Jinmahn Jo **(ID#: 915528897)**

Contents

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 approach, the running variable corresponds to the consumption level in a billing period and the outcome variable to the (relative or absolute) daily average consumption in the subsequent billing period.

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 (relative or absolute) 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 relative heating degree days (HDDs) and relative cooling degree days (CDDs)[2]; and $\epsilon_{i,0}$ is a stochastic error term.

[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 relative HDDs imply HDDs in period 1 relative to those in period 0. The relative CDDs are defined in the same way.

The treatment variable is a binary indicator of whether household i's electricity consumption in period 0 was 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 10,355,022 billing periods of 582,557 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, and 3) observations having overlapping billing periods within a pair of account and premise IDs.

[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 Relative Daily Average Consumption

	Dependent variable:							
	Daily Average Consumption in Period 1 relative to that in Period 0 (%)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Normalized Consumption in Period 0 relative to Base Usage Qty (%)	-0.135^{***}	-0.134^{***}	-0.127^{***}	-0.162^{***}	-0.103^{***}	-0.099^{***}	-0.092^{***}	-0.096^{***}
	(0.001)	(0.002)	(0.004)	(0.011)	(0.002)	(0.003)	(0.006)	(0.017)
1[Treated]	-0.632^{***}	-0.637^{***}	-0.674^{***}	-0.494^{***}	-0.375^{***}	-0.419^{***}	-0.460^{***}	-0.408^{***}
	(0.033)	(0.039)	(0.047)	(0.066)	(0.051)	(0.058)	(0.072)	(0.101)
Relative CDDs					-0.059^{***}	-0.057^{***}	-0.054^{***}	-0.054^{***}
					(0.001)	(0.001)	(0.002)	(0.002)
Relative HDDs					-0.935^{***}	-0.937^{***}	-0.938^{***}	-0.945^{***}
					(0.004)	(0.005)	(0.006)	(0.008)
Constant	102.681^{***}	102.670^{***}	102.690^{***}	102.568^{***}	115.407^{***}	115.435^{***}	115.452^{***}	115.430^{***}
	(0.019)	(0.021)	(0.026)	(0.037)	(0.034)	(0.039)	(0.049)	(0.066)
Bandwidth	20%	15%	10%	5%	20%	15%	10%	5%
Observations	10,355,022	7,854,823	5,263,359	2,676,721	4,789,520	3,622,779	2,424,627	1,228,555
R^2	0.004	0.003	0.001	0.001	0.102	0.102	0.101	0.102
Adjusted R^2	0.004	0.003	0.001	0.001	0.102	0.102	0.101	0.102
Residual Std. Error	27.208	27.169	27.148	27.108	28.033	27.988	27.958	27.926

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2: Regression Results: Using Daily Average Consumption

	Dependent variable:							
	Daily Average Consumption in Period 1 (kWh/Day)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Normalized Consumption in Period 0 relative to Base Usage Qty (%)	0.275^{***}	0.302^{***}	0.336^{***}	0.399^{***}	0.289^{***}	0.314^{***}	0.337^{***}	0.367^{***}
	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
1[Treated]	-0.501^{***}	-0.285^{***}	0.026^{**}	-0.271^{***}	-0.426^{***}	-0.318^{***}	0.003	0.091^{***}
	(0.010)	(0.010)	(0.012)	(0.017)	(0.014)	(0.015)	(0.017)	(0.024)
Relative CDDs					-0.003^{***}	-0.002^{***}	-0.001^{***}	-0.002^{***}
					(0.0003)	(0.0004)	(0.0004)	(0.001)
Relative HDDs					-0.204^{***}	-0.204^{***}	-0.203^{***}	-0.203^{***}
					(0.001)	(0.001)	(0.001)	(0.001)
Constant	24.755^{***}	24.661^{***}	24.500^{***}	24.617^{***}	27.443^{***}	27.390^{***}	27.192^{***}	27.074^{***}
	(0.005)	(0.006)	(0.007)	(0.009)	(0.008)	(0.009)	(0.011)	(0.014)
Bandwidth	20%	15%	10%	5%	20%	15%	10%	5%
Observations	8,603,348	6,967,163	4,875,476	2,503,757	3,985,458	3,197,910	2,221,863	1,141,761
R^2	0.111	0.121	0.133	0.143	0.190	0.195	0.199	0.200
Adjusted R^2	0.111	0.121	0.133	0.143	0.190	0.195	0.199	0.200
Residual Std. Error	7.493	7.504	7.512	7.514	7.210	7.245	7.281	7.317

