Exploring Economic Development
Instructions for Lab 3
ECON582 Economic Development

1 Overview

1.1 Lab Objective

The purpose of the lab is to help you think about what you have learned in lecture about productive inputs and also explore some univariate and bivariate graphical and numerical summaries in the context of economic variables. The lab will also introduce you to some first-level statistics in the context of regressions and tables.

1.2 Lab Procedures

In this lab, you will have access to data on economic variables from the World Bank (command: **wbopendata, language(en - English) topics(3 - Economy & Growth)**). The data allows for an analysis of various economic variables which are predictive of performance and level of income.

Open the data file titled "WB Economy Panel" located on Blackboard. This data looks at the composition of the economy (services, manufacturing, trade, etc.) and explores how that composition varies by level of income and geography. In addition to the composition of the economy, this data provides insights into some of the crucial development topics such as debt service and overseas development aid.

Data Analysis Tip: When constructing datasets, one often has to perform operations on variables. This will change the label of the variable sometimes making the label uninformative. Stata allows you to easily change the label of a variable so as to make it useful to understand the meaning of the variable: **label variable varname ["label"].**

2 Helpful Commands

Data Analysis Tip: It is often helpful to take variables with many values and discretize them. One example of this is to take the year variable and discretize year into decades. The command for this would be **gen decade=0** *and then replace the values of decade with 1 for 1960, 2 for 1970, and so forth:* **replace decade = 1 if year>=1960 & year<1970**.

Recall that it is possible to put multiple plots on a single graph. For instance, if you want to graph net overseas development aid as a scatter plot and also as a line, you would use the command: **graph twoway (scatter** *net_oda* **year) (lfit** *net_oda* **year)**.

What happens if you use line instead of lfit? What does lfit do?

If you notice, there does not seem to be a pattern for *net_oda* over time. Another helpful command is the "if" command, which allows you to take only a segment of the data. For instance, you could use **graph twoway (scatter net_oda year if regionname=="Sub-Saharan Africa") (lfit net_oda year if regionname=="Sub-Saharan Africa")**. Do you see a pattern now?

Another way to display this data would be to look at the boxplot over the variable "decade" created above. **graph box** *net_oda* **if regionname=="Sub-Saharan Africa", over(decade)**.

3 Questions

1. This question begins the exploratory analysis of the data.
   (a) Try looking at the pattern of *net_oda* by other regions. What do you notice?

   Data Analysis Tip: If you want to get some quick summary statistics to compare, it might be useful to try the **by [varname]: summarize [varname]** *and* **graph box [varname], over(varname) commands**.

   (b) In any analysis, it is important to check whether the means and SDs are strongly influenced by individual data points. For Stata's box plot, the box contains observations in the Inter-Quartile Range (IQR), aka the 25th and 75th percentiles. The lines extend an additional 1.5*IQR. Look at the debt service variable by region. Which regions are particularly influenced by outliers?

2. A comparison of means and standard deviations might be inadequate. For example, suppose one group has a right-skewed distribution, and the other group has a left-skewed distribution. Just reporting means and standard deviations does not inform the reader about such structure. Compare the distributions of *net_oda* of Sub-Saharan Africa and South Asia. Describe any differences between the two groups' distributions: compare locations of most of the data, the spreads of the distributions, and whether there are outliers. Reminder: histograms are useful for looking at distributions and you can name the graphs to combine them for a side-by-side comparison. You can also use the boxplot for side-by-side comparisons.

3. We chose to discretize year into decade earlier. What other variable might we choose to look at discretely? Try creating a discrete categorical variable from one of the continuous variables in the dataset.

4. Use the discrete variable as an explanatory variable to see if it has any predictive power on a dependent variable of interest (**reg dependent i.independent**). The i.independent means you would do something like i.decade. You can then use the **margins independent** command followed by the marginsplot command to get a neat visualization for the impact of increasing your discrete variable.

   Data Analysis Tip: When looking at an OLS regression (which is the result of the reg command in Stata), the coefficient represents the slope of a fitted line. This gives the researcher the effect of (in the binary case) going from the untreated group to the treated group. The p-value tells us whether the results are significant. If the p-value is 0.10 or less, then we have significant results.

   Data Analysis Tip: Researchers sometimes categorize continuous variables to simplify analyses. However, when there are strong linear relationships, categorization sacrifices information and can lead to inaccurate results. When categorizing, be sure to have a valid scientific rationale for choosing the end points of the categories.

5. Explore the behavior of services, manufacturing, and trade over time. What does this look like by region? What does the behavior of trade over time indicate about global integration of particular regions?

6. Sometimes we want to take advantage of the panel structure of the dataset. As such, it is useful to have a numeric variable that represents the country name. You can use the **egen id=group(countryname)** command to create this variable. Then, use **xtset id year** to get a balanced panel with country-year as the observation of interest. Now you can run panel regressions to examine variable relations (rather than reg, you can use xtreg).

7. An additional benefit of the xtset command is that Stata now understands time commands. This means you can lag your independent variables to see the impact of something like saving last year on capital formation this year. Use the lagged command (*l.gross_saving* for instance) in your regression to test the relationship between savings and capital (full command: **xtreg capital l.gross_saving**).

8. Sometimes, it is interesting to look at the log of a variable. Try taking the log of some of the larger variables like consumption and look at the distribution before and after the log (command **gen *l*_consumption=log(consumption)**.