The Effect of a Mandatory Time-of-Use Pricing Reform on Residential Electricity Use*

Katrina Jessoe † UC Davis, ARE

David Rapson ‡ UC Davis, Economics

Jeremy Blair Smith§ Boston University

November 27, 2012

Abstract

Time-of-use (TOU) electricity pricing has attracted attention as a potential policy to reduce peak electricity demand and thus address the engineering and environmental challenges associated with electricity production. However, convincing evidence of its effectiveness is lacking in the literature. In this paper, we analyze short-run household responses to a large-scale field deployment of TOU pricing. Households that breached a usage threshold were forced to switch from a flat-rate plan to one with a high electricity price during peak hours (noon through 8pm on weekdays) and a low price during all other hours. Features of the program implementation give rise to multiple natural experiments that we exploit within a regression discontinuity framework. We find that, after being switched to TOU, large households substantially reduced total electricity consumption during the summer months. However, some of the responses we find are inconsistent with static utility maximization. For example, households reduced their off-peak consumption at some times of the year when rates were such that a static optimizer should have unambiguously increased off-peak consumption. This indicates either that they were subject to additional constraints or were responding to incentives other than contemporaneous prices.

*We would like to thank Kevin Lang, Marc Rysman, Johannes Schmieder, Jim Bushnell, Lucas Davis, Michael Anderson, Alan Meier, Michael Manove, and the many other participants in seminars at Boston University, the UC Energy Institute, and UC Berkeley ARE for their comments. Any errors are our own.

†E-mail: kkjessoe@ucdavis.edu; Website: http://kkjessoe.ucdavis.edu/

‡E-mail: dsrapson@ucdavis.edu; Website: http://www.econ.ucdavis.edu/faculty/dsrapson

§JOB MARKET PAPER. PhD Candidate, Department of Economics, Boston University. Home address: 36 Harvard Avenue, Apartment 1, Brookline, Massachusetts, 02446, USA. Cell: 617-314-1526. E-mail: jersmith@bu.edu. Website: http://people.bu.edu/jersmith.

1 Introduction

Electricity consumption will be inefficiently high if rates do not reflect the associated social costs. While the most obvious social cost is environmental harm, a more subtle source of inefficiency in electricity markets is the mismatch between wholesale production costs, which can fluctuate widely by the minute, and the fixed retail rates faced by most consumers. Due to this second issue, price-based policies to achieve efficient levels of consumption would entail charging rates that vary in real time. In part because consumers are not accustomed to prices that fluctuate at such a great frequency, there has been a reluctance to adopt such measures.[1] Time-of-use (TOU) electricity pricing, in contrast, divides electricity use into just two or three blocks according to the time of day at which it is consumed, and applies higher rates to blocks corresponding to historically high-cost times. It is a small step towards aligning retail electricity prices with marginal production costs, and has the crucial advantage of being easy for consumers to understand and, in principle, respond to.

In this paper, we study short-run household responses to a large-scale mandatory residential TOU program, and seek to draw lessons about TOU's potential as a tool to achieve efficiency in electricity markets. Such potential could be realized in several ways. First, and of most concern to utilities, TOU could induce a reduction in peak consumption. Shedding peak load is desirable for utilities from a cost-containment perspective, and desirable more generally to the extent that it reduces reliance on dirtier generation sources. Second, though this is not an explicit goal of the policy, TOU might induce a general conservation effect depending on how the rates are set, which again would lead to a reduction in both production and environmental costs. Finally, TOU could provide a platform for regulators and rate-setters both to familiarize consumers with time-varying rates and to learn about the dimensions along which consumers respond to rate changes. This would be valuable in terms of guiding future policies aimed at correcting both sources of inefficiency.

We find evidence that the TOU program induced some efficiency improvements. Specifically, we find that, after being switched to TOU, the largest households in our sample substantially reduced their total electricity consumption in the summer months. The timing

[1] As noted by Joskow and Wolfram (2012), a 2010 survey conducted by the Federal Energy Regulatory Commission found that only about one percent of U.S. residential customers face a time-varying rate structure of any kind.

of this conservation effect is especially noteworthy, since the social value of usage reductions is highest during the peak summer usage season. However, this conservation effect was not exclusive to on-peak times of day, but rather reflects roughly proportional decreases in on-peak and off-peak usage. In fact, we find little evidence of load shifting across all household sizes in any summer months, and only weak evidence of load shifting in winter and spring months. Moreover, we find that, after being switched to TOU, the smaller households in our sample moderately increased their total electricity usage in some summer months. We provide speculation as to the incentives and decision-making processes that may have led to these responses, but cannot at this time formally test these hypotheses against one another. Nonetheless, we propose that the very questions that our results raise and the models of consumer behavior that they can rule out provide important insights to regulators wishing to pursue further policy innovations.

The TOU program that we study is the first large-scale residential field deployment of its kind. The program, implemented by a utility in the northeastern United States, forced households to switch irrevocably to a TOU tariff after breaching a monthly usage threshold. The threshold was initially set at 4000 kilowatt hours (kWh) per 30-day billing month in November 2006, and was reduced to 3000kWh in January 2008 and finally to 2000kWh in January 2009. The program thus targeted very large electricity users at first, but covered more moderate users over time as the threshold was lowered. It is the mandatory nature of the program that gives rise to a natural experimental setting, which we exploit within the regression discontinuity framework. We use monthly billing data on total usage and total expenditure for over 20,000 households between 2006 and 2011 to estimate treatment effects with a clear, causal interpretation.

We make contributions to the literature evaluating household responses to electricity rates along several dimensions. First, the mandatory nature of the program design eliminates concerns over selection effects that are present in evaluations of programs based on voluntary participation. Second, the field deployment of the program provides us with a natural experimental setting as well as a large sample size. Together, these features of the program facilitate a clear and transparent research design that permits the estimation of causal treatment effects that are free of the small-sample and selection concerns inherent

in many past studies. Finally, we develop a simple, reduced-form method to overcome a common deficiency in electricity billing data: since utilities have no need to meter usage by time of day when total usage is billed at a flat rate, it is usually impossible to retrieve the historical on-peak/off-peak breakdown of control households' usage. Our method uses data from multiple sources to estimate separate treatment effects for on-peak and off-peak usage.

While the mandatory nature, field deployment, and size of the program make our setting unique, we are not the first to examine the potential of time-varying prices to improve efficiency. An early theoretical literature described how incentives to reduce peak usage would translate into more efficient capital deployment.[2] Making use of technological developments (via high-frequency metering), some small field trials were implemented in the 1970s to study the potential of TOU pricing to shift load. Of particular relevance is Hausman, Kinnucan, and McFadden (1979), who estimated a demand system on data derived from 199 households that participated in a utility-run trial program of TOU pricing in Connecticut from October 1975 to October 1976. The dramatic financial incentives induced households to shift from peak to off-peak usage, and also led to overall energy conservation.[3]

More recently, several studies have evaluated consumer responses to more granular forms of time-varying pricing, mostly in limited experimental settings.[4] Of note are two studies that rely on randomized assignment to control and treatment to isolate the impact of time-varying pricing on usage. Wolak (2006) considers the impact of critical-peak pricing, whereby consumers receive rebates for conservation on a selected number of high-price electricity days and pay a standard rate at other times. Households assigned to the treatment group reduced usage by 12% compared to control groups during pricing events. Allcott (2011) evaluates the

[2] Steiner (1957) and Williamson (1966) are two of the most frequently-cited examples from this body of previous work.

[3] Treatment households were exposed to on-peak, intermediate, and off-peak prices of 16 cents/kWh, 3 cents/kWh and 1 cent/kWh, respectively, while control households faced a schedule on which they were charged time-invariant prices between 6.25 and 3.24 cents/kWh (with this flat rate decreasing with monthly usage). Due to the magnitude of the price changes, the period of maximum usage for treated households shifted from peak hours to off-peak hours. In addition, total usage declined by 5 percent for treated households.

[4] Even when set optimally, TOU rates are too coarse to capture more than a small fraction of wholesale price fluctuations, as the small number of different rate periods within a given day and the fact that the rates are set for months at a time foreclose the possibility of transmitting high-frequency price signals to retail customers. Borenstein (2005) and Jessoe and Rapson (2012a) show that TOU prices can at most capture 6 to 13 percent of wholesale price variation. For this reason, economists tend to favor prices that are more timely and granular, such as real-time pricing.

impact of real-time pricing, whereby the retail price of electricity varies hourly based upon day-ahead wholesale prices. He finds that households exhibited a price elasticity of -0.1 on average, and conserved electricity on net.[5] While our work is most closely related to these studies, our program setting and research design differentiates us from this strand of the literature along two dimensions: the regime change we study arose out of the rate-setting process and applied (with few exceptions) to the entire residential customer base, establishing a *natural* as opposed to a *framed* experimental setting, in the taxonomy of Harrison and List (2004); and the regime change was permanent.

2 Program Design and Data

We have obtained data from an electric distribution company in the northeastern United States that implemented a mandatory residential time-of-use program beginning in 2006. Prior to the introduction of this program, most residential customers were billed according to a seasonal flat rate, with the price of electricity varying seasonally but remaining constant within a day. However, approximately 12% had chosen to be placed instead on a seasonal TOU rate, with the price of electricity varying seasonally and within a day.[6] The program that we study was introduced in order to increase the take-up of this TOU rate by making it mandatory for high-use customers.

The TOU program was intended to provide customers with an economic incentive to shift electricity consumption from peak to off-peak periods and to invest in energy efficient technologies to reduce usage in the long run. One reason that TOU pricing was favored over other time-varying pricing programs was the recognition that, to effectively induce customers to shift load, a rate must be simple and readily understood. The rationale for the mandatory nature of the program was the concern that households opting into the TOU rate would mostly be those not needing to modify usage in order to reduce expenditure, so that

[5] Faruqui and Sergici (2010) provide a meta-analysis of 15 additional time-varying pricing pilots and experiments, including those with time-of-use, critical-peak, and real-time pricing components, conducted by utilities over the past decade. They find that households responded to these programs by reducing usage in general, though the magnitude of the response depended on the presence of enabling technologies and other factors. Further, they find that TOU programs induced reductions in on-peak usage of between 3% and 6%.

[6] In the analysis that follows, we exclude these voluntary adopters.

continuing to make adoption purely voluntary would have had little effect on system-wide peak consumption.

Under the policy, when a residential customer's electricity usage in any 30-day billing period exceeded a pre-determined threshold, the customer would be automatically placed onto TOU pricing. Further, this threshold would be lowered annually. Beginning November 2006, a household would be placed on TOU pricing by January of 2008 if usage in any 30-day billing period exceeded 4000 kWh. This threshold applied until December 31, 2007, following which a 3000 kWh threshold was introduced. Under the 3000 kWh threshold, a customer would be switched onto the TOU rate within six months of exceeding the 3000 kWh threshold. The threshold was lowered to 2000 kWh beginning in January of 2009, after which households exceeding the threshold would be switched onto TOU pricing within 6 months of exceeding the threshold.[7]

To inform customers of this program, the utility engaged in a limited education and outreach initiative in 2006. The goals of this campaign were to highlight (i) the equitability in this cost recovery program and (ii) the potential bill savings to customers should they shift load from peak to off-peak hours. The reason for the limited scope of the outreach effort was the assumption that most people understand the idea behind TOU pricing, as many retail products – e.g. phone service – are based on peak pricing structures.

The residential TOU rate plan charges a high per-kWh rate at on-peak times (noon through 8pm on weekdays) and a low per-kWh rate at off-peak times (all other times and days). Both rates are higher during the summer season (June through September inclusive) than during other seasons. Besides the seasonal rate changes on June 1 and October 1 of each year, the utility also makes periodic rate adjustments, often on January 1 and July 1 of each year. Table 1 shows the TOU rates that were in effect over the period of our analysis, and compares them to the corresponding non-TOU rates.[8] Although revenue neutrality is generally an overarching goal of utilities and regulators when planning rate design changes,

[7] It should be noted that households suffering from a serious illness or other life threatening situation necessitating the use of specialized electrical devices could apply to be exempted from this program.

[8] Until 2009, the summer non-TOU tariff had an increasing-block structure, with the first 500 kWh of usage in a billing month charged at a base "headblock" per-kWh rate and the remaining usage in that billing month charged at a higher "tailblock" per-kWh rate. The tailblock rate actually exceeded the TOU on-peak rate in the summer of 2008. We will revisit this point when interpreting our results below.

we do not know how the utility attempted to achieve this or what other issues they considered when setting these individual rates.

To evaluate the effect of this policy change on total usage, total expenditure, and load shifting, we rely on two separate datasets. Our first dataset contains monthly billing data from June 2006 to September 2011 on total usage, total expenditure, and rate class for a sample of about 35,000 households. This sample is made up of two groups of households, and was constructed based on usage levels in September 2010: the first group is the population of households with usage above 1500kWh in that month; and the second group is a random selection, of similar size as the first group, of households with usage between 1300 and 1500kWh in that month.[9] This is our primary dataset, which we employ to estimate our main results on the effect of mandatory TOU pricing on total electricity usage and total electricity expenditure.

A shortcoming of our primary billing dataset is that we cannot directly observe the on-peak/off-peak breakdown of household total usage. While we are able to use the billling data to impute this breakdown for households on TOU pricing (as discussed in Section 5.1), this imputation is not possible for households on a flat rate tariff. To estimate peak and off-peak usage for non-TOU households, we therefore rely on our second dataset, referred to as the load profile data. The load profile data comprise hourly usage data between January 2006 and October 2011 for a random sample of 1,300 households present for between 2 and 48 months. Using these data, we can directly calculate the ratio of peak to off-peak usage on a monthly basis for the flat rate households in this sample. We can then use these data to estimate the peak-to-off-peak usage ratio for certain non-TOU households in our primary billing dataset. This allows us to also assess the effect of TOU pricing on peak and off-peak usage separately.

We present a summary of some of our data in the following section. In order to frame this summary, we must first introduce some aspects of our experimental design and related concepts.

[9] The year 2010 was chosen for this rule so that the included households would be most representative of the utility's current customer base. September was chosen because it corresponded to the annual system peak that year.

3 Experimental Design

The purpose of this section is to explain in detail how the TOU program we study gives rise to a regression discontinuity experimental design. However, it should be noted at the outset that only part of our empirical strategy exploits this experimental design. A question of particular interest that we would like to address is whether the mandatory switch to time-of-use pricing induced households to reduce their electricity usage during on-peak hours. But, as noted above, we only observe *total* electricity usage (and *total* electricity expenditure) in our primary billing dataset. We can construct on-peak and off-peak usage on a monthly basis from the hourly usage data in our sencondary load profile dataset, but this sample contains only a small number of households and has a limited panel structure. As the following discussion will make clear, we require data on a large number of households that are present for a large number of consecutive months to leverage the experimental design. Only our billing dataset fulfills these requirements. Our primary empirical strategy will therefore be to identify changes in total usage and total bills that were caused by the switch to TOU by using the billing data to exploit the experimental setting. We will then use information from the load profile data to calculate the effects on peak and off-peak usage that are implied by these changes. We will describe the various components of our empirical strategy in more detail, as well as present the corresponding results, in the next two sections.

