

PUBPOL 639: ASSIGNMENT 4

Winter 2011

Due: Friday, April 1st at the start of class

This assignment continues to examine the concepts of omitted variable bias and also starts to examine the linearity assumption implicit in OLS. Be sure to copy your do and log files into the back of your solutions. Courier 8 or 9 pt font works well. Your log file may be long, so don't worry about cleaning/formatting it.

PART I: Media Access, Environmental Concerns, and Views about Women's Rights

In this assignment you will examine the relationship between media access (as proxied by television watching) and views about the environment and women's rights. Some people believe that increasing access to the media may increase people's awareness of global problems (e.g. global warming and environmental pollution) and promote more progressive values (e.g. gender equality). In this first part, your task is to examine this theory and write a short memo describing what you find.

Data

The data you will analyze come from the World Values Survey. As described on their website: *"The World Values Survey (WVS) is a worldwide network of social scientists studying changing values and their impact on social and political life. The WVS in collaboration with EVS (European Values Study) carried out representative national surveys in 97 societies containing almost 90 percent of the world's population. These surveys show pervasive changes in what people want out of life and what they believe. In order to monitor these changes, the EVS/WVS has executed five waves of surveys, from 1981 to 2007."* [<http://www.worldvaluessurvey.org>].

Our two key outcome variables are:

envir – this is a measure of respondent's willingness to pay higher taxes to address environmental pollution. The question reads:

I am now going to read out some statements about the environment. For each one read out, can you tell me whether you agree strongly, agree, disagree or strongly disagree?

"I would agree to an increase in taxes if the extra money were used to prevent environmental pollution."

The variable ranges from 4 (strongly agree) to 1 (strongly disagree). You should interpret this as a generic measure of concern about environmental pollution, relative to other factors.

womenwork – this is a measure of respondent's support of women in the workforce. The question reads:

For each of the following statements I read out, can you tell me how much you agree with each. Do you agree strongly, agree, disagree, or disagree strongly?

"A working mother can establish just as warm and secure a relationship with her children as a mother who does not."

The variable ranges from 4 (strongly agree) to 1 (strongly disagree). You should interpret this as a generic measure of support for women in the workforce.

The key explanatory variable is **tvhours**, which measures the average number of hours spent watching television each day. The question reads:

Do you ever watch television? If yes: How much time do you usually spend watching television on an average weekday (NOT WEEKENDS)?

It has been recoded into four categories (0, 1.5, 2.5, or 4 hours).

Though the world value survey includes information from a number of different countries and for a wide range of years, this subsample only includes data from Eastern European, former Soviet and Balkan countries during the late 1990s.

Format

For this part of the assignment you will turn in a one page (single-sided) memo. Your memo should address the following:

- What is the unadjusted relationship between television viewing and support for pollution reduction?
- How does this relationship change (if at all) when education, age, sex, and town size are controlled for?
- What is the unadjusted relationship between television viewing and support for women in the workforce?
- How does this relationship change (if at all) when education, age, sex, and town size are controlled for?
- Based on your findings, what is your overall assessment of the theory proposed in the introduction?
- What are some other variables that you would like to control for if you could? How do you think that omitting them has affected your estimates?
- What concerns would you have with applying your findings to the context of Latin American countries today?

You should present all your regression results in a table similar to the one you made for the last problem set (except now you have two outcome variables, so you may need two tables). The page limit includes your table of regression output, so the total amount of text and regression output should fit on a single page. Your do and log files do not count towards the page limit.

Your memo should be clear and concise and highlight the main findings. You don't need to discuss every minor detail in your regression output, but do discuss the important things. For instance, if you find that including a certain control variable completely changes your estimate [I am not saying that you will], then you should discuss why this happens.

Note on interpreting magnitudes. As always, you should provide a “real-world” interpretation to the magnitude of the relationship you estimate. This is particularly tricky in this case since the outcome variables do not have natural units. One approach would be to compare the magnitude of the coefficient on television watching to one of the other coefficients. For example, “an additional hour of television watching is associated with a change in support for environmental policy similar to X years of age or Y years of education”....or something like that.

PART II – The Conditional Expectation Function

You may recall from lecture that OLS fits a line to the means of Y for each X value. That is, it fits a line to the conditional expectation function. This part of the assignment asks you to verify that you get the same coefficient when you run the regression on individual data and when you run it on conditional means [using the appropriate weights]. To do this you will first need to “collapse” your dataset so that each observation corresponds to a value of `tvhours`. Since there are only four values for `tvhours`, you should have four observations. The collapsed outcome variables are the average across all individuals with a given level of `tvhours`. You will then run the same bivariate regression as before, but only on these four observations.

To collapse your data, you should type:

```
collapse (count) count=country (mean) enviro women, by(tvhours)
```

To make it consistent with the OLS on micro data, the aggregated OLS regression should be weighted by the number of observations in each tv-watching category:

```
regress Y X [aweight = count], robust
```

1. Create a table of regression output. There should be four columns, corresponding to the four bivariate regressions you have run (two with individual data, two with aggregate data. Do the individual and aggregate regressions give you the same coefficient estimates?
2. Plot the conditional expectation functions and the (appropriately weighted) fitted OLS lines. To do this, you should type:

```
twoway scatter enviro tvhours || lfit enviro tvhours [aw = count] ||  
    scatter womenwork tvhours || lfit womenwork tvhours [aw = count] ||, legend(  
label(1 "mean(enviro)") label(2 "fit(enviro)")  
    label(3 "mean(womenwork)") label(4 "fit(womenwork)"));
```

In these regressions, what have we implicitly assumed about the difference in outcomes between 0 and 1.5 hours of television daily and between 2.5 and 4 hours daily? From your graph, does this assumption seem like a good one? Explain.