Yale JACKSON SCHOOL OF GLOBAL AFFAIRS

Assignment 2: Firm Size Distributions in Ethiopia

GLBL 6595 Development Policy in the 21st Century

Due Friday, March 8

This assignment asks you to analyze the distribution of firms in Ethiopia. The file pulls together two data sources in 2002: Ethiopia's Large and Medium Scale Manufacturing Survey (sample==1) and Ethiopia's Small Scale Manufacturing Survey (sample==2). The dataset provides the firm-level information: total employment, total sales, total assets, age, and others. You can download the file in Assignment folder on canvas → Assignment 2 Firm Size folder).

Submit one pdf file per group to gradescope with graphs and answers to the questions. Gradescope will ask you to identify your groupmates. Append your code at the end of the write-up (for Stata code, change the font to Courier New font size 9 to make it readable).

Notes:

- The sales and assets variable are in Ethiopian Birr. Assume an exchange rate ETB 0.12 to USD 1

Please answer the following questions:

1. Present a way to describe the distribution of firm employment. Why would it be misleading to paint a picture of the true firm-size distribution in Ethiopia's manufacturing sector using Sample 1 only?
   - For this question, it may be useful to top code the employment, for example, at 100, when presenting the distribution.

2. Contrast the employment size distribution in (1) with the distribution of sales. List a few reasons why they may be different.

3. Present a way to examine the relationship between employment size (y-axis) and firm age (x-axis). What message do you infer from this relationship?

4. Assume that a firm *i*'s production function is given by: $y_i = A_i k_i^\alpha l_i^{1-\alpha}$, where $\alpha = 0.3$, *y* denotes sales, k denotes capital, and l denotes labor.
   a. Provide the expression for a firm's value marginal product of capita in terms of the firm's sales, α, and its capital stock.
   b. Remove the firms in the top 5% of VMPK values. The data are often noisy, and we don't want outliers to skew our conclusions. Report the mean and standard deviation of log(VMPK) after removing these outliers.
      - One way to remove outliers is to use the *winsor2* command in Stata.

c. What is the relationship between log(VMPK) and firm size, as measured by employment? If we believe that all firms produce using the production function above, what message can you draw from this relationship?

d. The variables "prob_major1", "prob_major2", "prob_major3" report the three major problems faced by firms. Construct a variable that is equal to one if a firm reports "working capital shortage" (the code is 5) as any of the responses to these three variables; otherwise, it takes a value of zero. Run a regression that compares the log(VMPK) between firms that report working capital shortages and those that do not, while controlling for firm size. Provide an interpretation of the finding.

e. Within Sample 2, there is a variable called "reason_not_solve_cap_prob" which takes on 6 possible values for why firms may be unable to solve their working capital problems. It's only been answered by 167 firms, but let's take the answers at face value and assume firms are telling the truth. Re-do the analysis in (d) comparing the log(VMPK) of firms controlling for firm size and the different responses to this question. What message emerges from the regression? (Hint, if you write "i.reason_not_solve_cap_prob", Stata will automatically create the dummy variables for you and drop one ("insufficient loan").