The key feature of the regression discontinuity design in general is that assignment to the treatment group is triggered by crossing some threshold. In the neighborhood of the threshold, assignment to the treatment group is effectively random, since idiosyncratic factors will push some individuals over the threshold but not others. Therefore, differences in outcomes between individuals on either side of the threshold can be interpreted as causal treatment effects. On the surface, the TOU program we study corresponds directly to this type of setting, since crossing a usage threshold determines a household's TOU status. However, two features of the program complicate the analogy: the sequential lowering of the threshold over time; and the varying lag across households between crossing a threshold and receiving the TOU treatment.

We address these complications first of all by separating our dataset into three distinct experiments based on the program design: one each corresponding to the 4000kWh, 3000kWh,

and 2000kWh thresholds. For each of these three experiments, we divide our sample into three time periods: the pre-experiment period; the qualification period; and the treatment period. The qualification period for a given experiment encompasses a subset of the months in which a household, by the rules of the program, should trigger its eventual assignment to the TOU treatment group by having total usage in excess of the threshold for that experiment in any 30-day billing period. The treatment period is the focus of our analysis, and encompasses selected months in which at least some households that crossed a threshold during the qualification period ("crossers") have been switched to TOU. The pre-experiment period is defined residually as the set of months before the qualification period back to the beginning of our sample.

We choose the qualification period to be the summer months of the corresponding calendar year. The first motivation for this restriction is that the households that first cross a threshold in the summer begin to be switched to TOU pricing by the fall. If we were to extend the qualification period to capture households that first cross a threshold in the fall, there would be an overlap between the qualification and treatment periods, which would raise the possibility of endogeneity between the variable determining treatment status (i.e. usage during the qualification period) and the TOU treatment itself.

The second motivation for restricting the qualification period to summer months only is that this maximizes the number of households included in the experiment. Given the first motivation, if we were to extend the qualification period to capture households that first cross the threshold early in the year, we would have to end the qualification period at the point that these households start to be switched on to TOU, which typically happens in the summer. This would force us to exclude a large number of households, since the households in our dataset generally have substantial usage peaks in the summer months, and these months are therefore associated with a large volume of households crossing the threshold for the first time.[10]

[10]Neither of these issues is a concern for the 4000kWh experiment, since there are no households in our analysis that were switched to TOU before February 2008, which was after the 4000kWh threshold was no longer binding and the 3000kWh threshold was already in place. We therefore define the qualification period for the 4000kWh experiment as the full set of months for which the 4000kWh threshold was in place, namely November 2006 through December 2007. The qualification period for the 2000kWh experiment includes the summer months of 2009 only, though the 2000kWh threshold remained binding in later years as well. It would be possible to define a separate and additional 2000kWh experiment with a qualification period

We choose the treatment period for a given experiment to be the set of months following the qualification period for which there is a substantial difference between crossers and non-crossers in their propensity to be treated. Though, by rule, all crossers should be switched to TOU pricing within six months of crossing a threshold for the first time, some are switched more quickly than this, some experience delays of several months before being switched, and a few are never switched within the timeframe covered by our sample.[11] At the same time, non-crossers in a given experiment may eventually receive the TOU treatment by virtue of being a crosser in a different experiment.[12] We therefore expect the propensity to be treated to increase over time for *both* groups, and thus for the difference in this propensity across groups to be substantial for a limited window only.

Unlike in a canonical "sharp" regression discontinuity setting, the preceding discussion demonstrates that crossing a usage threshold in the TOU program under consideration is not a perfect determinant of being in the treatment group in any given treatment-period month. Instead, each of the three individual experiments should be viewed as having been generated by the Fuzzy Regression Discontinuity (FRD) design, where the "fuzziness" refers to the imperfection of the crosser/non-crosser distinction as a predictor of a household's TOU status in a given experiment.

As in the general regression discontinuity framework, we will assume that a household's crossing status is exogenous in the neighborhood of the threshold for a given experiment.

corresponding to the summer months of 2010, but we have not explored this in detail, since the subsequent treatment period would be truncated due to our sample ending in the middle of 2011.

[11]The long delays between crossing and switching and the failure to switch some qualifying households altogether appear to be most frequently due to technical and administrative difficulties associated with installing requisite metering equipment. A small number of crossers were switched to TOU but eventually allowed to revert to a non-TOU rate due to a medical exemption; these households have been removed from the analysis. It is possible that some of the crossers that were never switched to TOU were granted a medical exemption pre-emptively, but we cannot observe this.

[12]For example, a household that did not cross the 3000kWh threshold in 2008 might cross the 2000kWh threshold early in 2009 and be switched to TOU on that basis shortly thereafter, potentially even before some of the 3000kWh crossers are switched. Also, while we exclude households from a given experiment that have already received the TOU treatment before the end of the qualification period, we do not exclude households that were crossers in a previous experiment that had not yet received the treatment. It is therefore possible that a non-crosser in a given experiment will receive the treatment due to having been a crosser in an *earlier* experiment. Note that customers could also voluntarily opt in to a TOU pricing regime at any time, so that it is theoretically possible that we could see a household that is never a crosser in any experiment but is nonetheless on TOU. However, we have removed all voluntary switchers from our analysis, since such self-selection into treatment would violate the experimental design and prevent us from identifying causal effects.

And as in the general framework, this will allow us to interpret differences in outcomes between crossers and non-crossers as causal treatment effects – but only after adjusting for the propensity for each group to be treated. These treatment effects can be estimated consistently only for treatment months in which a sufficiently high proportion of crossers is on TOU relative to the proportion of non-crossers on TOU (i.e. in which the crosser/non-crosser distinction is a strong instrument for treatment status).

Before turning to a more precise discussion of how we implement the estimation of these treatment effects, we present summary statistics for households close to the threshold in the qualifying period for each of the three experiments in Table 2.

4 Treatment Effects for Total Usage and Total Bills

4.1 Methods

We begin by comparing crossers to non-crossers along several dimensions, separately for each month in the entire sample and separately by experiment. Specifically, for a given experiment, we estimate

$$Y_i = \beta_0^{Yt} + \beta_1^{Yt} C_i + \beta_2^{Yt} f\left(\tilde{X}_i\right) + \beta_3^{Yt} C_i \times f\left(\tilde{X}_i\right) + \varepsilon_i^{Yt} \tag{1}$$

individually for each month (t) and for various dependent variables Y. The variable C_i is a dummy variable for whether household i is a crosser. The variable $\tilde{X}_i$ is the "forcing variable" that determines whether household i is a crosser. More precisely, $\tilde{X}_i$ is household i's maximum total usage across all 30-day billing periods during the qualification period net of the threshold qualifying level of kilowatt hours for a given experiment. Under the rules of the program, if $\tilde{X}_i$ is strictly greater than zero, household i is a crosser and is thus supposed to receive the TOU treatment eventually.[13] Both of these variables are constant

[13] Formally, let X_{it} be household i's total electricity usage in month t. Further, let usage on a standardized 30-day-billing-period basis be $\ddot{X}_{it} \equiv X_{it}/d_{it} \times 30$, where d_{it} is the number of total days actually in the billing period corresponding to household i's bill in month t. Then $\tilde{X}_i \equiv \left(\max_{t \in \mathbb{Q}}\{\ddot{X}_{it}\} - \bar{X}\right)$, where $\mathbb{Q}$ is the set of months in the qualification period and $\bar{X}$ is the threshold; and $C_i \equiv \mathbb{1}\left\{\tilde{X}_i > 0\right\}$, where $\mathbb{1}\{\}$ is the indicator function.

within households for a given experiment.[14]

The dependent variables we consider are total usage, total bills, and a dummy variable TOU_{it} indicating whether household i was on TOU pricing (i.e. was treated) in month t. Specification (1) allows for a flexible relation between the outcome variable of interest and the forcing variable through the function $f(\cdot)$, and allows this relation to differ for crossers and non-crossers.[15] The parameter β_1^{Yt} measures the effect of being a crosser on the level of outcome variable Y in month t as the distance from the threshold approaches zero, and is interpreted as the Intent to Treat effect (ITT).[16]

The fuzzy regression discontinuity treatment effect for outcome Y in any month t in the treatment period for a given experiment is defined as

$$\tau_{FRD}^{Yt} \equiv \frac{\beta_1^{Yt}}{\beta_1^{TOUt}}. \tag{2}$$

That is, the treatment effect for the outcome of interest is the ratio of the ITT for the outcome of interest to the ITT for the propensity to be treated. It can be estimated by applying two-stage least squares to the following system of equations for any outcome variable Y in a given treatment-period month t:

$$Y_i = \tau_0^{Yt} + \tau_1^{Yt} TOU_i + \tau_2^{Yt} f\left(\tilde{X}_i\right) + \tau_3^{Yt} C_i \times f\left(\tilde{X}_i\right) + \omega_i^{Yt} \tag{3}$$

$$TOU_i = \beta_0^{TOUt} + \beta_1^{TOUt} C_i + \beta_2^{TOUt} f\left(\tilde{X}_i\right) + \beta_3^{TOUt} C_i \times f\left(\tilde{X}_i\right) + \varepsilon_i^{TOUt}, \tag{4}$$

where $\hat{\tau}_{1,2SLS}$ is numerically equivalent to inserting the ITTs estimated via specification (1) into equation (2). Note that we apply two-stage least squares as a computational convenience,

[14] The households included in these regressions are only those with a value of the forcing variable $\tilde{X}_i$ within a selected bandwidth around zero, i.e. households "close to" the threshold. For the results that we present below, we first use a wide bandwidth to visually examine the data, then use an optimal bandwidth for each experiment to estimate the treatment effects. The optimal bandwidth selection is discussed in more detail in Section A.1.

[15] For the results that we present below, we first define $f(\cdot)$ as a fourth-order polynomial to visually examine the data, then as linear to estimate the treatment effects. Within the optimal bandwidth for each experiment, we have not found alternatives to the linear form to qualitatively affect our estimated treatment effects.

[16] These are causal effects by virtue of our assumption that, as the distance from the threshold approaches zero, a household's crossing status is exogenous. This assumption is discussed and evaluated in Section A.2.

not because we are attempting to address endogeneity concerns.[17] Nonetheless, there is a useful analogue to be drawn between our setting and an instrumental variables (IV) setting: just as weak instruments can lead to biased IV estimates, so too will our ability to estimate treatment effects consistently be hampered when crossing status is a weak predictor of TOU status. As discussed in the previous section, this concern will guide our precise definition of the treatment period for each experiment.

4.2 Results

We begin by visually examining the propensity to be treated, total billed amount, and total usage on each side of the threshold for each experiment. Specifically, we estimate specification (1) for selected months following the qualification period for each experiment, including households within a very wide range around the threshold and allowing the relation between the outcome variable of interest and the forcing variable to have a separate quartic form on each side of the threshold. This provides a first look at whether the relation exhibits a discontinuity at the threshold (i.e. an intent to treat effect), and allows us to diagnose any non-linearities in the relation that may complicate the identification of any discontinuity.

We then restrict specification (1) to be linear in the forcing variable and its interaction with crossing status, and include only households within a narrower, optimally-chosen bandwidth around the threshold.[18] This is the form that we use to identify ITTs for each dependent variable for several months before, during, and after the qualification period. To present the large set of results in as compact a form as possible, we graph time series of the set of estimated coefficients from (1) for each of the three dependent variables. More specifically, for dependent variable Y, we graph $\hat{\beta}_0^{Yt}$ – the estimate of outcome Y in month

[17] A household's time-of-use status in a given treatment-period month depends on its crossing status in the preceding qualification period and on unobservable factors. However, crossing status is exogenous at the threshold by assumption, and the unobservable factors are ostensibly exogenous issues related to various meter installation and administrative hurdles faced by the utility. Further, as mentioned previously, we have excluded all voluntary adopters from the analysis. We therefore do not consider concerns about endogeneity between TOU status and either total expenditure or total usage to be present.

[18] The method used to determine the optimal bandwidth is described in Section A.1. A larger bandwidth leads to more precise estimates of the discontinuity. However, a larger bandwidth also means that households further away from the threshold are being used to identify the discontinuity *at* the threshold, which can impart a bias; and this bias can be large if the relation is highly non-linear around the threshold. We choose an optimal bandwidth for a given experiment to apply uniformly for the estimation of all ITTs and treatment effects in each month of the treatment period for that experiment.

t for a non-crosser exactly at the threshold – and $\hat{\beta}_0^{Yt} + \hat{\beta}_1^{Yt}$ – the same for a crosser exactly at the threshold – for every month, also indicating when the difference between the two is statistically significant.

It should be noted that the number of parameters that we estimate separately across treatment months and experiments raises difficulties for the assessment of their statistical significance. In general, as the number of hypotheses tested increases, so does the likelihood of witnessing a rare occurence and rejecting a hypothesis by sheer chance. We do not make any corrections for this "simultaneous inference" problem. Instead, in our discussion of results in this and the following section, we focus on broad patterns that have the strongest statistical support, and avoid drawing conclusions based on conventional assessments of the statistical significance of individual estimates.

4.2.1 4000kWh Experiment

Figure 1 shows the estimated propensity to receive the TOU treatment for crossers and non-crossers in the 4000kWh experiment in July 2008. Being a crosser (i.e. being to the right of the threshold) is clearly a strong predictor of having received the TOU treatment by July 2008, as illustrated by the dramatic discontinuity at the threshold. However, it is not a perfect indicator, as some non-crossers just to the left of the threshold – i.e. whose maximum 30-day usage during the 4000kWh qualification period was very close but did not exceed the 4000kWh threshold – have a small but positive propensity to be treated (due to having crossed the 3000kWh threshold early in 2008). Likewise, a few crossers still had not received the TOU treatment by July 2008, as indicated by the high but less than 100 percent propensity to be treated for households to the right of the threshold.

Figure 2 graphs the analogous discontinuity for each individual month between June 2006 and January 2009 based on a linear specification and the optimal bandwidth for the 4000kWh experiment of 600kWh.[19] The months between the vertical lines delineate the qualification period, and the months further to the left are the pre-experiment period. Since we restrict our sample to households that never received the treatment before the end of the

[19]The bandwidth is symmetric, so encompasses households with a value of the forcing variable between -600kWh and 600kWh. Note that the data in Figure 1 have been smoothed for ease of presentation, so that each point represents several households. The point for July 2008 in Figure 2 is based on straight lines of best fit through the first 7-8 points on each side of the threshold in Figure 1.

qualification period, the propensity to be on time-of-use pricing is zero for both crossers and non-crossers throughout the pre-experiment and qualification periods by construction.[20]

The propensity for crossers to be on TOU pricing increases sharply in February 2008, and continues to increase thereafter. The propensity for non-crossers to be treated remains low until the summer of 2008, but jumps sharply in August and September of that year. This corresponds to households that crossed the 3000kWh threshold early in 2008 being switched to TOU on that basis. By November 2008, the difference in the propensity to be treated across the 4000kWh crossers and non-crossers is too small to support the consistent estimation of treatment effects for total bills and total usage. We therefore choose to limit the treatment period for the 4000kWh experiment to February 2008 through October 2008 inclusive.

