

I. RCT/RED: post-treatment data only

You have been contacted by RCT Power and Light (RCTP&L) about helping them design a field experiment. Under a recent PUC ruling, RCTP&L is now required to reduce electricity usage in its service territory to 10 percent of forecast levels by 2025. If the utility can demonstrate that consumers reduce demand in response to intervention, they can count this program towards compliance with this regulation. RCTP&L are proposing to run a pilot program to help them estimate the average effect of this intervention on monthly household electricity consumption. RCTP&L plans to provide their customers with following:

- An in-home monitor that captures and displays real time energy consumption (in terms of dollars and kWhs),
- An easy-to-use phone app that helps customers make sense of their consumption patterns and develop a strategy for reducing their energy consumption.

Pat the project manager has come to you with a research design and is hoping you can help him figure out how many customers he will need in the study. The proposed research design proceeds as follows:

1. Randomly select N customers from the larger population of residential customers.
2. Randomly assign these customers to either a control group or the treatment group.
3. Contact treated households with the offer of free devices.
4. Collect monthly billing information for both the treated and control groups (i.e. N customers) over a two-year period following the introduction of the treatment in the treated group.
5. Compare consumption patterns in the treated and control group.

Pat is asking you how big the sample should be in order to detect a treatment effect of 5% with reasonable confidence. He thinks customers would be crazy not to jump at this offer, and instructs you to assume all “treated” households will receive the intervention. Assume you will have 24 months of post-treatment data to work with.

- (1) Use the sample of customer billing data provided (and described at the end of this assignment) to estimate how big the sample of customers would need to be to detect a treatment effect of 5% with reasonable confidence.
 - a Begin by clearly stating the estimating equation you have in mind.
 - b Given this equation, compute the required sample size using the analytic formulae provided in the lecture notes on power calculations.

Estimating Equation:

$$Y_{i,t} = \alpha + \tau D_{i,t} + \sum_{s \in \mathcal{T}} \gamma_s \delta_{s,t} + \varepsilon_{i,t} \quad (1)$$

where:

- $Y_{i,t}$ = consumer i 's electricity consumption at time t
- $\alpha = \mathbb{E}[Y_{i,t}(0)]$
- $\tau = \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0)]$

- $\delta_{s,t} = 1$ if $s = t$, 0 otherwise
- \mathcal{I} and \mathcal{T} are the index sets of clusters (households) and year-months, respectively.

If we had customer-level characteristics, we could include a $\mathbf{x}_i\beta$ term to control for random differences in consumer characteristics that might contribute to their electricity usage. We also are assuming here that consumers are i.i.d. and the consumption or treatment of one consumer is not effecting others. Because we have no information on location or other group memberships, we cannot explicitly cluster on a group-level identifier.

For the calculation, I assume:

- level of statistical power $\kappa = 0.8$
- significance level $\alpha = 5\%$
- treatment proportion $P = 0.5$

I begin by estimating $\sigma_\varepsilon^2, \sigma_\epsilon^2, \sigma_v^2$, and ρ using the data provided. From the data we have, no treatment has been applied. I will assume that $\hat{\sigma}_\varepsilon$ from a fixed effects regression

$$\begin{aligned}\hat{\sigma}_\varepsilon: \quad Y_{i,t} &= \alpha + \sum_{s \in \mathcal{T}} \gamma_s \delta_{s,t} + \varepsilon_{i,t} \\ \hat{\sigma}_\epsilon: \quad Y_{i,t} &= \alpha + \sum_{j \in \mathcal{I}} \alpha_j \delta_{j,i} + \sum_{s \in \mathcal{T}} \gamma_s \delta_{s,t} + \epsilon_{i,t} \\ \hat{\sigma}_v &= \hat{\sigma}_\varepsilon - \hat{\sigma}_\epsilon \\ \hat{\rho} &= \frac{\hat{\sigma}_v}{\hat{\sigma}_v + \hat{\sigma}_\varepsilon}\end{aligned}$$

will be a good estimate of $\text{Var}(\varepsilon_{i,t})$ from (1).

```
1 library(tidyverse)
2 data1 = read.csv('PS1_data.csv')
3 FE1 = lm(kwh ~ factor(yearmonthid), data=data1)
4 FE2 = lm(kwh ~ factor(custid) + factor(yearmonthid), data=data1)
5 sigma2_vare = sum(FE1$residuals^2) / (summary(FE1)$df[2] - 1)
6 sigma2_e = sum(FE2$residuals^2) / (summary(FE2)$df[2] - 1)
7 sigma2_v = sigma2_vare - sigma2_e
8 rho = sigma2_v / (sigma2_v + sigma2_e)
9
10 sigma2_vare
11 [1] 855540.9
12 sigma2_e
13 [1] 333429.6
14 sigma2_v
15 [1] 522111.3
16 rho
17 [1] 0.6102704
```

