

Problem Set 2: Macroeconomics and Inequality in Uganda.

Albert Rodriguez-Sala and Jacob Adenbaum
Programming for Economics—ECNM10106
The University of Edinburgh

2023

For this problem set we will use the [2011-2012 of the Ugandan National Panel Survey](#). The survey is under the umbrella of the ISA-LSMS surveys and offers nationally representative household data on consumption, income, wealth and other key variables in Uganda. The file *UNPS_1112_PS2.xls* contains the data while the file *variables_description_UNPS_1112_PS2.xls* contains a description of the variables. Import the data and answer the following exercises.

Exercise 1. Exploring the data (25 points).

- Are there duplicate households in the data? That is check if there are repeated observations in the unique household identifier variable ('hhid'). How many observations are there in the data?
- Present some basic summary statistics for the following variables: *head_gender*, *head_age*, *familysize*, *consumption*, *income*, *wealth*. Comment your results in 2 lines. In particular, you might mention if there are missing observations or potential outliers for some of the variables.
- Using the *head_gender* variable, create a dummy variable for household head being female (1=female, 0=male). Be careful with missing values. What is the proportion of households where the head is female?
- Using the *groupby* method, compute the average consumption, average household size, and average household head age for households where the head is male vs where the head is female. Do we observe noticeable differences across the two groups?

Exercise 2. Inequality in Uganda (50 points).

- Create the variables *log_c*, *log_inc*, *log_w* that are the log of consumption, income, and wealth, respectively. Plot in the same graph the distribution of the log of consumption and the log of income. Do the distributions resemble some known distribution? Is inequality higher in consumption or in income?
- A commonly used statistic to measure inequality is the variance of the logs. Compute the variance of the log of consumption, of the log of income, and of the log of wealth. How do these measures of inequality in Uganda compare to the same measures of inequality in the United States? Use table 3, column 5–PSID in [De Magalhães, L., & Santaaulàlia-Llopis, R. \(2018\)](#) for the comparison.
- Measuring between rural and urban inequality in Uganda. Compute the average consumption, income, and wealth for rural and urban areas separately (*groupby*). Are the differences between the two areas large?
- Measuring within rural and urban inequality in Uganda. Compute the variance of the log of consumption, income, and wealth for rural and urban areas separately (*groupby*).
- Compute the Gini coefficient in consumption, in income, and in wealth in Uganda. Compare these values with the Gini coefficients in the United States—table 3, column 5–PSID in [De Magalhães, L., & Santaaulàlia-Llopis, R. \(2018\)](#)

- f. Compute the share of the wealth that the bottom 50 percent hold. Compute the share of the wealth that the top 10, 5, and 1 percent hold.
- g. Although in the last years, there has been a big debate on inequality, the debate has mostly focused on rich countries. From your results of this exercise, is inequality large in Uganda? Should we also be concerned about inequality in low-income countries such as Uganda?
- h. The few previous studies on income inequality in Africa have mostly used consumption measures to estimate income inequality. See, for example, [Alvaredo & Gasparini \(2005\)](#). From your results of this exercise, what do you think about these estimates? Are they overestimating or underestimating income inequality in Africa?

Exercise 3. The life-cycle of male vs female head households in Uganda (25 points).

Before going to the plots, you might want to drop ages for where there are few households. You might also want to group the ages in bins so that the plots are more smooth. The following code does these two steps.

```
# Drop extreme values (too few observations to get means within age)
data = data[data['head_age'] < 80]
data = data[data['head_age'] > 18]

#To smooth the lines, group the ages in bins
bins = [18, 30, 40, 50, 60, 80]
labels = [25, 35, 45, 55, 70]
data['age_bins'] = pd.cut(data['head_age'], bins=bins, labels=labels)
```

Then, using seaborn lineplot with the argument *hue='female'*, or any other variable to distinguish the gender of the head,

- a. Plot the lifecycle of the log of consumption for households where the head is male and for households where the head is female.
- b. Plot the lifecycle of the log of income for households where the head is male and for households where the head is female.
- c. Plot the lifecycle of the log of wealth for households where the head is male and for households where the head is female.
- d. What are the differences in the lifecycle of households where the head is a male versus the head is a female?