Figure 3 shows the estimated total billed amount on each side of the 4000kWh threshold in July 2008. The graph illustrates a dramatic discontinuity, suggesting that a crosser at the threshold had a substantially lower electricity bill than a non-crosser at the threshold (by about $100). While the relation exhibits some non-linearity, particularly for very high levels of the forcing variable, Figure 3 provides fairly clear evidence that this difference is indeed the result of a discontinuity, as opposed to a highly non-linear but continuous relation over the entire range.

Figure 4 shows the estimated ITTs on the total bill over time. The large discontinuity illustrated in Figure 3 for July 2008 is also present for the other summer months in 2008. However, no such discontinuity is present for other months in the treatment period. Figure 4 also illustrates that the estimated total bill was nearly identical for crossers and non-crossers at the threshold throughout the pre-experiment and qualification periods. Such balance on pre-determined covariates is consistent with the intent to treat being randomly assigned at

[20] In the case of the 4000kWh experiment, this restriction is mostly vacuous, since there was no mandatory TOU program in place before the qualification period for the 4000kWh experiment (and thus no way to administer the treatment on a non-voluntary basis) and since, as has been mentioned previously, we have removed voluntary TOU households from our sample. Households with a value of the forcing variable substantially higher than the upper bandwidth cut-off of 600kWh are more likely to have crossed the 4000kWh threshold for the first time early in 2007, and such households were required to have been switched to TOU before the end of 2007. A few of these households were indeed switched in late 2007, but most were not switched until February 2008, along with most of those crossing for the first time in the summer of 2007. The delay in rolling out the program for these larger households (that are not included within the bandwidth we consider in any case) appears to be due to unforeseen technical and administrative issues by the utility.

the threshold, and as such is an important component of a valid experimental design. In this specific case, it supports the proposition that the large ITTs we find in the summer of 2008 are not spuriously caused by some systematic difference in summer usage patterns between crossers and non-crossers.

Figure 5 shows estimated total usage on each side of the 4000kWh threshold in July 2008. Though less dramatic than that for the total billed amount, the graph illustrates a discontinuity for total usage as well. Again, the graph does not suggest that the difference can be attributed to non-linearities in the relation between the forcing variable and the outcome variable.

Figure 6 shows the estimated ITTs on total electricity usage over time for the 4000kWh experiment. Total usage was nearly identical throughout the pre-experiment and qualification periods, and remained so for most of the treatment period as well. Only in the summer of 2008 was there a significant difference in total usage, when crossers at the threshold had lower usage in June and July than non-crossers at the threshold.[21]

Table 3 shows the treatment effects, adjusted for the propensity to be treated, on total usage and total bills for each month in the 4000kWh treatment period. To give a better sense of magnitudes, Table 3 presents each treatment effect as a percentage of the level of the respective variable for non-TOU households at the threshold.[22] As was foreshadowed by the ITTs discussed above, we find that the switch to time-of-use pricing caused very large and statistically significant drops in total electricity expenditure in the summer months of 2008 – by more than 20% in each of the four months and as much as 30% in July. This is matched by

[21] Again, the absence of significant differences in total usage between crossers and non-crossers during the pre-experiment and qualification periods indicates that the differences in June and July 2008 are not driven by pre-existing differences between the groups. In this case, it also indicates that non-crossers were not purposely restraining their usage during the qualification period to avoid crossing the threshold, which would violate the random assignment assumption. Figure 6 does show that non-crossers at the threshold had slightly lower total usage than crossers in early 2007, but this difference is not statistically significant. Section A.2 discusses manipulation of the forcing variable in more detail, and tests for such manipulation more formally.

[22] That is, each entry shows $\hat{\tau}_1^{Yt}/\hat{\tau}_0^{Yt} \times 100$ from a separate two-stage least squares application of equations (3)-(4). Note that we calculate standard errors for the treatment effects based on non-parametric bootstrap methods. While the robust 2SLS covariance matrix is asymptotically valid, we opt to report bootstrapped standard errors for the total usage and total expenditure treatment effects for consistency with the on-peak and off-peak treatment effects that we present below, since the method we use to calculate these latter effects (to be discussed in Section 5.1) necessitates the retrieval of standard errors through bootstrap methods. We discuss the bootstrap methods that we employ in Section A.3.

statistically significant declines in total electricity usage in June and July of a proportionately smaller but nonetheless substantial 9-10%. In the remaining treatment months, the effects for both variables are estimated imprecisely, and are smaller in magnitude.

4.2.2 3000kWh Experiment

Figures 7 through 12 show the same series of ITTs for the 3000kWh experiment as discussed above for the 4000kWh experiment. The propensity of 3000kWh crossers to be treated in July 2009 exhibits some extreme non-linearities throughout the upper range of the forcing variable; the sources of this non-linearity are unknown, but neither is it suggestive of an invalid design.[23] Crossing is a reasonably strong predictor of receiving the TOU treatment between October 2008 and November 2009 inclusive, which is the set of months we therefore define as the 3000kWh treatment period.

Crossers have slightly lower bills and usage than non-crossers at the threshold in October and November 2008. However, this difference is also present in October and November 2007, suggesting that there may be some systematic difference in autumn electricity usage between crossers and non-crossers. We therefore do not focus on the first two months of the treatment period. As in the 4000kWh experiment, crossers at the 3000kWh threshold have lower bills than non-crossers in the summer months of the treatment period. However, unlike the 4000kWh experiment, the difference in summer bills is small and is not accompanied by any significant difference in total summer electricity usage.

Table 4 shows the treatment effects on total bills and total usage for the 3000kWh experiment. For parsimony and consistency across experiments, the table only includes the 3000kWh treatment months in 2009 corresponding to the months in 2008 for which treatment effects could be estimated for the 4000kWh experiment in Table 3. Between July and October 2009 inclusive, TOU caused a drop of 7-12% in total bills. The effects on total bills in other treatment months and on total usage in all months are negligible in magnitude and significance.

[23] The bandwidth we choose for the 3000kWh experiment is 400kWh on either side of the threshold. This corresponds to the first 6-7 bins on either side of the threshold in Figures 7, 9, and 11. The relations are approximately linear over this range in each case.

4.2.3 2000kWh Experiment

Figures 13 through 18 show the same series of ITTs for the 2000kWh experiment as discussed above for the 4000kWh and 3000kWh experiments. The propensity of 2000kWh crossers to be treated is positive but low in October and November 2009, directly following the qualification period. It then jumps above 60 percent in December 2009 and remains at about that level until November 2010. The propensity for 2000kWh non-crossers remains negligible for this period. This is sensible, as the 2000kWh threshold remained in place in 2010 as well, and it is unlikely that non-crossers in the summer of 2009 would exceed the same threshold later in 2009 or early in 2010. On the other hand, summer high temperatures were higher and occurred later in the year in the northeast in 2010 compared to 2009. Consistent with this, the propensity for households that did not cross the 2000kWh threshold in the summer of 2009 to be treated jumps to about 60 percent in December 2010, corresponding to many of them eventually crossing the same threshold late in the summer of 2010. Altogether, crossing in the summer of 2009 is a reasonably strong predictor of receiving the TOU treatment between December 2009 and November 2010 inclusive, which is the set of months we therefore define as the 2000kWh treatment period.[24]

As with the previous two experiments, we see some significant differences between the total bills of crossers and non-crossers in the summer months at the 2000kWh threshold. However, while this difference is negative in July 2010, consistent with July 2008 in the 4000kWh experiment and July 2009 in the 3000kWh experiment, it is positive in September 2010, in contrast to both of the previous experiments. Further, we see a significant ITT outside the summer months, in contrast with the previous two experiments, with the total bill of crossers at the threshold lower than that of non-crossers in March 2010. Crossers also had significantly lower total usage in July 2010 than non-crossers at the threshold, followed by significantly higher total usage in August through October 2010.[25]

Table 5 shows the treatment effects on total bills and total usage for the 2000kWh

[24]The bandwidth for the 2000kWh experiment is 80kWh, which corresponds with the first 4 bins on either side of the threshold in Figures 13, 15, and 17. This is a substantially smaller bandwidth than for the previous two experiments, which reflects the much greater density of consumers with usage in this range.

[25]Both the total bill and total usage of crossers at the threshold were slightly higher than those of non-crossers in May 2009, at the beginning of the 2000kWh qualification period. The sources of these small differences are not obvious, but we do not consider them to be indicative of a systematic difference between crossers and non-crossers that would affect the interpretation of the treatment period results.

experiment. Again, only February through October 2010 are shown, for consistency with Tables 3 and 4. TOU caused a 5% drop in total usage in July 2010, but increases in total usage of 5-10% in each of the following three months. This pattern is matched by the effects for total bills, though only the drop in July is significant. There is also suggestive evidence that TOU caused drops in total bills of about 5-7% in the spring months of 2010.

4.2.4 Summary

These estimated treatment effects on total usage and total bills will serve as the foundation of our estimates of treatment effects on peak and off-peak usage, to be discussed in the following section. But they point to some interesting and potentially important conclusions in themselves, which we attempt to synthesize here.

An encouraging observation is that TOU caused sizeable reductions in total usage in the summer months by the largest households. This is a desirable effect from the perspective of engineering and environmental goals, even if these largest households represent a very small proportion of the overall customer base. Further, TOU caused substantial bill reductions for large households in the summer. This follows naturally from the lower usage, but is also due in part to the increasing-block structure of the non-TOU tariff that was in place in the summers of 2008 and 2009, as shown in Table 1. The high tailblock rate that applied to most of the consumption of these large households implies that the switch to TOU was likely associated with a significant "rate-class discount" – i.e. that bills would be lower under TOU without any change to total usage or its timing. In other words, given the same underlying household behavior, the effects on total bills would not have been as dramatic if the utility had set a lower non-TOU tailblock rate or a higher TOU on-peak rate in these months. Nonetheless, lower expenditure is a desirable effect from the perspective of consumer satisfaction, regardless of its underlying sources.

On the other hand, a disappointing conclusion that emerges consistently across the three experiments is that TOU caused very little change in either total expenditure or total usage outside the summer months. This strongly suggests that there were also no effects on underlying on-peak and off-peak consumption in these non-summer months, which will be largely confirmed by the results presented below.

Finally, the smallest households examined exhibit a curious reversal in their response to TOU. As with the largest households, they display an encouraging drop in total bills and total usage due to TOU in July 2010. However, these negative TOU treatment effects turned disconcertingly and substantially positive for both variables in the late summer and unusually warm autumn of 2010.

5 Implied Treatment Effects for On-Peak and Off-Peak Usage

5.1 Methods

We cannot estimate treatment effects for on-peak and off-peak usage directly because we do not observe these variables in our billing dataset. However, we can take advantage of our load profile data and observed rates to calculate implied treatment effects for on-peak and off-peak usage that are consistent with the estimated effects for total usage and total bills.

We first note that we can use the structure of customers' electric bills to impute a household's on-peak and off-peak usage for months that it is on TOU. When household i is on TOU, its total billed amount E in month t is

$$E_{itT} = p_{tT}^{on} x_{itT}^{on} + p_{tT}^{off} x_{itT}^{off} + f_{tT} \tag{5}$$

where T indicates the TOU pricing regime and x^{on} and x^{off} represent the household's on-peak and off-peak usage respectively.[26] That is, bills depend on a fixed fee f, and on on-peak and off-peak per-kWh charges of p^{on} and p^{off} respectively. Combining this with the fact that on-peak and off-peak usage must sum to the household's observed total usage, X, i.e.

$$X_{its} = x_{its}^{on} + x_{its}^{off} \tag{6}$$

(for either pricing regime $s \in \{T, N\}$), gives two equations in two unknowns. This allows us

[26] We index usage and bills by T because, even though a household is on only one pricing regime in a given month in reality, we wish to contemplate how a household's behavior would change in a given month, varying the pricing regime only.

to solve for on-peak and off-peak usage as functions only of variables that we observe:

$$x_{itT}^{on} = \frac{E_{itT} - f_{tT} - p_{tT}^{off} X_{itT}}{p_{tT}^{on} - p_{tT}^{off}} \quad \text{and} \quad x_{itT}^{off} = \frac{p_{tT}^{on} X_{itT} - f_{tT} - E_{itT}}{p_{tT}^{on} - p_{tT}^{off}}. \tag{7}$$

Note that this imputation is, unfortunately, impossible for non-TOU household-months, as the non-TOU rate is the same for on-peak and off-peak usage, and the non-TOU analogues to the expressions in (7) are hence undefined. Instead, we develop an alternative strategy to estimate on-peak and off-peak usage levels for non-TOU household-months employing our load profile dataset.

Specifically, we first define the peak-to-off-peak usage ratio for non-TOU household-months as

$$\check{x}_{itN} \equiv \frac{x_{itN}^{on}}{x_{itN}^{off}}, \tag{8}$$

where N indicates the non-TOU pricing regime. Along with total usage, this ratio fully determines on-peak and off-peak usage levels, since, using equations 6 and 8,

$$x_{itN}^{on} = \frac{\check{x}_{itN}}{1 + \check{x}_{itN}} \times X_{itN} \quad \text{and} \quad x_{itN}^{off} = \frac{1}{1 + \check{x}_{itN}} \times X_{itN}. \tag{9}$$

We then estimate this ratio for representative non-TOU households by making use of our load profile dataset, in which we observe on-peak and off-peak usage for a small sample of households. To do so, we first calculate $\check{x}$ according to equation 8 for all non-TOU household-months, then calculate sample means by calendar month and household size.[27] We use total usage as a proxy for household size, and we choose the ranges of total usage over which to calculate sample means according to the range spanned by the non-TOU households in our billing data for a given treatment month of a given experiment. We discuss this estimation in more detail in Section A.4.

[27]Our motivation for considering household size and time of year is that we believe them to be important factors determining households' on-peak usage intensity. For households with electric air conditioning and gas or oil heating, which accurately describes most homes in the northeast, the summer months will naturally have higher on-peak usage intensity corresponding to daytime air conditioning. And this will be especially true of larger households, with larger living spaces to cool. Of course, many other factors will also be important determinants of inter-household differences in on-peak usage intensity, but we wish to restrict ourselves to variables that we observe in both our billing and load profile datasets.

