ECO 330: Problem Set 2 Professor Jack Rossbach

Due Date: Sunday, April 24 at 11:59PM EST Spring 2016

Instructions:

Complete the questions for each problem. Answers must be typed and uploaded to blackboard as either a word document or PDF to blackboard. Everybody needs to submit their own assignment and type up their own answers.

Problem 1: Settler Mortality as an Instrument for Institutions

There is a strong positive correlation in the data between the quality of a country's institutions and how developed a country is. Without additional investigation, however, it is difficult to determine whether there is a causal relationship and in what direction the causality goes. In this exercise we will use the instrumental variable approach of Acemoglu, Johnson, and Robinson (2001) to try and make a causal statement about whether institutions encourage growth and development.

SETUP) Download the file **Institutions.dta** from my website and open it in STATA. The file is adapted from the data appendix of the paper mentioned above.

This problem requires the use of **STATA**. STATA is available in all university computer labs.

Problem 1 Questions

1.a) Run a regression for the equation

$$\text{GDP}_i = \alpha + \beta_{\text{inst\_qual}} \times \text{Institution Quality}_i + \epsilon_i$$

where GDP is log GDP per Capita (PPP) in 1995 and Institution Quality is an index of Institution Quality in 1900 as, measured by average protection from expropriation risk.

Report the estimated regression coefficients and their confidence intervals. What is the interpretation of $\beta_{\text{inst\_qual}}$? What is the correlation between log GDP per Capita in 1995 and Institution Quality in 1900?

1.b) The values for Institution Quality in 1900 range from 3.5 to 10. According to the regression, what is the predicted value of GDP per Capita (**not logged!**) in 1995 for a country with a value of 10 for Institution Quality in 1900?

What is the predicted value of GDP per Capita in 1995 for a country with a value of 3.5 for Institution Quality in 1900?

1.c) Now we will run a 2SLS Instrumental Variable regression in STATA using Settler Mortality pre-1900 as an instrument for Institution Quality in 1900. What is the reasoning behind using Settler Mortality as an instrumental variable for institution quality? What is the correlation between the two variables?

Report the new estimate of $\beta_{\text{inst\_qual}}$ under the Instrumental Variable Regression. Why might the estimated regression coefficients differ?

1.d) What happened to the number of observations included in the regression compared to 1.a? What is $\beta_{\text{inst\_qual}}$ if we re-run the regression in 1.a using only observations with data on settler mortality? Do the results change much?

1.e) Now we will run an additional specification of the IV regression with controls. Include controls for how many years 1995 is from when the country first gained independence, whether the country is in Africa, how far away from the equator a country is (its latitude), and whether it is landlocked.

Report the coefficients for each of the controls and whether they are statistically significant or not? What happens to the confidence interval for our $\beta_{\text{inst\_qual}}$ coefficient compared to the regression in 1.c?

1.f) As one last check, run the regression from 1.a with all the controls from 1.e. What happens to the estimate for the coefficient on the Africa dummy variable and its statistical significance?

Based on the results from 1.e, how should we view whether geography or institutions are more important for long run development? Assuming we trust the IV results, what do they say about the view that there are factors unique to Africa (e.g. geography, ethnic differences, prevalence of tribal communities) that are needed to explain its lack of development over the past 100 years? Why is using an Instrumental Variable regression important for arriving at this conclusion?

Problem 2: Misallocation and Output

For this problem, we will explore the potential gains from reducing misallocation using data on individual furniture producers (individual factories/establishments) in Chile and India. The Chilean data is from the 2005 Annual National Industrial Survey (ENIA), and includes all manufacturing establishments in Chile with more than 10 employees. The Indian data is from the 2005 Annual Survey of Industries (ASI) which covers manufacturing establishments in the formal sector (around 80% of employment and 20% of value added in India is in the informal sector) with more than 10 employees for establishments with electricity and more than 20 employees for establishments that lack electricity. Values are in local currencies and not adjusted for inflation.

SETUP) Download the file **India_Misallocation_DATA.xlsx** from my website. The file contains the data described above, with outliers and observations with missing values removed.

