ECON2390

Final exam

Friday, Dec 13th - Monday, Dec 16th, 23:59

Submit your typed or scanned scripts via Assignments tab on the canvas course page.

Note: Take-home exam, open-book, allowed to access any online resources, but not allowed to post the questions on online Q&A forums. **This is not a group work**. You cannot communicate with your peers about the exam questions.

Best luck, everyone!

Answer all the questions, 1 - 3.

1. This question is based on the article by Agan, Doleac, and Harvey (2023, QJE), "Misdemeanor Prosecution," posted in the exam folder of the Canvas course page.

   (a) The authors use a residualized leave-out ADA leniency as an instrument for the nonprosecution decision. Explain how the instrument is constructed, what its value represents (e.g., how to interpret a leniency value of 0.05), and why it satisfies the Random Assignment assumption.

   (b) On page 1467, the authors state: "*As is standard to avoid the small-sample correlation between the ADA decision in this case and her average leniency, we then construct the leave-out mean measure of ADA nonprosecution (leniency).*" Explain why constructing leniency by leaving-out is important. How might failing to account for this issue affect the instrument exogeneity?

   (c) A potential caveat of the paper is that the authors do not directly observe nonprosecution and instead proxy for it by the absence of further case events. Could the use of a proxy introduce bias in the estimates? If so, explain in which direction and why. If not, provide an argument to support the validity of this approach.

   (d) Is the monotonicity assumption plausible in this setting? Explain. If it is plausible, discuss how the tests they do in the paper support your conclusion. If it does not seem plausible, how would your interpretation of the results change? What would the instrument be identifying instead?

(e) Is the exclusion assumption plausible in this context? The authors assess this assumption by investigating whether leniency predicts subsequent case outcomes for prosecuted defendants. However, on page 1473, they concede that "*it is also not clear that if we did find associations between ADA leniency and case outcomes that this would represent a violation of the exclusion restriction.*" What alternative approach could they use to evaluate not only the exclusion restriction but also the overall validity of the instrument? In your answer, giving the main idea of your approach suffices.

(f) Consider the results presented in Table III. Interpret the findings, focusing on what the coefficients imply about the impact of nonprosecution on subsequent criminal complaints. Given that the analysis uses a continuous, constructed instrument (ADA estimated leniency), who are the compliers in this setting?

(g) Seeing the 2SLS estimate from the paper, policymakers suggest a new policy eliminating the possibility of prosecuting misdemeanors for all defendants. What additional steps and/or assumptions do you need to estimate what would be the effect of this policy on future criminal complaints within two years? Discuss.

(h) A researcher points out that ADAs not only have very different prosecution rates but also claim conflicts of interest at very different rates.[1] Furthermore, they claim that anecdotal evidence suggests that ADAs with the highest prosecution rates are also those with the lowest rates of claiming conflicts of interest. If this concern is empirically valid, which assumption, if any, would it violate? If an assumption is violated, would you expect the main estimate of the paper to be biased upward or downward? Discuss.

[1]Before deciding whether to prosecute, ADAs must first decide if they have a conflict of interest with the case. If they do, the case is passed to another ADA.

2. This question is based on the analysis shown in "Social Media and Mental Health" by Braghieri, Levy, and Makarin (2022, AER), as posted on Canvas.

   Consider the following potential outcome framework. Let i denote students, c colleges, and t time periods. Let $D_{i,c,t}$ and $(Y_{i,c,t}(0), Y_{i,c,t}(1))$ respectively denote the treatment status and the potential outcomes without and with treatment of student i in college g at period t. Furthermore, let $G_{i,c} = \min\{t : D_{i,c,t} = 1\}$ denote the first time period that the students i in college c are exposed to the treatment.

   (a) Formally state the no anticipation and parallel trends assumption using the potential outcome framework and interpret them in the context of the paper.

   (b) On page 3664, the authors claim that "*our estimates capture the general equilibrium effects of introducing social media in an entire community. Such general equilibrium effects are arguably particularly important for technologies like social media that exhibit strong network externalities.*" Explain why and how this claim might imply the violation of the main identifying assumptions in the Difference-in-Differences framework.