With observed rates and levels of total usage and total expenditure, equation (7) can be used to calculate on-peak and off-peak usage for *any* TOU household-month. Likewise, observed levels of total usage and the load profile estimates of the peak-to-off-peak usage ratio can be inserted into equation (9) to calculate on-peak and off-peak usage for *any* non-TOU household-month.[28] Our goal, though, is to calculate on-peak and off-peak usage levels *specifically* for treated and non-treated households exactly at the threshold for a given experiment for the corresponding treatment-period months. To accomplish this, we use the 2SLS estimates of the respective levels of total usage and total expenditure for a given treatment month. That is, we set $X_{itT} = \hat{\tau}_0^{Xt} + \hat{\tau}_1^{Xt}$ and $E_{itT} = \hat{\tau}_0^{Et} + \hat{\tau}_1^{Et}$ in equation (7); and we use $X_{itN} = \hat{\tau}_0^{Xt}$ in equation (9) along with the load profile estimate $\hat{\hat{x}}_{tN}$. For the levels of on-peak and off-peak usage calculated with these estimates, we interpret the difference between the level for a treated household at the threshold and the level for a non-treated household at the threshold as a treatment effect consistent with the treatment effects for total usage and total bills.[29]

5.2 Results

5.2.1 4000kWh Experiment

Table 6 presents the implied treatment effects for on-peak and off-peak usage for the 4000kWh experiment. Consistent with the effects on total usage and total bills, TOU's effects on peak and off-peak usage are limited to the summer months for the most part. The switch to TOU caused substantial declines in on-peak usage in June and July 2008, by about 13%. However, TOU also caused significant though proportionately smaller declines in off-peak usage in the same months. There is also weak evidence of a decrease in on-peak usage in February 2008.

[28] We observe the rate variables in these equations and the peak-to-off-peak usage ratio on a calendar-month basis. In contrast, we observe total usage and total bills on a billing-month basis, i.e. for billing periods that are comprised of parts of consecutive calendar months as determined by the billing cycle that a given household is on. Wherever necessary, we can make use of billing cycle information to align calendar-month and billing-month variables according to the procedure discussed in Section A.5.

[29] A possible alternative to the final step of this method would be to calculate the imputed and estimated on-peak and off-peak usage levels for every household-month in our billing sample, and then to apply the 2SLS estimation procedure to these constructed data series. We have not pursued this alternative because we consider our preferred method to be a more efficient estimation procedure.

5.2.2 3000kWh Experiment

Table 7 presents the implied treatment effects on peak and off-peak usage for the 3000kWh experiment. Corresponding to the absence of any TOU effect on total usage for these households, there is no evidence of any TOU effect on either on-peak or off-peak usage. However, Table 7 does provide suggestive evidence that TOU did not cause a decline in on-peak or off-peak usage in the summer of 2009: the treatment effects are almost uniformly positive, though they are estimated very imprecisely, and a joint test against this one-sided alternative cannot be rejected at conventional significance levels.

5.2.3 2000kWh Experiment

Table 8 presents the implied treatment effects on peak and off-peak usage for the 2000kWh experiment. Once again, TOU's effects on peak and off-peak usage are limited to the summer months. The drop in total usage in July 2010 caused by TOU appears to be more heavily weighted towards a decline in off-peak usage. On the other hand, the increases in total usage in September and October 2010 do not appear to be disproportionately driven by either on-peak or off-peak usage: the conclusion that can be most strongly supported is that TOU caused a decline in neither on-peak nor off-peak usage in either month.

5.2.4 Load Shifting

Load shifting refers to the displacement of some on-peak usage to off-peak hours, or in other words, the substitution of off-peak for on-peak consumption. Such substitution is precisely how economists would predict households to respond to TOU in principle: increasing the on-peak price and decreasing the off-peak price steepens the budget constraint, and, other things equal, should lead utility-maximizing consumers to choose less on-peak consumption and more off-peak consumption. We will pursue this type of consumer theory line of analysis in the following section, but first present estimates of how much load shifting was actually induced by the TOU program.

To summarize the degree of load shifting implied by the on-peak and off-peak treatment effects, we calculate non-parametric elasticities of substitution. That is, we calculate the percentage change in the peak-to-off-peak usage ratio between a TOU and a non-TOU

household at the threshold for a one percent increase in the ratio of the on-peak rate to the off-peak rate, and we do so based on the treatment effects we have estimated and on observed rates, rather than by relying on a model.[30]

We can calculate such implied elasticities of substitution for any treatment month of any experiment. However, for parsimony and in an attempt to identify broad trends, we have focused on two times of year: a winter-spring shoulder period, encompassing February through April; and an extended summer period, encompassing May through October. For each of these two periods, we calculate the average elasticity of substitution (i.e. the elasticity based on the average across months of the underlying variables, rather than the average of monthly elasticities) for each experiment separately.

The average elasticities of substitution for the February-April period are presented in Table 9. There is very weak evidence of a moderate degree of load shifting across the three experiments, but the estimated elasticities are statistically insignificant at conventional levels. The average elasticities of substitution for the May-October period are presented in Table 10. In this period, the point estimate of the elasticity of substitution is largest in absolute value for the largest households, but still implies a drop by less than one tenth of one percent in the peak-to-off-peak usage ratio for a one percent increase in the peak-to-off-peak price ratio.[31] Further, the estimated elasticities of substitution are insignificantly different from zero for all three experiments, and the point estimates are positive for smaller households. That is, moderate degrees of reverse load shifting in the summer months cannot be ruled out.

Curiously, the period in which we see any evidence of load shifting is precisely the period in which we saw no evidence of effects on total bills or total usage. This may indicate that households treat their overall level of early spring electricity consumption as largely non-discretionary, but are nonetheless able to make small changes to its timing over the course of the day that are statistically indiscernible in Tables 6-8. This may also indicate that households treat the *timing* of their summer electricity consumption across hours of

[30] Formally, we calculate $\sigma = \%\Delta\hat{\check{x}}_{N\to T}/\%\Delta\check{p}_{N\to T}$, where $\check{p}_T \equiv p_T^{on}/p_T^{off}$ and $\check{p}_N = 1$.

[31] Based on the rates in Table 1, the TOU peak-to-off-peak price ratio was about 1.5 on average in the summer of 2008, or 50% higher than the non-TOU peak-to-off-peak price ratio. Combined with the point estimate of the 4000kWh elasticity of substitution for this period, this implies that TOU induced an average drop in the peak-to-off-peak usage ratio of about 4%.

the day as largely non-discretionary. We leave further interpretation and discussion of these estimates for the following section.

6 Interpretation

In this section, we take a closer look at the rates in Table 1 from the perspective of the consumption possibilities they afford to households on each rate plan.[32] We present three budget frontiers for July of 2008: based on non-TOU rates and the level of expenditure of the non-TOU household at the 4000kWh threshold (from the respective application of the 2SLS estimation); based on TOU rates and the expenditure of the non-TOU household at the threshold; and based on TOU rates and the expenditure of the TOU household at the threshold.

Figure 19 shows the first two of these budget constraints for the 4000kWh experiment (i.e. for July 2008). The TOU rates/non-TOU expenditure frontier represents the theoretical consumption possibilities available to a household that is switched to TOU pricing and retains the same level of electricity expenditure as before the switch. That is, it represents the set of on-peak/off-peak bundles that a static utility maximizing household would choose from.

Surprisingly, this budget frontier lies completely above the budget frontier corresponding to the same total bill but non-TOU rates. This is driven by the fact that, as shown in Table 1, the non-TOU tailblock rate (the marginal price of both on-peak and off-peak usage faced by these high-consumption households) exceeded *both* the TOU off-peak and on-peak rates.[33] This implies that, regardless of the starting non-TOU bundle, that same bundle would entail lower expenditure under TOU rates.

With any standard utility function, the optimal bundle implied by these budget constraints will involve an unambiguously higher level of off-peak consumption in the TOU scenario relative to the non-TOU scenario. What we see instead, as discussed above, is sup-

[32] These rates, and thus the analysis in this section, reflect fixed and delivery charges only. They do not include generation charges which may differ depending on the generation supplier that each household has selected. Approximately one-half of customers face a non-TOU generation charge, and thus the analysis holds unchanged for them. However, we anticipate that for households on bundled rate with the utility as provider, which imposes a small TOU price gradient, the qualitative conclusions will not change.

[33] All budget constraints account for the fact that households lose the low headblock rate when switched to TOU. This causes a small inward shift of the TOU frontier.

port for precisely the opposite conclusion: that TOU did not cause an increase in off-peak consumption in July 2008. In fact, as shown in Figures 19 and 20, the imputed TOU bundle lies not only within the TOU rates/non-TOU expenditure frontier, but also within the original non-TOU frontier.

One potential hypothesis that could explain this outcome is that households were not responding only to contemporaneous rates, but were also making durable goods investments in response to expected future rates. Alternatively, consumers may have been adjusting on-peak consumption in response to some combination of contemporaneous rates, expected rates, and other incentives, but at the same time, faced technological or attention constraints that prevented them from adjusting off-peak usage separately from on-peak usage.

7 Conclusion

This paper describes a large-scale deployment of time-varying electricity pricing among large residential electricity customers in the northeastern United States. The policy was implemented as part of a statewide energy plan which sought to (among other goals) decrease strain on the electricity grid during hours of peak load. It is, to our knowledge, the first field deployment of such a residential pricing policy in the United States that went beyond voluntary adoption; rules were set whereby high-use households would be mandated onto the TOU rate plan.

An important feature of the analysis is the natural experiment that arises from the structure and implementation of the program. Customers were placed on the TOU rate automatically after breaching the active threshold, creating an appropriate setting in which to apply a regression discontinuity design. This differentiates our research design from most recent existing studies of time-varying electricity pricing, which rely on framed field experiments in which participants are aware of their participation.[34] Thus, our paper offers a novel estimate of how residential consumers behave when exposed to a TOU pricing plan. We also propose a simple method for overcoming a common data shortcoming in this setting: we combine

[34] We refer here to the taxonomy of field experiments proposed by Harrison and List (2004). Wolak (2006), Allcott (2011), and Jessoe and Rapson (2012b) are examples of recent studies of the effect of time-varying pricing that are based on framed field experiments.

our baseline estimates from the regression discontinuity design with a secondary data source on customer load profiles in order to estimate peak and off-peak usage reductions and the elasticity of substitution which they imply.

Our results reveal that some consumers exhibit responses that are consistent with the policymakers' goals and expectations. The largest treated households reduced total usage during summer months by approximately 9 percent, including reductions during peak hours specifically of about 13 percent. Were similar responses also observed in smaller households, the overall capital efficiency gain associated with the program could have been quite large, as the need for oft-idle reserve capacity would have diminished. Unfortunately, this is not the case. The smaller households in our experiments exhibit mixed responses. Households treated at the 2000kWh threshold first reduced their usage in summer months, but then increased it (relative to control) in subsequent months.

Some of our results are not easily explained by standard economic models of consumer choice. For example, the standard model of static utility maximization unambiguously predicts that treated households at the 4000kWh threshold ought to have increased their off-peak usage, yet we can reject all but the smallest of increases. While we may conjecture a number of hypotheses for the responses we observe more generally, our main conclusion is that more research is needed to understand the mechanisms that underpin consumer behavior in this setting.

References

Allcott, H. (2011): "Rethinking Real-Time Electricity Pricing," *Resource and Energy Economics*, 33(4), 820–842.

Borenstein, S. (2005): "The Long-run Efficiency of Real-time of Electricity Pricing," *The Energy Journal*, 26(3), 93–116.

Cameron, C., and P. Trivedi (2009): *Microeconometrics Using Stata*. Stata Press.

Efron, B., and R. Tibshirani (2009): *An Introduction to the Bootstrap*. Chapman & Hall Inc.

Faruqui, A., and S. Sergici (2010): "Household Response to Dynamic Pricing of Electricity: A Survey of the Experimental Evidence," *Journal of Regulatory Economics*, 38, 193–225.

Harrison, G., and J. List (2004): “Field Experiments,” *Journal of Economic Literature*, 42(4), 1009–55.

Hausman, J., M. Kinnucan, and D. McFadden (1979): “A Two-Level Electricity Demand Model: Evaluation of the Connecticut Time-of-Day Pricing Test,” *Journal of Econometrics*, 10, 263–289.

Imbens, G., and K. Kalyanaraman (2012): “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” *Review of Economic Studies*, 79, 933–59.

Jessoe, K., and D. Rapson (2012a): “Commercial and Industrial Demand Response Under Mandatory Time-of-Use Electricity Pricing,” *Working Paper*.

——— (2012b): “Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use,” *Working Paper*.

Joskow, P., and C. Wolfram (2012): “Dynamic Pricing of Electricity,” *American Economic Review*, 102(3), 381–5.

Lee, D., and T. Lemieux (2010): “Regression Discontinuity Designs in Economics?,” *Journal of Economic Literature*, 48, 281–355.

McCrary, J. (2008): “Manipulation of the running variable in the regression discontinuity design: A density test,” *Journal of Econometrics*, 142, 698–714.

Steiner, P. (1957): “Peak Loads and Efficient Pricing,” *The Quarterly Journal of Economics*, 71(4), 585–610.

Williamson, O. (1966): “Peak-load Pricing and Optimal Capacity Under Indivisibility Constraints,” *The American Economic Review*, 56(4), 810–827.

Wolak, F. (2006): “Residential Customer Response to Real-Time Pricing: The Anaheim Critical-Peak Pricing Experiment,” *Working Paper*.

Figures and Tables

Table 1: Electricity Rates, Cents per kWh

	Non-TOU		TOU	
	Headblock	Tailblock	On-Peak	Off-Peak
2008				
Jan.	7.9		10.9	6.9
Jun.	7.9	11.8	11.4	7.6
Jul.	8.6	12.6	12.0	8.1
Oct.	8.6		11.5	7.5
2009				
Jan.	8.8		13.5	6.9
Jun.	8.9	10.9	14.0	7.3
Jul.	8.6	10.6	13.6	7.2
Oct.	8.2		12.3	6.7
2010				
Jan.	9.2		12.6	7.5
Jun.	9.6		14.0	7.5
Oct.	9.2		12.6	7.5

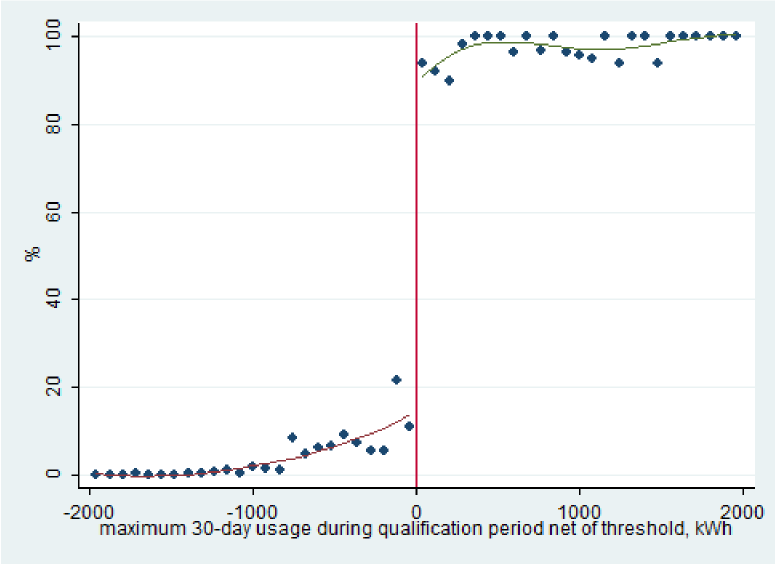
Notes: Prices include distribution, transmission, and delivery charges plus fees only. Actual customer bills also include generation charges, which can vary by customer-chosen supplier. Generation charges account for about 55-65% of the total bill. The headblock is the first 500kWh of total usage in the billing month.

