

Homework 8

Economics 7103

I have provided you access to a small subsample (<10%) of the electricity consumption data used in one of my papers in (*electric_matching.dta*). Each observation is for a zone/region i during hour t . The data span from 2014-2020. Table 1 describes the variables in the data.

Variable	Description
<i>datetime_beginning</i>	Time string
<i>time</i>	Stata time
<i>date</i>	Stata date
<i>month</i>	Month of year
<i>day</i>	Day of month
<i>year</i>	Year
<i>dow</i>	Day of week
<i>hour</i>	Hour of day
<i>mw</i>	Megawat hours consumed
<i>zone</i>	Zone/region
<i>temp</i>	Temperature (degrees F)
<i>pcp</i>	Precipitation

Table 1: Variable descriptions for homework 7.

You are going to study the effect of the pandemic on electricity consumption using a matching estimator.

1 Stata

1. Generate a log outcome variable and a binary treatment variable that is equal to one for all time periods March 1, 2020 and after.

- (a) Estimate the following regression:

$$\log(MW_{i,t}) = \alpha_{i,m,d,h} + \beta \text{treatment}_t + \gamma X_{i,t} + \varepsilon_{i,t} \quad (1)$$

where $\alpha_{i,m,d,h}$ includes indicator variables for zone, month of year, day of week, and hour of day and $X_{i,t}$ includes temperature and precipitation. Report the coefficient estimate and heteroskedasticity-robust standard error on treatment_t (not in a table).

- (b) Using the command `teffects nnmatch`, estimate the treatment effect on $\log(MW_{i,t})$ using the Mahalanobis distance norm and one nearest neighbor. Match on the continuous variables *temp* and *pcp*, and match exactly on zone, day of week, hour of day, and month of year. Report the coefficient estimate and heteroskedasticity-robust standard error (not in a table). *Hint*: You need to drop months 1 and 2 to get overlap for the matching estimate. The estimator may be slow, depending on your computer.
- (c) What issues do you think there are with these approaches?

2. Estimate the following regression:

$$\log(MW_{i,t}) = \alpha_{i,m,d,h,y} + \beta \text{treatment}_t + \gamma X_{i,t} + \varepsilon_{i,t} \quad (2)$$

where you have now added an indicator for year of sample.

- (a) Report the coefficient estimate and heteroskedasticity-robust standard error on $treatment_t$ (not in a table).
 - (b) How does this attempt to address the shortcoming in 1(c)?
3. Generate a new binary variable *year2020* equal to one during all of 2020. Using the command **teffects nnmatch**, match electricity consumption in 2020 to electricity consumption in 2019 using the most similar hour as defined by the Mahalanobis distance norm. Match on the continuous variables *temp* and *pcp*, and match exactly on zone, day of week, hour of day, and month of year. Generate a new variable *logmw_hat* equal to the matched electricity consumption. *Hint*: Use the **generate** option and in the case of ties, use the average matched consumption.
- (a) Denote $Y_{i,t} = \log(MW_{i,t})$ and $\hat{Y}_{i,t}$ the matched log electricity consumption (*logmw_hat*). Estimate the following regression on the 2020 data only:

$$Y_{i,t} - \hat{Y}_{i,t} = \alpha + \beta treatment_t + e_{i,t} \quad (3)$$

and report the estimate and heteroskedasticity robust standard error for the coefficient on $treatment_t$.

- (b) Why might you not trust the standard errors for the coefficient estimate from 3(a)?