Lectures 08-09:
Paper overviews

PPHA 34600
Prof. Fiona Burlig

Harris School of Public Policy
University of Chicago

Lecture 08 recap

TL;DR:

1. Instrumental variables are very powerful
2. ...but they require extremely strong assumptions!
3. Hashtag no free lunch

An example: Health impacts of air pollution

Policy issue:

- Pollution is probably bad...
- ...but *how* bad, exactly?
- What role do airports play in pollution?

Approach:

- (We're not actually evaluating a program here)
- We need a shock to air pollution conditions

→ We don't have randomization, so we use IV

- Instrument of choice: flight delays on the East Coast

→ Do we believe this? Hold that thought...

Estimating treatment effects of pollution on health

How does pollution affect hospitalizations (simplified)?

First stage:

$$\text{Pollution in CA}_{it} = \alpha + \gamma \text{Taxi times}_{kt} + \beta X_{it} + \eta_{it}$$

where
Pollution in CA_{it} is pollution at CA airport i in time t
Taxi times_{kt} is the taxi time at non-CA airport k
X_{it} are controls
η_{it} is an error term

First stage (tabular form)

	CO Pollution			NO_2 Pollution			O_3 Pollution		
Variable	**(1a)**	**(1b)**	**(1c)**	**(2a)**	**(2b)**	**(2c)**	**(3a)**	**(3b)**	**(3c)**
Taxi Time	40.37***	56.16***	49.44***	0.51***	0.65***	0.76***	-0.07	0.04	-0.11
	(4.83)	(9.61)	(8.79)	(0.09)	(0.16)	(0.17)	(0.09)	(0.11)	(0.16)
Taxi x Distance		-2.23*	-1.82		-0.02	-0.03		-0.02*	0.01
		(1.23)	(1.13)		(0.02)	(0.02)		(0.01)	(0.02)
Taxi x Angle_u			15.28***			0.30			-0.43**
			(5.75)			(0.19)			(0.17)
Taxi x Angle_d			1.07			-0.02			0.12
			(5.38)			(0.13)			(0.09)
Taxi x Speed			-0.50			-0.06**			0.09**
			(1.27)			(0.03)			(0.04)
Taxi x Distance x Angle_u			-1.27			-0.02			0.05**
			(0.79)			(0.03)			(0.02)
Taxi x Distance x Angle_d			0.26			0.00			-0.01
			(0.66)			(0.02)			(0.01)
Taxi x Distance x Speed			0.19			0.00			-0.01*
			(0.15)			(0.00)			(0.01)
Taxi x Angle_d x Speed			1.03			0.04			-0.09*
			(1.65)			(0.03)			(0.05)
Taxi x Angle_u x Speed			-9.65***			-0.17***			0.23***
			(2.37)			(0.06)			(0.08)
Taxi x Dist. x Angle_u x Speed			1.29***			0.02**			-0.03***
			(0.32)			(0.01)			(0.01)
Taxi x Dist. x Angle_d x Speed			-0.34			-0.00			0.01
			(0.21)			(0.00)			(0.01)
Observations	179580	179580	179580	179580	179580	179580	179580	179580	179580
Zip Codes	164	164	164	164	164	164	164	164	164
Days	1095	1095	1095	1095	1095	1095	1095	1095	1095
F-stat(joint sig.)	69.29	38.23	12.39	33.13	16.85	6.00	0.65	2.11	1.13
p-value (joint sig.)	3.26e-14	2.43e-14	1.09e-17	4.17e-08	2.23e-07	1.25e-08	.4223	.1251	.3373

Estimating treatment effects of pollution on health

How does pollution affect hospitalizations (simplified)?

First stage:

$$\text{Pollution in CA}_{it} = \alpha + \gamma \text{Taxi times}_{kt} + \beta X_{it} + \eta_{it}$$

where
Pollution in CA_{it} is pollution at CA airport i in time t
Taxi times_{kt} is the taxi time at non-CA airport k
X_{it} are controls
η_{it} is an error term

Second stage:

$$Health_{it} = \alpha + \delta \widehat{pollution}_{it} + \tau X_{it} + \eta_{it}$$

where
$\widehat{pollution}_{it}$ is the fitted values from the first stage

Second stage (OLS)