Table 2: Summary Statistics, July of the Qualification Period

	Total Usage (kWh)	Total Bill ($)	Crossers (%)	N
4000kWh Experiment (2007)	3,309	382	0.339	1,096
	[751]	[91]	[0.473]	
3000kWh Experiment (2008)	2,766	333	0.331	2,010
	[387]	[50]	[0.471]	
2000kWh Experiment (2009)	1,494	166	0.421	1,576
	[287]	[31]	[0.494]	

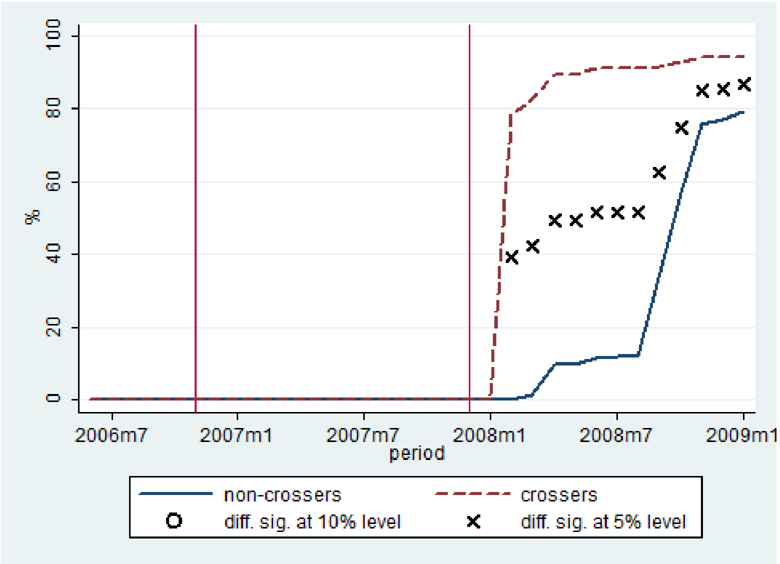
Notes: Standard deviations are in square brackets. The households included for each experiment are those within an optimally-chosen bandwidth around the threshold; see the text for details.

Figure 1: Intent to Treat Effect, Propensity to be Treated, 4000kWh Experiment, July 2008



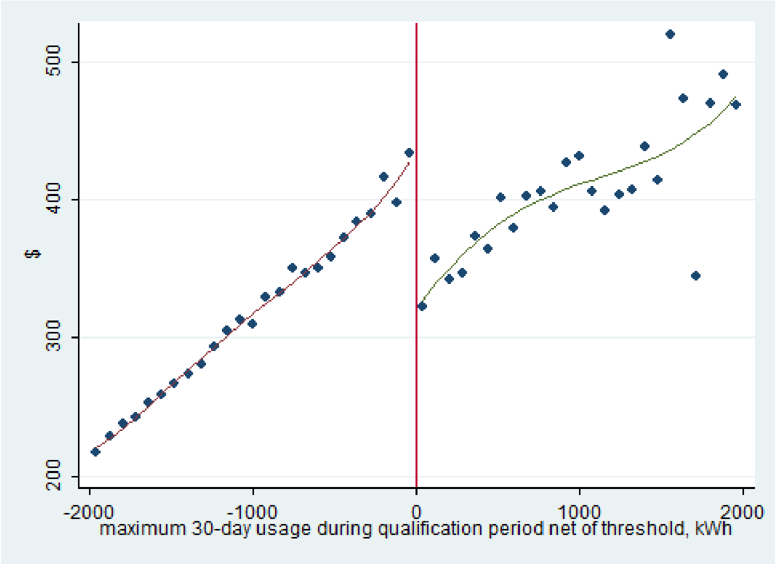
Notes: Data are smoothed into bins of width 80kWh.

Figure 2: Intent to Treat Effects, Propensity to be Treated, 4000kWh Experiment, 600kWh Bandwidth



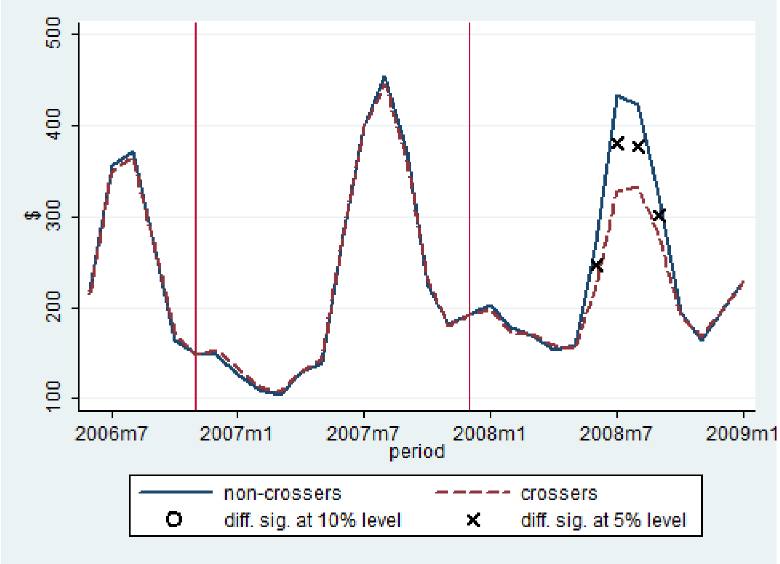
Notes: The qualification period is the set of months between the vertical lines.

Figure 3: Intent to Treat Effect, Total Bill, 4000kWh Experiment, July 2008



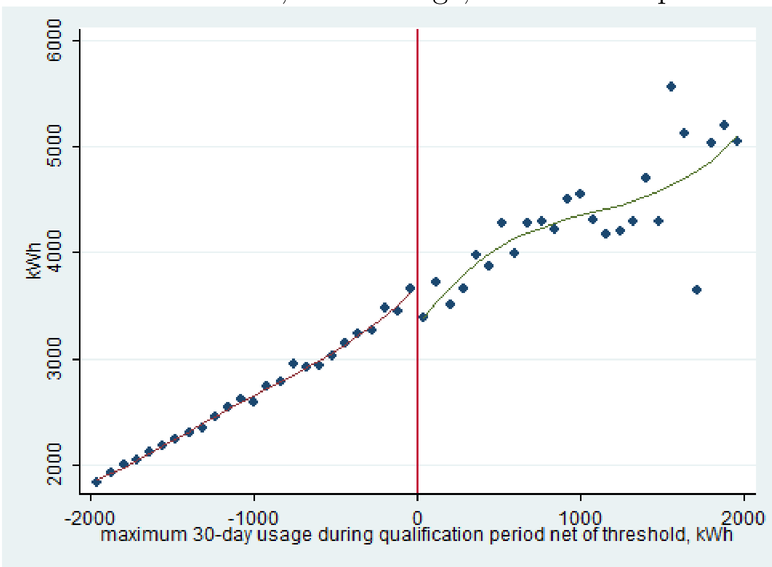
Notes: Data are smoothed into bins of width 80kWh.

Figure 4: Intent to Treat Effects, Total Bill, 4000kWh Experiment, 600kWh Bandwidth



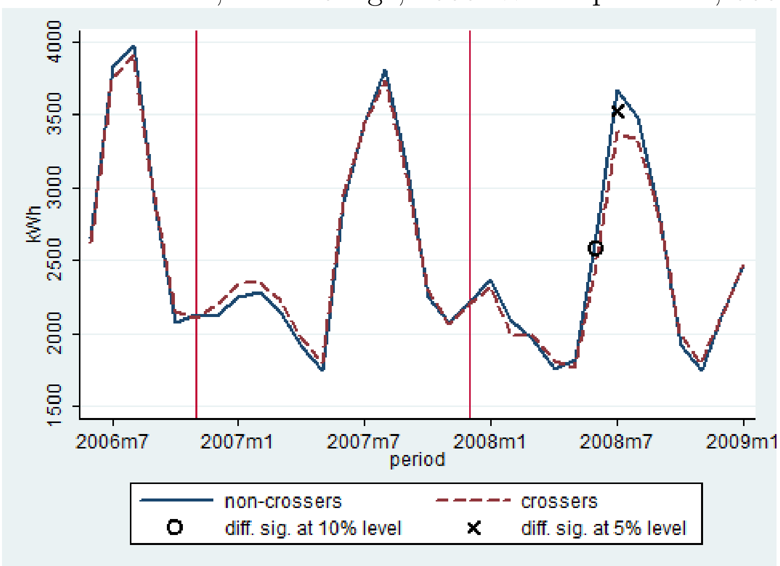
Notes: The qualification period is the set of months between the vertical lines.

Figure 5: Intent to Treat Effect, Total Usage, 4000kWh Experiment, July 2008



Notes: Data are smoothed into bins of width 80kWh.

Figure 6: Intent to Treat Effects, Total Usage, 4000kWh Experiment, 600kWh Bandwidth



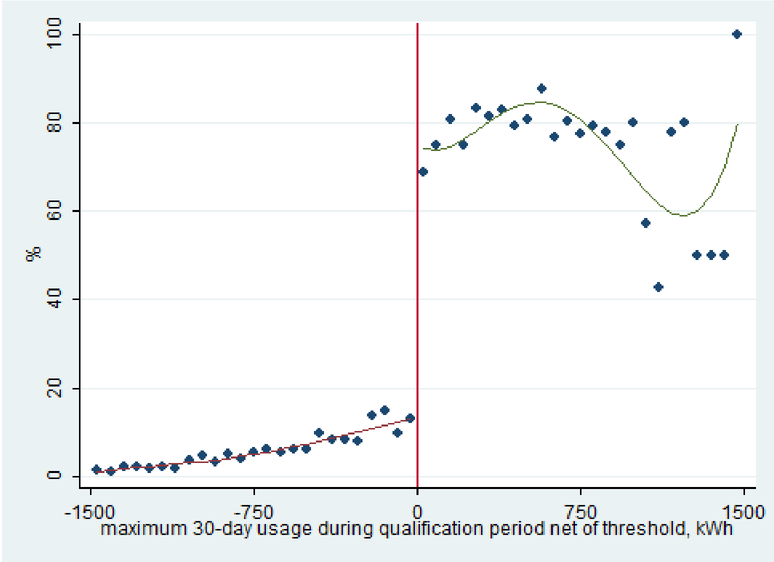
Notes: The qualification period is the set of months between the vertical lines.

Table 3: Treatment Effects (%), 4000kWh Experiment, 600kWh Bandwidth

	Total Usage		**Total Bill**		N
Feb. 2008	-6.21		-4.73		1,106
	(5.85)		(5.42)		
Mar. 2008	1.41		2.28		1,107
	(5.91)		(5.45)		
Apr. 2008	4.04		4.27		1,106
	(6.07)		(5.54)		
May 2008	-3.97		-2.88		1,108
	(5.54)		(5.12)		
Jun. 2008	-9.24	**	-21.50	***	1,105
	(4.69)		(4.19)		
Jul. 2008	-9.85	***	-30.06	***	1,105
	(3.73)		(2.96)		
Aug. 2008	-5.39		-26.15	***	1,107
	(4.04)		(3.08)		
Sep. 2008	-2.11		-22.31	***	1,095
	(5.82)		(4.36)		
Oct. 2008	11.71		-3.75		1,105
	(13.02)		(9.97)		

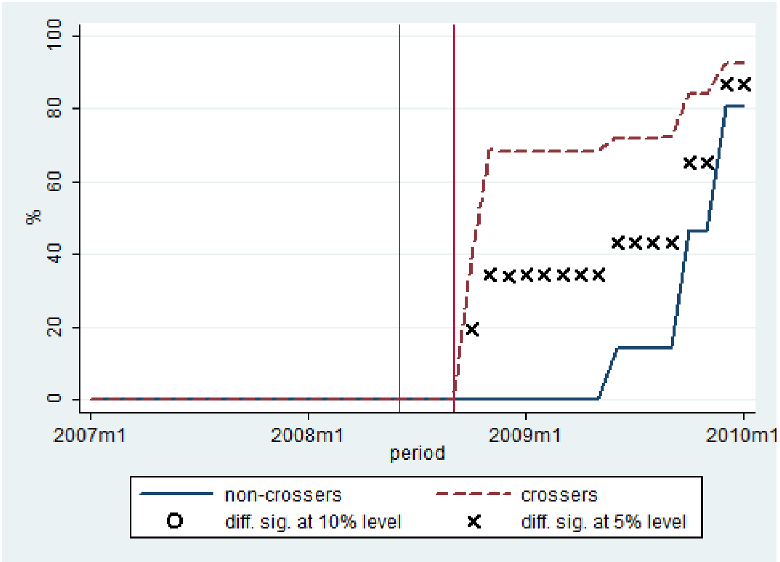
Notes: Standard errors (in parentheses) are based on a non-parametric bootstrap with 1,000 replications. Significance at the 1% (***), 5% (**), and 10% (*) levels is indicated. Each estimate is from a separate regression, and is the estimated TOU treatment effect as a percentage of the estimated non-TOU level at the threshold for the respective dependent variable. Of the 1,105 households included in the regressions for July 2008, 373 are crossers; and the distribution of households is similar in other months.

Figure 7: Intent to Treat Effect, Propensity to be Treated, 3000kWh Experiment, July 2009



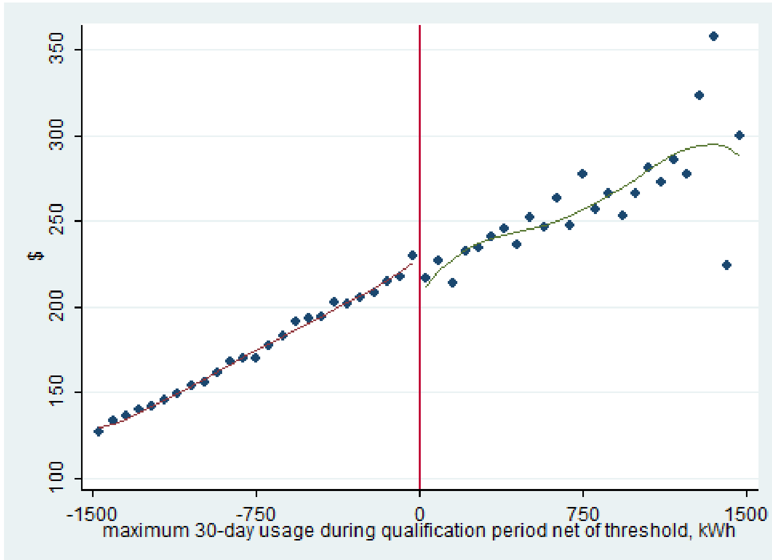
Notes: Data are smoothed into bins of width 60kWh.