Then, plugging in these estimated values and our assumed values for α, κ, P , and MDE , we get an estimate for the number of households we need to detect a significant 5% treatment effect 80% of the time.

```
1 alpha = 0.05
2 kappa = 0.8
3 P = 0.5
4 MDE = 0.05*mean(data1$kwh)
5 T = length(unique(data1$yearmonthid))
6 t_a = qt(alpha/2, J, lower.tail=FALSE)
7 t_k = qt((1-kappa), J, lower.tail=FALSE)
8
9 J_hat = (t_a + t_k)^2*sigma2_e/(P*(1-P)*MDE^2)*(rho + (1-rho)/T)
10 J_hat
11 [1] 2781.984
```

- (2) Now repeat the power analysis you conducted in part (2) using the data provided and a simulation-based approach. In other words, simulate the estimation exercise using different sample sizes until you achieve the desired statistical power.
- Please explain the specific steps you take to conduct this analysis (and provide the associated code).
 - Do you reach the same conclusion (i.e. the same sample size) as in part 2? If no, offer some explanation as to why not.

Steps:

1. Load data
2. Define the treatment proportion $P = 0.5$ and the significance level $\alpha = 0.05$
3. Define the number of simulations S to run for each sample size.
4. Define range of sample sizes $Jlist$ to test (number of households to sample from our data).
5. Let $J \in Jlist$.
6. For each one of the S simulations for that J , pick (with replacement) J households from the data.
7. Randomly pick (without replacement) $J \cdot P$ households from the sample to treat. Mark them with a treatment indicator D and reduce their electricity consumption Y by 5% (the desired MDE).
8. Regress Y on D and year-by-month fixed effects.
9. Save the p-value of treatment effect coefficient.
10. After the S simulations are done for J , calculate the proportion of the S simulations that let to a **rejected** null hypothesis of no treatment effect (i.e., what is the proportion of S where the p-value was $< \alpha = 0.05$?).
11. Plot this proportion for all $J \in Jlist$ to determine the J where we can safely see a significant treatment effect in more than $\kappa = 0.8$ of the simulations.

To do this, I define a simulation function (see comments in green):

```

1 # Define a function to run simulations
2 simulate <- function(data1, Jlist, Simulations, take_up=1) {
3   # data1 = dataframe of representative observations to sample from
4   # Jlist = list of sample sizes J (number of households)
5   # Simulations = Number of simulations to run for each J
6   # take_up = fraction of households assigned treatment that take up treatment
7
8   # Start a dataframe to keep track of simulation outputs
9   df = data.frame()
10
11  # Iterate over a list of sample sizes
12  for (J in Jlist) {
13    df1 = data.frame()
14
15    # Run 'Simulations' simulations
16    for (s in 1:Simulations){

```

```

17 # Generate sample of J households with replacement
18 sample_hhs = sample(data1$custid, J, replace=TRUE)
19 sample = bind_rows(lapply(sample_hhs, function(i) data1[data1$custid %in% c(i
), ]))
20 # create new HH ids in case a HH is repeated
21 sample$id = rep(1:J, each=24)
22
23 # Randomly treat P proportion of the observations
24 Jtreat = round(P*J*take_up)
25 idx = sample(1:J, Jtreat, replace=FALSE)
26 sample$kwh_new = sample$kwh
27 sample[sample$id %in% idx,]$kwh_new = sample[sample$id %in% idx,]$kwh*(1 -
0.05)
28 sample$D = 0
29 sample[sample$id %in% idx,]$D = 1
30
31 # Regress
32 reg = summary(lm(kwh_new ~ D + factor(yearmonthid), data=sample))
33
34 # Save p-value of average treatment effect estimate
35 coef = coef(reg)['D', 'Estimate']
36 pval = coef(reg)['D', 'Pr(>|t|)']
37 df2 = data.frame(J=J, coef=coef, pval=pval, reject=(coef < 0 && pval <= 0.05)
)
38 df1 = rbind(df1, df2)
39
40 }
41 # Save portion of rejected nulls (p < alpha)
42 m = mean(df1$reject)
43 se = (m*(1-m)/Simulations)^0.5
44 lower = m - 1.96*se
45 upper = m + 1.96*se
46 df = rbind(df, data.frame(J=J, portion=m, lower=lower, upper=upper, simulations
=Simulations))
47 }
48
49 return(df)
50 }
51
52 plotsim <- function(Simulations, Jlist, df, J_hat) {
53   png(file=paste0('1-2_simulations_', Simulations, '_', min(Jlist), '-to-', max(
Jlist), '.png'),
54       width=800, height=400)
55   xlower = min(Jlist)
56   xupper = max(Jlist)
57   ylower = min(df$lower)
58   yupper = max(df$upper)
59   x1 = xlower + 0.1*(xupper - xlower)
60   y1 = 0.8 + 0.05*(yupper-ylower)
61   x2 = J_hat - 0.12*(xupper - xlower)
62   y2 = ylower + 0.1*(yupper-ylower)
63   a_size = 6
64   ggplot(df, aes(x=J, y=portion)) +
65     labs(title=paste('Proportion of', Simulations, 'simulations rejecting the
null when sampling from J households'),

