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 1 continued on next page. . .

Part 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.

Solution

One core challenge with evaluating the causal effect of membership in a local Chamber of Commerce on firm revenues and profits is that we cannot observe a firm i in two different states of the world: one where the firm is a member, and one where it is not. Instead, we can compare the mean outcomes of two firm groups (members vs. non-members) to learn about the true, average treatment effect (ATE) of membership.

The ATE is defined as:

$$ATE = E(Y_i(1) - Y_i(0))$$

where $Y_i(1)$ represents the potential outcome (the annual sales) of a firm i if part of a local Chamber of Commerce and $Y_i(0)$ represents the annual sales for the **same** firm if not.

To compare the mean outcomes of the two groups of firms (members vs. non-members of a local Chamber of Commerce), we instead estimate:

$$E(Y_i(1) \mid X_i = 1) - E(Y_i(0) \mid X_i = 0)$$

where we condition our observation of Y_i on X_i, with $X_i = 1$ representing if firm i is part of a local Chamber of Commerce, and $X_i = 0$ if it is not. The above is not necessarily equal to the ATE defined earlier because X_i, or membership status, is *not independent* of potential outcomes (annual firm sales). We would have to show first that, in the absence of a local Chamber of Commerce, member and non-member firms would have to be on average identical (there would have to be no selection bias).

Unpacking the above expression we can see the effect of selection bias:

$$\begin{aligned}E(Y_i(1) \mid D_i = 1) - E(Y_i(0) \mid D_i = 1) = {} & \underbrace{E(Y_i(1) \mid D_i = 1) - E(Y_i(0) \mid D_i = 1)}_{\text{Average Treatment Effect on the Treated}} \\ & + \underbrace{E(Y_i(0) \mid D_i = 1) - E(Y_i(0) \mid D_i = 0)}_{\text{Selection Bias}}\end{aligned}$$

The first term is the average treatment effect on the treated (ATT), the difference between $Y_i(1)$ and $Y_i(0)$ for the group of firms that are members of a local Chamber of Commerce. The second term estimates the selection bias by comparing $Y_i(0)$, annual sales before being a member, between members, $X_i = 1$, and non-members, $X_i = 0$. $E(Y_i(0) \mid X_i = 1)$ cannot be observed directly for both of these terms, because we cannot rerun the experiment a second time where the treatment is not applied. The bias term appears because firms that are members of a local Chamber of Commerce may have different outcomes than firms that are not members, *even in the absence of a local Chamber of Commerce.*

Question 1 continued on next page. . .

Part 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.

Solution

In a RCT, treatment and control groups are randomly selected, and the average outcomes of these two groups are compared after the treatment. By randomly allocating treatment status, we can use the simple difference in outcomes between treatment and control groups to estimate the ATE. This is because the randomization removes selection bias by ensuring that member firms and non-member firms are comparable in terms of both observable and unobservable characteristics and that any difference between the two groups is not due to systematic differences. This implies that the ATE = ATT.

Connecting back to the potential outcomes framework, $D_i = 1$ indicates if firm i is a member of a local Chamber of Commerce (the treatment), and $D_i = 0$ if the firm is not (D_i indicates treatment status). $Y_i(1)$ are the annual sales (the potential outcome) for a firm i if treated (part of a local Chamber or Commerce), and $Y_i(0)$ are the annual sales for the same firm if not.

The ATE is then defined as:

$$ATE = E(Y_i(1) - Y_i(0))$$

If the treatment assignment D_i is assigned randomly and independent of the potential outcomes of the firms ($D_i \perp (Y_i(1), Y_i(0))$), we can compare the mean outcomes of the two groups (members vs. non-members of a local Chamber of Commerce):

$$\begin{aligned} E(Y_i(1) \mid D_i = 1) - E(Y_i(0) \mid D_i = 1) &= \underbrace{E(Y_i(1) \mid D_i = 1) - E(Y_i(0) \mid D_i = 1)}_{\text{Average Treatment Effect on the Treated}} \\ &+ \underbrace{E(Y_i(0) \mid D_i = 1) - E(Y_i(0) \mid D_i = 0)}_{\text{Selection Bias}} \\ &= E(Y_i(1) \mid D_i = 1) - E(Y_i(0) \mid D_i = 1) \quad \text{(selection bias is zero)} \\ &= ATT \end{aligned}$$

