mining sector in Appalachia. These three papers study different important aspects of the electricity sector, from upstream regulation of generation to end use pricing and consumption efficiency.

In the first chapter, I study how in electricity markets, the price paid by retail customers during periods of peak demand is far below the cost of supply. This leads to overconsumption during peak periods, requiring the construction of excess generation capacity compared to first-best prices that adjust at short time intervals to reflect changing marginal cost. In this paper, I investigate a second-best policy designed to address this distortion, and compare its effectiveness to the first-best. The policy allows the electricity provider to raise retail price by a set amount (usually 3 to 5 times) during the afternoon hours of a limited number of summer days (usually 9 to 15). Using a quasi-experimental research design and high-frequency electricity consumption data, I test the extent to which small commercial and industrial establishments respond to this temporary increase in retail electricity prices. I find that establishments reduce their peak usage by 13.4% during peak hours. Using a model of capacity investment decisions, these reductions yield \$154 million in welfare benefits, driven largely by

reduced expenditures on power plant construction. I find the current policy provides of the first-best benefits but that, with improvements in targeting just the days with the highest demand, a modified peak pricing program could achieve 80% welfare gains relative to the first-best pricing policy.

In the second chapter, I study energy efficiency retrofits programs, which are increasingly being used to both save on energy bills and as a carbon mitigation strategy. This paper evaluates the California Energy Savings Assistance program, which provides no-cost upgrades to low-income households across the state. I use quasi-experimental variation in program uptake to measure energy savings for a large portion of the treated population in the San Diego Gas & Electric service territory between 2007 and 2012. The results suggest that the overall program is ineffective at delivering energy savings and is not cost-effective. One challenge in implementing efficiency retrofit programs is that each upgrade must be customized to the housing unit on which it is installed. As a consequence, there is a wide range in efficiency upgrade potential across the population of candidate households. To better understand this heterogeneity in measure installation and its potential to drive program outcomes, I use discontinuities in program rules to identify key measure specific savings. This analysis shows that larger upgrades such as refrigerator replacements do provide cost-effective savings when considering the full set of social benefits. Households that do not receive larger upgrades generally see little or no savings. These results suggest that heterogeneity in upgrade potential can drive overall program outcomes when only a small portion of the treated population is eligible for cost-effective efficiency upgrades.

In the third chapter, I study the costs of Title IV of the Clean Air Act. This regulation put a cap on sulfur emissions from electric power plants, which reduced the demand for high-sulfur coal. Using a quasi-experimental research design, I estimate how coal mine employment and production in high-sulfur coal-producing counties were impacted by the regulation by comparing them to neighboring counties that produced low-sulfur coal. I find that coal production dropped by 20% and coal sector employment dropped by 14%. I find no evidence of spillovers to employment or wages in the non-coal sectors of the high-sulfur coal counties. The results suggest that the coal sector employment costs of Title IV of the Clean Air Act are highly concentrated in the coal industry, and that the decline does not detectably impact the overall regional economy.

Contents

List of Figures

List of Tables

Acknowledgments

I am grateful to many people for their help throughout the PhD process. Severin Borenstein has been an enormously important advisor, mentor, role model and teacher for me while at Berkeley. I consider myself lucky to have spent so many years learning from him and he has shaped the way I think about and study economics. Maximillian Auffhammer has been an inspiring advisor and mentor who taught me how to write papers, be an effective teacher, and how to work in an academic environment. I could not have asked for better advisors to guide me through this process.

I am thankful to everyone at the Energy Institute that helped create a special and unique academic environment. The lunchtime conversations about energy policy were pivotal to my education as an economist. I want to thank Lucas Davis, Meredith Fowlie, Catherine Wolfram, Jim Sallee, and Reed Walker for their input and willingness to discuss my research. Many graduate students played a critical role in my education. I want to thank Patrick Baylis, Judson Boomhower, Fiona Burlig, Walter Graff, Sylvan Herskowitz, Louis Preonas , Matt Woerman and Matthew Zaragosa-Watkins for their invaluable input on my research. I also want to thank Paula Pedro, Casey Hennig and Karen Notsund for their help and support while at the Energy Institute.

Finally, I want to thank my family and friends who supported me throughout the PhD process. My parents were incredibly supportive and were always willing to feed me a home cooked meal at a moment's notice. I couldn't have completed this degree without their ever-present support.

Chapter 1

Making the Best of the Second-Best: Welfare Consequences of Time-Varying Electricity Prices

1.1 Introduction

Supplying electricity during periods of peak demand is expensive. Because electricity storage is not cost effective, sufficient generation capacity must exist to satisfy demand at all moments in time. To avoid blackouts, electricity providers regularly invest in power plants that operate only on the few highest demand days of the year. Electricity prices, however, do not reflect the high cost of meeting peak demand. Most retail prices reflect the average cost of providing power and do not vary based on when this power is consumed. As a result, retail electricity customers are undercharged for their electricity at peak times, leading to inefficiently high consumption (Boiteux 1949; Steiner 1957).

In the long run, higher peak consumption necessitates additional generation capacity. In most U.S. electricity markets, capacity investment decisions are made by the regulator through the "resource adequacy" process (P. Joskow and Tirole 2007). The regulator uses past demand levels to determine generation capacity requirements for electricity providers. If retail prices were adjusted to reflect the full cost of generation during peak periods, this would reduce both peak demand and the regulator's capacity requirement. Borenstein and S. Holland (2005) and Borenstein

(2005) estimate the efficiency loss due to flat retail prices to be 5%-10% of wholesale electricity costs.

Inefficient peak demand pricing also occurs in other contexts. Vickrey (1963) and Vickrey (1969) outlines the problem of unpriced traffic congestion, where drivers do not pay the full external costs of using infrastructure at peak hours. Instead, drivers pay for the use of road networks through a flat gasoline tax, which is similar to the average pricing structure used for electricity. In the long run, this can lead to overinvestment in transportation infrastructure, as additional capacity is built to alleviate unpriced congestion instead of providing the greatest social marginal benefit.[1]

For both electricity markets and traffic congestion, the first-best policy is to charge a price that reflects the short-run marginal scarcity value during periods of peak demand. In the case of electricity, this policy is "real-time pricing" (RTP), under which the retail price changes hourly or more frequently. RTP is technologically feasible at low cost for most commercial and industrial customers due to the wide-scale deployment of smart meters. Despite its large potential benefits, however, real-time pricing remains politically infeasible. Because many customers receive large cross-subsidies under existing flat pricing schemes, mandatory real-time pricing would be difficult to implement without politically unpopular transfer payments (Borenstein 2007b).

The inability to implement real-time pricing suggests two important questions. First, how large are the potential benefits of real-time pricing? This depends on the extent to which customers would respond to short-run price changes. If demand is sufficiently price inelastic, then any potential costs of implementing real-time pricing could outweigh the benefits. Borenstein (2005), however, shows that, for most plausible elasticities, the benefits are very likely to outweigh the costs. Second, to what extent can second-best policies achieve the benefits of real-time pricing? This paper addresses the second question by examining a common second-best policy that raises electricity prices on high demand days, and by measuring this program's effectiveness compared to the first-best, real-time pricing policy.

I study the largest peak demand program to date in the U.S., which includes commercial and industrial (C&I) establishments in the Pacific Gas & Electric (PG&E) Northern California service territory. Programs like PG&E's "Peak Pricing" are among the most common time-varying pricing policies in the U.S. The popularity of such programs has grown with the recent deployment of advanced smart meter technology. The peak pricing implementation I study gives PG&E the ability to

[1] A related problem is seen in the provision of public transportation, where fares typically do not vary over time to reflect the marginal costs (Mohring 1972; Parry and Small 2009).

declare up to 15 "event days" per summer, during which the retail electricity price more than triples between 2:00 pm and 6:00 pm. Customers are notified one day before each event day, and they receive a small discount on all other summer consumption in exchange for their participation in the program. My analysis focuses on small commercial and industrial (C&I) establishments because the way in which the program was implemented for these customers created a quite similar, and exogenous, control group to which the treated group could be compared.

To identify the impacts of this peak pricing program, I leverage the rules that governed its rollout. Establishments were placed on peak pricing by default only after they satisfied a set of eligibility criteria. They were then allowed to opt out. I compare establishments that satisfied the eligibility criteria for the first wave of peak pricing to similar establishments that just missed being eligible by not satisfying the eligibility criteria. I provide supporting evidence that assignment to peak pricing is as-good-as-random, using data from before program implementation. I use both a panel fixed effects instrumental variables strategy and a regression discontinuity design to identify program impacts.

Using hourly electricity consumption data, I find that peak pricing reduces electricity consumption for non-coastal establishments by 13.4% on event days, compared to a control group. I estimate that the program will reduce PG&E peak demand by 118 MW among small C&I customers when fully implemented by the summer of 2018, thereby reducing the need for one or more specialized power plants that are constructed with the sole purpose of generating electricity during the highest demand hours of the year. To evaluate the welfare impacts of peak pricing, I model the regulatory resource adequacy process that governs the amount of capacity that is built specifically to meet peak demand. I find that the estimated reduction in peak demand increases welfare on the PG&E grid by $154 million over a 30-year period, due to avoided generation capacity investments.

To put these welfare effects in perspective, I compare my estimated peak demand reductions to a first-best real-time price. Using the empirical estimates of demand response, I calculate that the current program recovers 43% of the first-best welfare gains. I then consider simple adjustments to the policy to better target the highest demand days, and show that substantial welfare gains would likely result from reducing the number of event days and increasing the event day price. This better-targeted peak pricing policy could achieve 80% of first-best welfare gains.

The importance of the design of second-best policies has also been found in other contexts. Sallee and Slemrod (2012) and Ito and Sallee (2014) show that notched levels of fuel economy regulation can lead car makers to strategically manipulate their production decisions for favorable treatment. This behavior can lead to negative welfare outcomes compared to a differently designed, smooth fuel economy regulation.

Ramnath (2013) finds that the design of the Saver's Credit in the U.S. tax code distorts household income reporting behavior in a way that is potentially costly to the program.

This paper makes three distinct contributions to the economics literature. First, no previous academic research has estimated the impacts of peak pricing on the commercial and industrial sector, which is responsible for two-thirds of California and U.S. electricity demand (Energy Information Administration 2016). The program rollout that I study caused more business establishments to move to peak pricing than any similar program in the U.S. Previous empirical work has focused on peak pricing programs in the residential sector (Fowlie, C. Wolfram, et al. 2015; Ito, Ida, and Tanaka 2015; Bollinger and Hartmann 2015; Jessoe and Rapson 2014; Wolak 2010; Wolak 2007).

The paper also contributes to the literature on long-run investment efficiency in electricity markets. This literature typically relies on simulation and stylized models of power plant capacity construction to value the impacts of alternate pricing policies (Borenstein 2012; S. P. Holland and Mansur 2006; Borenstein and S. Holland 2005; Borenstein 2005). I depart from this approach, and instead focus on the mechanisms that actually drive power plant construction in a regulatory setting. Using this technique, I am able to estimate welfare impacts under a more realistic set of assumptions.[2] I am also able to evaluate which program design features specifically drive impacts, allowing me to propose improvements informed by the empirical estimates.

Finally, the paper contributes to the literature on second-best pricing policies under capacity constraints. The existing literature is mainly theoretical in nature, applying a range of assumed parameter values to stylized models. For example, Arnott, de Palma, and Lindsey (1993) use numerical examples to estimate Vickrey's (1969) model of traffic congestion and simulate outcomes under different pricing regimes. Similarly, research on airplane landing congestion relies on stylized numerical examples to analyze optimal pricing (Brueckner 2002; Brueckner 2005; Brueckner 2009). My paper contributes to this literature by estimating the causal effects of a peak pricing program focused on capacity constraints. My empirical estimates enter directly into the welfare calculations, while allowing evaluation of potential improvements to program design.

The rest of the paper is organized as follows: Section 1.2 discusses the electricity industry, related literature, and the peak pricing program in detail. Section 1.3 outlines the data used in the analysis. Section 1.4 describes the empirical strategy

[2] My approach complements the work of Boomhower and L. W. Davis (2016), who use capacity market payments to value the benefits of energy efficiency at peak hours.

and Section 1.5 presents results. Section 1.6 proposes a model for calculating the welfare impacts of peak pricing programs, discusses potential improvements, and benchmarks the outcomes to the first-best, real-time price. Section 1.7 concludes.

1.2 Background

Most electricity in the U.S. is sold to retail customers at a constant flat rate that does not reflect the time-varying marginal cost of producing another kilowatt-hour (kWh). In most cases, the marginal cost consists of two components. The first is the short-run marginal production cost (SRMC), which includes the fuel costs associated with producing an additional kWh. The second is due to a regulatory process in most states that requires an electricity supplier to demonstrate it controls adequate capacity to meet the peak demand it serves. These "resource adequacy" requirements are generally based on previous peak demand quantities. As a result, each additional kWh consumed on the highest demand days of the year increases future capacity requirements, adding significant costs.

The welfare costs of the current system of flat-rate pricing are well studied in the economics literature (Borenstein 2005; Borenstein 2012; Borenstein and S. Holland 2005). An efficient alternative to the current system is to vary the retail price of electricity to reflect the time-varying marginal cost of supply. This could be done by passing through the wholesale electricity market prices to retail customers in real time. Real-time pricing is technically feasible at low cost due to the wide-scale deployment of smart meters over the last decade (P. L. Joskow and C. D. Wolfram 2012).[3] Existing research shows that RTP could provide large, long-run efficiency gains compared to the current flat-rate pricing by reducing total quantity demanded (load) during high demand hours and increasing load when generation costs are low (S. P. Holland and Mansur 2006). By reducing peak demand, RTP reduces the need for costly power plants specifically built for the hottest few days of the year.

Despite the large potential welfare gains from RTP, implementation is politically challenging. Retail electricity prices are set through a regulatory process under political constraints. Some customers would face significantly higher energy bills under real-time pricing, creating a constituency opposed to the new pricing system. Borenstein (2007b) shows that substantial transfers would be required to keep many customers whole when transitioning to a real-time price. Other customers are wary

[3] Smart meter deployment is financially justified because the meters eliminate the need to pay employees to manually check electricity usage every month. As of 2014, the smart meter penetration for C&I customers in California and the rest of the US was 89% and 66% respectively (Energy Information Administration 2014).

of the potential volatility in electricity bills that could result from real-time pricing. Borenstein (2007a) shows that switching to a real-time price could increase the month-to-month bill volatility for commercial and industrial customers by a factor of two to four, but that simple hedging programs offered by the utility could reduce most of the variation.

In the absence of RTP, policymakers have introduced a number of other policies that pass through some portion of time-varying prices to customers without unexpected volatility. Time of Use (TOU) pricing adjusts the price of electricity in a prescribed manner by hour, day and season, but does not pass through high price events. For example, a previously flat retail price of \$.20/kWh could be changed to a TOU rate of \$.25/kWh between noon and 6:00 pm, when demand is generally high, and \$.15/kWh at night. These prices can capture some of the average shape of marginal costs, but they do not adjust when wholesale costs spike on the highest demand days of the year. Borenstein (2005) shows that TOU captures only a small amount of the efficiency gains of RTP.

Peak pricing programs, like the one studied in this paper, are designed to address the costs associated with the highest demand days of the year. To date, however, research on consumer response to peak pricing programs has focused on the residential sector. Existing studies find that households reduce their energy consumption when facing high prices during peak hours, though the estimated response magnitude varies across studies and depending on the use of automation technology. There has been no published research to date on peak pricing in the commercial and industrial sector. Because firms are responsible for twice the electricity usage of the residential sector, this is an important gap.

The existing residential peak pricing studies have been run as utility experiments and pilot programs. Fowlie, C. Wolfram, et al. (2015) partnered with the Sacramento Municipal Utility District in California to study the impacts of opt-in versus opt-out peak pricing programs. They find that households in the opt-out program reduce their electricity usage by 13.9% during peak pricing events. Households that chose to opt-in to peak pricing reduced their usage by 27.3%. Other residential peak pricing research has focused on the importance of information and technology in responding to peak pricing. Jessoe and Rapson (2014) find that providing households with detailed usage data results in substantially larger reductions than just the price alone. Bollinger and Hartmann (2015) investigate how automation technology that adjusts consumption in response to higher prices affects the response to peak pricing. They find that households are more than twice as responsive when they are given automation technology and higher prices along with price information, compared with price information alone. My paper is the first to investigate whether a similar overall response to peak prices is also found among commercial and industrial customers.

1.2.1 PG&E's Peak Pricing Program

The PG&E peak pricing program for small C&I customers raises the price of electricity from the normal price of \$.25/kWh to \$.85/kWh from 2:00 pm to 6:00 pm on 9 to 15 "event days" per year. The program runs between June 1st and October 31st each year. Enrolled establishments receive a discount of \$.01/kWh on all other consumption during the summer to compensate them for participating. PG&E determines when event days are called based on day-ahead weather forecasts.[4] for specific details on this process. Establishments are notified by 2:00 pm the day before an event via e-mail, text message and/or phone call. Establishments are told about Monday event days on the prior Friday.

Establishments are given "bill protection" for the first summer they are enrolled. This protection guarantees that customers do not pay more in their first summer as a consequence of the peak pricing rates. If their total utility bill is higher between June 1 and October 31 on peak pricing than it would have been if they had opted out, the customer is refunded the difference. Establishments were sent a letter by PG&E in November 2015 informing of them of how much money they saved or would have lost during the first year of the program. The letter explained that the bill protection credit would be dispersed on their November 2015 bill, and that they would no longer receive bill protection going forward. In Section 1.5, I discuss the potential impact of bill protection on the estimates of price response.

The enrollment data suggests that customers will remain in the peak pricing program after they no longer have bill protection. In the first summer of peak pricing, 89% of establishments in my sample would have lost money if not for bill protection. The average loss for these establishments was \$104 over the summer of 2015.[5] Despite these losses, only an additional 5.5% of establishments dropped out between their bill protection payment in November 2015 and the most recent data from October 2016. This suggests that, even after the first summer, when establishments no longer have bill protection, they do not choose to leave the peak pricing program.

I study the first wave of enrollments, in which 29% of small C&I accounts were placed on peak pricing and given the ability to opt out at any time using a simple web interface.[6] Only 5.9% of the establishments in the first wave opted out before the first summer. An additional 5.3% of establishments dropped out during the

[4]When the forecasted maximum temperature at a set of five specified weather stations exceeds a given "trigger" temperature, an event day is called. See Appendix Section A.1

[5]2015 was the first year that small C&I establishments were included in peak pricing. The program is designed to be revenue neutral with respect to enrollees, suggesting that the \$.01/kWh subsidy for non-event hours may need to be increased in future years.

[6]See Appendix Figure A.1 for an example of the letter sent to establishments 30 days before the program started, with directions on how to opt out.

first summer of the program. The high number of people remaining in the program reflects both the role of default bias and the impact of bill protection. There is a large economics literature documenting the impact that changing the default can have on choice, including Madrian and Shea (2001), Choi et al. (2004), Abadie and Gay (2006), and Johnson, Bellman, and Lohse (2002), among many others.

1.3 Data

I use confidential data provided by PG&E for this analysis. The data consist of hourly electricity usage data for 19,071 establishments for the summers of 2014 and 2015. These establishments are used in the analysis because their smart meter data started within 6 months of September 1, 2011, which is a key feature of the identification strategy and is described in the following section. I classify establishments in the sample as being in coastal or inland areas based on a PG&E designation. This classification is used frequently in my analysis of peak pricing because the two regions have vastly different climates. The coastal region, which runs the length of the coast in PG&E's service territory, has much milder summers compared to the inland region.[7]

To construct the final dataset, I combine hourly usage data with establishment characteristics. Exact establishment latitude and longitude coordinates were provided by PG&E, and are used to match establishments to hourly weather data obtained from Mesowest.[8] I observe when a customer was placed on the opt-out tariff and whether they decided to opt out. I also observe industry classification in the form of North American Industry Classification System (NAICS) codes for 89.2% of establishments in the sample.

PG&E categorizes its C&I customers based on electricity consumption. This paper focuses on the smallest non-residential PG&E rate, the A-1 tariff, because the peak pricing rollout for this group allows me to causally identify program impacts.[9] I remove smaller individual meters that consumed below 800 kWH/month in the

[7]Appendix A.2 describes the creation of the dataset in detail. See Appendix Figure A.2 for a map of the 7,435 establishments used in the primary specification and their region designation.

[8]The hourly weather station data were cleaned to remove any weather stations with unreliable data and are matched to the closest establishment. The final dataset contained measurements from 297 weather stations over 2014 and 2015.

[9]Establishments are placed on the A-1 tariff if they consume less than 150,000 kWh/year and if they have peak usage of less than 75 kW. The average PG&E residential customer consumes around 8,000 kWh/year. PG&E, like most utilities, imposes a demand charge for its larger non-residential customers. This charge is based on the customer's maximum flow of electricity in a given month. A-1 establishments do not pay a demand charge.

summer of 2014.[10] This leaves me with the 19,071 establishments used in the analysis. The average customer in the sample consumed 87 kWh/day and spent $560/month on electricity in the summer of 2014. This is a larger amount than the average residential household, which consumed 21 kWh/day. Figure 1.1 shows the average summertime hourly consumption profile of the establishments in the sample, where the vertical lines indicate the peak pricing window.

There are approximately 283,000 C&I customers of this size profile in the PG&E service territory. These establishments make up 82% of the load of the small C&I class. In total, small C&I customers constitute about 2,000 MW of peak load, which is around one-tenth of PG&E's total peak load. The customers in my sample are typically smaller businesses for which energy is not a major input, including restaurants, barber shops, bakeries, corner stores, small retail shops, strip mall storefronts, law offices and doctors' offices. Energy intensive establishments from industries such as food processing, cement manufacturing, aluminum smelting or commercial establishments with large refrigeration needs are on different tariffs and are not studied because they face different electricity prices and event day prices.[11]

1.4 Empirical Strategy

1.4.1 Natural Experiment in Peak Pricing Enrollment

The nature of the PG&E peak pricing program does not permit the use of an OLS selection-on-observables design to carry out a simple comparison between enrolled customers and those yet to be enrolled. That approach would likely result in a comparison between dissimilar establishments and therefore biased estimates of program impacts. To avoid potential bias, I use an instrument that leverages a natural experiment in the rollout of opt-out peak pricing for the summer of 2015.

PG&E used a set of rules to determine when an establishment would be placed on opt-out peak pricing. They evaluated their customer base once per year starting in November 2014 to determine which establishments were eligible. This paper examines the first wave of this rollout. The regulator required that an establishment had a history of high-frequency metering data before they were placed on peak pricing, so

[10] I drop low-usage meters because most are not associated with an establishment. For example, a single meter may be attached to a sign in a strip mall, but may not be associated with other uses of the business. A full accounting of how the final dataset was constructed and cleaned is provided in Appendix Section A.2.

[11] Most larger establishments were moved to peak prices using different criteria before 2015. As a result of how this was done, there is no way to reliably identify the impacts of peak pricing on their usage.

that customers could be informed about the potential price impacts and could make informed decisions.

Specifically, *establishments' smart meter data needed to have started before September 1, 2011 to be eligible for the 2015 rollover to opt-out peak pricing.* Figure 1.2 provides a timeline of this process. I classify establishments in two groups: those that were eligible for peak pricing in 2015 and those that were not. Those establishments with high-frequency data starting after September 1, 2011 were deemed ineligible for peak pricing in 2015.

The impacts of this eligibility status can be seen more than three years after the September 1, 2011 threshold, when treatment started.[12] In November, 2014, a portion of the eligible establishments were moved to opt-out peak pricing.[13] In contrast, none of the ineligible establishments were moved and will have to wait for subsequent rollovers.

To illustrate the transition to peak pricing, Figure 1.3 breaks down the eligible and ineligible groups by the week their smart meter data were first collected. The horizontal axis shows weeks relative to the September 1, 2011 cutoff. The vertical axis displays the percent of each bin that was placed on opt-out peak pricing for the summer of 2015. A portion of the establishments to the left of the September 1, 2011 threshold were moved to peak pricing, while no establishments to the right were moved.

The date an establishment's smart meter data began is based on when its smart meter was installed. PG&E started installing smart meters in 2008, long before planning began for the peak pricing program rollout. PG&E treated installations as general capital upgrades, with installation decisions based on factors such as labor availability and logistical constraints. Installations typically took 5-15 minutes, and did not require the account holder to be present. The smart meter installation date was not related to consumption or to any observable characteristics of a given establishment.[14]

The nature of the smart meter rollout suggests that establishments on either side of the September 1, 2011 threshold are similar. The peak pricing eligibility cutoff was not known when the smart meters were installed, suggesting that establishments had no reason to strategically adjust their installation date. While the installations are as

[12] The long time lag was due to a number of requirements that the regulator had given PG&E about the information that had to be available to an establishment before it was transitioned to opt-out peak pricing. See Appendix Section A.3 for more details on these requirements.

[13] Which establishments were moved depends on technological factors, which are described later in this section.

[14] See Appendix Section A.3.1 for more details on the smart meter rollout, including quotes from annual reports describing the process.

good as random over short periods of time, there are longer-term patterns to consider. Smart meters were installed across California during this time period, but certain areas of the state were emphasized earlier in the rollout compared to others. I select customers within an eight-week bandwidth of the September 1, 2011 threshold to avoid potential bias from long-term installation trends. This bandwidth is indicated as the dashed vertical lines in Figure 1.3, and cuts the sample to 7,435 establishments.[15] Table 1.1 shows the summary statistics for a number of characteristics broken out by peak pricing eligibility. The table shows that establishments within eight weeks of the September 1, 2011 cutoff are observationally similar to each other.

One noteworthy feature of the September 1, 2011 cutoff is that eligible establishments closer to the threshold were less likely to be rolled over. This pattern is due to technical requirements that govern when high-frequency usage data is considered usable. PG&E requires that the "remote meter reads become stable and reliable for billing purposes" before they can be used for any official purpose (Pacific Gas & Electric 2010).[16] The validation process can be quick for some establishments, but can take a number of months to complete for others.[17] For this reason, establishments that had high-frequency data for longer (farther to the left in Figure 1.3) are more likely to be placed on opt-out peak pricing in the summer of 2015. The eligible establishments that missed peak pricing in the summer of 2015 due to technical requirements were scheduled to be moved over for the summer of 2016.

I use two different identification strategies to estimate program impacts. I first instrument for program participation based on the high-frequency meter data start date eligibility criterion. Second, I use a regression discontinuity approach around the September 1, 2011 threshold. This explicitly controls for an establishment's distance in days from the September 1, 2011 discontinuity in the post period, by using a trend line. Both approaches use establishment fixed effects to control for time-invariant characteristics. The unit of observation is the establishment-hour. In most specifications, I limit the sample to 2:00 pm-6:00 pm on event days in the summer of 2014 and 2015. In the summer of 2014, event days were called by PG&E, but they did not apply to this customer class. This makes them an ideal set of pre-period control days that are similar to the 2015 event days.

[15] I consider alternate bandwidths in the results section as robustness checks.

[16] PG&E still sends employees to physically read the meters monthly until this validation is complete. See Appendix Section A.3.1 for more details on the validation process.

[17] Establishments to the right of the cutoff are assumed to have a similar pattern of data validation characteristics.

1.4.2 Instrumental Variables Approach

To identify peak pricing program impacts in the instrumental variables (IV) approach, I instrument for peak pricing participation with whether an establishment's smart meter was installed before September 1, 2011, limiting my sample to establishments getting smart meters within eight weeks. I estimate the impact of peak pricing using the following two equations via 2SLS:

$$Q_{it} = \beta_1 \widehat{Peak}_{it} + \beta_2 Temp_{it} + \beta_3 Temp_{it}^2 + \zeta_t + \gamma_{ihd} + \epsilon_{it} \tag{1.1}$$

$$Peak_{it} = \alpha_1 \{Eligible \times Post\}_{it} + \alpha_2 Temp_{it} + \alpha_3 Temp_{it}^2 + \zeta_t + \gamma_{ihd} + \eta_{it} \tag{1.2}$$

Equation (1.1) is the second stage. $\widehat{Peak}_{it}$ is an indicator of peak pricing enrollment for establishment i in hour-of-sample t, which is predicted in the first stage regression (Equation 1.2) using the eligibility instrument interacted with the 2015 dummy ($\{Eligible \times Post\}_{it}$).

Q_{it} is the log of electricity consumption for establishment i in hour-of-sample t. Hourly temperature is controlled for with $Temp_{it}$ and $Temp_{it}^2$.[18] Hour-of-sample fixed effects, which control for any contemporaneous shocks that affect all establishments, are captured with ζ_t. γ_{ihd} is a set of establishment fixed effects that control for time-invariant factors. Each establishment has a separate establishment fixed effect for each hour of day (h) and day of week (d) combination because these are both significant dimensions across which establishments change their energy consumption. β_1 is the coefficient of interest and represents the average hourly reduction across peak event hours in 2015. The identifying variation comes from within-establishment variation in peak electricity consumption following the implementation of the peak pricing program in 2015.

ϵ_{it} is the error term in the second stage and η_{it} is the error term from the first stage. The panel nature of this analysis makes each of the errors potentially correlated both over time and across establishments. To account for this two-way errors dependence, I two-way cluster at the establishment and hour-of-sample level, as suggested by Cameron, Gelbach, and Miller (2011). As a result, the errors are robust to both within-establishment and within-hour-of-sample correlation.

The identifying assumption underlying the 2SLS estimation is that peak pricing eligibility is not correlated with peak electricity consumption, conditional on fixed

[18]The temperature controls are used to increase precision, but the results are robust to their omission.

effects and temperature controls, through any other mechanism than being placed on opt-out peak pricing. Formally, this is written as cov($\mathit{Peak\ Eligibility}_{it}$, $\epsilon_{it} \mid X_{it}$) = 0, where X_{it} represents the covariates and fixed effects that are controlled for in Equation (1.1). The exclusion restriction could be violated if there are time-varying trends that differentially affect establishments in the two eligibility groups. The estimation also requires a valid first stage, for which I provide evidence in Section 1.5.1.

Evidence of the validity of the research design restriction is provided in Figure 1.4, which shows the average summer 2014 (pre-period) consumption by eligibility group, after controlling for establishment-level fixed effects. The consumption patterns are similar, indicating that the eligible and ineligible establishments function as good comparison groups. Table 1.1 shows summary statistics by eligibility group for establishments in the eight-week bandwidth on either side of the September 1, 2011 threshold. I cannot reject that eligible and ineligible establishments are statistically the same across all observables.

1.4.3 Regression Discontinuity Approach

This section introduces a regression discontinuity (RD) approach that explicitly controls for the distance in days an establishment is from the September 1, 2011 threshold. I estimate the impact of peak pricing with the following two equations via 2SLS:

$$
\begin{aligned}
Q_{it} = \beta_1 \widehat{Peak}_{it} + \beta_2 X_i Post_t + \beta_3 X_i \{Eligible \times Post\}_{it} + \beta_7 Temp_{it} + \\
\beta_8 Temp_{it}^2 + \zeta_t + \gamma_i + \epsilon_{it}
\end{aligned}
\tag{1.3}
$$

$$
\begin{aligned}
Peak_{it} = \alpha_1 \{Eligible \times Post\}_{it} + \alpha_2 X_i Post_t + \alpha_3 X_i \{Eligible \times Post\}_{it} + \\
\alpha_4 Temp_{it} + \alpha_5 Temp_{it}^2 + \zeta_t + \gamma_i + \eta_{it}
\end{aligned}
\tag{1.4}
$$

Equation (2.4) is the second stage equation. As above, $\widehat{Peak}_{it}$ is an indicator of peak pricing enrollment for establishment i in hour-of-sample t, which is instrumented for in the first stage (Equation 2.3) using the cutoff-based instrument interacted with the post period. I control for the distance in days from September 1, 2011 linearly, using X_i, as suggested by Gelman and G. Imbens (2014). γ_i controls for establishment fixed effects.[19] The remaining terms are the same as those found in

[19] The results are robust to using an establishment by hour-of-day by day-of-week fixed effect.

Section 1.4.2. Inference is complicated by the discrete nature of the distance from the threshold running variable. I cluster at the distance from threshold level based on the suggestion of Lee and Card (2008).[20]

The main difference between the RD and IV approach is that the RD controls for the distance from the threshold in the post period. This technique absorbs any linear relationship between the distance from the threshold and ϵ_{it}, which removes it as a potential confounding factor in the estimation of peak pricing impacts. Identification in the RD model comes from the assumption that the relationship between ϵ_{it} and the distance from threshold does not change discontinuously at the September 1, 2011 cutoff, conditional on controls and fixed effects.

Figure 1.5 presents graphical evidence that the observable characteristics are smooth through the discontinuity. Another concern is the potential manipulation of the running variable near the threshold. I do not expect this to be a factor because the September 1, 2011 threshold was not known to the establishments or PG&E staff at the time. The top right graph in Figure 1.5 shows the count of smart meter installations by bin. There is no visible spike before or after the September 1, 2011 threshold, which is evidence that establishments did not manipulate their starting date.

The main RD specification uses the same sample as the IV approach, where establishments are restricted to have high-frequency metering data that started within eight weeks of the September 1, 2011 cutoff. In alternate specifications, I use varying bandwidths and find similar results.

1.5 Results

1.5.1 Main Results

I use the IV and RD approaches to identify the impacts of peak pricing on electricity usage. Table 1.2 shows the first stage results from estimating Equations (1.2) and (2.3). Columns (1) and (2) show the results for the sample that spans the PG&E service territory. The first stage is significant for both identification strategies, and the IV approach has a larger coefficient. The discrepancy reflects the differences between the approaches: they are identifying different local average treatment effects (LATE). The RD approach estimates the vertical difference, conditional on fixed effects, at the September 1, 2011 cutoff, which is roughly 9 percentage points, as seen

[20]Individual establishments are nested within each distance from the threshold, meaning the errors are also robust to within-establishment correlation. See Appendix Section A.4.3 for alternate clustering specifications.

in Figure 1.3. The IV approach, on the other hand, estimates the average difference between eligible and ineligible customers, leading to a higher number. The F-statistic for the IV and RD approaches are 406 and 24 respectively, providing evidence of a valid first stage. Columns (3)-(6) report the first stages for the coastal and inland regions separately. The results show a significant impact of eligibility on peak pricing enrollment for all specifications except for the coastal RD.

Table 1.3 shows the IV and RD impacts of peak pricing on electricity consumption. The sample for analysis comprises the 7,435 establishments with high-frequency data starting within eight weeks of the September 1, 2011 cutoff. Columns (1) and (2) show the impacts for the IV and RD strategies. Both show reductions in peak usage, but with p-values of .10 and .31 for the IV and RD approaches respectively. Columns (3)-(6) split the results by region, showing that the impact of peak prices varies substantially by geography and temperature. Coastal regions, which are characterized by lower electricity usage and temperatures, show almost no response to peak prices. In contrast, inland establishments reduce peak usage by 13.4% and 24.6% in the IV and RD approaches respectively, and both are significant at the 5% level.[21]

The results provide evidence that, in the warmer inland regions of California, peak pricing significantly impacts electricity usage. Coastal customers, however, do not seem to be as responsive. The regional nature of the results is consistent with Ito (2015), who finds that inland households are more price-elastic than coastal customers.

Figure 1.6 graphically shows the reduced form impacts of peak pricing eligibility on peak usage using the RD approach for inland customers. The horizontal axis bins customers by when their smart meter data were first collected, similar to Figure 1.3. The vertical axis displays the difference between average 2015 event day consumption and 2014 event day consumption. The figure presents residuals after temperature, establishment, and hour-of-sample fixed effects are removed. Customers to the right of the September 1, 2011 cutoff were not on peak pricing, while a portion of customers to the left of the vertical line were on peak pricing. The figure shows a reduction in peak consumption for peak-pricing-eligible establishments to the left of the vertical line compared to the ineligible group to the right.[22] The reduced form impacts of peak pricing seen in this figure are visible but noisy, so I focus on regression analysis for the remainder of the results section.

[21] Percent reductions reflect antilog transformed coefficients. See Appendix Section A.4.1 for the non-instrumented OLS results, which show a smaller impact of peak pricing. Appendix Table A.5 shows the results as elasticities.

[22] I remove Monday event days from the figure because they typically have a noisier response due to being announced the Friday before. By removing Mondays, it is easier to see the effects in Figure 1.6.

The role of bill protection is important to consider when interpreting the results in this paper. Establishments know they cannot lose money in the first year of the program. This creates incentives similar to those in Ito (2015), where establishments, far from making money under the program, may choose to "give up," take the bill protection, and not respond to the price. The role of bill protection can be seen by examining the financial impacts of peak pricing in its first year. Only 11% of establishments in my sample saved money in the 2015 peak pricing program, with the remainder receiving the help of bill protection. As discussed in Section 1.2.1, only 5.5% of establishments dropped out between the time they received the bill protection credit in November 2015 and the end of the second year of the peak pricing program. The low dropout rate after most establishments would have lost money in the first year, combined with the lack of bill protection in future years, suggest that my results are a lower bound for future peak pricing impacts. If establishments are exposed to potential monetary losses, they have a larger incentive to reduce their usage. It is possible that additional establishments may opt out of peak pricing after losing money, which could reduce future aggregate impacts. However, the low observed opt-out rate after the first summer suggests that this impact may not be very large. Future years of program data are necessary to resolve the impact that opt-out behavior might have on program impacts.[23]

Both the RD and IV approaches use an eight-week bandwidth around the September 1, 2011 cutoff, but the results do not change substantially at different bandwidths, as shown in Figure 1.7. The results in this section are robust to a number of other specification and clustering choices, as shown in Appendix Section A.4.

1.5.2 Spillovers to Non-Event Hours

The analysis to this point has only focused on the change in usage between 2:00 pm and 6:00 pm on event days. This ignores the scope for establishments to re-optimize their usage during off-peak hours. Figure 1.8 shows the treatment effects for inland establishments by hour of day. The results suggest that establishments begin to reduce their energy usage around 11:00 am, with the reductions becoming statistically significant by 1:00 pm. This pattern of reductions is consistent with establishments making event day changes that spill over to non-event window hours. For example, an establishment may adjust its air conditioner set point from the normal 72 degrees up to 76 degrees on event days. This behavior would reduce the overall demand for cooling on event days, leading to the reductions seen before 2:00 pm.[24] Immediately

[23]I am not able to estimate the causal impact of peak pricing without bill protection in future years because the control group used in my identifications strategy will have rolled onto peak pricing.

[24]Appendix Section A.9 shows the impacts on non-event days between 2:00 pm and 6:00 pm.

after the event window ends, usage returns to the level of the control group. Many small C&I businesses close around 6:00 pm, which might explain the return to control consumption levels.

1.5.3 Impacts of Temperature

The outdoor temperature on event days is much higher in the inland regions of California than on the coast.[25] This suggests that temperature could play a role in an establishment's demand elasticity. Reiss and White (2005) show that residential customers with air conditioners have more elastic demand than those without. Ideally, I would measure the impacts of peak pricing on establishments with air conditioning separately from those without, but this is not possible with the data available. Instead, I focus on the role that temperature plays in event day reductions.

Table 1.4 presents the results for inland establishments from interacting the treatment effect in Equations 1.1 and 2.4 with temperature.[26] The negative sign on the interaction term shows that peak reductions get larger as temperature increases. The estimated impacts are relative to a 75 degree day.[27] The IV results show a statistically significant reduction, while the RD estimates have the same sign but with a p-value of .099.

The results show that, on average, higher reductions come from higher outdoor temperatures. As a consequence, the peak pricing program may provide larger reductions on the hottest event days when the grid is most stressed. This finding is relevant to program design, because, if event days occurred only on the hottest few summer days, then reductions might be higher than the average impacts under the current program.

1.5.4 Firm Heterogeneity

Small C&I establishments use electricity to produce a wide range of goods and services in their day-to-day operations. For example, retail establishments have different patterns of electricity usage than office spaces or doctors' offices (Kahn, Kok, and Quigley 2014). In this subsection, I use the industry classification information provided by PG&E to test how different types of establishments respond to peak pricing. Specifically, I test how customer-facing and non-customer-facing

[25]See Appendix Figure A.4 for a map showing temperatures on event days.

[26]For the RD specification, I interact temperature with treatment and the distance from the threshold terms. The results for the coastal region remain insignificant.

[27]I re-center temperature at 75 degrees for ease of interpretation; this does not impact the peak pricing times temperature coefficient.

establishments each respond to peak pricing. I hypothesize that customer-facing businesses such as retail establishments or movie theaters may be less likely to reduce air conditioning usage if it affects business. Customers may choose a different movie theater or store if the indoor temperature is above expectations. On the other hand, non-customer-facing establishments such as office spaces may be more willing to reduce peak usage if it is easier for employees to adapt.[28]

I classify establishments as customer-facing or non-customer-facing using the first two digits of their North American Industry Classification System (NAICS) industry code. To determine which two-digit industries are customer-facing, I use the U.S. Bureau of Labor Statistics classification of service-providing industries.[29] From this list, I define the set of service industries that are customer-focused. This list includes retail trade (NAICS 44-45), health care (NAICS 62), leisure and hospitality (NAICS 71), and accommodation and food services (NAICS 72). All other NAICS codes are classified as non-customer-facing. These include industries such as goods manufacturing (NAICS 11-31), transportation and warehousing (NAICS 48-49) and office spaces (NAICS 52-56).[30]

Table 1.6 shows the results from running the IV regressions separately for customer-facing and non-customer-facing industries. In all cases, the customer-facing industries do not show a significant response to peak pricing. This is in contrast to the non-customer-facing industries, where the impacts are larger than previously found when considering all industries together in Table 1.3. Inland customer-facing establishments show the largest response to peak pricing, reducing their peak usage by 17.9%. The result supports the hypothesis that customer-facing industries are less price-elastic. The result also highlights that most of the overall reductions from peak pricing are coming from the non-customer-facing establishments in inland California. In other states where peak pricing is not structured as an opt-out program, it may be optimal to target non-customer-facing establishments for enrollment to generate the largest program impacts.

1.5.5 Coastal Event Days

Event days are determined based on the day-ahead forecasts for weather stations in the inland regions of California. Temperatures in the coastal region of California,

[28]For example, an employer could inform their staff of an event day in advance and encourage them to dress for a warm office.

[29]http://www.bls.gov/iag/tgs/iag07.htm

[30]The NAICS codes that I have are often imprecise, which limits the ability to finely cut the data into many different industries. See Appendix Table A.3 for a breakdown of establishments by two-digit NAICS code.

however, are not highly correlated with inland temperatures. In many cases, peak hour average inland temperatures will reach 96 degrees Fahrenheit or more, while coastal temperatures remain below 70. In some years, this can result in a relatively cool set of coastal event days even with high inland temperatures. In 2014, inland temperatures were high while all but one of the event days on the coast were below 72.[31]

Previous sections have illustrated the role that temperature and air conditioning play in an establishment reducing usage on an event day. On cool event days on the coast, there is likely a lower demand for air conditioning on the coast. If air conditioning is playing a central role in an establishment's ability to respond to peak pricing, then it is possible that establishments are less responsive on cool event days. The relatively cool summer of 2014 on the coast suggests it may not be a good control group for estimating coastal peak pricing impacts.

To estimate program impacts on hot event days on the coast, I adjust my identification strategy to use only 2015 data. I replace the 2014 pre-period event days with a set of control days in 2015 when the temperature was relatively hot, but an event day was not called. The sample is limited to days where the average temperature for both event days and non-event days was above 72 degrees. This results in a set of the hottest 7 event and 15 control days of 2015, which I use to run the analysis.[32] I further limit establishments to those with consumption over 1600 kWh/month in the summer of 2014 in order to focus on establishments that are more likely to have air conditioning.

Table 1.5 shows the results from this modified regression specification for both the IV and RD specifications. The results show approximately an 8% reduction in usage for coastal customers when the appropriate set of control days is considered. These results highlight the role that temperature plays in an establishment's ability to respond to peak prices. When it is cool out, establishments run less air conditioning, which gives them a smaller margin on which to adjust their usage compared to a control group.