```

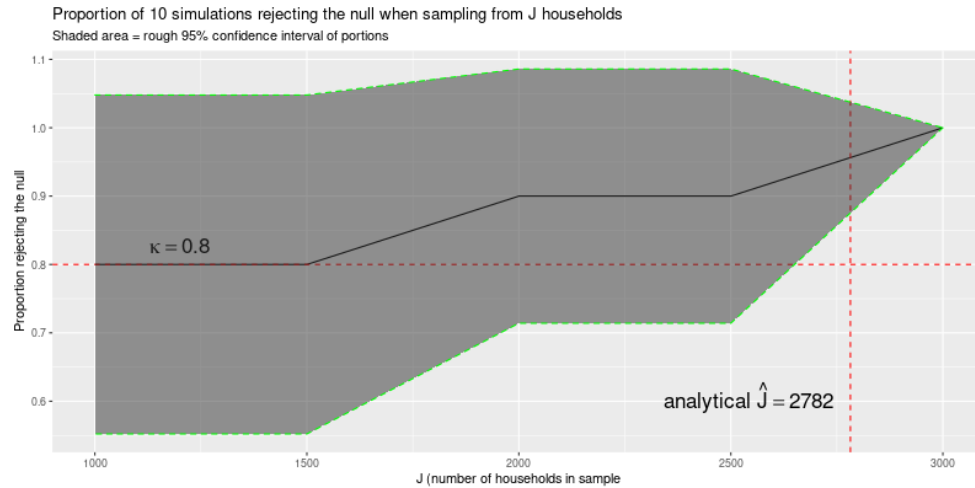
```

66     subtitle="Shaded area = rough 95% confidence interval of portions")
67   +
68     geom_line() +
69     xlab('J (number of households in sample') +
70     ylab(paste('Proportion rejecting the null')) +
71     xlim(range(min(Jlist), max(Jlist))) +
72     geom_hline(yintercept=0.8, linetype="dashed", color = "red") +
73     annotate("text", x=x1, y=y1, label=TeX("$\\kappa = 0.8$"), size=a_size) +
74     geom_vline(xintercept=J_hat, linetype="dashed", color = "red") +
75     annotate("text", x=x2, y=y2, label=TeX(sprintf('analytical $\\hat{J} = %d$'), round(J_hat))), size=a_size) +
76     geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.5, linetype="dashed", color="green")
77   dev.off()
78 }
79 Jlist0 = seq(from=1000, to=3000, by=500)
80 sim0 = simulate(data1, Jlist0, Simulations=10)
81 plotsim(S0, Jlist0, sim0, J_hat)
82
83 Jlist1 = seq(from=100, to=3500, by=100)
84 sim1 = simulate(data1, Jlist1, Simulations=100)
85 plotsim(S1, Jlist1, sim1, J_hat)
86
87 Jlist2 = seq(from=1000, to=3000, by=100)
88 sim2 = simulate(data1, Jlist2, Simulations=1000)
89 plotsim(S2, Jlist2, sim2, J_hat)
90
91 Jlist3 = seq(from=2700, to=2800, by=10)
92 sim3 = simulate(data1, Jlist3, Simulations=10000)
93 # This was taking too long (~ 3 hours per J)

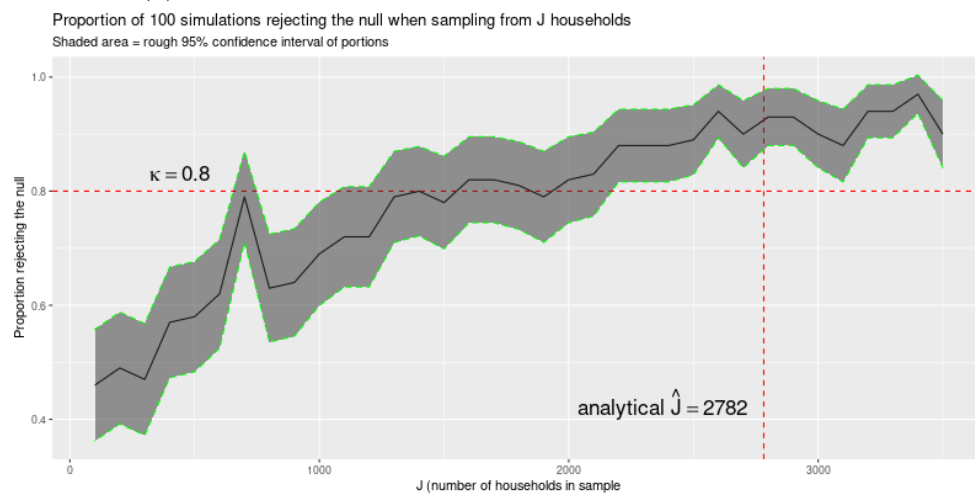
```

