

Problem Set 3

ECO3121 - Fall 2023

November 10, 2023

Due 11:59PM, 20/11/2023

Please remember to submit your Stata code and requested output as it will be graded.

In this problem set, we'll continue to examine the effects of land rental activity in rural households on output and aggregate productivity. Please continue using the main dataset "`aghousehold.dta`" from your first assignment from the blackboard site. It is the National Fixed Point Survey (NPFS) in the year of 2010.

We continue with our inquiry on the causal effect of land rental behavior on agricultural productivity. The researcher plans to follow the specifications in Question 1-7 in Assignment 1 and regress the yield on land_rental-out activities..

Model 1:

$$\text{logyield}_i = \beta_0 + \beta_1 \text{rental_out_share}_i + \mu_i$$

Question 1

1. List three plausible arguments why the point estimate in the regression model 1 could be biased, and the corresponding bias directions (**upwards** or **downwards**) relative to the true causal effect of household's rental activities on household yield. (6 points)
2. Now your friend suggests that the total land area (`d31` in the data) of each household could be a valid instrumental variable (IV) for your measure of household's land rental-out share `rental_out_sharei` (the proportion of rent-out land to total land area (`d31`)). List the assumptions that need to hold true for this to be correct and explain why this instrumental variable is true (or not true). (4 points)
3. Now your TA suggests that the rainfall (precipitation) could be a valid instrumental variable (IV) for your measure of household's land rental-out share `rental_out_sharei`. Try to merge the household's production dataset and precipitation dataset via `vl_id`, the specific village identifier, using stata command `merge` (Many-to-one merge, type "`help merge`" in Stata for assistance).

Note: drop the missing value for $rental_out_share_i$, av_rain and $logyield_i$ First!

You decide to use the average rainfall in 2010 (av_rain) as the IV for the household's rental-out land share. Verify if the assumption of instrument relevance is (or is not) satisfied using the first stage regression, and export the results. **Write down the first stage regression model** and interpret the result and statistical significance of your result. (5 points)

4. Now use Stata/R to estimate the 2nd stage IV point estimate (using linear probability model) as suggested by the TA, and export your result. **Write down the second stage regression model** and interpret the result and statistical significance of your result. (5 points)
5. Now your TA tells you that you can use *ivreg* command directly to replicate the results in question (4). Do you find any difference in the IV estimations (β_{IV}) **regarding the coefficients and standard errors** relative to (4). (3 points)
6. Now your Professor suggests that both rainfall (Z_1) and the implementation of land law (Z_2) which legalizes the land transaction could be valid instrumental variables (IVs) for your measure of household's land rental-out share $rental_out_share_i$. Try to merge the household's production dataset and land law dataset named "land-law.dta". To do the merge, you should take the first two digit number of vl_id in the aghousehold data by using the command "substr" and then generate a new variable called vl_id2 . Next, merge household data and landlaw data via vl_id2 , the specific province identifier, using stata command *merge* (Many-to-one merge, type "*help merge*" in Stata for assistance).

you can use *ivreg* command directly to get the estimations. (15 points)

- 1) Propose the tests for the relevance assumption and exogenous assumption under this context. **Write down the regression model** and each step of the two tests.
- 2) Conduct the tests using our data.