ECON 121, Applied Econometrics and Data Analysis   **Prof. Tom Vogl**

Fall 2023   **University of California San Diego**

PROBLEM SET 5: EFFECT OF MILITARY CONSCRIPTION ON CRIME

DUE BY 11:59 PM PST ON TUESDAY, 12/12

Many countries require young men to serve in the military. Proponents of these policies cite many benefits, including promoting national security and disciplining otherwise undisciplined young men. In this problem set, we will examine the second of these purported benefits by estimating the effect of military conscription on crime in Argentina.

Military service in Argentina was mandatory for young men throughout most of the twentieth century. The needs of the military varied over time, however, so it held an annual lottery to decide which newly eligible men would serve. The following paragraph (taken from a published source) describes the lottery in detail:

> *The eligibility of young males for military service was randomly determined, using the last three digits of their national IDs. Each year, for the cohort due to be conscripted the following year, a lottery assigned a number between 1 and 1,000 to each combination of the last three ID digits. The lottery system was run in a public session using a lottery drum filled with a thousand balls number 1–1,000. The first ball released from the lottery drum corresponded to ID number 000, the second released ball to ID number 001, and so on. The lottery was administered by the National Lottery and supervised by the National General Notary in a public session. Results were broadcasted over the radio and published in the main newspapers. After the lottery, individuals were called for physical and mental examinations. Later on, a cutoff number was announced. Individuals whose ID number had been assigned a lottery number higher than the cutoff number, and who had passed the medical examination, were mandatorily called to military service. Clerics, seminarians, novitiates, and any individual with family members dependent upon him for support were exempted from military service.*

To produce the dataset (https://github.com/tvogl/econ121/raw/main/data/crime_ps5.csv), researchers started with all men born in 1958-1962, divided them into cells by birth year and last three ID digits, and then calculated crime rates for each of these cells. Thus, each observation in the dataset represents a set of men with the same birth year and last three ID digits. (The data are aggregated in this way to ensure confidentiality.) The following table defines the variables in the dataset:

Variable name	Description
birthyr	Birth year
draftnumber	Draft number (1-1000)
conscripted	Fraction conscripted
crimerate	Fraction with criminal record by 2005
argentine	Fraction non-indigenous Argentinean
indigenous	Fraction indigenous Argentinean
naturalized	Fraction naturalized citizens

Our main outcome variable will be `crimerate`, which reflects the probability of ever having a criminal record.

To install R and RStudio, follow this link. You are encouraged to work in a group of up to 4 members. You may write code together, but you must write verbal answers yourself. Please use a Markdown template for your code. Write verbal answers in the comments within the Markdown file, so that you produce a single PDF with code, results, and writing, which you will upload to Gradescope.

1. List your group members.

2. Summarize the data. Are there differences in conscription rates or crime rates across birth years?

3. Use OLS to estimate the relationship between conscription rates and `crimerate`, controlling for observable covariates. Does the results reflect the causal effect of conscription? Describe possible biases.

4. The lottery assigned a draft number to each last three ID digit combination, and the military then set a cutoff based on the needs of the military, such that all draft numbers at or above the cutoff were eligible for conscription. Based on the following cutoffs, code a variable that equals 1 if eligible, 0 if not:

Year:	1958	1959	1960	1961	1962
Cutoff:	175	320	341	350	320

5. Estimate the "first stage" effect of eligibility on conscription. Think about the regression specification. Do you need to control for birth year indicators? Do you need to control for ethnic composition?

6. Estimate the "reduced form" effect of eligibility on `crimerate`.

7. Based on your results for questions (5) and (6), calculate the instrumental variables estimate of the effect of conscription on `crimerate`. You need only calculate a point estimate, not standard errors.

8. Confirm your calculations by running a two-stage least squares regression. Are there differences between the 2SLS (question 8) and OLS (question 3) results? Why or why not?

9. Given your knowledge of the Argentine draft, assess the validity of eligibility as an instrument for conscription. Does it satisfy the criteria for a valid instrument? (WORDS ONLY, NO CODING.)

10. Interpret the 2SLS result. Which sub-population's average treatment effect does it estimate? Is it

reasonable to call it a local average treatment effect? Is it reasonable to call it a treatment-on-the-treated effect? (WORDS ONLY, NO CODING.)

11. Suppose we are concerned that draft numbers are correlated with characteristics that raise a person's risk of committing crimes. Estimate the effect of conscription on `crimerate` in a fuzzy regression discontinuity design. Because the problem set is getting long, I will guide you through it.

    (a) To motivate the regression discontinuity design, draw scatter plots of conscription rates against draft numbers by birth year:

    ```
    ggplot(df, aes(draftnumber, conscripted)) +
    geom_point() +
    facet_wrap(~birthyr)
    ```

    (b) Generate a new variable `distance` that measures the distance from the birth-year-specific cutoff. Drop observations with `distance` greater than 100 or less than -100. This step effectively restricts our analysis to a bandwidth of 100.

    (c) Draw a scatter plot of `conscripted` against `distance`. Do your results suggest that crossing the cutoff raises conscription?

    (d) Draw a scatter plot of the `crimerate` against `distance`. Do your results suggest that crossing the cutoff raises conscription?

    (e) Now run a two-stage least squares regression to estimate the effect of `conscripted` on `crimerate` in a regression discontinuity design. Use local linear regression with a bandwidth of 100. Allow the slopes to be different above and below the cutoff. Report the first and second stage results:

    ```
    df <- df %>% mutate(distXelig = distance*eligible)
    fuzzy_rd <- feols(crimerate ~ distance + distXelig | conscripted ~ eligible, data = df, vcov = 'hetero')
    summary(fuzzy_rd, stage = 1)
    summary(fuzzy_rd, stage = 2)
    ```

    Explain the code, and interpret both the first- and second-stage results. Do the results confirm your interpretation of the scatter plots in parts (c) and (d)?

    (f) Why do you think these results are different from the results earlier in the problem set?