1.5.6 Aggregate Impacts

The previous subsections estimated the impacts of peak pricing on a subset of small C&I establishments in the PG&E service territory. Importantly, these customers are part of a utility-wide rollout that will place all small C&I establishments on peak pricing by 2018, which has the potential to generate large peak reductions.

[31] See Appendix Table A.2 for a breakdown of temperatures by event day and region.

[32] This approach would not work for inland customers; the event days are typically the hottest days of the summer, making the non-event days bad controls.

To better understand the impacts of the fully deployed peak pricing program, I extrapolate my savings to all small C&I customers. There are three main assumptions that I make for this calculation. First, both the IV and the RD approaches reflect local average treatment effects. It is possible that the average treatment effect across all small C&I customers could be smaller or larger than those found here. Observationally, the establishments in the eight-week bandwidth are similar to those in a 27-week bandwidth, which is my complete sample.[33]

Second, these estimates capture only the short-run impacts of peak pricing in the summer of 2015. It is possible that establishment demand will become more elastic as peak pricing continues. For example, customer-facing establishments may be able to reduce peak consumption by upgrading their air conditioners to more efficient models or improving insulation. Third, I assume the estimated savings reflect future program year savings when there is no bill protection. This could result in my estimates understating aggregate impacts, as discussed in Section 1.5.1.

I extend the savings from both the IV and RD estimates using the results for inland customers from Columns (5) and (6) of Table 1.3. I focus on the inland establishments because coastal establishments only reduce their usage on a subset of the hotter coastal event days. I assume that the establishments in the eight-week bandwidth are representative of all small inland C&I customers, and that long-run opt-out rates will be similar to those observed in the first two years of my sample. I combine this with customer count information provided by PG&E to estimate the projected total impacts when the program is fully rolled out by the summer of 2018.[34] Using this technique, I find that small C&I establishments will provide reductions of 118 MW and 216 MW in peak load for the IV and RD approaches respectively.

1.6 Welfare Impacts of Peak Pricing

In this section, I first introduce a model to evaluate the welfare impacts of the peak pricing program, which I calibrate using the empirical peak demand reductions from the previous section. Using this approach, I consider changes to the current peak pricing program to better target the long-run investment inefficiencies that result from flat-rate pricing. I find that, by changing when event days are called, and adjusting the event hour price, program outcomes can be greatly improved. I conclude by benchmarking the impacts of peak pricing against the first-best, real-time price using a simple theoretical energy pricing model.

[33] See Appendix Figure A.3 for a comparison of 2014 pre-period consumption across the two groups.

[34] A full accounting of the assumptions and calculations can be found in Appendix section A.5.1.

1.6.1 A Model of Welfare Impacts from Peak Prices

The model is based on the current regulatory process in California, which is responsible for capacity construction decisions. Most other states follow a similar process. In the model, peak pricing reduces the level of summer peak demand, which in turn reduces long-run capacity requirements and saves costs by avoiding power plant construction. This framework allows me to calculate the welfare benefits of peak pricing in a manner that reflects how capacity decisions in electricity markets are made. The existing literature typically calculates the welfare impacts of alternative pricing policies using a stylized model of electricity prices and power plant construction (Borenstein 2012; S. P. Holland and Mansur 2006; Borenstein and S. Holland 2005; Borenstein 2005). These models provide insight on the welfare impacts of alternative pricing policies, but use assumptions that do not realistically portray the nature of the binding capacity constraint in electricity markets.

The structure of electricity markets is defined by the lack of cost-effective storage, which requires supply and demand to be balanced in real time. This feature introduces a capacity constraint equal to the total capacity of generators; blackouts will result if demand exceeds this constraint at any time. The stylized models of electricity markets do not consider this constraint, and assume that the price and demand for electricity are able to adjust quickly enough to avoid shortfalls. Additionally, such models assume a cost to build new generators, but not construction time. In practice, it can take six years from the initial proposal for a power plant to begin generating electricity. Much of this process is governed by the regulator that sets the amount of generation capacity that a utility must have on hand to avoid blackouts. The stylized models used in the literature do not reflect the complexities of the regulatory process and how this impacts electricity market outcomes.

I introduce a model based on the actual "resource adequacy" process, where the regulator mandates how much peak generation capacity the utility must have on hand (P. Joskow and Tirole 2007).[35] These peak capacity requirements are typically met by building specialized "peaker" power plants, which have a low capital cost but a high marginal cost of generation. Some of these plants run for only a few hours on the hottest day of the year. A large amount of peaker capacity is expensive to build and maintain.

Typically, the regulator forecasts future peak demand using historical data and load growth projections. Using this forecast and a valuation of blackouts, they set a resource adequacy level for the utility in the coming year. The model I introduce

[35] The process I model is based on the California resource adequacy process, but is representative of how capacity requirements are set in most states.

is based on how peak pricing changes the resource adequacy process. The model proceeds in three steps that happen yearly.

In step 1, no peak pricing program has been implemented. The regulator has information about the distribution of historical peak loads and temperatures, which also includes information about the peak load from the previous summer. I denote this information set as L_0. The regulator uses this information to determine how much peak capacity is needed, using the decision function F(), which does not change over time.[36] I assume this is a well-defined process known to all market participants, and that the regulator sets capacity high enough that there will be no generation shortages in the coming year.[37] I define this peak capacity requirement as $X_1 = \mathrm{F}(L_0)$.

In step 2, the utility acquires capacity X_1 at cost $\mathrm{C}(X_1)$. I assume that the utility must fulfill this requirement and that the regulator perfectly observes the utility's behavior. The cost function is linear and does not change over time. For simplicity, it reflects the utility's yearly cost to acquire capacity.

Step 3 is when the demand for the year is realized. In the absence of peak prices, demand would have reached a peak level of L_1. However, with the implementation of peak prices and their corresponding demand reductions, the new load is $\widetilde{L_1}$, such that $\widetilde{L_1} < L_1$. For simplicity, I assume peak pricing reduces peak load on all event days in a summer by the same amount, and that this amount remains constant from year to year.

The majority of the benefits from peak pricing come from this reduction in summer peak demand. This can be seen in Panel A of Figure 1.9, using a simplified peaker capacity supply curve. By reducing the total peak demand, peak pricing reduces the total generation capacity necessary to satisfy demand. This saving in capacity cost reflects the high costs associated with building generation capacity and is the main driver of savings under the peak pricing program.

The second impact from peak pricing is the surplus loss that results from changes in customer behavior when paying higher prices on event days. For example, establishments may choose to run their air conditioner less, leading to a less comfortable

[36] If the regulator were an optimal social planner, the F() decision function would balance the benefits of reliability against the costs of acquiring capacity, and pick an optimal capacity requirement for the utility. In practice, most regulators are risk averse and put a very high cost on supply shortfalls that result in localized blackouts. As a consequence, regulators typically set very high reserve requirements for utilities. I do not take a stand on the exact approach the regulator should use. I simply assume they follow the same rule each year.

[37] Blackouts from demand exceeding capacity are rare. The current California process requires capacity at 1.15 times the projected peak load. This level is sufficiently high to assure that capacity limits will never be reached. See P. Joskow and Tirole (2006) and P. Joskow and Tirole (2007) for a discussion of optimal capacity with the possibility of rationing.

indoor environment. This impact can be seen graphically in Panel B of Figure 1.9. To calculate this impact, I first recognize that the electricity still sold at the peak price (to the left of Q_1) induces no change in total surplus, as it is just a transfer from consumers to producers. For units that go unsold due to the price increase (to the right of Q_1 and to the left of Q_0), the change in surplus is the area under the demand curve minus the resource savings from not producing these units. In this case, the resource savings are equal to the fuel savings of the peaker plants that would otherwise be used to generate this electricity.

I value the reduction in fuel used to run a peaker plant at its short-run marginal production cost (SRMC), which I assume to be $.102/kWh based on current natural gas prices (California Energy Commission 2015). I use the SRMC for this calculation because I assume the regulatory process dictates that sufficient capacity is available at all hours of the year, meaning the surplus losses are net of the short-run costs associated with running a peaker plant. This set of calculations leaves what I term the net consumer surplus loss, which is represented by the shaded triangle in Panel B of Figure 1.9. I use a linear demand curve for simplicity and because it provides a conservative upper bound on the net CS losses compared to other concave alternatives such as a constant elasticity of substitution demand curve. I define the net consumer surplus loss in year 1 as CS_1.

The process now restarts at step 1 in year 2. In the world with peak pricing, the regulator observes peak load $\widetilde{L_1}$ and sets peak capacity requirements for the coming year $\widetilde{X}_2 = \mathrm{F}(\widetilde{L_1})$. In the non-peak pricing scenario, the regulator observes peak load L_1, resulting in peak capacity requirement $X_2 > \widetilde{X}_2$. In step 2 of year 2, the utility must acquire capacity at cost $\mathrm{C}(\widetilde{X}_2)$ and $\mathrm{C}(X_2)$ for the peak and non-peak pricing scenarios, respectively. This process continues and repeats for both scenarios over time.

To calculate the welfare impacts of the peak pricing program, I subtract the costs in the peak pricing scenario from the costs in the non-peak pricing scenario. The benefits are calculated over T periods on N event days per year using discount rate r. The change in welfare from implementing peak pricing is defined as follows:

$$\Delta Welfare = \sum_{t=1}^{T} \frac{C(X_t) - C(\widetilde{X}_t) - N \times CS_t}{(1+r)^t} \tag{1.5}$$

This calculation compares the cost of peak generation capacity with standard pricing to the lower peak capacity needs when peak prices are used.

The model in this section leans on a stylized formulation of net consumer surplus losses. It considers only the surplus losses to establishments that occur between 2:00

pm and 6:00 pm on event days, when prices increase. It is possible that customers are responding to peak prices in ways that that are not reflected in these hours. For example, Section 1.5.2 shows that peak pricing enrolled establishments reduce their usage before the 2:00 pm-6:00 pm event window starts. The model presented here does not capture the net consumer surplus losses associated with this change in behavior before the event window, or any other non-event window impacts. Bill protection could also impact the magnitude of the net consumer surplus losses. If the price signal to establishments is potentially affected by bill protection, then the response to peak pricing may not reflect the true net consumer surplus impacts. I conduct robustness checks using different levels of CS losses to see how these factors impact the welfare estimates.

1.6.2 Calculating Welfare Impacts of PG&E's Peak Pricing Program

In this section, I calculate the welfare impact of the PG&E peak pricing program using the model from the previous section and my empirical results.[38] Some of the simplified assumptions in the model are adjusted to better reflect the PG&E service territory. In the model, the utility purchases capacity yearly at cost $C(X_t)$. In practice, the peaker plants that are used to satisfy peak demand typically last at least 30 years. To approximate the cost function, I use the construction cost of a single cycle peaker plant. The California Energy Commission (CEC) estimates it costs \$1,185,000/MW to build a natural gas combustion turbine peaker plant (California Energy Commission 2015).[39] Using these plant construction numbers and my empirical estimates, I find that the peak pricing program would provide a one-time saving of \$139 million in construction costs with the IV approach. I assume this cost savings occurs in year 1 of a 30 year program. To value the total impacts of the program, I include the discounted stream of annual costs and benefits. Reducing peaker capacity provides an annual benefit of avoided staffing and maintenance costs, which in this case totals \$3.05 million per year.

To make the CS loss calculation, I use a linear demand curve as discussed in the previous section. One important difference between the model and PG&E prices is that retail electricity rates for small C&I customers are set at \$.25/kWh. Retail prices are higher than the short-run marginal cost of production because fixed costs

[38]From this point forward, I present calculations using only the IV estimate from inland establishments for simplicity. Appendix Section A.5.2 outlines why I make this choice and the welfare benefit calculations using the RD estimates.

[39]All values used in this paper are in 2016 dollars. Original 2011 values are inflated using the IHS North American Power Capital Costs Index.

are recovered volumetrically in PG&E. In the previous section, I set retail rates at the short-run marginal cost of production, which I assume to be \$.102/kWh for a peaker plant (California Energy Commission 2015). Establishments were willing to pay the \$.25/kWh price for their electricity during these peak periods, meaning economic surplus is lost on event days when prices are increased and consumption is reduced. Graphically, this impact is represented by a rectangle between the \$.25/kWh electricity price and the \$.102/kWh short-run marginal cost of production. The total CS loss from peak pricing is this rectangle plus the triangle under the demand curve, shown in Figure 1.9. Using my empirical estimates, I find that the total net consumer surplus loss in 2015 equals \$3.14 million/year.

The PG&E peak pricing program gives enrolled establishments a \$.01/kWh discount on all non-event day electricity consumption. As a result, establishments will consume more electricity in off-peak hours, resulting in increased consumption across almost all summer hours.[40] Using my elasticity estimates and linear demand, I calculate these welfare gains to be \$0.84 million/year.[41]

To come up with a total welfare value, I take the construction costs and add on the discounted stream of costs and benefits detailed above. This results in total welfare benefits of \$154 million (2016 dollars) using a 3 percent real discount rate and a 30 year horizon.[42] These numbers represent the welfare benefits of running the peak pricing program every summer for 30 years. Embedded in this back-of-the-envelope calculation is the assumption that electricity supply and demand will not change in ways that affect the numbers calculated above. I also assume that the operation and maintenance costs stay constant over the life of the plant, which likely understates the costs as the plant ages. Furthermore, establishment demands are likely to become more elastic as they face peak prices over many summers.

The above welfare calculations only capture the negative net consumer surplus impacts from peak pricing that occur between 2:00 pm and 6:00 pm. Establishments may undertake behaviors that affect consumption outside of the event window, resulting in welfare impacts that are not captured with this model. Bill protection may also impact the welfare estimate by affecting consumer response to the peak

[40]Importantly, this price reduction is welfare-improving because the retail price of electricity for small C&I customers exceeds any reasonable social cost.

[41]This is a strong assumption because I am applying my demand curve estimates, derived for the period between 2:00 pm and 6:00 pm on event days, to all other hours in the summer. Using the empirical analysis on non-event hours in the summer of 2015, I can reject the level of responsiveness I am using for this calculation. Ultimately, the response from the off-peak CS gains is small and does not significantly impact outcomes.

[42]The results are not very sensitive to discount rate assumptions because most of the benefit is incurred upfront with the avoidance of capital construction costs. The other annual costs and benefits are roughly offsetting.

price. For robustness, I consider a scenario where the net consumer surplus impacts are double what is calculated above. Using this assumption, I find the total welfare impacts of the program to be $108 million. This shows that, even under a conservative set of assumptions, the welfare impacts of peak pricing remain positive.

PG&E is one of three major utilities in California, along with Southern California Edison and San Diego Gas and Electric, to implement peak prices under order of the Public Utilities Commission. As a result, most of California is in the process of implementing opt-out peak prices for C&I customers. I use my estimates to inform the impacts of the larger rollout across the state. The first row of Table 1.7 shows the welfare impacts of peak prices for small C&I customers. Columns (2) and (3) show the PG&E savings estimates extended to the three major investor-owned utilities (IOU) and for the full state, respectively. I find that the IOU-wide benefits total $394 million while the California-wide benefits are $573 million over a 30 year period. There are a number of assumptions used to make these welfare calculations. First, I assume that small C&I customers in PG&E are similar to those in other regions. This may be a reasonable assumption in California, but it is likely less true for the full U.S. grid. Column (4) shows the savings estimates extended out to the national grid, showing a potential $17 billion benefit of this policy. This number represents the impact only for small C&I establishments, which I assume to be 10% of peak load across the U.S. The magnitude of this estimate highlights the significance of the distortion caused by flat retail pricing.[43]

The second row of Table 1.7 considers the impacts for the full set of C&I customers. I assume the same 13.4% reduction in peak usage for all of these customers as I estimated for the small C&I customers.[44] The welfare estimate is likely a lower bound, since peak pricing adds $1.20/kWh to the price of electricity for large C&I customers on event days rather than the $.60/kWh for small establishments. Extending this estimate nationally results in a $82 billion savings estimate. This estimate assumes that the national C&I makeup of the U.S. reflects California. This large potential welfare benefit provides perspective on the size of the distortion that flat retail prices introduce.

1.6.3 Targeting the Capacity Constraint

The PG&E peak pricing program is designed in a manner similar to other peak pricing policies around the U.S. The utility has discretion over when to charge higher prices on 9 to 15 event days per summer. In this section, I consider the

[43] See Appendix Section A.5.3 for the data and assumptions used in these calculations.

[44] I adjust the savings estimates for the 43% of large C&I customers that have opted out of peak pricing since it was first introduced in 2010.

welfare implications of how event days are chosen and the price charged during event hours. I do this in the context of the current peak pricing program, where resource adequacy requirements guarantee that there will be sufficient capacity available to avoid blackouts.

PG&E calls event days using day-ahead weather forecasts. When the average forecasted temperature for inland California exceeds a trigger temperature of 96 or 98 degrees, an event day is called. The trigger temperature is based on how many event days have been called so far in a given summer and on historical weather trends.[45] This approach is effective at selecting the top 12-15 demand days each summer, but it is not designed to maximize the net benefits of the peak pricing program.

The typical summer in California has a small number of days with very high demand that are responsible for peak load. For example, the difference between the demand on the highest event day and the median event day in 2015 was 1,220 MW, more than 14% of total peak load.[46] The few highest demand days each summer drive resource adequacy requirements and the long-run construction of peaker plants. I define as "super-peak" days the set of days each summer for which calling an event day reduces the total summer peak load. The number of super-peak days each summer depends on both the level of reduction due to peak pricing and the number of high-demand days.[47] Most Northern California summers have between one and three super-peak demand days based on the estimated reduction due to the peak pricing program for small C&I customers.

The role of super-peak days in the 2015 program can be seen in the first two columns of Table 1.8. To project the impact that the peak pricing program for small C&I establishments will have once it is fully rolled out, I use the aggregate 118 MW reduction that projects outcomes for 2018. This reduction would lower the 2015 summer peak from 19,451 MW to 19,333 MW, which would become the new summer peak $\widetilde{L_t}$. In 2015, no event days other than the one with the highest demand will affect $\widetilde{L_t}$. For example, reducing the load on September 9, 2015 from 19,017 MW to 18,899 MW will not affect $\widetilde{L_t}$ and will not provide savings in long-run generation capacity investment.[48]

[45]The trigger temperature is adjusted every 15 days throughout the summer to hit the target number of 12 to 15 event days. See Appendix Section A.1 for more details.

[46]The same pattern holds for all years 2010-2015, with the difference between maximum and median peak load of 1,600 MW.

[47]If the reductions from peak pricing were larger, then there could be more super-peak days each summer where calling an event day would reduce peak load $\widetilde{L_t}$. For example, the number of super-peak days could go up if the large C&I establishments were included in the calculation.

[48]A list of the top 20 demand days in 2015 can be seen in Appendix Table A.4.

Table 1.8 shows the 2015 event days with the welfare impacts broken out. These values reflect the welfare impact associated with each event day in 2015, using the ex-post information about the realized demands. In practice, the peak pricing program is based on day-ahead forecasts, which introduces significant uncertainty about which event days will provide benefits when they are called. Column 3 shows the capacity value of reducing peak load. Only the highest demand event day of the summer provided capacity cost savings, because none of the other event days affect summer peak load. Column 4 shows the net consumer surplus loss figure of $209,000 per event day, which is reported in the same discounted manner to allow for easy comparison. All of the non-super peak event days reduce the welfare impacts of the program without providing capacity cost savings.

The cost of non-super peak event days quickly adds up. Each extra event day that does not provide capacity cost savings results in a loss of $4.2 million of net consumer surplus over the life of the program. A refinement to the peak pricing program would call just the super-peak event days each summer.[49] This approach is challenging with event day programs because it is not possible to forecast ex-ante which summer days will be super-peak (Borenstein 2012). Despite this limitation, there are a number of improvements that could be made to the current program using the day-ahead information that is available to PG&E.

One simple change to the peak pricing program is to tighten the criterion used to call an event day. The second to last column of Table 1.8 shows the day-ahead temperature forecasts for the inland region of California. An event day is called when this temperature equals or exceeds the "trigger temperature" set by PG&E, which is shown in the last column.[50] The current set of trigger temperatures typically calls the super-peak demand days each summer, but also includes a large number of additional days that do not provide capacity cost savings. A simple adjustment to the peak pricing program would move the trigger temperature to 101 degrees and remove the current 9 days per summer minimum. This approach uses the same day-ahead temperature forecast that PG&E currently uses to pick event days. It would result in a program that is better targeted at the super-peak event days, and would result in fewer low-demand event days each year.

In an electricity market with regulated resource adequacy requirements, the impacts of missing a super-peak event day are a higher summer peak $\widetilde{L}_t$ and the costs of building capacity in a future period. In most cases, the welfare loss from missing a super-peak event day is much higher than the benefit of avoiding a non-super

[49] It may be useful to set a minimum number of event days so that establishments do not forget they are on the program. I have not found any research that identifies the impact of using too few event days.

[50] See Appendix refPDP program details for more details on trigger temperatures.

peak event day. Any day-ahead program must take this tradeoff into account. The proposed 101-degree trigger temperature accurately selects the super-peak event days over the last five years using day-ahead temperature forecast data, but a different trigger may be preferred in the future. For example, climate change may impact the intensity and frequency of high temperature days, which could necessitate a further refinement.[51]

The second dimension of the peak pricing program that could be adjusted is the level of the event day price. Currently, small C&I establishments pay $.85/kWh during event windows, which is $.60 higher than their typical rate. Wholesale prices are routinely above $.85/kWh, and the peak price for large C&I PG&E establishments is set at $1.35/kWh. This level is designed to reflect the long-run value of capacity and is based on the regulator's avoided cost of capacity (California Public Utilities Commission 2001). There is no reason to charge different event day prices to different customer classes, because both are subject to the same capacity constraint that drives system costs. If $.85/kWh is below the efficient wholesale cost of electricity on event days, there are potential welfare gains from raising the event day price for small C&I establishments.

I quantify the welfare benefits of changing the number of event days and level of peak prices in Table 1.9. To estimate the impact of higher prices, I assume a linear demand curve, as in the previous section, and extend the results from the current peak pricing program. The current peak pricing program has an event price of $.85/kWh on 15 days per year and is shown in the top left entry. Column (2) shows outcomes if the small C&I peak price were raised to $1.35/kWh, the level paid by large establishments. It shows that using the current 15 event days per summer and increasing the event price from $.85/kWh to $1.35/kWh would increase the welfare benefits from $154 to $204 million. The third column shows the impacts of a peak price set at $1.85/kWh.[52] Moving down the table decreases the number of event days per summer from 15 to 8 to just the 3 super-peak days.[53] The table shows that moving to a 101 degree trigger and using the large C&I peak price of $1.35/kWh – both of which are realistic adjustments – could improve program outcomes by 87%.

[51] I also considered more complicated regression-based event day models using load forecasts, but this adds unnecessary complexity without additional insight. A lower trigger may also be optimal when including the reductions from large C&I establishments, which can increase the set of super-peak days through larger reductions.

[52] The calculations assume that peak wholesale prices are greater than or equal to the peak price in each column. If, for example, peak prices only reached $1.50/kWh, then the results in Column (3) would overestimate the benefits.

[53] The estimated impacts assume that the super-peak days are correctly called as event days under all three approaches. Forecasting errors could reduce the benefits if a super-peak event day is missed.

1.6.4 Comparing Peak Pricing to First-Best Policy

To put the second-best peak pricing program in perspective, I compare outcomes to the first-best alternative. Real-time pricing has been shown by previous research to result in efficient long-run outcomes, making it a useful benchmark (Borenstein 2005; Borenstein and S. Holland 2005). For this exercise, I consider a theoretical energy-only electricity market where electricity supply and demand are cleared continuously with a uniform price auction.[54] I assume long-run capacity construction decisions are made through the resource adequacy process outlined in the previous sections.[55]

With real-time prices, customers face retail rates that change every five minutes to reflect the real-time wholesale cost of electricity. I assume customers are fully informed about the real-time price they are paying, and that their usage reflects the five minute price.[56] To allow a simple comparison between peak pricing and a real-time price, I assume the wholesale price takes on two distinct values. The low price reflects the marginal cost of generation at high-efficiency natural gas power plants, which I set at \$.10/kWh. When demand exceeds the capacity of the low-cost plants, the price of electricity spikes to the high level.[57] The high price reflects the long-run cost of generation, which includes the costs of building and running peaker power plants to meet demand. I assume a high real-time price of \$1.35/kWh, which corresponds to the peak price paid by large commercial and industrial customers in the existing program. When demand drops to a level where the base load capacity is sufficient to balance load, the price returns to the low price level.

I benchmark outcomes under the peak pricing program against the benefits under real-time pricing. I first consider the existing peak pricing program, where prices are increased to \$.85/kWh between 2:00 pm and 6:00 pm on 15 event days per summer.[58]

[54] I use a simplified market design because the California electricity market has a wholesale price cap of \$1.00/kWh, as well as secondary markets for providing capacity, which make the comparison challenging.

[55] The first-best outcome reflects any inefficiencies that may exist in the resource adequacy capacity investment process.

[56] Real-time pricing programs could have prices vary as frequently as every minute or in larger 15-30 minute increments. P. Joskow and Tirole (2006) suggest that customers may not respond to short-run changes in electricity price if transaction costs are too high. They suggest that this cost will be reduced through the use of advanced technologies that can quickly take advantage of price variation.

[57] This pricing structure reflects a retail electricity model with fixed charges, where the retail rate reflects the marginal cost of generation. Depending on natural gas prices, the cost at a high-efficiency natural gas power plant may be lower than \$.10/kWh. A full accounting of the assumptions can be found in Appendix Section A.5.4.

[58] I assume 15 event days will be called each summer, to reflect the summer of 2015 for which peak pricing impacts were estimated.

This can be seen in Panel A of Figure 1.10. The well-targeted peak pricing program uses the optimal event hour price of \$1.35/kWh on only eight event days per year.[59] This can be seen in Panel B of Figure 1.10. In both scenarios, I assume there are three super-peak event days each summer that provide capacity savings, and that these will be called as event days under both systems. During the super-peak days, I assume the price is at the high level for five minutes between 2:00 pm and 6:00 pm.[60]

The model outlined for this calculation is stylized in nature and makes a number of simplifying assumptions that could impact outcomes. Real-time prices typically vary throughout the day and year to reflect transmission constraints, variation in short-run marginal generation costs or other system costs. Peak prices, by having only two possible price levels on a set number of event days, are not able to capture the benefits from this type of price variation. Ultimately, these impacts are likely small compared to large capacity savings benefits from reducing peak load.

To compare peak-pricing to first-best, I use my empirical estimates to calculate the welfare gains under peak pricing and compare them to the outcomes under real-time pricing.[61] For the existing peak pricing program, the difference between first and second-best comes from two sources. First, by charging an event price below \$1.35/kWh, peak pricing will generate lower capacity construction savings than will real-time pricing. Second, under the current peak pricing program, the event price will be charged for 60 hours per year compared to just 15 minutes under the real-time price. I choose a short period of time during which real-time prices are at the high level in order to remain conservative in reporting the benefits of peak pricing compared to real-time pricing. The longer the high event price is charged while real-time prices are low, the lower the relative benefits that peak pricing provides. The well-targeted peak pricing program, by setting peak prices at \$1.35/kWH, provides the same capacity construction savings as the real-time price. The lower number of event hours each year (32) also reduces the extra net consumer surplus loss that comes from non-super peak event days.

Using this approach, I find that the current peak pricing program provides 43% of the welfare benefits of the first-best approach. The result illustrates that the current

[59]It may be ideal to set the peak price slightly below \$1.35/kWh due to the net consumer surplus loss caused by peak pricing. For simplicity, I assume the well-targeted peak price is set at \$1.35/kWh. I use the temperature trigger proposed in Section 1.6.3 to select eight event days per summer.

[60]The short length of time at peak is a conservative assumption with respect to the value of peak pricing. I consider longer periods of peak prices as a robustness check.

[61]I use the same elasticity for both the peak pricing and real-time price reductions. I assume customer response to the high price will be the same whether they face the high price for a short time or the full peak window. Wolak (2011) showed that, for residential customers, the response to peak prices was similar using both a short and a long event window.

program is providing some value, but performs poorly compared to the first-best policy. The well-targeted peak pricing program is able to make significant welfare improvements, delivering 80% of the first-best outcome. This result underscores the value of targeting. In markets with a binding capacity constraint, directly targeting the distortion caused by the constraint is an effective tool to improve welfare.

The results of the benchmarking analysis depend on both the empirical estimates and the modeling assumptions. The empirical estimates inform the level of capacity savings and the net consumer surplus loss from charging a higher price. There are a number of modeling assumptions that can affect the levels of the benchmarking numbers, but do not significantly shift the qualitative results. First, I assume the peak price is set at $1.35/kWh. This level is based on a PG&E valuation of capacity, but it may not reflect the true long-run cost of supply. If the optimal event day price were higher, it would reduce the effectiveness of the peak pricing program compared to real-time pricing. Second, the assumption that real-time prices are at peak for only 15 minutes per year underestimates the relative value of peak pricing compared to real-time pricing.[62] Third, the current approach measures net consumer surplus loss only between 2:00 pm and 6:00 pm on event days. If net consumer surplus losses from peak pricing were higher, the relative benefits of peak pricing would be lower.[63]

The benchmarking model is useful in understanding the impacts of poorly targeting the peak pricing program. Table 1.10 shows a number of alternate scenarios that consider how an incorrectly targeted peak pricing program might perform. As before, I assume the high real-time price is $1.35/kWh. Column (1) mirrors the current program, where peak prices are set at $.85/kWh; Column (2) shows the results for the correctly chosen peak price; and Column (3) shows the impacts if peak prices were set too high, at $1.85/kWh. The first row shows the outcome when eight event days are chosen per year using the 101 degree trigger of the well-targeted program. The second row shows the outcomes with 15 event days. The bold entries correspond to the current and well-targeted peak pricing programs discussed previously. The other entries show the consequences of poorly targeting the peak pricing program.[64]

The benchmarking model shows that, while the returns to targeting can be large, the downsides to incorrectly targeting are also significant. Setting the wrong price or calling too many days reduces program effectiveness. For example, calling 15

[62]This assumption has a relatively small impact on outcomes. For robustness, I consider the extreme case where real-time prices are at the peak for all four hours between 2:00 pm and 6:00 pm. Using this assumption, the benefits of the current program would be 46% of the first-best policy.

[63]I find that, when net consumer surplus losses are doubled, the current program and the well-targeted program provide 32% and 60% of the benefits, respectively.

[64]See Appendix Table A.12 for a robustness check where prices hit the peak for the full four-hour period between 2:00 pm and 6:00 pm.

event days per year at a price of $1.85/kWh would capture only 19% of the first-best outcome. The results highlight the value that empirical research can provide in measuring program outcomes and using these estimates to further improve the program. The ability to design a well-targeted program depends on the aggregate peak pricing reductions, combined with knowledge of institutional details to value the costs and benefits of the program. Using these insights helps inform the best way to target the costly capacity constraints by observing the underlying structure of peak-event days on the PG&E grid. The estimation of the short-run electricity demand curve allows me to balance the net consumer surplus losses under peak pricing against the capacity cost savings from higher prices. Taken together, these results suggest that it is possible to achieve four-fifths of the first-best outcome using a well-targeted, second-best policy.

1.7 Conclusion

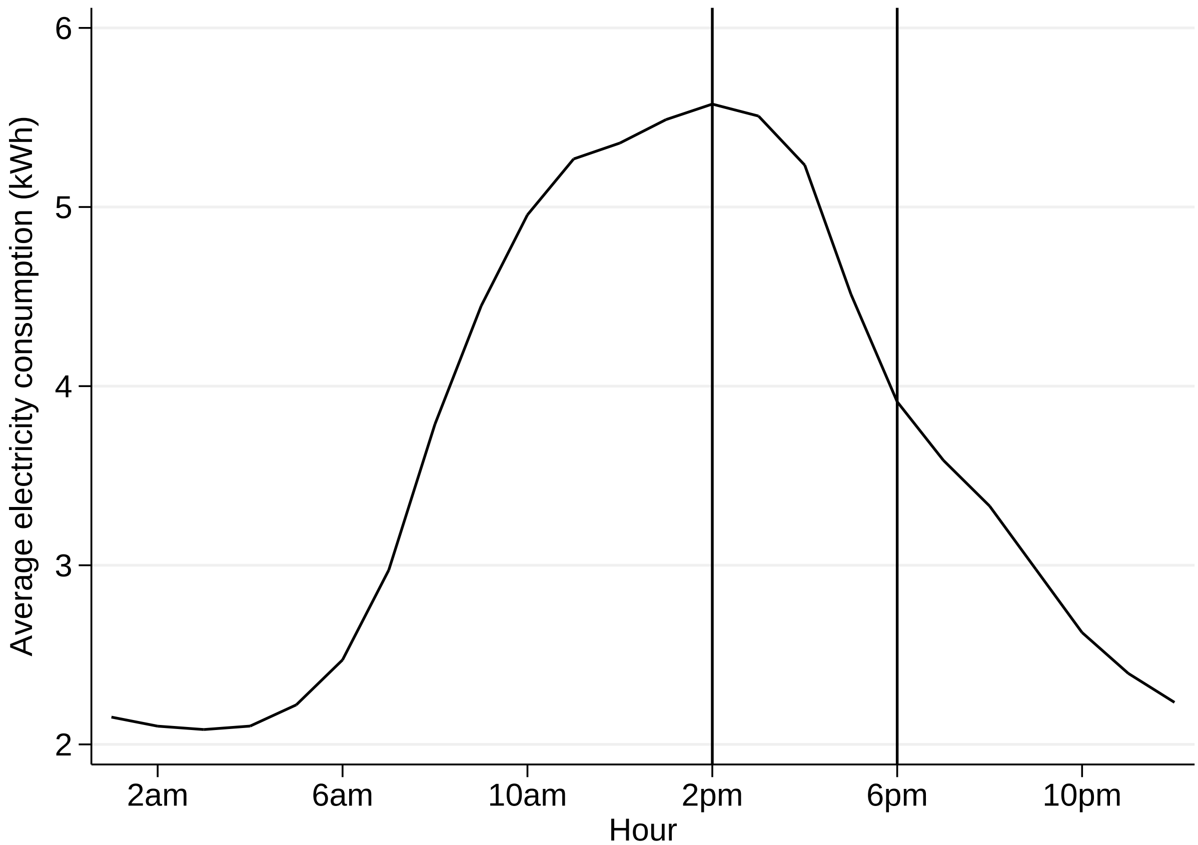
Retail electricity customers in the U.S. are typically charged a flat price per kWh consumed. This time-invariant price does not reflect the cost of capacity at peak demand hours. This paper studies a policy, peak pricing, that charges higher prices to retail customers on high-demand days when it is more costly to supply marginal units of electricity. Using quasi-random variation in program implementation and two different identification strategies, I find that establishments reduce their usage between 2:00 pm and 6:00 pm by 13.4%. In the aggregate, the peak pricing program will provide 118 MW of peak demand reductions in the PG&E service territory when fully implemented. The peak savings reduce the amount of generation capacity required at peak, yielding $154 million of welfare benefits. I compare outcomes to a theoretical first-best, real-time pricing policy, finding that the current program captures 43% of the benefits. I show that a well-targeted peak pricing program could provide greatly improved outcomes, equaling 80% of the first-best outcome.

This paper fills an important gap in the literature by providing the first evidence of how commercial and industrial customers respond to peak pricing. This is particularly important as the popularity of peak pricing programs continues to grow, fueled by the installation of low-cost, advanced metering technology. Further research is required to better understand the impacts of peak pricing on large C&I customers. They constitute over 50% of California and national electricity demand, making their response to peak prices important for future energy policy.

The approach I take in this paper is relevant to a wide range of markets where prices do not reflect the cost of capacity constraints at peak demand periods. I use empirical estimates of a second-best pricing policy to make welfare calculations and compare outcomes to the first-best alternative. This framework could be used to evaluate and improve other second-best policies. For example, most bridge tolls do not adjust to accurately reflect congestion costs at peak commute hours. In the long run, this may lead to the construction of excess transportation infrastructure, similarly to the manner in which flat electricity pricing leads to excess generation capacity. More work in these and related settings will help to validate these insights and can lead to the improvement of other second-best policies.

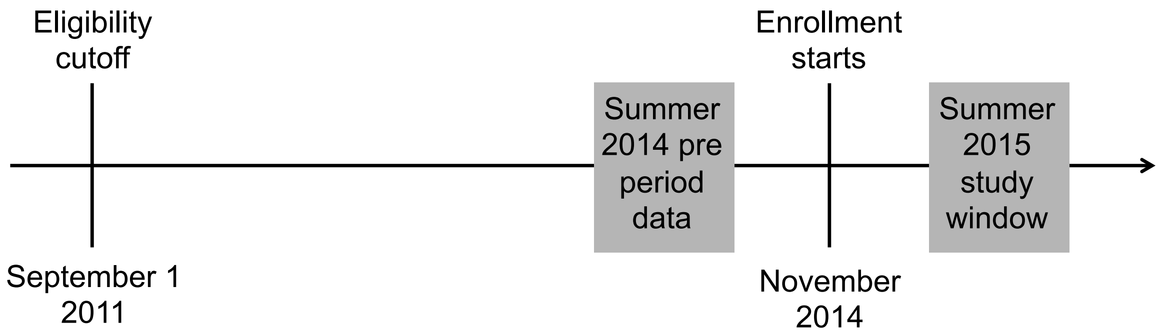
Figures

Figure 1.1: Average Consumption Profile of Small Commercial and Industrial Establishments in Sample



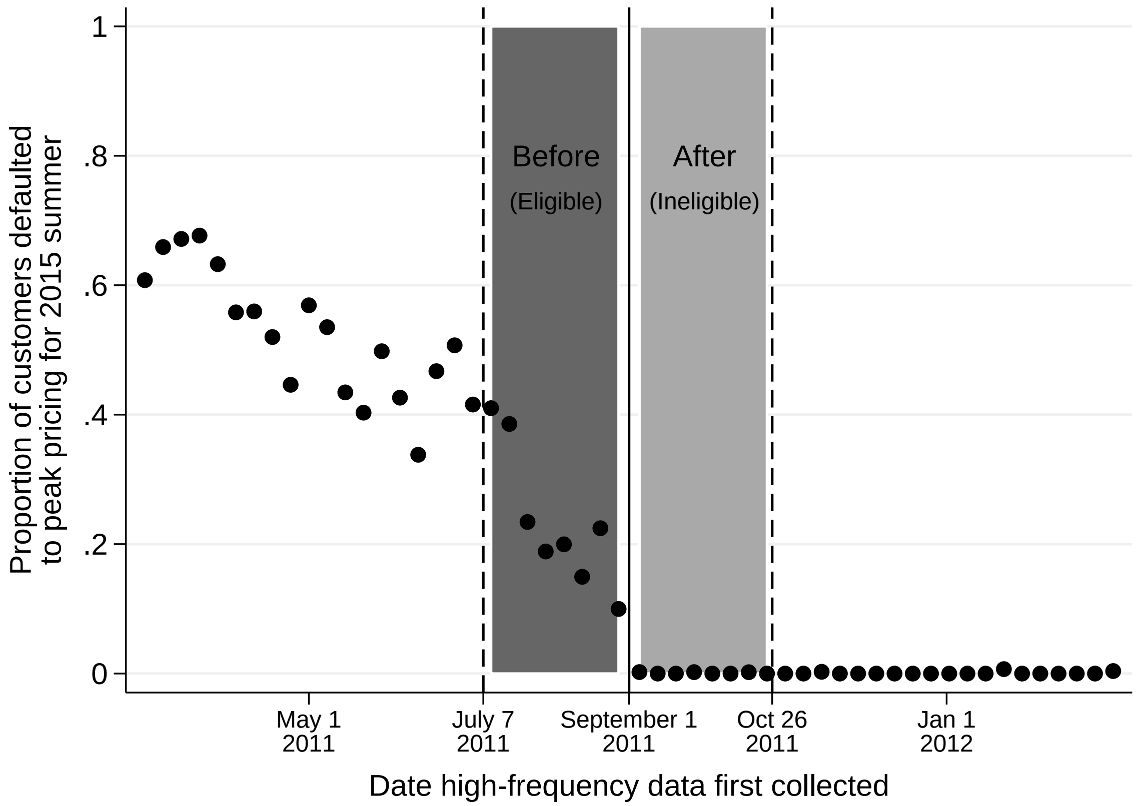
Note. - This figure shows the average consumption profile of the establishments in my analysis for all weekdays during the summer of 2014. The vertical lines signify the beginning (2:00 pm) and end (6:00 pm) of the peak event window. The system peak demand for the PG&E grid typically is between 4:00 pm and 6:00 pm.

Figure 1.2: Timeline of Peak Pricing Rollout



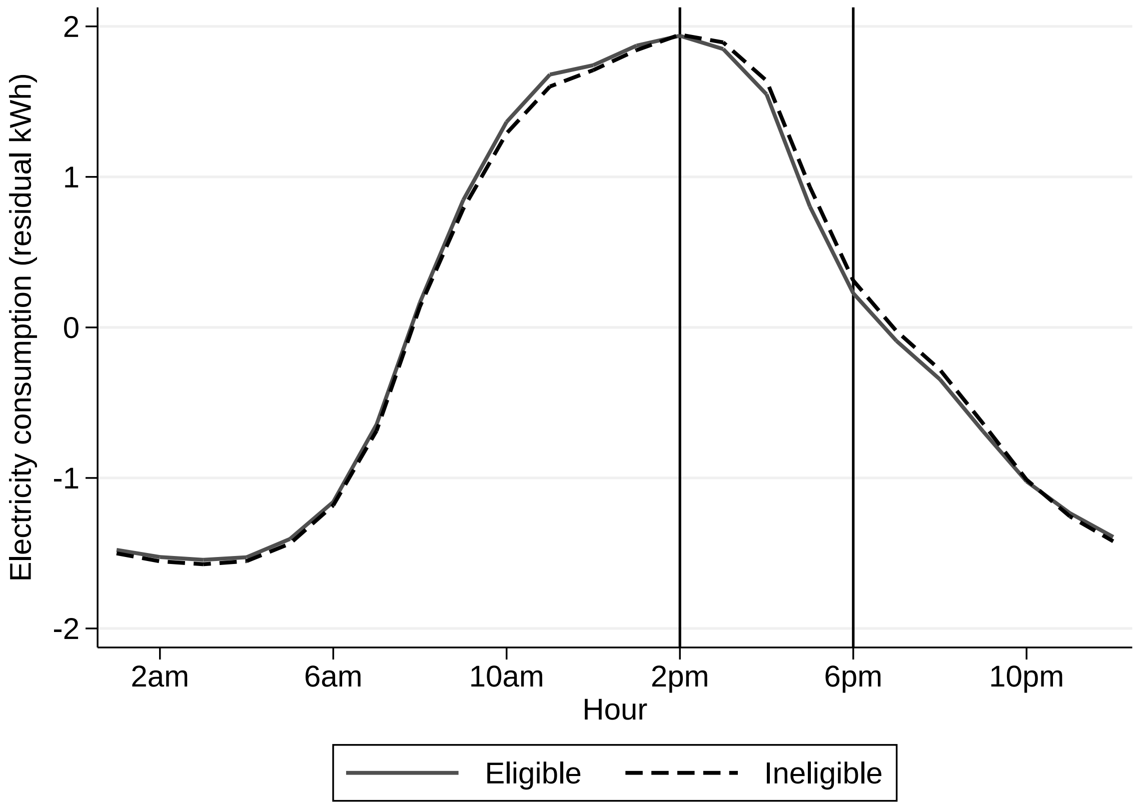
Note. - This figure shows the timeline of peak pricing implementation. I classify establishments as eligible for peak pricing in 2015 if their high-frequency metering data began before September 1, 2011. Enrollment in opt-out peak pricing starts November of 2014 for the summer of 2015. Final treatment status is determined by eligibility and technical requirements, which are described in Section 1.4.1.

