Exploring Technology
Instructions for Lab 4
ECON582 Economic Development

1 Overview

In this lab, we explore some of the data available at the World Bank on technology. This includes looking at the following variables:

- TX.VAL.TECH.MF.ZS (high tech exports as a percent of manufactured exports)
- SP.POP.TOTL (total population)
- TX.VAL.TECH.CD (high tech exports in current US dollars)
- SP.POP.SCIE.RD.P6 (researchers in R&D)
- GB.XPD.RSDV.GD.ZS (R&D as a percent of GDP)
- IP.TMK.TOTL (total patent applications)
- IP.TMK.RESD (patent applications by residents)
- IP.TMK.NRES (patent applications by non-residents)

1.1 Lab Procedures

Data Setup
If you have a Stata license on your computer, you can use the wbopendata command in the following manner to quickly obtain the panel dataset you want.

```
wbopendata, indicator(TX.VAL.TECH.MF.ZS; SP.POP.TOTL;
TX.VAL.TECH.CD; SP.POP.SCIE.RD.P6; GB.XPD.RSDV.GD.ZS;
IP.TMK.TOTL; IP.TMK.RESD; IP.TMK.NRES) long clear
```

You should then rename the variables appropriately. Alternatively, you may download the data from Blackboard. If you have Stata from the virtual lab, you will need to download the data available on Blackboard called Stata Lab 4 Data.

In the following sections, you will explore some aspects of research and development across the world. These empirical analyses complement the theory that we learned in class. In order to fully understand this analysis, please review your notes on regressions.

2 Helpful Commands

2.0.1 Regressions

- Recall that we can do various kinds of regressions in Stata. Given the nature of World Bank data, we want to do panel regressions where we set the cross-section and time variables (country and year specifically).The code to set the panel is **xtset country year** but first we need to create the country variable.

- To do this, we will need to generate a numeric cross-sectional variable. The code to do this is **egen country = group(countrycode)**. Alternatively, you can use the **encode** command in order to keep the 3-digit country codes. This would look like **encode countrycode, gen(country)**.

- We can run a regression using the **xtreg** command. This allows you to think about the impact of an independent variable across time and country. For instance, if you want to look at how countries vary in R&D spending by income level, you can perform the following analysis:

  1. Generate a numeric variable for income level.
     (a) **gen income = 0**
     (b) **replace income = 1 if incomelevel==“LIC”**
     (c) **replace income = 2 if incomelevel==“LMC”**
     (d) **replace income = 3 if incomelevel==“UMC”**
     (e) **replace income = 4 if incomelevel==“HIC”**
  2. **xtreg** *rd_per* **i.id**
  3. **margins i.income**
  4. **marginsplot**

  Note: you can use the graphing options with marginsplot. Put a comma after the marginsplot command and you are able to use the xtitle(“text”), ytitle(“text”), and title(“text”) commands.

Data Analysis Tip: when we look at the margins plot, it appears that the coefficients are different from each other. We should actually run a test. For instance, if you want to test that the coefficient on low income countries is different than the coefficient on high income countries, you can type **test 1.income = 4.income***. This gives a chi-squared test with the null hypothesis that the difference in coefficients is zero.*

- **Twoway fixed effects**: this controls for time trends and country trends. Time trends are things that happen over time to all countries and country trends are things that happen within a country over time. To run a twoway fixed effects model in Stata, you should use **xtreg dependent independent i.year, fe**.

2.0.2 Graphs

- Recall that there are several helpful ways to visualize data including boxplots, bar graphs, scatterplots, and line graphs.
  (a) Boxplot: **graph box continuous varname, over(discrete varname)**
  (b) Bar: **graph bar continuous varname, over(discrete varname)**
  (c) Scatterplots: **scatter varlist**
  (d) Line graphs: **line varlist**

2.1 A Little Econometrics

When dealing with panel data, it is good to think about potential confounding factors that drive correlations. It may be the case that you have country effects, where there is something occurring within a country and across time. It may also be the case that you have time effects where some trend is occurring across time over all countries. These trends are dealt with using "twoway fixed effects".

$Y_{it} = \beta X_{it} + \alpha_i + \gamma_t + \epsilon_{it}$ where α are the country fixed effects, γ are the time fixed effects, and ϵ is the error term. It is important to note that if you have a variable that does not change over time or by country, it will be omitted from your regression in a twoway fixed effects model.

3 Questions

1. Create a boxplot of the R&D percent of GDP by region and a boxplot by income level. What pattern do you see emerge? Represent this information with a bar graph as well.

2. Create a margins plot for income using the margins command to see the impact income has on tech exports as a percent of manufactured exports. What can you see in this plot?

3. Is there a correlation between the number of patents and the level of tech exports? What graph best shows this? How can you adjust the titles on the x-axis and y-axis?

4. Look at tech exports as a percent of manufactured exports using boxplots by region and by income. What kind of patterns do you see?

5. Regress tech exports as a percent of manufactured exports on the logged number of patents (create a logged variable for patents if you have not already) and use robust standard errors (put comma r after the panel regression). Now compare this with a twoway fixed effects regression. What happens to the significance and the standard errors?