	Asthma (1a)	Acute Respiratory (1b)	All Respiratory (1c)	All Heart (2)	Stroke (3)	Bone Fractures (4)	Appen-dicitis (5)
			Panel A: CO Pollution - All Ages				
No Controls	0.070***	0.265***	0.353***	0.035	-0.002	-0.022***	-0.001
	(0.017)	(0.041)	(0.053)	(0.028)	(0.006)	(0.007)	(0.001)
Time Controls	0.030	0.058	0.070	-0.022	-0.014*	-0.008	0.001
	(0.024)	(0.057)	(0.075)	(0.040)	(0.008)	(0.010)	(0.001)
Time + Weather	0.070**	0.071	0.097	0.004	-0.004	-0.010	-0.001
	(0.029)	(0.070)	(0.094)	(0.054)	(0.010)	(0.012)	(0.001)
Time + Weather + Zip Code FE	0.011	0.049***	0.078***	0.030***	-0.000	-0.006	0.002*
	(0.007)	(0.019)	(0.023)	(0.008)	(0.003)	(0.004)	(0.001)
			Panel B: NO_2 Pollution - All Ages				
No Controls	3.1***	10.7***	14.6***	4.3***	0.6***	-0.3	0.1**
	(0.5)	(1.3)	(1.7)	(1.1)	(0.2)	(0.2)	(0.0)
Time Controls	1.7**	6.0***	7.9***	1.0	-0.1	0.6*	0.1**
	(0.7)	(1.5)	(2.1)	(1.4)	(0.3)	(0.3)	(0.0)
Time + Weather	4.6***	9.0***	12.3***	3.2	0.8*	0.9*	0.0
	(1.1)	(2.7)	(3.8)	(2.5)	(0.5)	(0.5)	(0.1)
Time + Weather + Zip Code FE	0.1	1.1*	2.4***	1.1***	0.1	0.0	0.1**
	(0.2)	(0.6)	(0.8)	(0.3)	(0.1)	(0.2)	(0.0)

Second stage (IV)

	Asthma (1a)	Acute Respiratory (1b)	All Respiratory (1c)	Heart Problems (2)	Stroke (3)	Bone Fractures (4)	Appen-dicitis (5)
Panel A: All Ages							
Model 1: CO	0.341***	0.607***	0.828***	0.475***	0.059	-0.031	0.007
	(0.072)	(0.179)	(0.230)	(0.148)	(0.042)	(0.069)	(0.016)
Model 2: CO	0.330***	0.592***	0.812***	0.444***	0.048	-0.032	0.002
	(0.066)	(0.179)	(0.234)	(0.137)	(0.040)	(0.070)	(0.016)
Model 3: CO	0.203***	0.415***	0.534***	0.233***	0.020	-0.041	0.003
	(0.049)	(0.130)	(0.172)	(0.082)	(0.031)	(0.042)	(0.011)
Model 1: NO_2	29.2***	52.0**	70.9***	40.7***	5.1	-2.7	0.6
	(8.0)	(20.7)	(26.4)	(13.1)	(3.7)	(6.1)	(1.4)
Model 2: NO_2	28.7***	51.3**	70.3***	39.0***	4.4	-2.7	0.3
	(7.8)	(20.6)	(26.6)	(12.9)	(3.6)	(6.3)	(1.4)
Model 3: NO_2	11.9***	16.2	19.4	16.0**	0.6	-0.8	0.5
	(4.0)	(10.5)	(13.7)	(7.2)	(2.2)	(2.9)	(0.9)

Second stage (IV): Vulnerable populations

			Panel B: Ages Below 5				
Model 1: CO	0.606**	2.137*	2.956**	0.166*	0.019	0.047	-0.009
	(0.262)	(1.232)	(1.485)	(0.088)	(0.023)	(0.147)	(0.035)
Model 2: CO	0.621**	2.095*	2.846*	0.124	0.021	0.069	-0.019
	(0.252)	(1.202)	(1.476)	(0.082)	(0.025)	(0.141)	(0.038)
Model 3: CO	0.727***	2.300***	2.639***	0.076	0.023	-0.030	-0.009
	(0.173)	(0.800)	(0.990)	(0.058)	(0.015)	(0.126)	(0.023)
Model 1: NO_2	48.8*	172.0	237.9*	13.3*	1.5	3.8	-0.7
	(25.0)	(115.8)	(143.5)	(7.5)	(1.9)	(11.7)	(2.8)
Model 2: NO_2	50.0**	168.9	229.5	10.1	1.7	5.5	-1.5
	(24.2)	(113.0)	(142.3)	(7.1)	(2.1)	(11.1)	(3.0)
Model 3: NO_2	47.9***	116.9*	132.1*	4.6	2.8**	1.6	0.8
	(14.8)	(64.9)	(78.9)	(4.7)	(1.2)	(9.6)	(2.1)
			Panel C: Ages 65 and Older				
Model 1: CO	0.930***	1.620***	2.523***	3.888***	0.551*	0.478*	0.019
	(0.341)	(0.485)	(0.710)	(1.098)	(0.321)	(0.262)	(0.030)
Model 2: CO	0.864***	1.505***	2.423***	3.700***	0.503	0.417	0.017
	(0.298)	(0.451)	(0.695)	(1.035)	(0.326)	(0.260)	(0.030)
Model 3: CO	0.529**	0.734**	1.496***	2.011***	0.187	0.182	-0.031
	(0.213)	(0.326)	(0.545)	(0.642)	(0.259)	(0.169)	(0.028)
Model 1: NO_2	78.0***	135.9***	211.6***	326.1***	46.2	40.1*	1.6
	(26.8)	(41.9)	(65.5)	(93.2)	(28.5)	(21.4)	(2.6)
Model 2: NO_2	77.9***	135.6***	211.5***	326.0***	46.1	39.9*	1.6
	(26.8)	(42.0)	(65.7)	(93.4)	(28.5)	(21.4)	(2.6)
Model 3: NO_2	35.3**	35.4	66.2	122.8***	0.9	9.5	-1.3
	(14.4)	(24.3)	(41.7)	(47.7)	(16.1)	(12.1)	(1.8)

