

Inclass work: INSTRUMENTAL VARIABLES ESTIMATION

## 1 Question 1

- This question is based on Angrist and Evans (1998, American Economic Review).

$$hours = \beta_0 + \beta_1 kids + u \quad (1)$$

for the population of women with at least two children (so  $kids \geq 2$ ).

- (1) What are two conditions that IV must satisfy in eq. (1)?
- (2) What is outcome variable and what is policy variable?
- (3) What is the potential IV?
- (4) What are confounding factors? (Please provide the definition). Since the IV have to be randomly assigned between treat and control groups, it is not correlated with any other confounders except the treatment.

- (5) Draw a figure that describe relationship between outcome, policy variable (i.e., treatment), the IV, and confounding factors.

(6) Fill in the blank for the following statement.

The IV does not have causal effect on the ( BLANK ).

The IV does have causal effect on the ( BLANK ).

The IV is randomly assigned between ( BLANK ).

(7) The *same-sex* is the IV that is proposed by Angrist and Evans. On behalf of Angrist and Evans, please argue for the validity of the IV.

## 2 Question 2

**EXAMPLE:** Estimating the Return to Schooling Using CARD.DTA

This question is based on Card (1993, NBER).

$$\ln(wage) = \beta_0 + \beta_1 educ + u \quad (2)$$

where *educ* is years of education.

• A binary indicator, *nearc4<sub>i</sub>*, equal to one if the man was near a four-year college in high school can be used as an IV.

(1) What is outcome variable, what is treatment variable (or policy variable) and what is instrument?

(2) What is confounding factors here?

(3) What are two conditions for the IV?

(4) Please run the following regression.

(5) Draw a figure that describe relationship between outcome, policy variable (i.e., treatment), the IV, and confounding factors.

```
. use card
```

```
. sum educ nearc4
```

Variable	Obs	Mean	Std. Dev.	Min	Max
educ	3010	13.26346	2.676913	1	18
nearc4	3010	.6820598	.4657535	0	1

```
. reg educ nearc4, robust
```

Linear regression

Number of obs = 3010  
 F( 1, 3008) = 60.37  
 Prob > F = 0.0000  
 R-squared = 0.0208  
 Root MSE = 2.6494

		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
educ						
nearc4	.829019	.1066941	7.77	0.000	.6198182	1.03822
_cons	12.69801	.0902199	140.75	0.000	12.52112	12.87491

```
. * educ and nearc4 are strongly enough related: being near a 4-year college  
. * increases educ by almost a year. t statistic is pretty large.
```

```
. ivreg lwage (educ = nearc4), robust
```

Instrumental variables (2SLS) regression

Number of obs = 3010  
 F( 1, 3008) = 51.75  
 Prob > F = 0.0000  
 R-squared = .  
 Root MSE = .55686

---

			Robust				
	lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
	educ	.1880626	.0261426	7.19	0.000	.1368035	.2393217
	_cons	3.767472	.346742	10.87	0.000	3.087596	4.447347

---

Instrumented: educ

Instruments: nearc4

---

```
. * Note that the list of exogenous variables in the lwage equation is empty.
. * Estimated return to education seems too large. CI is wide, but lower
. * bound is still 13.7%.
```

. \* For comparison, OLS:

. reg lwage educ, robust

Linear regression

Number of obs = 3010  
 F( 1, 3008) = 321.16  
 Prob > F = 0.0000  
 R-squared = 0.0987  
 Root MSE = .42139

-----							
			Robust				
lwage		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----							
educ		.0520942	.0029069	17.92	0.000	.0463946	.0577939
_cons		5.570882	.0390935	142.50	0.000	5.49423	5.647535
-----							

. \* 18.8% versus 5.2%!



- Calculate  $\frac{1}{Corr(z,x)}$ . What is calculated blown-up factor?

```
. corr nearc4 educ
(obs=3010)
```

		nearc4	educ
-----+-----			
nearc4		1.0000	
educ		0.1442	1.0000

```
. corr nearc4 IQ
(obs=2061)
```

		nearc4	IQ
-----+-----			
nearc4		1.0000	
IQ		0.0765	1.0000

```
. corr educ IQ
(obs=2061)
```

		educ	IQ
-----+-----			
educ		1.0000	
IQ		0.5103	1.0000

```
. di .0765/.1442  
.53051318
```

- If we assume  $u = IQ$ , the “bias” terms are essentially the same: .51 for OLS and .53 for IV. So maybe there is measurement error in *educ* or we need to control for more factors.
- The IV standard error is .0261 compared with .0029 for OLS, or a factor of 9. The rough rule-of-thumb for the “blowing up” factor gives

$$\frac{1}{\widehat{Corr}(z, x)} \approx \frac{1}{.1442} \approx 7$$

- (6) There are a few ways to implement the IV method.

What are the two most popular ways to implement the IV method?

- (7) Now we will implement the IV method using 2SLS.

(Step 1)

Now,  $z_i$  must be *partially* correlated with  $x_i$ . Easiest to test with the regression

$$x_i \text{ on } z_i, \mathbf{r}_i$$

and reject the coefficient on  $z_i$  is equal to zero. This is called the **first-stage regression**. What are  $\mathbf{r}_i$ ?

