Hydro Ottawa Rates

Brandon Edwards

10/12/2020

Introduction

The Ontario government recently made a decision about switching back to time-of-use hydro rates, rather than charging the same "off-peak" rate through the summer (because of COVID-19). Hydro Ottawa is giving customers the choice of continuing with the Time-of-Use (TOU) rates, or switching to the Tiered rates. The rates are as follows:

TOU

- Off-Peak: 10.5 cents per kilowatt hour
- Mid-Peak: 15.0 cents per kilowatt hour
- On-Peak: 21.7 cents per kilowatt hour

Tiered (Residential)

- First 1000 kWh: 12.6 cents per kilowatt hour
- Above 1000 kWh: 14.6 cents per kilowatt hour

To date, I have only received three hydro bills from Hydro Ottawa, because I have only lived in Ottawa since September. Each bill tells me the number of billing days, and the number of kilowatt hours used, separated by time-of-use by default. One bill was split into COVID-19 billing and non-COVID-19 billing, so it contains 2 sets of days and kilowatt hour usage. To determine whether it is more cost effective to continue with **TOU** or to switch to **Tiered** billing, I will use the limited information I have to simulate multiple bills, and determine the average cost per simulated bill.

Data

Billing seems to run roughly from the 19th of a month to the 20th of the next month, give or take a day on either side. I have received bills for the end of September, October, and November. The bill received at the end of November is split into two: one for billing using COVID-19 rates, and one for billing using the previous rates. For all bills, the usage is split up into the off-peak, mid-peak, and on-peak levels. Let's create the data:

```
days <- c(18, 31, 11, 19)

off_kwh <- c(20.709548, 43.417831, 11.000865, 23.538913)
mid_kwh <- c(11.989209, 19.607409, 5.741327, 12.542706)
on_kwh <-  c(9.223242, 20.196430, 7.036137, 12.703381)
```

Analysis

I could choose to make this a fairly complicated analysis, but because I'm pretty short on time right now and have work that needs my attention, I'll do a super simple analysis to start, and maybe one day I'll come back to this with more gusto (and more data).

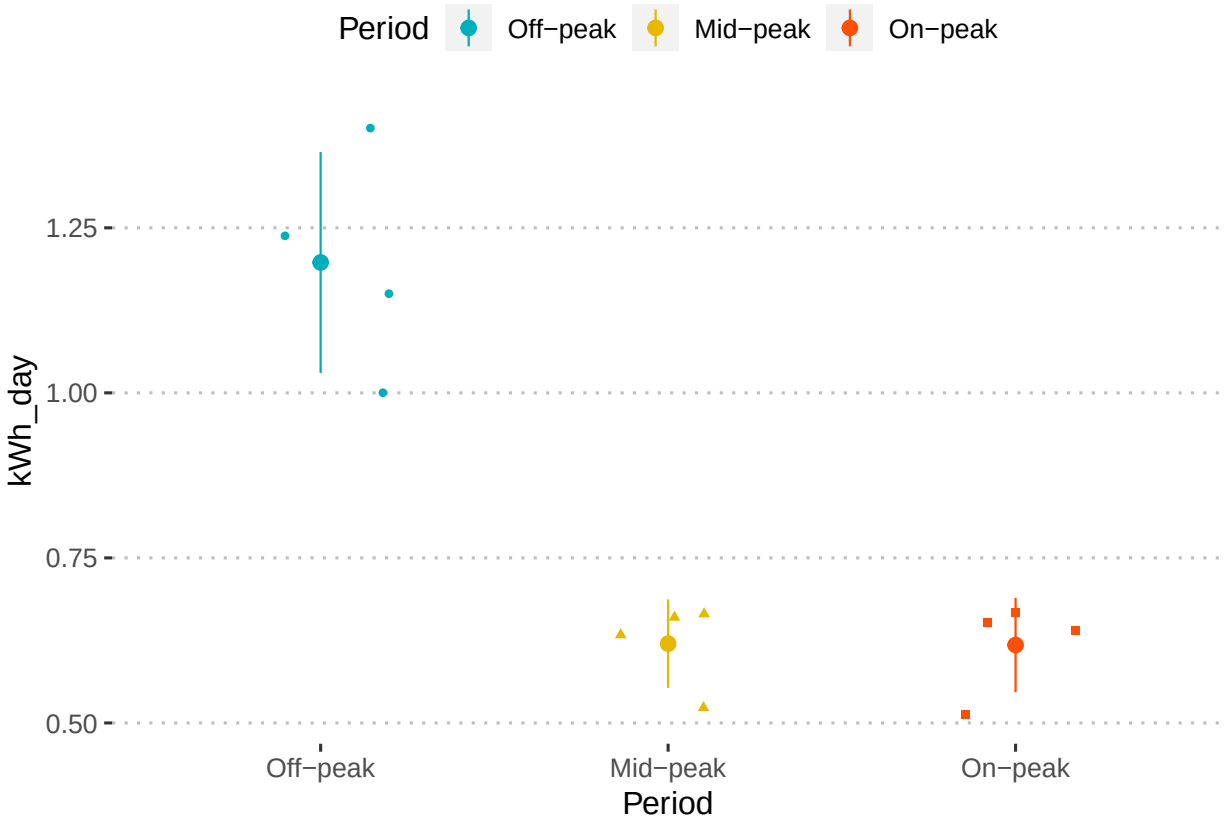
The kWh mentioned above are totals for the respective billing cycles. It might make more sense to standardize them by kWh / day. We'll do that, put that in a data frame, and plot the mean and standard deviations of