Figure 1.3: The Effect of Eligibility on Peak Pricing Treatment Status



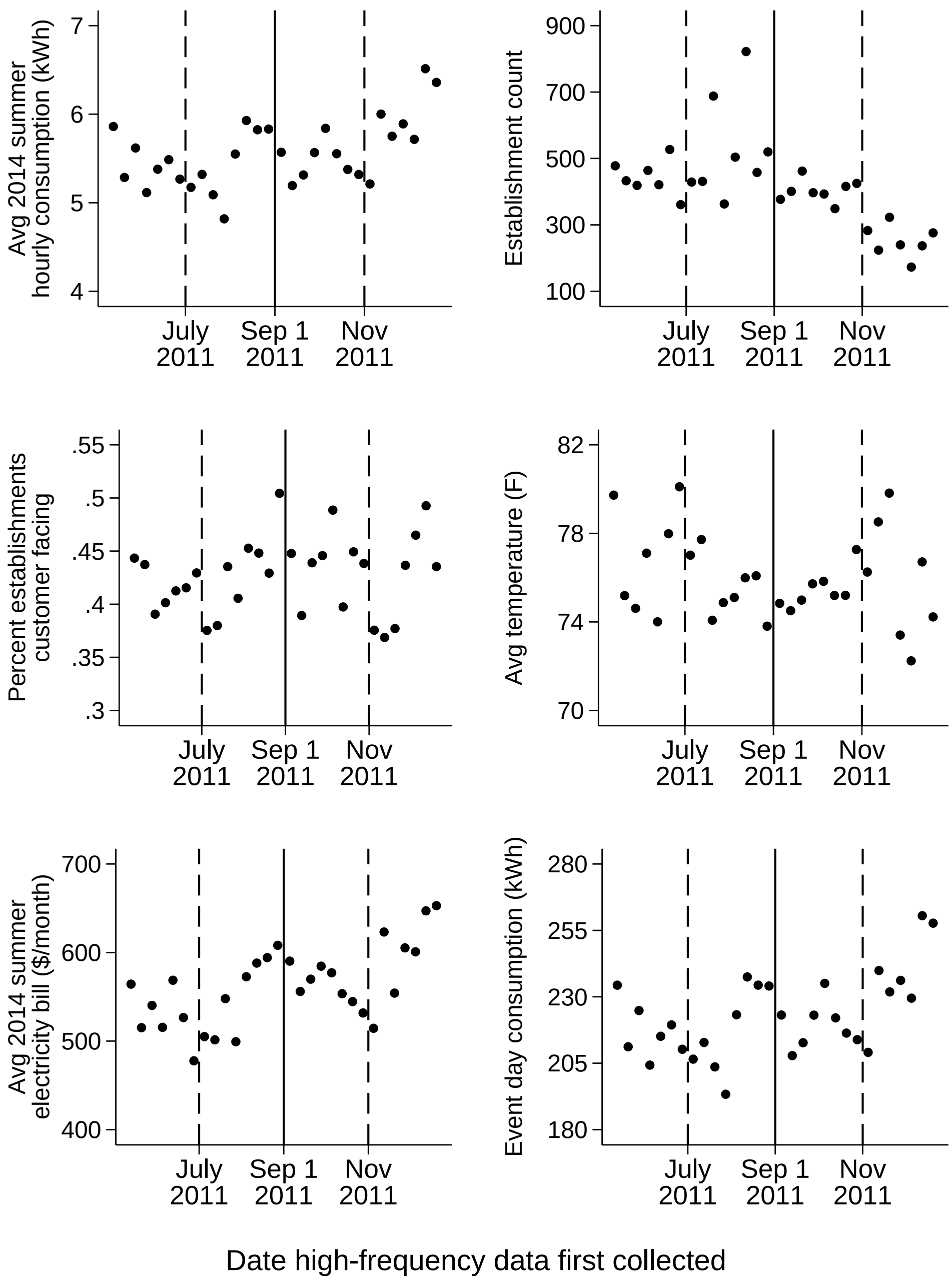
Note. - This figure shows the impact of peak pricing eligibility on treatment. Establishments are binned by the week their high-frequency data began. Establishments to the left of the September 1, 2011 threshold are peak pricing eligible. There are around 500 establishments per bin. The figure shows 27 weeks in each direction from the threshold to show the larger default patterns. The vertical dashed lines represent the eight-week bandwidth used in the main specification.

Figure 1.4: Pre-Period Electricity Consumption by Eligibility Group



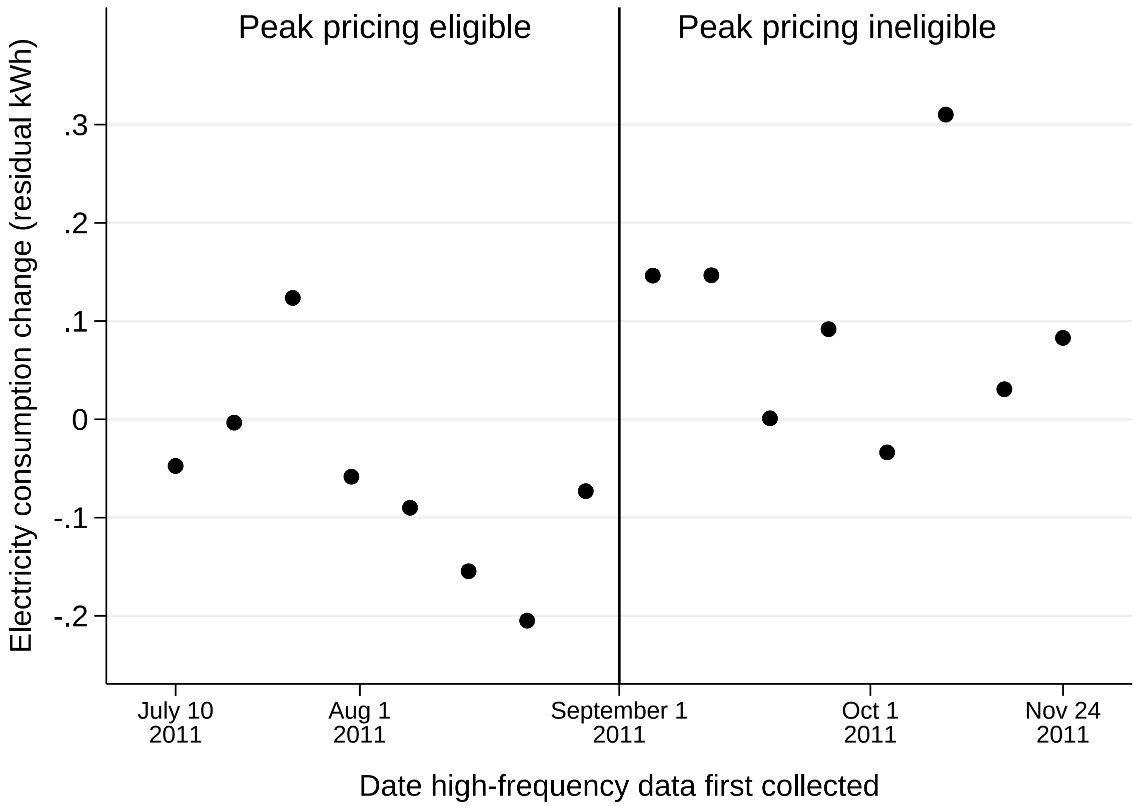
Note. - This figure shows the 2014 pre-period average hourly consumption for peak pricing eligible and ineligible establishments. Consumption is shown conditional on establishment fixed effects. I cannot statistically reject that the pre-period consumption is the same for both groups using hour-by-hour t-tests. The vertical lines signify the beginning (2:00 pm) and end (6:00 pm) of the peak event window.

Figure 1.5: Smoothness of Observable Characteristics through the September 1, 2011 Threshold



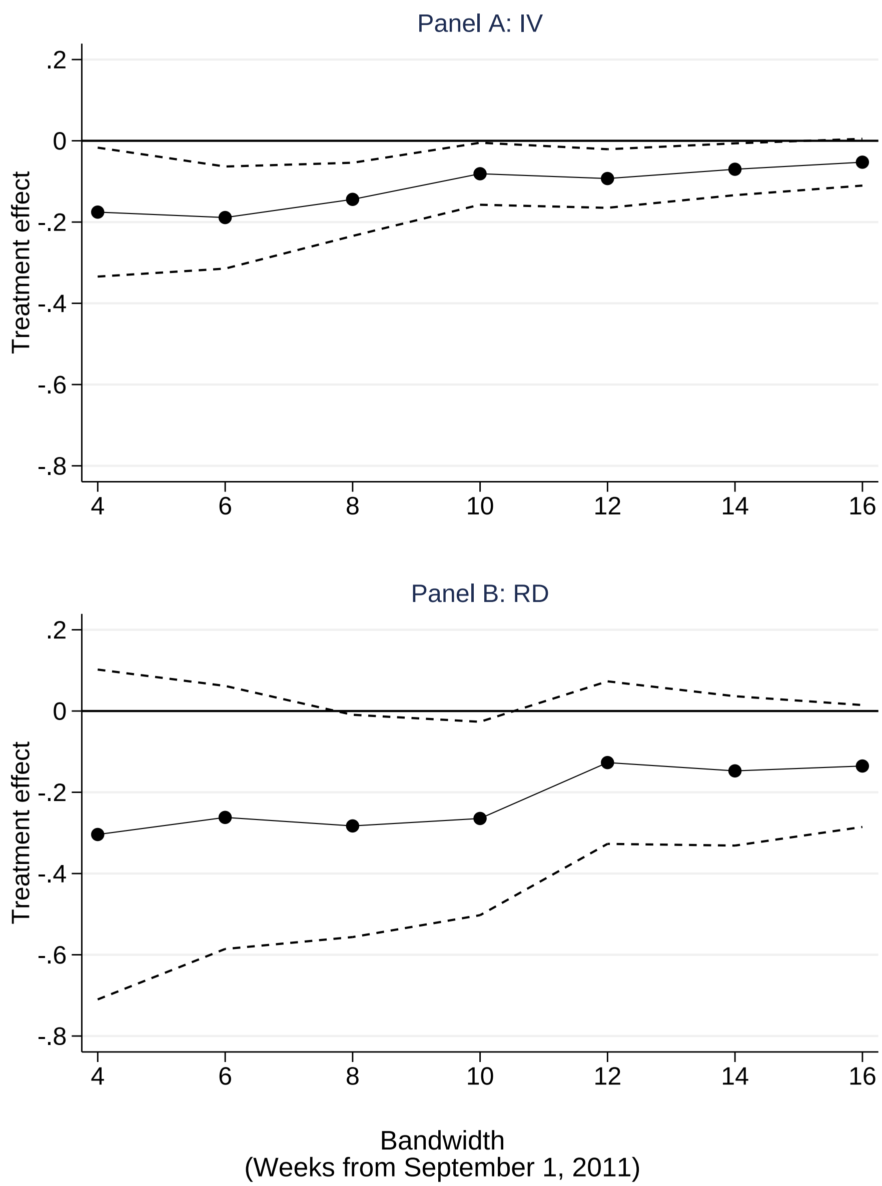
Note. - This figure shows trends in observable characteristics near the September 1, 2011 discontinuity, shown with the solid black vertical line. The vertical dashed lines indicate the eight-week bandwidth used in the main specifications.

Figure 1.6: The Impact of Peak Pricing Eligibility on Inland Establishment Peak Consumption (Reduced Form)



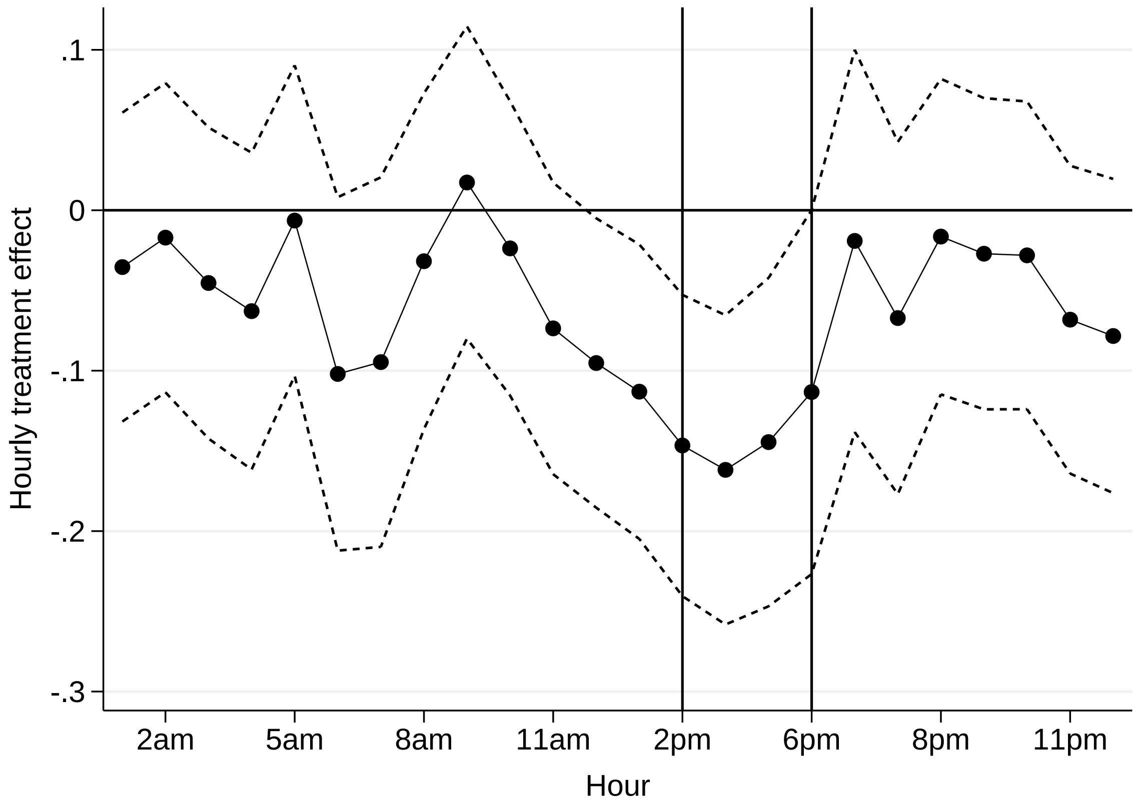
Note. - This figure shows the reduced form impact of peak pricing eligibility on consumption between 2:00 pm and 6:00 pm on event days. Each dot represents the difference between 2015 and 2014 peak consumption by bin, conditional on establishment and hour-of-sample fixed effects. The figure shows the reduced form impacts of the peak pricing policy, which is 6.2% and is significant at the 5% level. Establishments to the left of the September 1, 2011 cutoff are eligible for peak pricing and show a reduction in peak usage.

Figure 1.7: Treatment for Inland Establishments Effects Estimated at Varying Bandwidths



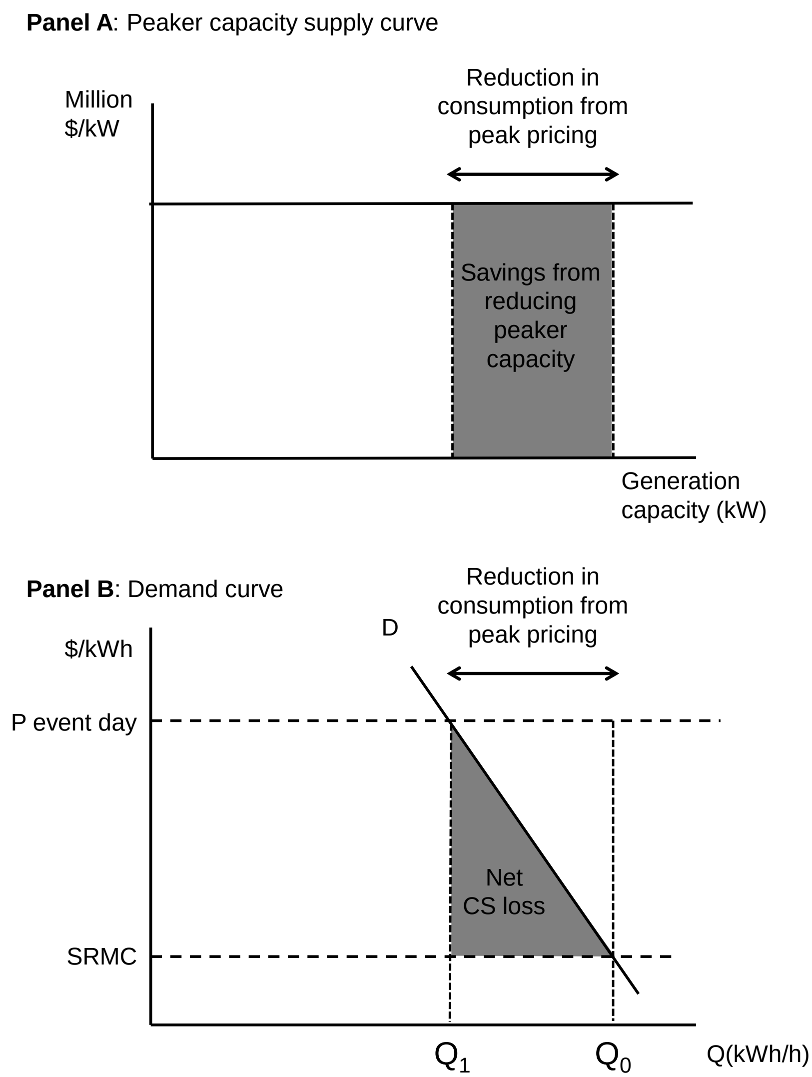
Note. - Each panel on this figure shows the coefficient from seven different regressions estimating the impacts of peak pricing on usage. Each dot represents an individual regression. Panel A shows the results from estimating Equation (1.1) for inland establishments using bandwidths between 4 and 16 weeks from the September 1, 2011 threshold. Panel B does the same using the RD specification from estimating Equation (2.4). The dotted lines are the 95% confidence interval. The estimate at eight weeks is the same as the results in Columns (5) and (6) in Table 1.3.

Figure 1.8: Effect of Peak Pricing on Inland Establishment Electricity Consumption



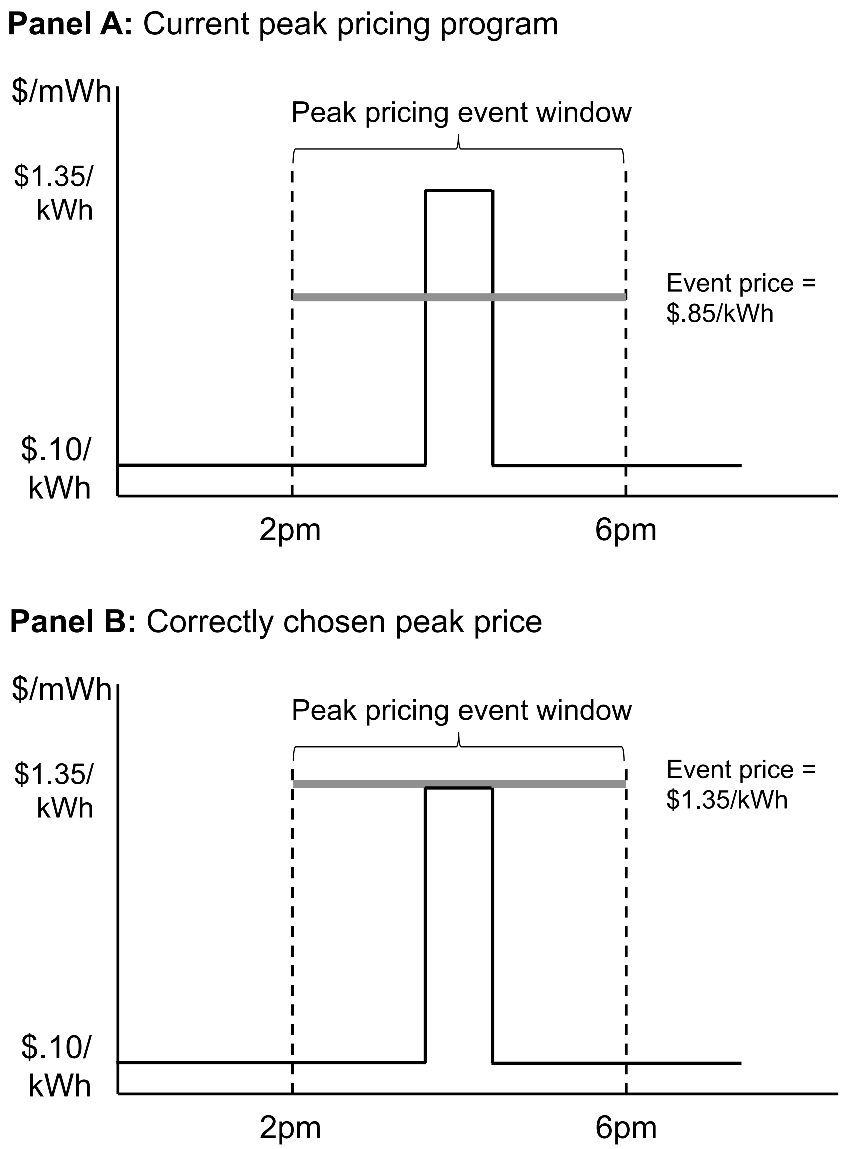
Note. - This figure shows the results of a regression estimating the hourly impacts of peak pricing on event days. Each dot corresponds to an hourly treatment effect comparing treated establishments to the control group. The dotted lines signify the 95% confidence interval. The vertical lines signify the beginning (2:00 pm) and end (6:00 pm) of the peak event window. The regression is estimated on inland establishments using the IV approach. The average impact between 2:00 pm and 6:00 pm reflects the coefficient in Column (5) of Table 1.3. The results show that establishments begin reducing their electricity usage in the hours before the event window starts. This pattern suggests that some establishments are adjusting their consumption over the whole event day and not just between 2:00 pm and 6:00 pm.

Figure 1.9: Benefits and Costs of Peak Pricing



Note: - This figure graphically shows the benefits and costs of peak pricing. Panel A shows the capacity supply curve for fossil generation. Reducing peak demand lowers the need for peaker power plants. I assume a constant cost of $1.2 million/MW to build a peaker plant, using California Energy Commission estimates to value the benefits. Using the IV estimate for inland establishments, I find an aggregate reduction of 118 MW, which translates into a reduction of $139 million in capacity costs. Panel B shows the hourly net consumer surplus (CS) loss from calling an event day. The horizontal axis is in kWh per hour (kWh/h), which is equivalent to kW. Short-run marginal production costs (SRMC) are $.102/kWh and reflect the fuel cost at marginal power plants during peak hours. Q_0 is the quantity of electricity consumed during an event hour without peak prices, and Q_1 is the quantity consumed during an event hour with peak prices. I assume a linear demand curve and find that each event day reduces welfare by $209,000. See Section 1.6.2 for a full discussion of the welfare impacts of peak pricing.

Figure 1.10: Comparison of Peak Pricing to the First-Best, Real-Time Price



Note. - This figure compares the first-best, real-time price (solid black line) to the peak pricing program on one super-peak event day. The real-time price takes on two values. The low value is \$.10/kWh and represents the marginal cost of generation at a low-cost power plant. I assume the price jumps to the high level of \$1.35/kWh on three super-peak event days per summer. This high price reflects both the marginal cost and capacity costs on the high demand days. The dashed vertical lines signify the beginning (2:00 pm) and end (6:00 pm) of the peak event window. Panel A shows the current peak pricing program, where the event price is set at \$.85/kWh between 2:00 pm and 6:00 pm, which I assume happens on 15 event days per year. Panel B shows the well-targeted version where the event price is set at \$1.35/kWh, which I assume happens on eight event days per year. The current program provides 43% of the first-best benefits while the well-targeted program achieves 80%.

Tables

Table 1.1: Characteristics of Establishments by Peak Pricing Eligibility Status

Variable	Ineligible	Eligible	P value of difference
Summer 2014 avg peak hourly consumption	5.17	5.19	.87
	(3.79)	(3.8)	
Summer 2014 max peak hourly consumption	9.92	10.00	.61
	(6.82)	(6.86)	
Summer 2014 event consumption	218	219	.80
	(165)	(166)	
Summer 2014 non-event consumption	12,412	12,280	.51
	(8,958)	(8,958)	
Summer 2014 electricity expenditure	\$563	\$557	.54
	(396)	(385)	
Percent of establishments customer facing	.44	.43	.73
	(.5)	(.5)	
Money saved if program run on 2014 usage	-\$10	-\$12	.16
	(58)	(57)	
Average peak hour temperature (F)	73.24	73.38	.41
	(7.55)	(6.96)	
Establishment count	3,188	4,190	

Notes. - This table shows the mean and standard deviation of the observable characteristics by peak pricing eligibility status for establishments within eight weeks of the September 1, 2011 threshold. Standard deviations are shown in parentheses. Customer-facing establishments are defined based on North American Industry Classification System codes, as discussed in Section 1.5.4. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 1.2: The Effect of Peak Pricing Eligibility on Enrollment (First Stage)

	All PG&E		Coastal		Inland	
	(1)	(2)	(3)	(4)	(5)	(6)
	FS IV	FS RD	FS IV	FS RD	FS IV	FS RD
Eligible × Post	0.2230***	0.0932**	0.1547***	0.0538	0.3654***	0.2258***
	(0.0064)	(0.0359)	(0.0068)	(0.0361)	(0.0129)	(0.0449)
Establishments	7,435	7,435	5,096	5,096	2,339	2,339
F statistic	406	24	174	15	268	45

Notes. - This table reports regression coefficients from six separate first-stage regressions. The dependent variable in all regressions is a binary indicator if an establishment is enrolled in the peak pricing program. Eligible × Post is an interaction of an establishment's eligibility for peak pricing and 2015. The coefficients show the impact of peak pricing eligibility on program enrollment. "FS IV" and "FS RD" correspond to the first stage of the IV and RD approaches estimated using Equations (1.2) and (2.3). All regressions control for temperature and include hour-of-sample fixed effects and establishment fixed effects. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 1.3: The Effect of Peak Pricing on Peak Electricity Consumption (2SLS results)

	All PG&E		Coastal		Inland	
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	RD	IV	RD	IV	RD
Peak pricing	−0.0695*	−0.2152	0.0084	−0.0584	−0.1441***	−0.2828**
	(0.0412)	(0.2102)	(0.0708)	(0.4227)	(0.0454)	(0.1379)
Establishments	7,435	7,435	5,096	5,096	2,339	2,339
Event day kWh usage	5.55	5.55	5.03	5.03	6.70	6.70
Average temperature	78	78	71	71	92	92

Notes. - This table reports regression coefficients from six separate 2SLS regressions. The dependent variable in all regressions is the log of establishment hourly kWh consumption. Peak pricing is an indicator for enrollment in peak pricing, for which I instrument with eligibility status. The coefficients show the impact of peak pricing on peak consumption between 2:00 pm and 6:00 pm. "IV" and "RD" correspond to the instrumental variables and regression discontinuity approaches estimated using Equations (1.1) and (2.4). For inland establishments, the IV coefficient corresponds to a 13.4% reduction in usage. All regressions control for temperature and include hour-of-sample fixed effects and establishment fixed effects. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 1.4: The Effect of Peak Pricing on Peak Electricity Consumption for Inland Establishments: Temperature Interaction

	Inland	
	(1)	(2)
	IV	RD
Peak pricing × Temperature (F)	−0.01120**	−0.03622*
	(0.00450)	(0.02168)
Peak pricing	0.04018	0.18102
	(0.07371)	(0.13676)
Temperature	0.01210***	0.01513***
	(0.00168)	(0.00294)
Temperature squared	−0.00013***	−0.00010
	(0.00004)	(0.00006)
Establishments	2,339	2,339
Event day kWh usage	6.73	6.73
Average temperature	92	92

Notes. - This table reports regression coefficients from two separate 2SLS regressions for inland establishments where treatment is interacted with temperature. The dependent variable in both regressions is the log of establishment hourly kWh consumption. "IV" and "RD" correspond to the instrumental variables and regression discontinuity approaches estimated using Equations (1.1) and (2.4). Peak pricing × Temperature (F) is the interaction between the treatment variable and hourly establishment temperature. Temperature has been re-centered at 75 degrees for scaling purposes. The coefficients show that peak pricing impacts are larger on hotter inland event days. In the IV specification, the peak pricing impacts become positive around 79 degrees, which is lower than the temperature for all inland event days. Both regressions include hour-of-sample fixed effects and establishment fixed effects. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 1.5: The Effect of Peak Pricing for Coastal Establishments on Hot Event Days: Alternate Control Day Approach

	Coastal	
	(1)	(2)
	IV	RD
Peak pricing	-0.0783^{**}	-0.0824^{**}
	(0.0361)	(0.0409)
Establishments	2,991	2,991
Event day kWh usage	6.96	6.96
Average temperature	76	76

Notes. - This table reports regression coefficients from two separate 2SLS regressions for coastal establishments. The dependent variable in both regressions is the log of establishment hourly kWh consumption. Peak pricing is an indicator for enrollment in peak pricing, for which I instrument with eligibility status. To identify the impacts of peak pricing for coastal customers on the hottest event days, I use hot 2015 non-event days instead of 2014 event days as controls. This approach is used because 2014 was a relatively cool summer on the coast, making it a bad control group with which to identify coastal program impacts on hot event days. See Section 1.5.5 for more details on this approach. The coefficients show the impact of peak pricing on coastal establishment consumption on hot event days between 2:00 pm and 6:00 pm. "IV" and "RD" correspond to the instrumental variables and regression discontinuity approaches estimated using Equations (1.1) and (2.4). Both regressions control for temperature and include hour-of-sample fixed effects and establishment fixed effects. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 1.6: The Effect of Peak Pricing on Peak Electricity Consumption for Inland Establishments: Industry Classification

	All PG&E		Coastal		Inland	
	(1)	(2)	(3)	(4)	(5)	(6)
	Customer facing	Non-cust facing	Customer facing	Non-cust facing	Customer facing	Non-cust facing
Peak pricing	0.0351	-0.1261**	0.0849	-0.0477	-0.0160	-0.1967***
	(0.0555)	(0.0586)	(0.0981)	(0.1021)	(0.0518)	(0.0659)
Establishments	2,889	3,745	2,133	2,468	756	1,277
Event day kWh	6.34	5.15	5.66	4.61	8.25	6.19
Average temp	76	78	71	72	92	92

Notes. - This table reports regression coefficients from six separate 2SLS regressions broken down by industry. The dependent variable in all regressions is the log of establishment hourly kWh consumption. Peak pricing is an indicator of enrollment in peak pricing, for which I instrument with eligibility status. The coefficients show the impact of peak pricing on peak consumption between 2:00 pm and 6:00 pm. All regressions use the instrumental variables approach estimated using Equation (1.1). Establishments are classified as customer- facing or non-customer-facing by their industry classification code, as described in Section 1.5.4. For inland establishments, the non-customer-facing coefficient corresponds to a 17.9% reduction in usage. All regressions control for temperature and include hour-of-sample fixed effects and establishment fixed effects. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 1.7: Total Welfare Benefits of Peak Pricing

Scenario	(1) PG&E welfare benefits	(2) IOU welfare benefits	(3) California welfare benefits	(4) National welfare benefits
Small C&I customers	\$154	\$394	\$573	\$16,616
All C&I customers	\$1,320	\$1,940	\$2,820	\$81,754

Notes. - This table shows the welfare benefits (in millions of 2016 dollars) of the peak pricing program over a 30 year horizon under a number of scenarios. Welfare benefits are calculated using aggregate peak load reduction values informed by empirical estimates. Benefits are primarily due to the reduction in generation capacity necessary to meet peak demand. Costs include the net consumer surplus loss from higher prices on event days. The top left entry shows the estimated savings for the current program in the PG&E service territory. Moving to the right scales this welfare impact for larger regions of the country. The IOU column corresponds to the three major Investor Owned Utilities in California, all of which will implement peak pricing over the next five years. The bottom row extends the peak pricing program to all C&I customers and assumes the same percent reductions for all C&I customers, with adjustments for establishments opting out. Moving both down and to the right, the estimates require more out-of-sample assumptions.

Table 1.8: Welfare Impacts for 2015 Event Days

Event day	PG&E max load	Annual capacity cost savings (discounted)	Annual net consumer surplus loss (discounted)	NWS day ahead max temperature forecast	Trigger temperature
8/17/2015	19,451	$10,000,000	-$209,000	101	96
6/30/2015	19,320	$0	-$209,000	101	96
7/29/2015	19,248	$0	-$209,000	104	98
8/28/2015	19,233	$0	-$209,000	96	96
9/10/2015	19,230	$0	-$209,000	104	98
9/9/2015	19,017	$0	-$209,000	102	98
7/28/2015	18,403	$0	-$209,000	101	98
8/27/2015	18,328	$0	-$209,000	97	96
6/25/2015	18,114	$0	-$209,000	103	96
9/11/2015	18,019	$0	-$209,000	101	98
6/26/2015	17,950	$0	-$209,000	100	96
7/30/2015	17,750	$0	-$209,000	100	98
7/1/2015	17,734	$0	-$209,000	100	98
8/18/2015	17,372	$0	-$209,000	96	96
6/12/2015	17,275	$0	-$209,000	99	96

Note. - This table shows the two main welfare impacts of the 2015 event days. The annual capacity cost savings shows the benefits of reducing peak load. Annual capacity cost savings includes both the plant construction and operating costs, amortized over the assumed 30 year power plant life. There are non-zero savings numbers only for the super-peak event days of each summer. In 2015, only the highest load day was super-peak. The annual net consumer surplus loss shows the negative welfare consequences of charging higher prices during event hours and is displayed in the same units as capacity cost savings. The values are the same for all event days because the estimate is based on the average impact of peak pricing. NWS day-ahead maximum temperature forecast is the day-ahead temperature used by PG&E to call event days. It is based on the average of five National Weather Service weather stations. When the day-ahead maximum forecast equals or exceeds the trigger temperature, an event day is called.

Table 1.9: Welfare Impacts of Peak Pricing Under Alternate Scenarios

Scenario	(1) $.85/kWh peak (current price)	(2) $1.35/kWh peak (large C&I peak price)	(3) $1.85/kWh peak (high price)
15 days/summer	$154	$204	$200
101 degree trigger (8 days)	$184	$288	$349
Super-peak days (3 days)	$205	$349	$455

Note. - This is a table that shows the welfare benefits (in millions of 2016 dollars) of the peak pricing program under different program design scenarios. Column (1) shows outcomes under the current $.85/kWh peak price. Column (2) shows the estimated outcomes if the peak price were set at $1.35, which is the level of large commercial and industrial customers and is based on a PG&E valuation of capacity at peak. Column (3) shows the impacts if the price was set at $1.85/kWh. The first row reflects the current 15 event days per summer and the entry in the top left shows the welfare impacts estimated for the current program. The middle row reflects the proposed alternate 101 degree trigger for event days, and the bottom row shows the hypothetical scenario when only the three super-peak event days each year could be called. The welfare calculations assume that peak wholesale prices are greater than or equal to the peak price in each column.

Table 1.10: Welfare Impacts of Peak Pricing Compared to First-Best, Real-Time Price

Event days called per summer	(1) $.85/kWh peak price (peak price < RTP)	(2) $1.35/kWh peak price (peak price = RTP)	(3) $1.85/kWh peak price (peak price > RTP)
8 event days (well targeted)	49%	**80%**	57%
15 event days (current)	**43%**	62%	19%

Note. - This table compares the peak pricing program to the first-best, real-time price across a number of scenarios. The percent values reflect the percent of the welfare benefits the peak pricing scenario can achieve compared to the first-best alternative. For this table, the optimal peak price is set at $1.35/kWh for five minutes on three super-peak days per summer. Column (1) reflects the current program, where peak prices are set at $.85/kWh, which is below the optimal level. Column (3) shows the impacts when prices are set above this level. The top row reflects the outcomes when eight event days are called per year. The bottom row shows the results for the current program, in which I assume 15 event days are used each summer. The current program achieves 43% of the first-best policy, while the well-targeted program could achieve 80% of the benefits.

Chapter 2

Energy Efficiency Retrofit Heterogeneity and Program Outcomes: Evidence from the California Energy Savings Assistance Program

2.1 Introduction

Energy efficiency is a large part of the U.S. strategy in addressing climate change. Some view it as a win-win policy, where energy savings and reduced carbon emissions can be easily achieved. Many of these projections are done using engineering models that estimate the benefits from energy efficiency to be large and cost-effective (McKinsey & Company 2009). The analyses present energy efficiency as an attractive and efficient policy tool that can be used to effectively address climate change in the absence a carbon policy.

The growing importance of energy efficiency is reflected in the increasing funding it has received. Gas and electric energy efficiency budgets totaled 8.7 billion dollars for 2014, up from 3.7 billion in 2008. California is a major contributor to the total, with 1.7 billion allocated to energy efficiency programs in 2014 (CEE 2014).

Energy efficiency spending is expected to grow as the Clean Power Plan (CPP), the major piece of national environmental policy aimed at reducing U.S. carbon emissions, includes energy efficiency as a carbon mitigations strategy. The CPP fact sheet highlights that:

> Demand-side EE is an important, proven strategy that states and utilities are already widely using, and that can substantially and cost-effectively lower CO2 emissions from the power sector. EPA anticipates that, thanks to their low costs and large potential in every state and region, demand-side EE programs will be a significant component of state compliance plans under the Clean Power Plan. The CPP's flexible compliance options allow states to fully deploy EE to help meet their state goals (EPA 2015).

The EPA designed the CPP with energy efficiency as one of the main low-cost mitigation mechanisms available to states. One challenge in implementing energy efficiency as a carbon mitigation policy is how to credit states for their energy efficiency programs. The current CPP program rules will credit energy efficiency by calculating program specific savings and counting them against the hypothetical energy demand in the state if the energy efficiency program was not in effect. This calculation is based on the measurement of efficiency program savings, making it imperative that states are credited the appropriate amount for their energy efficiency programs.

The Clean Power Plan also provides second level of credits for energy efficiency implemented in low-income communities named the Clean Energy Incentive Program (CEIP). This optional program grants double credits for demand side management in low-income communities conducted during 2020 or 2021.[1] The targeted nature of the CEIP credits shows that low-income energy efficiency is a national policy priority and will continue to play a part in climate policy going forward.

One such program that provides energy efficiency retrofits to low-income households is the Energy Savings Assistance (ESA) program in California. The program has been in existence in various forms since the 1980s, and currently provides no-cost energy efficiency retrofits to around 300,000 low-income California households per year. The program is administered through the three major investor owned utilities and provides energy efficiency upgrades such as CFL lightbulbs, weatherstripping, water conservation technologies and appliance replacements to qualifying households. This paper focuses on estimating the energy savings from the ESA program run by San Diego Gas and Electric (SDG&E) and Southern California Gas (SoCalGas) between 2007 and 2014.

In order to credibly identify energy savings from the ESA program, the empirical analysis exploits quasi-random variation in program uptake. The ESA program, like many other low-income weatherization programs, requires income certification to demonstrate that a customer qualifies. Previous work on a similar program

[1] These credits are granted for early action projects, since the mandatory goals of the Clean Power Plan do not start until 2022. Renewable energy construction is also included as a category that can receive CEIP credits.

has shown high non-monetary costs associated with this process that likely reduce program participation (Fowlie, Greenstone, and C. Wolfram 2015a). To reduce the burden associated with signing up for ESA, SDG&E modified its program to allow customers that lived in areas with sufficiently low incomes to self-certify that they met the income requirements without providing documentation. The self-certification increases program participation in the areas where it was implemented. Importantly, the assignment to self-certification was done at a very small geographic level, leading to similar households in geographic proximity being assigned to different certification regimes. The increase in ESA uptake from self-certification allows for program evaluation across this quasi-random change in program participation.

Using the variation in self-certification requirements as an instrument, I find the ESA program does not save a significant amount of energy for households close to the self-certification threshold. I am able to rule out up to a 2.5 percent decrease in kWh usage and a .7 percent decrease in therm usage for these customers.[2] These low overall program savings results are similar to the other main analysis conducted on low-income energy retrofits by Fowlie, Greenstone, and C. Wolfram (2015b).

One defining feature of residential energy efficiency programs is that retrofits must be adapted to each individual housing unit. Some households have a high potential for efficiency measures while others may have little or no possibility for cost-effective upgrades. This heterogeneity in housing stock creates a challenge for policymakers trying to design and implement a cost-effective policy. The ESA program exhibits this characteristic variation in upgrade potential, with 19 percent of households receiving a large appliance or insulation upgrade, and the remainder only eligible for smaller upgrades.

To better understand the heterogeneity in housing upgrade potential and its effects on outcomes, the analysis turns to estimating measure-specific savings. These estimates are identified by exploiting a source of quasi-experimental variation in program implementation. The ESA program has a strict set of guidelines that govern which households are eligible for specific upgrades. Some of the guidelines have discontinuous eligibility thresholds that allow for identification of measure specific effects. Refrigerator replacements are given to households at no-cost if their existing unit was manufactured in 1992 or earlier. This cutoff allows for regression discontinuity analysis of energy savings across this refrigerator age threshold by comparing the energy usage of households with refrigerators that qualified for replacements to those that just missed the cutoff. Similarly, high efficiency washing machines are provided to households with more than 4 members that satisfy a number of criteria. Importantly

[2] Using a more restrictive set of assumptions and sample trimming the analysis rejects a 5.7 percent decrease in kWh usage and an 8.6 percent decrease in therm usage.

for this analysis, appliance characteristics are collected for all program participants even if they do not qualify for appliance replacement.

The results of this appliance specific analysis finds that the 1 in 10 households who receive a replacement refrigerator experience a significant electricity savings of 36 kWh/month. High efficiency washing machines are also installed in around 3.4 percent of households, and similarly show significant therm savings as a result. When considering the full social benefits of the upgrade, refrigerator replacements are found to deliver cost-effective savings for high usage recipients. The ability to conduct these measure specific analyses shows that while the ESA program may not provide large savings for most customers, some customers do see statistically and economically significant savings as a result of the program. Furthermore, it shows how heterogeneity in upgrade potential can dramatically affect overall program savings. By primarily retrofitting households that were only eligible for small upgrades, the ESA program yielded average outcomes that were not cost-effective.

This is the first paper to estimate both program level effects and measure specific savings. Past analyses have not had variation in both program uptake and measure installation that allows for this type of identification. The ability to estimate both effects for the same large-scale program gives this paper a high degree of policy relevance. Understanding the impact of heterogeneity in retrofit upgrade potential on overall program savings offers policymakers ways to improve efficiency programs going forward.

The ESA program is an important policy since it has the potential to serve as a model program for other states implementing low-income energy efficiency retrofits. Currently, the federal Weatherization Assistance Program is the main provider of low-income no-cost weatherization nationally, but it relies on federal funding and community agencies for implementation. The ESA program is the largest state funded and administered low-income program in the country. As states create policies to comply with the Clean Power Plan and take advantage of the CEIP credit, it is important that the existing programs provide good examples of well designed and evaluated policies that provide cost-effective savings. Furthermore, the Clean Power Plan requires measurement of program effects to appropriately credit implementing states. Without careful evaluation of energy efficiency programs, the carbon reductions promised by the Clean Power plan will not deliver the intended savings.

, the date of installation, and a number of demographic characteristics for enrolling households including number of occupants, household size, household income and language spoken. This includes 120,244 individual household upgrades between 2007 and the end of 2012. The ESA program collected detailed information on the condition of many existing appliances in all households that participated in the program. Data on refrigerator age, model number and serial number were collected for all households regardless of if they qualified for a replacement.

Similar data was used for SoCalGas to estimate the effect of high efficiency washing machines. Billing data on 23,415 customers from 2012-2014 was used in this analysis. These data were merged with ESA upgrade data and supplemental information that specified if a household had a washing machine and its condition.

Both the billing data and ESA program data was provided through an internship with the California Public Utilities commission.

