

Question 1.

An important issue in the field of economics of education is to what extent performance differs between students in public and private schools. To investigate this issue, two American researchers present empirical results based on data for a sample of students in the US where some of the students attended public schools and some students attended catholic private schools. Their aim is to estimate the causal effect on student performance of attending catholic schools. Descriptive statistics for the variables in their sample is presented in Table 1 below. Dependent variable is math12, the student's math test score in high school. female is a dummy equal 1 if student is female, 0 otherwise. motheduc and fatheduc are years of schooling for father and mother, respectively. lincome is the natural logarithm of yearly income for the student's family. cathhs is a dummy equal to 1 if the student attended a catholic highschool, 0 if attended a public highschool. parcath is dummy variable taking the value 1 if one of the student's parents attended a catholic school. N is the number of observations. The sample average of math12 for students attending public and private schools is 51.9230 and 55.3854, respectively. The researchers present a number of regression results in Table 2 below.

- a) What is the interpretation of the estimated coefficients in col (1) in Table 2?
- b) Interpret the coefficients in column (2) in Table 2. What does the results tell you about the effect of catholic schools on student performance? Construct a 90 percent confidence interval for the effect of catholic schools.
- c) A commentator argues that the gain from attending catholic schools differs between boys and girls. Use the results in Table 2 to investigate this hypothesis
- d) The researchers are concerned that students attending catholic schools are a selected group of students. How would that affect the credibility of the results in column (1) and (2) in Table 2?
- e) Column (6) presents results using the instrumental variable method using parcath as an instrument. What is the argument for using the instrumental variable method in this case? Explain how the estimation is done and interpret the effect of attending a catholic high school using the IV method.
- f) What is your interpretation of the equations estimated in column (4) and (5) in Table 2? In particular, what is the interpretation of the estimated effect of parcath in column (4) and (5), respectively?
- g) Use the results in Table 2 to evaluate the validity of the instrumental variable method. Discuss possible reasons why the instrumental method used here might possibly fail to identify the causal effect of attending catholic schools.
- h) Another commentator asks you to illustrate the size of the effect of attending a catholic high school on student math performance in terms of math test score standard deviations. Provide this information to him using the information in the tables.
- i) The commentator also wonders if it is possible to test whether cathhs is endogeneous or not. Explain to him how you would conduct such a test.

Table 1. Descriptive statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
math12	7,430	52.13	9.459	29.50	71.37
female	7,430	0.517	0.500	0	1
motheduc	7,430	13.36	2.006	8	18
fatheduc	7,430	13.67	2.268	8	18
lfaminc	7,430	10.35	0.795	6.215	12.35
cathhs	7,430	0.0608	0.239	0	1
parcath	7,430	0.346	0.476	0	1

Table 2. Regression results. Estimated standard errors are in parenthesis

Estimation method	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) IV
VARIABLES	math12	math12	math12	math12	cathhs	math12
cathhs	3.4624 (0.4574)	1.4162 (0.4171)	1.4903 (0.5671)			3.9511 (1.4648)
lfaminc		1.8252 (0.1424)	1.8248 (0.1424)	1.8360 (0.1423)	0.0182 (0.0038)	1.7642 (0.1467)
motheduc		0.7120 (0.0619)	0.7119 (0.0620)	0.7246 (0.0621)	0.0039 (0.0017)	0.7092 (0.0621)
fatheduc		0.8898 (0.0560)	0.8899 (0.0560)	0.9002 (0.0560)	0.0066 (0.0015)	0.8743 (0.0568)
female		-1.1681 (0.1982)	-1.1584 (0.2044)	-1.1742 (0.1982)	-0.0091 (0.0053)	-1.1382 (0.1994)
Cathhs·female			-0.1602 (0.8305)			
parcath				0.5645 (0.2088)	0.1429 (0.0056)	
Constant	51.9230 (0.1128)	12.0779 (1.3175)	12.0771 (1.3176)	11.5500 (1.3135)	-0.3137 (0.0350)	12.7892 (1.3782)
Observations	7,430	7,430	7,430	7,430	7,430	7,430
R-squared	0.008	0.188	0.188	0.188	0.096	