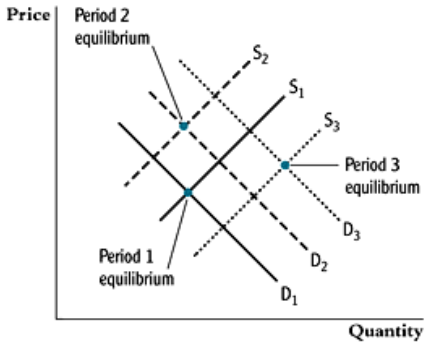
Instrumental Variables and Two Stage Least Squares

1. **Review: Instrumental Variables vs. Measurement Error Bias**
2. **Instrumental Variables vs. Endogeneity Bias**
3. **IV in general**
4. **IV vs. Heterogeneity Bias: Draft Lottery Example**
5. **Two Stage Least Squares**
6. **Weak Instruments**
7. **Overidentification**

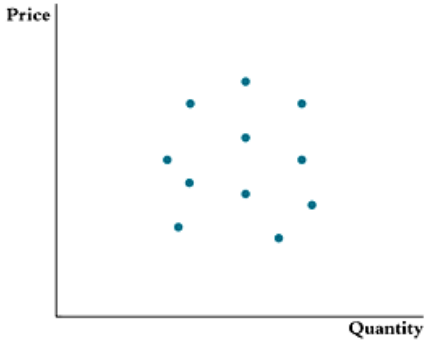
FIGURE 10.1

(a) Price and quantity are determined by the intersection of the supply and demand curves. The equilibrium in the first period is determined by the intersection of the demand curve D_1 and the supply curve S_1. Equilibrium in the second period is the intersection of D_2 and S_2, and equilibrium in the third period is the intersection of D_3 and S_3.



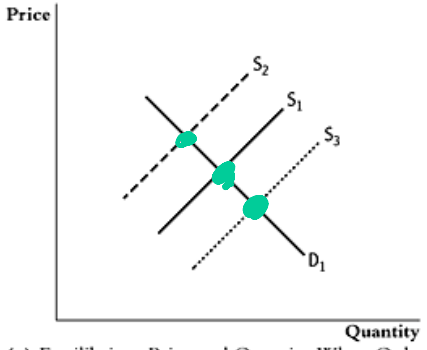
(a) Demand and Supply in Three Time Periods

(b) This scatterplot shows equilibrium price and quantity in eleven different time periods. The demand and supply curves are hidden. Can you determine the demand and supply curves from the points on the scatterplot?



(b) Equilibrium Price and Quantity for Eleven Time Periods

(c) When the supply curve shifts from S_1 to S_2 to S_3 but the demand curve remains at D_1, the equilibrium prices and quantities trace out the demand curve.



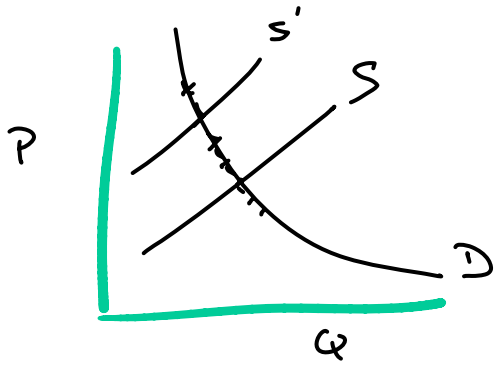
(c) Equilibrium Price and Quantity When Only the Supply Curve Shifts

The Endogeneity Problem:

When both supply and demand can shift the residual is correlated with price.

Solution: A supply shifter is needed to estimate the slope of a demand curve.

IV vs. Endogeneity Bias: Cigarette Tax Example



The Cigarette Consumption Panel Data Set

The data set consists of annual data for the 48 continental U.S. states from 1985–1995. Quantity consumed is measured by annual per capita cigarette sales in packs per fiscal year, as derived from state tax collection data. The price is the average retail cigarette price per pack during the fiscal year, including taxes. Income is per capita income. The general sales tax is the average tax, in cents per pack, due to the broad-based state sales tax applied to all consumption goods. The cigarette-specific tax is the tax applied to cigarettes only. All prices, income, and taxes used in the regressions in this chapter are deflated by the Consumer Price Index and thus are in constant (real) dollars. We are grateful to Professor Jonathan Gruber of MIT for providing us with these data.

