Contents

List of Tables

1 Regression Discontinuity (RD) Design

In this preliminary analysis about electricity consumption of residential consumers, the impact of electricity utilization in billing period 0, which determines the unit price in the subsequent billing period, on that in billing period 1 is estimated by applying a (sharp) RD design. That is, in the RD design, the running variable corresponds to the consumption level in a billing period (i.e., in period 0) and the outcome variable to the daily average consumption in the subsequent billing period (i.e., in period 1).

1.1 Treatment Status Determination Mechanism

The treatment status to which residential consumers are assigned, and hence the unit price they are charged in period 1, is determined by the consumption level in period 0. To be specific, while the treatment status of households whose consumption in period 0 is less than or equal to the monthly base usage quantity is *Control*, that of households whose consumption in period 0 is greater than the monthly base usage quantity is *Treatment*.

1.2 Assumptions

Followings are assumed in this preliminary analysis:

- Both groups of residential consumers (i.e., *Control* and *Treatment* groups) are expected to be very similar along observed and unobserved characteristics but experienced very different unit prices.[1]
- Households infer prices from their recent past bill statements.

1.3 Econometric Model

To implement the RD design, I exploit following regression model:

$$DAC_{i,1} \ = \ \beta_0 \ + \ \beta_1 Treatment_{i,0} \ + \ f(\overline{C}_{i,0}) \ + \ \mathbf{X'}\boldsymbol{\gamma} \ + \ \epsilon_{i,0} \tag{1}$$

where $DAC_{i,1}$ corresponds to daily average consumption in period 1 for household i; $\overline{C}_{i,0}$ corresponds to the running variable, household i's normalized consumption in period 0; $\mathbf{X}$ are covariates, including daily average heating degree days (HDDs) and daily average cooling degree days (CDDs)[2]; and $\epsilon_{i,0}$ is a stochastic error term.

The treatment variable is a binary indicator of whether household i's electricity consumption in period 0 was

[1]Under the assumption, all observable and unobservable variables should evolve smoothly around the threshold, and any jump in consumption in period 1 can be attributed to the discontinuous increase in the unit price.

[2]The daily average HDDs are obtained by dividing HDDs in a billing period by the length of that billing period. The daily average CDDs are computed in the same way.

greater than the base usage quantity. That is, it is determined as:

$$Treatment_{i,0} = \begin{cases} 0 & \text{if } \overline{C}_{i,0} \leq 0 \\ 1 & \text{if } \overline{C}_{i,0} > 0 \end{cases} \tag{2}$$

The parameter β_1 in (1) captures the average effect of barely surpassing the threshold, once I control for the running variable by using a flexible function (i.e., f). In this analysis, I simply use a linear term for normalized consumption in period 0 (i.e., $f(\overline{C}_{i,0}) = \overline{C}_{i,0}$).

2 Data

For this preliminary analysis, I utilize two data sets: 1) billing data of Sacramento Municipal Utility District's (SMUD's) residential consumers, and 2) local Climatological Data (LCD) of National Oceanic and Atmospheric Administration (NOAA).

The primary data set of the preliminary analysis consists of panel data of household-level monthly billing records from 2004 to 2013. Each monthly record includes a residential customer's account and premise IDs, rate schedule code, billing start and end dates, total consumption with consumption by each tier, and fixed and variable charges. Because the billing data does not include price and base usage quantity information, I collect historical price schedules and base usage quantities from documents presented by SMUD. After dropping several observations that could undermine the quality of data used in the analysis, the final data set includes 29,513,151 billing periods of 475,126 residential consumers.[3]

NOAA's LCD for Sacramento Metropolitan Airport during the period between 2004 to 2013, which includes daily HDDs and daily CDDs, is utilized to compute each billing period's accumulated HDDs and CDDs.[4]

[3]To be specific, I drop 1) observations whose length of billing period is either less than 27 or greater than 34, 2) observations with negative values for quantities or charges, 3) observations having overlapping billing periods within a pair of account and premise IDs, 4) observations for households that do not cross the threshold in their billing history (i.e., always-light-users and always-heavy-users), and 5) observations whose number of days from the previous billing period is greater than 14. In addition, I exclude, from the sample, households whose number of billing periods is less than 24.

[4]There are six missing observations in the LCD. I complete the missing observations by exploiting NOAA's Global Surface Summary of the Day (GSOD) data set.

