

# Problem Set 1

## ECO3121 - Fall 2023

September 25, 2023

**Due 5PM, 08/10/2023**

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

### Question 1

In this and later problem sets, we'll try to replicate a research paper titled "Property Rights, Land Misallocation and Agricultural Efficiency in China" (Chari, Liu, Wang, and Wang, 2021) step by step. The paper examines the impact of a property rights reform in rural China that allowed farmers to lease out their land. They examine the effects of land rental activity in rural households on output and aggregate productivity.

To replicate this paper, please download main dataset "`aghousehold.dta`" from the blackboard site and load into STATA. The main data we use is the National Fixed Point Survey (NPFPS), which is a nationally representative panel dataset (unbalanced) of roughly 20,000 households in 360 villages between 1986 and 2013. It is collected by the Ministry of Agriculture and Rural Affairs of China. Since we are learning simple linear regression and multiple linear regression at the current stage, we provide a cross sectional dataset this time which only includes households in the year of 2010.

It will be a univariate model, we'll be estimating regressions of the form:

(1)

$$yield_i = \alpha_1 + \beta_1 rental\_in_i + \mu_i$$

$$yield_i = \alpha_2 + \beta_2 rental\_out_i + \mu_i$$

First, let's analyze the effect of land rental activity (rent in and rent out) on agricultural yield.

1. You can generate variable `yield` (output per unit of land) via  $\frac{d32}{d31}$ , and variable `rent in` through variable `c10` and `rent out` via variable `c13`. Visualize and export a table that lists the number of observations, the mean, the standard deviation, the minimum value and the maximum value for these three variables in the dataset (edit and include the table in your written up answer). Briefly describe what we learn from the table about these households. (5 points)

2. What sign do you predict  $\beta_1$  and  $\beta_2$  will have? Why? (3 points)
3. Run the regression and report the estimated coefficients, their standard errors, and  $R^2$ . Does the estimated value of  $\beta_1$  and  $\beta_2$  agree with your predictions? (3 point)
4. Interpret the value of the estimated parameters  $\hat{\beta}_1$  and  $\hat{\beta}_2$  from the models. (3 points)
5. Compute the fitted value  $\widehat{yield_i}$  and the residual  $\hat{\mu}_i$  for each observation for both regressions, and verify that the residuals (approximately) sum to 0. (2 points)
6. How much of the variation in yields for these farmers is explained by their rent-in and rent-out activities? (2 points)
7. A professor asks you to create two new variables which measure the proportion of rent-in and rent-out land to total land area (*d31*) (**remember multiply these fractions by 100**). She also asks you to use these two measures as independent variables to re-run the two regressions,

(2)

$$yield_i = \alpha_1 + \beta_1 rental\_in\_share_i + \mu_i$$

$$yield_i = \alpha_2 + \beta_2 rental\_out\_share_i + \mu_i$$

Please report and interpret the new estimated parameters  $\hat{\beta}_1$  and  $\hat{\beta}_2$  from the models. (4 points)

8. The professor sees your analysis. She asks you to explore alternative functional forms for this relationship. Using the provided data, you estimate the following equations:

\* A log-linear relationship:

(3)

$$\log(yield_i) = \alpha_1 + \beta_1 rental\_in\_share_i + \mu_i$$

$$\log(yield_i) = \alpha_2 + \beta_2 rental\_out\_share_i + \mu_i$$

We take natural log here. In Stata, the coding is "gen newvar == log(variable)"

Run the regressions recommended by the professor. Interpret the parameters  $\hat{\beta}_1$  and  $\hat{\beta}_2$  in each of these equations. (4 points)

9. Compare model (2) in q7 and model (8) in q8, which do you prefer? And why? (2 points)
10. What are the SLR4 assumptions under this context. Propose two reasons that might cause biased estimation of  $\beta_1$  and  $\beta_2$  (4 points)
11. Think about two mechanisms that why land rental activities can increase yields or lower yields (depending on the sign of your  $\beta_1$  and  $\beta_2$ ) (4 points)

## Question 2

**[Potential Outcome Framework]** Background knowledge: the land rental contract law grants farmers the legal right to rent out and rent in land, outlining rules for leasing, transferring leases, and how to address land leasing disputes. Prior to this law reform, land rental activities are very rare since they were not protected by a formal law.

Suppose the Ministry of Agriculture and Rural Affairs has hired you as part of their impact evaluation team. Your first assignment is to evaluate a Randomized Control Trial (randomized experiment) that they have implemented before you arrived. Two years before you arrive, they have implemented a land rental contract law to **100 rural villages**, which they randomly selected from a list of **200 villages** that all had expressed their interest in participating in the project.

The board asks you to estimate the Average Treatment Effect of their intervention (land rental contract law) on **village-level** yields.

1. What data (which variables) do you require from their project team to answer this question? (2 points)
2. Which regression equation would you want to use these data for? Describe each variable and subscripts that indexes basic research entity and explain the interpretation of the regression coefficients. (3 points)
3. After you have started working on this, the former leader of their project implementation team tells you that she is concerned that the program may not actually have randomly allocated treatments across the 200 villages, and that some selection may have gone on (she heard reports that the richest villages were more likely to be put into the treatment group). What concern would this bring to your estimation? (3 points)