Homework 7

Roshani Bulkunde

March 2023

1 STATA

1.1 Question 1

1.1.1 Question 1 (a)

VARIABLES	(1) lnmw
treatment	-0.0655***
	(0.00136)
Temperature (degrees F)	0.000240***
	(4.71e-05)
Precipitation	-0.0186
	(0.0120)
Constant	6.851***
	(0.00285)
Observations	122,718
R-squared	0.985

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Estimation results of equation 1: the coefficient estimate and heteroskedasticity-robust standard error

The coefficient on $treatment_t$ is -.065 and heteroskedasticity-robust standard error are .001359 with p-value 0 at 1% significance level. That means, electricity consumption decreased by 6.5% after period March 1, 2020.

1.1.2 Question 1 (b)

VARIABLES	(1) lnmw
r1vs0.treatment	-0.0703***
	(0.00100)
Observations	102,800

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Estimates of 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

Estimates of the treatment effect on $log(MW_{i,t})$ using the Mahalanobis distance norm and one nearest neighbor is -0.0703 with heteroskedasticity-robust standard error 0.001. That means, electricity consumption decreased by 7% after period March 1, 2020.

1.1.3 Question 1 (c)

In question 1(a), the control group, i.e., the observation before the period March 1, 2020 may not be the true counterfactual for the treatment group, i.e., the observation after the period March 1, 2020. Because there might be other time-variant factors that needs to be considered in order to avoid endogeinity problem.

In question 1(b), we are not excluding the exact matching on year. As per my intuition, that might be the problem in selecting the nearest neighbour to the observations in treatment group.

1.2 Question 2

1.2.1 Question 2 (a)

The coefficient on $treatment_t$ is .025 and heteroskedasticity-robust standard error are .0027.

1.2.2 Question 2 (b)

In equation 2, we are adding the indicator variables for zone, month of year, day of week, and hour of day, and an indicator for year of sample. That might solve the problem of selecting the nearest neighbour counterfactual observations to the observations in treatment group.

1.3 Question 3

1.3.1 Question 3 (a)

Table 3 use command teffects nnmatch and match electricity consumption in 2020 to electricity consumption in 2019 using the most similar hour as defined by the Mahalanobis distance norm. Table 3 estimates of the treatment effect on $log(MW_{i,t})$, the coefficient on $treatment_t$ is -0.0436 and heteroskedasticity-robust standard error are .00105.

Table 3 reports the coefficient estimate on $treatment_t$ and heteroskedasticity-robust standard error using 2020 data. The electricity consumption is decreased by -17818.29 magawat hours after the period March 1, 2020 as compared to the previous preiod March 1, 2020.

VARIABLES	(1) lnmw
r1vs0.y2020	-0.0436***
	(0.00105)
Observations	35,084

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Average treatment effect: the coefficient estimate and heteroskedasticity-robust standard error

VARIABLES	(1) diff_lgmw
treatment	-17,818***
	(76.46)
Constant	-2,610***
	(31.74)
Observations	17,568
R-squared	0.421

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Estimation of equation 3 using 2020 data: the coefficient estimate and heteroskedasticity-robust standard error

1.3.2 Question 3 (b)

The heteroskedasticity-robust standard errors on the coefficient are 76.46, which is very large. I might not trust the standard errors because the outcome variable was the difference between $Y_{i,t} = log(MW_{i,t})$ and $Y_{i,t}$ the matched log electricity consumption (logmw hat) and that makes these standard errors large.