14.771 - Problem Set 3

MIT

You would like to develop an RCT to evaluate the impact of nutrition, as measured by caloric intake, on worker's productivity.

Exercise 1 Before running an experiment, you perform a literature review and try to understand if the question was already answered. Although you find no RCT on the subject, you discover that a series of papers have shown, using non-experimental variation, that caloric intake is positively associated with higher productivity and labour supply in developing countries. Cite at least three reasons (max. 3 lines each) why this association might *not* be causal.

Exercise 2 You decide to run an experiment where you will incentivise participants to increase their caloric intake. This experiment will be run in small scale, affecting only a negligible share of the population in the area you are studying. Your first thought is to provide individuals with a grant that can only be used to buy food. To enforce that, you randomise a subset of the participants of your experiment to receive a food stamp that allow them to purchase food at grocery stores, but not other type of good. Let Z_i^1 be an indicator of whether the participant was randomised to receive a food stamp. Answer the following questions:

1. What is the name given to the variable Z_i? (1 line)
2. What are the four classical assumptions behind IV (Imbens and Angrist, 1994) with heterogeneous treatment effects (a.k.a., LATE)? Explain each of them in words (3 lines max. per assumption).
3. Assuming your randomisation works perfectly and there is no attrition in the study. Which of the four LATE assumptions are guaranteed by random assignment *when there is no attrition*? (max. 3 lines)
4. Suppose that after running the experiment you realise that individuals in the control group were significantly more likely to refuse to participate in an endline survey where you would measure Body Mass Index and labour outcomes. When you analyse your data, you will of course need to condition on the participants who did not drop out from the study. How does this affect the assumption you pointed out on item (2) above? (max. 4 lines)

Exercise 3 The first test you would like to do is whether you have a First Stage. In this question, you will use the dataset *pset_771_3_2021_d1.csv* provided with the problem set. The instrument here is Z^1 (coded in the data set as Z_1), from Exercise 2 and there are two variables you believe could measure nutrition: (i) Food expenditure ($food\_exp$), which you can measure very precisely; and (ii) caloric intake *cal_int*, which you can measure with some noise. Answer the following questions:

1. Write down the first stage equation in this context.
2. Estimate the first stage for both variables and present your results in a two-column table. Argue how we can interpret your results as evidence of a failure of the First Stage hypothesis for caloric intake, but a success for food expenditure. (max. 5 lines)
3. You know for a fact that your measure of food expenditure is reliable (i.e, does portray an accurate description of food consumption by the participant). Explain what type of consumption behaviour could explain how food expenditure could increase without an increase in caloric intake. (max. 5 lines)

Exercise 4 Since you failed to find a first stage effect in your first pilot, you run a second pilot. Now, you try to increase caloric intake by randomising a second type of food stamp. Now, you restrict the goods that can be purchased by the participants to rice, which is the staple food in the region you are studying. Answer the following questions:

1. Using the dataset *pset_771_3_2021_d2.csv*, run the first stage with your new instrument, Z_2. Report the equation you estimated and the results in a two-column table similar to the one in Exercise 3.
2. Rationalise your results using classic consumer (price) theory. (No need to do mathematics here, words suffice.). (tip 1: You may find it useful to check Jensen and Miller (2008) to answer this question; Tip 2: there are other variables in this dataset you might want to check)[1]

Exercise 5 Let us assume now that you give financial incentives for participants to increase their caloric intake. In particular, you randomise participants into an experimental and a control condition, where the experimental group receives a small payment if they increase their caloric intake. Assume you can still measure the participants' caloric intake. Using the dataset *pset_771_3_2021_d3.csv*, where work productivity is measured by the variable *prod* and the instrument is measured by Z_3, please answer the following questions.

1. What is the intent-to-treatment (ITT, sometimes called reduced form) equation you would like

[1] Jensen, R.T. and Miller, N.H., 2008. Giffen behavior and subsistence consumption. American economic review, 98(4), pp.1553-77.

to estimate?

2. Please estimate the intent-to-treat equation and present and interpret the results.
3. Assuming you trust the randomisation, can you conclude that the intent-to-treat association is causal?
4. Estimate equation 1 by Two-Stage Least Squares (2SLS), instrumenting cal_int_i with Z_3, and present your result.

$$prod_i = \beta_0 + \beta_1 cal\_int_i + \varepsilon_i \tag{1}$$

5. What is the relationship between $\hat{\beta}_1^{2SLS}$, the 2SLS estimator of equation 1, and $\hat{\alpha}^{ITT}$ and $\hat{\alpha}^{FS}$, the ITT and the First Stage estimators, respectively?
6. You are very enthusiastic about your findings, and so you present them in Dev/PE Lunch. An audience member is concerned that your instrument may be weak. To investigate this concern, please do the following:
   (a) Run the first stage and report the *F*-statistic. What can you conclude from this?
   (b) Produce an Anderson-Rubin confidence interval. This can be implemented with *weakiv* in Stata. Report the results. Is the confidence interval wider than the one you estimated above?
   (c) Produce a confidence interval using the adjustment factor in Lee et al (2021).[2]
   (d) Discuss briefly whether your 2SLS estimate is robust to weak IV.

   Hint: Look at Table 3a of Lee et al (2021). Find the F-statistic from your first stage in the top row (feel free to interpolate), and then take the corresponding adjustment factor from the second row. Then, compute a confidence interval as $\hat{\beta}^{2SLS} \pm 1.96 \cdot \widehat{SE}(\hat{\beta}) \cdot r(F)$*, where* $r(F) \geq 1$ *is the adjustment factor from the table. You might notice that the adjustment factor gets smaller as F gets larger, with* $r(F) = 1$ *for* $F = 104.7$*.*

Exercise 6 You are asked to estimate the calorie-productivity intake curve. This curve plots worker's average productivity by caloric intake level.

1. Argue why you cannot estimate this curve, except perhaps on two points, with the instrument from Exercise 5, even if you believe it to be a valid instrument (Max. 5 lines).
2. Can you propose a modification to your design in order to estimate this curve for more points? (tip: how could you modify the instrument to achieve this goal?)

[2] Lee, D.S., McCrary, J., Moreira, M.J. and Porter, J.R., 2021. Valid t-ratio Inference for IV (No. w29124). NBER.

14.771: Development Economics

Fall 2021