Problem Set 1

Due 2/16/2004 in TA's mail folder by noon.

The Clean Air Act has been one of the most controversial interventions into the marketplace by the federal government in the post-World War II period. Much of this controversy is generated by an absence of information on the costs and benefits of these policies. Recently, this controversy has centered on the wisdom of regulating small particulates air pollution. The EPA contends that these small particulates have pernicious effects on human health, yet many find the evidence to be inconclusive. I am in the midst of a research project that examines the effects of total suspended particulates (TSPs) air pollution on infant health. This problem set will involve you in the scientific process of discovery in "real time".

Feel free to work cooperatively but each person is required to turn in their own problem set that provides the solutions in their own words. The answer key will be posted immediately after the problem sets are due; hence late problem sets will not be accepted.

For those of you who become interested in this topic, you might be interested in reading:

Chay, Kenneth, and Michael Greenstone. (2003). "Air Quality, Infant Mortality and the Clean Air Act of 1970." Available at: http://dsl.nber.org/papers/w10053.pdf.

Chay, Kenneth Y., and Michael Greenstone, "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," Quarterly Journal of Economics, CXVIII (2003), 1121-1167.

The first paper may be useful in completing this problem set.

Data Source: tspsimr7172.dta (available on the assignment section)

This data file contains observations from 501 counties. Each observation records the change between 1972 and 1971 (i.e., the 1972 minus the 1971 value) for each of the variables. There are two exceptions:

     a. tbirth1 which equals the sum of the 1971 and 1972 number of births. **This variable should be used as a weight in ALL regressions in this exercise (you can use what STATA refers to as "analytical weights").**

     b. reg_tsp which equals 1 for tsps nonattainment counties and 0 otherwise. The centerpiece of the Clean Air Act was its efforts to clean-up "dirty" counties. These counties were designated as nonattainmnet (i.e., their air quality did not attain the federal standards). All other counties were considered attainment. The key issue for this problem set is that polluters faced more stringent regulatory oversight in the nonattainment counties.

The relevant variables are listed in four separate categories and all have descriptions in quotations:

1. Base Variables:

reg_tsp "=1 if county regulated for tsps"
dimr7271 "# inf death per 1000 births 72-71"
damtsp1 "change in tsps 72-71"
tbirth1 "total births 71 & 72"

2. Basic Natality Variables:

dwhite "% births, white mom 72-71"
dothr "% births, nonwhite/nonblack mom 72-71"
dfemale "% female births 72-71"
dedudad "father yrs of ed 72-71"
dedumom "mother yrs of ed 72-71"
dmaried "% mother married 72-71"
dagemom "mother age 72-71"
dpcare0 "% mom never received prenatal care 72-71"
dpcare2 "% mom began prenatal care in 3rd month 72-71"
dpcare3 "% mom began prenatal care in 4-6th month 72-71"
dpcare4 "% mom began prenatal care in 7-9th month 72-71"

3. Unrestricted Natality Variables:

dmedu1 "=1 if mother < 12 years of education 72-71"
dmedu2 "=1 if mother 12 years of education 72-71"
dmedu3 "=1 if mother 13-15 years of education 72-71"
ddedu1 "=1 if father < 12 years of education 72-71"
ddedu2 "=1 if father 12 years of education 72-71"
ddedu3 "=1 if father 13-15 years of education 72-71"
dnhosp1 "=1 if delivery outside hospital 72-71"
dphys1 "=1 if physician present at delivery 72-71"
dmage1 "=1 if mother's age 10-14 72-71"
dmage2 "=1 if mother's age 15-19 72-71"
dmage3 "=1 if mother's age 20-24 72-71"
dmage5 "=1 if mother's age 30-34 72-71"
dmage6 "=1 if mother's age 35-39 72-71"
dmage7 "=1 if mother's age 40+ 72-71"
dagedad "father age 72-71"
dtwins "not a singleton birth 72-71"
dcare "number of prenatal care visits 72-71"
ddead1 "=1 if 1 previous fetal death 72-71"
ddead2 "=1 if 2 or more previous fetal deaths 72-71"
dllive "=1 if last pregnancy resulted in a live birth 72-71"
dlterm "=1 if last pregnancy resulted in a fetal death 72-71"
dord2 "=1 if 1 previous live birth 72-71"
dord3 "=1 if 2 or more previous live births 72-71"
dpreg1 "=1 if 1-11 months since termination of last pregnancy"

dpreg2 "=1 if 12-17 months since termination of last pregnancy"
dpreg3 "=1 if 18 or more months since termination of last pregnancy"
dfdth1 "=1 if 1-11 months since termination of last fetal death"
dfdth2 "=1 if 12-17 months since termination of last fetal death"
dfdth3 "=1 if 18 or more months since termination of last fetal death"
dintrv1 "=1 if 1-11 months since termination of last live birth"
dintrv2 "=1 if 12-17 months since termination of last live birth"
dintrv3 "=1 if 18 or more months since termination of last live birth"

4. Per-Capita Income
dpcinc "per-capita income 72-71"

The data file is available at the assignments section. Please include a constant term, analytic weights (using the tbirth1 variable) and robust standard errors in all of your regressions.