Note: *p<0.1; **p<0.05; ***p<0.01

3 Results

In this preliminary analysis, I utilize two different dependent variables: 1) daily average consumption in period 1 relative to that in period 0, which is measured by percentage, and 2) daily average consumption in period 1, which is measured in kWh per day.

Table 1 shows regression results that are obtained by using the first dependent variable (i.e., relative daily average consumption in period 1) for different bandwidths. In the last four columns, relative CDDs and HDDs are added as controls. The estimated effect of *Treatment* that lies in the range of -0.637 to -0.375 is a discontinuous decrease in relative daily average consumption. The estimated treatment effect, in all eight columns in Table 1, is statistically significant even at the 0.01 level, while the sizes of estimates are quite small.

For four different sizes of bandwidth, Table 2 presents regression estimates that are obtained by using the second dependent variable (i.e., absolute daily average consumption in period 1). As in Table 1, Column 5, 6, 7, and 8 add controls for relative CDDs and HDDs. Contrary to Table 1, Column 3, 7, and 8 in Table 2 demonstrate positive values for the estimates of the treatment effect.

One interesting point from the results in Table 1 and Table 2 is that estimated coefficients on relative CDDs and HDDs are negative, which is opposite to my expectation. One possible explanation could be that households tend to reduce their electricity consumption in order to offset increasing burden from natural gas consumption during winter seasons.

A Appendix

A.1 Robustness Check

Table 3: Regression Results: Using Relative Monthly Consumption

	Dependent variable:							
	Consumption in Period 1 relative to that in Period 0 (%)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Normalized Consumption in Period 0	-0.135^{***}	-0.134^{***}	-0.127^{***}	-0.162^{***}	-0.103^{***}	-0.099^{***}	-0.092^{***}	-0.096^{***}
relative to Base Usage Qty (%)	(0.001)	(0.002)	(0.004)	(0.011)	(0.002)	(0.003)	(0.006)	(0.017)
1[Treated]	-0.632^{***}	-0.637^{***}	-0.674^{***}	-0.494^{***}	-0.375^{***}	-0.419^{***}	-0.460^{***}	-0.408^{***}
	(0.033)	(0.039)	(0.047)	(0.066)	(0.051)	(0.058)	(0.072)	(0.101)
Relative CDDs					-0.059^{***}	-0.057^{***}	-0.054^{***}	-0.054^{***}
					(0.001)	(0.001)	(0.002)	(0.002)
Relative HDDs					-0.935^{***}	-0.937^{***}	-0.938^{***}	-0.945^{***}
					(0.004)	(0.005)	(0.006)	(0.008)
Constant	102.681^{***}	102.670^{***}	102.690^{***}	102.568^{***}	115.407^{***}	115.435^{***}	115.452^{***}	115.430^{***}
	(0.019)	(0.021)	(0.026)	(0.037)	(0.034)	(0.039)	(0.049)	(0.066)
Bandwidth	20%	15%	10%	5%	20%	15%	10%	5%
Observations	10,355,022	7,854,823	5,263,359	2,676,721	4,789,520	3,622,779	2,424,627	1,228,555
R^2	0.004	0.003	0.001	0.001	0.102	0.102	0.101	0.102
Adjusted R^2	0.004	0.003	0.001	0.001	0.102	0.102	0.101	0.102
Residual Std. Error	27.208	27.169	27.148	27.108	28.033	27.988	27.958	27.926