Plotting these simulations (Fig. 1), we can see that the safe sample size J to ensure a significant treatment effect is in the same neighborhood as our analytical version (somewhere between 2000 and 3000 households). Using a larger number of simulations per sample size would result in a smoother curve and we can see that more simulations lead to more monotonically increasing curves. The safe threshold sample size from the simulation (about 1800-2500) is a bit less than the analytical sample size (2782), but it is reassuring that it's nearby.

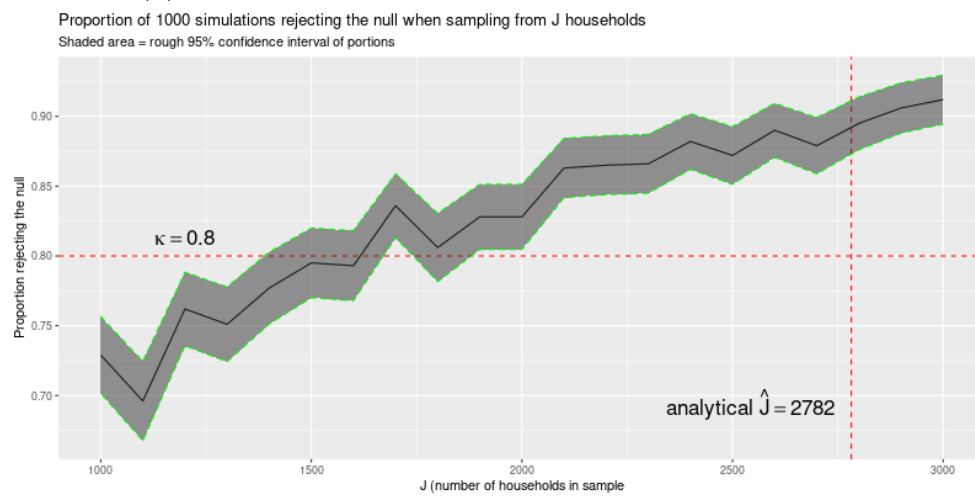
For a real experiment, I would want to run more simulations. If we perfectly specified the analytical version (using the more complicated panel data version), I would expect that the simulated answer would converge to the analytical version as the number of simulations used increased.



(a) Portion of 10 simulations over 1000-3000 households.



(b) Portion of 100 simulations over 100-3500 households.



(c) Portion of 1000 simulations over 1000-3000 households.

Figure 1: Portion of simulations resulting in a rejected null hypothesis of no treatment effect at a 95% confidence interval (estimated 95% confidence intervals of the estimated proportion is shaded).

- (3) Pat takes the power analysis you did in parts 1 and 2 to the RCTP&L board. They are rightly concerned about Pat's assumption of 100% take up. They point to other exciting residential programs that have take-up rates around 10%. You explain that you can redo the power calculations to reflect a 10% take up rate. You also suggest an "opt-out" design where customers receive these devices in the mail by default. This could help engage a larger share of customers.

Using the RED framework we discussed in class, modify the power analysis conducted above (you can choose to modify either the analytical or simulation-based approach) to account for the opt-in and opt-out alternatives, respectively. Assume an opt-in rate of 10%. Assume an opt-out rate of 50%. By how much does the required sample size increase under these two designs?

To adjust the analytical version, we need only multiply by a factor of $\frac{1}{(c-s)^2}$, where c is the take-up rate of those encouraged (10% and 50%), and s is the share of controls that take up treatment (0% here because they will not be able to access the devices if they are not sent them). So we would expect the analytical J to increase by a factor of $\frac{1}{0.1^2} = 100$ and $\frac{1}{0.5^2} = 4$ for each of the designs. This is about 278,200 for the first design, and 11,128 for the second.

