Econ 270C
Problem Set 1 – due 3/3/2005

The goal of this problem set is to credibly estimate the impact of deworming treatment (drugs that fight intestinal worm infections, like hookworm) on school attendance among Kenyan primary school children. Intestinal worm (or helminth) infections are extremely widespread in many less developed countries, and it has long been claimed that they may have adverse impats on schooling.

You should use the STATA dataset (or EXCEL sheet) posted on the course website to do this, following the suggestions below. I recommend you use STATA for the econometric analysis, since it is quite standard in economics, but you can use another statistical package if you prefer. For each of the 22,202 students in the sample there are two observations, one for 1998 and one for 1999 (for a total of 44,404 observations).

(a) Present the summary statistics (mean, standard deviation, minimum, maximum, observations) for all of the variables in the dataset. Briefly characterize this population. **[1 point]**
(STATA command hints: "use", "summarize, detail", "twoway (scatter y x)")

(b) Regress *attendance* on *took_drug, girl,* and *year_1999*. Make sure to account for the possibility that disturbance are correlated within the same school. Interpret the results. Why might this sort of specification be subject to omitted variable bias? In which direction (positive or negative) would you expect this bias to go? **[2 points]**
(STATA command hints: "regress", "cluster")

(c) One way to deal with omitted variable bias is by controlling for unobserved but fixed individual characteristics. Re-organize the dataset such that the dependent variable is the difference between school attendance in 1999 and in 1998 for the same student (call this variable *attendance_difference*). Call the difference in drug take up between 1999 and 1998 *drug_difference*. Now regress *attendance_difference* on *drug_difference*, once again taking into account correlated disturbance terms within schools (as you should do throughout). How does the result differ from your findings in part b above? Also present a scatter plot of pupil school attendance in 1998 (x-axis) versus 1999 (y-axis). **[1 point]**
(STATA command hints: "reshape", "generate")

Another way to deal with omitted variable bias / selection bias is through the use of instrumental variable methods. In the Kenyan setting we are studying, the schools that were provided deworming drugs were selected using a computer random number generator: among the 75 schools in the sample, 25 "Group 1" schools received treatment in 1998 and 1999, 25 other "Group 2" schools received treatment only in 1999, and the final 25 "Group 3" schools did not receive drugs in either year.

(d) Regress *took_drug* on *treat_school*, *girls*, and *year_1999*. What would we call this specification in the IV framework? Briefly interpret the results. **[1 point]**

(e) The intention-to-treat estimator (ITT): Regress *attendance* on *treat_school*, *girls*, and *year_1999*. What would we call this specification in the IV framework? Briefly interpret the results. **[1 point]**

(f) The treatment effect on the treated estimator (TOT): Regress *attendance* on *took_drug*, *girls,* and *year_1999*, using *treatment_school* as an instrument for *took_drug*. Interpret and discuss the main finding. **[2 points]**
(STATA command hints: "regress y x1 x2 (z1 z2)")

Like many other health problems, worms are an infectious disease: they are transmitted from one person (or more generally, disease vector) to others. In particular, worms are transmitted through infected fecal matter. People who live in areas with poor hygiene and sanitation are particularly likely to become infected. Deworming kills the worms that are in the body, in which case that individual no longer passes along worm eggs through fecal matter.

(g) Imagine that these deworming treatment externalities are large for children who attend the same school (but assume that cross-school infection externalities are zero). Discuss what this means for the instrumental variable identification strategy discussed above. Which (if any) of the IV conditions may break down in this case? Describe other approaches for identifying the overall impact of deworming in the presence of these externalities. **[2 points]**

(Extra question) Is the impact of assignment to deworming significantly different for girls and boys in this population?