these values.

```
# Average kWh / Day
off_kwh_day <- off_kwh / days
mid_kwh_day <- mid_kwh / days
on_kwh_day <- on_kwh / days

# Put into data frame
period <- c(rep("Off-peak", length(off_kwh_day)),
            rep("Mid-peak", length(mid_kwh_day)),
            rep("On-peak", length(on_kwh_day)))
real_kwh_day <- data.frame(kWh_day = c(off_kwh_day, mid_kwh_day, on_kwh_day),
                           Period = factor(period,
                                           levels = c("Off-peak", "Mid-peak", "On-peak")))

# Plot it
p <- ggplot(real_kwh_day, aes(x = Period, y = kWh_day)) +
  geom_jitter(aes(shape = Period, color = Period),
              position = position_jitter(0.2),
              size = 1.2) +
  stat_summary(aes(color = Period),
               fun.data="mean_sdl",
               fun.args = list(mult=1),
               geom = "pointrange",
               size = 0.4) +
  scale_color_manual(values =  c("#00AFBB", "#E7B800", "#FC4E07"))
print(p)
```



Nice to see that most of my hydro use seems to be off-peak! That would mean that I would mostly be paying the cheapest rate if I stuck with the **TOU** billing.

I don't have very much data at the moment, so it's difficult to get an idea of distribution of these points. For now, I will just assume that the process that generates the average daily kWh usage is probably a normal distribution. Let's generate 1000 random samples for each of the off-peak, mid-peak, and on-peak daily kWh, using the mean and standard deviation of the data I currently have:

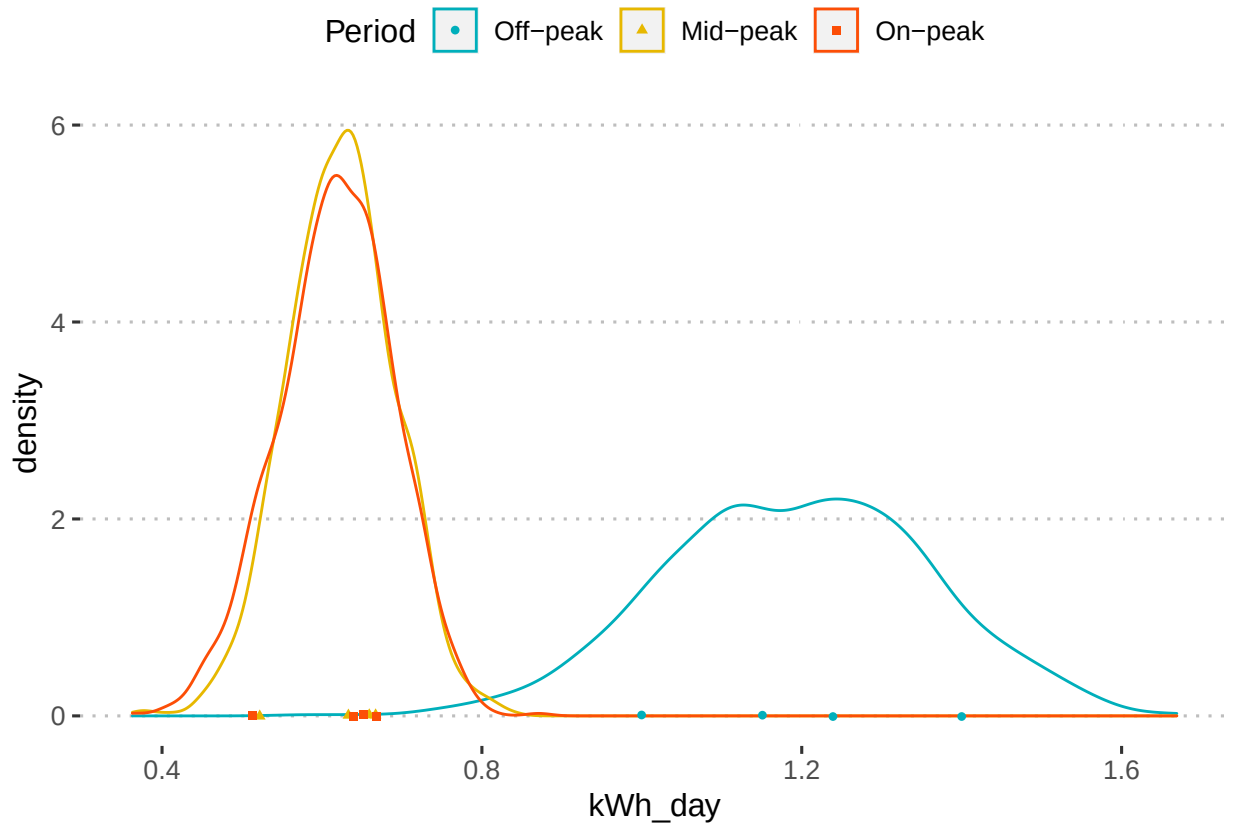
```
off_kwh_day_sample <- rnorm(n = 1000,
                            mean = mean(off_kwh_day),
                            sd = sd(off_kwh_day))
mid_kwh_day_sample <- rnorm(n = 1000,
                            mean = mean(mid_kwh_day),
                            sd = sd(mid_kwh_day))
on_kwh_day_sample <- rnorm(n = 1000,
                           mean = mean(on_kwh_day),
                           sd = sd(on_kwh_day))
```