Estimating the reduced form

How does taxi time affect health (simplified)?

Reduced form:

$$\text{Health}_{it} = \alpha + \theta \text{Taxi time}_{it} + \pi X_{it} + \eta_i$$

Reduced form

	Asthma (1a)	Acute Respiratory (1b)	All Respiratory (1c)	All Heart (2)	Stroke (3)	Bone Fractures (4)	Appen-dicitis (5)
			Panel A: All Ages				
Taxi Time	14.03***	24.98***	34.07***	19.54***	2.44	-1.28	0.27
	(2.74)	(7.88)	(10.03)	(5.24)	(1.71)	(2.89)	(0.68)
			Panel B: Ages Below 5				
Taxi Time	24.27**	85.57	118.38*	6.63*	0.75	1.88	-0.35
	(11.31)	(52.12)	(63.47)	(3.49)	(0.95)	(5.83)	(1.39)
			Panel C: Age 65 and Above				
Taxi Time	37.51***	65.34***	101.73***	156.77***	22.22*	19.28*	0.78
	(11.45)	(16.46)	(25.31)	(36.96)	(12.99)	(9.89)	(1.22)
Observations	179580	179580	179580	179580	179580	179580	179580
Zip Codes	164	164	164	164	164	164	164
Days	1095	1095	1095	1095	1095	1095	1095

The exclusion restriction is the key to any IV

You should always ask:
What is the exclusion restriction in this analysis saying?

The exclusion restriction is the key to any IV

You should always ask:
What is the exclusion restriction in this analysis saying?

Do we believe this? Why or why not?

Second stage (IV)

	Asthma (1a)	Acute Respiratory (1b)	All Respiratory (1c)	Heart Problems (2)	Stroke (3)	Bone Fractures (4)	Appen-dicitis (5)
				Panel A: All Ages			
Model 1: CO	0.341***	0.607***	0.828***	0.475***	0.059	-0.031	0.007
	(0.072)	(0.179)	(0.230)	(0.148)	(0.042)	(0.069)	(0.016)
Model 2: CO	0.330***	0.592***	0.812***	0.444***	0.048	-0.032	0.002
	(0.066)	(0.179)	(0.234)	(0.137)	(0.040)	(0.070)	(0.016)
Model 3: CO	0.203***	0.415***	0.534***	0.233***	0.020	-0.041	0.003
	(0.049)	(0.130)	(0.172)	(0.082)	(0.031)	(0.042)	(0.011)
Model 1: NO_2	29.2***	52.0**	70.9***	40.7***	5.1	-2.7	0.6
	(8.0)	(20.7)	(26.4)	(13.1)	(3.7)	(6.1)	(1.4)
Model 2: NO_2	28.7***	51.3**	70.3***	39.0***	4.4	-2.7	0.3
	(7.8)	(20.6)	(26.6)	(12.9)	(3.6)	(6.3)	(1.4)
Model 3: NO_2	11.9***	16.2	19.4	16.0**	0.6	-0.8	0.5
	(4.0)	(10.5)	(13.7)	(7.2)	(2.2)	(2.9)	(0.9)

Lecture 09 recap

TL;DR:

1. Instrumental variables are very powerful
2. With the right assumptions...
3. ...we can handle OVB and ME (and simultaneity)

An example: Early-life rainfall and health

Policy issue:

- Early-life shocks may be very important
- With bad harvests, kids may not get the proper nutrition

Approach:

- (We're not actually evaluating a program here)
- We want to estimate the effect of rainfall on health
- **Measurement of rainfall is poor in Indonesia**
- Instrument of choice: rainfall at weather stations $j \neq i$