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

A.2 Plots

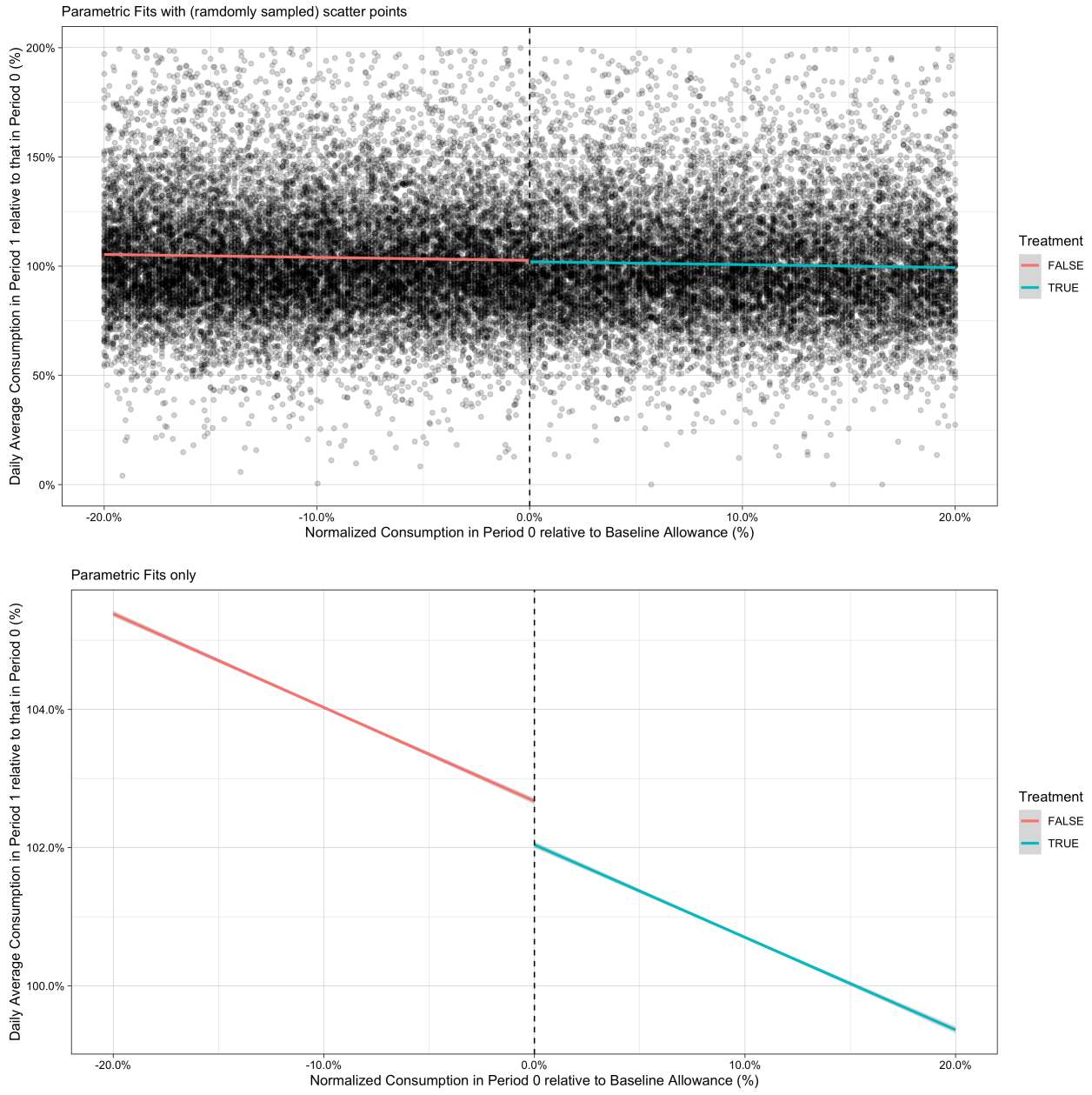

Figure 1: Changes in Electricity Consumption: Using Relative Daily Average Consumption

Notes: Parametric fits are generated from regressions that include a constant, a treatment dummy, and a linear term for normalized consumption in period 0. Only use observations whose daily average consumption in period 1 relative to that in period 0 is less than or equal to 200%.

