

Stata Class 1

Alexander C. Lembcke

Exercise 1

You should start a do-file where you collect all the commands, so you can replicate the analysis at a later date. The data we use comes from the paper *Americans do I.T better: US Multinationals and the Productivity Miracle* by Nick Bloom, Rafaella Sadun and John van Reenen, forthcoming American Economic Review.

- a) Open the data set "replicate.dta".
- b) Use the "describe" command to find out how many observations are in the data set and what the name of the variable containing information on "people management" scores is.
- c) Find the mean value of the people management scores
- d) Use the "tabulate" command to find out what countries are in our sample, what years and how many observations there are for each year and country.
- e) Save the do-file

Exercise 2

Keep using the same do-file and add the commands to reproduce the following exercises.

- f) What are the mean, the standard deviation and the number of observations of the level of employment in UK firms. Split this statistic by US Multinationals, Other Multinationals and UK domestic establishments, replicating the first column of Table 1 in the paper.
- g) Create the correlation matrix of all management scores separately for each country.
- h) Find the mean value of management scores for each country and year.
- i) Create a horizontal bar graph that depicts the average people management score in each country, replicating Figure 3a in the paper.
- j) Create the same graph again, this time only for those firms that are US multinational subsidiaries.
- k) Use the "graph export" command to save the plot.

Exercise 3

Keep using the same do-file and add the commands to reproduce the following exercises.

- l) Generate a variable that is equal to the total number of hours worked in a firm.
- m) List the first 10 observations to see whether your new variable is correctly defined.
- n) Create a dummy variable, i.e. a variable that takes the values 0 and 1, that is 1 if a firm has at least 1 employee that is a union member and 0 otherwise. Hint use generate and replace in combination with "if".
- o) Create a variable that is the sum of all individual management scores. Compare it to the existing variable that combines the management scores. Why do they differ? Account for the reason for the difference and check again whether scores differ.
- p) Save the current data set and import the file "additional.csv". Save the file in Stata format and reopen the original data set. Use the "merge" command to add the additional variables. Save the new data set.

Exercise 4

Keep using the same do-file and add the commands to reproduce the following exercises.

- q) Start a log-file at the beginning of your do-file, make sure you close it at the end of your do-file.
- r) Run a regression of log sales on people management scores.
- s) Predict the fitted value and create a graph where you combine the scatterplot of log sales and people management scores with the fitted line.
- t) Run the same regression again, this time constraining the sample to establishments that are i) UK domestic, ii) US multinationals, iii) Other multinationals. Plot the three predicted values inside the scatter plot.
- u) Create dummy variable interactions with the people management score and run the regression of log sales on the five appropriate variables to replicate the results from the 3 separate equations. Test whether the people management score has the same impact on log sales among the three ownership types.