Homework 2

Environmental Economics II
Maghfira Ramadhani

January 25, 2024

In this assignment, we are given imaginary data on an energy-efficiency retrofit program in Atlanta. The research hypothesis is whether the program reduced energy use. The experiment done in such way that after recruiting the households for the program, we assigned them to treatment and control groups. Treatment homes received the retrofits on the first of the month and control homes did not have any work done.

1 Python part

1.1 Balance table

To support the argument that the randomization worked, we can do a simple difference-in-means test between the control and treatment group. The word 'worked' means twofolds here. If the household characteristics in the control and treatment group looks similar, we may say that the treatment is independent conditional on the covariates of household characteristics, i.e. square feet of home and outdoor average temperature. And consequently, since the conditional independence assumption is justified, we can compare the difference in means of monthly electricity consumption between the treatment and control group and say something about the impact of the retrofit program.

Comparing the difference-in-means of the covariates in Table 1, there is no evidence of selection in the assignment. In arguing whether the program really affect household electricity consumption, there is strong evidence that it successfully reduces monthly consumption. Can I say that by satisfying the independence assumption then the difference-in-means estimates the average treatment effect? In this probably not, because square feet of home might and outdoor average temperature are likely to be correlated with electricity consumption which result in biased estimates if it is not included as control.

Table 1: Balance table from Python

	Control	Treatment	P-value
Electricity	1181.33	1086.75	0.001
	(454.31)	(423.96)	[3.403]
Square feet of home	1633.05	1657.55	0.572
	(682.90)	(686.27)	[-0.566]
Outdoor average temperature	79.89	79.89	0.987
	(2.16)	(1.97)	[-0.016]
Observations	501	499	

Notes Columns 1 and 2 report the mean and standard deviation in parentheses of the variables for the control group and treatment group consecutively. The third column show the p values for difference-in-means test and its corresponding t-statistics in square bracket.

1.2 Graphical evidence

Figure 1 shows the density plot for the control and treatment groups. The distribution of electricity consumption for home with retrofit is first order stochastically dominated, or in other words across the support

on average the average consumption for home with retrofit is less than those without.

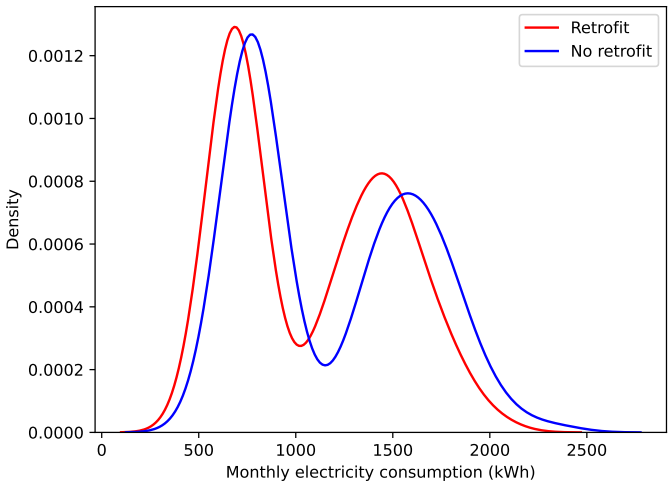

Figure 1: Density plot of electricity consumption

1.3 OLS regression

There are different computation methods to obtain the OLS estimates. I describe the method used to compute the OLS estimates inside the `OLS.py` file within the `manualOLS` class. The class takes three argument `useRobust, addIntercept,` and `method`. Options of `method` to use are `'byhand'` to use matrix inversion, `'stasmodels'` to used the python canned routing, and `'leastsquares'` to use the least square minimizer. The argument `useRobust` is by default set to `False` while `addIntercept` is by default set to `True`. User can use `useRobust=True` to get the heteroscedasticity robust OLS standard error estimator. The class return the coefficient estimates under the properties `.beta()`. The standard error of estimates can be obtained from the properties `.beta_std()`.

Table 2 and 3 shows the estimates from using different computation method which gives identical coefficient estimates.

Table 2: OLS estimates for different computation methods

	By Hand	Stats Model	Least Squares
=1 if retrofit	-109.666	-109.666	-109.666
	(7.948)	(7.948)	(7.948)
Square feet of home	0.615	0.615	0.615
	(0.006)	(0.006)	(0.006)
Outdoor average temperature	3.255	3.255	3.255
	(1.924)	(1.924)	(1.924)
Constant	-83.603	-83.603	-83.593
	(154.360)	(154.360)	(154.360)
MSE	125.652	125.652	125.652

The s.e. estimates is obtained using the square-root of the diagonal element of the following covariance matrix:

$$\hat{V}(\hat{\beta}) = \frac{1}{n-k}\sum_{j=1}^{k} \underbrace{\hat{e}_j^2}_{y_i - \mathbf{x}_i\hat{\beta}} (\mathbf{X'X})^{-1}$$