Figure 8: Intent to Treat Effects, Propensity to be Treated, 3000kWh Experiment, 400kWh Bandwidth



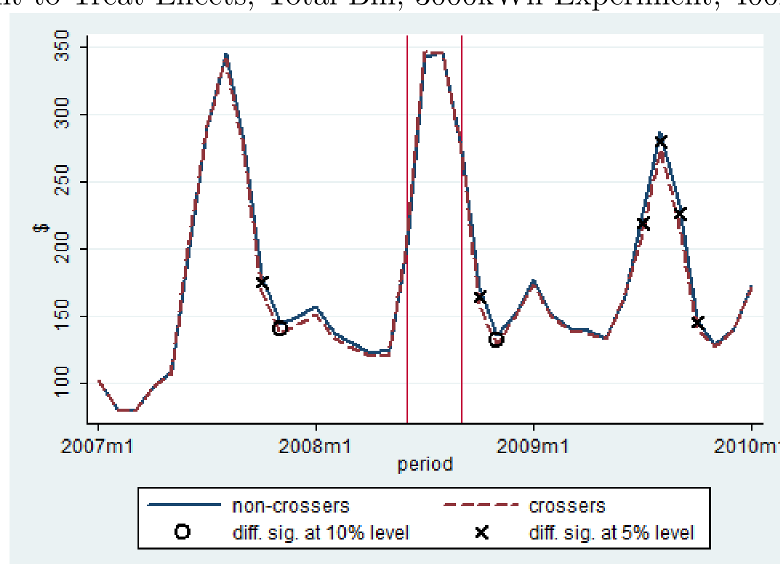
Notes: The qualification period is the set of months between the vertical lines.

Figure 9: Intent to Treat Effect, Total Bill, 3000kWh Experiment, July 2009



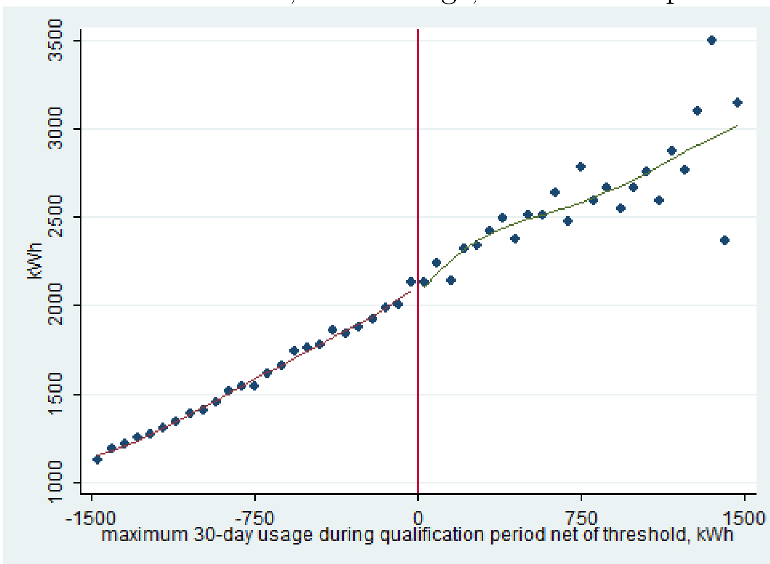
Notes: Data are smoothed into bins of width 60kWh.

Figure 10: Intent to Treat Effects, Total Bill, 3000kWh Experiment, 400kWh Bandwidth



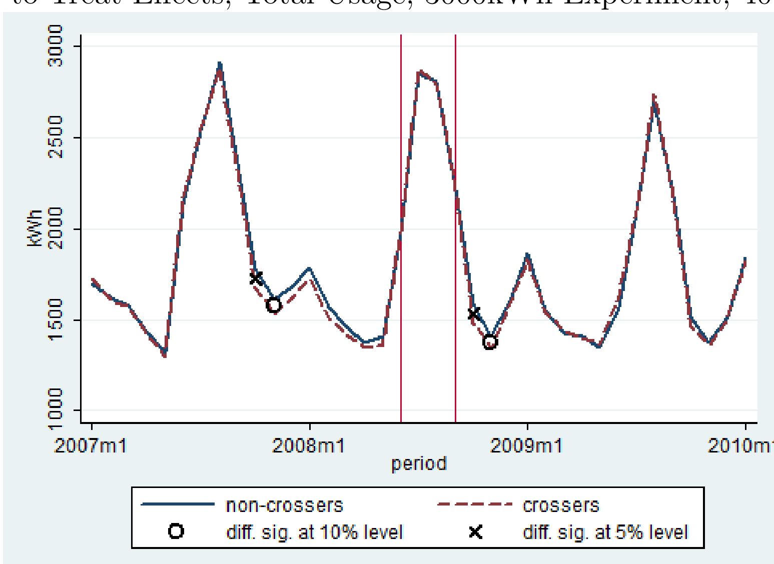
Notes: The qualification period is the set of months between the vertical lines.

Figure 11: Intent to Treat Effect, Total Usage, 3000kWh Experiment, July 2009



Notes: Data are smoothed into bins of width 60kWh.

Figure 12: Intent to Treat Effects, Total Usage, 3000kWh Experiment, 400kWh Bandwidth



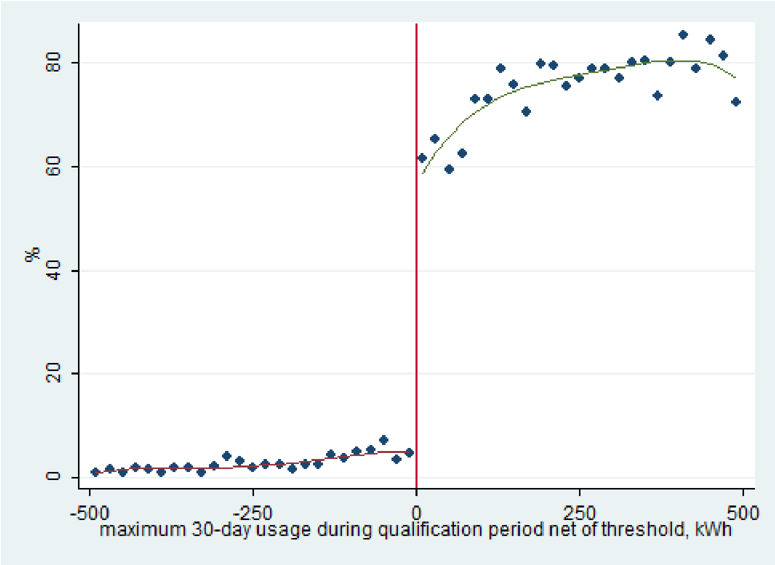
Notes: The qualification period is the set of months between the vertical lines.

Table 4: Treatment Effects (%), 3000kWh Experiment, 400kWh Bandwidth

	Total Usage	**Total Bill**		N
Feb. 2009	-0.66	-2,05		2,011
	(4.59)	(4.04)		
Mar. 2009	1.35	-0.62		2,012
	(4.37)	(3.84)		
Apr. 2009	-0.37	-2.43		2,011
	(3.97)	(3.48)		
May 2009	1.71	-0.09		2,009
	(3.93)	(3.45)		
Jun. 2009	6.26	-0.13		2,009
	(4.65)	(4.20)		
Jul. 2009	-0.33	-10.14	***	2,010
	(3.91)	(3.45)		
Aug. 2009	2.98	-7.28	**	2,009
	(3.38)	(2.97)		
Sep. 2009	0.22	-9.01	***	2,009
	(3.83)	(3.32)		
Oct. 2009	-8.49	-11.60	**	2,007
	(6.23)	(5.54)		

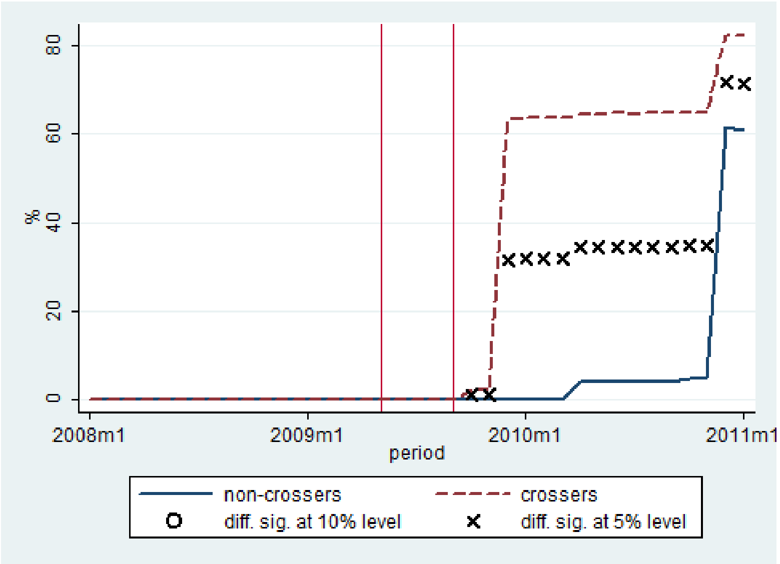
Notes: Standard errors (in parentheses) are based on a non-parametric bootstrap with 1,000 replications. Significance at the 1% (***), 5% (**), and 10% (*) levels is indicated. Each estimate is from a separate regression, and is the estimated TOU treatment effect as a percentage of the estimated non-TOU level at the threshold for the respective dependent variable. Of the 2,010 households included in the regressions for July 2009, 665 are crossers; and the distribution of households is similar in other months.

Figure 13: Intent to Treat Effect, Propensity to be Treated, 2000kWh Experiment, July 2010



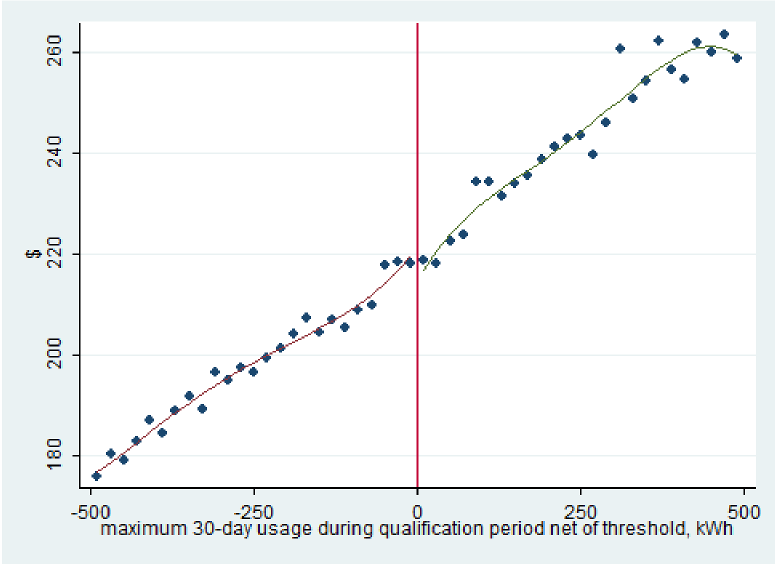
Notes: Data are smoothed into bins of width 20kWh.

Figure 14: Intent to Treat Effects, Propensity to be Treated, 2000kWh Experiment, 80kWh Bandwidth



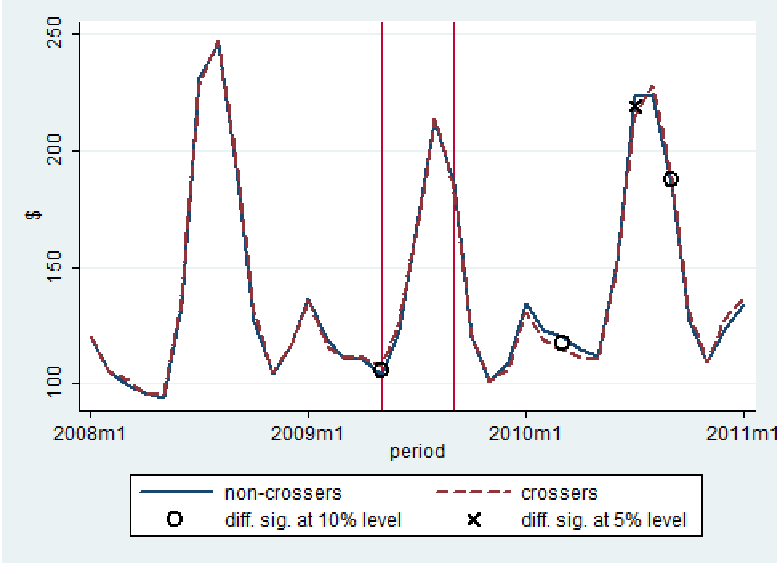
Notes: The qualification period is the set of months between the vertical lines.

Figure 15: Intent to Treat Effect, Total Bill, 2000kWh Experiment, July 2010



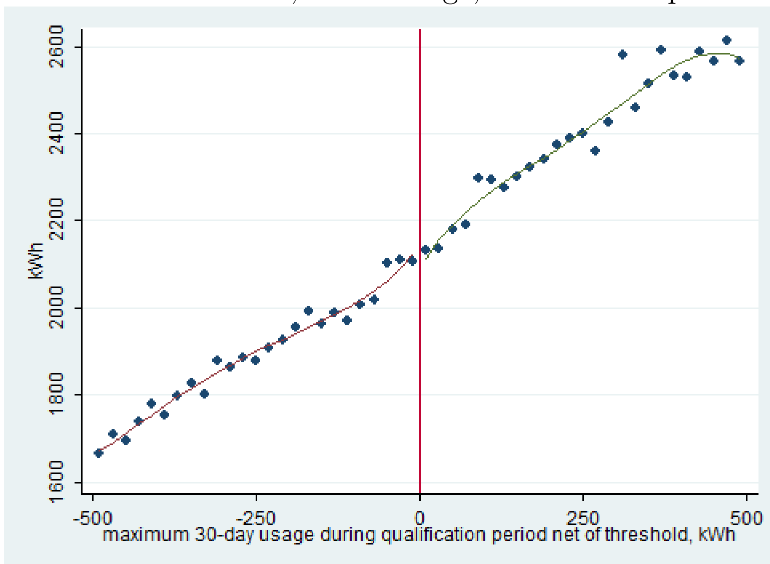
Notes: Data are smoothed into bins of width 20kWh.

