Homework 3

Environmental Economics II
Maghfira Ramadhani

January 27, 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. Suppose that for for a home i, the underlying relationship between electricity use and predictor variable is given by

$$y_i = e^{\alpha}\delta^{d_i} z_i^{\gamma} e^{\eta_i} \tag{1}$$

where y_i is the electricity use, e is Euler's number or the base of the natural logarithm, d_i is a binary variable equal to one if home i received the retrofit program, z_i is a vector of the other control variables, and η_i is unobserved error, and $\{\alpha, \delta, \gamma\}$ are the parameter to estimates.

   (a) Show that $\ln(y_i) = \alpha + \ln(\delta)d_i + \gamma \ln(z_i) + \eta_i$
   Answer
   Rewrite equation (1) by taking natural log on both sides and distribute the natural log on the right hand side of the equation, we have

$$\begin{aligned}
&\ln(y_i) = \ln\left(e^{\alpha}\delta^{d_i} z_i^{\gamma} e^{\eta_i}\right) \\
\Leftrightarrow\ &\ln(y_i) = \ln(e^{\alpha}) + \ln(\delta^{d_i}) + \ln(z_i^{\gamma}) + \ln(e^{\eta_i}) \\
\Leftrightarrow\ &\ln(y_i) = \alpha + \ln(\delta^{d_i}) + \gamma\ln(z_i) + \eta_i \\
\Leftrightarrow\ &\ln(y_i) = \alpha + \ln(\delta)d_i + \gamma\ln(z_i) + \eta_i. \qquad (2)
\end{aligned}$$

$\square$

   (b) What is the intuitive interpretation of δ?
   Answer
   Let's take the expectation of Equation (1) conditional on d_i, we have

$$\mathrm{E}[y_i|d_i = 1] = e^{\alpha}\delta z_i^{\gamma} e^{\eta_i} \tag{3}$$
$$\mathrm{E}[y_i|d_i = 0] = e^{\alpha} z_i^{\gamma} e^{\eta_i} \tag{4}$$

   Dividing the first equation by the second equation, we have

$$\delta = \frac{\mathrm{E}[y_i|d_i = 1]}{\mathrm{E}[y_i|d_i = 0]} \tag{5}$$

   which means δ translates to the ratio of the expected electricity use of the treatment group to the expected electricity use of the control group. The intuitive interpretation of δ related to the context of the program is as follows. If the program improve the energy efficiency of the house, then the electricity use of the treatment group will be lower than the control group, and δ will be less than one. If the program does not improve the energy efficiency of the house, then the electricity use of the treatment group will be higher than the control group, and δ will be greater than one. If the program does not have any effect on the energy efficiency of the house, then the electricity use of the treatment group will be the same as the control group, and δ will be equal to one.

(c) Show that $\frac{\Delta y_i}{\Delta d_i} = \frac{\delta - 1}{\delta^{d_i}} y_i$. What is the intuitive interpretation of $\frac{\Delta y_i}{\Delta d_i}$?

Answer

Note that since d_i is a binary variable, the average marginal effect of d_i is the difference in the expected value of y_i between the treatment and control group, i.e:

$$\frac{\Delta y_i}{\Delta d_i} = \mathrm{E}[y_i|d_i = 1] - \mathrm{E}[y_i|d_i = 0].$$

By using previous result in equation (5), we have

$$\begin{aligned} \mathrm{E}[y_i|d_i = 1] - \mathrm{E}[y_i|d_i = 0] &= \delta \mathrm{E}[y_i|d_i = 0] - \mathrm{E}[y_i|d_i = 0] \\ &= (\delta - 1)\mathrm{E}[y_i|d_i = 0] \end{aligned} \tag{6}$$

Since we only observed y_i from the data we need to know what exactly is $\mathrm{E}[y_i|d_i = 0]$. Since the relationship between y_i and d_i is non-linear, it might be useful to apply potential outcome framework with equation (2). Suppose y_{0i} is the potential outcome of y_i if $d_i = 0$ and y_{1i} is the potential outcome of y_{1i} if $d_i = 1$. From the underlying relationship, note that for every house i, we have $y_{1i}/y_{0i} = \delta$. Applying the potential outcome framework, we have

$$\begin{aligned} &\ln(y_i) = \ln(y_{0i}) + [\ln(y_{1i}) - \ln(y_{0i})]d_i \\ \Leftrightarrow &\ln\left(\frac{y_i}{y_{0i}}\right) = d_i \ln\left(\frac{y_{1i}}{y_{0i}}\right) = d_i \ln(\delta) \\ \Leftrightarrow &\frac{y_i}{y_{0i}} = e^{d_i \ln(\delta)} = \delta^{d_i} \\ \Leftrightarrow &y_{0i} = \frac{1}{\delta^{d_i}} y_i \end{aligned} \tag{7}$$

Substituting equation (7) to (6), we can obtain the average marginal effect of d_i as follows:

$$\frac{\Delta y_i}{\Delta d_i} = \mathrm{E}[y_i|d_i = 1] - \mathrm{E}[y_i|d_i = 0] = \frac{\delta - 1}{\delta^{d_i}} y_i.$$

$\square$

The intuitive interpretation of $\frac{\Delta y_i}{\Delta d_i}$ is the reduction in monthly electricity consumption in kWh if a house received a retrofit.

(d) Show that $\frac{\delta y_i}{\delta z_i} = \gamma \frac{y_i}{z_i}$. What is the intuitive interpretation of $\frac{\delta y_i}{\delta z_i}$ when z_i is the size of the home in square feet?

Answer

Since z_i is a vector of control variables, we can apply the chain rule to equation (2) with respect to z_i as follows:

$$\frac{\partial \ln(y_i)}{\partial z_i} = \frac{\partial \ln(y_i)}{\partial y_i}\frac{\partial y_i}{\partial z_i} = \frac{1}{y_i}\frac{\partial y_i}{\partial z_i} = \frac{\gamma}{z_i} \Rightarrow \frac{\partial y_i}{\partial z_i} = \gamma \frac{y_i}{z_i} \tag{8}$$

$\square$

If z_i is the size of the home in square feet, then the intuitive interpretation of $\frac{\delta y_i}{\delta z_i}$ is the change in monthly electricity consumption in kWh if the size of the home in square feet increase by one square feet. We can expect the sign of γ to be positive. The larger the size of the home, the more electricity it consumes.

(e) Estimate the log-transformed equation via OLS on the transformed parameters. Save the coefficient estimates and the average marginal effects estimates of z_i and d_i $\left(\frac{\delta y_i}{\delta z_i} \text{ and } \frac{\Delta y_i}{\Delta d_i}\right)$. Bootstrap the 95% confidence intervals of the coefficient estimates and the marginal effects estimates using 1000 sampling replications. Display the results in a table with three columns (one for the variable name, one for the coefficient estimates, and once for the marginal effect estimate). Show the 95% confidence intervals for each estimate under each number.

(f) Graph the **average marginal effects** of outdoor temperature and square feet of home with bands for their bootstrapped confidence intervals so that they are easy to interpret and compare.