Econometrics -2 PS6

Krishna Srinivasan

May 5, 2019

Discussed with Miriam.

Problem 6.1
The data is given in the Figure below, and also available on OLAT as an Excel spreadsheet, Berkeley.xls, for convenience.

In this application, the outcome variable is accept (yes/no) and the treatment is the gender of the applicant, i.e., male (yes/no).

a. Can gender be considered a treatment? Explain.

Yes, as there is some variation in the population.

b. Assuming full independence between treatment and potential outcomes, what is the average treatment effect of gender on admission?

The ATE is .1416.

c. Do you find the assumption of full independence plausible? Why, or why not?

No it is not plaussible as potential outcomes are not independent of gender. In particular, although poeple can't select into genders, two the genders may select into different fields of study that have different rates of acceptance.

d. A weaker assumption is conditional independence. Assume that treatment and potential outcomes are independent conditional on the field of study and compute the unconditional average treatment effect (ATE), as well as the treatment effect on the treated (ATT). These are a matching estimators.

The average treatment effect is -.04263. The ATT is -.0709

e. Why do your answers in b. and d. differ?

As matching gives different weights to different sectors while OLS gives equal weights to sectors. Women seem to have a higher chance of admission in certain fields, which is given a higher weight in the matching estimator.

f. Define five dummy-variables field# that are equal to 1 if the
field $x = B, \ldots, E$ and run the regression

$$accept = \beta_0 + \beta_1 male + \beta_2 fieldB + \cdots + \beta_6 fieldE + \epsilon \tag{1}$$

It is useful to start with a data set of $6 \times 2 \times 2 = 24$ observations. Each line represents a unique combination of accept, male, and
field, and the last variable is the number of observations, n, or duplicity, of this combination. In a next step, you can use the STATA expand command to generate n identical copies of each line. The final data set should have 4526 observation.

Alternatively, you can use the frequency fw option without expanding the data.

Paramter estimates can be found in Table 1.

g. Read chapter 2.3.2 of Angrist and Krueger (1999)[1]. Are your results from OLS and matching the same? If not, why?

In each cell, the matching estimator gives a weighted proportional to the number of obsrvations in each cell while the OLS gives a weight proportional to the variance in each cell.

6.2 Application: Instrument Variables

In this problem we estimate health returns to education using US census data. In particular, we address the issue of potential omitted variables that lead to upward biases in estimates of returns to education by using instrumental variables. Please read Angrist and Krueger (1991)[2], henceforth AK91, for their instrumental variables approach to solving the omitted variables problem in their estimation of wage returns to education.Please also read two papers that point out fundamental problems with AK91: Bound, Jaeger and Baker (1995)[3] , henceforth BJB95, describes issues with weak instruments and nite sample bias and Bound

[1]Angrist, Joshua D., and Alan B. Krueger. "Empirical Strategies in Labor Economics", in Orley C. Ashenfelter and David Card (Eds.): Handbook of Labor Economics, Chapter 23 Vol. 3, Part A, 1999, 1277-1366.

[2]Angrist, Joshua D., and Alan B. Krueger. "Does Compulsory School Attendance Affect Schooling and Earnings?" The Quarterly Journal of Economics , vol. 106, no. 4, 1991, pp. 979-1014.

[3]Bound, John, et al. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable Is Weak." Journal of the American Statistical Association , vol. 90, no. 430, 1995, pp. 443450

Table 1: Parameter estimates from 2b

	(1) accept
	b/se
male	-0.019
	(0.015)
1.major	0.000
	(.)
2.major	-0.010
	(0.023)
3.major	-0.303***
	(0.022)
4.major	-0.310***
	(0.022)
5.major	-0.402***
	(0.025)
6.major	-0.586***
	(0.023)
_cons	0.660***
	(0.020)
N	4527

and Jaeger (1996)[4], henceforth BJ96, discusses the problems related to the exclusion restriction. You will be using a sample from the 1980 US census data for this problem set (uploaded on Olat as datarest.dta). AK91 and BJB95 also used a sample from that census in their respective papers, but we are going to focus on the health returns to education, not on the monetary returns.