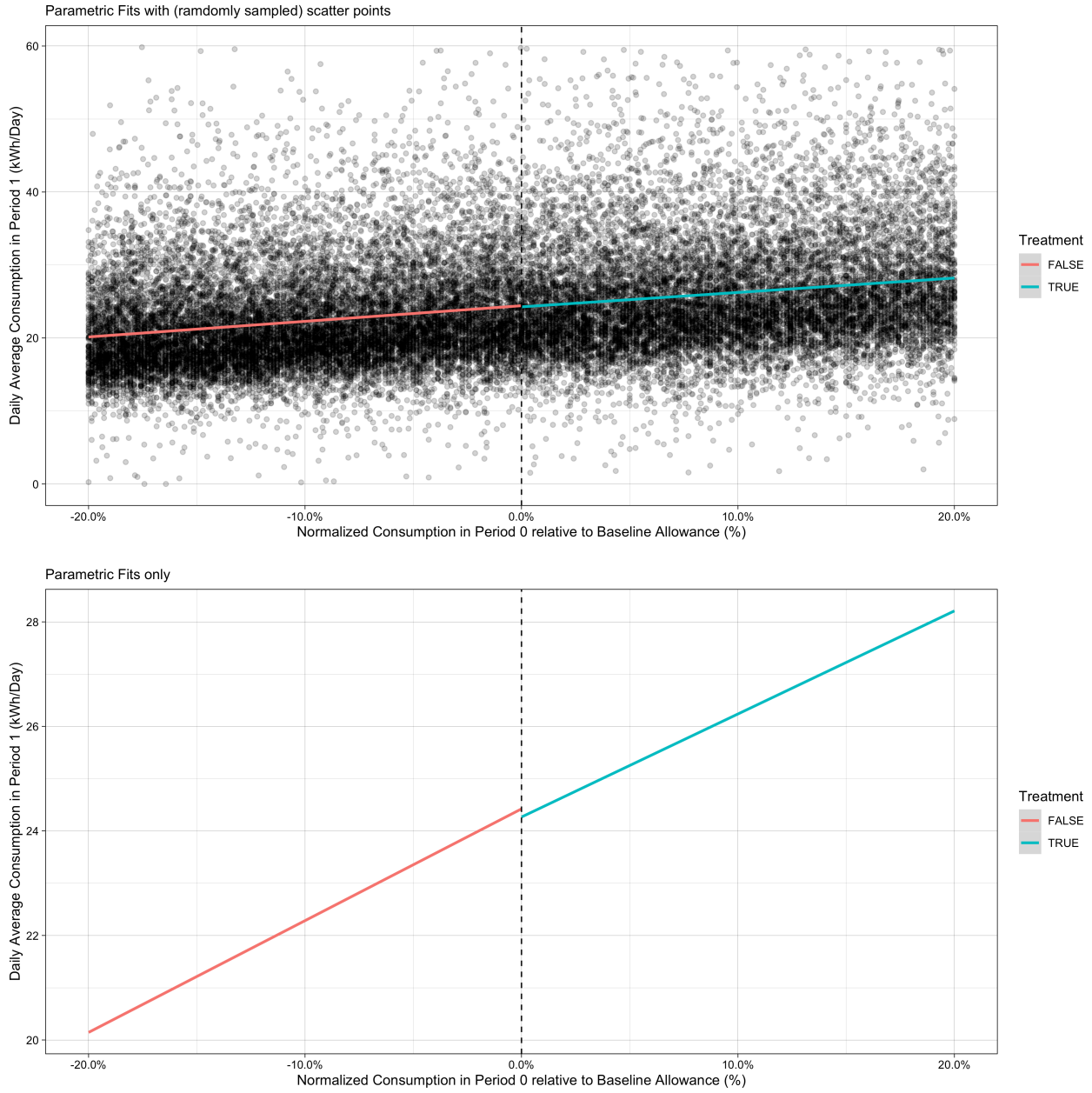

Figure 2: Changes in Electricity Consumption: Using Daily Average Consumption

Notes: Parametric fits are generated from regressions that include a constant, a treatment dummy, and a linear term for normalized consumption in period 0. Only use observations whose daily average consumption in period 1 is less than or equal to 60 kWh/Day.