Table 1: Regression Results: Using Daily Average Consumption as the Dependent Variable

	Dependent variable:														
	Daily Average Consumption in Period 1 (kWh/Day)														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Normalized Consumption	0.191^{***}	0.211^{***}	0.210^{***}	0.213^{***}	0.194^{***}	0.185^{***}	0.202^{***}	0.202^{***}	0.205^{***}	0.185^{***}	0.125^{***}	0.098^{***}	0.095^{***}	0.094^{***}	0.084^{***}
in Period 0 relative to	(0.001)	(0.0005)	(0.001)	(0.001)	(0.004)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.003)	(0.00004)	(0.0003)	(0.001)	(0.001)	(0.003)
Base Usage Qty (%)															
1[Treated]	0.947^{***}	-0.082^{***}	-0.075^{***}	-0.088^{***}	0.008	0.641^{***}	-0.093^{***}	-0.090^{***}	-0.103^{***}	-0.003	-0.492^{***}	-0.008	0.004	0.010	0.060^{***}
	(0.100)	(0.011)	(0.012)	(0.015)	(0.021)	(0.097)	(0.010)	(0.012)	(0.014)	(0.020)	(0.004)	(0.008)	(0.009)	(0.011)	(0.016)
Daily Average CDDs						0.897^{***}	0.771^{***}	0.770^{***}	0.768^{***}	0.766^{***}	1.114^{***}	1.084^{***}	1.091^{***}	1.094^{***}	1.087^{***}
						(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)
Daily Average HDDs						0.390^{***}	0.304^{***}	0.302^{***}	0.300^{***}	0.299^{***}	0.462^{***}	0.418^{***}	0.419^{***}	0.419^{***}	0.413^{***}
						(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0003)	(0.0004)	(0.0005)	(0.001)	(0.001)
Bandwidth	N/A	20%	15%	10%	5%	N/A	20%	15%	10%	5%	N/A	20%	15%	10%	5%
FEs: Account-Premise IDs	No	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	29,513,151	9,350,655	7,086,939	4,744,344	2,410,342	29,513,151	9,350,655	7,086,939	4,744,344	2,410,342	29,513,151	9,350,655	7,086,939	4,744,344	2,410,342
R^2	0.605	0.079	0.046	0.021	0.005	0.637	0.157	0.127	0.103	0.088	0.767	0.556	0.553	0.563	0.602
Adjusted R^2	0.605	0.079	0.046	0.021	0.005	0.637	0.157	0.127	0.103	0.088	0.764	0.533	0.522	0.517	0.520

Note: The "N/A" bandwidth means that all observations are used for estimating coefficients, $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

3 Results

Table 1 shows regression results of the estimating model (1) for different bandwidths. Except the first five columns, daily average CDDs and HDDs are added as controls. And for the last five columns, account-id-by-premise-id fixed effect are added to eliminate household-specific fixed effects. That is, bill-to-bill within-household variation is exploited as the source of identifying variation in column (11)–(15).

For the positive bandwidths (i.e., 5%, 10%, 15%, and 20%), the estimated effect of *Treatment*, which lies in the range of -0.103 to 0.060, implies a discontinuous change in relative daily average consumption. The estimated coefficients on the treatment variable have a negligible size, even they are statistically significant at the 0.01 level.

Table 1 also demonstrates several important points:

- From column (1), (6), and (11), it is inferred that there exists large across-households variation in electricity consumption.
- The last five columns seem to imply that residential consumers respond only to sizable bill shocks.

A Appendix

A.1 Robustness of Regression Results

I utilize, instead of daily average consumption in period 1, monthly consumption in period 1 as the dependent variable to check my estimation model's robustness. The estimates obtained by using the dependent variable is presented in Table 2.

A.2 Regression Results with Additional Bandwidths

Table 3 demonstrates regression results of model (1) with controls and fixed effects for additional bandwidths. And Table 4 shows, for each bandwidth, distribution of observations between control and treatment groups.