Now, let's plot a density curve for each of these samples, and let's overlay the real data points on the plot as well:

```
# Put into data frame
period_sample <- c(rep("Off-peak", length(off_kwh_day_sample)),
                   rep("Mid-peak", length(mid_kwh_day_sample)),
                   rep("On-peak", length(on_kwh_day_sample)))
sim_kwh_day <- data.frame(kWh_day = c(off_kwh_day_sample,
                                      mid_kwh_day_sample,
                                      on_kwh_day_sample),
```

```
                          Period = factor(period_sample,
                                          levels = c("Off-peak", "Mid-peak", "On-peak")))

# Plot it
p <- ggplot(sim_kwh_day, aes(x = kWh_day, colour = Period)) +
  geom_density() +
    geom_jitter(data = real_kwh_day,
                aes(x = kWh_day, y = 0, shape = Period, color = Period),
                size = 1.2,
                height = 0.01) +
  scale_color_manual(values =  c("#00AFBB", "#E7B800", "#FC4E07"))
print(p)
```



Okay great! So it looks like the mid-peak and off-peak may have skewness in the data. It appears that the points seem to cluster to the left of their respective peaks. Again, I only have four data points, so hard to make a judgement call for distributional assumptions here. This is good enough for now!

Time for some calcuations. We can get a distribution of simulated costs with all the simulated data that I have created above. Since the samples are kWh / day, I will assume a 31 day month, and multiply each by their respective cost per time period. The cost in dollars will simply be the sum of these, divided by 100.

First, calculate the cost for **TOU** pricing:

```
off_tou <- off_kwh_day_sample * 31 * 10.5
mid_tou <- mid_kwh_day_sample * 31 * 15.0
on_tou <- on_kwh_day_sample * 31 * 21.7
```

```
tou_cost <- (off_tou + mid_tou + on_tou) / 100
```