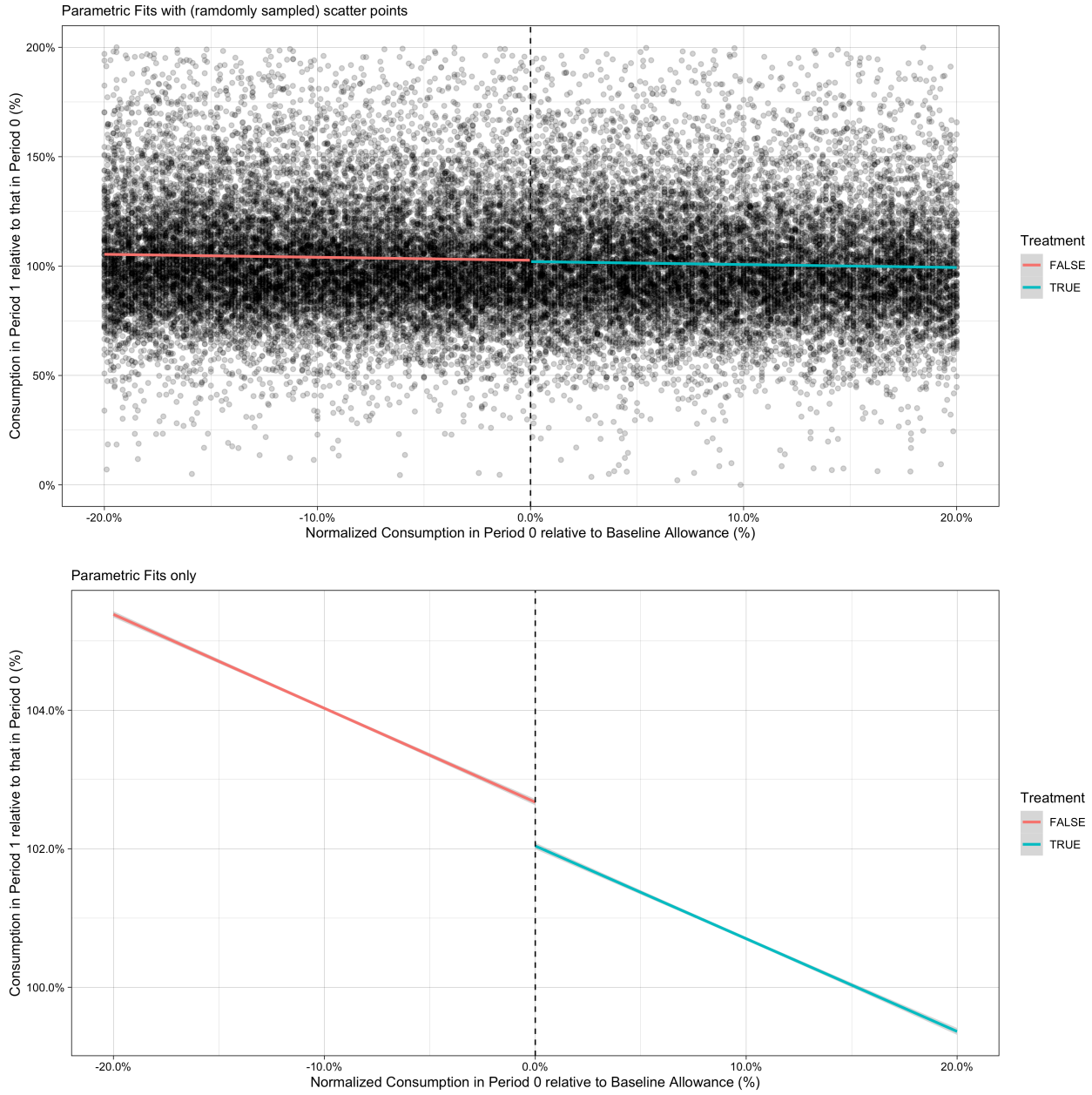

Figure 3: Changes in Electricity Consumption: Using Relative Monthly Consumption

Notes: Parametric fits are generated from regressions that include a constant, a treatment dummy, and a linear term for normalized consumption in period 0. Only use observations whose consumption in period 1 relative to that in period 0 is less than or equal to 200%.