To adjust the simulation, we just need to specify the take up rate. This adjusts the proportion of the sample that is selected for treatment. Thus, a simulation with 100 simulations per sample size, and a range of sample sizes from 2000 to 6000 would look like:

```
1 # Opt-in design ==> take_up rate of 10%
2 Jlist4 = seq(from=1000, to=3000, by=100)
3 sim4 = simulate(data1, Jlist2, Simulations=1000, take_up=0.10)
4 plotsim(1000, Jlist2, sim4, J_hat)
5 # Opt-out design ==> take_up rate of 50%
6 sim5 = simulate(data1, Jlist4, Simulations=1000, take_up=0.50)
7 plotsim(1000, Jlist2, sim5, J_hat)
```

From the simulations below, we can see that the J for which the simulation fraction is safely over a $\kappa = 0.8$ is well below the analytical suggestion. For 10% take up, the simulations suggest somewhere between 10,000 and 20,000 households should be enough (as opposed to $\approx 278,200$ from the analytical adjustment). For the 50% take up, simulations suggest somewhere between 3000 and 4000, again for below the 11,128 suggested analytically. This suggests that either my simulations are buggy or that there is something unaccounted for in my analytical calculations.

In general, I would trust my simulations more because it's a simpler method to test: I artificially treat my sample and regress.

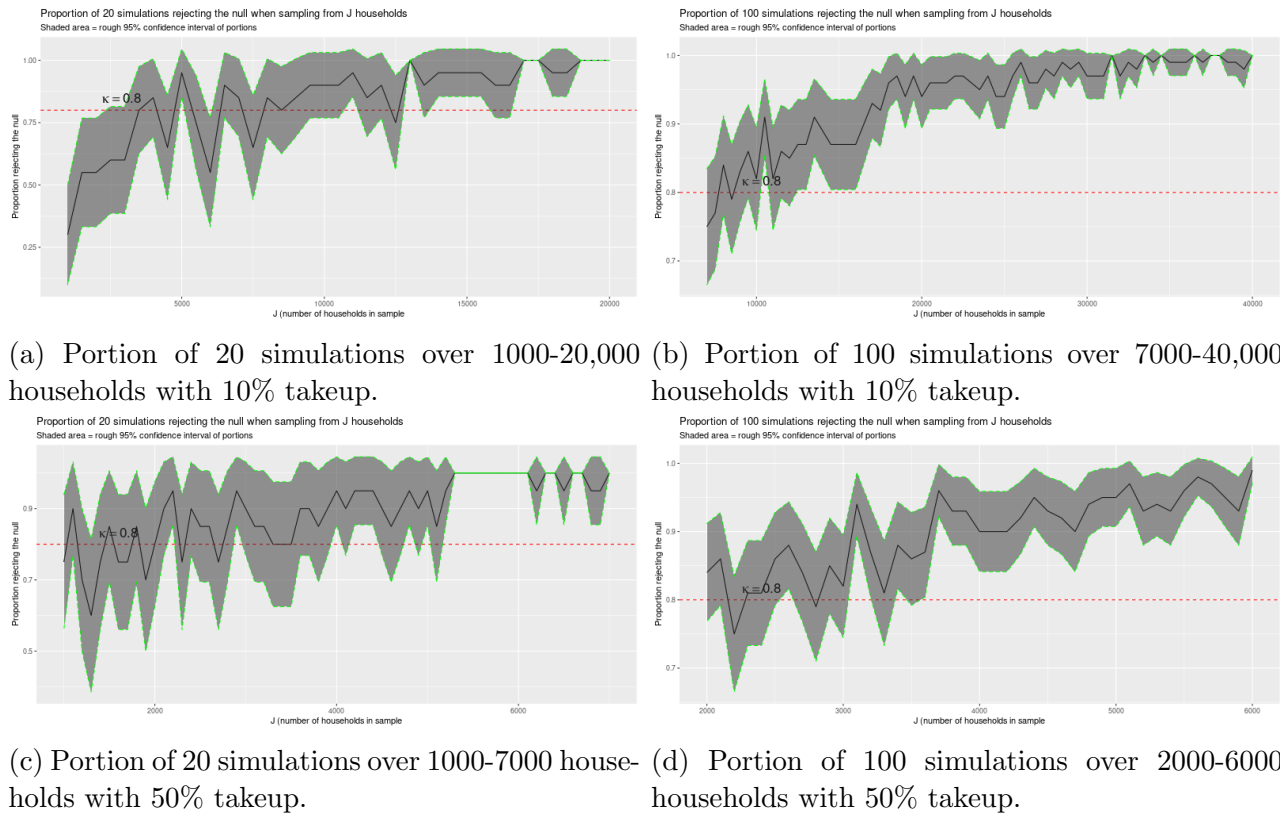


Figure 2: Take-up considered: Portion of simulations resulting in a rejected null hypothesis of no treatment effect at a 95% confidence interval with a 10% and 50% take-up percentages (estimated 95% confidence intervals of the estimated proportion is shaded).

