Ec240a, Fall 2019

Professor Bryan Graham

Problem Set 5

Due: December 16th, 2019

Problem sets are due at 5PM in the GSIs mailbox. You may work in groups, but each student should turn in their own write-up (including a printout of a narrated/commented and executed Jupyter Notebook if applicable). Please also e-mail a copy of any Jupyter Notebook to the GSI (if applicable).

Model based instrumental variables

This problem set provides an opportunity to implement the model based instrumental variables estimator introduced by Imbens & Rubin (1997). For a textbook discussion see Imbens & Rubin (2015) and the provided lecture notes.

1 Dataset simulation

You will use simulated data to complete this problem set. Let $X \in \{0,1\}$ be a binary encouragement that is randomly assigned to $i = 1, \ldots, N$ randomly sampled units according to a fair coin flip (i.e., $\Pr(X=1) = \Pr(X=0) = \frac{1}{2}$). Let the population frequencies of *never-takers*, *compliers* and *always-takers* be, respectively $\pi_N = \frac{1}{3}$, $\pi_C = \frac{1}{3}$ and $\pi_A = \frac{1}{3}$. Let A be a 3×1 vector corresponding to a random draw from a Multinomial$(1, \pi_N, \pi_C, \pi_A)$ distribution. The four compliance-strata-specific potential outcome distributions are

$$\begin{aligned}
Y(0)|\,A = (1,0,0)' &\sim \mathcal{N}(0,1) \\
Y(0)|\,A = (0,1,0)' &\sim \mathcal{N}(1,1) \\
Y(1)|\,A = (0,1,0)' &\sim \mathcal{N}(3,1) \\
Y(1)|\,A = (0,0,1)' &\sim \mathcal{N}(4,1).
\end{aligned}$$

Simulate $i = 1, \ldots, N$ random draws of $\{X_i, D_i, Y_i\}$ according to the following algorithm (for $N = 1000$).

1. Let $A_i \sim$ Multinomial$(1, \pi_N, \pi_C, \pi_A)$ and $X_i \sim$ Binomial$(1, \frac{1}{2})$.

2. Let $D_i = 1$ if the unit chooses the active treatment and zero otherwise. Specifically

$$\begin{aligned} D_i =& 0 \cdot \mathbf{1}\left(A_i = (1,0,0)'\right) + 0 \cdot (1 - X_i) \cdot \mathbf{1}\left(A_i = (0,1,0)'\right) \\ &+ 1 \cdot X_i \cdot \mathbf{1}\left(A_i = (0,1,0)'\right) + 1 \cdot \mathbf{1}\left(A_i = (0,0,1)'\right). \end{aligned}$$

   Observe that actual treatment is deterministic given a unit's compliance-strata/type and encouragement.

3. Let Y_i be a random draw from the relevant type-and-treatment specific potential outcome distribution.

4. Summarize your simulated dataset. What is the average outcome difference across treated and control units? How does this difference compare with the population LATE? Explain/discuss.

2 Estimation

1. Compute the WALD-IV estimate of the local average treatment effect of D on Y:

$$\beta_0^{\text{LATE}} = \mathbb{E}\left[Y(1) - Y(0) \middle| A = (0,1,0)'\right].$$

   Derive the asymptotic sampling distribution of $\sqrt{N}\left(\hat{\beta}_{\text{WALD}}^{\text{LATE}} - \beta_0^{\text{LATE}}\right)$ and use this distribution to construct an approximate standard error for $\hat{\beta}_{\text{WALD}}^{\text{LATE}}$. Can you reject the null of a zero LATE?

2. Compute the maximum likelihood estimate of β_0^{LATE} based on the complete data log-likelihood.

3. Compute the maximum likelihood estimate of β_0^{LATE} based on the observed log-likelihood using the EM-Algorithm.

4. Compare and discuss your three different LATE point estimates. Provide a verbal description/explanation of the EM-Algorithm in the present context.

3 Parametric Bootstrap

The EM-Algorithm will deliver estimates of all the parameters used to simulate your dataset except those generating the encouragement/instrument. We will condition on the observed/simulated encouragements in what follows. For each of $b = 1, \ldots, B$ bootstrap replications simulate a new dataset according to the algorithm outlined in Section 1 above, but using the parameters estimated in question 3 of Section 2 above instead of those indexing the "true" data generating process (for $B = 1000$). Leave the configuration of encouragements fixed across each bootstrap replication at whatever was simulated initially in Section 1.

1. For each bootstrap simulation compute the maximum likelihood estimate of β_0^{LATE} based on the observed log-likelihood using the EM-Algorithm. Let $\hat{\beta}^{(b)}$ be the estimate associated with the b^{th} bootstrap simulation.

2. Construct a 95 percent confidence interval for $\hat{\beta}_{\text{MLE}}^{\text{LATE}}$ (as estimated in question 2 of Section 2) using the 0.025 and 0.975 sample quantiles of $\left\{\hat{\beta}^{(b)}\right\}_{b=1}^{B}$.

3. Construct a histogram for $\left\{\hat{\beta}^{(b)}\right\}_{b=1}^{B}$. Discuss.

4. Repeat your analysis but now let the population frequencies of *never-takers*, *compliers* and *always-takers* be, respectively $\pi_N = \frac{4}{9}$, $\pi_C = \frac{1}{9}$ and $\pi_A = \frac{4}{9}$. Discuss any changes in your results.

References

Imbens, G. W. & Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variable models. *Review of Economic Studies*, 64(4), 555 – 574.

Imbens, G. W. & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge: Cambridge University Press.