Table 2: Regression Results: Using Monthly Consumption as the Dependent Variable

	Dependent variable:														
	Consumption in Period 1 (kWh)														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Normalized Consumption	5.576^{***}	6.189^{***}	6.173^{***}	6.258^{***}	5.674^{***}	5.402^{***}	5.895^{***}	5.890^{***}	5.977^{***}	5.352^{***}	3.572^{***}	2.683^{***}	2.599^{***}	2.546^{***}	2.213^{***}
in Period 0 relative to	(0.034)	(0.014)	(0.022)	(0.040)	(0.111)	(0.033)	(0.013)	(0.021)	(0.038)	(0.104)	(0.001)	(0.011)	(0.016)	(0.030)	(0.084)
Base Usage Qty (%)															
1[Treated]	29.476^{***}	-2.468^{***}	-2.249^{***}	-2.677^{***}	0.307	19.117^{***}	-2.818^{***}	-2.719^{***}	-3.154^{***}	0.010	-15.250^{***}	-0.253	0.128	0.314	1.964^{***}
	(2.928)	(0.328)	(0.378)	(0.463)	(0.649)	(2.828)	(0.309)	(0.356)	(0.436)	(0.612)	(0.134)	(0.237)	(0.273)	(0.338)	(0.492)
CDDs						1.041^{***}	0.930^{***}	0.929^{***}	0.928^{***}	0.927^{***}	1.248^{***}	1.231^{***}	1.239^{***}	1.242^{***}	1.237^{***}
						(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)
HDDs						0.520^{***}	0.439^{***}	0.438^{***}	0.437^{***}	0.436^{***}	0.582^{***}	0.541^{***}	0.543^{***}	0.544^{***}	0.538^{***}
						(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0003)	(0.0004)	(0.0005)	(0.001)	(0.001)
Bandwidth	N/A	20%	15%	10%	5%	N/A	20%	15%	10%	5%	N/A	20%	15%	10%	5%
FEs: Account-Premise IDs	No	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	29,513,151	9,350,655	7,086,939	4,744,344	2,410,342	29,513,151	9,350,655	7,086,939	4,744,344	2,410,342	29,513,151	9,350,655	7,086,939	4,744,344	2,410,342
R^2	0.579	0.071	0.041	0.019	0.004	0.624	0.178	0.151	0.130	0.117	0.756	0.561	0.560	0.570	0.608
Adjusted R^2	0.579	0.071	0.041	0.019	0.004	0.624	0.178	0.151	0.130	0.117	0.752	0.538	0.529	0.525	0.528

Note: The “N/A” bandwidth means that all observations are used for estimating coefficients, *p<0.1; **p<0.05; ***p<0.01

Table 3: Regression Results: Using Daily Average Consumption as the Dependent Variable with Additional Bandwidths

	Dependent variable:									
	Daily Average Consumption in Period 1 (kWh/Day)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Normalized Consumption	0.094^{***}	0.098^{***}	0.102^{***}	0.106^{***}	0.110^{***}	0.113^{***}	0.116^{***}	0.119^{***}	0.122^{***}	0.125^{***}
in Period 0 relative to	(0.001)	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00004)
Base Usage Qty (%)										
1[Treated]	0.010	−0.008	-0.023^{***}	-0.033^{***}	-0.062^{***}	-0.115^{***}	-0.186^{***}	-0.274^{***}	-0.374^{***}	-0.492^{***}
	(0.011)	(0.008)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Daily Average CDDs	1.094^{***}	1.084^{***}	1.065^{***}	1.047^{***}	1.031^{***}	1.022^{***}	1.019^{***}	1.022^{***}	1.027^{***}	1.114^{***}
	(0.001)	(0.001)	(0.001)	(0.0005)	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0005)
Daily Average HDDs	0.419^{***}	0.418^{***}	0.413^{***}	0.409^{***}	0.407^{***}	0.407^{***}	0.409^{***}	0.412^{***}	0.417^{***}	0.462^{***}
	(0.001)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Bandwidth	10%	20%	30%	40%	50%	60%	70%	80%	90%	N/A
FEs: Account-Premise IDs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,744,344	9,350,655	13,630,734	17,451,285	20,655,094	23,133,851	24,887,731	26,050,278	26,847,959	29,513,151
R^2	0.563	0.556	0.577	0.608	0.639	0.663	0.684	0.698	0.708	0.767
Adjusted R^2	0.517	0.533	0.562	0.597	0.631	0.656	0.678	0.692	0.703	0.764

Note: The "N/A" bandwidth means that all observations are used for estimating coefficients, $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 4: Distribution of Observations

Bandwidth	Control	Treatment	Total
10%	2,497,475	2,246,869	4,744,344
20%	5,061,435	4,289,220	9,350,655
30%	7,598,821	6,031,913	13,630,734
40%	9,946,284	7,505,001	17,451,285
50%	11,917,597	8,737,497	20,655,094
60%	13,375,725	9,758,126	23,133,851
70%	14,288,718	10,599,013	24,887,731
80%	14,763,871	11,286,407	26,050,278
90%	14,999,973	11,847,986	26,847,959
N/A	15,170,855	14,342,296	29,513,151

Notes: The "N/A" bandwidth means that all observations are used for estimating coefficients.