2.2 IV empirical design and analysis

2.2.1 Empirical setting

Starting in 2007, SDG&E began to use an income self-certification system across their service territory in an effort to increase program participation. If a neighborhood was determined to be self-certification eligible, households were not required to provide income verification when they enrolled in the ESA program. The idea behind this strategy is to reduce barriers to entry for customers that were likely to satisfy the income criteria based on where they lived. To implement this program, SDG&E classified zip+4 codes using a product which they purchased from the Nielson Company called PRIZM codes. This product classifies each zip+4 into one of 66 geo-demographic characteristics. The codes roughly correspond to income, going from 1 (richest) to 66 (poorest), but they also contain other demographic information such as race, employment status, age group, household tenure, household composition and education level. These other demographic characteristics generally are not ordered based on the 1 to 66 ranking.

PRIZM codes are calculated using a proprietary formula and data sources which are not available to the public. An important feature of the PRIZM codes purchased by SDG&E is that the level of designation, the zip+4, is a small geographic unit usually consisting of on average 8 households. This results in large spatial variation in PRIZM code assignment across even small geographic areas.

SDG&E, with the permission of the PUC, used PRIZM codes to classify customers as eligible for self-certification. Households living in zip+4s designated with the numbers 46 through 49 and 52 through 66 do not have to provide income documentation when they enroll in the program. Households that do not live in these PRIZM codes can still qualify for the ESA program using the standard income verification process. For example, ESA participants can live in PRIZM code 1 neighborhoods, but they must provide documentation to show they meet the income qualifications for the program.

2.2.2 Self-certification as an instrument for program enrollment

There are a number of challenges in evaluating the effects of an energy efficiency retrofit program. First, households self-select into the ESA program. Choosing to enroll in a time consuming intensive program like ESA likely makes them different than the average low-income California utility customer. Second, the opt-in nature of the program makes the timing of enrollment potentially correlated with observed and unobserved factors. Both of these problems make it difficult to evaluate the ESA program using a standard billing regression framework typically used by past 3rd party evaluators (Evergreen Economics 2013).

The self-certification mechanism in the ESA program provides an opportunity to evaluate program outcomes in a causal, unbiased manner. Households that are eligible for self-certification have an easier time enrolling in the ESA program compared to households that must provide income verification. This makes them more likely to enroll in the program than similar households that cannot self-certify based on where they live. This variation in program uptake is then used to identify the program outcomes.

For this identification strategy to provide unbiased estimates, assignment to self-certification zip+4s must be as good as random conditional on observables and fixed effects. If this is not the case, the results could be biased as the self-certification areas could systematically differ from the non-self-certification areas. Two strategies are used to address this concern. First, households that live in lower numbered PRIZM code areas far below the self-certification cutoff are dropped from the analysis. This removes customers that may qualify for the ESA program, but live in higher income zip+4s. The main specification limits PRIZM codes to 18-66.

The second technique exploits the geographic nature of the self-certification designation. A unique feature of the data used for this analysis is that the billing data contains the latitude and longitude of all customers in SDG&E's service territory. Using this precise location, I construct a distance to household of the opposite certification regime that can be used to select a sample of similar households. The distance algorithm works by taking all of the self-certification households and finding each of their closest location match to a non-self-certification household. Each household then has a "distance to other regime" value that can be used to trim the data. The process is repeated for non-self-certification households and their distance to self-certification households. By limiting the sample to households within a small geographic distance, the differences between the self-certification and non-self certification households is reduced.

To better understand the geographic nature of the data, figure 2.4 maps all CARE eligible households in SDG&E's service territory.[3] The blue area represent zip+4s that are not eligible for self-certification due to their PRIZM code designation. It is important to note that these colors are based on the zip+4 designation, and many customers living in ineligible self-certification areas may still qualify for the program through traditional certification methods. The orange areas are eligible for self-certification, and all households living in those areas qualify for the ESA program.

To reduce the significant geographic variability of zip+4s show in 2.4, households within 75 meters of a household in the opposite regime are shown in figure 2.5. A zoomed in picture of the more concentrated areas of San Diego with the 75m distance can be see in figure 2.6. Both of these figures show that the ineligible and eligible households are close to each-other, and that the 75 meter limited sample reduces the geographic variability of households in the sample.

One concern is that the large concentrated orange eligible zip+4s in figure 2.6 might be somewhat different from the blue ineligible areas on their periphery. This turns out to not be the case, since the nature of the zip+4 assignment to self-certification creates a large amount of variation throughout the sample. Figure 2.7 shows the same zip+4s as figure 2.6 color coded by their distance to the opposite regime. The blue dots indicate that the large concentrations of zip+4s are generally within 20-30 meters of the opposite regime.

The variation in self-certification status can be further seen by zooming in on a small area and showing individual houses. Figure 2.8 does this for a small patch of SDG&E's territory. As before, the blue dots correspond to households that are ineligible for self-certification and the orange dots are eligible. This figure shows the sample of CARE customers before trimming for any distance requirements. Even at the non-trimmed level, it show significant mixing between the self-certification eligible and ineligible households at the block by block level. Figure 2.9 takes figure 2.8 and keeps units that are within 75 meters of the opposite regime. The result is that the larger continuous blocks of orange and blue in the previous figure are dropped, and all that remains are ESA eligible households that are close together of opposite certification regimes. It is also possible to see how small of a level zip+4s and eligibility status vary. Some blocks have houses on one side that are ESA eligible while houses on the other side are not.

[3]SDG&E's territory stretches further east, but the majority of the population lives in the area shown in figure 2.4

2.2.3 Empirical strategy

In the IV approach, identification of ESA program savings comes from using eligibility for income self-certification as an instrument for program participation. The first stage uses OLS to estimate:

$$1\{ESA\}_{izmt} = 1\{SelfCertify\}_{izmt} + \zeta_{mtz} + \gamma_{im} + \eta_{izmt} \tag{2.1}$$

where $1\{ESA\}_{izmt}$ is an indicator for if the customer i received ESA treatment in zip code z in month m and year t. $1\{SelfCertify\}_{izmt}$ is an indicator for if a customer lived in a self-certification zip+4, and switches to 1 for households in eligible zip+4s when the self-certification policy starts in 2007. To control of weather, ζ_{mtz} is included as a zip code by month-of-sample fixed effect. Household-by-month fixed effects, γ_{im}, are used to control for time invariant differences in household energy consumption across months of the year, and η_{izmt} is the error term.

The first stage estimated in equation (2.1) gives an estimate for the effect of self-certification on ESA program enrollment once self-certification was started in 2007. Using the instrument above, I estimate the second stage with OLS as:

$$Y_{izmt} = \beta_1\{ESA\}_{izmt} + \zeta_{mtz} + \gamma_{im} + \epsilon_{izmt} \tag{2.2}$$

where Y represents the energy source (kWh or therms) being measured. $\{\widehat{ESA}\}_{izmt}$ from the first stage is substituted for $\{ESA\}_{izmt}$ to obtain the IV estimate of β_1. The estimate of this coefficient is the local average treatment effect (LATE) of the ESA program close to the self-certification threshold. Said differently, β_1 estimates the effect of the ESA program on the subpopulation of ESA households where the ability to self-certify brought them into the program. The estimate of β_1 should not be be interpreted as the average savings generated by the ESA program. Instead, it should be thought of as a savings estimate for a large subset of the ESA eligible population.

The identification of an unbiased estimate of β_1 depends on the validity of self-certification as an instrument for program enrollment. Importantly, the self-certification instrument cannot be correlated with the outcome variable (energy consumption) through any other mechanism than ESA program uptake. This cannot be empirically tested, but the sample trimming based on PRIZM codes and geographic distance limits the possibility that neighborhood characteristics or any other factor could be correlated with energy usage.

To better understand the effect of limiting the distance used in this analysis, figure 2.10 plots the density of pre-period kWh consumption for the 1000 meter (panel A) and 75 meter sample (panel B) used in the analysis. The overlap between the two improves when the distance is cut to 75 meters, but the differences do not go away entirely. Panel C of figure 2.10 shows the residual density after removing account-by-month and zip-by-month of sample fixed effects to better reflect the specifications run in equations (2.1) and (2.2). The significant overlap in pre-period kWh consumption residuals shows that the main specifications used in this analysis remove almost all of the differences between the two groups. Similar patterns and results can be seen when looking at the pre-period therm usage in figure (2.11).

2.2.4 Self-certification IV results

Table 2.3 presents the results of the first stage regression shown in equation (2.1). All results in this section limit the sample to households living in zip+4s with prizm code 18-66. The first column shows the results from the regression when households are limited to 75 meters of households of the opposite regime. The results show that the self-certification instrument is a significant driver of program uptake. Households that live in self-certification areas have a 12.8 percentage point higher program participation than those that cannot self-certify over the sample window.

Column (2) shows the results for the same regression run at 1000 meters. The coefficient is somewhat larger at 15.6 percent and still shows the same strong ESA program uptake.[4] These outcomes show that the self-certification instrument is correlated with program uptake, a necessary condition to it being used as an instrument.

There are two likely factors that are driving this large increase in participation among self-certification eligible households. The first is that the lower documentation requirements remove a significant barrier to entry for program enrollment. On a ride-along I participated in, a scheduled enrollment contractor visits was canceled due to the customer not having the income verification documentation. Program coordinators I spoke to confirmed that this was a common occurrence. The second factor that likely drives higher uptake is that contractors target households that they know are eligible for self-certification. Contractors are compensated based on the number of jobs they complete, and they have access to data that indicates which households can self-certify. This has lead to a more focused targeting of self-certification eligible customers by contractor outreach, leading to higher enrollment rates. The distinction between these two mechanisms is interesting from a policy implementation perspective, but it does not affect the validity of the identifications strategy.

[4]distances between 75 and 1000 meters also show the same result.

The IV results from the estimation of equation 2.2 with electricity as the outcome variable are shown in the first two columns of table 2.4. Column (1) presents the the IV estimate at 75 meters. The results show that the ESA program does not provide a significant amount of electricity savings. The results are reported in kWh reductions per month. The negative -.839 kWh/month savings is statistically indistinguishable from zero and can be used to reject a 5.7 percent reduction in energy usage. Column (2) shows the results at 1000 meters, and shows a similar results. The larger sample used in the 1000m sample allows for the rejection of a 2.5 percent reduction in electricity usage. Levels are used instead of logs due to the nature of the ESA program in giving incremental measures to households based on what they are eligible for. The results are robust to being run in logs.

The same regressions as column (1) is run for column (3) with therm usage as the dependent variable. The result show that the program does not provide significant therm savings. Column (4) presents the same regression at 1000 meters and finds the same zero savings results. Using these estimates, I can reject a savings of 8.6 and .7 percent for the 75m and 1000m results respectively. These results show that the ESA program on average does not reduce energy consumption.

It is worth noting that these results only estimate the energy impacts of the ESA program. There are numerous potential non-energy benefits that can result from an energy efficiency upgrade. These include increased comfort, improved air quality, increased safety and better all around living conditions. The natural gas safety checks that are an integral part of the ESA program frequently identify unsafe living conditions and work to improve them.[5] The non-energy benefits are potentially a source of value generated by the ESA program, but they are typically challenging to measure.

Another important consideration is that the effects estimated in these regressions do not reflect the average treatment effect of the ESA program. Instead, the results reflect a LATE for households that entered the ESA program as a consequence of income self-certification. It is possible that households which come into the program due to self-certification have lower savings on average. The instrument used, however, does affect a large portion of the SDG&E enrolling ESA population. Between 2007 and 2012, 54 percent of households self-certified their income when they enrolled in the program. This large percent of self-certifiers shows that the LATE estimates in this section reflect a large portion of the potential ESA population.

To test for the effects of heterogeneity in upgrades on program outcomes, an interaction term for households receiving refrigerators was included. These results can

[5]During a site visit I was on, the house did not pass the natural gas safety check. As a consequence, the local utility was contacted and brought out to shut off the dangerous gas leak.

be seen in columns (1) and (2) of table 2.5. The results show is that those households that did receive a new refrigerator had a total savings of around 36 kWh per month. Households that did not receive a fridge, however, still experienced no energy savings.

The refrigerator results are suggestive of heterogeneous upgrade savings, but there are potential problems with the identification of these coefficients. Households that receive replacement refrigerators might be different than those that did not qualify. For example households that did receive a new refrigerator might also be more likely to receive attic insulation and other significant savings measures. The refrigerator indicator could then be picking up these other factors, leading to a biased estimate of refrigerator savings. The next section presents a different methodology to estimate refrigerator savings estimates without the biases present in table 2.5.

2.3 Refrigerator savings estimates

Identifying heterogeneity in energy efficiency retrofit programs is typically not possible due to the non-random nature of the efficiency measures that are installed. Programs generally install all cost-effective retrofits are given to households that enroll. This creates the potential for bias, where the eligible retrofits could be correlated with energy usage and outcomes. Even with a well identified random assignment of treatment such as in Fowlie, Greenstone, and C. Wolfram (2015b), it is not possible to recover unbiased estimates of specific measures without exogenous variation in measure instillation.

2.3.1 Institutional details

The ESA program presents an opportunity to identify measure level effects by exploiting discontinuities in program rules that govern when certain appliances can be replaced. Households in the ESA program are eligible to have their refrigerator replaced for free if their current refrigerator was manufactured before 1993.[6] Importantly, the ESA program collects the refrigerator age for all households that enter the program, regardless of if they qualify for a new unit. This allows the identification of households who just qualified for a refrigerator replacement and a comparison group who just missed the eligibility cutoff.

The analysis is based off of 77,373 households that received ESA treatment between April 2007 and July 2014. The sample is further limited to 9,652 households that own refrigerators manufactured between 1990 and 1995, with 3,463 receiving replacements.

[6]The household must also have a properly grounded outlet to qualify. Starting in December 2012, refrigerators manufactured before 1999 were also accepted into the program.

Billing data is taken for these households between 2007 and 2014. It is important to note that the measure specific effects estimated using program rules are a local average treatment effect for participants in the ESA program close to the 1993 cutoff. The results will not be directly comparable to the to the estimated overall program effects since they are on a different population of customers.

2.3.2 Regression discontinuity specification

I use a RD design to estimate the effects of a replacement appliance. Appliance take-up is estimated in the first stage using the following equation:

$$1\{Refrigerator\}_{izmt} = \beta_1 1\{Eligible\}_{izmt} + \beta_2\, D_{izmt} + \beta_3\, D \cdot 1\{Eligible\}_{izmt} + \zeta_{mtz} + \eta_{izmt} \tag{2.3}$$

where $1\{Refrigerator\}_{izmt}$ is an indicator for if household i received a replacement refrigerator through the ESA program in zip code z in month m and year t. $1\{Eligible\}_{izmt}$ is an indicator for if a household has a refrigerator of model year 1992 or earlier, making them eligible for a replacement. The difference in refrigerator age from the 1993 threshold is normalized to zero and represented by D. Zip code-by-month fixed effects, ζ_{mtz}, control for weather and η_{izmt} is the error term.

The fuzzy RD regression is estimated using 2SLS using the following equation:

$$\begin{aligned} Y_{izmt} = {} & \beta_1\{\widehat{Refrigerator}\} \cdot P_{izmt} + \beta_2\, D \cdot P_{izmt} + \beta_3\, D \cdot 1\{Eligible\} \cdot P_{izmt} \\ & + \alpha_1 1\{Eligible\}_{izmt} + \alpha_2\, D_{izmt} + \alpha_3\, D \cdot 1\{Eligible\}_{izmt} + \zeta_{mtz} + \epsilon_{izmt} \end{aligned} \tag{2.4}$$

where $\widehat{Refrigerator}$ is the fitted value from the first stage regression. The regression is similar to the methodology used by G. W. Imbens and Lemieux (2008), except the standard equation is interacted with an indicator for the post period, represented as P. β_1 is the coefficient of interest, which represents the effects of having an ESA upgrade conducted (P) and receiving a refrigerator as part of the upgrade. Importantly, all households in the sample receive ESA treatment. The estimate β_1 identifies the incremental improvement in program savings for households that just qualified for an appliance replacement compared to similar households that did not.

Both equation 2.3 and 2.4 are estimated with local linear regression using a uniform kernel (G. W. Imbens and Lemieux 2008). Higher order polynomial specifications are avoided based on the suggestions of Gelman and G. Imbens (2014). The main specification uses a bandwidth of 3 refrigerator model years with robustness checks for larger and smaller bandwidths.

2.3.3 Validity of research design

This section uses graphical evidence to assess the validity of the regression discontinuity research design and to test for any manipulation across the eligibility threshold.

Discontinuity in manufacturing age

Figure 2.12 plots the refrigerator replacement rate by refrigerator model age. The dots represent the percent of households that received a new refrigerator binned by the age of their existing refrigerator. The red line represents the 1993 model year cutoff, where refrigerators manufactured before this year were eligible for replacement. The figure shows a discontinuous increase at the threshold indicating that the refrigerator age discontinuity does increase refrigerator uptake.

Validity of running variable

Manipulation of the running variable is a frequent concern when doing RD based analysis. The structure and implementation of the ESA program makes this unlikely since a number of different contractors are involved in the refrigerator evaluation and replacement process.[7] To show that refrigerator age manipulation is not a concern, figure 2.13 plots the distribution of refrigerator ages for all customers who received the ESA program. The red line shows the 1993 cutoff. If manipulation of refrigerator age was taking place, I would expect bunching on the left side of each of the red lines. This does not seem to be the case, indicating that there is no gaming of the running variable.

Another concern in RD analysis is that covariates change discontinuously at the eligibility threshold. These changes could lead to sorting or other behavior that might drive program outcomes at the discontinuity. To test for this, figure 2.14 plots all available covariates by refrigerator manufacture date. All of the panels in figure 2.14 have a smooth distribution across the threshold, showing that these variables are not changing discontinuously and potentially driving changes in energy usage.

2.3.4 Graphical RD results

Figure 2.15 plots energy savings from the ESA program against refrigerator age manufacture date. The red line indicates the 1993 cutoff, where households to the left

[7]Corruption is unlikely since this would require multiple contractors involved in the process to jointly manipulate program rules. The payoffs are likely not high enough to justify a customer or contractor engaging in this behavior.

qualify for the refrigerator replacement and households to the right do not. Each blue dot represents the average savings for households that go through the ESA program. This is calculated as the difference between the 6 month pre-upgrade consumption and the 6 month post-upgrade consumption. Crossing the red line greatly increases the savings from the ESA program as the probability of receiving a refrigerator increases.

Figure 2.16 presents an alternative version of the results. In this figure, the pre-period and post-period are displayed separately, showing how energy savings from the ESA program changed differentially on either side of the 1993 cutoff. The dots in figure 2.15 are calculated from as the difference between the blue and orange values in each refrigerator age bin. This alternate view illustrates a few additional points about the results. First, the pre-period consumption of the pre-1993 refrigerator owners is significantly higher than that of the post-1993 refrigerators. One potential explanation is that the federal minimum refrigerator efficiency standards, which set guidelines for all refrigerators made each year, were increased in 1993. The typical refrigerator manufactured in 1993 used 690 kWh/year compared to the 1990 standard which was set at 903 kWh/year. This means that the average 1993 refrigerator will consume much less energy than the average 1992 refrigerator (Koomey et al. 1999). The 1990 minimum efficiency standards can also be seen in this figure as the 1989 and earlier refrigerators consume even more energy than the 1990-1992 models. Once households replaced these older standard refrigerators, their consumption dropped to a similar level of other households with newer refrigerators.

The second noteworthy aspect of figure 2.16 are the small savings to the right of the 1993 cutoff. All households in this figure received ESA treatment, but the difference between pre-period and post-period is quite small for households that did not qualify for a refrigerator. These small differences between pre and post upgrade consumption for non-refrigerator eligible households should only be interpreted as an event study, but the results are similar in magnitude to the effects estimated in section 2.2.

2.3.5 Regression results

I now turn to the numerical results that correspond to the graphical evidence in the previous section. Table 2.6 shows the results of the estimation of the first stage shown in equation 2.3. Column (1) shows the effect of having a pre-1993 eligible refrigerator on refrigerator replacement without the inclusion of zip-by-month fixed effects. Column (2) shows the same equation with fixed effects included. The 72-73 percent increase in uptake corresponds to the jump in refrigerator uptake seen in figure 2.12.

The main IV results can be seen in table 2.4. The results are reported in kWh savings per month. Consistent with the graphical evidence, replacing a refrigerator generates significant savings. Column (1) shows the results of estimating equation (2.3) without fixed effects. Column (2) shows the same equation estimated with zip-by-month FE to control for weather. The fixed effects are not necessary if the RD assumption is valid, but their inclusion can help increase the precision of the estimate. Both show around a 36 kWh/month reduction in energy usage from the installation of a new refrigerator. These numbers are almost identical to the savings estimates from the previous section shown in table 2.5, where fridge savings are interacted with overall program savings. Column (3) replicates column (2) but is run in logs to report the percent reductions from the instillation of a new refrigerator.

When compared to the ex-ante estimates of refrigerator savings, the estimates found here are much lower. For example, the 2012 projection finds that a replacement refrigerator will save around 57 kWh/month. The savings numbers are not directly comparable, since the ESA ex-ante number is the average for all replaced fridges, while the results in this section reflect a LATE around the 1993 threshold. Nevertheless, it is striking how much larger the program projects refrigerator savings to be.

Other empirical studies of refrigerator replacement programs have found much smaller savings estimates. L. W. Davis, Fuchs, and Gertler (2014) examined the Mexican Cash for Coolers program, where participants were given subsidies to replace their old refrigerator with a new energy efficient model. The authors find that the program saved 11 kWh/month for the average program participant. There are a number of important differences that help explain some of the disparity between the L. W. Davis, Fuchs, and Gertler (2014) estimates and the savings found in my analysis. First, the ESA program replaced refrigerators that were at least 15 years old, which is much longer than the 10 year limit requirement in the Mexican program. Second, the ESA program provided all recipients with the same basic energy star model with no additional features. In the Cash for Coolers program, the customers had a choice of what model they purchased with their subsidy and the replacement fridges tended to have features that increased electricity consumption. As a consequence, the ESA program refrigerator estimates reflect savings with less behavioral responses to consider.

To put the 36 kWh/month reductions in in context, I calculate the bill savings experienced by households from a refrigerator replacement. California uses increasing block pricing, where the per-kWh charge goes up as the total monthly usage increase. In SDG&E, CARE households pay between 9.99 and 17.55 cents/kWh for their energy depending on what tier they are on and the season.[8] These prices translate to between a 3.63 and 6.38 dollars/month savings for the households who receive an upgraded refrigerator. These are significant savings for households, especially considering the no-cost nature of the upgrades.

To verify that the estimates shown are not sensitive to bandwidth choice, table 2.11 shows the regression from column (2) estimated with all bandwidths between 2 and 5 years. The results show that the estimates vary between 34.65 kWh of savings with a 4 year bandwidth to 41.00 kWh of savings with a 5 year bandwidth. This small shift in savings estimates shows that the bandwidth selection is not determining the RD regression outcomes. The 3 year bandwidth is chosen as the preferred specification because minimum efficiency standards were increased in 1990. Any bandwidth above 3 years would include refrigerators from that early efficiency era.

2.4 High efficiency washing machine savings estimates

2.4.1 Institutional details

High efficiency (HE) washing machines provide energy savings through two mechanisms. First, HE washing machines use less water than older units, and require less energy to heat that smaller amount of water. Second, HE washing machines generally do a better job at removing water from the washed clothes, reducing the amount of energy used by the dryer. To qualify for a free washing machine, ESA enrollees must meet a strict list of requirements. These include: the household must have 4 or more members, the customer must own the washing machine, the customer must have a dryer, the dwelling must have an individual water heater billed to the unit, both washing machine and dryer must be functioning and the old washing machine must be manufactured before 2004. HE washing machines were introduced into the ESA program starting in 2012.

Similar to the work on refrigerators, the analysis conducted on washers is limited to households that participate in the ESA program. The identification strategy will exploit the 4 occupant requirement for a HE washer replacement and compare outcomes to otherwise similar households with 3 occupants. In some ways this is similar to the refrigerator RD regressions run in the previous section, except the variation in take-up will be used in a different manner. In particular, the sample will

[8]These prices are for the first 3 months of 2012. The prices change slowly over time, but not in any significant amount during the sample window.

be limited to households that meet the list of qualifications that make them eligible for a new HE washing machine. The sample will also be limited to households with 3 and 4 occupants in order to minimize the difference between the eligible and ineligible ESA customers.

The analysis on HE washing machines will be conducted using SoCalGas data as opposed to the SDG&E data used in the earlier sections of the paper. This is done for a number of reasons. First, the supporting data provided by SoCalGas provides more insight on which households have the prerequisite conditions such as having an existing washer and dryer. Second, SoCalGas is a larger service territory with more customers, and as a consequence has installed more HE washers in their program. SDG&E installed 5,376 washers in the sample window and only 2,361 of those to households with 4 occupants. SoCalGas installed 49,587 HE washing machines, with 21,672 going to households with 4 occupants. This allows for a more precise identification of the effects of HE washers on energy usage. There is no reason to believe that the results from SoCalGas would vary significantly from those in SDG&E.

The dataset used includes the billing records for 23,415 households between 2012 and 2014. The analysis is conducted only on households with 3 or 4 members that meet the program criteria such as owning a washer and dryer. There are 6,969 3 occupant households in the sample that satisfy this criteria but will not be eligible for for an upgrade due to household size. Of the 16,446 households that have 4 occupants, 11,730 of them receive HE washing machines.

2.4.2 Empirical strategy

The identification of savings from HE washers comes from the variation in uptake between households with 3 and 4 members, with the later qualifying for an upgrade. Unlike the previous section where savings are identified with a regression discontinuity design, this estimation will use a difference-in-difference setup. The pre-period signifies the time before a household receives ESA treatment and will be interacted with if the household has 4 or more occupants to estimate the effects of HE washers on therm usage. A RD approach was not utilized in this context because there is a limited support on the number of occupants on either side of the cutoff, making inference at the threshold challenging.

The first stage uses OLS to estimate the effect of having 4 occupants on HE washer uptake using the specification:

(2.5)
$$1\{Washer\}_{izmt} = \alpha_1 1\{4\, occupants\} \cdot 1\{ESA\, upgrade\}_{izmt} + \alpha_2 1\{ESA\, upgrade\}_{izmt} + \zeta_{mtz} + \gamma_{im} + \eta_{izmt}.$$

$1\{Washer\}$ is an indicator of if household i received a replacement HE washer through the ESA program in zip code z in month m and year t. $1\{4\, occupants\}$ is an indicator for if a household has 4 occupants and is eligible for the replacement HE washer. When a household recieves an ESA upgrade, the $1\{ESA\, upgrade\}$ turns to 1. There is no standalone $1\{4\, occupants\}$ term since it is absorbed by the account-by-month fixed effects. Similar to previous estimation equations, zip code-by-month fixed effects, ζ_{mtz}, control for weather and η_{izmt} is the error term.

The second stage is estimated with 2SLS using the following equation:

$$Therm_{izmt} = \beta_1 1\{\widehat{Washer}\} \cdot 1\{ESA\, upgrade\}_{izmt} + \beta_2 1\{ESA\, upgrade\}_{izmt} + \zeta_{mtz} + \gamma_{im} + \epsilon_{izmt} \tag{2.6}$$

where $\widehat{Washer}$ is the fitted value from the first stage regression. β_1 is the coefficient of interest which is the interaction term between receiving an ESA upgrade and receiving a HE washer.

The causal identification of the therm savings from HE washers depends on the exclusion restriction holding, which requires that the only mechanism through which the $1\{4\, occupants\}$ instrument changes energy consumption is from the instillation of an HE washer. In this case, the violation of the exclusion restriction must be time varying because household-by-month fixed effects absorb the time-invariant differences between the 3 and 4 occupant groups. For example, if households with 4 occupants saved more energy from the non-HE washer ESA upgrades than 3 occupant participants, then the estimate of β_1 could be biased upwards.

Another important consideration is if selection into the program could be driven by HE washer eligibility. This might lead to 4 occupant ESA households being systematically different than 3 occupant households. It is unlikely, however, that prospective ESA participants knew what measures they might qualify for. The appliance eligibility criteria was not advertised or public on any of the utility web sites. Households only discovered what they were eligible for after they completed the enrollment contractor step. Even if a household was eligible based on all of the criteria, the weatherization specialist (the second contractor to visit) usually had the final say on if a household could accommodate a replacement HE washer.

2.4.3 HE washer results

Table 2.9 presents the results from the estimating the first stage equation (2.5). It shows a significant increase of in washer uptake of 66 percentage points for households with 4 occupants. Figure 2.17 presents this first stage graphically by showing the HE washer uptake by number of occupants. This figure helps confirm that the uptake pattern is driven by this occupant threshold, and the pattern continues for smaller and larger households.

The results for the IV regression can be seen in Table 2.10, which reflects the estimation of equation (2.6). Column (1) reports the results of the regression run in levels, and column (2) has the same result in logs. The results show that the instillation of an HE washer significantly reduced therm usage in households that received them. This 1.26 therm/month saving from the new HE washers was around a 4.5 percent reduction in therm usage.

The economic significance of these reductions must also be considered. SoCalGas CARE customers pay between 55.94 and 76.74 cents per therm depending on their monthly usage.[9] The 1.26 therm/month reduction translates to a .70 to .97 dollar/month discount on a households bill. Even though this is a small savings amount, it is important to note that the therm reduction is only one part of the savings from an HE washer. The new washing machines likely provide water savings, as well as some electricity savings for households that use electric dryers. Unfortunately, the water data is not available for analysis in this paper. Electricity data should be available for subsequent drafts.

2.5 Discussion

2.5.1 Costs vs benefits

Comparing the costs and benefits of the ESA program is important to understanding its success as a low-income energy retrofit program. As discussed in the conceptual framework, the ESA program must be evaluated as both a transfer program and a carbon mitigation policy. As an in kind transfer program, it is strictly worse from a welfare perspective than an equivalent cash transfer. Furthermore, it is not able to dynamically provide relief in times of need such as economic downturns.

To quantify the benefits of the ESA program as an energy policy, this paper empirically estimates the savings the program generates. There are a number of

[9] SoCalGas has a two tier rate schedule where customers pay a lower marginal cost for a given portion of their usage, and a higher rate for all remaining therms. The numbers cited here reflect the Tariff in January 2012 with the 20 percent CARE discount.

different ways to value the energy savings to society. First, households receive private benefits in the form of reduced energy bills. Second, ESA customers are also on the CARE rate. All reductions in energy usage by this population will reduce the amount of CARE subsidies that must be paid out. This in turn reduces the amount of fees that must be collected to fund the CARE program. Third, the reduction in energy usage also reduces the amount of carbon emissions associated with eletricity generation. These three main savings from the ESA program are considered in turn and the total is reported to evaluate the cost-effectiveness of the ESA program. This draft has not considered the comparison to the true marginal cost of energy, which will be included in subsequent drafts.

Private benefits

One of the main goals of the ESA program is to reduce energy usage and bills for low-income households. The savings estimates presented in section 5 show that the program likely does not provide significant electricity or gas savings to participants. It is important to highlight that those savings are a LATE around the self-certification threshold, and do not represent the average effect of the program on a new enrollee. Even considering that caveat, it is striking that a large part of the eligible ESA population can expect few savings when they enroll in the ESA program.

It is hard to back out a specific cost against which one can compare savings estimates to. One method is to compare measure costs to program savings. For example, SDG&E spent $14.4 million dollars on measures in 20,888 homes in 2012. This works out to $691/household for just the measures alone. An additional $7.5 million was spent on costs such as outreach and assessment ($3.5 million), advertising ($1.2 million), administration costs ($1.9 million). Factoring in total program expenditures, which is the best metric to compare savings against, the program cost $1,053 per household. Considering that a large portion of households likely does not see any significant savings, this is a noteworthy overall program expenditure.

The heterogeneity in upgrade potential present in the ESA program suggests that overall cost numbers are not always the best metric to compare against savings. Some households likely cost more than $1053 per household, and many cost less. To better understand these individual expenditures, I use a set of calculators generated for the ESA program. These are developed by the consulting firm E3 and give measure costs for each of the main measures installed. They estimate that the gross cost, which includes purchase price and retrofit expenditures, totals to $1,025 for a replacement refrigerator.

Discounted private benefits over the lifetime of the refrigerator are shown in the top part of panel A of table 2.12. The nature of California energy billing, where households pay on increasing blocks based on how much they use, necessitates private benefits being presented as a range. The low value in the left column of table 2.12 corresponds to energy savings for customers on the first tier and the high numbers on the right to customers on the highest tier. For refrigerators, this works out to \$3.63/month for low tier usage and \$6.38/month for high tier usage.

The calculations show the discounted stream of energy savings to households at a 10, 15 and 20 year horizon at a 3 and 6 percent discount rate in a similar manner to Fowlie, Greenstone, and C. Wolfram (2015a). The 15 year horizon corresponds to the estimated appliance life used in the E3 calculators. The savings numbers in the left and right sections display the lower and upper bounds respectively. The results show that users on the lower electricity tiers, and thus paying the lower electricity rate, do not see savings that exceed the \$1,025 cost of the refrigerator. Users on the highest tier that pay the most for electricity only see cost-effective savings for the longest time horizon and lowest discount rate. These outcomes show that when only counting private benefits, refrigerator replacements are not necessarily a cost-effective investment.

Using similar metrics, HE washers are estimated to cost \$749 per replacement. These can be compared to the the energy savings estimates for HE washers shown in the bottom portion of panel A in table 2.12. These numbers show that the natural gas savings are much lower than the replacement cost of a washer. These estimates, however only reflect one aspect of the savings that HE washers can generate. By using less water, HE washing machines can help lower a household's water bill. Electricity bills may also be reduced for households with an electric dryer since HE washing machines are better at removing moisture from clothing. Unfortunately, this analysis cannot presently capture the electricity benefits. With this limitation in mind, the numbers presented in table 2.12 present a lower bound for total savings from HE washers.

Benefits for reductions in CARE usage

The CARE program, which provides subsidized energy for low-income households, and the ESA program are closely linked in California energy policy. The two programs share the same eligibility limits and ESA treated households are automatically enrolled in CARE. The ESA program can reduce the amount of CARE expenditures by lowering low-income household demand for energy (and CARE subsidies) going forward. In SDG&E's territory, the CARE rates are 4.35 cents/kWh lower for users on the first tier and 10.14 cents/kWh lower for users on the highest tier. The discount in SoCalGas is 13.99 cents/therm for users on the lowest tier, and 19.19 cents/therm for high tier users.[10] By lowering the demand for energy, the ESA program can reduce the amount of subsidies it pays out to CARE customers.

Table 2.12 panel B shows the range of potential CARE savings from installing a refrigerator and a HE washing machine. The calculations show that there are significant savings from reduced CARE payouts for a refrigerator replacement. For example households consuming on the top tier at a 6 percent discount rate over 15 years can reduce the amount of CARE subsidies by \$456 over the lifetime of the refrigerator. The results for HE washing machines are more modest. This is due to the smaller savings numbers for HE washers, and the lower relative CARE discount for natural gas.

Social and total benefits

Reducing carbon emissions is an integral goal of the ESA program. To account for the amount of carbon emissions avoided, I use the estimates of carbon intensity in J. S. Graff Zivin, Kotchen, and Mansur (2014). They estimate the marginal carbon emissions rate of generation on specific parts of the U.S. grid for every hour of the day. I use their estimate of .8lbs/kWh, which is the daily average emissions rate of the Western Interconnection grid. Avoided emissions of CO_2 are valued at \$38 ton as calculated in Greenstone, Kopits, and Wolverton (2013) and used in a similar manner in Fowlie, Greenstone, and C. Wolfram (2015a).

The value of total abated carbon can be seen in panel C of table 2.12. The carbon savings estimates are modest and highlighting that the climate benefits of the ESA program are much smaller than the direct energy cost savings. For example, the carbon savings from a refrigerator replacement are 8 percent the value of the private benefits for high tier users. It is important to note that the carbon savings numbers should only be counted as benefits for time periods when California does not have a binding cap-and-trade program in effect. Under a binding carbon cap, the total carbon emitted in the state is capped, so emissions reductions from energy efficiency will not reduce total state emissions. There was no binding carbon cap during the period studied in this analysis.

Adding up the private, CARE and carbon reduction benefits gives a good approximation of the value of each appliance replacement to society. These estimates can be seen in panel D of table 2.12. The results show that refrigerators are a cost-effective replacement under most scenarios when reducing high tier electricity usage. This is not the case when refrigerators are displacing first tier energy usage, where replacement costs are mostly higher than benefits. High efficiency washing

[10] As before, these rates reflect prices in January 2012.

machines show lower savings numbers than the $749 replacement cost. These values only reflect the therm savings, and not electricity or water savings, making them not directly comparable to the full $749 cost.

The total benefits in table 2.12 do not include non-energy benefits. They are outside the scope of this analysis, and typically are challenging to estimate. Any non-energy benefits would add on to the estimated energy benefits in table 2.12, making refrigerator replacements cost-effective under a larger set of situations.

2.5.2 Comparison to third party evaluation

The ESA program is periodically evaluated by a third party consulting firm to assess program savings. The last evaluation was written in 2013 and covered the results for program year 2011 (Evergreen Economics 2013). The study found that the ESA program saved 278.57 kWh/year and 26.06 therms/year for the average SDG&E customer. Both of these estimates are outside the 95 percent confidence intervals of the results found in this paper.[11]

The Evergreen Economics (2013) study also estimated measure specific savings. The refrigerator estimates show a replacement unit saved 640.42 kWh/year, which is substantially larger than the 436.58 kWh/year savings estimated in this paper. The therm savings from HE washing machines in SoCalGas's territory was found to be 30.88 Therms/year which is more than twice the 15.12 savings found in this analysis.[12]

The methodology used in the Evergreen Economics (2013) study does not leverage quasi-random variation in program uptake to estimate program effects. Furthermore the appliance estimates in this paper are LATEs around the implementation thresholds, meaning they do not provide an average savings estimate for all refrigerator or HE washer replacements. As a consequence, the estimates in this analysis are not directly comparable to the Evergreen Economics (2013) study. With these caveats in mind, the existing Evergreen Economics (2013) report finds significantly larger savings estimates for the ESA program. Past 3rd party evaluations have found similarly large program savings, highlighting the importance of further careful evaluation going forward.

[11] The 75 meter 1000 meter estimates can reject an annual savings of 245 and 122 kWh/year respectively. The therm savings estimates can reject a 23.7 and 2.1 therm/year for the same distances.

[12] Both the refrigerator and HE washing Evergreen estimates lay outside the 95 percent confidence interval of the estimates in this paper. Standard errors were not reported in the Evergreen savings estimates.

2.6 Conclusion

This paper focuses on evaluating the Energy Savings Assistance program in California, and provides insight on the effectiveness of efficiency retrofit programs aimed at low-income households. The prevalence of similar programs will continue to increase with the Clean Power Plan and the CIEP credit mechanism incentivizing their deployment. States that implement energy efficiency programs to meet their carbon reduction goals will be credited based on their projected savings. If evaluations are not conducted appropriately, the climate goals of the Clean Power Plan could be undermined by programs not delivering on their ex-ante projections.

This paper focuses on the largest state-run low-income energy efficiency retrofit policy, the California Energy Savings Assistance program. The results find that the overall program is ineffective at delivering energy savings to most participants. Some households, however, were eligible for larger upgrades and had energy savings as a result. The 1 in 10 households that received a refrigerator replacement experienced between \$3.63 and \$6.38 in bill savings per month. Furthermore, the refrigerator replacements were cost-effective when considering the full set of societal benefits.

The analysis in this paper is the first to both estimate program effects for a large portion of the treated population and to estimate measure specific savings numbers. Previous studies have either estimated overall program effects, or analyzed appliance replacement, but never both for the same retrofit program. The results provide important insights on program design that have not previously been empirically testable.

More broadly, this paper provides strong evidence of heterogeneity in upgrade potential affecting overall program outcomes. Many households are not well suited for energy efficiency retrofits, and their inclusion in upgrade programs can erode the overall cost-effectiveness of a program. In the ESA program, 81 percent of households received smaller upgrades and likely did not see significant savings. This challenge is present in most energy efficiency retrofit programs, and must be carefully considered when designing efficiency policies.