   (c) The NCHA data that the authors used to construct the outcome variable is a repeated cross-section of students instead of a panel (*i.e.*, we do not observe the same students over time). What additional assumption is required to identify the treatment effect in this setup? Express this assumption using the potential outcome framework above.

   (d) Consider the two-way fixed effects (TWFE) specification (1) on page 3671. Now, assume a simplified two-period setup ($t = 1, 2$) with two colleges $c = 1, 2$. College $c = 1$ gets treated in the first period (*i.e.*, $G_{i,1} = 1$) and college $c = 2$ gets treated in the second period (*i.e.*, $G_{i,2} = 2$). What does the coefficient β identify in this setup? Derive the expression for β, explain it, and discuss whether you observe anything unusual about it.

   (e) More generally (*i.e.*, with many time periods and many groups being treated at different times), what issues arise with specification (1) on page 3671 when treatment effects are heterogeneous across either time or units? What is the source of this problem?

3. A centralized government has decided to implement a tax policy to eliminate income inequality, setting a common target income level, z, for all citizens. Due to high levels of informality and limited data availability, the government can only estimate individual incomes with some degree of uncertainty. For simplicity, assume the government has the authority and resources to adjust individual incomes through taxes and transfers without political or financial constraints. This allows us to focus on the technical aspects of designing a rule to minimize inequality.

   Formally:

   - The population consists of n individuals, indexed by $i = 1, \ldots, n$.
   - Each individual's net income is $y_i \in \mathbb{R}$, and the vector of net incomes is $\boldsymbol{y} = (y_1, \ldots, y_n)$. Net income accounts for gross income minus liabilities or debts, so it can take negative values.
   - The government evaluates the performance of income transfer policy using the following function, which penalizes deviations from the target income z:

     $$L(\boldsymbol{\tau}, \boldsymbol{y}) = \sum_{i=1}^{n} (z - y_i - \tau_i)^2,$$

     where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)$ are the adjustments made to individuals' net incomes.

   The government observes only noisy estimates of incomes, $\hat{\boldsymbol{y}} = (\hat{y}_1, \ldots, \hat{y}_n)$, where each $\hat{y}_i$ is a random variable such that:

   $$\hat{y}_i \overset{\text{iid}}{\sim} N(y_i, 1).$$

   The variance reflects the uncertainty due to measurement errors, informality, or incomplete records.

   (a) Define the key elements of the statistical decision problem: state space, observations, action space, decision function, loss function and risk.

   First consider the case of a single individual i, where the loss function is simplified to $L_i = (z - y_i - \tau_i)^2$ and you only observe $\hat{y}_i$.

   (b) Let the decision rule be defined as $\tau_i^{(c)}(\hat{y}) = c\hat{y}_i$, where $c \in \mathbb{R}$. Find the risk of this decision rule.

   (c) Show that for $c > 1$ the decision rule $\tau_i^{(c)}$ is inadmissible.

(d) Find the worst-case risk of $\tau_i^{(c)}$. Among all decision rules of the form $\tau_i^{(c)}$, determine the value of c that minimizes the worst-case risk.

(e) Suppose that we have a normally distributed prior for y_i centered at 0 and with variance γ^2. Find the Bayes risk of $\tau_i^{(c)}$. Among all decision rules of the form $\tau_i^{(c)}$, determine the value of c that minimizes the Bayes risk.

Now consider the case of n individuals, each with true income y_i, noisy estimate $\hat{y}_i \sim N(y_i, 1)$, and the same target income z. The loss function is now $L(\boldsymbol{\tau}, \boldsymbol{y}) = \sum_{i=1}^{n}(z - y_i - \tau_i)^2$.

(f) Consider the decision rule $\tau_i^{(c)}(\hat{y}) = c\hat{y}_i$ for all i. Derive the risk under this decision rule for the entire population.

(g) Given $(y_1, \ldots, y_n)$, determine the value of c that minimizes the total risk across all n individuals. Compare the resulting optimal c^* to the value of c you found in part (d). How does the optimal c differ between the single-individual case and the multi-individual case?

(h) Propose an estimator $\hat{c}^*$ for the optimal value of c^* you found in (g). Compare your proposed estimator, $\tau_i^{(\hat{c}^*)}(\hat{y}) = \hat{c}^*\hat{y}_i$, with the James-Stein estimator.