

FINAL EXAM

There are 2 questions, each with multiple parts. All parts are given equal weight. You may not use any notes, books, electronic devices, or other aids (including collusion). And most important: don't panic!

1. Be sure to read all parts of this question carefully before you begin (your answer to part d will depend on how you choose to answer part c, so choose carefully).
 - a. What is selection bias? Explain how and why selection bias causes problems for estimating the causal effect of a treatment D on an outcome of interest, Y .
 - b. Suppose the university decides to offer free weekly tutoring sessions to any SFU student who wants tutoring. Let Y_i denote the GPA of student i at graduation. Let $D_i = 1$ if student i participates in the free tutoring sessions, and $D_i = 0$ if student i does not. Let X_i denote a set of other observable characteristics of student i that are thought to affect GPA (e.g., major, number of hours worked for pay, etc.). Suppose we estimate the regression, $Y_i = \beta X_i + \delta D_i + \varepsilon_i$. Does δ estimate the causal effect of the free tutoring sessions on GPA at graduation? Is there a selection bias problem here? If so, what is the nature of the selection bias, and what is its consequence? Explain.
 - c. Describe a better econometric methodology for estimating the causal effect of the tutoring program on GPA. Feel free to change the tutoring program (e.g., who is eligible for tutoring, what kind tutoring is offered, how often, etc.) if doing so is helpful. Be sure to explain what kind of data your method requires, what model you would estimate, how your method deals with the selection bias problem, and what assumptions are required for your method to give an unbiased or consistent estimate of the causal effect of interest. If you choose to change the tutoring policy as part of your answer, be sure to clearly explain the change(s).
 - d. Choose one paper that we discussed this semester, and briefly describe how it applies your method from part c to estimate a causal effect. Be sure to explain what causal effect the authors are trying to estimate, what data they use, what model they estimate, how their method deals with the selection bias problem, and how it identifies the causal effect of interest.

2. During the financial crisis and recession of 2008-2009, the government of Canada introduced the Career Training Assistance (CTA) initiative. Roughly speaking, the CTA was a cash subsidy given to unemployed workers who did some training and met some conditions. The objective of the CTA was to encourage unemployed workers to get some training, in the hope that training would improve their subsequent labour market performance. Suppose that you have been hired by the government of Canada to estimate the effect of the CTA on unemployment duration (i.e., how long an unemployed person remains unemployed).

Suppose the CTA came into effect on February 1 2009. You have data on a random sample of 10,000 workers who became unemployed between August 1 2008 and July 31 2009. Your sample is randomly selected from all 10 provinces. For each unemployed worker i in the sample, you observe the length of time that he/she was unemployed (Y_i), the date he/she became unemployed (T_i), the unemployment rate in his/her city (U_i), and a set of background characteristics (X_i , such as age, gender, etc.). In addition, each city in the country was assigned a random lottery number (L_i) between 1 and 100.

- a. Suppose the government chose to introduce CTA only in cities where U_i was more than 8 percent. Explain how you would measure the causal effect of CTA on unemployment duration (Y_i) in this case. What assumptions are required for your method to estimate the causal effect?
- b. Suppose instead that the government chose to introduce CTA only in cities where $L_i \leq 50$. Explain how you would measure the causal effect of CTA on unemployment duration in this case. What assumptions are required for your method to estimate the causal effect?
- c. Suppose instead that the government used a function of both the unemployment rate and the lottery numbers to choose the cities in which CTA would be introduced. In particular, the **probability** that CTA was introduced in a particular city was $p = L_i/100 + U_i/20$. Explain how you would measure the causal effect of CTA on unemployment duration in this case. What assumptions are required for your method to estimate the causal effect?
- d. Suppose instead that the government chose to implement CTA in **all** cities. Can you measure the causal effect of CTA on unemployment duration in this case? If yes, explain how and be sure to clearly state your method's assumptions. If not, explain why not.
- e. Which of your four estimators from parts a-d do you prefer, and why? Explain.