Figures

Figure 2.1: In kind transfers

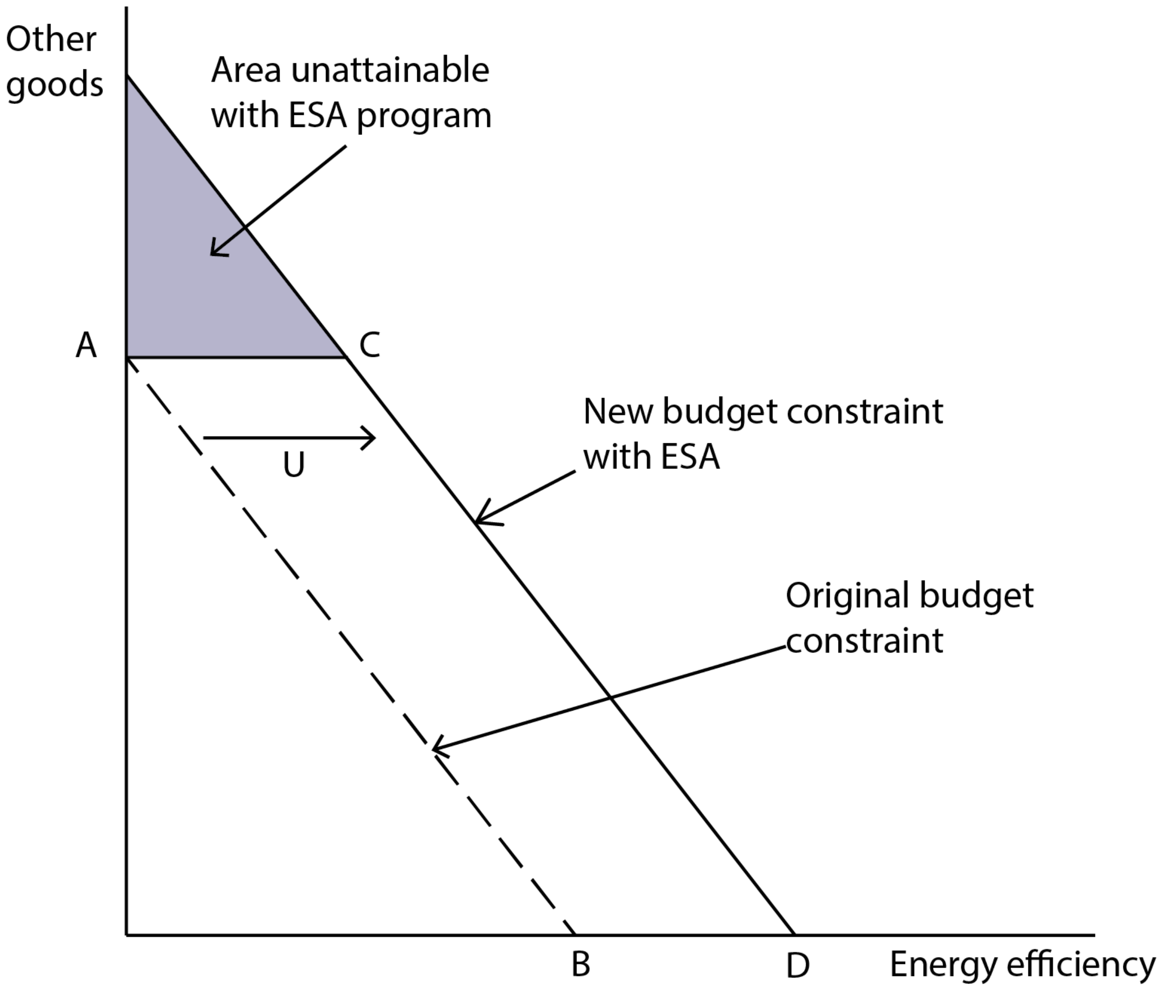


Figure 2.2: In kind transfers

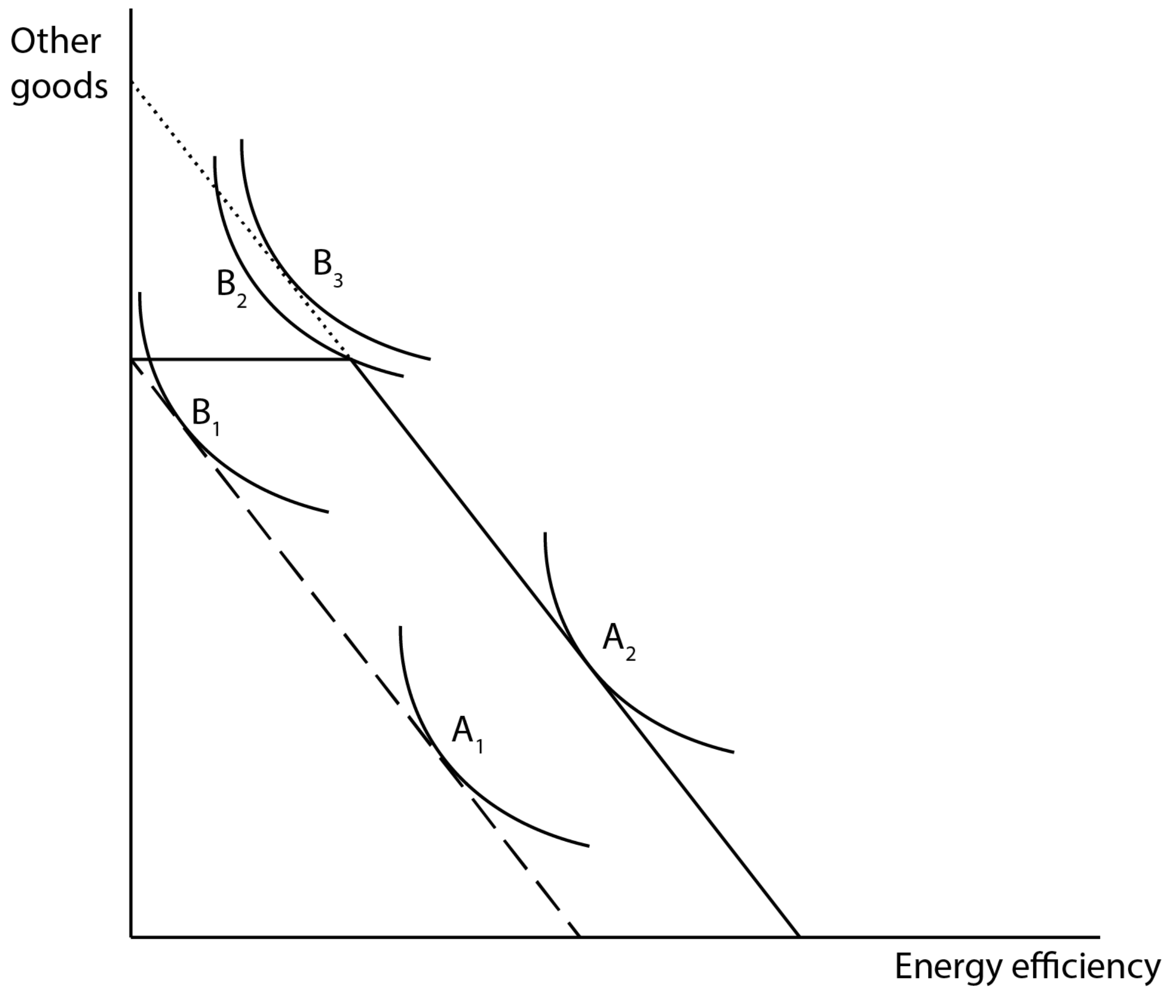


Figure 2.3: Care enrollments and unemployment rate in SDG&E's service territory

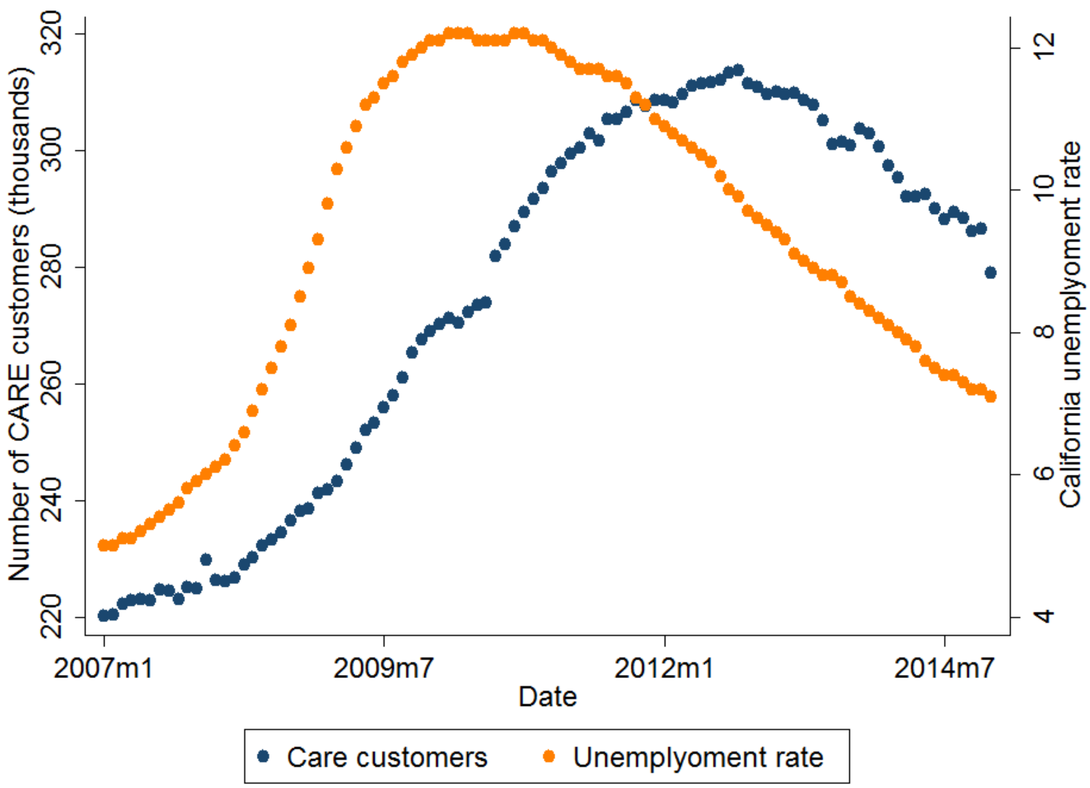


Figure 2.4: Map of all zip+4s in SDG&E's service territory by certification status

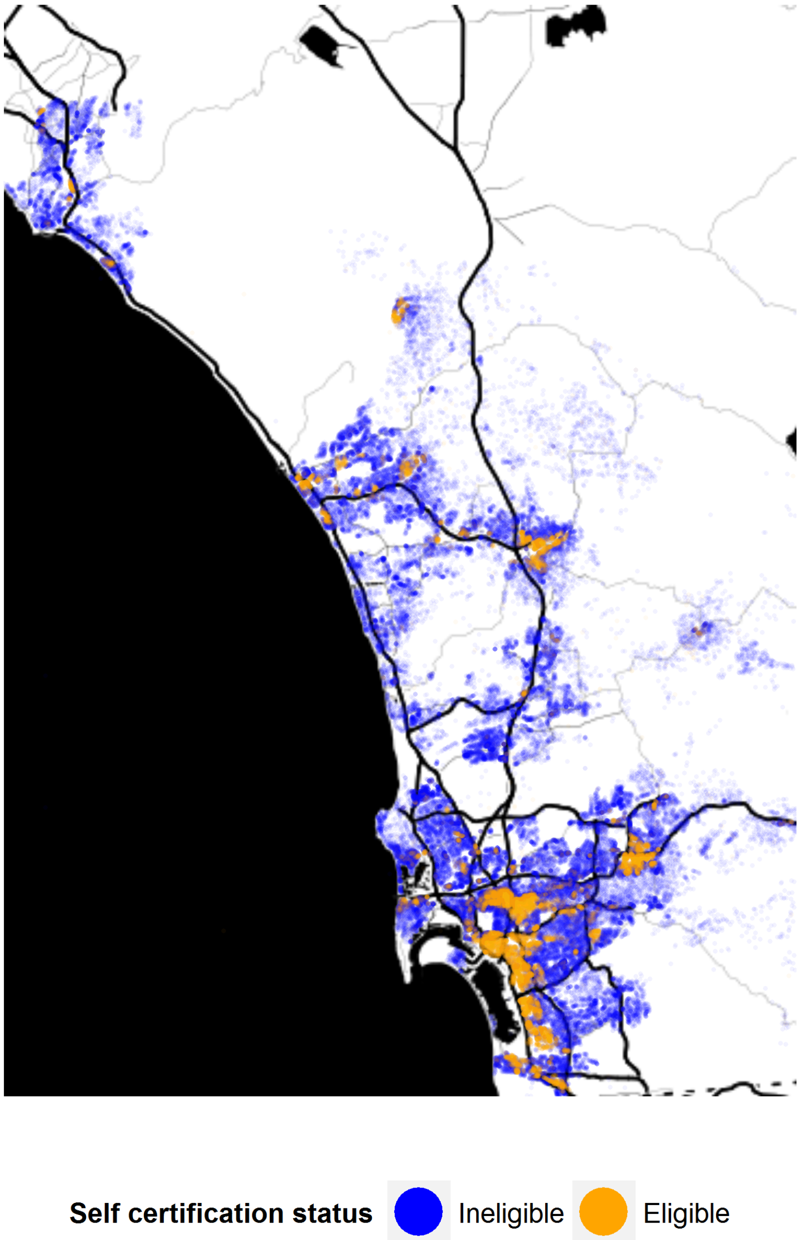


Figure 2.5: Map of zip+4s in SDG&E's service territory trimmed to 75 meters by certification status

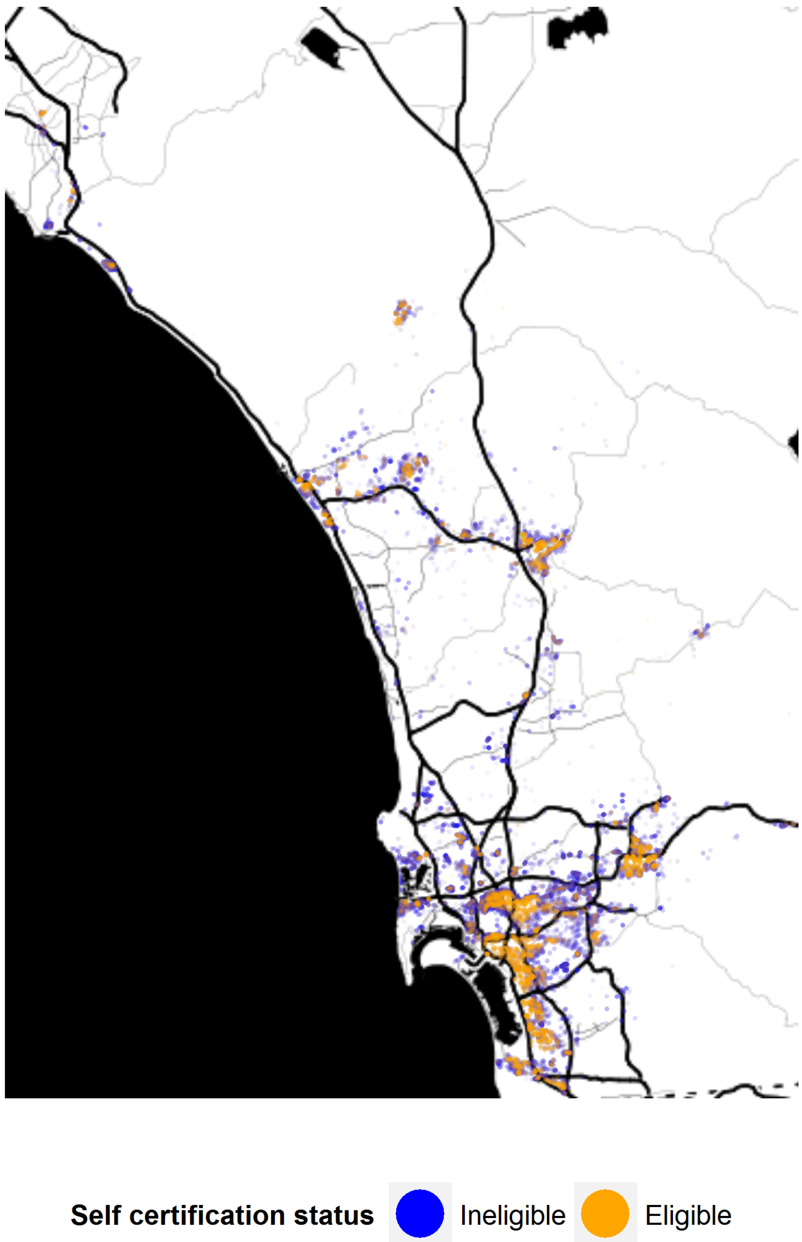


Figure 2.6: Zoomed in map of zip+4s trimmed to 75 meters by certification status

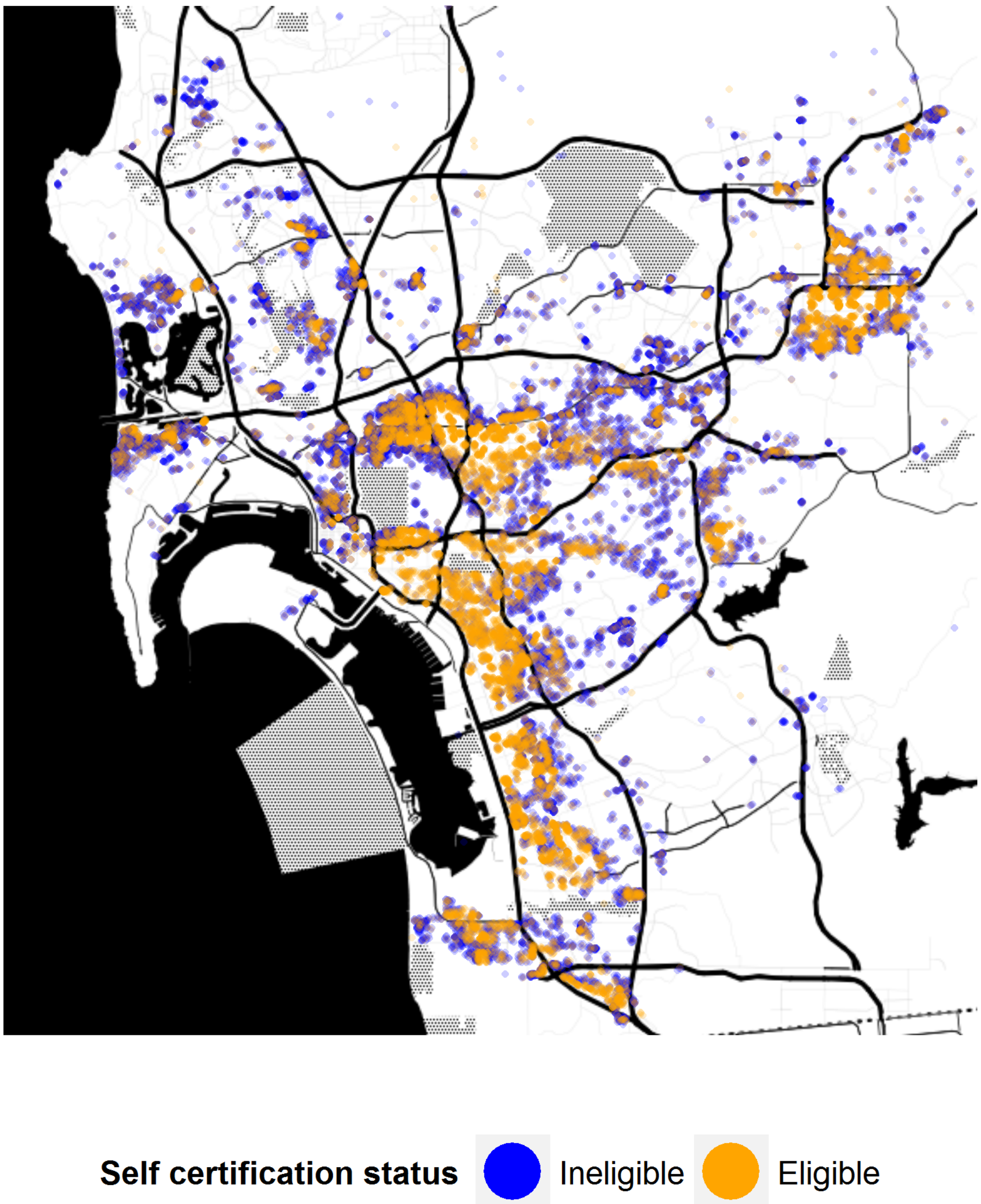


Figure 2.7: Zoomed in map of distances between zip+4s of opposite certification regimes trimmed to 75 meters

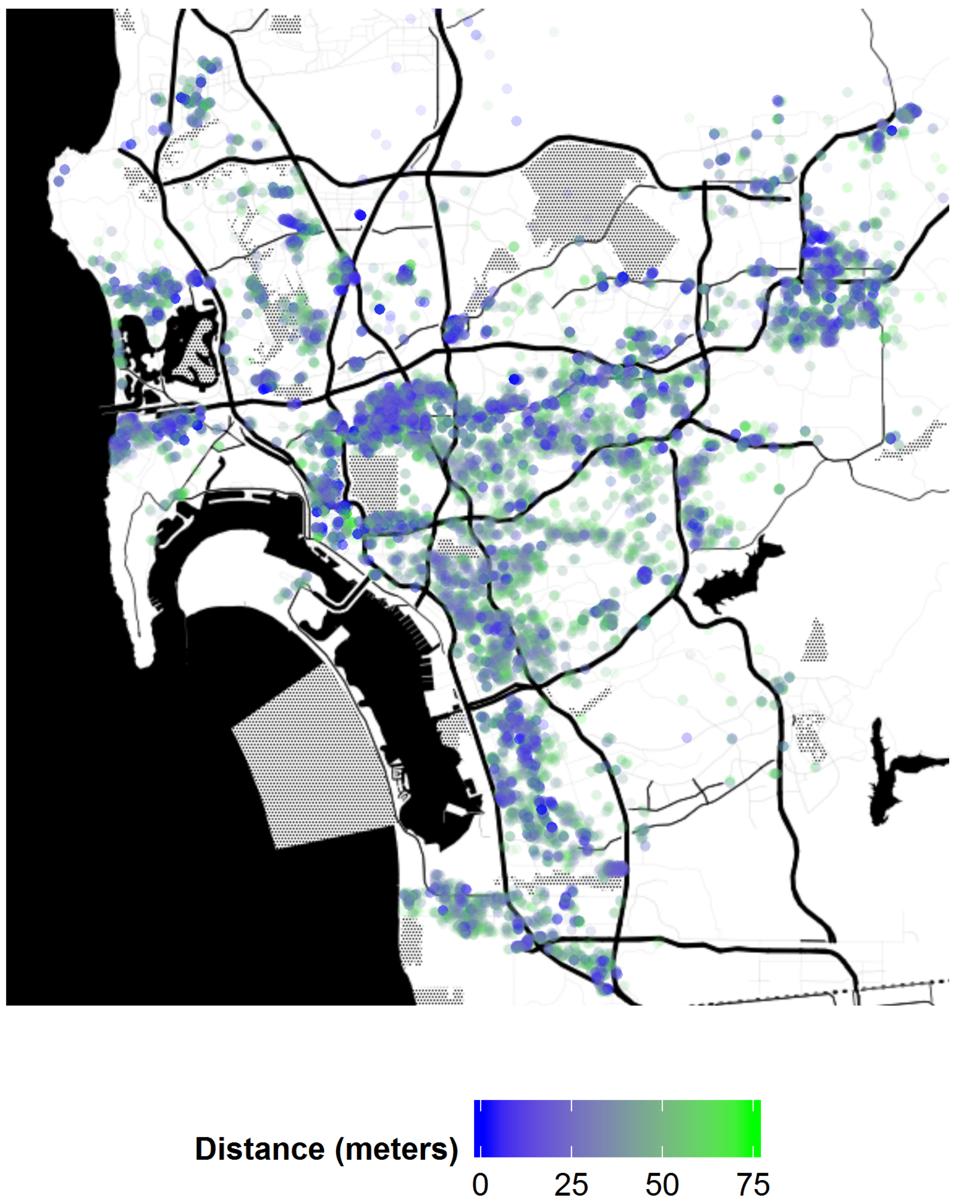


Figure 2.8: Individual housing units by self-certifications status

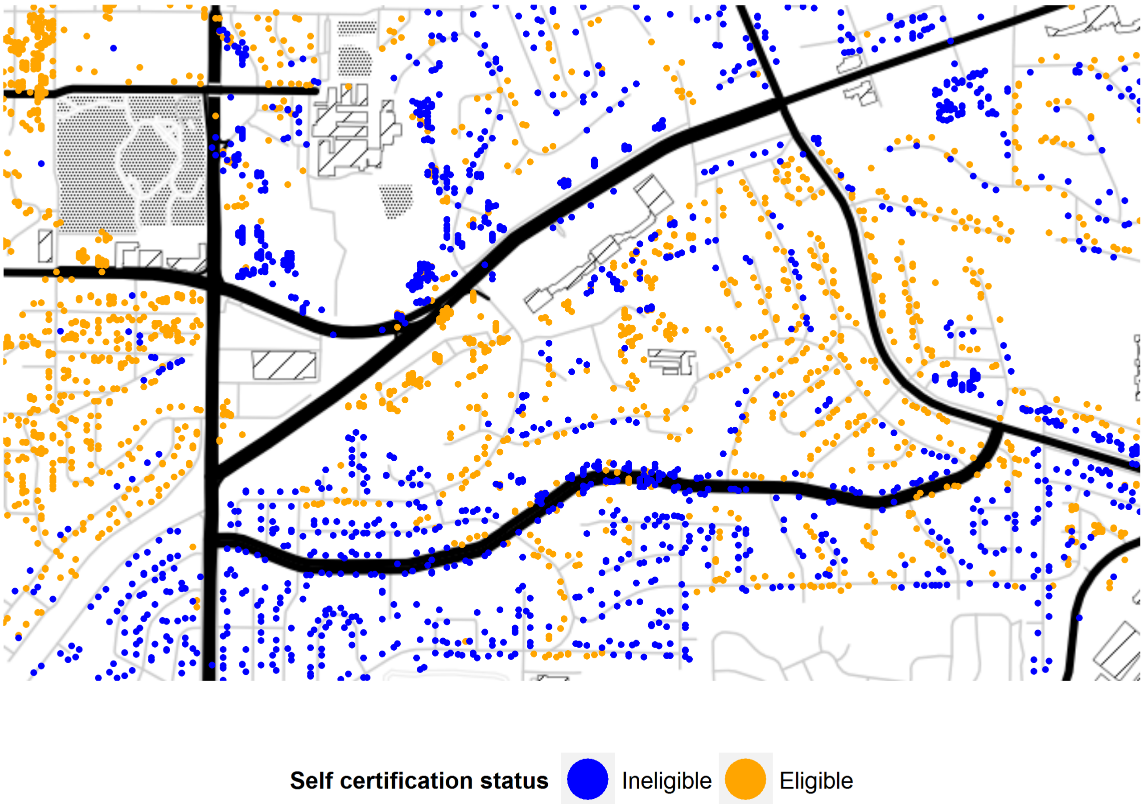


Figure 2.9: Individual housing units by self-certifications status - trimmed to 75m

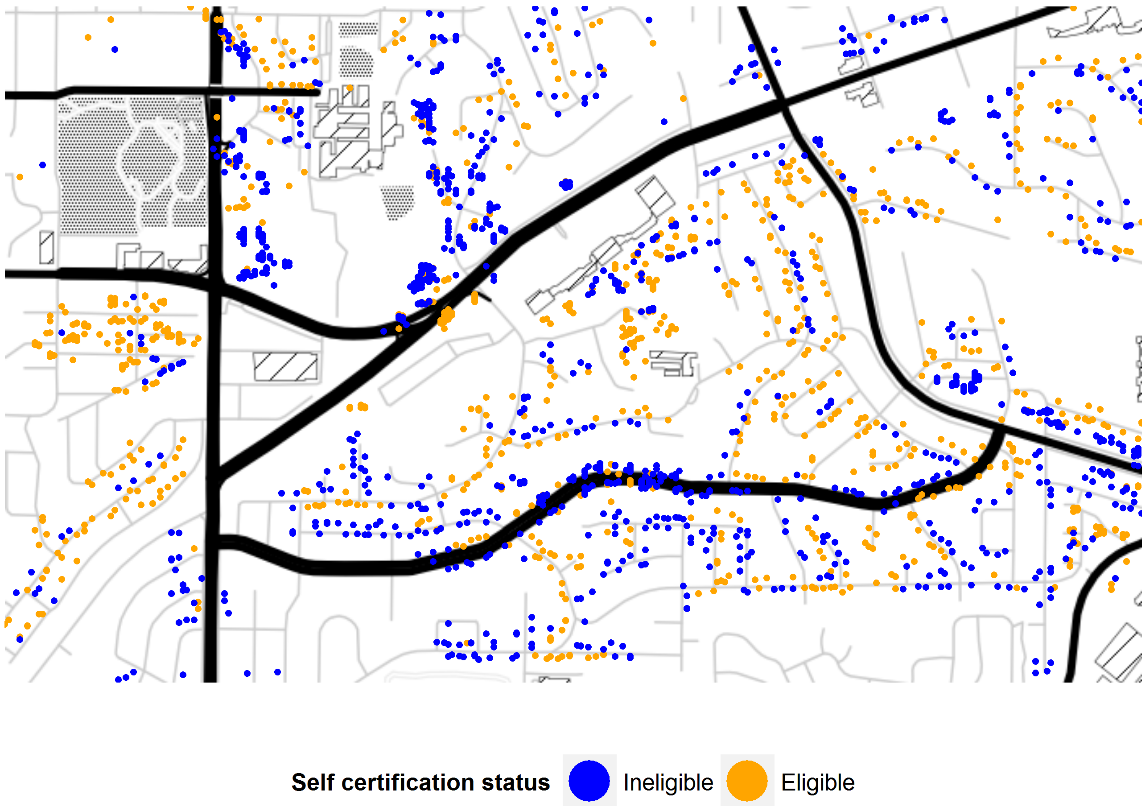


Figure 2.10: Density of pre-period kWh consumption by eligibility status

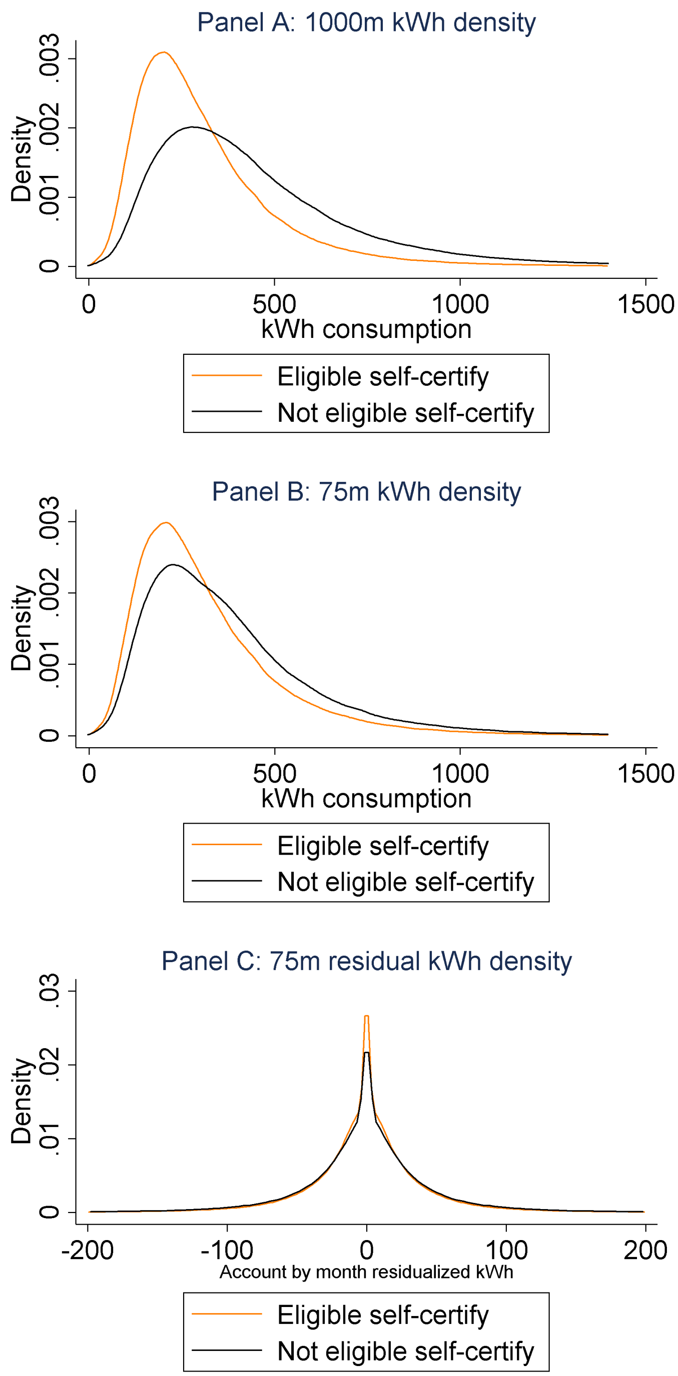


Figure 2.11: Density of pre-period therm consumption by eligibility status

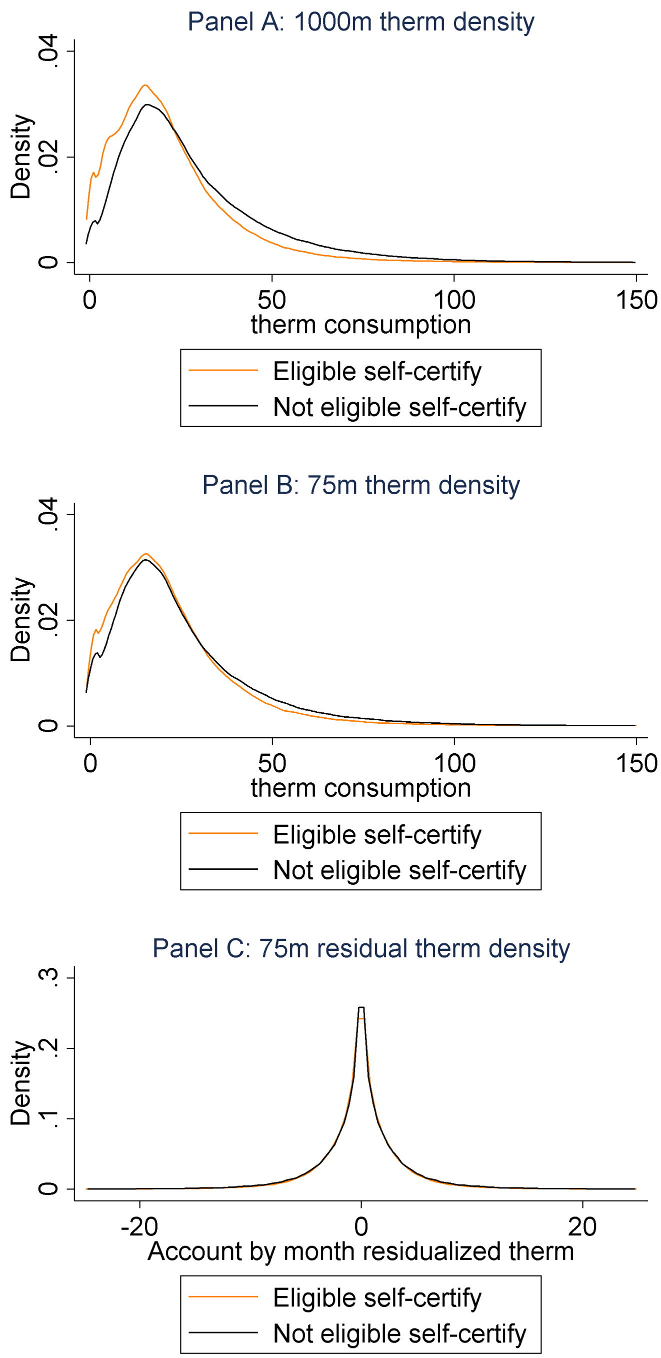


Figure 2.12: First stage of fridge analysis

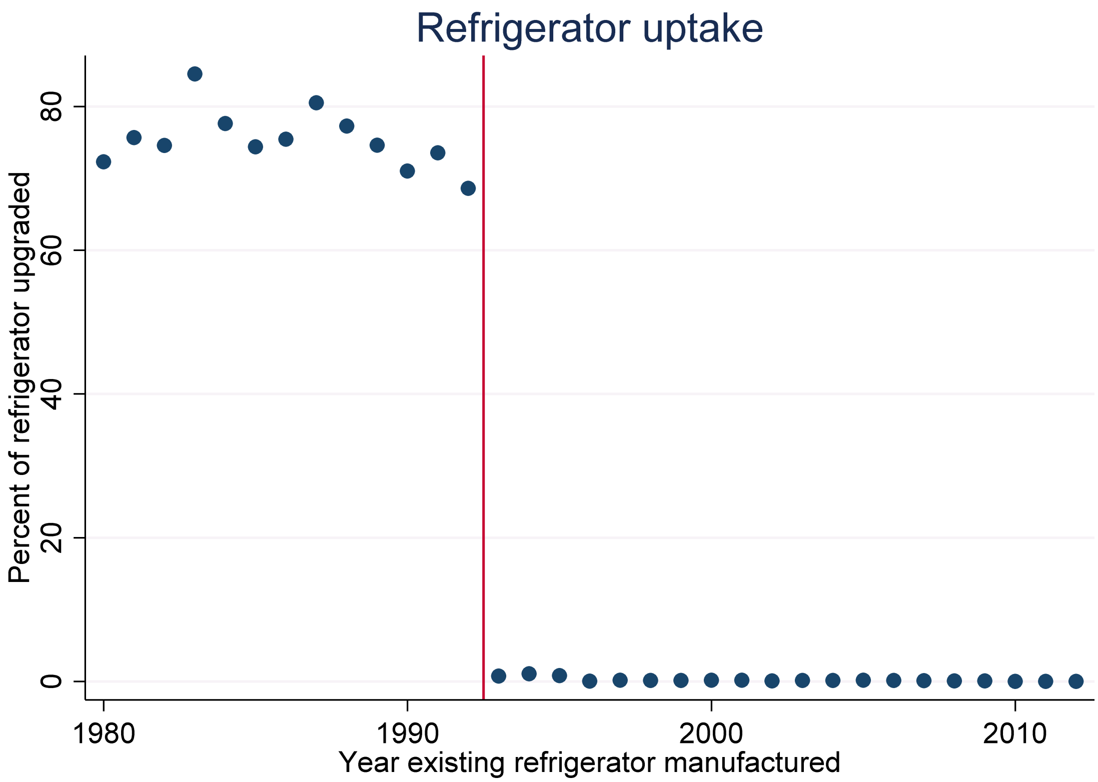


Figure 2.13: Distribution of fridge vintages

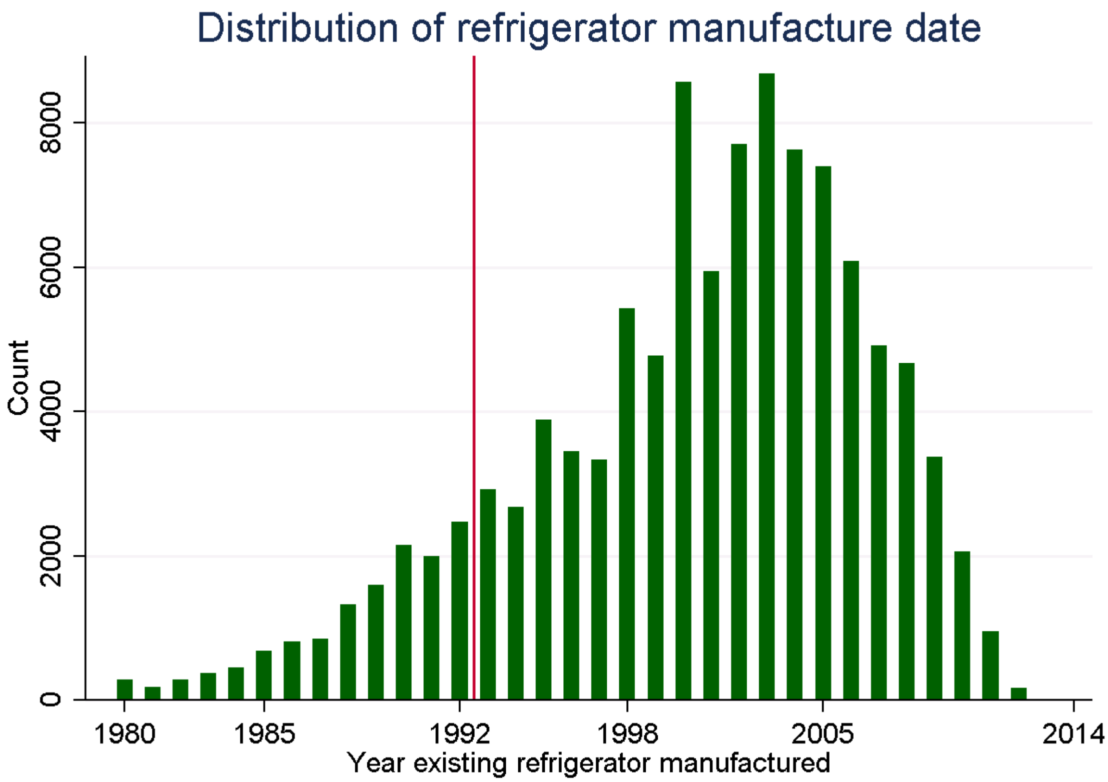


Figure 2.14: Covariates graphed through 1993 fridge age discontinuity

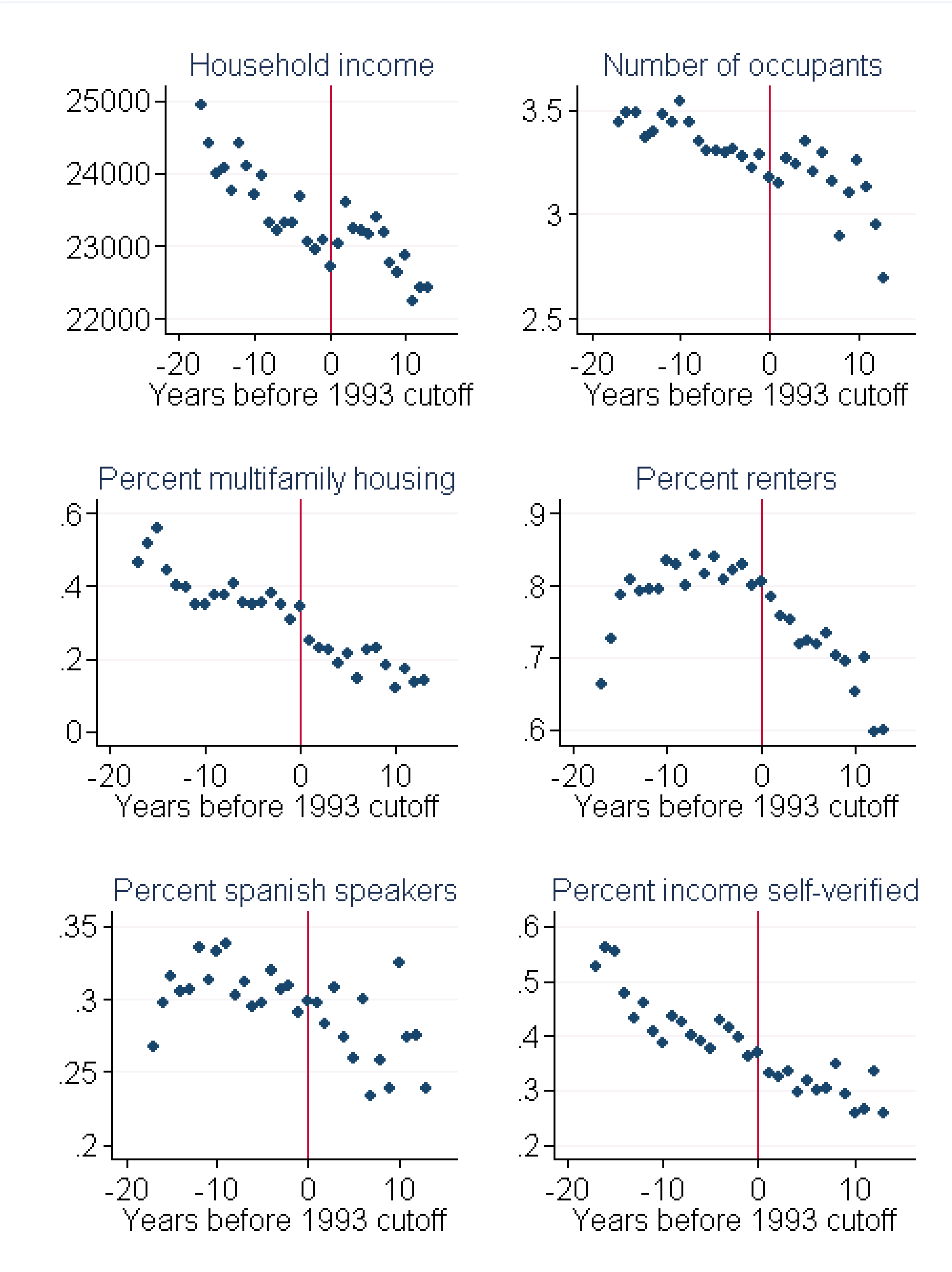


Figure 2.15: Savings from fridge replacement

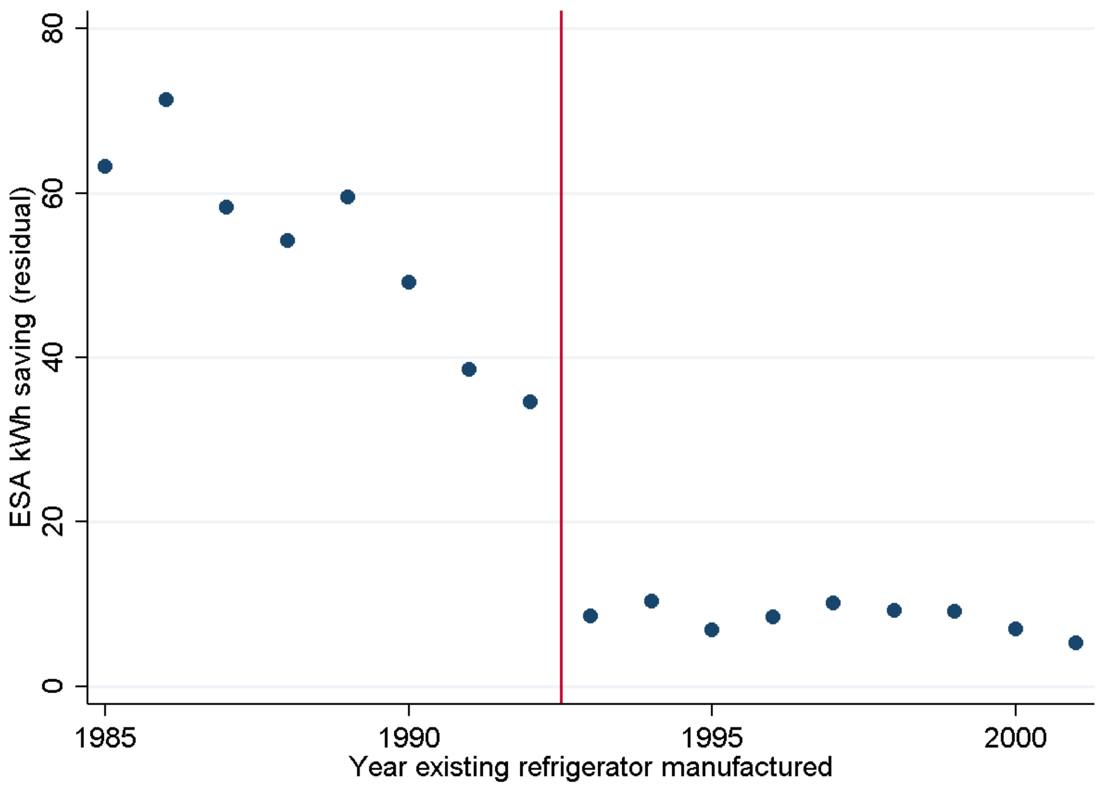


Figure 2.16: Savings from fridge replacement - alternate specification

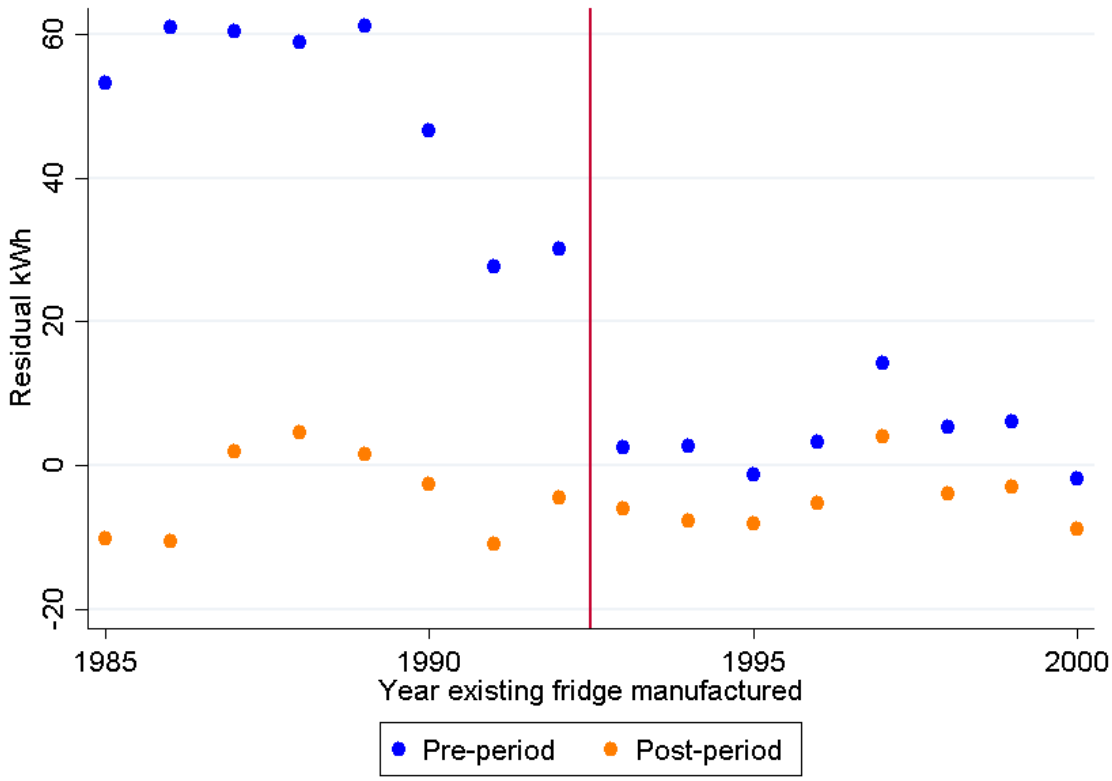
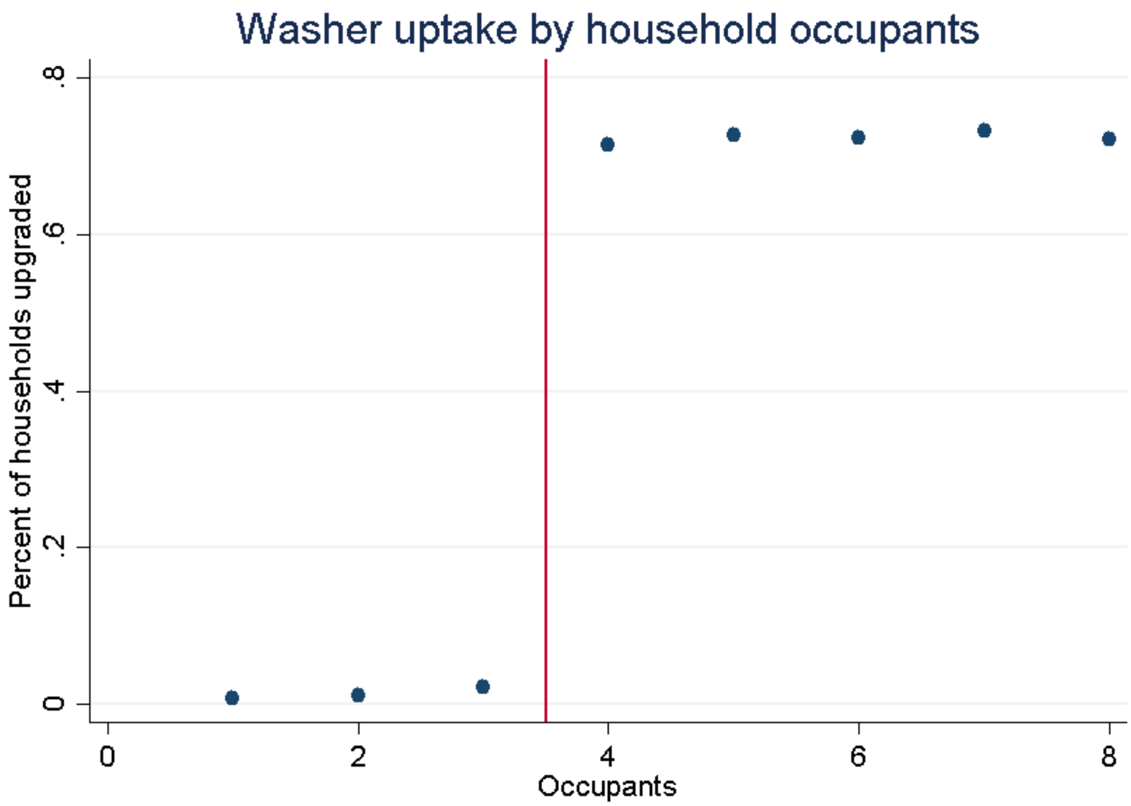


