ECON 4261: Final

May 6, 2023

Instructions

- Write your **name** and **recitation number/instructor** on your answer booklet. Your exam will not be graded if you don't provide this information clearly.
- Read every question **carefully**!
- Please ensure that your answers are **neat** and **legible**.
- There are 40 points available. Good luck!

Question 1 - 14 Points

In this section you are going to use some data to diagnose economic inequality. You have a cross-sectional dataset:

$$\{W_n, F_n, B_n, Yf_n, Y_n, G_n\}_{n=1}^N$$

where:

- W_n is log annual earnings in dollars
- $F_n \in \{0, 1\}$ indicates if person n has ever had children.
- B_n is person n's year of birth.
- Yf_n is the year in which person n had their first child (data missing if $F_n = 0$)
- Y_n is the calendar year for the observation.
- $G_n = \{M, F\}$ indicates gender.

We want to use this dataset to analyze the relationship between fertility, gender, and wages. Let $X_n = (F_n, B_n, Yf_n, Y_n)$.

(a)[4 points] Start with the model:

$$\mathbb{E}[W_n|X_n, G_n = g] = \sum_a \mathbf{1}\{Age_n = a\}\mu_{a,g} + \sum_y \mathbf{1}\{Y_n = y\}\gamma_{y,g} + \alpha_g F_n$$

Describe how you would estimate this model and calculate standard errors for the coefficients. Provide enough details that someone could follow your instructions. *Do not* just say what command you would use in `R`.

(b)[3 points] In this model, which combination of parameters represents the *difference* in the gender gap between individuals with children ($F_n = 1$) and those without ($F_n = 0$):

$$\begin{aligned}\text{Diff in Gender Gap} = &(\mathbb{E}[W_n|X_n, F_n = 1, G_n = M] - \mathbb{E}[W_n|X_n, F_n = 1, G_n = F]) \\ &- (\mathbb{E}[W_n|X_n, F_n = 0, G_n = M] - \mathbb{E}[W_n|X_n, F_n = 0, G_n = F])\end{aligned}$$

(c) [3 points] Explain how you would calculate a 95% confidence interval for the difference in the gender wage gap (defined in part (b)) using your estimates and standard errors from part (a). You can assume that the set of male-specific coefficients is uncorrelated (i.e. zero covariance) with the set of female-specific coefficients.

(d)[2 points] Explain how you would change the model to allow for the difference in the gender wage gap (defined in part(b)) to have a separate value for each calendar year. You only have to write the model, you do not have to describe estimation.

(e)[2 points] Explain how you would change the model to allow for the difference in the gender wage gap (define in part(b)) to grow linearly with the number of years since birth of the first child ($Y_n - Yf_n$). You only have to write the model, you do not have to describe estimation.

Question 2 - 14 Points

Consider a staggered rollout of a federal subsidy for community college across counties in the US. Suppose you have the following data:

$$\{E_{c,t}, Y_{c,t}, P_{c,t}\}_{c=1,t=1}^{C,T}$$

where

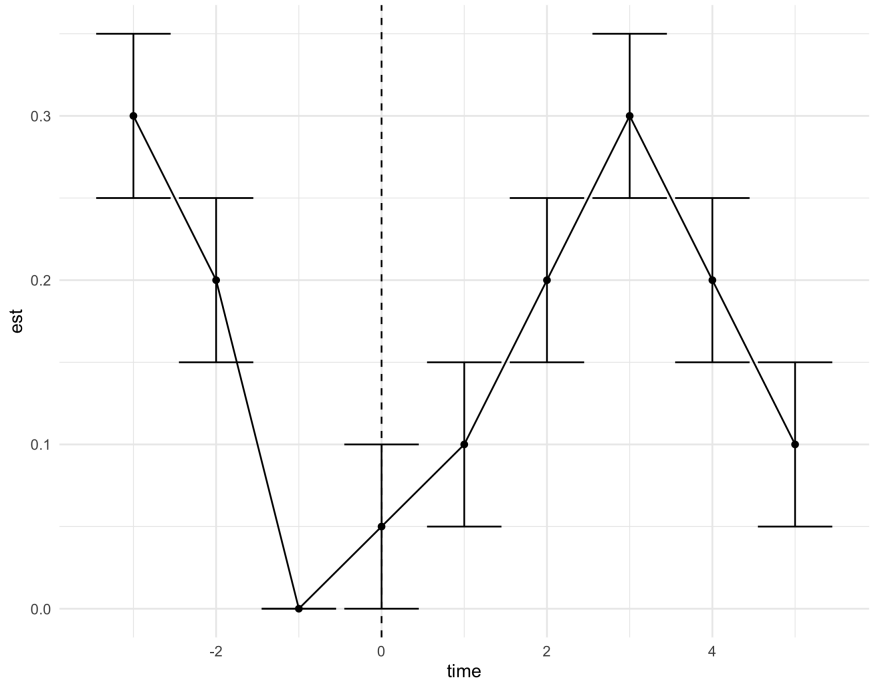
- $P_{c,t}$ is a binary variable that indicates whether the policy has been introduced in county c at time t
- $E_{c,t}$ is the community college enrollment rate in county c at time t.
- $Y_{c,t}$ is average earnings in county c at time t.

(a) [3 points] Propose a difference-in-differences strategy for estimating the effect of the tuition subsidy policy on community college enrollment. Write a model where the effect of the policy is α, which is the same across counties and is immediate after the policy is introduced.

(b) [3 points] Write an event-study model of the effect of the policy on enrollment and earnings where α_s is the effect of the policy s periods after introduction, and the effect of the policy is constant after 10 periods.

(c) [4 points] Describe how you would estimate this event-study model along with standard errors. You can be brief but provide enough details that someone could follow the steps. *Do not* just write what command you would use in `R`.

(d) [2 points] Suppose you estimate the event-study model and you plot your coefficients, $\hat{\alpha}_s$, with 95% confidence intervals with the below result:



where the x-axis indicates the number of periods since introduction of the policy ("event time"). Suppose that the policy introduction is completely unanticipated. Does this picture propose any challenges to your model assumptions? Which assumption does it appear to violate?

(e) [2 points] Suppose that the policy is introduced *simultaneously* in every county in period 3 ($t = 3$). Will this cause problems for your difference-in-differences study? Which condition is violated?

Question 3 - 12 Points

Consider again the tuition subsidy from the previous question You are now going to estimate a TSLS model:

$$Y_{n,t} = \mu_c + \kappa_b + \gamma_t + \alpha D_{n,t}$$
$$D_{n,t} = \tilde{\mu}_c + \tilde{\kappa}_b + \tilde{\gamma}_t + \delta P_n$$

where

- $Y_{n,t}$ is the earnings of person n at time t
- $D_{n,t}$ is a dummy that indicates whether person n at time t has been to community college.
- μ_c and $\tilde{\mu}_c$ are county fixed effects.
- γ_t and $\tilde{\gamma}_t$ are calendar time fixed effects.
- κ_b and $\tilde{\kappa}_b$ are year of birth fixed effects.
- P_n is an indicator for whether the tuition policy was in place in person n's county in the year that person n turned 18.

(a) [4 points] Describe how you would estimate the parameter α along with standard errors. You can be brief but provide enough details that someone could copy your steps. *Do not* just say what command you would use in `R`.

(b) [3 points] In the model, what is the logic for including the county and birth fixed effects? What might go wrong if we don't include these parameters?

(c) [2 points] Suppose that the effect of community college on earnings is in fact heterogeneous across individuals. How can you interpret the TSLS estimand, α, in this case?

(d) [2 points] Suppose now that

1. You run a diff-in-diff analysis as in Question (2) and you find that the subsidy policy caused the price of community college tuition to increase; and
2. The subsidy is only offered to individuals in the county who are below a certain income threshold.

What condition could be violated that was necessary for your interpretion in part (c)? Why?