AREC 345: Global Poverty and Economic Development

Problem Set 3

Department of Agricultural and Resource Economics
University of Maryland
Fall 2016

Problem Set 3 is due at the start of section on October 7. Problem sets turned in more than 5 minutes after the start of section will be marked as late. All problem sets must be turned in as hard copies; points will be deducted if multiple pages are not stapled together.

This problem set uses the data set `arec345ps3data.xlsx`. The data are from Nathan Nunn's (2010) paper "Shackled to the Past." The data set is also available on his web site, though in a different format. The data set includes information about 52 countries in Africa, including small island nations.

Several of the questions ask you to run a series of linear (OLS) regressions using the `LINEST` command in Excel. The `LINEST` command estimates the best linear fit for a given data set, allowing us to explore the relationships between the independent and dependent variables. To run the `LINEST` command, start with an Excel worksheet in which your dependent variable is in the first column and your independent variables are just to the right of the dependent variable. For example, if you wanted to estimate the linear regression equation

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + e$$

to explore the relationship between independent variables x_1 and x_2 and dependent variable y, you'd have the variable y in Column A, x_1 in Column B, and x_2 in Column C. The data set we're using contains 52 observations, so the dependent variable, for example, might be contained in cells `A2:A53`. In this example, you are estimating three parameters: a (the constant), b_1, and b_2. To use the `LINEST` command, highlight a block of cells three cells across and two down. Then type the following:

```
=LINEST(A2:A53,B2:B53:C2:C53,TRUE,TRUE)
```

and hit `CONTROL-SHIFT-ENTER`. Your block of six cells now contains the following **in this order**: the coefficient estimate of b_2 with the estimated standard error below it (in the first column of the highlighted block of cells), the coefficient estimate of b_1 with the estimated standard error below it (in the second column), and finally the estimated constant and associated standard error.

Please see the TA for assistance if you are having trouble using Excel to run the regressions. You are welcome to use Stata instead of Excel if you are already familiar with it.

1. The variable `lnexportslaves` is the natural log of Nunn's estimate of the total number of slaves exported from each country, normalized by that country's total land area. Use the `LINEST` command in Excel to run a linear regression of log GDP per capita in 2000 (`lngdp2000`) on log slave exports. Cut and paste the Excel results into your problem set write-up. Interpret the coefficients. Can we reject the hypothesis that the slave trade is unrelated to GDP per capita? At what level of confidence?

2. Re-run the regression above, this time including a control (i.e. additional variable) for whether a country is landlocked (`landlocked`). How does the coefficient in this regression compare to the previous estimate?

3. Calculate the difference in the average level of `lngdp2000` in non-landlocked vs. landlocked countries. Generate a new variable that is equal to `lngdp2000` for countries that are not landlocked and equal to `lngdp2000` plus the difference in means for landlocked countries. If you have done this correctly, you should see that the average level of the new variable is the same for landlocked and non-landlocked countries.

4. Do the same re-scaling for the `lnexportslaves` variable: create a variable that is equal to `lnexportslaves` for non-landlocked countries and equal to `lnexportslaves` + the difference in means for landlocked countries.

5. Now regress your new normalized GDP variable on your new normalized slave exports variable. Compare your results to those reported in Question 2. (If you have done this correctly, you should get the same coefficient on slave exports as in Question 2; if not, go back and work through the problems to figure out where you went wrong.) Report your coefficients.

6. Drawing on lecture and the readings, discuss one reason Nunn's estimates may over-state the number of slaves taken from inland countries, and one reason they may under-state the number of slaves taken from inland countries.

7. What factors predict participation in the slave trade? Nunn argues that relatively developed regions were targeted by slave traders, and that distance to the Atlantic ocean and other slave trading centers were key determinants of participation. To assess these claims, we've estimated a regression of log slave exports on log population density in 1400 (`lnpopdensity1400`) and the minimum distance to the Atlantic Ocean (`atlanticdistance`), the Indian Ocean (`indiandistance`), the Sahara Desert (`saharadistance`), and the Red Sea (`redseadistance`). The results are presented in the table below. Discuss the coefficient estimates on the population density variable. Is it significantly different from zero? At what level of statistical confidence? How should we interpret the magnitude of the coefficient? Evaluate Nunn's claims in light of the regression results.

Dependent Variable: Log of Slave Exports

	(1)	(2)
Population Density in 1400	1.554***	1.169***
	(0.364)	(0.391)
Minimum Distance to the Atlantic Ocean	.	-0.619
		(0.638)
Minimum Distance to the Sahara	.	-1.495
		(1.283)
Minimum Distance to the Indian Ocean	.	-0.446
		(0.783)
Minimum Distance to the Red Sea	.	0.261
		(0.793)
Constant	3.490***	15.568
	(0.484)	(12.575)
Observations	47	47
R^2	0.289	0.439

Note: coefficients significantly different from zero at .99 (***), .95 (**) and .90 (*) confidence levels.