Figure 2.17: First stage of high efficiency washer analysis



Tables

Table 2.1: Income limits for ESA program

Household Size	Care Income Limit
1 to 2	\$30,500
3	\$35,800
4	\$43,200
5	\$50,600
6	\$58,000
Each additional	\$7,400

Note: Limits reflect 2008-2009 cutoffs

Table 2.2: Summary of ESA upgrades

Upgrade category	Count	Percent of households receiving upgrades
Minor upgrade only	129455	81.25
Refrigerator	16715	10.49
Lighting	152650	95.8
Microwave	16006	10.05
Furnace replacement	4048	2.54
Room AC	2440	1.53
Attic insulation	4538	2.85
Washer - electric	607	.38
Central AC	61	.04
Washer - gas	5367	3.37
Water heater	845	.53
Building seal	107356	67.38
Water conservation	94308	59.19
Furnace tune up	28527	17.9
Total number of houses upgraded	159336	100

Notes: Data presents all upgrades in SDG&Es territory from 2007-2012. These aggregate categories were constructed by the author to best reflect the nature of the upgrades. Minor upgrades signifies households that received only one of the following: lighting, microwave, building seal, water conservation and furnace tune up.

Table 2.3: Program savings: First stage

	(1)	(2)
Self-certification instrument	0.12587***	0.15359***
	(0.00440)	(0.00312)
Average take-up rate	0.259	0.239
Distance trimming	75 meters	1000 meters
Number of accounts	38,323	109,199
Number of observations	2,953,999	8,111,111
F statistic	820	2,425

Notes:All regressions have account by month FE and zip by month of sample FE. Standard errors clustered at account level

Table 2.4: Program savings: IV results

	(1) kWh	(2) kWh	(3) Therms	(4) Therms
ESA program install	-0.839	0.948	-0.172	0.761
	(9.976)	(5.658)	(0.921)	(0.478)
Mean kWh	355.951	398.839	22.813	24.991
Max distance	75 meters	1000 meters	75 meters	1000 meters
Number of accounts	38,320	109,189	24,808	78,978
Number of observations	2,956,417	8,126,628	1,934,202	5,868,280

Notes:All regressions have account by month FE and zip by month of sample FE. Standard errors clustered at account level

Table 2.5: Program savings: IV results interacted with refrigerator installation

	(1) kWh	(2) kWh
ESA program install	3.798	5.152
	(10.716)	(6.007)
Fridge installed	-40.271***	-41.991***
	(8.534)	(5.149)
Mean kWh	355.951	398.839
Max distance	75 meters	1000 meters
Number of accounts	38,320	109,189
Number of observations	2,956,417	8,126,628

Notes:All regressions have account by month FE and zip by month of sample FE. Standard errors clustered at account level

Table 2.6: Refrigerator RD: first stage results

	(1) No FE	(2) Zip-by-month FE
Eligible for fridge	0.72457***	0.73181***
	(0.01236)	(0.01223)
Distance by post	0.00360***	0.00579***
	(0.00064)	(0.00091)
Distance by eligible by post	-0.00764	-0.01177
	(0.00967)	(0.00956)
Number of accounts	7,668	7,668
Number of observations	99,684	99,684
F statistic	2,826	2,892
Bandwidth	3 years	3 years

Notes:All regressions have month of sample FE. Standard errors clustered at account level

Table 2.7: Refrigerator RD: IV results

	(1) kWh	(2) kWh	(3) Log kWh
Fridge monthly kWh savings	-36.17366***	-36.38198***	-0.11646***
	(7.58157)	(8.41046)	(0.01972)
Distance by post	0.92664	1.20173	0.00499
	(1.89563)	(2.20018)	(0.00494)
Distance by eligible by post	6.51369**	4.37848	0.00900
	(3.25937)	(3.67810)	(0.00834)
Post	-5.18822	-14.41889***	-0.04312***
	(4.18866)	(5.12823)	(0.01232)
Distance	-3.97252	-3.36659	-0.01043
	(4.27086)	(4.04895)	(0.00984)
Distance by eligible	-4.94705	-2.15524	-0.01443
	(7.30631)	(7.14531)	(0.01571)
Eligible	21.49186*	24.07697**	0.04972*
	(12.42827)	(11.68784)	(0.02694)
Average kWh consumption	391.718	391.718	5.782
Number of accounts	7,668	7,668	7,667
Number of observations	99,684	99,684	99,647
Bandwidth	3 years	3 years	3 years
Account-by-month FE	No	Yes	Yes
Zip-by-month FE	No	Yes	Yes

Notes:All regressions have month of sample FE. Standard errors clustered at account level

Table 2.8: Refrigerator RD: bandwidth robustness checks

	(1) 2 year bandwidth	(2) 3 year bandwidth	(3) 4 year bandwidth	(4) 5 year bandwidth
Fridge monthly kWh savings	-39.11551***	-36.38198***	-34.64998***	-41.00426***
	(12.17608)	(8.41046)	(7.07341)	(6.25784)
Average kWh consumption	391.962	391.718	394.015	396.465
Number of accounts	4,769	7,668	10,129	12,429
Number of observations	61,997	99,684	131,666	161,561
Bandwidth	2 years	3 years	4 years	5 years

Notes:All regressions have month of sample FE. Standard errors clustered at account level

Table 2.9: HE washing machines: first stage results

	(1)
Four occupants by ESA upgrade	0.65699***
	(0.00449)
ESA upgrade	0.00255
	(0.00214)
Number of accounts	23,415
Number of observations	770,211
F statistic	12,703

Notes:All regressions have month of sample FE. Standard errors clustered at account level

Table 2.10: HE washing machines: IV results

	(1) Therm	(2) Log therm
Washer installed	-1.25635***	-0.04448***
	(0.20366)	(0.00695)
ESA upgrade	-0.38798***	-0.01734***
	(0.12888)	(0.00428)
Average monthly therm usage	.	.
Number of accounts	23,415	23,410
Number of observations	770,211	763,457

Notes:All regressions have month of sample FE. Standard errors clustered at account level

Table 2.11: Fridge savings bandwidth checks

	(1) 2 year bandwidth	(2) 3 year bandwidth	(3) 4 year bandwidth	(4) 5 year bandwidth
Fridge monthly kWh savings	-39.11551***	-36.38198***	-34.64998***	-41.00426***
	(12.17608)	(8.41046)	(7.07341)	(6.25784)
Average kWh consumption	391.962	391.718	394.015	396.465
Number of accounts	4,769	7,668	10,129	12,429
Number of observations	61,997	99,684	131,666	161,561
Bandwidth	2 years	3 years	4 years	5 years

Notes:All regressions have month of sample FE. Standard errors clustered at account level

Table 2.12: Discounted benefits from appliance upgrades

	Low tier costs		High tier costs	
Discount Rate	3 percent	6 percent	3 percent	6 percent
Panel A - Private Benefits				
Fridges				
10 years	$383	$340	$673	$598
15 years	$536	$449	$942	$789
20 years	$668	$530	$1,174	$932
HE washing machines				
10 years	$74	$66	$102	$91
14 years	$104	$87	$143	$119
20 years	$130	$103	$178	$141
Panel B - CARE benefits				
Fridges				
10 years	$167	$148	$389	$345
15 years	$234	$196	$544	$456
20 years	$291	$231	$678	$538
HE washing machines				
10 years	$19	$17	$25	$23
15 years	$26	$22	$36	$30
20 years	$32	$26	$44	$35
Panel C - Carbon benefits				
Fridges				
10 years	$53	$47		
15 years	$74	$62		
20 years	$92	$73		
HE washing machines				
10 year	$25	$22		
15 year	$35	$30		
20 year	$44	$35		
Panel D - Total Savings				
Fridges				
10 years	$603	$535	$1,115	$990
15 years	$844	$706	$1,560	$1,306
20 years	$1,052	$834	$1,945	$1,543
HE washing machines				
10 years	$118	$105	$153	$136
15 years	$165	$138	$214	$179
20 years	$206	$163	$266	$211

Chapter 3

Environmental Air Quality Regulation and Coal Employment Costs: Evidence from Appalachia

3.1 Introduction

Starting with its passage in 1963, The Clean Air Act (CAA) is the main legislation responsible for regulating air quality in the U.S. The Environmental Protection Agency (EPA) estimated that in 2010 the CAA will provided $1.3 trillion in benefits, most of which were from the reduction in premature deaths due to airborne particulate matter. The costs of the policy in 2010 are estimated to be $53 billion per year, the majority of which are estimated to come from shifting technologies to reduce emissions in the electricity utility and transportation sectors (Environmental Protection Agency 2011).

One potentially important cost that is not included in the official EPA estimates is the cost of dislocated workers from industries impacted by the regulation (Environmental Protection Agency 2011). If displaced employees in regulated industries are able to easily find jobs with similar wages in other sectors, then this cost may be small. W. R. Walker (2013) investigates this question and shows that the reallocative costs of the 1990 CAA amendments are significant. He finds that workers in regulated sectors experience a total lifetime earnings loss of 20% compared to those in unregulated sectors. The existing literature has focused on the cost of environmental regulation to the sectors and plants explicitly covered by the regulation. These analyses capture a large portion of the costs of environmental regulation, but they do not include the costs to industries that supply the regulated industries.

In this paper, I estimate the cost of the CAA on a prominent upstream sector – the coal mining industry. I focus on Title IV of the 1990 CAA amendments, which put a cap on sulfur dioxide emissions permitted from electricity generating plants. The regulation reduced the demand for high-sulfur coal, which had large impacts on the coal mining industry. In this paper I focus on the employment impact of Title IV on communities in the Appalachian region of the U.S. that previously mined much of this high-sulfur coal.

To study the impacts on counties that mine high-sulfur coal, I leverage a natural experiment created by Title IV of the CAA. One of the main ways electric power plants complied with the cap on sulfur dioxide emissions was to switch to a low-sulfur coal (Carlson et al. 2000). I focus on Appalachia in the eastern U.S. because within that region there are both low-sulfur and high-sulfur coal mines. This allows for the comparison between counties with high-sulfur coal mines that were negatively impacted by Title IV, and counties with low-sulfur coal mines that were not similarly affected. Using this approach, I am able to identify the impacts of Title IV on employment in the high-sulfur coal sector and to investigate spillovers to other sectors of the economy.

I find that Title IV has a large negative impact on coal industry in the high-sulfur counties. Production drops by 20% in the five years after the regulation. Coal mines appear to have reduced their output in part by laying off employees. Coal mining employment in high-sulfur counties drops by 14% relative to the neighboring low-sulfur counties. The regulation also impacts how mines function. I find that Title IV causes a reduction in labor efficiency (defined as tons mined/employees) for the high-sulfur coal regions relative to the low-sulfur coal regions. This result is consistent with a long-run reduction in value for high-sulfur coal decreasing the incentive to invest. It also could follow from a lowered economy of scale at now smaller coal mines. The overall reductions in production, employment and labor efficiency are striking considered the coal mining industry was not directly regulated under Title IV of the CAA.

Coal mining has historically been a large part of the Appalachian economy. In 1980, coal mining accounted for 15.5% of the West Virginian GDP. By 1994, this number had dropped to 7.8%, mostly through growth in non-coal sectors. Sociologists have noted that the coal industry actively promotes the image of coal mining as being part of the regional identity through industry funded organizations such as "Friends of Coal" (Bell and York 2010). The large role of the coal industry in Appalachia creates the concern that coal-mining job losses will spillover to the non-coal sectors. I test this hypothesis by using quarterly county level employment data and comparing non-coal industry outcomes across high and low-sulfur counties. I find no evidence that the Title IV regulations significantly impacted the Appalachian economy. I can rule out a reduction in total employment and wages in the high-sulfur coal counties of 2.8% and 1.2% respectively.

The results from this analysis highlight a previously unmeasured costs of the CAA on industries that are not directly regulated. The large impacts show that these upstream employment costs should be included in the official cost-benefit analysis of environmental regulation. This likely will be a future requirement. In 2016, a federal judge required that the EPA consider the impacts of the Clean Power Plan (a CAA regulation) on jobs in the coal industry.

There are two important policy implications from this analysis. First, the concentrated nature of the job losses in the coal mining industry suggest that a targeted federal program to compensate the losers from Title IV could ameliorate some of the the negative consequences of the regulation. To date, little or no funds have been directed to coal miners in this region as compensation for the employment costs of the regulation. Second, in the decades since the implementation of Title IV, the Appalachian region has continued to lose coal jobs at a high rate. Much of these changes are driven by technological improvements in mining and cheap natural gas, but the employment costs to coal miners remains significant. The results in this paper suggest that the non-coal sectors in Appalachia may not be significantly impacted by the continued decline of the coal mining industry.

The rest of the paper is organized as follows: Section 2 discusses the Clean Air Act, the related literature and the coal industry. Section 3 outlines the data used in the analysis. Section 4 describes the empirical strategy and Section 5 presents results. Section 6 discusses the results, policy implications and future directions for the research. Section 7 concludes.

3.2 Background

3.2.1 Clean Air Act

The Clean Air Act was first passed in 1963 with the goal of reducing harmful air pollution in the U.S. Major amendments were added in 1970 that introduced regulatory programs to reduce pollution from both stationary and mobile sources. The CAA was greatly expanded again in 1990 with a new set of amendments designed to increase the scope of the existing regulation. Title IV of the 1990 Amendments added a new program to regulate sulfur emissions from electric power plants that contribute to acid rain. This was done through a "cap-and-trade" system, where a cap was set at the total quantity of sulfur emissions and regulated entities were able to trade permits.

Phase 1 of the acid rain program began in 1995 and required 110 power plants across 21 states to comply with the cap. These power plants were the largest and dirtiest coal plants in the country, making them big contributors to sulfur emissions that caused acid rain. A second set of plants was brought under the cap-and-trade program in 2000, further expanding the restrictions on sulfur emissions (Busse and Keohane 2007).

Power plants have two main options to comply with the regulation. The first is to install "scrubbers" that remove the sulfur from the exhaust stream before they are released from the power plant. This can be an expensive investment and includes ongoing operation costs. Power plants also have the option to switch to coal with a lower sulfur content. Before Title IV, high-sulfur coal was generally preferred in power plants because it has a much higher energy content (around 12-13,000 Btu/ton) than low-sulfur coal (around 8-10,000 Btu/ton). Transportation of coal is a large portion of the cost, so this difference in energy density made high-sulfur coal the preferred input.

Installing scrubbers and switching to lower-sulfur coal were both used to comply with the Title IV cap-and-trade program. One of the consequences of Title IV was to increase the demand for low-sulfur coal across the U.S. This furthered the growth of major coal mining operations in the Powder River Basin of Wyoming, which is now the largest source of low-sulfur coal in the U.S. Busse and Keohane (2007). Title IV also reduced the value of high-sulfur coal by putting a price on the embedded sulfur content. The decrease in the value of high-sulfur coal had impacts on the Appalachian region, which mined a large portion of the high-sulfur coal in the U.S. This paper seeks to estimate a portion of the costs to the employees who worked in the coal industry and to the wider regional economy.

3.2.2 Costs and benefits of air pollution regulation

The EPA is required to estimate the projected costs and benefits of environmental regulation before it is implemented. The most recent retrospective report on the CAA conducted in 2011 estimates that the benefits are 25 times the costs (Environmental Protection Agency 2011). The official EPA analyses are conducted using a computable general equilibrium model that assumes the economy is at full employment, implying it does not capture the costs of dislocated workers. Other research has addressed the employment costs of environmental regulation using an ex-post econometric approach. Greenstone (2002) estimates that the 1970 CAA amendments were responsible for 590,000 jobs lost across the U.S.

W. R. Walker (2013) uses detailed, worker level data to track the impacts of the 1990 CAA amendments on wages for employees in regulated sectors. He estimates that the wage cost for workers in regulated industries totaled more than $5.4 billion. The majority of this lost income is from workers leaving the regulated industries and receiving lower wages in other sectors. Other analyses have investigated various other costs of environmental regulation including where industry locates due to regulation (Kahn and Mansur 2013) and the transition away from regulated industries and locations (W. R. Walker 2011).

A large literature documents the harmful impacts of air pollution on a variety of health outcomes. Infants are particularly sensitive to air pollution, and infant mortality has been shown to drop as air quality improves (Chay and Greenstone 2003; Currie and Neidell 2005). Worker productivity for agricultural laborers has also been shown to decline at high levels of ozone pollution (J. Graff Zivin and Neidell 2012; Chang et al. 2016). Air pollution is also responsible for respiratory and heart-related problems among the general public that cause expensive hospital visits and early death (Schlenker and W. R. Walker 2016). Many other papers (e.g. (Currie, L. Davis, et al. 2015; Isen, Rossin-Slater, and W. R. Walker 2014; Currie and R. Walker 2011)) document the negative consequences of air pollution to human health and the benefits from reductions through government action. The EPA has estimated that the value of the reductions in air pollution caused by the CAA totals $1.3 trillion/year in 2010 (Environmental Protection Agency 2011). To date, all CAA studies have found that the benefits are an order of magnitude larger than the costs. This suggests that even if some portion of the costs are not counted, the net benefit the regulation remains high.

3.2.3 Coal mining techniques

Low-sulfur coal is usually defined as having below 1% sulfur by weight. High and low-sulfur coal are each mined using different approaches. High-sulfur coal is found in coal seams that are hundreds of feet below the surface. To recover this coal, a variety of techniques that involve underground tunnels are used. This is in contrast to low-sulfur coal, which occurs in larger, less dense coal seams closer to the surface. The most economic method for extracting this coal is through surface mining, where coal is exposed by removing the earth above the coal seam. Depending on the characteristics of the terrain, this can be done using strip mining, contour mining or mountaintop removal mining. In 1994, before Title IV was implemented, 39% of coal mined in the U.S. was underground.

The differences in mining approaches can lead to different scale mining operations. Low-sulfur surface mines typically require larger scale operations to make them economic due to the less energy dense coal they mine and the earth-moving capital required. In contrast, underground high-sulfur coal mines are usually smaller operations. In 1994, the average underground mine produced 57% as much coal as a surface mine by volume. High-sulfur mines are also thought to be more flexible in their production. P. L. Joskow (1987) argues that low-sulfur mines, by virtue of their scale and capital requirements, are "lumpier" in their ability to adjust production. High-sulfur mines, by contrast, have less economies of scale and can can more easily expand or contract production by changing capital deployment.

3.3 Data

For the analysis I use data from the Mine Safety and Health Administration (MSHA), which has quarterly data on coal mine production, employment and wages. This data are reported directly to MSHA by mine operators using Form 7000-2. I aggregate the mine-level data to the county level for analysis. From the Bureau of Labor Statistics, I use the Quarterly Census of Employment and Wages. The dataset reports quarterly employment, wages and establishment count at the county level by industry. I use the data which is classified using the Standard Industrial Classification (SIC) system.[1] I take county level annual population from the Census Bureau's County Intercensal Dataset.

To proxy for the sulfur content of the coal in a given county, I use data from Federal Energy Regulatory Commission (FERC) Form 423. Form 423 reports monthly deliveries of fossil fuels to electricity generators around the U.S. For coal deliveries, it reports many statistics including the county of origin, quantity delivered, and sulfur content. Coal delivered to power plants is the relevant subset of coal production for this analysis, because that is what was directly impacted by Title IV. To calculate the average county level sulfur content, I take the quantity weighted average of a given county's coal production for the 5 years before Title IV went into effect (1990-1994). This approach gives an imperfect measure of the sulfur content of the coal in a county because there can be multiple mines in a given county, each with different sulfur contents.

The analysis in this paper focuses on the coal-producing counties in the Appalachian region of West Virginia, Kentucky, Ohio and Virginia.[2] In this region, there are both high and low-sulfur coal resources. I classify 23 of the 92 coal-producing counties in the Appalachian region studied as producing low-sulfur coal based on the FERC Form 423 data. The sulfur content in the high and low-sulfur counties

[1] I use the SIC classified data in this version of the paper. Future analysis will use North American Industry Classification System (NAICS) coded employment data.

[2] Tennessee and Pennsylvania coal country was omitted because it not contiguous with the Appalachian coal region studied.

averages 1.85% and .85% by weight respectively. Figure 3.1 shows a histogram of the sulfur content of the relevant counties. Low-sulfur counties produce more coal than higher sulfur counties. This is due to the different scale of high versus low-sulfur coal mines as discussed in the previous section.

3.4 Empirical strategy

3.4.1 Natural experiment in coal demand

The empirical approach in this paper exploits the decrease in demand for high-sulfur coal that results from Title IV of the CAA amendments of 1990. The exogenous shift in demand affected only a subset of coal-producing Appalachian counties. To identify the impacts of Title IV, I compare outcomes in counties with high-sulfur coal to neighboring counties with low-sulfur coal resources.

Figure 3.2 shows a map of the region studied. It includes all coal producing counties in West Virginia, Ohio, Kentucky, Maryland and Virginia.[3] The shaded counties are those with significant coal deposits that were active during this time period. The darker blue counties have a lower sulfur content. Figure 3.3 shows the same map with a binary indicator for counties with high-sulfur coal. Both maps show that the sulfur content changes across the region, with high-sulfur counties located close to low-sulfur neighbors.

The empirical approach allows for the comparison of similar counties with different sulfur contents. Table 3.1 shows the summary statistics for a number of characteristics across the high and low-sulfur regions during the period before Title IV was enacted. The table shows that the high and low-sulfur counties have similar populations and non-coal employment. One main difference in the non-coal sectors is that the high-sulfur counties have a larger manufacturing sector than the low-sulfur counties. In the coal industry, the two regions differ significantly on the sulfur content, number of coal employees and quarterly coal production. This dissimilarity is due to the different mining techniques used for high and low-sulfur coal discussed in Section 3.2.3. Low-sulfur coal, by virtue of having larger mining operations, has larger production and employs more workers per county.

The impacts of Title IV on coal prices are hard to measure with the available data. Figure 3.4 shows the publicly available pricing data split between the high and low-sulfur counties in the sample. The figure shows some divergence in the price after Title IV is enacted, with high-sulfur coal becoming relatively less valuable.

[3]Counties in Tennessee and Pennsylvania were not included because they are not contiguous with the group of counties examined.

Unfortunately, this data is not available before 1994. I instead turn to data on quantity produced to show how Title IV impacted the Appalachian coal industry.

The total coal produced in this region is split relatively evenly between high and low-sulfur coal before the implementation of Title IV of the CAA. Figure 3.5 shows the total coal production by quarter from 1989-2000. The vertical black line indicates when Title IV of the CAA went into effect. After this date, coal production in high-sulfur regions dropped relative to the low-sulfur regions.

Figure 3.6 shows the total employment in the coal industry in this region over time. Before the implementation of Title IV of the CAA, the high and low-sulfur regions employed a similar number of workers. After the regulation went into effect, employment in high-sulfur coal counties dropped relative to low-sulfur counties. This figure suggests that the reduction in coal production due to Title IV had an impact on employment in the coal sector.

One similarity in the coal industry across high and low-sulfur counties is the ratio of coal mined to employees before the implementation of Title IV. Figure 3.7 shows this ratio over time, which is similar in the pre period and diverges after the implementation of the regulation. The following section investigates the changes in outcomes after Title IV was implemented using a regression analysis.

3.4.2 Regression framework

To identify the impacts of changing coal demand on employment, I leverage the exogenous impact of Title IV on the relative demand for high and low-sulfur coal. I estimate the effects of the regulation using the following equation:

$$Y_{it} = \beta_1 + \beta_2 High\ Sulfur_i \times Post_t + \beta_3 Population_{iy} + \zeta_t + \gamma_i + \theta_{st} + \epsilon_{it} \tag{3.1}$$

Y_{it} represents the outcome of interest in county i and quarter of sample t in logs. $High\ Sulfur_i$ is an indicator for county i having a sulfur content above 1% before 1995. $Post_t$ is an indicator for the implementation of Title IV, which turns to 1 after January 1, 1995. The interaction of $High\ Sulfur_i$ and $Post_t$ is the treatment indicator. $Population_{iy}$ controls for county level population in year y. ζ_t is a quarter of sample fixed effect, which absorbs any contemporaneous impacts to all coal counties. γ_i is a county fixed effect that absorbs time-invariant county characteristics. θ_{st} is a state time trend, which controls for any trends in outcomes at the state level. ϵ_{it} is the error term, which is clustered at the county level.

The identifying assumption is that the high and low-sulfur counties have parallel trends in the pre-period. It is not possible to prove this assumption, but examining pre-period trends can be informative. Figures 3.5, 3.6 and 3.7 show that total coal production and employment are at similar levels before Title IV was implemented.

The high and low-sulfur counties are also similar in their non-coal industry makeup. Figure 3.8 shows the average county level employment in the high and low-sulfur counties. It shows that before Title IV, employment was increasing in both high and low-sulfur counties at a similar rate.[4] The same parallel trends can be seen in wages in Figure 3.9, which shows the average weekly nominal wage by high and low-sulfur counties. Before the implementation of Title IV, they were increasing at a similar rate for both high and low-sulfur counties.

The balance in observable characteristic trends suggests that high and low-sulfur counties are similar along most dimensions. The main area where they differ is in the size of the coal industry, where low-sulfur coal counties have higher production and employment per county than their high-sulfur neighbors. This difference in county level industry sizes means that the impacts estimated using Equation 3.1 must be interpreted in this context. The estimated coefficients reflect what happens when there is a relative shift in demand from high to low-sulfur coal.

3.5 Results

I use Equation 3.1 to identify the impacts of Title IV on high-sulfur coal production, employment and wages in both the coal and non-coal sectors in the Appalachian region. Table 3.2 shows the impact of Title IV on coal production in high-sulfur regions compared to low-sulfur regions in the five years after implementation. Column 1 shows a significant reduction of 20% in tons per quarter mined in the high-sulfur counties compared to the low-sulfur counties. Column 2 estimates the impacts in levels, and shows that Title IV is associated with a 528 thousand ton reduction in output per quarter. These results show that Title IV resulted in a significant drop in coal production in the Appalachian region as seen in Figure 3.5. It is important to interpret these results as the difference in production between high and low-sulfur counties. If, for example, low-sulfur coal production increased due to the Title IV regulations, then the estimated impact reflects both this increase and the decrease in the high-sulfur counties.

Table 3.3 shows the estimated impacts of Title IV on workers in the coal industry. Column 1 shows that high-sulfur coal production dropped by 14% after the regulation went into effect. This impact is similar to the drop in total coal production and reflects the shift in employment seen in Figure 3.6 after Title IV is implemented.

[4] The large changes in employment for low-sulfur counties in 1996 and later may be a data quality issue. Further investigation is needed to resolve those shifts.

Column 2 estimates how the hours worked in the coal industry changed as a result of the regulation. It shows a 16% decline, which is similar to the total decline in employment. Column 3 tests if the coal industry readjusted the hours worked per employee after the regulation. The results show a small and insignificant change, which suggests that the coal industry primarily adjusted its output by firing works, not by reducing their hours.

One important question is do the high-sulfur counties adjust their business practices after the implementation of Title IV? Figure 3.7 shows that both the high and low-sulfur counties were increasing their tons of coal produced per employee at a steady rate in the pre period. One result of the reduction in tons mined at high-sulfur mines could be an increase in mining efficiency. Borenstein and Farrell (2007) found that gold-mining firms accumulate inefficient "fat" when times are good, and then scale back production and cut fat when commodity prices are low. An analogous situation could happen in the coal industry. Mines, facing lower demand for their product, could contract and mine the lowest cost coal available, leading to an increase in efficiency. Conversely, the long-run reduction in the value of high-sulfur coal could reduce the incentive to invest in efficient production. Table 3.4 tests this outcome by looking at the impact of Title IV on tons of coal produced per employee. Column 1 shows that there is a 7% reduction in coal produced per worker after 1995. Column 2 tests the impact using total hours worked, showing that there is a 5% reduction in that measure with a p value of .064.

Figure 3.7 illustrates what may be causing these changing labor input ratios. Before Title IV, both the high and low-sulfur counties were increasing their ratio of output to employees at similar and steady rates. The low-sulfur counties continued this throughout the time period, but the high-sulfur counties level off after the regulation. This suggests that high-sulfur counties are not adjusting to the lower demand by increasing their efficiency of production with respect to labor. One interpretation is that production in high-sulfur counties declined more quickly than the number of employees, which would produce this impact. The two year lag after the implementation of Title IV makes this explanation less likely. Another interpretation is that both high and low-sulfur counties were investing in increased efficiency before Title IV, which can be seen in the increasing ratio of tons/hours worked even as the total number of employees in the sector dropped. After Title IV, the high-sulfur counties may have reduced their investment in efficiency, which can be seen in the ratio flattening out and slightly declining after 1997. It is not possible to separate these explanations with the current data, but it shows that Title IV did have significant impacts on both the size and composition of the high-sulfur coal industry.

The previous results show that the coal industry in high-sulfur counties was significantly impacted by the implementation of Title IV. The next question is what happens to the overall economy in high-sulfur counties after the regulation? Are the negative consequences limited to the coal industry or do they spillover to non-coal sectors? Table 3.5 estimates the impact on employment in all sectors including the coal industry. Column 1 shows that there is not a significant change in overall employment due to Title IV. I can reject a 5.9% decrease in employment using the 95% confidence interval. Column 2 shows that average wages across all industries are also not changing as a result of the regulation. This impact is estimated more precisely than the employment change, allowing me to reject a 3% decrease in wages.[5] Figure 3.9 supports this result and shows that the wages in low-sulfur counties are higher by around $30-$50/week, this amount does not change after Title IV.

3.6 Discussion

The results in the previous section suggest that Title IV had a negative impact on the coal industry in the high-sulfur counties. Production, employment and hours worked all dropped in high-sulfur counties relative to their low-sulfur neighbors. High-sulfur counties may also have reduced their investment in improved labor efficiency, as illustrated in Figure 3.7.

The impacts of Title IV appear to be limited to the coal sector. Overall employment and wages in the high-sulfur counties did not change as a result of the regulation, which has a number of implications for the interpretation of the costs of Title IV. First, while the regulation did hurt the high-sulfur coal industry, it does not detectably depress employment in the whole region. Opponents of environmental regulation sometimes argue that regulation not only hurts the industry it targets, but it creates ripple effects through the economy as the industries that support the coal industry also lose out. Coal was only 3.3% of total employment before section IV was implemented in high-sulfur counties in Appalachia, but it is thought of as one of the economic divers of the region.[6] The results in this paper show that the overall impacts were not large enough to detectably affect employment throughout the region.

There are a number of important caveats to consider when interpreting the results. First, these impacts are estimated over a short time horizon in the five years after the implementation of Title IV. It is possible that over a longer period the overall impacts will be larger. Second, as discussed previously, the results reflect the relative differences between the high and low-sulfur regions in Appalachia after Title IV. If

[5]Both of these lower bounds are likely overestimates of the potential impacts because the estimation strategy reflects increases in employment in the low-sulfur counties.

[6]I use the West Virginian GDP as a proxy for the Appalachian economy because it makes up the majority of the counties studied.

Title IV caused an increase in low-sulfur employment and a decrease in high-sulfur county employment, then the estimates reflect both of these shifts. Third, some of the county-level employment data are omitted from the publicly available non-coal employment data for confidentiality reasons. This happens in small counties where information about individual establishments could be recovered from the aggregate data. Further work will be needed to appropriately use this data for analysis that deals with individual non-coal sectors.

3.6.1 Policy implications

The concentrated impact of Title IV on the coal industry suggests that government intervention may be effective at reducing the costs of the policy for this limited population. If it is only the coal miners that are hurt by the regulation, then targeted financial compensation for those who lost their jobs may be able to ameliorate some of the negative consequences. To my knowledge, there has not been funding directed at coal miners that have lost their jobs due to the Clean Air Act regulation. The Appalachian Regional Commission, founded in 1965 as a federal-state partnership to address poverty in the greater Appalachian region, has assisted in some coal communities, but that is not its primary goal. The agency covers 13 states and is relatively small (less than $100 million/year), meaning its limited resources are not able to directly compensate coal miners in these regions.

Coal miners that lose their jobs due to regulation are limited to the same programs as all other unemployed Americans. Disabled coal miners suffering from Black Lung are eligible for benefits through the Federal Black Lung Program, but this does not help all regulation-affected workers. One model to help unemployed coal miners due to regulation is the Federal Trade Adjustment Assistance (TAA) program. TAA provides compensation and retraining to U.S. workers that lost their job due to trade. There are many shortcomings to this approach, including how a worker proves they lost their job to trade (or regulation) and the benefits of retraining programs. (Baicker and Rehavi 2004). A similar program for coal workers hurt by regulation could reduce some of the employment consequences of environmental regulation. To avoid some of the shortcomings of TAA, the program could be designed as a cash transfer to be used at the recipients discretion.

3.6.2 Future directions

There are a number of additions that will be made to the analysis in future drafts. First, I intend to conduct the analysis using sector-level non-coal employment data from the Bureau of the Economic Analysis. This will allow me to see if certain sectors gained or lost employment as a result of the drop in coal sector employment. The results will help inform if coal miners can easily move to different sectors, and which sectors they choose. Second, I intend to include non-coal producing counties in Appalachian region as potential controls for robustness checks. This approach would allow me to compare coal-producing counties to similar non-coal counties, giving a second estimate of the non-coal sector employment impacts.

Third, I hope to have better price data for high and low-sulfur coal during the time period studied. Unfortunately, new data confidentiality policies at the Energy Information Administration have made it challenging to access FERC 423 data that include mine level price information. Fourth, I plan to estimate the total costs of the employment dislocation to affected workers in the form of forgone wages. This exercise will require a number of assumptions, but it will provide an overall cost estimate for the employment impacts of Title IV on coal workers. Finally, I hope to look at how the decreased demand for high-sulfur coal impacted mining practices using data from the Mine Health Safety Administration. I plan to test if, for example, there more citations or injuries at high-sulfur mines that could result from cost cutting measures.

3.7 Conclusion

Environmental air quality regulations are responsible for large gains in human health and productivity. The costs of this regulation, while usually much smaller than the benefits, are frequently concentrated in a small number of regulated industries. This paper considers the costs of Title IV of the Clean Air Act, which sets limits on total sulfur emissions from power plants, on an upstream sector not directly regulated – the coal mining industry. Understanding the indirect impacts to upstream industries is important to calculating the full costs of the regulation. The existing literature has focused on the industries and areas directly regulated, but have not covered upstream industries. The upstream costs of the CAA have recently become relevant as a 2016 ruling by a federal district court judge required the EPA to explicitly consider the employment costs to the coal industry.

This paper measures a portion of the employment impacts of Title IV of the CAA on the coal industry in the Appalachian region of the Eastern U.S. using a natural experiment introduced by the regulation. I find that coal production in high-sulfur counties in Appalachia dropped by 20% compared to low-sulfur counties after the implementation of Title IV. This led to a drop in coal worker employment in high-sulfur coal counties of 16%. The results show that industries not directly regulated by the CAA can be significantly impacted. I then examine non-coal employment, finding no evidence of spillovers to employment in other sectors in the Appalachian region. Overall employment and wages in the high-sulfur counties do not change relative to the low-sulfur counties, suggesting the employment costs are limited to the coal industry.

The findings in this paper have important policy implications. First, I provide the first well-identified evidence that Title IV of the CAA impacts upstream sectors that are not directly regulated by the policy. The size of the impact to the coal industry shows that official analyses should consider both the direct and upstream costs of environmental regulation. Second, the findings suggest that it is possible to reduce some of the costs of Title IV by directly compensating or retraining workers displaced from the coal industry. Further research will be required to better understand the magnitude of these costs, and what happens to the displaced coal workers in the long run. This research can also be informative about the consequences of continued coal mining employment decline that is driven by market forces (e.g. low natural gas prices) or future regulation (e.g. the Clean Power Plan). The lack of spillovers from the coal industry to non-coal sectors suggest that the long-term decline of the coal industry in Appalachia may not have large impacts on the region's economic health. Further research is required to better understand the future of the Appalachian coal industry and its role in the regional economy.