Table 3: OLS estimates with robust S.E. for different computation methods

	By Hand	Stats Model	Least Squares
=1 if retrofit	-109.666	-109.666	-109.666
	(7.943)	(7.943)	(7.943)
Square feet of home	0.615	0.615	0.615
	(0.007)	(0.007)	(0.007)
Outdoor average temperature	3.255	3.255	3.255
	(1.932)	(1.932)	(1.932)
Constant	-83.603	-83.603	-83.593
	(154.695)	(154.695)	(154.695)
MSE	125.652	125.652	125.652

The s.e. estimates is obtained using the square-root of the diagonal element of the following covariance matrix:

$$\hat{V}(\hat{\beta}) = (\mathbf{X'X})^{-1} \left(\frac{n}{n-k} \sum_{j=1}^{k} \hat{e}_j^2 \mathbf{x}_j \mathbf{x}_j \right) (\mathbf{X'X})^{-1}$$

2 Stata

2.1 Balance table

Table 4 is produced from Stata and show exactly the same result as in Table 1.

Table 4: Balance table from Stata

	Control	Treatment	P-value
Electricity	1181.33	1086.75	0.001
	(454.31)	(423.96)	[3.404]
Square feet of home	1633.05	1657.55	0.572
	(682.90)	(686.27)	[-0.566]
Outdoor average temperature	79.89	79.89	0.987
	(2.16)	(1.97)	[-0.016]
Observations	501	499	1,000

Notes Columns 1 and 2 report the mean and standard deviation in parentheses of the variables for the control group and treatment group consecutively. The third column show the p values for difference-in-means test and its corresponding t-statistics in square bracket.

2.2 Scatterplot

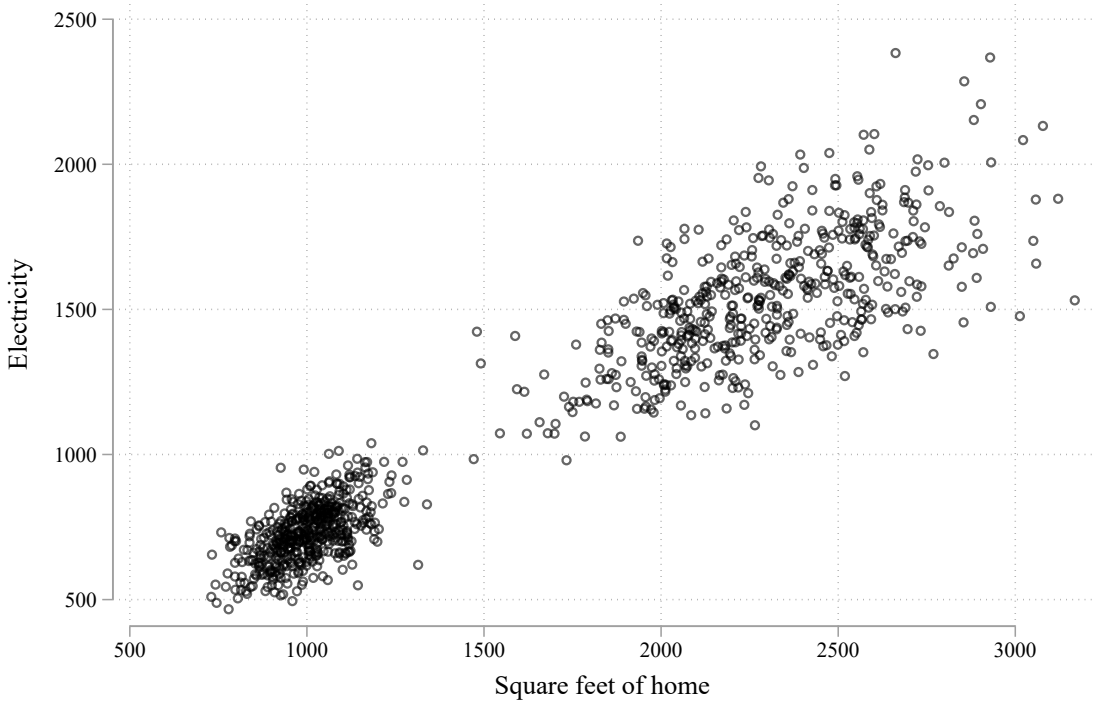

Figure 2: Scatterplot of electricity consumption and square feet of home

2.3 OLS regression

Table 5 is produced from Stata and shows the same result with the Python version.

Table 5: Estimates from Stata

	Non-robust S.E.	Robust S.E.
=1 if retrofit	-109.666	-109.666
	(7.948)	(7.943)
Square feet of home	0.615	0.615
	(0.006)	(0.007)
Outdoor average temperature	3.255	3.255
	(1.924)	(1.932)
Constant	-83.603	-83.603
	(154.360)	(154.695)
Observations	1,000	1,000

The non-robust and robust standard error estimates is exactly the same as in Table 2 and 3.