because the treatment status is independent of outcomes, we can also show:

$$ATT = E(Y_i(1) - Y_i(0)) = ATE$$

This means that if we randomly assign treatment status (membership in a local Chamber of Commerce) and control status (non-membership), we can estimate the ATE by comparing mean sales outcomes across the two groups of observation. We can report the estimated treatment effect of the RCT design from the OLS regression, $ln(Y_i) = \beta_0 + \beta_1 D_i + u_i$. The β_1 coefficient is the difference in means between the treatment group relative to the control group. Dividing this difference by its standard error produces a t-statistic that allows us to determine whether the observed difference in annual sales between firms part of a local Chamber of Commerce and firms that are not is statistically significant.

Part 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?

Solution

In the case where assigned treatment status, D_i, cannot be enforced on firms, T_i represents the effective treatment status of a firm i. Some firms may not comply with their assigned treatment status: firms assigned to become members of a local Chamber of Commerce may opt not to, and firms assigned to be non-members might become members. However, firms' choices to deviate from their assigned treatment are not random, and T_i may be correlated with annual sales (potential outcomes). $T_i \perp (Y_i(0), Y_i(1))$ is no longer true, which means the difference between the mean outcomes of the control and treatment groups may be influenced by selection bias.

However, the initial assignment status, D_i, is still randomly assigned and may have strong predictive power on actual treatment status, T_i. So, to address our original concern, we can use D_i as an instrumental variable for T_i to estimate the causal effect. To be a valid instrumental variable, D_i must satisfy three conditions: instrument relevance, instrument exogeneity, and the instrument exclusion restriction.

Instrument relevance. The instrument needs to be related strongly enough with the endogenous, explantory variables of interest T_i. So, as long as the assignment protocol is partially followed, then the actual treatment status will be partially determined by the assigned treatment status, making D_i a relevant instrumental variable.

Instrument exogenity. The instrument must be uncorrelated with any observed variables that also affect Y_i in the error term. That is, $Cor(D_i, u_i) = 0$. Since initial treatment assignment is random, then D_i is distributed independently of u_i, so the instrument is exogenous.

Instrument exclusion restriction. The instrument must not have a direct effect on Y_i except through its relationship with T_i. That is, conditional on T_i, D_i has no effect on Y_i. Again, since the initial treatment assignment is random, D_i is distributed independenty of any other variable than T_i that can have an effect on Y_i, satisfying the exclusion restriction.

Therefore, the original random treatment assignment is a valid instrumental variable, and the following is the two-stage least squares regression:

$$\begin{aligned} T_i &= \beta_0' + \beta_1' D_i + u_i' \quad \text{(first stage)} \\ ln(Y_i) &= \beta_0 + \beta_1 \hat{T}_i + u_i \quad \text{(second stage)} \\ ln(Y_i) &= \beta_0'' + \beta_1'' D_i + u_i'' \text{ (reduced form)} \end{aligned}$$

In contrast to our original regression, our point estimate from the second stage is: , which gives us a weighted average of the treatment effect β_{1i}, where the weights increase with the degree to which the instrument D_i influences T_i in the first stage.

$$\hat{\beta_{IV}} = \frac{\beta_1''}{\beta_1'} = \frac{E(Y_i|D_i = 1 - Y_i|D_i = 0)}{E(T_i|D_i = 1 - T_i|D_i = 0)} = \frac{\text{Difference in mean outcomes between groups}}{\text{Difference in fraction treated between groups}}$$

Question 1 continued on next page. . .

Part 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.

Solution

The concern behind spillover effects is that, even if we were to isolate the treatment (membership in a local Chamber of Commerce) to firms only in the treatment group and restrict membership for those in the control group, we still have to worry about the effects that the treatment has on the control group. For example, in the case where membership in a local Chamber of Commerce improves firm sales for members, increased sales may contribute to overall economic surplus and increas sales for firms in the control group. This means that the control group is no longer a reference in which nothing has changed.

One solution would be to

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, $\sum_{i=0}^{k} a_i$ 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 $\int_1^\infty \frac{1}{x^2}\mathrm{d}x$.