Econometrics: Problem Set #2

Due on April 5, 2024 at 4:00pm

Professor Ben Faber Section 101

Zachary Brandt

Question 1

The Ministry of Commerce in a large country wants to know the causal effect of membership in a local Chamber of Commerce on firm revenues and profits. Firms pay for their membership and they supposedly benefit from the network of information and contacts that the local Chambers of Commerce offer them. But usually, it is only a small minority of all firms that end up paying for the membership.

The Ministry plans to estimate the causal effect of being a member in a local Chamber of Commerce by letting the Ministry's staff estimate the average percentage change in annual firm sales between firms that are members and the rest of the firms that are non-members of their local Chamber of Commerce.

A) Write down the OLS regression specification that the Ministry's staff could use to implement their analysis described above. Interpret what the intercept and slope coefficients would capture in such a specification.

B) Using notation from the Potential Outcomes Framework, briefly explain the concept of the Average Treatment Effect (ATE) to the Minister, and how what they plan to estimate in A) relates to this definition.

C) Referring to the expressions you use in your answer to B), explain why a randomized control trial (RCT) could be useful, and very briefly describe the basics of the RCT design for how the Ministry could set this up.

D) The Ministry mentions that it has no legal authority to force firms to become members in their local Chambers of Commerce. Using notation from the Potential Outcomes Framework, explain why this information could be important for the interpretation of the results from the RCT relative to the ATE, and how the Ministry should address this concern in the RCT analysis?

E) The Ministry talked to other economists, and now it is worried about spillover effects on the control group. The staff don't fully understand what the concern is, however. Briefly explain to them the intuition behind this concern, and explain how they could potentially address it when designing the RCT.

Question 1 continued on next page. . .

Part A

Write down the OLS regression specification that the Ministry's staff could use to implement their analysis described above. Interpret what the intercept and slope coefficients would capture in such a specification.

Solution

The following OLS regression specification is one that the Ministry could implement:

$$ln(Y_i) = \beta_0 + \beta_1 D_i + u_i$$

where

the subscript i runs over the observations, $i = 1, \ldots, n$;

Y_i is the *dependent variable*, annual firm sales

D_i is the *dummy variable*, $D_i = 1$ if the firm is a local Chamber of Commerce member and 0 otherwise

β_0 is the *intercept* of the regression, the population mean value of annual, non-member firm sales

β_1 is the coefficient on D_i, associating a change in D_i by one unit with a $100\beta_1\%$ change in Y_i

$\beta_0 + \beta_1$ is the population mean value of annual, member firm sales

u_i is the *error* term, all the factors responsiblze for the difference between predicted and observed values

Question 2

Suppose we would like to fit a straight line through the origin, i.e., $Y_i = \beta_1 x_i + e_i$ with $i = 1, \ldots, n$, $\mathrm{E}[e_i] = 0$, and $\mathrm{Var}[e_i] = \sigma_e^2$ and $\mathrm{Cov}[e_i, e_j] = 0, \forall i \neq j$.

Part A

Find the least squares esimator for $\hat{\beta}_1$ for the slope β_1.

Solution

To find the least squares estimator, we should minimize our Residual Sum of Squares, RSS:

$$\begin{aligned} RSS &= \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 \\ &= \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 x_i\right)^2 \end{aligned}$$

By taking the partial derivative in respect to $\hat{\beta}_1$, we get:

$$\frac{\partial}{\partial \hat{\beta}_1}(RSS) = -2 \sum_{i=1}^{n} x_i (Y_i - \hat{\beta}_1 x_i) = 0$$

This gives us:

$$\begin{aligned} \sum_{i=1}^{n} x_i (Y_i - \hat{\beta}_1 x_i) &= \sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} \hat{\beta}_1 x_i^2 \\ &= \sum_{i=1}^{n} x_i Y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \end{aligned}$$

Solving for $\hat{\beta}_1$ gives the final estimator for β_1:

$$\hat{\beta}_1 = \frac{\sum x_i Y_i}{\sum x_i^2}$$

Question 2 continued on next page. . .

Part B

Calculate the bias and the variance for the estimated slope $\hat{\beta}_1$.

Solution

For the bias, we need to calculate the expected value $\mathrm{E}[\hat{\beta}_1]$:

$$
\begin{aligned}
\mathrm{E}[\hat{\beta}_1] &= \mathrm{E}\left[\frac{\sum x_i Y_i}{\sum x_i^2}\right] \\
&= \frac{\sum x_i \mathrm{E}[Y_i]}{\sum x_i^2} \\
&= \frac{\sum x_i(\beta_1 x_i)}{\sum x_i^2} \\
&= \frac{\sum x_i^2 \beta_1}{\sum x_i^2} \\
&= \beta_1 \frac{\sum x_i^2 \beta_1}{\sum x_i^2} \\
&= \beta_1
\end{aligned}
$$

Thus since our estimator's expected value is β_1, we can conclude that the bias of our estimator is 0.

For the variance:

$$
\begin{aligned}
\mathrm{Var}[\hat{\beta}_1] &= \mathrm{Var}\left[\frac{\sum x_i Y_i}{\sum x_i^2}\right] \\
&= \frac{\sum x_i^2}{\sum x_i^2 \sum x_i^2} \mathrm{Var}[Y_i] \\
&= \frac{\sum x_i^2}{\sum x_i^2 \sum x_i^2} \mathrm{Var}[Y_i] \\
&= \frac{1}{\sum x_i^2} \mathrm{Var}[Y_i] \\
&= \frac{1}{\sum x_i^2} \sigma^2 \\
&= \frac{\sigma^2}{\sum x_i^2}
\end{aligned}
$$

Question 3

Prove a polynomial of degree k, $a_k n^k + a_{k-1} n^{k-1} + \ldots + a_1 n^1 + a_0 n^0$ is a member of $\Theta(n^k)$ where $a_k \ldots a_0$ are nonnegative constants.

Proof. To prove that $a_k n^k + a_{k-1} n^{k-1} + \ldots + a_1 n^1 + a_0 n^0$, we must show the following:

$$\exists c_1 \exists c_2 \forall n \geq n_0,\ c_1 \cdot g(n) \leq f(n) \leq c_2 \cdot g(n)$$

For the first inequality, it is easy to see that it holds because no matter what the constants are, $n^k \leq a_k n^k + a_{k-1} n^{k-1} + \ldots + a_1 n^1 + a_0 n^0$ even if $c_1 = 1$ and $n_0 = 1$. This is because $n^k \leq c_1 \cdot a_k n^k$ for any nonnegative constant, c_1 and a_k.

Taking the second inequality, we prove it in the following way. By summation, $\sum\limits_{i=0}^{k} a_i$ will give us a new constant, A. By taking this value of A, we can then do the following:

$$\begin{aligned} a_k n^k + a_{k-1} n^{k-1} + \ldots + a_1 n^1 + a_0 n^0 = & \\ & \leq (a_k + a_{k-1} \ldots a_1 + a_0) \cdot n^k \\ & = A \cdot n^k \\ & \leq c_2 \cdot n^k \end{aligned}$$

where $n_0 = 1$ and $c_2 = A$. c_2 is just a constant. Thus the proof is complete. □

Question 18

Evaluate $\sum_{k=1}^{5} k^2$ and $\sum_{k=1}^{5}(k-1)^2$.

Question 19

Find the derivative of $f(x) = x^4 + 3x^2 - 2$

Question 6

Evaluate the integrals $\int_0^1 (1 - x^2)\mathrm{d}x$ and $\int_1^\infty \frac{1}{x^2}\mathrm{d}x$.