- (4) You present your cost and power calculations in part 4 to the energy efficiency division at RCTP&L. They want to know if you will be able to test whether the response from consumers who choose to participate is different from the customers who were nudged/defaulted into the program. Can you separately identify these effects with your design?

We can identify these effects with the experimental setup (treatment encouragement design), but the statistical model I have proposed would not suffice to parse these effects. To assess these effects, we can run two different regressions: one on just the encouraged group (where treatment is actual treatment take-up) and one on the control group (where treatment is treatment take-up). This works in this instance because, presumably, the agency knows who has been actually treated (they can see who is using the in-home monitors).

- (5) Think about some of the assumptions and complications we discussed in class: the monotonicity assumption; SUTVA; the Hawthorne effect. Which (if any) of these might you be concerned about in this empirical setting?

Monotonicity: I can imagine some customers might resent being sent a device (in the opt-out design) and might turn on all their lights and leave some windows open just to spite the environmental justice warriors that are trying to force them to decrease their electricity usage. This violates monotonicity of treatment effect.

SUTVA: I'm not worried about SUTVA because I doubt the average electricity consumer is talking to their neighbors about an in-home monitor they receive from the electricity company.

Hawthorne effect: I think this actually might be the treatment effect here – we are interested in knowing how receiving a monitor might impact electricity usage. The Hawthorne effect is a change in behavior because the treatment unit is aware of being monitored. But in a treatment where monitoring is the treatment, the Hawthorne effect is the actual treatment effect (or at least the part of treatment effect where the consumer is aware that others are watching, vs the part where they are more aware of their usage because they are monitoring their usage, but these are two interconnected effects of the monitor, and I think we are interested in the combined effect).

II. Panel data!

Another utility contacts you about designing a small field experiment. They have equipped a non-random sample of 97 households with smart meters. These households eagerly opted into this experiment. They now want to equip these households with enabling devices that provide real-time feedback by displaying smart-meter data on an in-home device. The electric data will be aggregated to the daily level, and they will give you one year of pre-treatment data and one year of post-treatment data. They want to know the minimum effect that they could detect with this sample. Assume full compliance with treatment (i.e. everyone who is offered the free monitor will take it).

- (1) You have data aggregated to the daily and monthly level. Write down your estimating equations for both the daily and monthly data, respectively.

Estimating Equations:

$$\text{Daily: } Y_{i,y,d} = \beta_0 + \tau \mathbb{1}(y = 2014) + \beta_d + c_i + \varepsilon_{i,y,d}$$

$$\text{Monthly: } Y_{i,y,m} = \beta_0 + \tau \mathbb{1}(y = 2014) + \beta_m + c_i + \varepsilon_{i,y,m}$$

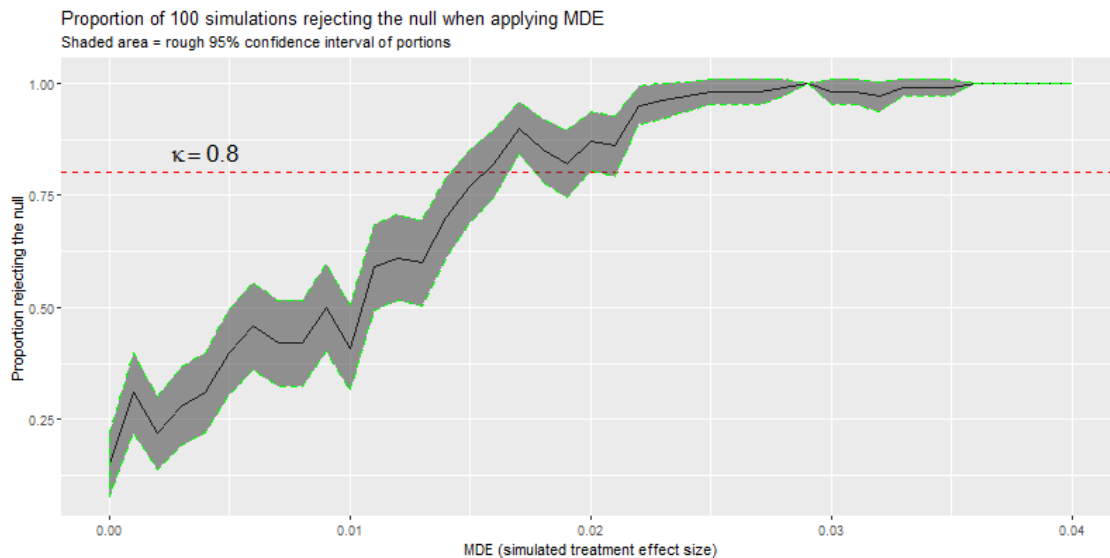
where:

- $Y_{i,t}$ = consumer i 's electricity consumption at time t
- τ = estimated treatment effect
- β_t = period-specific effect for period $t \in \{day, month\}$
- c_i = i 's household-specific effect

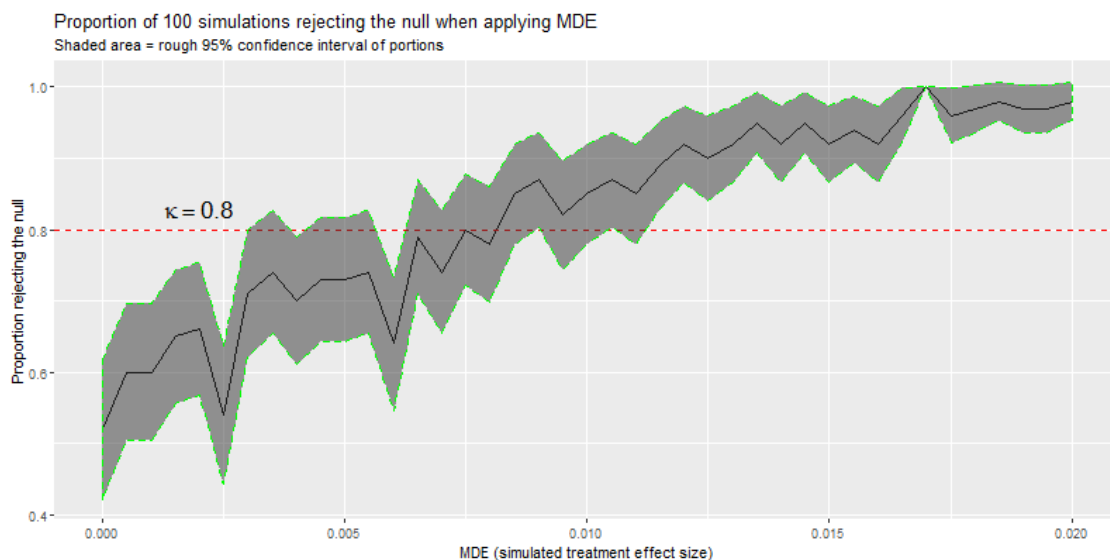
- (2) Using the sample of customer billing data provided (daily and monthly), estimate the minimum detectable effect on electricity consumption using a simulation-based power analysis. Assume a sample of 97 households with one year of pre-treatment data and one year of post-treatment data. Comment on the relative magnitude of the MDEs you estimate for the daily versus monthly data.

Modifying the previous simulation function to iterate over imposed treatment effects instead of sample size, we can see the simulation results below. While the monthly data results in an MDE of 2-2.5%, the daily data results in an MDE of 1-1.5%, about one-half the size.

I expected the relative size of the MDE to be much smaller for daily data relative to monthly (because we have roughly 30 times more data) but I am also using more time fixed effects (on the day-of-year level).



(a) Monthly data: Portion of 100 simulations over imposed, randomly assigned treatment effect from 0 to 4%.



(b) Daily data: Portion of 100 simulations over imposed, randomly assigned treatment effect from 0 to 2%.

Figure 3: Portion of simulations resulting in a rejected null hypothesis of no treatment effect at a 95% confidence interval, when imposing a treatment effect of various sizes (estimated 95% confidence intervals of the estimated proportion is shaded). MDE = treatment effect imposed on data.

```

1 ## QUESTION 2: PANEL DATA =====
2 library(lfe) # fixed effects
3 library(latex2exp) # greek text
4 library(tidyverse) # data manipulation (like rbind)
5 set.seed(194842)
6
7
8 # Define a function to run simulations

