ECON 121 PRACTICE FINAL EXAM - FALL 2025

For people living in poor countries, migration is a leading way to find economic opportunity. You will estimate the economic return to international migration in a sample of workers from a lower middle income country. The country has an agreement with a richer country to make available a fixed number of temporary visas to work in the richer country, which it allocates through a lottery. Male workers apply to the lottery, and then the government randomly chooses some of them to receive the visas. For political reasons, the government sets the probability of receiving a visa to be proportional to the population of the applicant's subdistrict. That is to say, receiving a visa is random within a subdistrict but not between subdistricts.

The template R Markdown file loads data from a survey that researchers carried out to learn about the impact of the visa program. The researchers randomly chose localities (which are smaller than subdistricts) and then interviewed all applicants who lived in these localities before the lottery.[1] The survey included questions about geography (`subdistrict` and `locality`), the lottery outcome (`winner`), migration (`migrated`), log monthly income before and after the lottery (`ln_income_pre` and `ln_income_post`), indicators for getting married before or after the lottery (`married_pre` and `married_post`), and a number of demographic characteristics (`muslim` to `children`).

You may refer to books, internet resources (including AI), notes, and code from problem sets, solutions, and classroom examples. You may **not** communicate with any person but Prof. Vogl or Regina during the exam; evidence of such communication will result in failure. If you are taking the exam remotely, record yourself with Zoom and submit the recording on Canvas; if you have questions during the exam, you may e-mail Prof. Vogl.

Questions that require coding have a **C** next to them; you should explain your coding choices in comments, but no interpretation of the results is necessary. Questions that require words have a **W** next to them; here is where you will interpret and assess your results. Every question is worth the same number of points.

1. Estimate the effect of winning a visa on log monthly income after the lottery. Include covariates as required by the research design, but do not include covariates that are not required. Compute standard errors as required by the sampling strategy. You should continue to use the same basic covariates and standard error throughout your analysis, but you will only be graded on those choices here. **C**
2. Interpret the result from question 1. How much does winning a visa affect income? Express your answer in *percent* or *percentage points*, as appropriate. Is the estimated effect statistically significant? (2-3 sentences) **W**
3. A causal interpretation depends on whether program visas were successfully randomized. As a first check

[1] For applicants who had migrated since the lottery, the researchers obtained phone numbers from family members who still lived in the sample localities and then interviewed the applicants by phone.

on random assignment, assess whether average log income *before* the lottery differed between winners and losers. **C**

4. As a second check on random assignment, assess whether the result in question 1 is robust to the inclusion of predetermined covariates. **C**
5. Interpret your results from questions 3 and 4. Do they suggest any concerns about selection bias? (3-4 sentences) **W**
6. Using the model you estimated in question 4, test whether the income effect of winning is larger than the income gap between literate and illiterate workers. **C**
7. Suppose you are confronted by a critic who thinks the visa lottery was botched due to corruption. This critic is unwilling to believe that winners were randomly chosen but *is* willing to assume that the incomes of winners and losers would have followed parallel trends in the absence of the lottery. Run a regression that estimates the effect of winning on income under the critic's alternate assumption. **C**
8. Go back to the original research design. Estimate the effect of winning a visa on migrating abroad. Because migration is a binary dependent variable, check whether the estimated effect is similar in a linear probability model, a logit model, and a probit model. **C**
9. Interpret your results from question 8. How much does winning affect migration? Is the estimated effect statistically significant? Do the linear probability, logit, and probit models estimate similar effects? If we plan to use `winner` as an instrument for `migrated`, what do we call the regression in question 8? (4-5 sentences) **W**
10. Applying simple arithmetic to your earlier results, compute an instrumental variables (IV) estimate of the effect of migration on income. **C**
11. Now use two-stage least squares (TSLS) to obtain an IV estimate of the effect of migration on income. **C**
12. Interpret the magnitude and statistical significance of the estimated effect in question 11. Should this estimate be exactly the same as the one you computed in question 10 using simple arithmetic? Is it? Explain why or why not. (4-5 sentences) **W**
13. To help interpret the IV estimates, assess whether "never takers" and "always takers" exist in this research design. **C**
14. Interpret your results in question 13. Do "never takers" and "always takers" exist here? Given this conclusion and our IV setup, what term describes the effect estimated in question 1? (3-4 sentences) **W**
15. If the effect of migration on income is heterogeneous, is it possible for the IV estimates to represent an average effect of migrating on income? If so, what would we call this average effect, and which workers would be represented in this average? (2-3 sentences) **W**

16. Which assumptions do we need for the interpretation you described in question 15? For each assumption, explain what it means in this context and whether your earlier results contain any evidence on its validity. (5-7 sentences) **W**
17. The return to migration may be different for the rich and poor. To gain insight into this issue, generate a variable indicating which income quartile the individual was in before the lottery. Separately for winners and losers, plot *mean post-lottery log income* on the vertical axis against *pre-lottery income quartile* on the horizontal axis. Also plot *share migrating* against *pre-lottery income quartile*, again separately for winners and losers. **C**
18. Which quartile exhibits the largest return to migration? Explain the intuition based on the graphs alone. If you wanted to test for a difference in returns between the lowest and highest quartiles, could you do it by running separate TSLS regressions by quartile and then comparing the estimates? Why or why not? (4-6 sentences) **W**
19. A possible cost of migrating is that it may get in the way of life events like getting married. Use two-stage least squares to estimate the effect of migrating on the probability of getting married after the lottery. For reference, also estimate the overall rate of getting married after the lottery. **C**
20. Interpret your results from question 19. What is the effect of migration on the probability of getting married? Is the estimated effect statistically significant? How large is this effect relative to the overall marriage rate? (2-3 sentences) **W**