The Exam

EC 425/525: Econometrics

Typed, readable submissions due at 5PM on Friday, 14 June 2019

Instructions

DUE Your submission is due at 5PM on Friday, 14 June 2019 on Canvas.
Your submission **must be typed with R code and output/figures beneath your responses**.
E.g., `knitr` and `R Markdown`. Word files are fine.

IMPORTANT **You may not work with anyone else**. You may not communicate with anyone about this exam except Jenni and Ed. Any evidence of a violation of this rule will result in a grade of zero.

IMPORTANT$_2$ **All materials are allowed**—notes, previous assignments/solutions, Google, *etc.*.

IMPORTANT$_3$ You may use any functions/packages you would like on this exam—and any statistical package (R, `Stata`, *etc.*). I can only guarantee technical support from Jenni and me in R.

OBJECTIVE This exam will test your knowledge and help you continue building intuition and skills.

POINTS There are 104 points possible on this exam.

LASTLY Good luck and thank you for a great quarter.

Section 1: ATE, TOT, ITT, or LATE

32 points (4 points per part)

Without explanation Full credit for correct answers; 0 points for incorrect answers.

With explanation 50% for correct answer. Up to 50% for incorrect answer with good explanations.

In this section, choose whether the described point estimate is the ATE, TOT, ITT, or LATE. There may be multiple answers; in those cases choose the most general anser. Assume all estimates can be interpreted as causal. Treatment effects can be heterogeneous.

1. Consider the OLS-based estimate of τ from the model

$$\mathrm{Y}_i = \alpha + \tau \mathrm{D}_i + \varepsilon_i$$

Which type of treatment effect does $\hat{\tau}$ represent?

2. Consider the case of blocking on propensity scores. With n individuals we calculate

$$\hat{\tau} = \frac{\sum_b n_b \hat{\tau}_b}{n}$$

where $\hat{\tau}_b$ is the estimated treatment effect in block b and n_b is the number of observations in block b.

Which type of treatment effect does $\hat{\tau}$ represent?

3. Let X_i be a continuous variable. When X_i exceeds the value t, the probability that individual i participates in a program increases from 0.2 to 0.4.

We regress our outcome Y_i on an indicator for $\mathrm{X}_i > t$. Which type of treatment effect does the coefficient on this indicator represent?

4. Using nearest-neighbor matching on a set of covariates, we match the outcome for each of our treated individuals Y_i to the outcome for her nearest control neighbor $\mathrm{Y}_j^{(i)}$. We then calculate the average difference between Y_i and $\mathrm{Y}_j^{(i)}$, *i.e.*,

$$\hat{\tau} = Avg\left(\mathrm{Y}_i - \mathrm{Y}_j^{(i)}\right)$$

Which type of treatment effect does $\hat{\tau}$ represent?

5. A company approves credit-card applications if and only if the applicant's credit score exceeds 600. To test the effect of credit-card approval on consumption, you estimate the linear model

$$\mathrm{Consumption}_i = \beta_0 + \beta_1 \left(\mathrm{Score}_i - 600\right) + \beta_2 \mathrm{Approved}_i + \beta_3 \left(\mathrm{Score}_i - 600\right) \times \mathrm{Approved}_i + \varepsilon_i$$

Which type of treatment effect does $\hat{\beta}_2$ represent?

6. You estimate τ via instrumental variables. Which type of treatment effect does $\hat{\tau}$ represent?

7. You estimate τ via instrumental variables **and** there are no always takers. Which type of treatment effect does $\hat{\tau}$ represent? *Hint:* You answer should differ from **6**.

8. You regress your outcome Y_i on an instrument Z_i. What does the coefficient on Z_i represent?

Section 2: The value of a business degree

42 points (6 points per part)

Supopse we want to estimate the causal effect of a business degree (Business_i) at UO on earnings 10 years after graduation (Income_i). The university gives you a random sample of undergraduates at UO who applied to be a major in the business school. Assume a homogeneous treatment effect.

For each individual, you observe

- Income_i, the individual's income 10 years after graduation
- Business_i, an indicator for whether the student was accepted by the business school
- SAT_i, the individual's SAT score (from their application to attend UO)
- GPA_i, the individual's GPA at the time of their application to the business school
- Econ_i, an indicator for whether the student passed the introductory economics sequence

9. Your first impulse is to regress Income_i on Business_i, SAT_i, GPA_i, and Econ_i. Under which conditions will the coefficient on Business_i be interpretable as the causal effect of a business degree at UO on income?