Figures

Figure 3.1: Distribution of County Level Sulfur Content

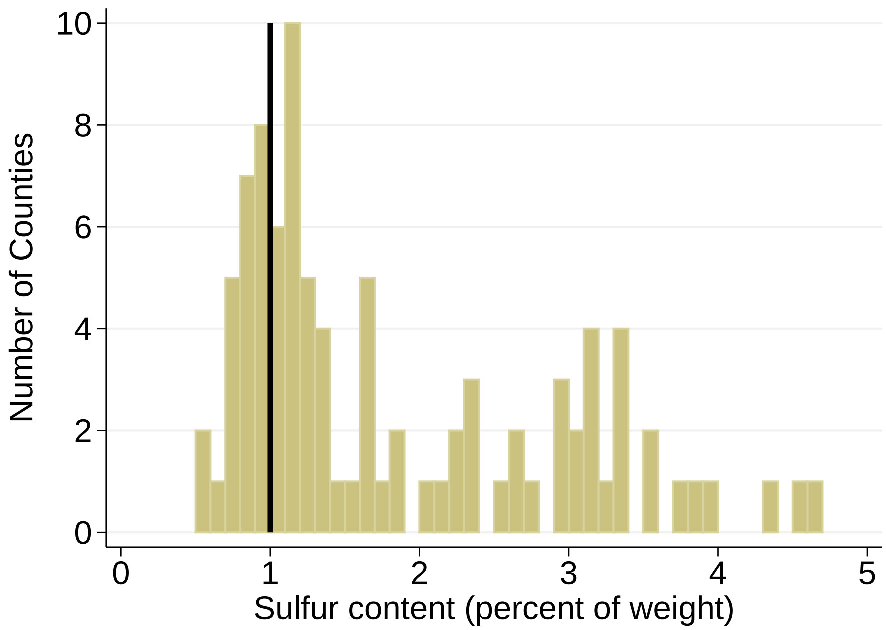


Figure 3.2: Sulfur Content of Appalachian Counties in Study

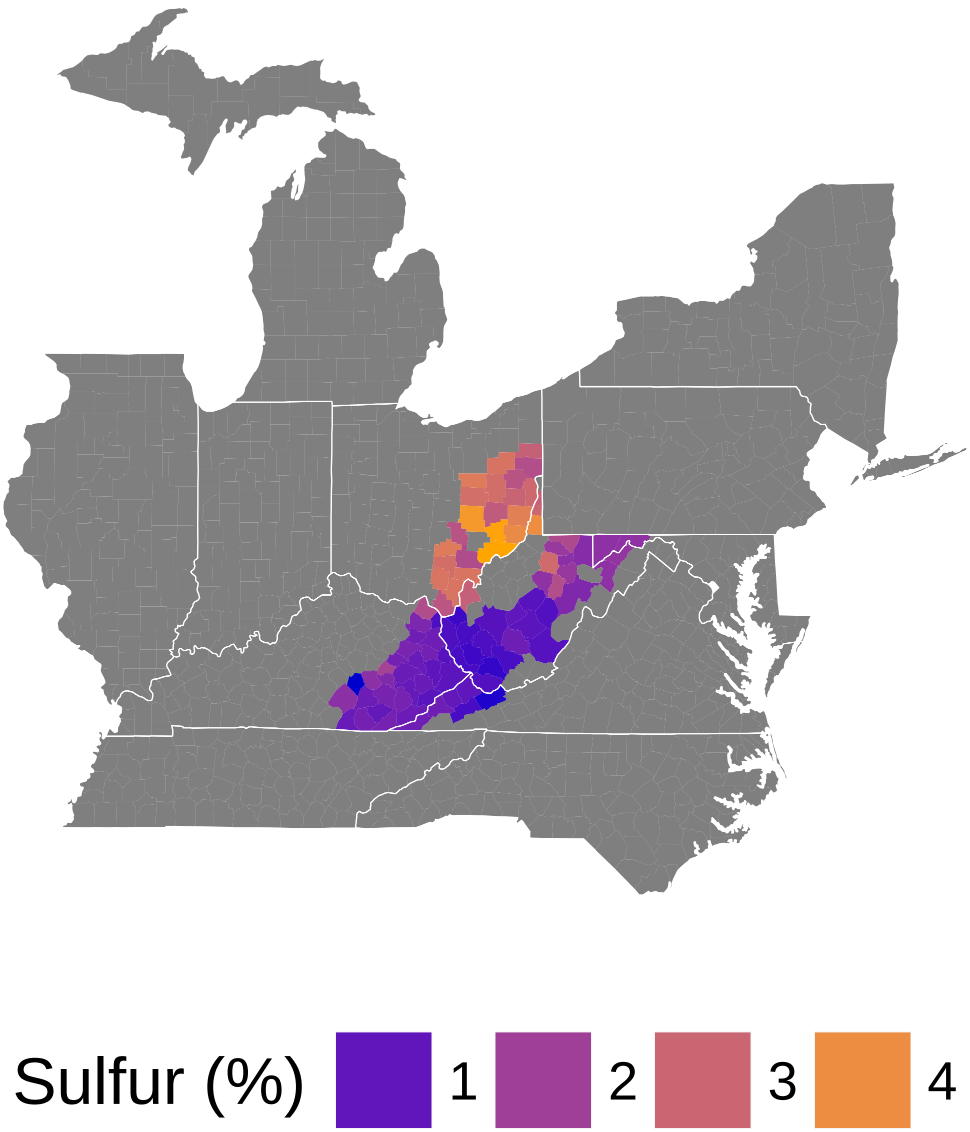


Figure 3.3: High sulfur (>1%) Designation of Appalachian Counties

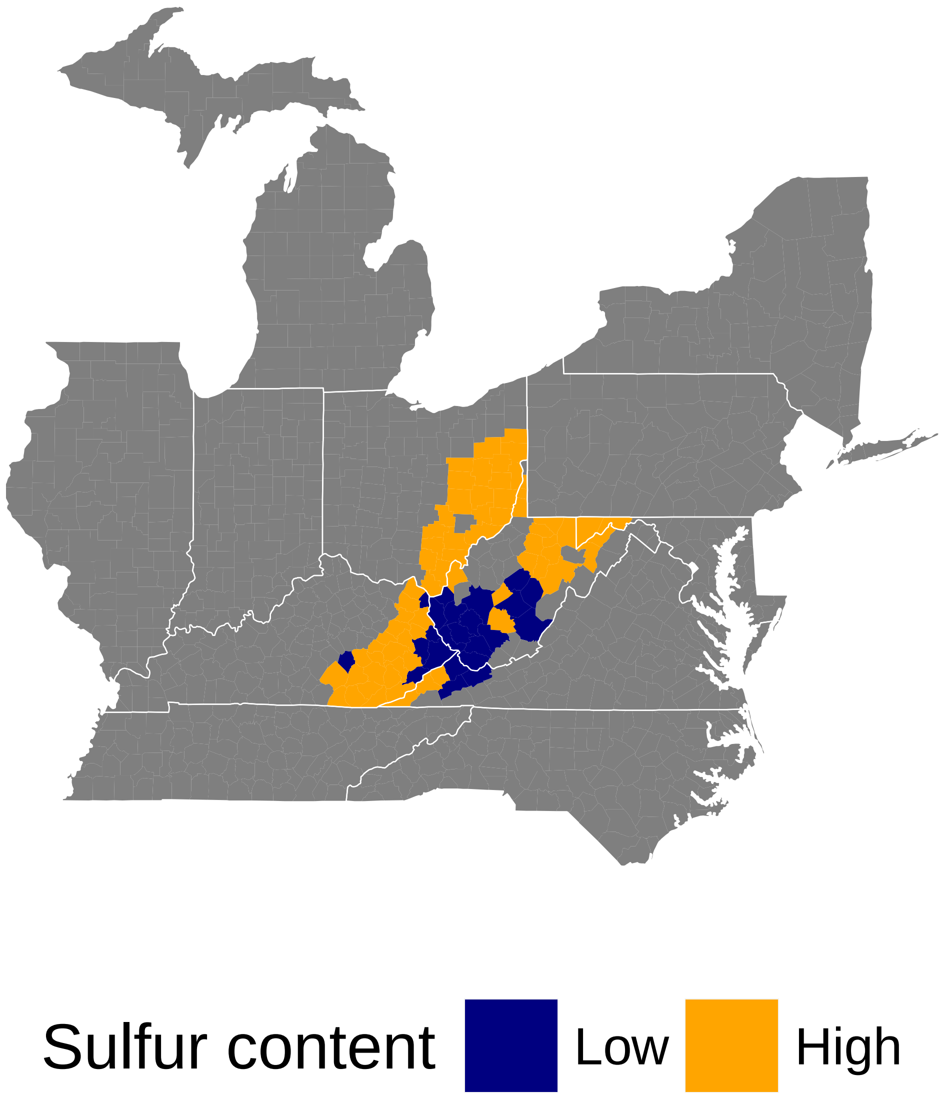


Figure 3.4: Average Appalachian Coal Price by Sulfur Content

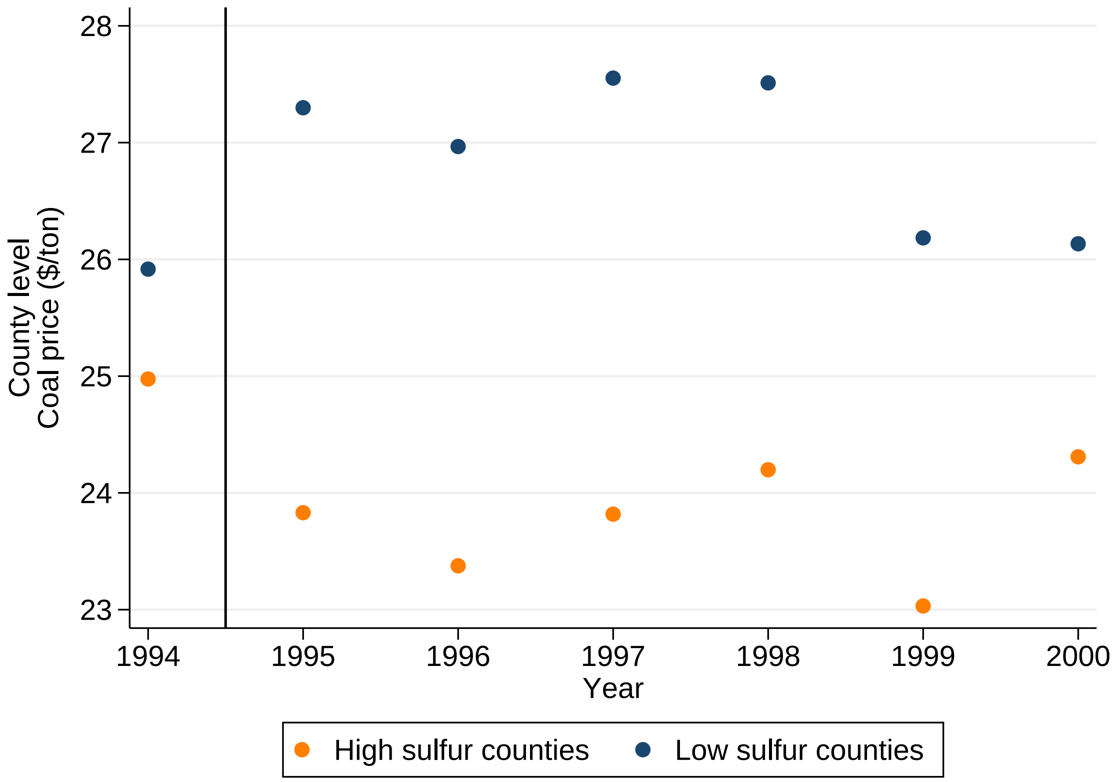


Figure 3.5: Total Appalachian Coal Production by Sulfur Content

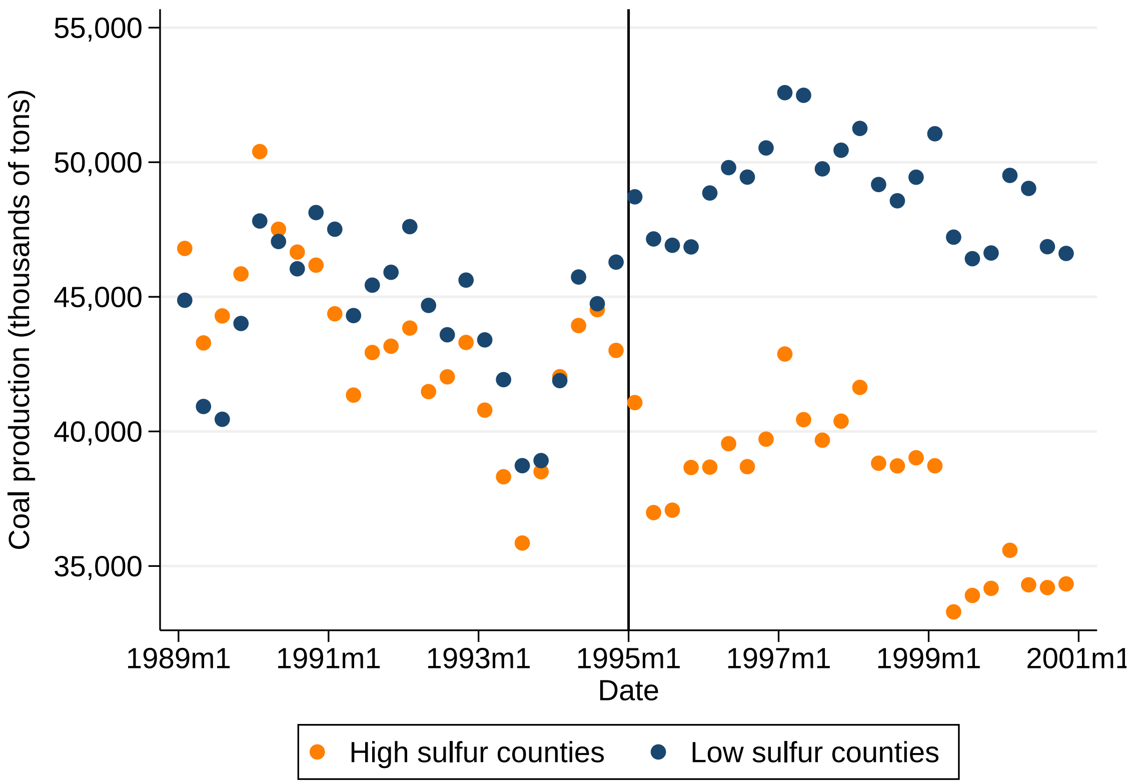


Figure 3.6: Total Appalachian Coal Employment by Sulfur Content

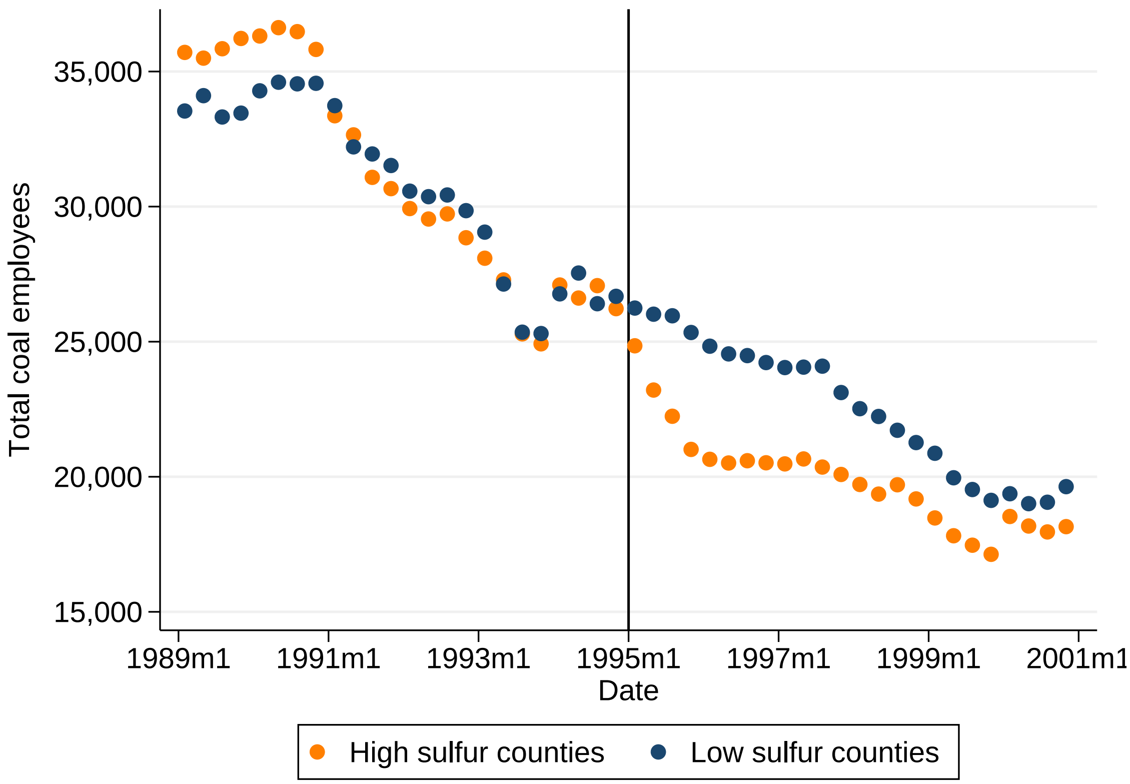


Figure 3.7: Coal Production Labor Efficiency by Sulfur Content

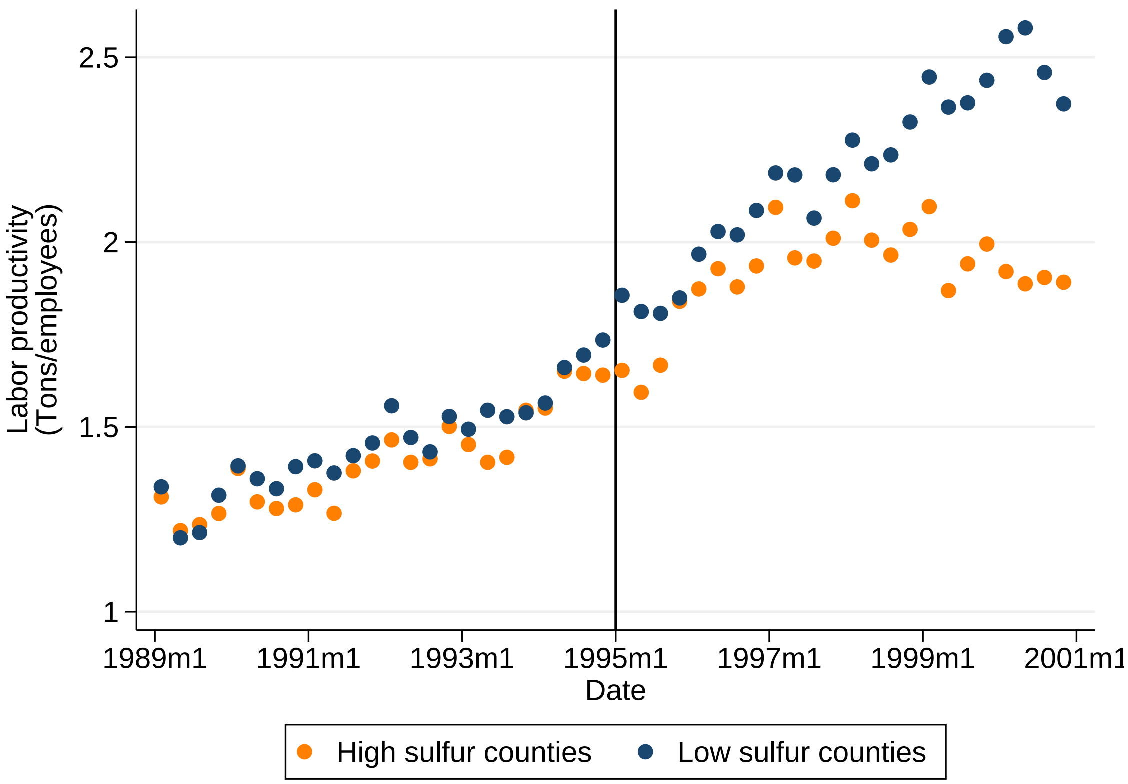


Figure 3.8: Average Appalachian County Level Employment (all sectors)

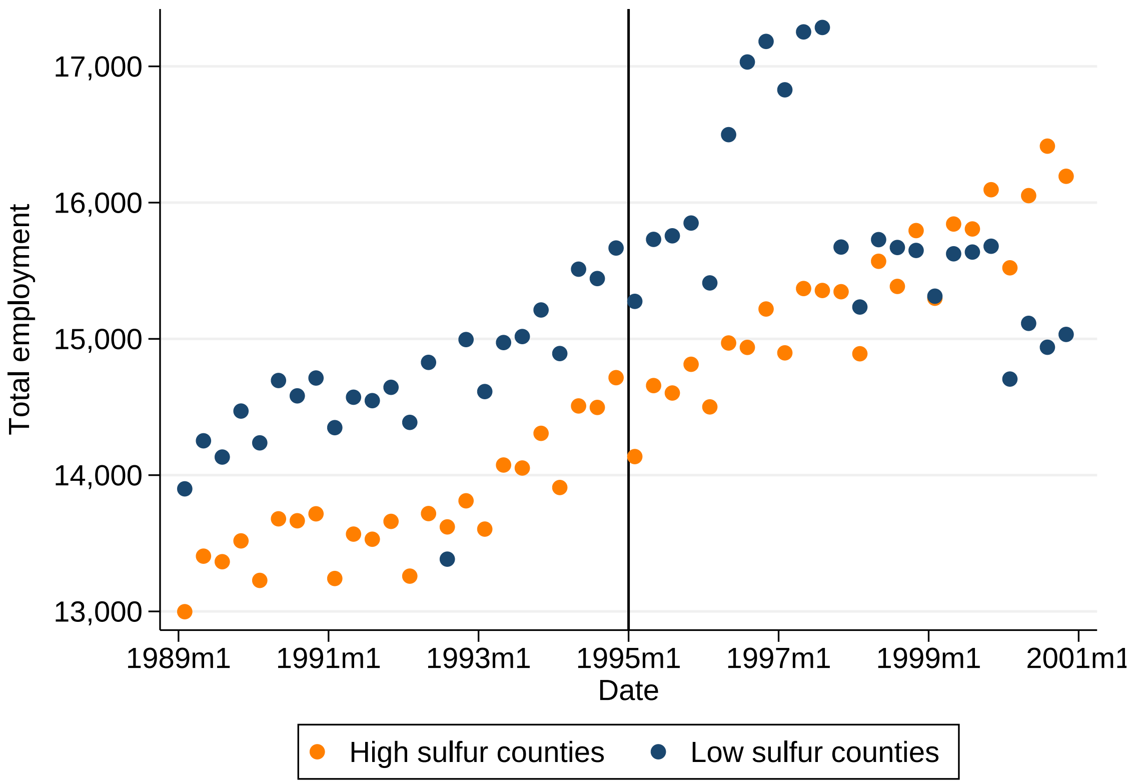
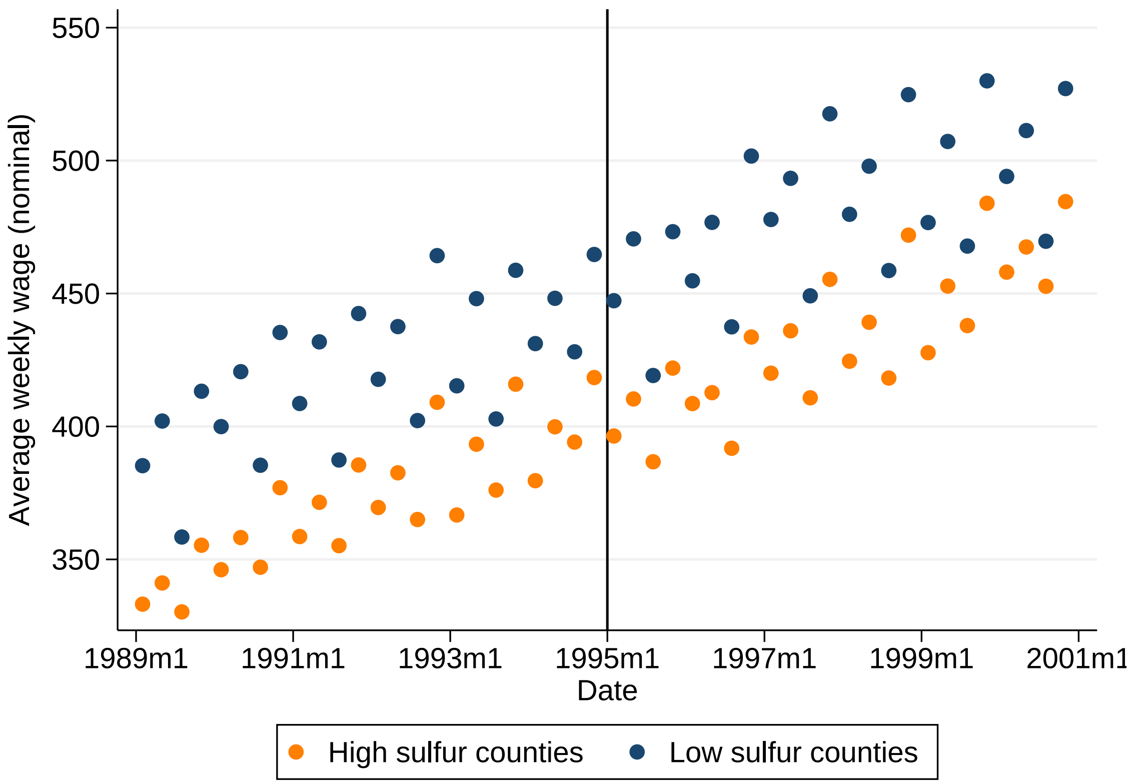


Figure 3.9: Average Appalachian County Level Wage (all sectors)



Tables

Table 3.1: Summary Statistics by Sulfur Content

Variable	High-Sulfur counties	Low-Sulfur counties	P value of difference
Coal sulfur content (%)	2.19	.821	.00 ***
Coal employees	459	1338	.00 ***
Quarterly coal production	640	1933	.00 ***
County population	42547	43040	.87
All employed	13733	14669	.67
Construction employment	545	599	.13
Manufacutring employment	2606	1458	.03 **
Public sector employment	630	839	.93
Wholesale employment	606	721	.15
Retail employment	2818	2903	.44
Finance employment	518	607	.46
Service employment	3015	3265	.32

Table 3.2: Impact of Title IV on High-Sulfur Coal County Production

	(1) ln(coal production)	(2) Coal production
High sulfur county	-0.2195**	-528.4882**
	(0.0861)	(264.2537)
Percent change	-19.70	
High sulfur average production	625	625
High sulfur counties	68	68
Low sulfur counties	23	23

Notes: Dependent variable is quarterly coal production in thousands of tons. All regressions include county fixed effects, quarter of sample FE and state specific time trends. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 3.3: Impact of Title IV on High-Sulfur Coal County Employment

	(1) ln(coal employment)	(2) ln(hours worked)	(3) ln(hours/employees)
High sulfur county	-0.1521**	-0.1787**	-0.0266***
	(0.0765)	(0.0797)	(0.0099)
Percent change	-14.11	-16.37	-2.62
High sulfur county employment	392	392	392
High sulfur counties	68	68	68
Low sulfur counties	23	23	23

Notes: Dependent variable is the log of quarterly coal employment in Column 1. Dependent variable is the log of quarterly hours worked in coal sector in Column 2. Column 3 is the log of the ratio of hours worked to employees per county each quarter. All regressions include county fixed effects, quarter of sample FE and state specific time trends. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 3.4: Impact of Title IV on High-Sulfur Coal County Labor Productivity

	(1) ln(tons/coal employees)	(2) ln(tons/hours worked)
High sulfur county	-0.0727**	-0.0469
	(0.0363)	(0.0341)
Percent change	-7.01	-4.59
High sulfur county employment	392	392
High sulfur counties	68	68
Low sulfur counties	23	23

Notes: Dependent variable is the log of quarterly coal production divided by hours worked in Column 1. Dependent variable is the log of quarterly coal production divided by coal employees in Column 2. All regressions include county fixed effects, quarter of sample FE and state specific time trends. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 3.5: Impact of Title IV on High-Sulfur Coal County Total Employment

	(1) ln(all employment)	(2) ln(all wages)
High sulfur county	0.0361	0.0256
	(0.0329)	(0.0193)
High sulfur county employment	14,489	14,489
High sulfur counties	68	68
Low sulfur counties	23	23

Notes: Dependent variable in Column 1 is the log of quarterly employment in all sectors. Dependent variable in Column 2 is the log of average quarterly wage by county for all sectors. All regressions include quarter of sample FE and county specific time trends. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Bibliography

Abadie, Alberto and Sebastien Gay (2006). "The Impact of Presumed Consent Legislation on Cadaveric Organ Donation: A Cross-Country Study." In: *Journal of Health Economics* 25.4, pp. 599–620. DOI: `10.1016/j.jhealeco.2006.01.003`.

Arnott, Richard, André de Palma, and Robin Lindsey (1993). "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand." In: *The American Economic Review* 83.1, pp. 161–179.

Baicker, Katherine and M. Marit Rehavi (2004). "Policy Watch: Trade Adjustment Assistance." In: *The Journal of Economic Perspectives* 18.2, pp. 239–255. DOI: `10.1257/0895330041371196`.

Bell, Shannon Elizabeth and Richard York (2010). "Community Economic Identity: The Coal Industry and Ideology Construction in West Virginia." In: *Rural Sociology* 75.1, pp. 111–143. DOI: `10.1111/j.1549-0831.2009.00004.x`.

Boiteux, M. (1949). "La Tarification Des Demandes En Pointe: Application de La Théorie de La Vente Au Coût Marginal." In: *place Henri-Bergson.* Translated into English in Boiteux, M (1960) "Peak-Load Pricing." The Journal of Business 33.2, pp. 157-179.

Bollinger, Bryan and Wesley Hartmann (2015). "Welfare Effects of Home Automation Technology with Dynamic Pricing." In: *Working Paper.*

Boomhower, Judson and Lucas W. Davis (2016). "Do Energy Effciency Investments Deliver at the Right Time?" In: *EI @ Haas working paper WP-271.* URL: `http://e2e.haas.berkeley.edu/abstractWP023.html`.

Borenstein, Severin (2005). "The Long-Run Efficiency of Real-Time Electricity Pricing." In: *The Energy Journal* 26.3, pp. 93–116.

— (2007a). "Customer Risk from Real-Time Retail Electricity Pricing: Bill Volatility and Hedgability." In: *The Energy Journal* 28.2. DOI: `10.5547/ISSN0195-6574-EJ-Vol28-No2-5`.

— (2007b). "Wealth Transfers Among Large Customers from Implementing Real-Time Retail Electricity Pricing." In: *The Energy Journal* 28.2, pp. 131–149.

Borenstein, Severin (2012). “Effective and Equitable Adoption of Opt-In Residential Dynamic Electricity Pricing.” In: *Review of Industrial Organization* 42.2, pp. 127–160. DOI: 10.1007/s11151-012-9367-3.

Borenstein, Severin and Joseph Farrell (2007). “Do Investors Forecast Fat Firms? Evidence from the Gold-mining Industry.” In: *The RAND Journal of Economics* 38.3, pp. 626–647. DOI: 10.1111/j.0741-6261.2007.00104.x.

Borenstein, Severin and Stephen Holland (2005). “On the Efficiency of Competitive Electricity Markets With Time-Invariant Retail Prices.” In: *The RAND Journal of Economics* 36.3, pp. 469–493. URL: https://www.researchgate.net/publication/24049308_On_the_Efficiency_of_Competitive_Electricity_Markets_With_Time-Invariant_Retail_Prices.

Brueckner, Jan K. (2002). “Airport Congestion When Carriers Have Market Power.” In: *American Economic Review* 92.5, pp. 1357–1375. DOI: 10.1257/000282802762024548.

— (2005). “Internalization of Airport Congestion: A Network Analysis.” In: *International Journal of Industrial Organization* 23 (7–8), pp. 599–614. DOI: 10.1016/j.ijindorg.2005.03.007.

— (2009). “Price vs. Quantity-Based Approaches to Airport Congestion Management.” In: *Journal of Public Economics* 93 (5–6), pp. 681–690. DOI: 10.1016/j.jpubeco.2009.02.009.

Busse, Meghan R. and Nathaniel O. Keohane (2007). “Market Effects of Environmental Regulation: Coal, Railroads, and the 1990 Clean Air Act.” In: *The RAND Journal of Economics* 38.4, pp. 1159–1179. DOI: 10.1111/j.0741-6261.2007.00130.x.

California Energy Commission (2015). *Estimated Cost of New Renewable and Fossil Generation in California*.

California Public Utilities Commission (2001). *Rulemaking 02-06-001 - Adopting a Business Case Analysis Framework for Advanced Metering Infrastructure*.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011). “Robust Inference With Multiway Clustering.” In: *Journal of Business & Economic Statistics* 29.2, pp. 238–249. DOI: 10.1198/jbes.2010.07136.

Carlson, Curtis et al. (2000). “Sulfur Dioxide Control by Electric Utilities: What Are the Gains from Trade?” In: *Journal of Political Economy* 108.6, pp. 1292–1326. DOI: 10.1086/317681.

CEE, CEE (2014). “CEE Annual Industry Report.” In:

Chang, Tom et al. (2016). “Particulate Pollution and the Productivity of Pear Packers.” In: *American Economic Journal: Economic Policy* 8.3, pp. 141–169. DOI: 10.1257/pol.20150085.

Chay, Kenneth Y. and Michael Greenstone (2003). “The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." In: *The Quarterly Journal of Economics* 118.3, pp. 1121–1167. DOI: 10.1162/00335530360698513.

Choi, James et al. (2004). *Perspectives in the Economics of Aging*. University of Chicago Press.

Correia, Sergio (2016). "REGHDFE: Stata Module to Perform Linear or Instrumental-Variable Regression Absorbing Any Number of High-Dimensional Fixed Effects." In: Statistical Software Components s457874, Boston College Department of Economics, revised 25 Jul 2015. URL: https://ideas.repec.org/c/boc/bocode/s457874.html#biblio.

Currie, Janet, Lucas Davis, et al. (2015). "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings." In: *American Economic Review* 105.2, pp. 678–709. DOI: 10.1257/aer.20121656.

Currie, Janet and Matthew Neidell (2005). "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?" In: *The Quarterly Journal of Economics* 120.3, pp. 1003–1030. DOI: 10.1093/qje/120.3.1003.

Currie, Janet and Reed Walker (2011). "Traffic Congestion and Infant Health: Evidence from E-ZPass." In: *American Economic Journal: Applied Economics* 3.1, pp. 65–90. DOI: 10.1257/app.3.1.65.

Davis, Lucas W., Alan Fuchs, and Paul Gertler (2014). "Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico." In: *American Economic Journal: Economic Policy* 6.4, pp. 207–38. DOI: 10.1257/pol.6.4.207.

Energy Information Administration (2014). *Form EIA-861.*

— (2016). *Monthly Energy Review*. URL: http://www.eia.gov/totalenergy/data/monthly/.

Environmental Protection Agency (2011). *Benefits and Costs of the Clean Air Act 1990-2020*. URL: https://www.epa.gov/clean-air-act-overview/benefits-and-costs-clean-air-act-1990-2020-report-documents-and-graphics.

EPA (2015). *Clean Power Plan: State at a Glance California*. URL: http://www3.epa.gov/airquality/cpptoolbox/california.pdf (visited on 10/15/2015).

Evergreen Economics (2013). "PY2011 Energy Savings Assistance Program Impact Evaluation." In:

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram (2015a). *Are the Non-Monetary Costs of Energy Efficiency Investments Large? Understanding Low Take-up of a Free Energy Efficiency Program*. Working Paper BFI-2015-01. Becker Friedman Institute for Research In Economics. URL: https://ideas.repec.org/p/bfi/wpaper/bfi-2015-01.html.

— (2015b). *Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program*. Working Paper 21331. National Bureau of Economic Research. URL: http://www.nber.org/papers/w21331.

Fowlie, Meredith, Catherine Wolfram, et al. (2015). "Default Effects, Follow-on Behavior and Welfare in Residential Electricity Pricing Programs." In: *Working Paper*.

Gelman, Andrew and Guido Imbens (2014). "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." In: *NBER Working Paper* no. 20405. URL: http://www.nber.org/papers/w20405.

Graff Zivin, Joshua S., Matthew J. Kotchen, and Erin T. Mansur (2014). "Spatial and Temporal Heterogeneity of Marginal Emissions: Implications for Electric Cars and Other Electricity-Shifting Policies." In: *Journal of Economic Behavior & Organization* 107, Part A, pp. 248–268. DOI: 10.1016/j.jebo.2014.03.010.

Graff Zivin, Joshua and Matthew Neidell (2012). "The Impact of Pollution on Worker Productivity." In: *American Economic Review* 102.7, pp. 3652–3673. DOI: 10.1257/aer.102.7.3652.

Greenstone, Michael (2002). "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures." In: *Journal of Political Economy* 110.6, pp. 1175–1219. DOI: 10.1086/342808.

Greenstone, Michael, Elizabeth Kopits, and Ann Wolverton (2013). "Developing a Social Cost of Carbon for US Regulatory Analysis: A Methodology and Interpretation." In: *Review of Environmental Economics and Policy* 7.1, pp. 23–46. DOI: 10.1093/reep/res015.

Holland, Stephen P. and Erin T. Mansur (2006). "The Short-Run Effects of Time-Varying Prices in Competitive Electricity Markets." In: *The Energy Journal* 27.4, pp. 127–155.

Imbens, Guido W. and Thomas Lemieux (2008). "Regression Discontinuity Designs: A Guide to Practice." In: *Journal of Econometrics*. The regression discontinuity design: Theory and applications 142.2, pp. 615–635. DOI: 10.1016/j.jeconom.2007.05.001.

Isen, Adam, Maya Rossin-Slater, and W. Reed Walker (2014). *Every Breath You Take - Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970*. Working Paper 19858. National Bureau of Economic Research. DOI: 10.3386/w19858.

Ito, Koichiro (2015). "Asymmetric Incentives in Subsidies: Evidence from a Large-Scale Electricity Rebate Program." In: *American Economic Journal: Economic Policy* 7.3, pp. 209–37. DOI: 10.1257/pol.20130397.

Ito, Koichiro, Takanori Ida, and Makoto Tanaka (2015). "The Persistence of Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand." In: *NBER Working Paper* no. 20910. URL: http://www.nber.org/papers/w20910.

Ito, Koichiro and James M. Sallee (2014). “The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel-Economy Standards.” In: *NBER Working Paper* no. 20500. URL: http://www.nber.org/papers/w20500.

Jacobsen, Mark R. et al. (2016). *Sufficient Statistics for Imperfect Externality-Correcting Policies*. Working Paper 22063. National Bureau of Economic Research. URL: http://www.nber.org/papers/w22063.

Jessoe, Katrina and David Rapson (2014). “Knowledge Is (Less) Power: Experimental Evidence from Residential Energy Use.” In: *American Economic Review* 104.4, pp. 1417–38. URL: https://ideas.repec.org/a/aea/aecrev/v104y2014i4p1417-38.html.

Johnson, Eric J., Steven Bellman, and Gerald L. Lohse (2002). “Defaults, Framing and Privacy: Why Opting In-Opting Out.” In: *Marketing Letters* 13.1, pp. 5–15. DOI: 10.1023/A:1015044207315.

Joskow, Paul L. (1987). “Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets.” In: *The American Economic Review* 77.1, pp. 168–185.

Joskow, Paul L. and Catherine D. Wolfram (2012). “Dynamic Pricing of Electricity.” In: *The American Economic Review* 102.3, pp. 381–385.

Joskow, Paul and Jean Tirole (2006). “Retail Electricity Competition.” In: *The RAND Journal of Economics* 37.4, pp. 799–815. DOI: 10.1111/j.1756-2171.2006.tb00058.x.

— (2007). “Reliability and Competitive Electricity Markets.” In: *The RAND Journal of Economics* 38.1, pp. 60–84. DOI: 10.1111/j.1756-2171.2007.tb00044.x.

Kahn, Matthew E., Nils Kok, and John M. Quigley (2014). “Carbon Emissions from the Commercial Building Sector: The Role of Climate, Quality, and Incentives.” In: *Journal of Public Economics* 113, pp. 1–12. DOI: 10.1016/j.jpubeco.2014.03.003.

Kahn, Matthew E. and Erin T. Mansur (2013). “Do Local Energy Prices and Regulation Affect the Geographic Concentration of Employment?” In: *Journal of Public Economics* 101, pp. 105–114. DOI: 10.1016/j.jpubeco.2013.03.002.

Koomey, Jonathan G. et al. (1999). “Projected Regional Impacts of Appliance Efficiency Standards for the US Residential Sector.” In: *Energy* 24.1, pp. 69–84. DOI: 10.1016/S0360-5442(98)00065-6.

Lee, David S. and David Card (2008). “Regression Discontinuity Inference with Specification Error.” In: *Journal of Econometrics*. The regression discontinuity design: Theory and applications 142.2, pp. 655–674. DOI: 10.1016/j.jeconom.2007.05.003.

Madrian, Brigitte C. and Dennis F. Shea (2001). "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." In: *The Quarterly Journal of Economics* 116.4, pp. 1149–1187. DOI: 10.1162/003355301753265543.

McKinsey & Company (2009). "Unlocking Energy Efficiency in the US Economy." In:

Mohring, Herbert (1972). "Optimization and Scale Economies in Urban Bus Transportation." In: *The American Economic Review* 62.4, pp. 591–604.

Pacific Gas & Electric (2010). *Advanced Metering Infrastructure Semi-Annual Assessment Report SmartMeter™ Program Quarterly Report.*

Parry, Ian W. H. and Kenneth A. Small (2009). "Should Urban Transit Subsidies Be Reduced?" In: *American Economic Review* 99.3, pp. 700–724. DOI: 10.1257/aer.99.3.700.

Ramnath, Shanthi (2013). "Taxpayers' Responses to Tax-Based Incentives for Retirement Savings: Evidence from the Saver's Credit Notch." In: *Journal of Public Economics* 101, pp. 77–93. DOI: 10.1016/j.jpubeco.2013.02.010.

Reiss, Peter C. and Matthew W. White (2005). "Household Electricity Demand, Revisited." In: *The Review of Economic Studies* 72.3, pp. 853–883. DOI: 10.1111/0034-6527.00354.

Sallee, James M. and Joel Slemrod (2012). "Car Notches: Strategic Automaker Responses to Fuel Economy Policy." In: *Journal of Public Economics*. Fiscal Federalism 96 (11–12), pp. 981–999. DOI: 10.1016/j.jpubeco.2012.06.005.

Schlenker, Wolfram and W. Reed Walker (2016). "Airports, Air Pollution, and Contemporaneous Health." In: *The Review of Economic Studies* 83.2, pp. 768–809. DOI: 10.1093/restud/rdv043.

Steiner, Peter O. (1957). "Peak Loads and Efficient Pricing." In: *The Quarterly Journal of Economics* 71.4, pp. 585–610. DOI: 10.2307/1885712.

Vickrey, William S. (1963). "Pricing in Urban and Suburban Transport." In: *The American Economic Review* 53.2, pp. 452–465.

— (1969). "Congestion Theory and Transport Investment." In: *The American Economic Review* 59.2, pp. 251–260.

Walker, W. Reed (2011). "Environmental Regulation and Labor Reallocation: Evidence from the Clean Air Act." In: *American Economic Review* 101.3, pp. 442–447. DOI: 10.1257/aer.101.3.442.

— (2013). "The Transitional Costs of Sectoral Reallocation: Evidence From the Clean Air Act and the Workforce*." In: *The Quarterly Journal of Economics* 128.4, pp. 1787–1835. DOI: 10.1093/qje/qjt022.

Wolak, Frank A. (2007). "Residential Customer Response to Real-Time Pricing: The Anaheim Critical Peak Pricing Experiment." In: *Center for the Study of Energy Markets Working Paper* no. 151. URL: http://escholarship.org/uc/item/3td3n1x1.

Wolak, Frank A. (2010). "An Experimental Comparison of Critical Peak and Hourly Pricing: The PowerCents DC Program." In: *Working Paper*. URL: https://www.researchgate.net/publication/228978483_An_Experimental_Comparison_of_Critical_Peak_and_Hourly_Pricing_The_PowerCents_DC_Program.

— (2011). "Do Residential Customers Respond to Hourly Prices? Evidence from a Dynamic Pricing Experiment." In: *American Economic Review* 101.3, pp. 83–87. DOI: 10.1257/aer.101.3.83.

Appendix A

All the robustness checks I couldn't fit in my job market paper

A.1 Peak pricing program details

The PG&E peak pricing program was created in 2008. In 2010 and 2011, the California Public Utilities commission issued decision 10-02-032 and 11-11-008 respectively, which ordered that small and medium C&I customers be placed on opt-out peak pricing once they had sufficient hourly billing data available.[1] Prior to these decisions, peak pricing was structured as an optional opt-in program, but enrollment generally was low. The first wave of small and medium C&I customers were placed on the peak pricing tariff in November 2014. Customers were notified of their enrollment via mail and e-mail, and were easily given the ability to opt out at any time via a simple web interface. Figure A.1 shows the letter that was sent to all establishments that were opted in October 2014 that includes clear directions on how to opt out of the program through their PG&E online billing interface.

Event days are chosen using the day ahead maximum temperature forecasts at 5 National Weather Service (NWS) stations located in the inland regions of California.[2] When the average of maximum temperatures across all 5 stations exceeds a trigger temperature, typically 96 or 98 degrees, an event day is called for the following day. Appendix Table A.1 lists all of the event days between 2013 and 2015. The second column shows the forecasted average maximum temperature from the 5 NWS weather stations. The trigger temperature is based on historical weather patterns and is

[1] Large PG&E customers, defined as having demand charges above 200kW/month, were transitioned to opt-out peak pricing starting in 2010.

