**Advanced Econometrics**

Assignment 1: Causality, basic econometrics, hypothesis testing, interpretation of variables

All answers should be prepared and submitted using MS word or latex (or any other editor you prefer). Some questions ask you to produce regression output, but many also ask you to interpret findings and talk about the paper’s idea as a whole. Regression output should be in a table format along with other answers - do not submit stata or other software log files. Stata’s outreg and outreg2 commands help translate regression output into publication-quality tables. I am sure there other similar scripts to convert software output to excel or word in other softwares like R, SAS, SPSS. Please take a look at these commands to produce good quality tables and include them in your assignment document. Submit a stata .do file along with the assignment answers, or other program codes if available. Produce the code with neat comments indicating which part of the code belongs to what question (including sub-parts).

1. Does being able to work with a computer raise productivity and earnings? In an important paper, Krueger (1993) included a measure indicating whether a worker uses computers at work in standard wage regressions (“How Computers Have Changed the Wage Structure" Quarterly Journal of Economics, February 1993, 33-60). The paper’s Table II (page 38) contains the main results. The paper found a considerable wage differential for using computers as opposed to not using computers. This has been taken as evidence in favour of the view that introduction of new technologies improve productivity and wages. This question asks you to evaluate this claim. Open d\_p.dta, the dataset used in another study “The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?”, by John E. DiNardo and Jorn-Steffen Pischke published in the Quarterly Journal of Economics. This is based on pooled cross sections of the Germany’s West German Qualification and Career Survey. Use desc and tab commands to understand the variables.
2. Find the difference in the mean log wage of computer users and non-computer users using the 1979 data.
3. What are the assumptions under which this difference would be an unbiased estimate of the effect of computers on wages?
4. Examining Krueger’s table 2, how does the effect of computer usage on wages change when the regressions include nothing else (column 1 and 4) and when the regressions include years of education and other variables (columns 2 and 3, 5 and 6). What does this “change” tell you about the relation between the schooling of a worker and his or her computer use on the job? Which computer coefficient is likely to be a better measure of the causal effect of computer skills? Why?
5. Using the 1985 data in dp.dta, run a regression similar to Krueger's table 2 (columns 1 and 2 only). Some variables Krueger uses don't exist in the German data like race or veteran status. Ignore those variables unavailable. Show regression results in a table. Compare your results for Germany to the U.S. results and comment on them.
6. Use the entire dataset d\_p.dta – which contains a pooled cross-section from 1979 and 1985. Run the same regression as in 1.d, but now include a dummy variable for year 1985, in addition to the other covariates, and an interaction term between computer use and your dummy variable for 1985. Add your results in a new column to the table your prepared for 1.d.
7. How can you perform a formal test of the hypothesis that the association between computer use and wages is the same in 1979 and 1985? Clearly state your null and alternate hypothesis. Perform your test in stata. What is the result of your test?
8. Till now, we have controlled for variables like schooling, labor market experience, gender, etc. But there could be occupation specific unobservable characteristics that can also affect computer usage and wages. For example, workers in higher paying occupations may be more likely to use computers. Ability may be higher for workers higher wages jobs, and they are the ones likely to use a computer on their desk. It is therefore worthwhile to control for occupation or ability measures in the regressions. These variables are available only for 1979.
9. Repeat your regressions from 1.d for the 1979 data, but additionally include occupation fixed effects. Since there are numerous occupation categories and you may not want to see a coefficient specific to each occupation, you can use the areg command in stata (with the absorb option) to see a simplified output. Show regression results in a table. Type “help areg" or “man areg” to obtain help. Also, write a sentence or two about how occupation fixed effects are going to help us address some of the endogeneity issues.
10. German and math scores, referring to grades in German and mathematics classes, are indicators of ability. The variable father contains occupation code for the father. Use these three as additional variables in the regression for 1979 data, and show results in the same table as you produced for g.i, but in a new column. These variables only have few categories, so we can include separate dummies for each of these categories in the regression (use the “i.” operator). What is the rationale for adding these extra variables? How does it address endogeneity issues?
11. Other variables indicating what people do/use on their job are also available: whether they use a calculator or cash register (calc), a telephone (teleph), writing materials (the variable is called pencil), whether they work mostly while sitting (sit), or whether they use blue collar tools like a hammer, screw driver, drill (hammer) etc. Run separate regressions for 1979 and for 1985 where you include these variables one by one in the equation instead of the computer variable. Include again occupation dummies, grades, and family background variables into all of these regressions. Prepare two tables for 1979 and 1985 each. How do the results compare to the computer effects? Do you think these coefficients are associated with the causal effect of a particular skill on earnings? How do these results make you feel about the interpretation of the computer coefficient? How are the non-computer tools affected by this?
12. Write up a conclusion based on your results. What have you learned about the causal effect of computers on wages?