

# Università degli Studi di Milano

## Global Firms and Markets – Development Module

### STATA Assignment

Academic Year 2022-2023

Our training sessions on STATA (class n.1 to n.5) were dedicated to introducing some of the basics you will need in virtually any analysis involving the use of data. With that toolbox we then replicated some of the results presented in a working paper by Ashraf et al. (2019). Now it's your turn!

Use the dataset provided “panel\_data\_avg\_ADG” to obtain the output according to the instructions below. You will have to provide essentially two files:

- a) .do file with the code you created;
- b) PDF file collecting your results (graphs, tables, regression output etc.).

As discussed in our classes, the .do file should allow the end user to replicate your results simply by changing the directory string to match that of the local computer. Furthermore, the output presented (tables, graphs and regression output) in the PDF file must be complemented by a few lines of intuition and your original interpretation where applicable. These two aspects are also determinants in the evaluation of your work.

#### **Task 1 – Getting started with your data**

Create your .do file and open the dataset from your local directory.

Open a .txt log file so that you can easily store your output in a dedicated text file. It will make easier for you to extract the results you want to present. Don't forget to close the log at the end of your exercise. The log file is for your internal use only and is not to be handed in as part of the assignment.

Inspect your dataset and get a list of the countries included in the dataset. How many observations do you have?

Report your answers in the PDF file.

## **Task 2 – Refreshing your memory on descriptive statistics**

Obtain a minimum set of descriptive statistics for the share of firms with an absolute majority of female proprietorship across countries using the most recent observations available. From what you see, what is the mean and the standard deviation of `wb_majFemOwned_mr`?

Now repeat the same procedure for the discrimination in the family index using the most recent year of observation available for each country in the dataset.

Report your answers in the PDF file.

## **Task 3 – Correlation Matrices**

Obtain a cross-sectional correlation matrix for all the available variables of discrimination in the family and restricted physical integrity (i.e. first year obs., average across years of obs. and most recent obs.). Now focus on the most recent observations for the two variables. What is the correlation between them? What interpretation can you provide for this result? Bear in mind the basics of causality.

Report the output and your comments in the PDF file.

## **Task 4 – Bar Graphs**

Replicate the bar graph depicted in Figure 2 in Ashraf et al (2019) pag. 48. Sort the countries in ascending order as we did during our tutorials.

Now repeat the exercise plotting the WBI rule of law index (most recent observation) over the countries. What is the median value of `rulelaw_mr`?

Report the output in the PDF file, choose style and colours you prefer.

## **Task 5 – Scatter Plots**

Replicate the scatter plot reported in Figure 4 in Ashraf et al. (2019) pag. 50. Use the most recent observations for female ownership (absolute) and discrimination in the family for each country. Include a linear prediction, then try with a quadratic fit.

Report the output and your comments both for the linear and the quadratic fit in the PDF file, choose style and colours you prefer.

Now obtain an estimate of the regression line you depicted with the linear prediction. What is the constant term? Is the coefficient associated with discrimination statistically significant?

Report the output of the regression in the PDF file.

Now repeat the exercise using the most recent observations for each country for female ownership (absolute) `wb_majFemOwned_mr` and overall WJP score (most recent) `wjp_index_new_mr`. Include a linear prediction line and complete the figure with a plot of the 95% confidence interval around the regression line. What can you say regarding its statistical significance? Provide an interpretation for the relation you just illustrated.

Report the output and your comments in the PDF file, choose style and colours you prefer.

### **Task 6 – Multiple Entries Table**

Create a multiple entries table in the style of Table 1(a) in Ashraf et al. (2019) pag. 34 where rule of law `rulelaw_mr` sets the rows and the index for restricted physical integrity `Restrictedphysintegrity_mr` the columns. In each intersection you shall include the frequency and the percentage of countries that are *above or below the median* values for the two variables you are using.

Remember, to get there you need to create two categorical variables from the original ones. Do not replace the values, create two new variables instead and label the values you obtained as shown in Table 1(a) (i.e. High/Low).

Report your output in the PDF file. Choose the solution you prefer to export your results from STATA (.txt, .xls etc.).

### **Task 7 – Creating a percentile variable**

Create a new variable using the appropriate commands in order to have for each country its percentile positioning with respect to `majFemOwned_manuf_mr`. (*Hint: rank your observations in ascending order with respect to the relevant variable and then scale everything back to a hundred*).

What are the percentile positions of Italy and Zambia? Report your answer in the PDF file.

### **Task 8 – Regression Analysis**

Replicate the regression analysis described in columns (2), (4) and (5) in Table 2 in Ashraf et al. (2019). Comment on the results you obtained and provide an interpretation.

Report the output of the regressions along with your comments in the PDF file.