[2] The five stations used for the average are Red Bluff (KRBL), Sacramento (KSAC), Fresno (KFAT), Concord (KCCR), and San Jose (KSJC).

adjusted every 15 days throughout the summer. The trigger temperature starts at 96 degrees earlier in the summer and adjusts based on how many event days are called. For example, if many event days are called in the first part of the summer, the trigger will be revised upward to save the remaining event days for the hottest days.[3] Monday event days must be called the Friday before to give businesses that may be closed on the weekend time to prepare.[4]

A.2 Data appendix

For this analysis I combined PG&E data from many sources to create the final dataset for analysis. Interval billing data requires extensive cleaning and a number of assumptions to collapse it down to the establishment level. The following sections detail the process for how the final dataset was constructed.

A.2.1 Interval usage data

I was initially given interval usage data for a large sample of non-residential non-agricultural establishments for 2010-2014. From this dataset, I requested the 2015 data for the subset of establishments that I use in my analysis. This gave me a dataset of 2014 and 2015 consumption for 54,458 accounts that I proceeded to clean. The usage data is collected from establishments at the 15 minute level, which I aggregated to the hourly level for analysis. From this point forward, the sample I discuss refers to 2pm-6pm for all summer non-holiday weekdays (June-October) in 2014 and 2015.[5]

Using this data, I created a balanced panel of establishments that did not move or change ownership during the summer of 2014 or 2015. This step dropped 10,231 establishments, leaving 44,227 in the balanced panel. I required that at least 23% of the establishments have non-zero usage over the 2014-2015 sample. This dropped an additional 4,603 establishments leaving 39,624. I dropped all establishments that never consumed 1 kWh in any peak hour and I dropped establishments that consumed less than 800 kWh/month during the summer of 2014.

[3]The goal of this approach is to be more stringent earlier in the summer as there is more uncertainty over the remaining weather of the summer.

[4]My conversations with the staff involved with calling the peak pricing program is aware of this shortcoming, but it is a requirement for how the program must be run.

[5]I drop all establishments that voluntarily opted into the peak pricing program. More details on this can be found in section A.4.4.

These requirements are to remove any smaller usage meters that may not be directly associated with an establishment's main electricity usage.[6] For example, there are cases where a meter was installed to power a single light in a strip mall, but was not associated with any of the establishments there. In some cases, the light was paid for by the owner of the strip mall and not a business establishment, making too small it too small to consider in this analysis. The 1 kWh/hour restriction dropped 5,145 accounts and the 800 kWh/month restriction dropped another 14,224, leaving a dataset of 19,318. Establishments that consumed more than 10,000 kWh/month in the summer of 2014 were also dropped due to their large size and the likelihood that they would graduate to a higher tariff in the near future. Only 272 establishments met this criteria. Despite the 800 and 10,000 kWh/month restrictions dropping a large number of customers, the remaining customers account for 82% of the load in this group.

Industry classifications in the form of NAICS codes were provided for 89.2% of the establishments in the sample. The NAICS classifications have shortcomings because PG&E does not closely maintain or update this data field. Classifications are typically done at the firm level, meaning that the NAICS code assigned to a given establishment may not reflect its actual business. For example, the office space associated with a food packing plant may also be classified as a food packing plant due to the overall firm classification. Despite these limitations, it still provide useful information for the data cleaning process. Appendix Table A.3 shows the breakdown of establishments by 2 digit NAICS prefix. I dropped establishments with the 2 digit prefix 22 and 51. NAICS code 22 signifies "utilities", but for small C&I establishments it typically signified irrigation systems run by city governments. Only 166 of these establishments were in the dataset. The NAICS prefix 51 corresponds to the"information" industry classification, which in my dataset meant cell transmission towers run by companies such as AT&T and Verizon. The 702 establishments with this classification had flat consumption profiles and were usually located in fields or on top of buildings. The results in this paper are robust to the inclusion of these two NAICS codes.

The final cleaned dataset contains interval usage data for 19,071 establishments in the summers of 2014 and 2015.

[6]This step is due to the fact that the data is provided at the account level, and must be aggregated to the establishment level. Some small usage accounts are not associated with an establishment, and are dropped in this step. Appendix Section A.2.2 discusses the establishment definition in more detail.

A.2.2 Classifying establishments

Establishments are defined as a business at a single location where the utility bills are all paid by the same entity. An establishments may be owned by an individual, set up as a franchise or owned by a larger company. Around 70% of the establishments in the sample have a single account that pays for just that establishment and no other locations.

PG&E interval usage data is reported at the meter-account level which does not map directly to the establishment level that I use for analysis. The majority (83%) of establishments have one meter associated with each location, making the mapping of meter-account to establishment level data straightforward. Around 9% of the total establishments had multiple meters clearly at the same location, making it easy to collapse down to the establishment level. Another 8% of customers have meters that may be at the same premise, but where the smart meters were installed on a different date. These customers are harder to collapse down to the establishment level, so they are left in the sample as individual establishments. Around 2.3% of establishments share a premise with a meter that is of a higher tariff. For example, the office space that administers a food processing plant may be on the A-1 tariff and is a part of my sample. The food processing plant, which uses a lot more electricity, may be on the much higher E-19 tariff and is not.

To test for the impact of establishment classification, Appendix Table A.6 shows the result from the main regression in Table 1.3 with all potentially ambiguously classified establishments dropped. The results show that dropping these establishment has little impact on the estimated outcomes.

A.3 Time of use pricing

The California Public Utilities Commission established a set of data requirements for each establishment before it was placed on opt-out peak pricing. They were designed so that establishments would have a history of interval metering data before they were presented with a new pricing system. The September 1, 2011 cutoff is due to two different but related requirements. First, establishments needed to be on mandatory time of use (TOU) pricing for two years before they were eligible for peak pricing. Second, establishments needed to be presented with a billing analysis by PG&E before they were moved on to mandatory TOU pricing. The billing analysis needed to be given to establishments at least 45 days before being placed on TOU prices, and it required one full year of data to conduct. These two requirements combined to require that an establishment had interval usage data before September 1, 2011 to be eligible for opt-out peak pricing in the summer of 2015.

The TOU pricing structure for small commercial & industrial establishments does not change prices significantly between peak and off-peak periods. The peak period, which runs between noon and 6pm, charges customers \$.248/kWh, while the low-cost off-peak period price is set at \$.212/kWh.[7] This is contrast to large C&I establishments, where the off-peak price is almost half the peak price.

The establishments that were placed on peak pricing in November 2014 are the same establishments that were placed on TOU in November 2012. At the time of peak pricing treatment in the summer of 2015, these establishments had been on the TOU rate for around 2.5 years. The establishments in my sample that were not placed on peak pricing in November 2014 were rolled over to TOU in November 2013, putting them on TOU pricing for 1.5 years in the summer of 2015.[8] Importantly, establishments were on TOU pricing for both the summer of 2014 and 2015, but some had been on TOU longer than others.

I empirical test the impacts of TOU on peak consumption by examining the impact during the first year it is rolled out. I leverage the same September 1, 2011 threshold used in the main identification strategy to test how TOU impacted usage. I compare establishments that are eligible for TOU in November 2012 to those that just missed the cutoff and were rolled over in 2013. This design compares establishments on the first year of TOU to those that are still on flat-rate prices. I use the instrumental variables approach outlined in section 1.4.2 and look at the same 2pm-pm window as the peak pricing analysis.

Appendix Table A.7 shows the results of these TOU regressions. I conduct the analysis for both the full summer and for just the event days called that summer. The results across all of the specifications show that TOU does not significantly affect peak electricity consumption during the summer of 2013 when the program was first implemented. If TOU does not significantly change an establishment's consumption compared to the flat rate, then it seems unlikely that being on the tariff for 2.5 years versus 1.5 years will significantly affect usage. This result suggests that TOU will not impact the peak pricing evaluation during the summer of 2014 and 2015.

[7]The part-peak rate is set between at \$.239/kWh. The non-TOU flat rate that customers previously paid was \$.228/kWh.

[8]This group includes all peak pricing ineligible establishments and the eligible establishments that were not moved to peak pricing.

A.3.1 Smart meter background

Analog meters have been used since the late 1800's to measure how much electricity an establishment consumes. These meters were read monthly a "meter reader',' a utility employee that manually checked an establishment's usage once a month. Analog meters are limited to tracking total kWh consumption, and for some customers they also measure peak monthly kW usage. Smart meters were first installed across the PG&E service territory starting in 2008. Smart meters automatically transmit data to PG&E via a radio network, eliminating the need for manual checking and allowing for the collection of more detailed usage and billing data.

Most PG&E establishments have smart meters as of mid-2013, with some residential customers remaining on traditional meters by request. Smart meter installations require a utility worker to visit a business and swap out the old meter. A replacement typically takes 5-15 minutes, does not require the account holder to be present, and only results in a brief interruption in power. Smart meters were deployed across California simultaneously, with deployments first being heavier in the central valley in the earlier years then moving to the rest of the state. Conversations with employees at PG&E have indicated that the deployment pattern of smart meters was based on the availability of contractors and resources, and generally not related to establishment characteristics. A PG&E report on the deployment described:

> The deployment schedule is dependent upon the availability of a trained workforce, an effective supply chain to maintain an efficient installation process, and customer premise access to make the necessary changes at each service location. Deployment planning adjustments may be required due to any number of factors, including adverse customer impacts, supply chain considerations, labor availability, and technology considerations, which could affect the scheduling of meter endpoint installations (Pacific Gas & Electric 2010).

The smart meter network functions by the meter interfacing with a series of network access points on utility polls throughout PG&E's service territory. After a smart meter is installed, it takes between 60 and 90 days for the meter to sync up with the network and for the data to become available in the PG&E system. Furthermore, a series of data quality checks is conducted by the PG&E system to verify that the data is of billing quality, and there are no holes in the data. During this time period, the meter reader would continue to manually check the usage on the smart meter to verify the transmission system worked as intended. Once this process is complete, the establishment is transitioned to full smart meter interval usage data collection. This process is summarized in the PG&E documentation as:

> After installation, gas and electric meters transition when: (1) the communications network infrastructure is in place to remotely read them; (2) the meters are installed, remotely read, and utilize smart meter data for billing; (3) and the remote meter reads become stable and reliable for billing purposes. Once enough customers on a particular "route string" transition to smart meter billing, manual reading of the meters on that "route string" ceases, and those meters are considered activated (Pacific Gas & Electric 2010).

This transition process explains why a large portion of the establishments that were peak pricing eligible, did not end up in the program for the summer of 2015. If an establishment did not have a full year of "stable and reliable" billing data to allow for a billing analysis to be conducted, then they were not moved to TOU pricing in November 2012. The interval meter start date data used in this paper reflects when the interval data was first collected, not when it was declared "stable and reliable." As a result, the eligibility status does not perfectly predict peak-pricing enrollment in the summer of 2015.

A.4 Results robustness

A.4.1 OLS results

Appendix Table A.8 shows the results for the IV approach run with OLS. This regression compares the roughly 13% of establishments with peak prices to all the other establishments in my sample. The results are smaller than what was found using the IV approach. This is consistent with the story that specific establishments opted out because peak pricing would be costly for them. The control group is then unrepresentative of the average usage of an establishment, meaning the treatment effect is subject to bias. The smaller coefficient suggests that the control group in the OLS specification may be increasing its electricity usage, resulting in a downward biased treatment effect.

A.4.2 Non-event day impacts

The results in this paper focus on establishment behavior on event days when prices increase. Establishments may make changes in response to peak pricing that spills over into non-event days. Appendix Table A.9 show the results of the regressions run on summer non-event weekdays between 2pm and 6pm. The results show a small reduction for the inland region using the IV approach that is only significant at the

10% level. This suggests that establishments do not undertake responses to peak pricing that have significant impacts on non-event days.

A.4.3 Clustering robustness

This section considers an alternate level of clustering of the errors in this paper. In the main analysis, the IV specification clusters errors at the establishment and hour of sample level. In the RD specification, errors are clustered at the distance from threshold level. One alternative option is to cluster errors at the weather station level. The hourly weather data comes from 297 weather stations across Northern California, with establishments distance matched to the closest station.[9] Establishments are matched to the same weather station for the full sample, meaning the establishment clusters are contained within each weather station cluster.

Appendix Table A.10 shows the results with errors clustered at weather station level. For the IV specification I cluster at the weather station and hour of sample level, and in the RD specification I cluster at the distance from threshold and weather station level. The change has a negligible impact on the standard errors, showing that the results are robust to higher levels of clustering.

A.4.4 Opt-in establishments

In the primary analysis in this paper, I do not include establishments that voluntary opted into the peak pricing program. I do this because these establishments opted in to peak pricing at different times during 2014 and 2015, meaning they faced a different treatment than the majority of establishments. 48 of the 234 establishments that opted into peak pricing did so before the summer of 2014, meaning they did not have bill protection in the summer of 2015. Another 5 establishments chose to enroll in peak pricing during the summer of 2015. The remaining 181 establishments enrolled in peak pricing in April and May of 2015, giving them much less time to prepare for the program.

I include the opt-in establishments in the main analysis to test if their presence impacts the results. Appendix Table A.11 shows the main specification estimated with the 234 opt-in establishments included. The results show that including these opt-in customers has a small impact on the overall results. Column (6) shows that the inland RD specification is no longer significant at the 5% level, but the point estimate does not change much.

[9] I use a balanced panel of weather stations, no weather stations enter or leave during the sample.

A.5 Calculations

A.5.1 Calculating PG&E wide savings estimates

This section provides details on the PG&E wide savings calculations discussed in Section 1.6. PG&E does not release data that splits peak load out by different customer classes. To proceed with the calculations in this section, I make a number of informed assumptions about the consumption patterns of small C&I PG&E customers.[10]

First, I calculate the total number of inland establishments on the A1 tariff based on demographic data provided by PG&E. I adjust this number downward to reflect that my sample only includes customers that consume between 800 kWh/month and 10,000 kWh/month during the summer months. This results in 157,000 inland establishments which are like those I study in my analysis. I adjust for establishments that will opt out of peak pricing using the overall observed opt out rate from 2015 of 16.7 %. I assume subsequent waves of establishments will opt out at the same rate.

I assume that the establishments used in the main estimation sample reflect the average consumption for all C&I establishments. Appendix Figure A.3 shows that this is true when comparing establishments within 8 weeks of the September 1, 2011 cutoff to those within 27 weeks. It shows a similar pattern of usage, helping to validate this assumption.

To estimate total savings, I multiply this average consumption by the implied percent reductions that from columns (5) and (6) of Table 1.3 respectively. This approach leads to a total savings estimate that is is conservative in nature. I am only considering the savings for inland customers with summer consumption between 800 kWh/month and 10,000 kWh/month. This leaves out a large number of smaller establishments and a few much larger establishments that likely reduce their usage under peak pricing.

The savings estimate also ignores reductions from coastal customers. I chose do to this because the main empirical strategy did not find significant reductions for the coastal establishments. Section 1.5.5 provides evidence that on hotter coastal days, larger firms reduce their usage. I choose not to include these savings numbers in my final MW savings estimates because there is a level of uncertainty about how these event days line up with the highest demand event days of the summer. In subsequent sections, I use the aggregate savings numbers in welfare calculations that assume the total savings number is reflected on all event days, not just hot coastal event days.

[10] I cannot use my interval consumption data to make these calculations because I only have a sample of small C&I establishments usage.

A.5.2 Welfare impacts of peak pricing using RD approach

I use the IV estimates of program impacts for the welfare calculations in this paper for two reasons. First, I prefer to use the IV estimate, because it is more likely representative of the average treatment effect across all small C&I establishments. The local average treatment effect for the IV approach is the full 8 week bandwidth, while the RD approach is more focused on the September 1, 2011 discontinuity.[11] Second, the IV approach yields a smaller coefficient, making my welfare estimates more conservative.

In this section I calculate the welfare impacts for the RD estimates for completeness. I use the estimated 216 MW of peak reductions reductions derived in Section 1.6. The net consumer surplus losses from higher prices total $3.14 million/year, and the benefits from the $.01/kWh discount on non event day consumption are 0.84 million/year. Taken together this results in a total welfare benefits of $283 million (2016 dollars) using a 3 percent real discount rate and a 30 year horizon.

A.5.3 Extending savings estimates to other California Utilities

This section describes the method used to extend the welfare estimates from small PG&E C&I customers to larger regions and customer classes. To make the calculation, I use data on system peak load from EIA form 861. In this form, every utility in the U.S. is required to file their summer and winter peak load with the EIA. To extend the welfare benefits to the larger regions, I simply scale up the PG&E welfare benefits by the ratio of peak loads. The IOU welfare benefits column in Table 1.7 extends the welfare benefits to the other major investor owned utilities in California Southern California Edison and San Diego Gas & Electric. These utilities provide 69% of the electricity in California, and are all in the process of implementing opt-out peak pricing for their C&I customers. The California welfare benefits further extends the savings to all California utilities, which includes the larger Los Angeles and Sacramento municipal utility districts.

There are a number of assumptions underlying these calculations. The savings estimates are for the peak pricing program run for small C&I establishments. I assume that the small C&I establishments in the other regions have similar consumption patterns and demand reductions when facing peak pricing. It also assumes that the small C&I customer class is around the same percent of peak load in the other regions. This assumption is reasonable in California, where the climate and income is similar

[11] In the limit, the RD approach estimates the impact directly at the discontinuity.

across the state. However, the assumption may be less plausable at the national level, where I find that the national welfare benefits of small C&I peak pricing would be over $17 billion. This exact number should be interpreted with caution based on the assumptions used in its calculation, but the large value does highlight the magnitude of the distortion caused by flat retail pricing.

The bottom row of Table 1.7 extends the welfare savings to the full C&I customer class. These customers were moved over to opt-out peak pricing over the previous four years. I assume they would have the same 13.4 % reduction in usage as the small C&I customers. I adjust for the opt out behavior observed for the large C&I establishments in the PG&E program, where 42% of establishments opted out.

A.5.4 Assumptions for comparison of peak pricing to real-time pricing

This section describes the assumptions used to make the calculations in Section 1.6.4. I compare the outcomes under peak pricing to the first-best outcomes under a theoretical real-time price scenario. I do this for two may reasons. First, there isn't a market price in California that can be used for the real-time price comparisons. The existing wholesale market has a number of distortions including a price cap, a capacity market and the regulator resource adequacy requirements. These policies distort the wholesale price, making it a poor real-time price. Second, the simple setup allows for a transparent comparison between the two that isn't dependent on institutional details of the California market.

The theoretical market I use is structured as an energy only market without any price caps. Real-time prices take on two values. The low value is set at $.10/kWh, which roughly reflects the marginal cost of a natural gas combined cycle generator. The high value is set at $1.35/kWh and reflects both the generation and capacity cost of peaker plants.[12] I assume prices spike to the high level sometime between 2pm and 6pm on 3 super-peak days per year. Customers are charged a fixed fee in the RTP market to recover the remaining fixed costs associated with transmission and distribution.

The peak pricing alternative is set up in a comparable manner. Retail prices are set at the low RTP price and fixed charges are used to recover any remaining costs including capacity costs, transmission and distribution charges. During event hours, the price is raised to the high RTP level between 2pm and 6pm. I assume 8 event days are called per year based on the 101 degree trigger temperature described in

[12] The $1.35 value is based on the large C&I peak price. PG&E based this value on their internal value of capacity number.

section 1.6.3. 3 of these event days are on the super-peak days, while the other 5 are called on low price days to capture the uncertainty in choosing the correct event days. By design, the peak pricing program will collect more revenue than the RTP program because of the longer and more frequent periods at the high price. I assume this money is reflected in adjustments to fixed charges for the subsequent year.

Appendix figures

Figure A.1: Letter Sent to Establishments 30 Days Before Peak Pricing

21PDPDEFLT2
PDP Final Default Letter (Ltr #3c) **Page 1 of 2**
<<DATE>>

Important information regarding your transition to a Peak Day Pricing electric rate plan. Please read to learn more.

[Customer Name1]
[Customer Name2, if exists]
[Mailing Address2, if exists]
[Mailing Address1]
[Mailing Address City, State Zip]
[Mailing Address Country1]

Re: 30 Day Notification of Switch to Peak Day Pricing Electric Rate Plans for Business

Dear [Customer Name]:

Last month, we sent a letter to notify you that starting in November, one or more of your business accounts is scheduled to transition from a time-of-use electric rate plan to a Peak Day Pricing rate plan. This is part of a requirement by the California Public Utilities Commission to encourage conservation when energy demand is higher. This is a reminder of the upcoming move to a new electric rate plan.

Peak Day Pricing works in conjunction with your existing time-of-use rate, applying higher energy prices on 9 to 15 Event Days per year in exchange for discounted energy rates at all other times from May 1st through October 31st.*

This rate plan transition will affect the Service ID(s) referenced on the following page(s).

Peak Day Pricing includes automatic Bill Protection
Bill Protection lets you try Peak Day Pricing risk-free for a full year. After 12 months, we will compare your costs on Peak Day Pricing to what your costs would have been on your time-of-use rate plan. If your costs on Peak Day Pricing are higher, you will automatically receive a bill credit for the difference. You can opt out and return to your time-of-use rate plan at any time.

Make your decision today
PG&E is here to help you understand this new rate plan and decide what is best for your business. A personalized rate analysis can help you estimate how your electric bills may change with Peak Day Pricing.

- Access your online rate analysis anytime at **pge.com/myrateanalysis**

If you want to enroll early, or opt-out of the transition to Peak Day Pricing you can do so before your eligible Service ID(s) are automatically enrolled in November.

- Enroll early, or opt out of transition by visiting **pge.com/pdpchoice**

Update your notification preferences
If you plan to enroll, or have already enrolled in Peak Day Pricing, please be sure to update your Peak Day Pricing notifications, so you don't miss any Event Day notices. Update your notifications at **pge.com/myalerts**.

We value you as a customer and understand you may have some questions. For more information about the transition to Peak Day Pricing, visit **pge.com/pdp30day**.

Sincerely,

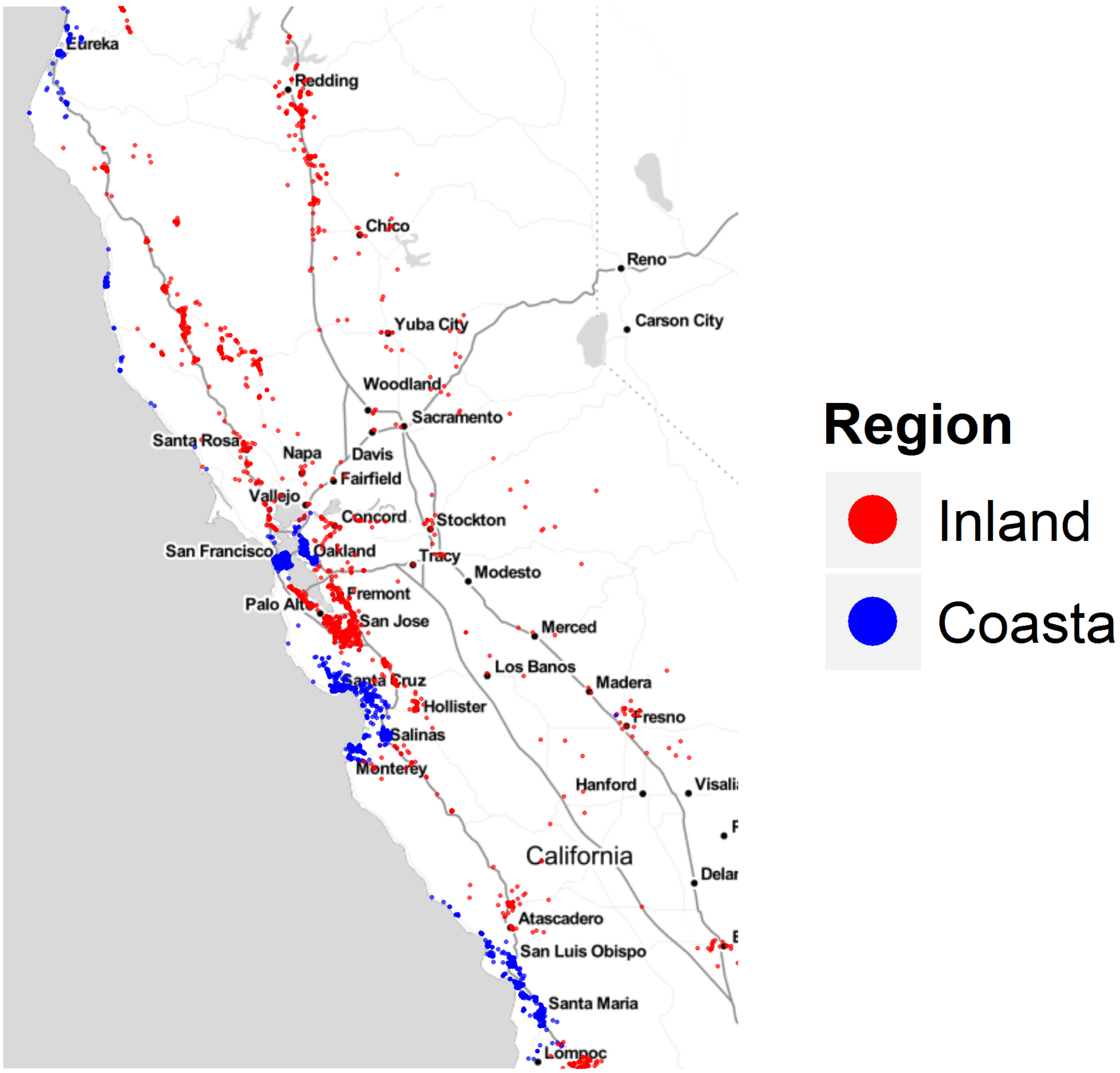
Maril Pitcock
Director, Pricing Products
Pacific Gas and Electric Company

P.S. Remember to stay away from downed power lines and never touch or try to move them. Always assume a downed power line is live and report them immediately by calling **911** and PG&E at **1-800-743-5000**.

* Effective summer rates are lower after Peak Day Pricing credits have been applied, but effective rates are higher during Peak Day Pricing Event Hours. "PG&E" refers to Pacific Gas and Electric Company, a subsidiary of PG&E Corporation. 21PDPDFLT2

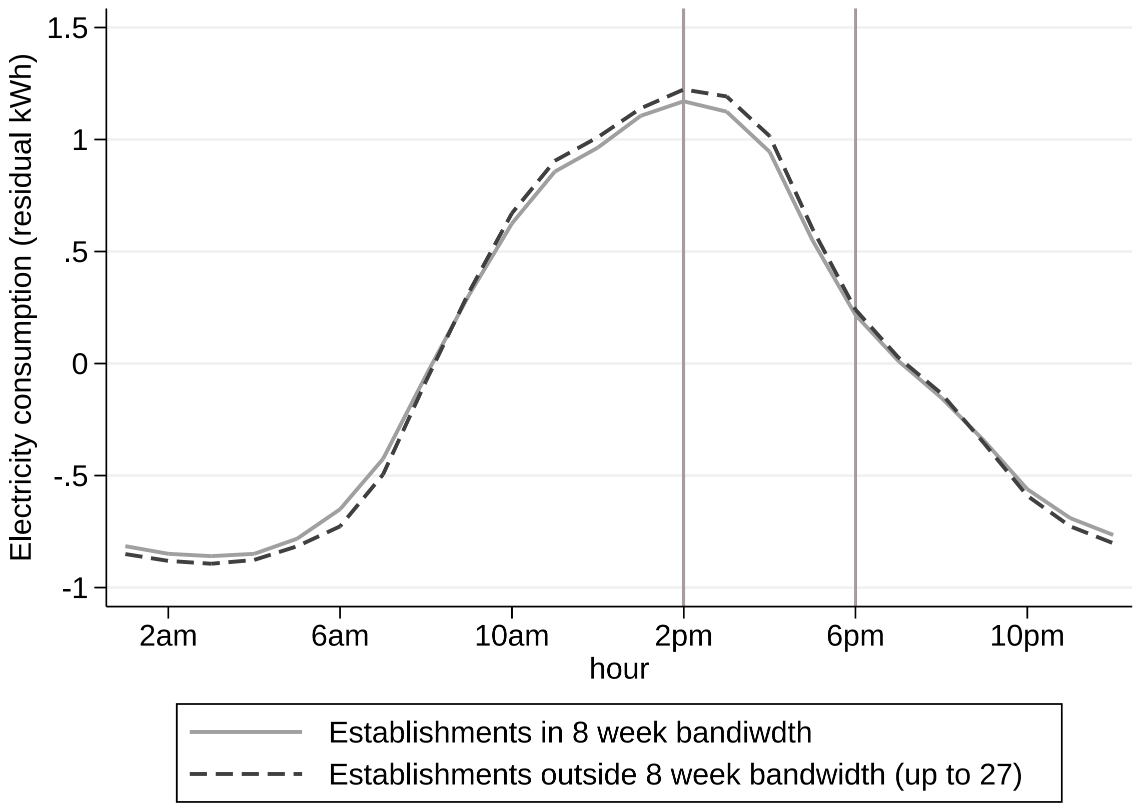
Note. - This letter is a sample of what was sent to every establishment 30 days before peak pricing started. It provides information on how to opt out at the web site "pge.com/pdpchoice." It also describes bill protection and directs establishments how to set their event day notification preferences.

Figure A.2: Map of Establishments in Sample by Region



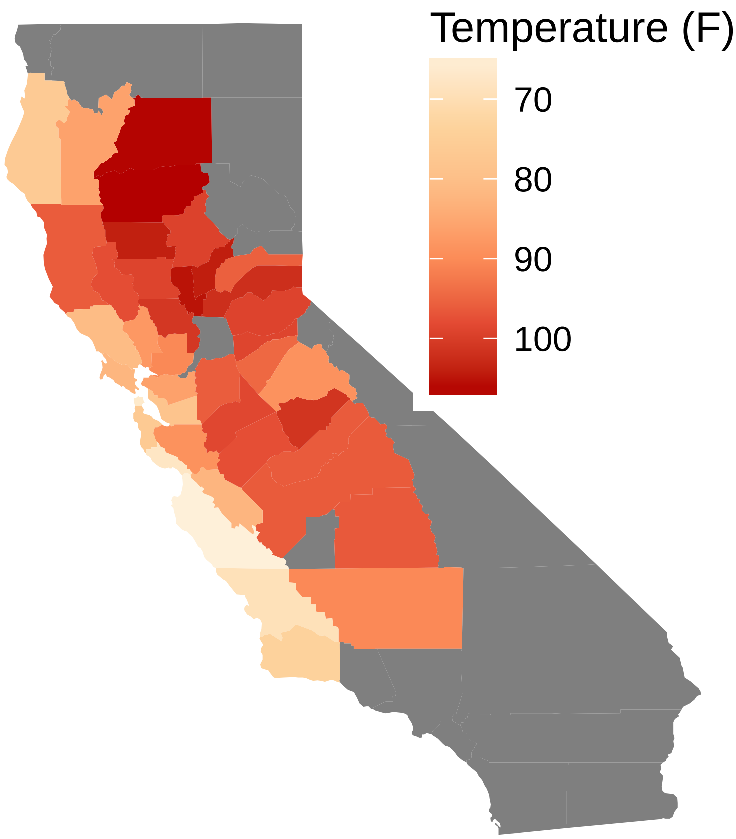
Note. - This figure shows all 7,435 establishments in primary sample. Each dot corresponds to an individual establishment. Inland vs coastal designation is based on baseline territory as defined by PG&E and reflects climate conditions.

Figure A.3: Comparison of Establishment Consumption Between Primary Sample and Establishments in Larger Bandwidths



Note. - This figure compares summer 2014 hourly kWh usage for establishments in the 8 week bandwidth to establishments in up to 27 weeks away from the September 1, 2011 cutoff. Values show residuals after establishment fixed effects are removed. Formal statistical comparison between the two cannot reject that the two groups are the same.

Figure A.4: Average Temperature on Event days by County



Note. - This figure shows the average temperature on event days in 2015 at the county level. Temperatures reflect the average temperature across all Mesowest weather stations in a county. Weather stations are weighted based on how many establishments they are distance matched to in the main analysis.

Appendix Tables

Table A.1: Event Days with Day Ahead Forecast and Trigger Temperature

Event date	NWS day ahead max temperature forecast	Trigger temperature
6/7/2013	98	96
6/28/2013	99	96
7/1/2013	107	96
7/2/2013	106	96
7/9/2013	96	96
7/19/2013	98	98
8/19/2013	94	96
9/9/2013	97	94
9/10/2013	94	94
10/18/2013	82	89
6/9/2014	100	96
6/30/2014	102	96
7/1/2014	96	96
7/7/2014	101	96
7/14/2014	99	96
7/25/2014	101	96
7/28/2014	97	96
7/29/2014	97	96
7/31/2014	98	96
9/12/2014	96	98
6/12/2015	99	96
6/25/2015	103	96
6/26/2015	100	96
6/30/2015	101	96
7/1/2015	100	98
7/28/2015	101	98
7/29/2015	104	98
7/30/2015	100	98
8/17/2015	101	96
8/18/2015	96	96
8/27/2015	97	96
8/28/2015	96	96
9/9/2015	102	98
9/10/2015	104	98
9/11/2015	101	98

Notes. - This table shows the day ahead maximum temperature forecast used by PG&E for all event days between 2013 and 2015. NWS corresponds to 5 National Weather Service stations PG&E uses for its forecasting. An event day is called when the day ahead forecast equals or exceeds the trigger temperature.

Table A.2: Average Outdoor Temperature by Event day

Event date	All PG&E average temperature	Coastal establishments average temperature	Inland establishments average temperature
6/9/2014	74.76	67.01	91.66
6/30/2014	75.79	68.05	92.67
7/1/2014	71.28	64.92	85.15
7/7/2014	73.26	66.89	87.15
7/14/2014	73.49	66.85	87.99
7/25/2014	80.98	74.76	94.54
7/28/2014	76.67	71.12	88.77
7/29/2014	76.93	70.71	90.41
7/31/2014	76.00	68.85	91.58
9/12/2014	75.55	68.69	90.50
6/12/2015	75.03	67.57	91.29
6/25/2015	77.30	70.18	92.81
6/26/2015	72.94	65.12	89.98
6/30/2015	81.08	73.57	97.44
7/1/2015	75.89	69.38	90.05
7/28/2015	80.86	74.34	95.06
7/29/2015	77.21	69.55	93.90
7/30/2015	76.86	70.50	90.69
8/17/2015	77.94	70.66	93.77
8/18/2015	75.65	70.37	87.14
8/27/2015	83.97	80.18	92.21
8/28/2015	82.88	78.52	92.38
9/9/2015	86.73	81.66	97.77
9/10/2015	82.79	76.54	96.40
9/11/2015	80.62	74.40	94.17
Average	77.70	71.21	91.82

Notes. - This table shows the average temperature between 2pm and 6pm on all event days in 2014 and 2015. The values reflect the establishment weighted average temperature. Temperatures do not reflect official National Weather station temperatures used to call event days.

Table A.3: Establishment Industry Classifications

Naics 2 digit code	Establishment count	Percent of establishments
11	104	1.4%
23	232	3.1%
31	168	2.3%
32	107	1.4%
33	226	3%
42	224	3%
44	749	10%
45	286	3.8%
48	73	.98%
52	213	2.9%
53	650	8.7%
54	307	4.1%
56	157	2.1%
61	106	1.4%
62	655	8.8%
71	131	1.8%
72	1,068	14%
81	963	13%
92	215	2.9%
Not available	801	11%

Notes. - This table shows the first two digits of the North American Industry Classification System (NAICS) industry classification for all 7,435 establishments in the sample. These 2 digit NAICS codes are used to classify establishments as customer facing or non-customer facing in section 1.5.4. The PG&E data did not have NAICS code information for the 11% of establishments classified as "Not available."

Table A.4: Highest PG&E System Demand Days of 2015

Date	Event day	PG&E max load	Hour of max load
8/17/2015	yes	19,451	4pm-5pm
6/30/2015	yes	19,320	4pm-5pm
7/29/2015	yes	19,248	4pm-5pm
8/28/2015	yes	19,233	4pm-5pm
9/10/2015	yes	19,230	4pm-5pm
9/9/2015	yes	19,017	4pm-5pm
7/20/2015	no	18,546	4pm-5pm
6/8/2015	no	18,441	6pm-7pm
7/28/2015	yes	18,403	5pm-6pm
9/21/2015	no	18,398	4pm-5pm
8/27/2015	yes	18,328	4pm-5pm
8/16/2015	no	18,197	6pm-7pm
6/25/2015	yes	18,114	4pm-5pm
9/11/2015	yes	18,019	4pm-5pm
6/26/2015	yes	17,950	4pm-5pm
9/8/2015	no	17,875	4pm-5pm
7/30/2015	yes	17,750	4pm-5pm
7/1/2015	yes	17,734	2pm-3pm
8/18/2015	yes	17,372	4pm-5pm
6/12/2015	yes	17,275	5pm-6pm

Notes. - This table shows the top 20 PG&E system load dates in the summer of 2015. PG&E max load reported by the California Independent System Operator (CAISO) at from oasis.caiso.com. The hour of max load signifies in what hour the load occurred.

Table A.5: The Effect of Peak Pricing on Peak Electricity Consumption: Demand Elasticities

	All PG&E		Coastal		Inland	
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	RD	IV	RD	IV	RD
Peak pricing - elasticity	−0.0565*	−0.1749	0.0068	−0.0474	−0.1171***	−0.2298**
	(0.0335)	(0.1708)	(0.0576)	(0.3434)	(0.0369)	(0.1120)
Establishments	7,435	7,435	5,096	5,096	2,339	2,339
Event day kWh usage	5.55	5.55	5.03	5.03	6.70	6.70
Average temperature	78	78	71	71	92	92

Notes. - This table reports elasticity coefficients from 6 separate 2SLS regressions. The dependent variable in all regressions is the log of establishment hourly kWh consumption. The coefficients show the impact of peak pricing on peak consumption between 2pm and 6pm. "IV" and "RD" correspond to the instrumental variables and regression discontinuity approaches estimated using equations 1.1 and 2.4. All regressions control for temperature and include hour of sample fixed effects and establishment fixed effects. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table A.6: Main results dropping establishments with ambiguity in establishment definition

	All PG&E		Coastal		Inland	
	(1)	(2)	(3)	(4)	(5)	(6)
	IV FE	IV RD	IV FE	IV RD	IV FE	IV RD
Peak pricing	−0.0543	−0.2284	0.0346	−0.0039	−0.1423***	−0.3479**
	(0.0436)	(0.2379)	(0.0711)	(0.4764)	(0.0508)	(0.1677)
Establishments	6,247	6,247	4,330	4,330	1,917	1,917
Event day kWh usage	5.47	5.47	4.92	4.92	6.73	6.73
Average temperature	77	77	71	71	92	92

Table shows the results from Table 1.3 with the ambiguously classified establishments dropped. Appendix Section A.2.2 outlines the establishment classification process and which establishments are dropped. All regressions control for temperature and include hour of sample FE. IV regressions include account by hour of day by day of week FE. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level.***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table A.7: Robustness : Impacts of TOU when it was first implemented

	All PG&E		Coastal		Inland	
	(1)	(2)	(3)	(4)	(5)	(6)
	All days	Event days	All days	Event days	All days	Event days
TOU	0.0363	0.0383	0.0343	0.0235	0.0390	0.0478
	(0.0296)	(0.0464)	(0.0538)	(0.0830)	(0.0327)	(0.0532)
Establishments	7,383	7,383	5,059	5,059	2,324	2,324
Event day kWh usage	4.99	5.45	4.75	4.96	5.52	6.52
Average temperature	71	76	66	69	79	90

Note. - Table shows the impact of TOU pricing on peak electricity consumption the first year it was implemented. This specification uses the same empirical strategy as the IV approach on the same sample of establishments. See Appendix Section A.3 for details. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table A.8: OLS results

	(1)	(2)	(3)
	All PG&E	Coastal	Inland
Peak pricing	-0.0469***	-0.0272	-0.0589**
	(0.0149)	(0.0176)	(0.0244)
Establishments	7,435	5,096	2,339
Event day kWh usage	5.59	5.03	6.70
Average temperature	78	71	92

Notes. - Table shows results from OLS regression of electricity usage on peak pricing status using the IV approach. All regressions control for temperature and include hour of sample FE. IV regressions include account by hour of day by day of week FE. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level.***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table A.9: Non event day 2pm-6pm impacts

	All PG&E		Coastal		Inland	
	(1)	(2)	(3)	(4)	(5)	(6)
	IV FE	IV RD	IV FE	IV RD	IV FE	IV RD
Peak pricing	−0.0170	−0.0380	0.0333	0.1175	−0.0682*	−0.1148
	(0.0341)	(0.1401)	(0.0585)	(0.3401)	(0.0370)	(0.1104)
Establishments	7,669	7,669	5,272	5,272	2,397	2,397
Event day kWh usage	5.13	5.13	4.87	4.87	5.71	5.71
Average temperature	73	73	69	69	81	81

Notes. - Table shows results from main specifications run with 2-6pm usage on non-event weekdays between June 1st and October 31st 2015. Establishments do not face high peak prices during these hours. "IV FE" and "IV RD" correspond to the 2SLS estimation of the IV and RD approach respectively. All regressions control for temperature and include hour of sample FE. IV regressions include account by hour of day by day of week FE. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level.***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table A.10: Main results with errors clustered at weather station level

	All PG&E		Coastal		Inland	
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	RD	IV	RD	IV	RD
Peak pricing	−0.0695*	−0.2152	0.0084	−0.0584	−0.1441***	−0.2828**
	(0.0401)	(0.2046)	(0.0729)	(0.4070)	(0.0438)	(0.1274)
Establishments	7,435	7,435	5,096	5,096	2,339	2,339
Event day kWh usage	5.55	5.55	5.03	5.03	6.70	6.70
Average temperature	78	78	71	71	92	92

Notes. - This table reports regression coefficients from 6 separate 2SLS regressions. The dependent variable in all regressions is the log of establishment hourly kWh consumption. The coefficients show the impact of peak pricing on peak consumption between 2pm and 6pm. "IV" and "RD" correspond to the instrumental variables and regression discontinuity approaches estimated using equations 1.1 and 2.4. All regressions control for temperature and include hour of sample fixed effects and establishment fixed effects. IV errors two-way clustered at the weather station and hour of sample levels. RD errors clustered at the distance from threshold and weather station levels. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table A.11: Robustness: Opt-in peak pricing establishments included

	All PG&E		Coastal		Inland	
	(1)	(2)	(3)	(4)	(5)	(6)
	IV FE	IV RD	IV FE	IV RD	IV FE	IV RD
Peak pricing	−0.0616	−0.1493	0.0145	0.1177	−0.1377***	−0.2765*
	(0.0431)	(0.1958)	(0.0726)	(0.4408)	(0.0481)	(0.1466)
Establishments	7,669	7,669	5,272	5,272	2,397	2,397
Event day kWh usage	5.54	5.54	5.02	5.02	6.71	6.71
Average temperature	78	78	71	71	92	92

Table shows the results from Table 1.3 with the 234 establishments that voluntarily opted in to peak pricing included. All regressions control for temperature and include hour of sample FE. IV regressions include account by hour of day by day of week FE. Standard errors are in parentheses. IV errors two-way clustered at the establishment and hour-of-sample levels. RD errors clustered at the distance from threshold level.***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table A.12: Robustness: Comparison of differently peak pricing targeting scenarios to RTP with long period of high peak prices

Event days called	(1) $.85/kWh peak price (peak price < RTP)	(2) $1.35/kWh peak price (peak price = RTP)	(3) $1.85/kWh peak price (peak price > RTP)
8 days (well targeted)	51%	**87%**	73%
15 days (current)	**46%**	69%	35%

Note. - Table shows the percent of the welfare gains for various peak pricing scenarios compared to the efficient real-time price policy when peak prices hit $1.35/kWh for 4 hours on 3 super-peak days per summer. The center column reflects when the peak prices are correctly set at $1.35/kWh. The left column shows the effectiveness of peak compared to real-time prices when the peak prices are set at $.85/kWh but prices actually hit $1.35/kWH. Similarly, the far right column shows the impacts when peak prices are set too high. The first row shows the impacts with the well targeted 8 days per summer, while the bottom row shows the impacts when 15 are called.