- Card argues that, while *nearc4* is not uncorrelated with ability (*IQ*), it is after controlling for region of the U.S. (where the man lived at age 16). He also includes a race indicator, living in an SMSA (both currently and at age 16), and living in the south (currently). Experience is included as in the usual Mincer equation.

(Step 2) Implement the **first-stage regression**.

```
. reg educ nearc4 exper expersq black smsa south smsa66 reg662-reg669, robust
```

		Robust					
	educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
nearc4		.3198989	.0850763	3.76	0.000	.153085	.4867128
exper		-.4125334	.0320751	-12.86	0.000	-.4754249	-.3496418
expersq		.0008686	.0017076	0.51	0.611	-.0024795	.0042167
black		-.9355287	.0925281	-10.11	0.000	-1.116954	-.7541037
smsa		.4021825	.1112278	3.62	0.000	.1840918	.6202731
south		-.0516126	.1419604	-0.36	0.716	-.3299623	.2267371
smsa66		.0254805	.1106315	0.23	0.818	-.1914409	.2424019
reg662		-.0786363	.1858739	-0.42	0.672	-.4430898	.2858171
reg663		-.027939	.1793411	-0.16	0.876	-.3795833	.3237053
reg664		.117182	.2075839	0.56	0.572	-.2898395	.5242035
reg665		-.2726165	.2243154	-1.22	0.224	-.7124443	.1672114
reg666		-.3028147	.2367287	-1.28	0.201	-.766982	.1613526
reg667		-.2168177	.2394968	-0.91	0.365	-.6864128	.2527773
reg668		.5238914	.2568717	2.04	0.041	.0202284	1.027554
reg669		.210271	.1993703	1.05	0.292	-.1806456	.6011876
_cons		16.63825	.2153815	77.25	0.000	16.21594	17.06056

```
. * So nearc4 is still partially correlated with educ and can be used  
. * as an IV. (Partial correlation not as strong as simple correlation.)
```

```
. ivreg lwage exper expersq black smsa south smsa66 reg662-reg669
      (educ = nearc4), robust
```

Instrumental variables (2SLS) regression                      Number of obs =     3010

-----							
		Robust					
lwage		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----							
educ		.1315038	.0541436	2.43	0.015	.0253414	.2376663
exper		.1082711	.0234089	4.63	0.000	.062372	.1541702
expersq		-.0023349	.0003488	-6.69	0.000	-.0030188	-.0016511
black		-.1467757	.0525019	-2.80	0.005	-.2497193	-.0438322
smsa		.1118083	.0311448	3.59	0.000	.0507409	.1728757
south		-.1446715	.0291429	-4.96	0.000	-.2018136	-.0875294
smsa66		.0185311	.0205651	0.90	0.368	-.021792	.0588542
reg662		.1007678	.0365488	2.76	0.006	.0291045	.1724311
reg663		.1482588	.0355971	4.16	0.000	.0784615	.218056
reg664		.0498971	.0436162	1.14	0.253	-.0356238	.1354179
reg665		.1462719	.0492259	2.97	0.003	.049752	.2427919
reg666		.1629029	.0517655	3.15	0.002	.0614034	.2644025
reg667		.1345722	.0505568	2.66	0.008	.0354427	.2337017
reg668		-.083077	.0572432	-1.45	0.147	-.195317	.029163
reg669		.1078142	.0410761	2.62	0.009	.0272739	.1883545



_cons		3.666151	.91096	4.02	0.000	1.87998	5.452322
-------	--	----------	--------	------	-------	---------	----------

---

Instrumented: educ

Instruments: exper expersq black smsa south smsa66 reg662 reg663 reg664  
reg665 reg666 reg667 reg668 reg669 nearc4

---

```
. reg lwage educ exper expersq black smsa south smsa66 reg662-reg669, robust
```

Linear regression

Number of obs = 3010

-----							
		Robust					
lwage		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----							
educ		.0746933	.0036462	20.48	0.000	.0675439	.0818427
exper		.084832	.0067548	12.56	0.000	.0715875	.0980765
expersq		-.002287	.0003194	-7.16	0.000	-.0029133	-.0016608
black		-.1990123	.0181644	-10.96	0.000	-.2346282	-.1633964
smsa		.1363845	.0192172	7.10	0.000	.0987042	.1740648
south		-.147955	.0280346	-5.28	0.000	-.202924	-.092986
smsa66		.0262417	.0185908	1.41	0.158	-.0102102	.0626937
reg662		.0963672	.0350964	2.75	0.006	.0275518	.1651826
reg663		.14454	.0338217	4.27	0.000	.078224	.210856
reg664		.0550756	.041204	1.34	0.181	-.0257154	.1358665
reg665		.1280248	.042915	2.98	0.003	.0438789	.2121707
reg666		.1405174	.0451252	3.11	0.002	.0520378	.228997
reg667		.117981	.045614	2.59	0.010	.028543	.207419
reg668		-.0564361	.0505995	-1.12	0.265	-.1556494	.0427773
reg669		.1185698	.0387784	3.06	0.002	.0425347	.1946048
_cons		4.620807	.074229	62.25	0.000	4.475262	4.766352

---

. \* Discrepancy is smaller now, but still large: 13.2% for IV, 7.5% for OLS.