10. You had such a great time coding up the matching estimator on the first problem set in 525, you decide to do it again. Specifically, match admitted individuals to non-admitted individuals who have identical values of SAT_i, GPA_i, and Econ_i. Within each combination of SAT_i, GPA_i, and Econ_i, you calculate the difference between the average income for admitted individual and the average income for rejected individuals.

a. Under which conditions can we interpret this estimate as a causal effect?
b. Would you ever prefer the estimate in **9** to this estimate in **10**?

11. After several emails, UO offers to provide you with $\text{GPA}_i^{\text{Final}}$, the GPAs of individuals when they graduate from UO. Your contact thinks this variable could be important, as individuals with higher college GPAs tend to make more money.

a. Are there advantages or disadvantages to replacing GPA_i with $\text{GPA}_i^{\text{Final}}$ in **9**?
b. Are there any advantages or disadvantages to adding this new variable $\text{GPA}_i^{\text{Final}}$ as an additional control in the regression described in **9**?

12. Suddenly, you remember how cool propensity-score matching seemed. Using a logitstic regression, you regress Business_i on SAT_i, GPA_i, and Econ_i. You then add the predictions from this regression ($\hat{p}_i$) to the regression described in **9**.

Under which conditions will this estimate be interpretable as causal?

13. During a conversation with an administrator from the business school, you learn that the business only admits students if two conditions are true: (**1**) their GPA is above 3.0, and (**2**) the student passed the introductory economics sequence.

Describe an estimation strategy in which you could credibly estimate the causal effect of admittance into the business school using the variables described at the start of this section. Include any assumptions your strategy makes.

Section 3: A simulation (of course!)

30 points (5 points per part)

Back to instrumental variables. We're going to explore some violations/bastardizations of IV/2SLS.

The data-generating process

Set up a data-generating process

$$\mathrm{Y}_i = (\beta_0 = 3) + (\beta_1 = 2)\,\mathrm{X}_i + u_i \tag{1}$$

where X_i is correlated with u_i. In your data-generating process, include a valid instrument Z_i for X_i. Set up the variance-covariance matrix Σ among the variables X_i, u_i, and Z_i

$$\Sigma = \mathrm{Cov}(\mathrm{X}_i,\, u_i,\, \mathrm{Z}_i) = \begin{bmatrix} 1 & 0.5 & 0.7 \\ 0.5 & 1 & 0 \\ 0.7 & 0 & 1 \end{bmatrix}$$

Each variable is mean zero and comes from a multivariate Normal distribution. For each iteration (run at least 1,000 iterations), sample $N = 50$ individuals from this defined DGP.

Finally, generate the variable $\mathrm{W}_i = \mathrm{X}_i + \varepsilon_i$ where $\varepsilon_i \sim N(0, 1)$.

The models

In each iteration, estimate the effect of X_i on Y_i using each of the following five models:

- **A.** The OLS estimate for β_1 in equation (1)
- **B.** The 2SLS estimate of β_1 in equation (1), in which Z_i instruments for X_i
- **C.** The 2SLS estimate of β_1 in equation (1), in which the indicator $\mathbb{I}\{\mathrm{Z}_i > 0\}$ instruments for X_i
- **D.** The OLS estimate of β_1 in (1) when we add Z_i as a control.
- **E.** The 2SLS estimate of β_1 in (1) using W_i as a control (in both stages).

The actual questions

14. Which elements of Σ tell us that Z_i is a valid instrument?

15. Which models appear to provide unbiased estimates? Provide graphical and numeric summaries of each model (with labeled axes and models clearly distinguished).

16. Which models do we expect to consistently estimate the causal effect of X_i on Y_i? Explain why.

17. How does the model described in **D.** differ from the typical IV/2SLS setup? Explain how this difference could lead to biased/inconsistent estimates for β_1.

18. What is wrong with the model described in **E.**? Explain.

19. In this setting, does the indicator variable $\mathbb{I}\{\mathrm{Z}_i > 0\}$ provide a valid instrument? How does it affect efficiency, relative to the 2SLS model described **B.**? Explain.

The end!