Estimating the effects of rainfall on health

The authors will run a (simplified) version of:

$$Y_i = \tau Rainfall_i + \varepsilon_i$$

Where:

Y_i is a health outcome of interest

$Rainfall_i$ is rain in location i

- (They'll actually do this in a series of lags)

ε_i is an error term

Estimating the effects of rainfall on health

The authors will run a (simplified) version of:

$$Y_i = \tau Rainfall_i + \varepsilon_i$$

Where:

Y_i is a health outcome of interest

$Rainfall_i$ is rain in location i

- (They'll actually do this in a series of lags)

ε_i is an error term

A big concern

- $Rainfall_i$ is measured with error
- We are likely to understate the true effect
- **Solution:** $Z_i = Rainfall\ Nearby_i$!

First stage estimates

Dependent variable: Rainfall in birthyear and birthdistrict (deviation of log rainfall in birth district from log of 1953-1999 district mean rainfall)

	Women	Men
Birthyear/birthdistrict rainfall, 2nd-closest station	0.138 (0.024)***	0.120 (0.023)***
Birthyear/birthdistrict rainfall, 3rd-closest station	0.144 (0.039)***	0.158 (0.035)***
Birthyear/birthdistrict rainfall, 4th-closest station	0.088 (0.053)	0.081 (0.044)*
Birthyear/birthdistrict rainfall, 5th-closest station	0.125 (0.025)***	0.158 (0.039)***
Number of observations	4,615	4,277
R-squared	0.59	0.59
F-statistic: Joint significance of all four rainfall variables	31.61	28.80
P-value	0.000	0.000

2SLS estimates

Table 2—Effect of Birth Year Rainfall on Adult Outcomes: Women and Men Born 1953–1974
(Instrumental variables estimates. Coefficients (standard errors) in regression of outcome on rainfall in individual's birth year and birth district. Instrumental variables for birth year/birth district rainfall are rainfall measured at second- through fifth-closest rainfall stations to respondent's birth district.)

	Women	Men
Self-reported health status very good (indicator)	0.101 (0.058)* [4,613]	−0.029 (0.072) [4,270]
Self-reported health status poor/very poor (indicator)	−0.192 (0.082)** [4,613]	−0.100 (0.098) [4,270]
Ln (lung capacity)	−0.044 (0.049) [4,454]	−0.073 (0.062) [3,907]
Height (centimeters)	2.832 (0.821)*** [4,495]	0.998 (1.795) [3,924]
Days absent due to illness (last four weeks)	−1.175 (0.831) [4,611]	0.515 (0.779) [4,267]
Completed grades of schooling	1.086 (0.453)** [4,598]	−0.474 (1.490) [4,259]
Ln (expenditures per capita in household)	0.095 (0.204) [4,615]	−0.274 (0.301) [4,277]
Asset index	0.876 (0.324)** [4,613]	−0.279 (0.507) [4,276]
Ln (annual earnings)	0.065 (0.988) [2,332]	−0.202 (0.350) [3,963]

2SLS estimates

TABLE 3—EFFECT OF RAINFALL IN YEARS BEFORE AND AFTER BIRTH: WOMEN BORN 1953–1974
(Instrumental variables estimates. Rainfall in individual's birth year and birth district instrumented with rainfall measured at second- through fifth-closest rainfall stations to respondent's birth district.)

Dependent variable	Self-reported health status very good (indicator)	Self-reported health status poor/very poor (indicator)	Height (centimeters)	Completed grades of schooling	Asset index
Coefficient on rainfall in:					
Year −3	0.025	−0.114	1.505	−0.065	0.003
	(0.084)	(0.120)	(1.572)	(0.992)	(0.424)
Year −2	−0.037	−0.013	0.854	−0.852	−0.426
	(0.103)	(0.075)	(1.813)	(1.670)	(0.721)
Year −1	−0.080	−0.045	3.338	0.104	−0.380
	(0.123)	(0.088)	(2.155)	(1.332)	(0.530)
Year 0	0.090	−0.179	3.833	1.598	0.750
	(0.067)	(0.093)*	(1.420)**	(0.675)**	(0.399)*
Year 1	−0.008	−0.096	0.676	1.083	0.203
	(0.053)	(0.067)	(1.592)	(0.769)	(0.272)
Year 2	−0.041	−0.015	1.666	0.117	−0.229
	(0.043)	(0.068)	(0.984)	(0.840)	(0.452)
Year 3	−0.020	−0.104	1.996	−0.135	0.088
	(0.116)	(0.067)	(1.774)	(0.802)	(0.232)
Observations	4,613	4,613	4,495	4,598	4,613