This problem requires the use of **Excel**. Excel is available in all university computer labs. The university has a license to allow all students to install it on their personal computers for free. See here and here.

Problem 2 Questions

2.a) Compute the following variables for each establishment in the Indian Data

$$\textbf{Total Costs} = \text{Fixed Capital} + \text{Labor Costs} + \text{Intermediate Inputs}$$

$$\textbf{Fixed Capital Share} = \frac{\text{Fixed Capital}}{\text{Total Costs}}$$

$$\textbf{Labor Share} = \frac{\text{Labor Costs}}{\text{Total Costs}}$$

$$\textbf{Intermediate Share} = \frac{\text{Intermediate Inputs}}{\text{Total Costs}}$$

$$\frac{\textbf{Sales}}{\textbf{Costs}} = \frac{\text{Total Sales}}{\text{Total Costs}}$$

Report Summary Statistics for each of the above variables. The summary statistics should include the average value, standard deviation, first quartile, third quartile, min, and max. Which factor of production is the most important (has the largest share)?

2.b) To study the potential gains from reducing misallocation, we will be conducting an analysis based on Hsieh and Klenow (2009). To conduct this analysis we need to compute the following things:

$$\textbf{Capital Wedge} = \left(\frac{\alpha}{1-\alpha}\right) \times \frac{\text{Labor Costs}}{\text{Fixed Capital} + \text{Intermediate Inputs}}$$

$$\textbf{Revenue Wedge} = \left(\frac{\sigma}{\sigma-1}\right) \times \frac{\text{Labor Costs}}{(1-\alpha) \times \text{Total Sales}}$$

$$\textbf{Productivity (TFPQ)} = \frac{(\text{Total Sales})^{\left(\frac{\sigma}{\sigma-1}\right)}}{(\text{Fixed Capital} + \text{Intermediate Inputs})^{\alpha} \times (\text{Labor Costs})^{1-\alpha}}$$

$$\textbf{Revenue Productivity (TFPR)} = \frac{\text{Total Sales}}{(\text{Fixed Capital} + \text{Intermediate Inputs})^{\alpha} \times (\text{Labor Costs})^{1-\alpha}}$$

Where α is computed for you in the worksheet and you need to compute σ as

$$\boldsymbol{\sigma} = \frac{1}{\text{Average Gross Margin}}; \text{ where Gross Margin} = \frac{\text{Total-Sales} - \text{Total Costs}}{\text{Total Sales}}$$

The Average Gross Margin is also computed for you in the worksheet. What is σ? What is the value of each of the four variables for the establishment identified by Factory ID Code 64082?

2.c) Since we only observe values, and not prices and quantities independently, our Productivity (TFPQ) term is not actually correct, as it needs to be multiplied by an unknown constant that depends on prices. This constant cancels, however, when we look at the Productivity of each establishment relative to the Average Productivity across establishments. Compute Relative Productivity for each firm

$$\text{Relative Productivity} = \frac{\text{Productivity (TFPQ)}}{\text{Average Productivity (TFPQ)}}$$

and create smoothed density chart showing the distribution of Capital Wedges and Relative Productivities across firms. Have the horizontal axis be log base 2 scale, and range from 1/32 to 4. Make the chart look professional.

2.d) After you complete the previous steps, the worksheet will automatically complete the Misallocation Exercise to compute actual output relative to potential output ($Y/Y_{\text{Potential}}$). This can be viewed as a measurement of how close the economy is to full capacity. What is this value for the furniture industry in India in 2005?

Now, use that value to compute the % gain from eliminating misallocation in the model according to the formula

$$\%\ \textbf{Gain} = 100 \times \left(\frac{Y_{\text{Potential}}}{Y} - 1\right)$$

What is the potential % gain from eliminating misallocation in the model?

2.e) Follow the link to the original paper here. Look at Table IV on page 19 of the PDF (numbered 1421 in upper right corner). What are the authors' calculated TFP gains from eliminating misallocation across all industries in India in 1994, China in 2005, and the United States in 1997?

What are some examples of things that may lead to misallocation? Why might misallocation be greater in India than it is in developed countries such as the United States?