1.   Graph dimr7271 against damtsp1.

   a) Does it look like there is an association between changes in infant mortality and tsps? (Please label the axes.)

   b) A nice feature in Stata is that it allows for the size of the points on a graph to vary with a variable; this is done by setting the weight = the variable of interest. Now graph dimr7271 against damtsp1 and set weight equal to tbirth1. Does there appear to be a relationship? (It is easier to see the graph if you use hollow markers rather than filled in ones.)

2.   We now want to prove the robustness of the findings in 1. Regress dimr7271 on damtsp1.

   a) Is there evidence of a relationship here?

   b) Now add the "Basic Natality" control variables. Are any of these variables important determinants of infant mortality?

   c) Have your conclusions about the TSPs/infant mortality relationship changed?

3.   One explanation for this result is that there are many unmeasured/unobservable confounding factors that determine both dimr7271 and damtsp1. Examples of these types of variables include: health insurance status, rates of smoking across mothers, and parents' income.

   a) Now interpret your findings in 1.

        If changes in TSPs were randomly assigned, we could be certain that the 'omitted variables bias' problem described in (3) was unimportant. Obviously, a randomized (or classical) experiment is unethical. In the absence of a classical experiment, one solution is to use a "natural experiment", where the researcher exploits differences in outcomes between a

treatment group and a control group, just as in a classical experiment. In the case of a natural experiment, however, treatment status is determined by nature, politics, or some other action beyond the researcher's control. Despite the "nonrandom" nature of the treatment status, it is still possible to draw valid inferences from the differences in outcomes between the treatment and control groups in a natural experiment, provided certain (potentially testable) assumptions are met.

The premise of this exercise is that the 1970 Clean Air Act Amendments might qualify as a valid "natural experiment". This legislation set air quality standards for TSPs that all counties are required to obtain. In counties that did not meet the standards, TSPs emitters were subject to harsh regulations that required them to reduce their emissions of TSPs (we can think of this as the treatment), while emitters in "clean" counties were relatively free from regulation (we can think of this as the control). I have collected the information for whether a county was regulated heavily and coded it as a variable that is equal to 1 for heavily regulated counties and 0 otherwise. The variable name is reg_tsp.

4.   Run a regression to determine whether TSPs declined more in heavily regulated counties.

   a)   What does the parameter on the regulation variable mean in words?

   b)   Does the result remain when you control for the covariates that were specified in (2)?

   c)   What is your conclusion about the effect of the Clean Air Act Amendments on TSPs concentrations?

5.   Run a regression to determine whether infant mortality rates declined more in heavily regulated counties.

   a) What does the parameter on the regulation variable mean in words?

   b) Does this result remain when you control for the covariates that were specified in (2)?

   c) What is your conclusion about the effect of the Clean Air Act Amendments on infant mortality rates?

6.   Two-stage least squares provides a method to combine the results from (4) and (5) in order to infer the relationship between infant mortality and TSPs (recall, this is the relationship that we are actually interested in). The intuition is that the regulations may isolate changes in TSPs that are unrelated to the sources of omitted variables bias that were discussed in (3); thus, the regulations may serve as a natural experiment.

   a)   Run a two-stage least squares regression (which is equivalent to instrumental variables estimation here) to find the effect of changes in TSPs on changes in IMRs, without controlling for any other explanatory variables. That is, you want to regress the change in infant mortality on the change in TSPs, instrumenting the change in TSPs with

the regulation variable (reg_tsp). Don't forget to include the weights and to use the "robust" command for robust standard errors.
{**Read the STATA manual to make sure that you understand the commands being used here.**}. What do the results suggest? What does the parameter on damtsp1 mean in words?

b) Now control for the "Basic Natality Variables." Did your findings change?

c) Now control for the "Unrestricted Natality Variables" and "Per-Capita Income". Did your findings change?

d) Are the results in (a) sensitive to controlling for a wide set of available predictors of infant mortality? What do you conclude about the relationship between infant mortality and tsps?

7.   With the benefit of hindsight, should the Clean Air Act Amendment have been passed in 1970? Why or why not?