14.771 Development Economics: Microeconomic Issues and Policy Models
Fall 2008

14.771, Development Economics
Problem Set 2 - Bleakley (2008)

Carefully read Bleakley's article "Malaria in the Americas: A Retrospective Analysis of Childhood Exposure". The following questions will make sure that you are following the author's arguement and econometric techniques. This problem set will use an extract with U.S. data very similar to that used in the paper - it can be downloaded from the Stellar website. You should hand in your Stata code and output with your problem set. This data is not identical to that used in the actual paper, so don't worry if your results don't exactly match Bleakley's. (For example, we have malaria mortality per 10K population rather than fraction of total mortality, so the regression results will be of a different order of magnitude). The following variables can be found in the file:

year	Census year
bplg	State of birth
yob	Year of birth
sei	ln(Duncan SEI)
occscore	ln(Occupation Income Score)
cellsize	Number of Obs in Microdata Cell
wtbpl	sqrt(cellsize)
age	Age
hookworm	Hookworm inf rate, Kofoid
lebergott99	Average wage, farm laborers, 1899
malmort1890	Malaria mortality per 10k pop, 1890
south	Dummy for US South
exp_mal	Fraction age 0-20 exposed to malaria eradication campaign

Question 1

This paper attempts to measure the impact of widespread malaria reduction campaigns on adult outcomes.

1. As economists, why are we interested in this estimate?

2. Suppose a colleague suggests simply running a regression of adult outcomes (say, wages) on regional malaria intensity. Why might this be problematic? List some factors you think may bias this estimate, list the direction of bias you'd expect from each factor, and explain your reasoning.

3. What is the key feature of the identification strategy in this paper? What assumptions must be satisfied for estimates to be unbiased? Do you believe them? What pieces of evidence would you like to see?

Question 2

First, define the "young" as those who were born in 1920 or later (exp_mal==1). They should have benefitted from the malaria eradication progam since birth. Define the "old" as those who were born in 1899 or earlier (exp_mal==0). They should have been 21 years old (or older) in 1920 and thus should not have benefited fom the program at all as children. For the moment, we'll ignore the transitional cohorts born between 1900-1919. Next, define "high" malaria areas as those in the 90^{th} pecentile or higher in terms of "malmort1890". Define "low" malaria areas as those in the 10^{th} percentile or lower. Ignore the intermediate percentiles for now.

1. Now, generate a new variable - sei_diff - which is the difference in the log Duncan SEI index between high and low malaria areas (you will have to collapse your data to do this). Plot this difference on the y axis with cohorts on the x axis - make a separate plot for the old and the young. If you like, you can use Stata's "twoway lfit" command to add a regression line to your plots. What do you see? How does this pattern fit with the author's arguement?

2. Now calculate the average difference in the Duncan SEI measure between the high and low malaria areas for the old and the young. Subtract your measure for the old from your measure for the young. (Hint: you can do this with a regression and get standard errors on your differences - don't forget to limit your sample appropriately!). Do the same for the occupational income score What do we call this estimator?

3. What are your above estimate and what do they tell us? What are the identification assumptions necessary for this estimate to be valid?

4. What is your assessment of the validity of the identification assumptions in light of the plots you produced in question 2.1?

5. Now define two more groups - the very old - who were born in 1860 or earlier, and the very young - those who were born in 1940 or later. Perform the same exercise you did in part 2.2 between the old and very old and the young and very young (don't forget to limit your sample!). What should you see if the identifying assumptions hold? What do you see in practice? Does this change your assessment of the author's arugment?

Question 3

We created dummies partitioning cohorts and malaria intensity, but we actually have a lot more information than this. We will start be taking account of the fact that we have continuous data on malaria intensity.

1. Try and reproduce the first two OLS estimates in Table 2 (we do not have the data to reproduce the IV estimates). What are they telling you?

2. In author includes sets of specifications with additional controls. Why? What do you think about the results? Do you see any potential problems with any of the controls?

3. Suppose a colleague reads this paper and interprets Bleakley's .62 estimate for Brazil (on page 20) as "the causal effect of eliminating childhood malaria exposure (for Brazilian adults who lived in areas where malaria rates were reduced by the campaigns) is to raise adult wages by 62%." What do you think about this statement?

Question 4

There is still more variation that we can use. Above, we just compared the young to the old, but there are also 20 cohorts who benefited from the campaign for some fraction of their youth. *Note:* as with question 3, you will not be able to reproduce these results in terms of magnitude, but you should get things that are qualitatively the same.

1. Try and reproduce figure 4. What is the author plotting? What does this tell us? Why do you think he chose to display the information in this way?

2. You may have noticed that the dots are more spread out for earlier cohorts. Why is this? In his estimation, does the author do anything to address this?

3. Now try and reproduce the two-step estimate in table 1, panel A. (Hint: take the βs that you saved to recreate figure 4 - these go on the left hand side. The right hand side includes cohort level variables as indicated in the text. You will find the formula for Exp_k in footnote 11.) What are the results telling us and how are they related to figure 4?

4. Why does the author renormalize the coefficients to facilitate interpretation? How should we interpet the coefficients before and after renormalization?

5. What is the author trying to test by controling for polynomials of cohort age? Why might this be important?

6. Can you think of another specification test you could do? Outline your test and do it. What are your results? Assume for a moment you could get any data you wanted - can you think of any other specification tests you'd like to see?

Question 5

Now let's take a step back and think about the paper as a whole.

1. What do you think about the magnitude of the author's estimates - are they small, large? Do they seem reasonable? Think of this in light of the different channels he discusses. Can you think of any other explanations for the results? These can be causal or due to bias. How would you test your theories if you had access to unlimited data?

2. Jeffrey Sachs has prominently argued that large malaria burdens play an important role in Africa's longstanding economic malaise, and proposed that successful malaria control campaigns would significantly improve African economic growth. Do you think these estimates bolster this argument, or do they go against it?