Figure 16: Intent to Treat Effects, Total Bill, 2000kWh Experiment, 80kWh Bandwidth



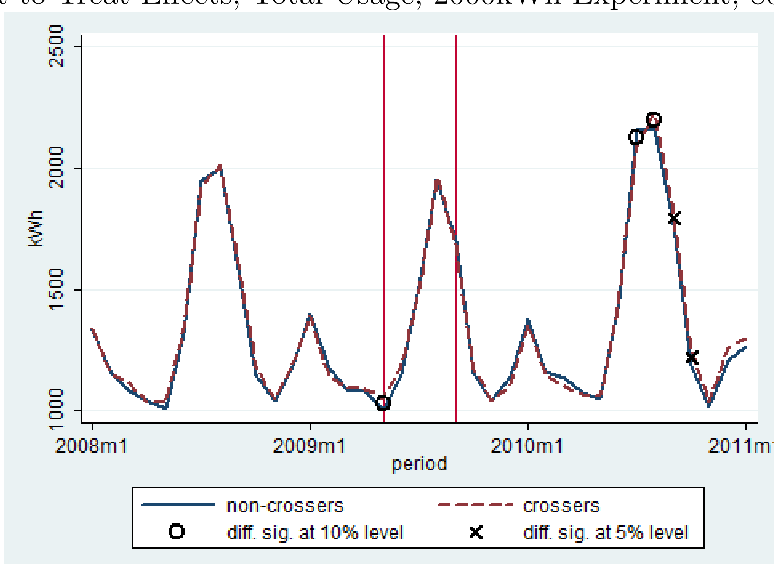
Notes: The qualification period is the set of months between the vertical lines.

Figure 17: Intent to Treat Effect, Total Usage, 2000kWh Experiment, July 2010



Notes: Data are smoothed into bins of width 20kWh.

Figure 18: Intent to Treat Effects, Total Usage, 2000kWh Experiment, 80kWh Bandwidth



Notes: The qualification period is the set of months between the vertical lines.

Table 5: Treatment Effects (%), 2000kWh Experiment, 80kWh Bandwidth

	Total Usage		**Total Bill**		N
Feb. 2010	-2.24		-5.44		1,574
	(4.85)		(4.11)		
Mar. 2010	-4.52		-7.18	*	1,570
	(4.55)		(3.86)		
Apr. 2010	-1.25		-4.60		1,574
	(4.51)		(3.77)		
May 2010	1.99		-2.02		1,574
	(4.49)		(3.74)		
Jun. 2010	-0.63		-3.19		1,574
	(3.54)		(3.22)		
Jul. 2010	-5.17	*	-6.44	**	1,576
	(3.07)		(2.85)		
Aug. 2010	4.68	*	3.12		1,574
	(2.56)		(2.38)		
Sep. 2010	5.60	**	3.91		1,575
	(2.63)		(2.38)		
Oct. 2010	9.34	**	4.78		1,576
	(4.22)		(3.60)		

Notes: Standard errors (in parentheses) are based on a non-parametric bootstrap with 1,000 replications. Significance at the 1% (***), 5% (**), and 10% (*) levels is indicated. Each estimate is from a separate regression, and is the estimated TOU treatment effect as a percentage of the estimated non-TOU level at the threshold for the respective dependent variable. Of the 1,576 households included in the regressions for July 2010, 664 are crossers; and the distribution of households is similar in other months.

Table 6: Implied Treatment Effects (%), 4000kWh Experiment, 600kWh Bandwidth

	On-Peak Usage	**Off-Peak Usage**
Feb. 2008	-9.81 *	-4.90
	(5.76)	(6.07)
Mar. 2008	-0.48	2.07
	(6.09)	(6.06)
Apr. 2008	2.46	4.57
	(6.68)	(6.13)
May 2008	-6.51	-3.08
	(6.29)	(5.54)
Jun. 2008	-13.27 *	-7.68 *
	(6.80)	(4.46)
Jul. 2008	-12.45 *	-8.78 **
	(6.35)	(3.46)
Aug. 2008	-6.64	-4.86
	(6.13)	(3.99)
Sep. 2008	-4.83	-1.01
	(8.11)	(5.57)
Oct. 2008	7.16	13.39
	(13.80)	(13.03)

Notes: Standard errors (in parentheses) are based on a non-parametric bootstrap with 1,000 replications. Significance at the 1% (***), 5% (**), and 10% (*) levels is indicated. Each estimate is from a separate application of the method discussed in the text, and is the estimated TOU treatment effect as a percentage of the estimated non-TOU level at the threshold for the respective variable.

Table 7: Implied Treatment Effects (%), 3000kWh Experiment, 400kWh Bandwidth

	On-Peak Usage	**Off-Peak Usage**
Feb. 2009	-3.32	0.32
	(4.70)	(4.72)
Mar. 2009	-0.86	2.13
	(4.66)	(4.41)
Apr. 2009	-2.19	0.26
	(4.42)	(3.98)
May 2009	1.32	1.85
	(4.43)	(3.93)
Jun. 2009	5.53	6.55
	(5.89)	(4.51)
Jul. 2009	-1.50	0.16
	(5.18)	(3.72)
Aug. 2009	5.14	2.03
	(4.50)	(3.33)
Sep. 2009	4.49	-1.54
	(5.33)	(3.79)
Oct. 2009	-8.00	-8.67
	(7.09)	(6.16)

Notes: Standard errors (in parentheses) are based on a non-parametric bootstrap with 1,000 replications. Significance at the 1% (***), 5% (**), and 10% (*) levels is indicated. Each estimate is from a separate application of the method discussed in the text, and is the estimated TOU treatment effect as a percentage of the estimated non-TOU level at the threshold for the respective variable.

Table 8: Implied Treatment Effects (%), 2000kWh Experiment, 80kWh Bandwidth

	On-Peak Usage	**Off-Peak Usage**
Feb. 2010	-6.89	-0.51
	(5.23)	(4.86)
Mar. 2010	-5.50	-4.17
	(5.15)	(4.48)
Apr. 2010	-2.15	-0.94
	(5.12)	(4.46)
May 2010	-0.10	2.72
	(5.22)	(4.42)
Jun. 2010	0.69	-1.14
	(4.56)	(3.44)
Jul. 2010	-4.00	-5.66 *
	(3.96)	(3.05)
Aug. 2010	5.23	4.43
	(3.64)	(2.75)
Sep. 2010	9.18 **	4.09
	(4.00)	(2.90)
Oct. 2010	7.65	9.98 **
	(4.92)	(4.34)

Notes: Standard errors (in parentheses) are based on a non-parametric bootstrap with 1,000 replications. Significance at the 1% (***), 5% (**), and 10% (*) levels is indicated. Each estimate is from a separate application of the method discussed in the text, and is the estimated TOU treatment effect as a percentage of the estimated non-TOU level at the threshold for the respective variable.

Table 9: Implied Average Elasticities of Substitution, February through April

4000kWh Experiment (2008)	-0.059
	(0.051)
3000kWh Experiment (2009)	-0.032
	(0.022)
2000kWh Experiment (2010)	-0.046
	(0.033)

Notes: Standard errors (in parentheses) are based on a non-parametric bootstrap with 1,000 replications. Significance at the 1% (***), 5% (**), and 10% (*) levels is indicated.

Table 10: Implied Average Elasticities of Substitution, May through October

4000kWh Experiment (2008)	-0.083
	(0.089)
3000kWh Experiment (2009)	0.018
	(0.030)
2000kWh Experiment (2010)	0.014
	(0.034)

Notes: Standard errors (in parentheses) are based on a non-parametric bootstrap with 1,000 replications. Significance at the 1% (***), 5% (**), and 10% (*) levels is indicated.

Figure 19: Budget Lines, Utility Maximization, 4000kWh Experiment, July 2008

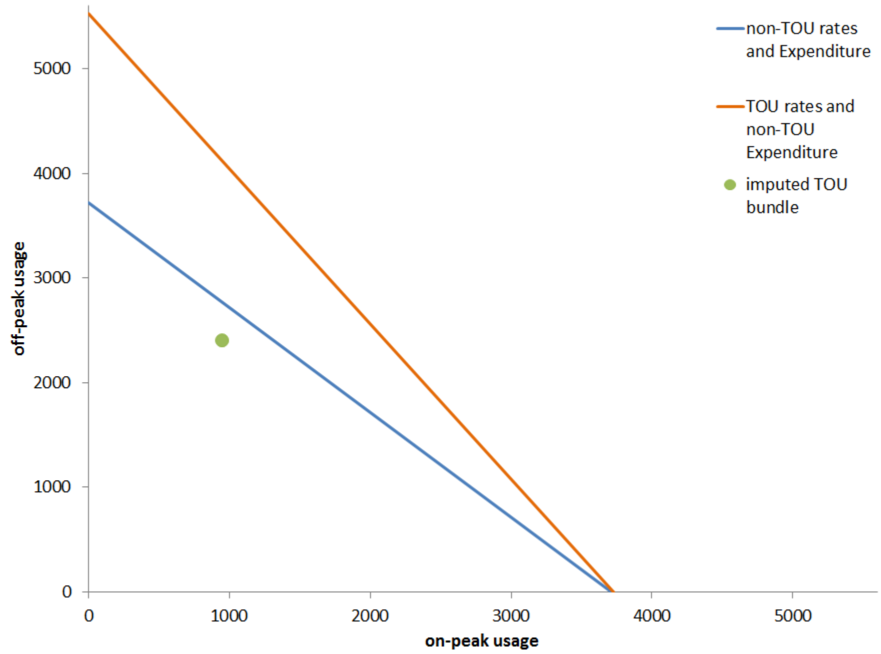


Figure 20: Budget Lines, Expenditure Minimization, 4000kWh Experiment, July 2008

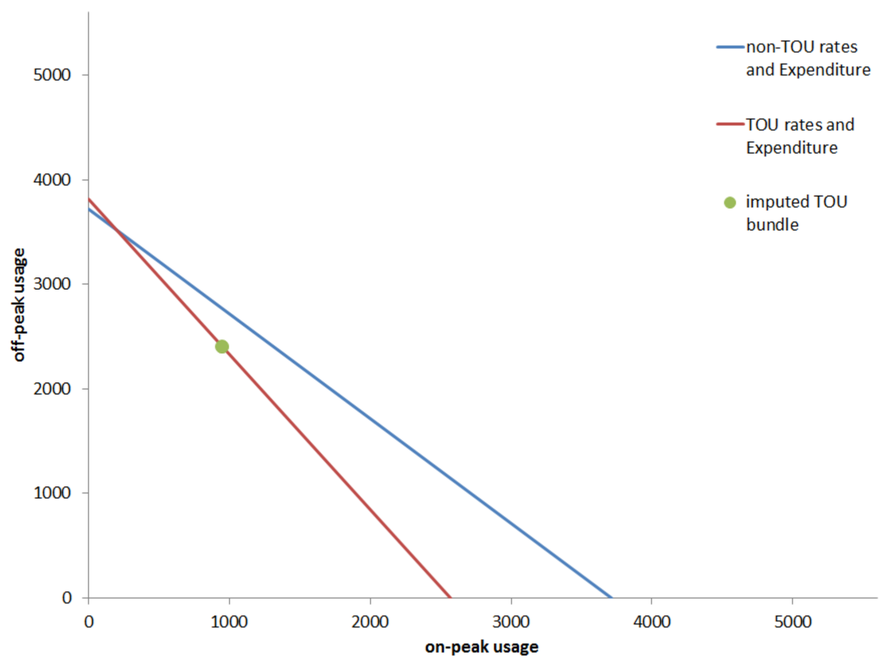


Figure 21: Sensitivity to Bandwidth, 4000kWh Experiment, Total Bill, July 2008

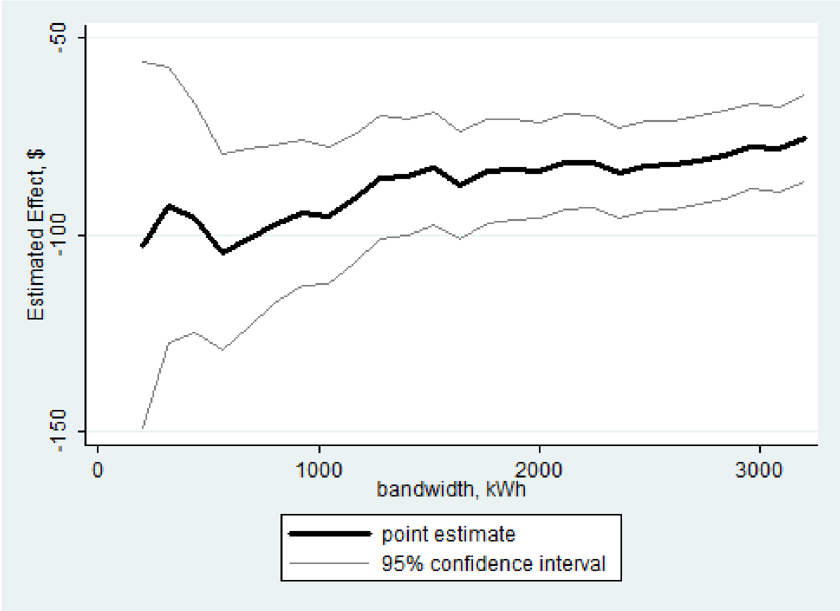
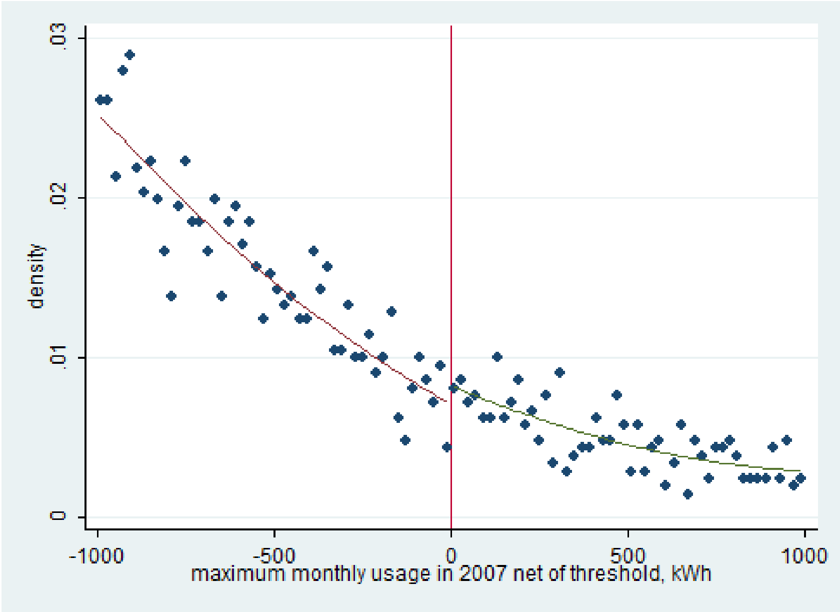
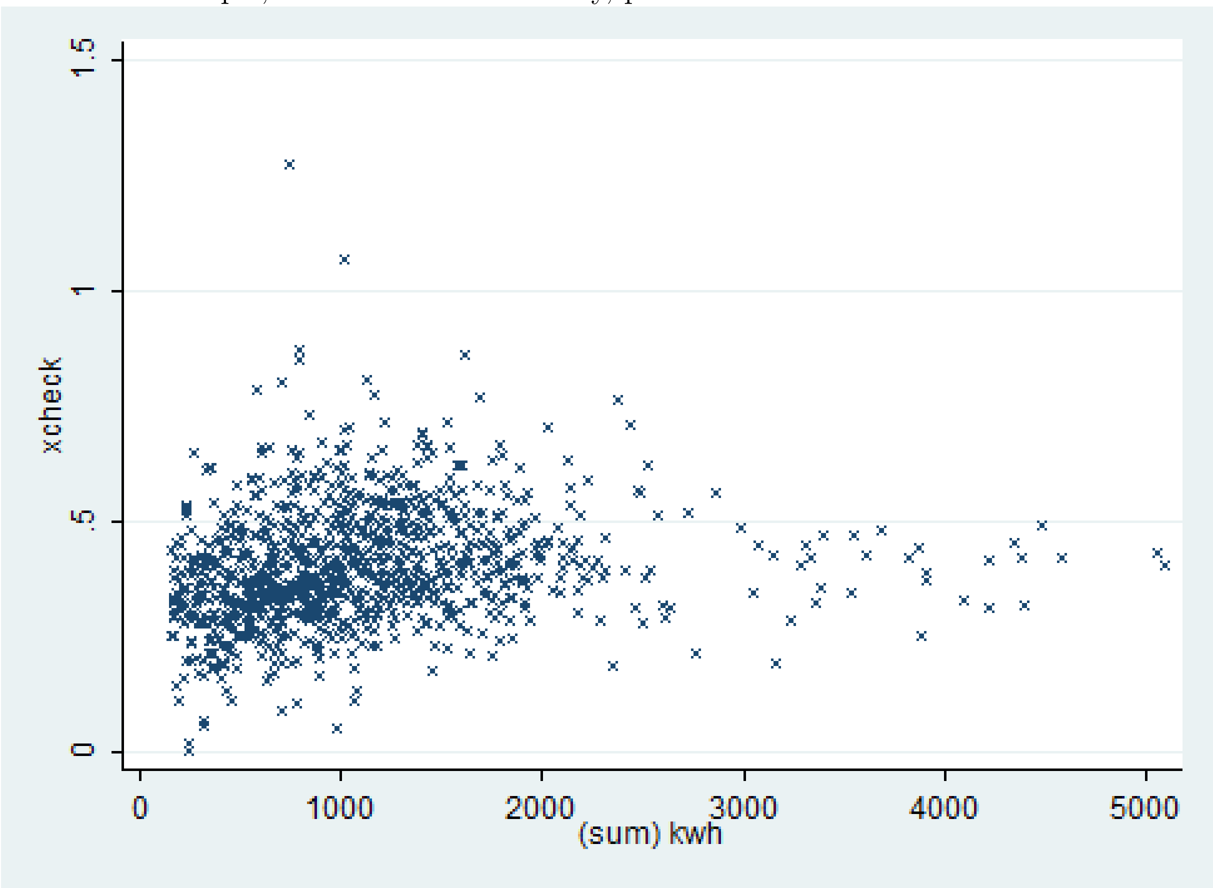