Consider a reduced form regression:

$$Health_i = \beta_1 + \beta_2 Educ_i + \beta_3 X_i + \mu_i$$

where X_i is a vector of individual chracteristics, and μ_i is a classical error term.

a. What are the assumptions (expressed in the standard structural notation, not the potential outcomes one) needed for β_2 to be consistently estimated? Describe why these assumptions are not likely to be satised in this regression using our sample from the census.

We need independence between $Educ_i$ and μ_i, or the assumption of exogeniety. This assumption is not likely to be satisfied as there are several ommitted variables, such as income, that lead to higher educational status and higher health outcomes.

Assume that the only endogenous variable in the previous model is education. Then, we can specify a two-equation model in the following form:

$$h = M\beta + \mu$$
$$M = Z\pi + v$$

where M is a matrix containing a variable for education and a vector X , and Z a matrix containing the available instruments for education and X (instruments for themselves).

b. Theoretically, if a valid instrument is found, can this model solve the problems mentioned in point a?

Yes.

c. What are the necessary assumptions for your instrumental variables strategy to yield consistent estimates for the health returns to education? Present the assumptions in our particular case both in the structural and in the potential outcomes framework.

[4]Bound, John and Jaeger, D.A. "On the validity of season of birth as an instrument in wage equations: A comment on Angrist and Kruegers Does compulsory school attendance affect schooling and earnings?", Working Paper No 5835, 1996, National Bureau of Economic Research, Cambridge.

- SUTVA:
  The treatment effect of an individual does not depend onthe effect for other individuals
- Relevance: The instrument is correlated with education
- exogeniety: The instrument is uncorrelated with ommitted factors that might affect educational outcomes
- Exclusion restriction: The only channel through which the instrument affects health outcomes is through education

d. Describe who the compliers, defiers, always takers and never takers are in AK91. What did you assume about them in point c?

Compliers are those born earlier (later) in the year but that have a lower (higher) level of education
Defiers are those born earlier (later) in the year but that have a higher (lower) level of education
Always takers are those have higher level of education regardless of quarter of birth.
Nevrer takers are those that have lower level of education regardless of quarter of birth.

We assume that there are no defiers.

e. AK91 uses Quarter of Birth as an instrument for education. They argue that birth quarter allows for an exogenous variation in the within-cohort year educational levels induced by age-based compulsory schooling laws. Discuss the arguments made by BJB95 and BJ96 concerning the validity of using quarter of birth as an instrument.

BJB95 and BJ96 claim the quarter of birth instrument is weak and thus raises questions about the validity of the relevance assumption. This also creates a small sample sample bias. They also claim that the exclusion restriction assumption may not be valid for two reasons. First, quarter of birth may have a direct effect on earnings. Second, quarter of birth may affect earnings through channels. This is evidenced by the fact that quarter of birth was correlated with earnings for earlier cohorts that preeceed the compulsory schooling law.

f. Theoretically, do you think that compulsory schooling laws (instead of quarter of birth) can be used directly as instruments? What assumption would the laws have to satisfy?

Compulsory schooling laws
If there were some states that had compulsory schooling laws and some that didn't, then the law could indeed be used directly as an instrument. The law would have to satisfy SUTVA. In particular, if it is not random as to which individuals (or conditionally random) recieve

the law then it may not be a good instrument as it may also violate exogeniety.

g. Describe some potential mechanisms through which education might affect health.

Individual's with higher education may be more aware of the benefits of healthcare and may be more likely to take care of their health. Another channel may that educated inidividuals may be less likley to follow certain religious practices that enforces an abstinance of modern medicine.

Empirical part
Dependent Variable : use the two health variables (disabwrk : reported disability that limits or prevents working, and disabtrn : reported disability that prevents the individual to use public transportation) to create a single dummy variable equal to one if the individual reports any type of disability. For the first stage, use education in completed years.
Covariates: SMSA (1=central city metro area), Married (=1 if married with spouse present) and 8 Regional dummies (exclude Pacic Division).

Sample : restrict your sample to make it more comparable with AK91. Keep only: men, born between 1930 and 1939, whites, born in one of the 50 states or in the District of Columbia (drop if bpl 90), and with no missing values for years of education and no missing values for at least one of the two disability variables.

h. Run OLS regressions:

- Health on Education with no covariates
- including the covariates mentioned above and also the variable you created for date of birth,
- including the covariates mentioned above and year of birth dummies

Present the results in the same table.

