

INSTRUCTIONS:

- Total points: 100, extra credit: 10 points
- Complete this final exam between Midnight December 3rd (12.01 am) and Midnight on Thursday December 9th (11.59 pm).
- A late submission (turned in after 11.59 pm on 12/9) will be graded for 50% of the final exam value. Every additional day late will lead to another 50% grading penalty. No extensions will be offered.
- For the full grade, this should be typed up and neatly formatted - working on L^AT_EX is not required, but a typed up and formatted pdf file is required.
- Send me your final code (do file or R script) and dataset that you used. Provide clear instructions on how I should run your do file / script with proper descriptions on each step that you run. This will be graded as well.
- There will be some challenging questions on this exam, but in grading I will focus on being generous with partial credit; I want to give you an opportunity to demonstrate what you know. The goal of the exam is to get you to think about how econometrics can be used as a tool for doing causal research.
- You can consult your notes or online sources during the exam.
- You can also help each other. However, *do not* share code and notes directly. You are expected to turn in your own work and I would like you to mention whom you worked with on as well.
- I will be having my usual Friday (12/3) office hours (10 am to noon) and I will be available for office hours on Monday (12/6) and Tuesday (12/7) (3 pm to 4 pm). I will answer clarifying questions only. We can meet one-on-one or I will respond to email during those dates and times *only*.

1. For the **data replication exercise**, download the data “card1993.dta” and replicate the following tables in the paper “Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling.” [Link to Paper](#). **(50 points)**

- (a) What is the main hypothesis of the paper? What is the data being used? What is the main identification that they are trying to solve?

Replicate the following parts of the paper (you can be around +/-0.01 of the final values)

- (b) Table 1 - column 3. Include the standard deviations, minimum, maximum and number of observations for each value. You don't have to replicate the regional distributions in this particular table
 - (c) Replicate table 2 and ensure that your estimates are within 0.01 of the estimates that you see. Please read the paper carefully and understand how the author treats all the variables before you run the final regression.
 - (d) Present table 2 as has been shown in the paper including all formatting - footnotes, table headers etc
 - (e) Replicate columns 1,2,3 and 4 in table 3 including all the formatting.
 - (f) Replicate rows 1, 3, 4, 5, and 7 in table 5 - including all the formatting etc.
 - (g) Send me your final code (do file or R script) and dataset that you used with clear instructions on each step.
2. In 100 words or less, **explain what issue is caused** by this problem, and what if any, solutions there are to it. **(20 points)**

- Multicollinearity
- Heteroscedasticity
- Omitted variable bias
- Missing not at random

3. **Thought experiment.** In this section, you can choose to respond to any **one** of these four research ideas. Note - you can read up literature and suggest experiments based on what you read, but make sure to include citations. Your answer should include four sections 1. Hypothesis - identify which mechanisms are at play and what you are planning to address, 2. Proposed method - DID, IV, RCT - plus an empirical estimation strategy written out clearly, 3. Proposed data collection strategy - what are the outcomes and the covariates you will measure and 4. Clear explanation on why your chosen variable provides a causal explanation. **(30 points)**

- In developed countries, there exists an obesity penalty which is associated with lower wages. Design a study to measure the impact of the penalty on labor force participation.
- Women earn less than men. Design a study that causally measures the effect of gender on wages.
- Diversity on boards impacts financial performance of a firm. Design a causal study that shows that a more diverse board room has implications on firm's financial performance
- Climate change impacts crop productivity. Design a causal study that shows that changes in climate impact yields and farm revenues.
- Note - Keep your answer to less than 3 pages. Include all citations. Provide clear explanations.

4. **Extra Credit (10 Points):** Read the following paper - Comfort, A. B., Krezanoski, P. J. (2017). The effect of price on demand for and use of bednets: evidence from a randomized experiment in Madagascar. *Health Policy and Planning*, 32(2), 178–193. <https://doi.org/10.1093/heapol/czw108>. [Link to Paper](#)

- Write a few lines on the study design - how was treatment assigned and how was control assigned.
- What were some of the dependent variables (Y) and independent variables (X) that were measured
- Provide the economical and statistical significance of the coefficient “Net Price” in Table 5
- Provide the economical and statistical significance of the coefficient “Net Price * No Fever” in Table 5
- Comment on the model fit
- Are the variables jointly significant or not
- Do you think this result is causal. Provide a few reasons on why this is the case or why it may not be the case?
- In the discussion section authors highlight some limitations of the study. Do these impact internal validity or the external validity of the study?