14.771 Development Economics: Microeconomic Issues and Policy Models
Fall 2008

14:771: Recitation Handout #3
Acemoglu and Johnson (2006), Field, Robles, and Torero (2007)

Acemoglu and Johnson - "Disease and Development"

Background

You may have found this paper very similar to Bleakley in spirit, although it asks a somewhat broader question - "what is the impact of increased life expectancy on economic performance?". This is definitely a big questions, and like to many big questions, it entails the use of cross-country regressions.

This paper exploits what the authors refer to as the "international epidemiological transition" - a set of international health interventions, public health measures, and new chemicals and drugs which were introduced in the 1940s and significantly increased life expectancy, especially in lower income countries (see figure 1, plus footnote caveat). Key innovations and the diseases that they impacted include:

- Penicillin - effective against a range of bacerial infections (like pneumonia). This only became widely available during the mid 1950s
- Streptomycin - an antibiotic effective against tuberculosis
- Yellow fever vaccine
- DDT - cited as "the major chemical innovation of this era" - we should all now know that this is very effective in preventing malaria.
- Establishing the WHO (they spearheaded public health campaigns in developing countries)
- The authors also cite a "change in international values" that helped with the diffusion of health interventions.

The authors, much like Bleakley, argue that these innovations were exogenous and driven by scientific progress. Also, the major causes of mortality at the onset of the 1940s were tuberculosis, pneumonia, and malaria.

Data and Identification Strategy

The authors amass life expectancy, population, GDP, and disease-specific mortality data from a number of sources for 75 countries (this is their base sample). Africa is not in their base sample due to concerns about data quality and reliability, although robustness to the inclusion of African countries is examined. The authors also document that population and GDP are both significantly correlated with life expectancy, although log GDP per capital is negatively correlated with life expactancy. *Why might we see this?*

The identification of the paper hinges on the author's "predicted mortality instrument". Here is the formula:

$$M_{it}^{I} = \sum_{d \in D} \left[\left(1 - I_{dt}\right) M_{di40} + I_{dt} M_{dFt} \right]$$

where d indexes diseases, t indexes years, and i indexes countries. M_{di40} is mortality in country i from disease d in 1940. I_{dt} is a dummy indicating that an intervention for disease d had occurred at time t or before. M_{dFt} is the mortality rate from disease d at the "health frontier" of the world at time t. So what is this? Basically mortality due to disease d in a country is set to its own country-specific 1940 level before a health intervention, and then to the global standard thereafter (in their baseline instrument, the authors set this global standard to 0, which would correspond to full disease erradication). The regressions in this paper are long difference regressions - that is they difference each variable (including the instrument) according to 1980-1940 or 2000-1940. *So what is the instrument then? What do we need for this instrument to be valid? Do you think that this is likely to be satisfied in practice?*

The authors then take the time to document that everything works. First, they check that the health interventions that they identify are actually correlated with improvements in life expectancy (this is good - otherwise there is no instrument - you can see this on table 4). They then present the first stage. As an aside - any paper that does IV should always present the first stage, so the reader can verify that there is unlikely to be a weak/many instruments problem. We can see this in Table 5. *What do we see? What are the control variables?* The regression equation for the first stage is:

$$x_{it} = \psi M_{it}^{I} + \zeta_{\iota} + \mu_t + Z_{it}'\beta + u_{it}$$

Where x_{it} is life expectancy at time t, M_{it}^{I} is predicted mortality, ζ_i is a country fixed effect, μ_t is a year fixed effect, and Z_{it} are other exogenous covariates.

The authors then test for mean reversion in life expectancy. *What is mean reversion, and why might it be a problem?* Table 6 presents their test, which is a regression of log life expectancy on predicted mortality and lags/leads of predicted moreality, as well as lagged log GDP per capita, and lagged log life expectancy. *Is there evidence for mean reversion? What do you think the most important regression on this table is, in terms of the authors' argument?*

Let's look at the core results. The second stage regression is

$$y_{it} = \pi \hat{x}_{it} + \zeta_i + \mu_t + \sum_{t=1940}^{1980} c_i' w_t + \varepsilon_{it}$$

where c_i includes average institutions or initial (1930) log population. *We can see the core results on tables 8 and 9. What do we see? Is any of this surprising to you? Are there any specification checks you'd like to see?*

Let's check out the specification check, which is actually presented earlier in the paper, in table 7. *What is the check? Does it pass? How do you feel about the results now?*

A Reconciliation?

This paper finds that major disease eradication actually lead to decreases in per capita income due to large changes in population growth. This serves as a strong contrast to what Bleakley found with malaria (which is one of the "big three" in the Acemoglu and Johnson paper). Do you think both of these papers can be right? If so, what is your story? If not, which one are you more inclined to believe? Do you prefer one identification strategy to another? What are the differences? What additional evidence would you like to see to clarify which paper rings closer to the truth?