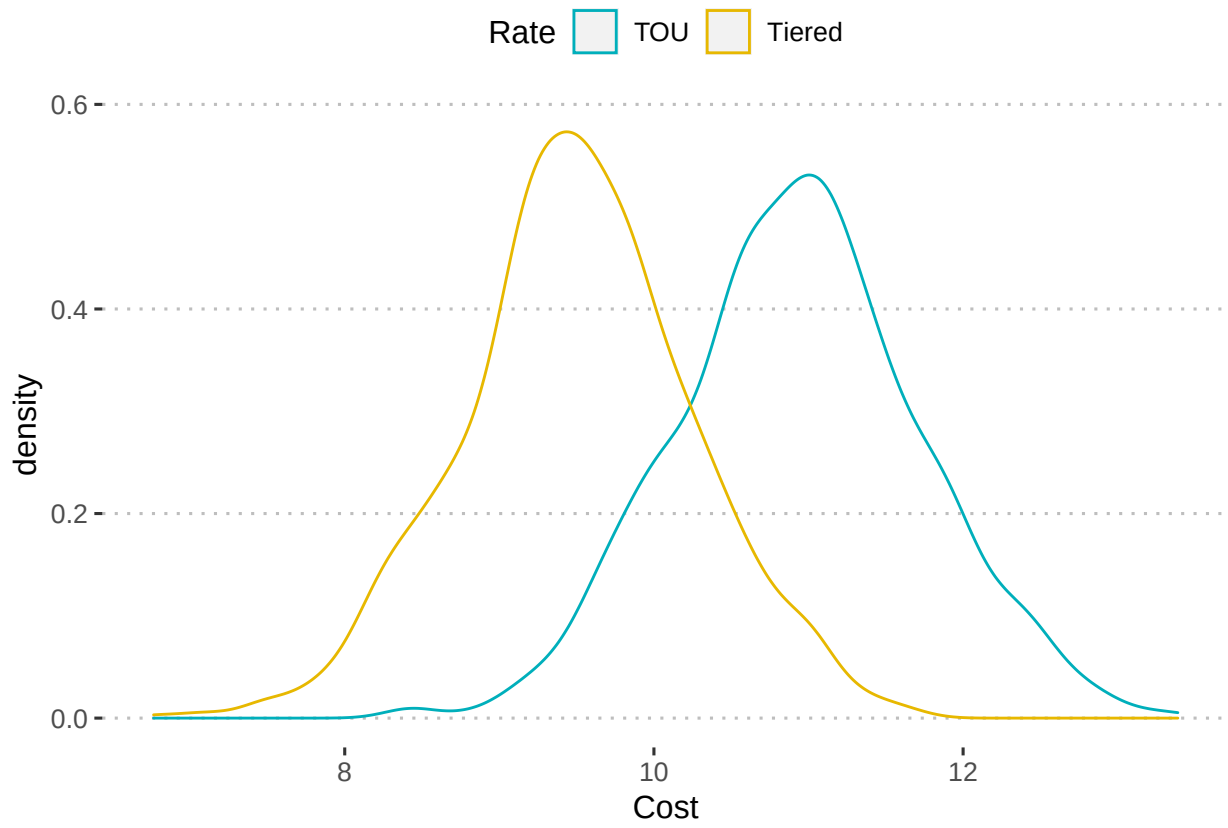
Now for the **Tiered** pricing:

```
sum_kWh <- off_kwh_day_sample + mid_kwh_day_sample + on_kwh_day_sample
tiered_cost <- ifelse(sum_kWh <= 1000,
                      (sum_kWh * 31 * 12.6) / 100,
                      (sum_kWh * 31 * 14.6) / 100)
```

Now for the moment of truth! Let's create a data frame of both of these simulated costs, and plot the densities to see what they look like:

```
rate <- c(rep("TOU", length(tou_cost)),
          rep("Tiered", length(tiered_cost)))
costs <- data.frame(Cost = c(tou_cost, tiered_cost),
                    Rate = factor(rate, levels = c("TOU", "Tiered")))

p <- ggplot(costs, aes(x = Cost, colour = Rate)) +
  geom_density() +
  scale_color_manual(values =  c("#00AFBB", "#E7B800"))
print(p)
```



That's certainly telling! Looks like the **Tiered** cost structure may be the way to go. Of course, we can do simple statistical tests, such as a t-test, to determine whether these distributions "significantly differ" from one another:

```
t.test(tou_cost, tiered_cost, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  tou_cost and tiered_cost
## t = 41.494, df = 1992.8, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.361527 1.496614
## sample estimates:
## mean of x mean of y
## 10.936244  9.507174
```

No surprise there, they do differ!

Discussion

Here, I presented the problem of whether I should make the switch to a **Tiered** rate for my hydro billing, or stick with the **Time-of-Use (TOU)** rate. **TOU** charges different depending on the time of day, and **Tiered** charges based on total usage.

Clearly, I have shown that I do not use very much hydro at all. I am a single person, with large windows for my apartment; the only lgiht I need is at night, when it's mostly off-peak. I do run a computer with a large monitor all day, so that likely accounts for most of my mid- and on-peak usage. Laundry is covered by the building, so I do not have to worry about when the dryer is run. As far as **TOU** goes, I clearly use most of my electricity in the off-peak timing, so I will mostly be charged the cheapest price. As far as **Tiered** goes, I will likely not even come close to hitting 1000 kWh in a month, so I will always be charged its cheapest price, which is roughly in between the off-peak and mid-peak pricing. This was the main motivation behind this investigation, as it seemed like it might be "too close to call".

Overall, it appears that I should switch to **Tiered** billing. However, there are a number of factors that I did not consider here, that should be considered in a future investigation. The main one that I can think of is a seasonal effect. In the winter, when it gets darker earlier, I need lights on for longer. Additionally, I am likely home more often in the winter, which means more usage in general. A future investigation might simulate daily rates based on a linear function that considers time of day.

I haven't looked into the terms of this switch yet, but I wonder how long they will be offering this switch. Can I switch back part way through the year, if its more beneficial for me to have one structure in the summer and one in the winter? All stuff to look into!