Figure 22: Density of Forcing/Running Variable, 4000kWh Experiment



Note: Data are smoothed into bins of width 20kWh. Separate quadratic predictions on each side.

Figure 23: Peak-to-Off-Peak Usage Ratio by Total Usage for Non-TOU Households in the Load Profile Sample, Calendar Month of July, pooled over 2006-2011



A Appendix

A.1 Bandwidth

The trade-off involved with increasing the bandwidth is as follows: on the positive side, the precision of the estimate is improved; on the negative side, a bias is imparted on the estimate of the effect *at* the threshold by including observations further away from the threshold. As discussed by Lee and Lemieux (2010), when the relationship between the forcing variable and the outcome variable is approximately linear on both sides of the threshold, the bias concern becomes less prominent (and, therefore, the optimal bandwidth exercise less useful).

Lee and Lemieux (2010) suggest a plug-in rule-of-thumb bandwidth that we implement in order to derive the optimal bandwidths used in the text and figures. Imbens and Kalyanaraman (2012) provide a completely data-driven approach to selecting an optimal bandwidth, which we propose to implement in the future. In either case, we wish to adopt a uniform bandwidth for every month, dependent variable, and estimator (ITT or treatment effect) for a given experiment. To do so, we apply the two-stage rule-of-thumb procedure with a quartic form common to each side of the threshold repeatedly for various treatment months of a given experiment and the ITT specification with total usage and total expenditure as dependent variables. From the set of optimal bandwidth estimates thus produced, we informally choose one in the lower range to apply uniformly in the estimation of all ITTs and treatment effects for that experiment.

Figure 21 shows the ITT on the total bill in July 2008 for the 4000kWh experiment, with 95% confidence bounds, for bandwidths ranging from 200kWh to 3200kWh. The graph shows a rapid tightening of the confidence interval and relative stability in the absolute magnitude of the point estimate up to a bandwidth of about 1000kWh; followed by a decrease in the absolute magnitude of the point estimate and moderate tightening of the confidence interval thereafter. Correspondingly, as shown in Figure 3, beyond a value of the forcing variable of about 1000kWh to the right of the threshold, the relation with expenditure becomes quite non-linear, indicating, along with Figure 21, that bias is becoming a more prominent concern than precision. On the other hand, Figure 15 shows a nearly linear relation over a very wide range for the 2000kWh experiment. Indeed, we find little variation in point estimates as we increase the bandwidth for this experiment, though neither do we find a dramatic improvement in the precision of the treatment-period estimates. Unsurpisingly, we

also do not find the optimal bandwidth exercise to be very worthwhile for this experiment, as the rule-of-thumb estimate varies widely by month and outcome variable.

A.2 Manipulation of the Forcing Variable

In order for our Fuzzy Regression Discontinuity treatment effect estimates to be valid, a household's crossing status during the qualification period must be exogenous to the dependent variables of interest during the treatment period. For example, there must be no systematic differences across the groups that would lead to some households being both less likely to have crossed the 4000kWh threshold in 2007 and less likely to have had lower bills in the summer of 2008. One possible difference could arise if some large households were able to manipulate their usage in order to avoid crossing the 4000kWh threshold in 2007, but then did not engage in such manipulation in 2008, perhaps because breaching the lower 3000kWh threshold would have been too difficult to avoid.

If there were "bunching" in the density of households just below the threshold, this might indicate that a substantial number of households were able to manipulate their usage along these lines in order to avoid crossing. However, it would seem reasonable to assume that households have only imprecise control over their exact usage in any 30-day billing cycle, since precise control would likely require sophisticated equipment for monitoring and regulating usage. And as described by Lee and Lemieux (2010), any such lack of precise control implies that assignment (to the group of crossers or non-crossers, in our case) is "as good as random" at the threshold.

We present visual evidence that supports the exogeneity assumption in Figure 22 for the 4000kWh experiment, and additional figures that we have suppressed to save space provide similar evidence for the 3000kWh and 2000kWh experiments as well. McCrary (2008) suggests a more formal test, which we propose to implement in the future.

A.3 Bootstrapped Standard Errors

We use nonparametric bootstrap methods to perform statistical inference at several points in our analysis. The treatment effects for total usage and total bills are estimated in levels but reported as percent changes, the standard errors for which are calculated via bootstrap. Inference on imputed treatment effects for on-peak and off-peak usage, and the substitution elasticities which they imply, are also retrieved via bootstrap. These statistics are calculated

from estimates derived from two different datasets, making analytical solutions for standard errors difficult (if not impossible) to retrieve. Bootstrap methods allow us to circumvent these challenges, and in this section we describe the sampling method that we have used.[35]

Let us first discuss the estimation of standard errors for the 2SLS treatment effects specifications, which rely solely on the billing dataset. Let w_i denote the full time series of data for household i, $w_i = (X_i, E_i, TOU_i, C_i, \tilde{X}_i)$ (corresponding to the notation in equation 1, where Y referred generically to either total usage (X), total expenditure (E), or the treatment indicator (TOU)). We draw a bootstrap sample of size N by sampling $w_1, \ldots, w_N$ with replacement at the household level from the subsample of the billing dataset corresponding to the optimal bandwidth restriction for a given experiment. Denoting the bootstrap sample by $w_1^*, \ldots, w_N^*$, we calculate an estimate, $\hat{\theta}^*$, of the vector of parameters of interest, θ, and repeat this for B separate bootstrap samples.[36] Given the B bootstrap estimates, $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$, we calculate the bootstrap estimate of the variance-covariance matrix according to

$$\hat{V}_{Boot}(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}_b^* - \bar{\hat{\theta}}^*\right) \left(\hat{\theta}_b^* - \bar{\hat{\theta}}^*\right)' \tag{10}$$

where $\bar{\hat{\theta}}^* = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b^*$.

For the analyses that draw upon both the billing data and the load profile data, an additional layer of simulation is required. The billing data bootstrap sample is as described above, and we follow an identical protocol for the load profile bootstrap sample. Note that we draw with replacement at the household level for the load profile samples despite using the pooled estimation method outlined in Section A.4, allowing us to account for within-household variation in an appropriate (conservative) way. Let v_{it} correspond to an observation for load-profile household i in month t, such that $v_{it} = (\breve{x}_{it}, X_{it}, month)$. Recall that we use the load profile data to calculate sample means of the peak-to-off-peak usage ratio ($\breve{x}$) by calendar month and various ranges of total usage.

Having retrieved B bootstrap samples by once again sampling with replacement, we randomly create pairs of billing and load profile bootstrap samples such that each of the samples is used once (as is appropriate for independent draws).[37] Together, these allow us to calculate B bootstrap estimates, $\hat{\gamma}^*$, of the true parameter vector γ. Having now retrieved B

[35]In both notation and procedure, what follows draws upon Cameron and Trivedi (2009).

[36]For all of the estimates in the text, we draw $B = 1000$ bootstrap samples.

[37]See Efron and Tibshirani (2009), pp88-90.

bootstrap estimates $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ and $\hat{\gamma}_1^*, \ldots, \hat{\gamma}_B^*$, we are able to combine individual parameters as necessary to calculate B draws from the distribution of peak and off-peak usage changes and the corresponding substitution elasticities. We then calculate the second moment of the distribution of these random variables non-parametrically as in equation 10.

A.4 Estimation of the Peak-to-Off-Peak Usage Ratio

As discussed in Section 5.1, the information we wish to obtain from the load profile data, conceptually, is the peak-to-off-peak usage ratio for a representative non-TOU household for a given treatment month of a given experiment. We have explored various methods for implementing the estimation of this parameter, and have presently settled on simply estimating sample means by calendar month for ranges of total usage. One concern that this simple approach mitigates is the sparseness of data in the load profile dataset in the total usage ranges we are most interested in. This sparseness is illustrated in Figure 23, which shows that most observations in the load profile data for all months of July between 2006 and 2011 inclusive correspond to total monthly usage below 1500kWh.

We therefore pool all household-month observations to maximize the number at our disposal, and calculate sample means by month-of-year, but not for month and year separately, and not for month-year combinations. This has the disadvantage of treating a calendar month in one year symmetrically with other years, which may not be warranted if average temperatures over a season vary from year to year; but it at least improves the precision of the month estimates by pooling observations across years and households. We choose to calculate sample means over fairly wide ranges of total usage rather than estimate a parametric relation over such ranges or the entire range because no such relation is apparent from Figure 23 or its counterpart for other calendar months. Further, the sparseness of the data corresponding to high usage levels would make such estimation susceptible to overfitting. Finally, it should be noted that the load profile households are, in general, not present in the sample for a long enough period to facilitate controlling for intra-household variation.

For a given treatment month of a given experiment, we examine the distribution of total usage in that treatment month across all non-TOU households in the billing data with a level of the forcing variable within the bandwidth for that experiment. We then define the range of interest as that between the 5th and 95th percentiles of this distribution, and calculate the sample mean of the peak-to-off-peak usage ratio over this range and for the calendar month

corresponding to the given treatment month in the load profile data. Finally, we use this sample mean (adjusted for billing cycle, as discussed in the following section) as an input to equation (9) when carrying out the final step of the procedure discussed in Section 5.1 to estimate non-TOU on-peak and off-peak usage levels at the threshold.

A.5 Billing Cycles

There were 17 distinct billing cycles for residential customers over the period covered by our dataset. Each billing cycle corresponds to a given day of the month (which can change by a couple of days in either direction depending on month and year, due to weekends and holidays) on which the meter is read and the billing period for customers on that billing cycle closes. For customers on billing cycle 1, the total usage and total bill data for "July 2008", for example, correspond to usage that mostly happened in the calendar month of June; on the other hand, for customers on billing cycle 17, total usage and total bill data for "July 2008" correspond to usage that mostly happened in the calendar month of July. There is thus heterogeneity in our billing data in what "July 2008" (and every other month) refers to. This is relevant because we only have rate information on a calendar-month basis. So the total bill in "July 2008" depends on a weighted average of the rates that were in place in the calendar month of June and those that were in place in the calendar month of July, with the appropriate weight depending on which billing cycle a household is on. We describe here how we retrieve billing cycle weights by household-month, and how we apply the weights thus retrieved to align variables observed on a calendar-month basis with variables observed on a billing-month basis.

We reconstruct the total billed amount for all non-TOU household months based on the observed rates, total usage, and the unknown weight; then solve for the weight that exactly aligns the reconstructed total billed amount with the observed total billed amount for each individual household-month. (We cannot do the same for TOU household-months because we do not observe the on-peak/off-peak breakdown of total usage. We can also not perform the calculation for months in which there was no rate change from the previous month.) A few households chronically had weights outside the sensible 0-1 range in the months for which weights could be calculated, and have been dropped completely from all analysis; a few remaining households occasionally had a month with a nonsensical weight, in which case it was just the single household-month observation that was dropped.

Finally, we calculate average billing cycle weights by billing cycle-month-year group over all household-months we could calculate the initial weights for; fill in the missing month-years (i.e. months across which there were no rate adjustments) with annual averages; then apply the appropriate billing cycle-month-year average to every corresponding non-TOU and TOU household-month. (We observe which billing cycle each household was on in September 2010, and know that households are supposed to always stay on the same billing cycle.)

We need to account for billing cycles in the imputation of on-peak and off-peak usage for the TOU household at the threshold and the estimation of the same for the non-TOU household at the threshold. The observed rates are for calendar months, and the estimation of $\check{x}$ for representative non-TOU households in the load profile dataset has also been performed (by necessity, due to the lack of billing cycle information on the households in the load profile sample) on a calendar-month basis. Meanwhile, the households used in the 2SLS estimation of treatment effects are on heterogeneous billing cycles, so that "July" actually includes a small part of the calendar month of June for some households and a larger part for other households. We align billing-month estimates with calendar-month rates and peak-to-off-peak usage ratio estimates by taking a weighted average of the latter across the relevant months. For consistency with the rest of the procedure we use to estimate on-peak and off-peak treatment effects, the weight we use in the calculation must be the billing cycle weight for a non-TOU or a TOU household (as appropriate) *at the threshold*. This is furnished by once more applying 2SLS estimation to equations (3)-(4), this time with average billing cycle weights as the outcome variable of interest.