Field, Robles, and Torero - "Iodine Deficiency and Schooling Attainment in Tanzania"

Background

Iodine is an important micronutrient for humans - maternal iodine deficiency is known to have severe and irreversible effects on the fetal brain development, particularly in the first trimester of pregnancy. Some experimental evidence suggests that this effect may be larger for females. Moreover, the problem is widespread - estimates indicate that around one billion people are at risk of brain damage due to iodine deficiency. Deficiency is caused by lack of iodine in the diet (areas with old soil and groups that consume little seafood are particularly at risk), and exacerbated by consumption of goitrogens - foods which impede iodine absorbtion by the body (these include cabbage, legumes, chaya leaves, and cassava). This paper seeks to shed light on the impact of maternal iodine deficiency on human capital accumulation.

In order to combat iodine deficiency, many countries have undertaken massive salt iodization campaigns. Tanzania has historically had significant problems with iodine deficiency - a survey conducted in the 1970s incidated that about 25% of the population (over 5 million people) suffered from iodine deficiency disorder (IDD). As such, Tanzania was targed for iodine supplementation via distribution of iodized oil caplets as a short term measure intended to improve IDD rates before the phase in of iodized salt, which was to take place in the mid 1990s. Program districts were selected on the visible goiter rate (VGR), which is an important symptom of IDD. Districts with VGR$\geq$ 10% were selected for the program (25 treatment districts encompassing 25% of the population).

Women of child bearing age were identified as highest priority for iodine supplementation. They were to receive one 380mg capsule every two years (iodine is retained in fatty tissue for quite some time). In practice, there were some major program delays - 10 districts began supplementation by 1988, and three did not start until 1992. In addition, districts were not always reached every two years to maintain appropriate supplementation levels. The authors had access to a good deal of administrative reports and records detailing the timing of the intervention district by district.

Data and Identification Strategy

The authors use data from two different sources - the 2000 Tanzanian Household Budget Survey (THBS) and the 2004 Tanzanian Demographic and Health Survey (TDHS). The TDHS has the nice feature that it records the birth-month of children. The sample was restricted to all children aged 10-13 in the 2000 survey and aged 10-14 in the 2004 survey who could be linked to mothers in the household. The lower bound of 10 is the modal age of school enrollment (nearly all children attend primary school in Tanzania). The maximum sample age was capped at 13 for the 2000 sample because this is the oldest cohort exposed to the treatment, and capped at 14 for the 2000 sample because children begin to leave parental households at high rates at age 15 (to marry or attend secondary school). *Why might this be a problem?*

Since most children attend school up through the end of primary school, the major outcome in this paper is educational progress (children are often held back at certain grades) rather than dropout. *Look at table 2 - what can you say about intervention and non-intervention districts? How might this inform your identification strategy?*

The authors used administrative information on program implementation and data on birth date to construct a variable equal to the estimated probability that a child was fully covered by the iodine supplementation program during the first trimester in utero. *Given this information, think of regressions that we could run. Who should be the treatment group? Who should be the control group?*

The core regression equation. It is given by:

$$grade_{if} = \alpha + \beta_1 T_{if} + \beta_2 A_{if} + \beta_3 X_{if} + \mu_f + \varepsilon_{if}$$

Where T_{if} =likelihood of program exposure in first trimester, A_{if} =a vector of birth year dummies, X_{if} =gender and birth order controls, f indexes families and i indexes individuals. *What are the two main types of regressions that the authors run? Which do you prefer? Tables 3 and 4 contain the core results. What do we see here? What do you think is driving the difference in the household vs. district level fixed effects regressions? What are some treats to identification that you can imagine?*

Specification Checks

This paper has a ton of specification checks, which is a nice thing to see given the identification strategy. *Before we discuss them, what types of identification checks would you perform? What would you expect to see from your check if identification held?* For brevity - I'll enumerate some of the checks in a list

1. "Placebo regression" - run the regressions using kids who were 10-13 in 1988 and thus too old to benefit from the program (Appendix D)

2. I really like this one - use program gaps and the individual fixed effects regressions - see if older siblings to relatively better when the older one got treated due to gaps, and see if younger siblings gained when the younger sibling got treated and the older one did not. (Figure 6, Table 7)

3. Restrict control to children born during a program year (Table 6)

4. Compare estimated treatment effect for children in utero when mothers first took the pill to children in utero two years after mothers first took the pill. *Why might this matter?* (Table 6)

5. Health impacts - if the story is iodine deficiency, the primary hypothesized causal mechanism should be reduced cognition rather than health effects. They generally find this, but do find a reduction in child mortaility after age 1 associated with treatment (Table 8). *What are some reasons why this could be happening?*

Other Stuff

The authors also use less good data to look at the pass rates of a secondary school admissions exam (fail rates are very high, particularly for girls) and run some cross country regressions - results are in line with their hypothesis, but I won't go into any more detail.

Parting Thoughts

- What did you think of the paper? Was the argument convincing?
- Is there anything you would have liked to see that they did not present?
- Did you find anything extraneous or unnecessary?
- How does this paper contribute to the existing literature?
- Can you think of other questions that could be answered in a similar way? Anything you've been thinking about for your own research?