ECON 484
Summer 2011
Prof. Simon Woodcock

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 fitness classes to any SFU student who wants to participate. Let Y_i denote a measure of the overall health of student i at graduation. Let $D_i = 1$ if student i participates in the free fitness classes, 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 overall health (e.g., age, diet, etc.). Suppose we estimate the regression, $Y_i = \beta X_i + \partial D_i + \varepsilon_i$. Does ∂ estimate the causal effect of the free fitness classes on health 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 fitness classes on student health. Feel free to change the fitness class policy (e.g., who is eligible to take the fitness classes, what kind classes are 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 fitness class 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. In 2010, the province of B.C. introduced a policy that increased the length of the school day for Kindergarten students from a half day to a full day. Suppose that you have been hired by the province of B.C. to measure the effect of full day Kindergarten (FDK) on the number of hours that young mothers work in the labour force.

    You have data on a random sample of 4000 young mothers in B.C.: 1000 who had a child enrolled in Kindergarten in 2008; 1000 who had a child enrolled in Kindergarten in 2009; 1000 who had a child enrolled in Kindergarten in 2010; and 1000 who had a child enrolled in

Kindergarten in 2011. Your sample is randomly selected from all 60 school districts in the province. For each mother *i* in the sample, you observe the number of hours that she worked (Y_i), the year (T_i), the average family income in her school district (F_i), and a set of background characteristics (X_i, such as her age, education, etc.). In addition, each **school district** was assigned a random lottery number (L_i) between 1 and 60.

To manage the cost of the policy change, the government considered several different options. Under most of these options, the government sought to spread the cost of the policy change over 2 years by introducing FDK in half of B.C.'s school districts in 2010, and then introducing FDK in the remaining school districts in 2011.

a. Suppose the government chose to introduce FDK in all school districts where F_i was **less** than the B.C. average in 2010, and in all school districts where F_i was **more** than the B.C. average in 2011. Explain how you would measure the causal effect of FDK on mothers' hours of work in this case. What assumptions are required for your method to estimate the causal effect?
b. Suppose instead that the government chose to introduce FDK in all school districts where $L_i \leq 30$ in 2010, and all school districts where $L_i > 30$ in 2011. Explain how you would measure the causal effect of FDK on mothers' hours of work 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 average family income and the lottery numbers to choose which districts would implement FDK in 2010, and which districts would implement FDK in 2011. In particular, the **probability** that FDK was introduced in a particular district in 2010 was $p = L_i/120 + F_i/200{,}000$. Explain how you would measure the causal effect of FDK on mothers' hours of work in this case. What assumptions are required for your method to estimate the causal effect?
d. Suppose instead that the government chose to implement FDK in **all** school districts in 2010. Can you measure the causal effect of FDK on mothers' hours of work in this case? If yes, explain how and be sure to clearly state your method's assumptions. If not, explain why not.
e. Same as part d, but now suppose you also have a sample of young mothers (covering the same years 2008-2011) who had a child enrolled in Grade 1. Note that FDK did not change the length of the school day for Grade 1 students (it was a full day both before and after the introduction of FDK). Can you use this extra information to measure the causal effect? If yes, explain how and be sure to clearly state your method's assumptions. If not, explain why not.
f. Which of your four estimators from parts a-e do you prefer, and why? Explain.