```

```

9 simulate2 <- function(data, MDElist, Simulations, P=0.5, take_up=1, sample_size=97,
10   day_or_month='day') {
11   # data = dataframe of representative observations to sample from
12   # MDElist = list of sample sizes J (number of households)
13   # Simulations = Number of simulations to run for each J
14   # take_up = fraction of households assigned treatment that take up treatment
15   # Start a dataframe to keep track of simulation outputs
16   df = data.frame()
17   print(paste('START:', Simulations, "simulations for MDE sizes", min(MDElist), 'to
18     ', max(MDElist)))
19   time_start = proc.time()
20
21   # Iterate over a list of sample sizes
22   for (MDE in MDElist) {
23     print(paste(format(Sys.time(), "%X"), ': Starting', Simulations, 'simulations
24       for MDE =', MDE))
25     df1 = data.frame()
26
27     # Run 'Simulations' simulations
28     for (s in 1:Simulations){
29       # Generate sample of J households with replacement
30       sample_hhs = sample(data$hhid, sample_size, replace=TRUE)
31       sample = bind_rows(lapply(sample_hhs, function(i) data[data$hhid %in% c(i),
32         ]))
33       # create new HH ids in case a HH is repeated
34       obs_per_hh = 24
35       sample$time = sample$month
36       if (day_or_month=='day') {
37         obs_per_hh = 2*365
38         sample$time = sample$day
39       }
40       sample$HHid = rep(1:sample_size, each=obs_per_hh)
41
42       # Treat second year of observations
43       sample[sample$year == 2014,]$kwh = sample[sample$year == 2014,]$kwh*(1 - MDE)
44       sample$treat = ifelse(sample$year == 2014, 1, 0)
45
46       # Regress
47       # reg = summary(lm(kwh ~ D + factor(t) + factor(hhid), data=sample))
48       reg = summary(felm(kwh ~ treat | time + HHid, data=sample))
49
50       # Save p-value of average treatment effect estimate
51       coef = coef(reg)['treat', 'Estimate']
52       pval = coef(reg)['treat', 'Pr(>|t|)']
53       df2 = data.frame(MDE=MDE, coef=coef, pval=pval, reject=(coef < 0 && pval <=
54         0.05))
55       df1 = rbind(df1, df2)
56     }
57
58     # Save portion of rejected nulls (p < alpha)
59     m = mean(df1$reject)
60     se = (m*(1-m)/Simulations)^0.5
61     lower = m - 1.96*se
62     upper = m + 1.96*se
63     print(paste0('Portion of simulations that lead to rejecting the null: ', m,

```

```

59     ' [, lower, ', ', upper, ' ]'))
60     df = rbind(df, data.frame(MDE=MDE, portion=m, lower=lower, upper=upper,
61                               simulations=Simulations))
62   }
63   print(paste('DONE: Simulation duration:', round((proc.time() - time_start)[3]/60,
64     2), 'min'))
65   return(df)
66 }
67 plotsim2 <- function(df, day_or_month) {
68   Simulations = df$simulations[1]
69   xlower = min(df$MDE)
70   xupper = max(df$MDE)
71   ylower = min(df$lower)
72   yupper = max(df$upper)
73   x1 = xlower + 0.1*(xupper - xlower)
74   y1 = 0.8 + 0.05*(yupper-ylower)
75   a_size = 6
76
77   png(file=paste0('2-2_', day_or_month, '_simulations_', Simulations, '_', xlower, '
78     -to-', xupper, '.png'), width=800, height=400)
79   print(ggplot(df, aes(x=MDE, y=portion)) +
80     labs(title=paste('Proportion of', Simulations, 'simulations rejecting the
81       null when applying MDE'),
82       subtitle="Shaded area = rough 95% confidence interval of portions")
83     +
84     geom_line() +
85     xlab('MDE (simulated treatment effect size)') +
86     ylab(paste('Proportion rejecting the null')) +
87     xlim(range(xlower, xupper)) +
88     geom_hline(yintercept=0.8, linetype="dashed", color = "red") +
89     annotate("text", x=x1, y=y1, label=TeX("$\\kappa = 0.8$"), size=a_size) +
90     geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.5, linetype="
91     dashed", color="green"))
92   dev.off()
93 }
94
95 data_monthly = read.csv('pecanstreet_monthly.csv')
96 data_daily = read.csv('pecanstreet_daily.csv')
97
98 # Monthly
99 MDElistmonth = seq(from=0, to=0.04, by=0.001)
100 mde_sim_month = simulate2(data_monthly, MDElistmonth, Simulations=100, day_or_
101   month='month')
102 plotsim2(mde_sim_month, 'month')
103
104 # Daily
105 MDElistday = seq(from=0, to=0.02, by=0.0005)
106 mde_sim_day = simulate2(data_daily, MDElistday, Simulations=100, day_or_month='day
107   ')
108 plotsim2(mde_sim_day, 'day')

```