Paramter estimates can be found in Table 2

i. Write down your 2SLS model (one equation for the health variable and one for education), justifying your choice of variables. Would you also add income as a covariate, why or why not?

$$Health_i = \beta_0 + \beta_1 Educ_i + \beta_2 X_i + \mu_i$$
$$Educ_i = \gamma_0 + \delta birthqtr + \gamma_1 X_i + v_i$$

Table 2: OLS for question 2g and 2h

	(1) disab b/se	(2) disab *b/se*	(3) disab *b/se*	(4) disab *b/se*	(5) disab *b/se*	(6) disab *b/se*	(7) disab *b/se*
educrec	-0.027***	−0.026***	−0.025***	−0.073***	−0.070***	−0.068***	−0.049***
	(0.000)	(0.000)	(0.000)	(0.011)	(0.011)	(0.010)	(0.007)
SMSA		0.001	0.001		0.012***	0.011***	0.006***
		(0.001)	(0.001)		(0.003)	(0.003)	(0.002)
married		−0.092***	−0.093***		−0.077***	−0.078***	−0.084***
		(0.001)	(0.001)		(0.004)	(0.004)	(0.003)
_cons	0.282***	0.355***	0.374***	0.596***	0.680***	0.662***	0.546***
	(0.002)	(0.002)	(0.003)	(0.077)	(0.077)	(0.069)	(0.048)
birthyear dum	No	*No*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
region dum	No	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
state dumm	No	*No*	*No*	*No*	*No*	*No*	*Yes*
N	446241	446241	446241	446241	446241	446241	446241

Notes: (1) -(3) are OLS, (4)-(6) are the second stage from a 2sls

No we would not add income as a covariate since it is a bad control.birth quarter affects income and income affects health. Thus, you would violate the exclusion restriction assumption that the only channel through which birthquarter affects health is through education.

j. Run 2SLS regressions:

- With no covariates and using the rst three quarter of birth dummies as instruments,
- With covariates, including also year of birth dummies as covariates and using the quarter of birth dummies as instruments,
- With covariates, including also year of birth dummies as covariates and using the interaction between quarter of birth dummies and year of birth dummies as instruments,
- With covariates, including also year of birth dummies and state of birth dummies as covariates and using the interaction between quarter of birth dummies and state of birth dummies as instruments (for the state of birth dummies use Wyoming as omitted category and include Washington DC).

Add the results to the OLS table, compare your estimates for the health returns to education.

The estimates from the 2SLS are slightly larger in absolute value. This indicates that the traditional OLS was biased upward. This indeed makes sense as people with higher education have other characteristics that make them less likley to have a disability.

k. BJB95 discuss the issues that arise when confronted with both weak instruments and nite sample bias, in particular in the face of even small correlation between the instrument and the outcome variable. Discuss how these issues can confound your analysis. In your analysis, how strong is your rst stage regression; is nite sample bias an issue; is the exclusion restriction likely to be satised? (Tip for Stata: look at the testparm command).

From the four 2SLS, our F statistic is 67.04, 703.37, 316.81, 362.97, and 154.28. We can use the rule of thumb ($F > 10$) and claim that we indeed have a strong first-stage.

Although the relevance assumption holds since we have a strong first stage, exclusion restriction may not hold and hence there may be a finite sample bias.

We run a regression of diab on birthquarter dummies and educrec. An F-test shows that the birthquarter dummies are significiantly different from zero indiciting that the exclusion restriction is not likley ot hold. However, this is not conclusive since no correlation does not imply the lack of independence.

l. If your instrument is valid, would you expect your 2SLS estimate to higher or lower than estimates form OLS? You could base your answer on measurement error, omitted variables or both.

A positive correlation between the educrec and the error term would imply that the former is upward biased. Which is the case in our situation. Thus, the 2SLS would be lower that the estimate from OLS.

Similarly, the 2SLS would be higher than the OLS estimate if there is a negative correlation between the ommited variables and the variable of interest in the OLS regression.

m. Interpret your 2SLS estimate in the face of heterogeneous treatment effects. In particular, whose return does the instrument capture? Considering this, would you still expect the same relationship between the 2SLS and OLS estimates for health returns to education?

n. Is there a relationship between education and health? Are you convinced by your analysis? Why?

No. The variation in education is very low. It is not likely that just about year of education would drastically change health outcomes. Moreover, there may be other channels through which quarter of birth influences health.