Cigarette Tax Results

TABLE 10.1 Two Stage Least Squares Estimates of the Demand for Cigarettes Using Panel Data for 48 U.S. States

Dependent variable: $\ln(Q_{i,1995}^{cigarettes}) - \ln(Q_{i,1985}^{cigarettes})$

Regressor	(1)	(2)	(3)
$\ln(P_{i,1995}^{cigarettes}) - \ln(P_{i,1985}^{cigarettes})$	−0.94** (0.21)	−1.34** (0.23)	−1.20** (0.20)
$\ln(Inc_{i,1995}) - \ln(Inc_{i,1985})$	0.53 (0.34)	0.43 (0.30)	0.46 (0.31)
Intercept	0.21 (0.13)	0.45** (0.14)	0.37** (0.12)
Instrumental variable(s)	Sales tax	Cigarette-specific tax	Both sales tax and cigarette-specific tax
First-stage *F*-statistic	33.70	107.20	88.60
Overidentifying restrictions *J*-test and *p*-value	-	-	4.93 (0.026)

These regressions were estimated using data for 48 U.S. states (48 observations on the ten-year differences). The data are described in Appendix 10.1. The *J*-test of overidentifying restrictions is described in Key Concept 10.6 (its *p*-value is given in parentheses), and the first-stage *F*-statistic is described in Key Concept 10.5. Individual coefficients are statistically significant at the *5% level or **1% significance level.

3. IV in General

The General Instrumental Variables Regression Model and Terminology

The general IV regression model is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \beta_{k+1} W_{1i} + \cdots + \beta_{k+r} W_{ri} + u_i, \qquad (10.12)$$

$i = 1, \ldots, n$, where:

- Y_i is the dependent variable;
- u_i is the error term, which represents measurement error and/or omitted factors;
- $X_{1i}, \ldots, X_{ki}$ are k endogenous regressors, which are potentially correlated with u_i;
- $W_{1i}, \ldots, W_{ri}$ are r included exogenous regressors, which are uncorrelated with u_i;
- $\beta_0, \beta_1, \ldots, \beta_{k+r}$ are unknown regression coefficients;
- $Z_{1i}, \ldots, Z_{mi}$ are m instrumental variables.

The coefficients are overidentified if there are more instruments than endogenous regressors ($m > k$); they are underidentified if $m < k$; and they are exactly identified if $m = k$. Estimation of the IV regression model requires exact identification or overidentification.

Conditions for Valid Instruments

The Two Conditions for Valid Instruments

A set of m instruments $Z_{1i}, \ldots, Z_{mi}$ must satisfy the following two conditions to be valid:

1. Instrument Relevance

- *In general*, let $\hat{X}_{1i}^{*}$ be the predicted value of X_{1i} from the population regression of X_{1i} on the instruments (Z's) and the included exogenous regressors (W's), and let "1" denote a regressor that takes on the value "1" for all observations (its coefficient is the intercept). Then $(\hat{X}_{1i}^{*}, \ldots, \hat{X}_{ki}^{*}, W_{1i}, \ldots, W_{ri}, 1)$ are not perfectly multicollinear.
- *If there is only one* X, then at least one Z must enter the population regression of X on the Z's and the W's.

2. Instrument Exogeneity

The instruments are uncorrelated with the error term, that is, $\text{corr}(Z_{1i}, u_i) = 0, \ldots, \text{corr}(Z_{mi}, u_i) = 0$.

4. Draft Lottery Example

- How does military service affect earnings?

X_i { 1 veteran, 0 not ; Y_i - earnings after service.

$Y_i = \beta_0 + \beta_1^a X_i + \varepsilon_i$

- X not R.A.: 1) Some volunteer (self-selection) (−)
  2) Military selection (+)

β_1^c is the estimate you get if X R.A.

We need an I.V., Z,

A) Relevant $Cov(Z, X) \neq 0$

B) Valid $Cov(Z, \varepsilon) = 0$.

Z_i { 1 draft eligible, 0 not

You could estimate ACA effect:

(using I.V. as a R.H.S variable)

"Reduced Form":

$Y_i = \alpha_0 + \alpha_1 Z_i + u_i$

$$b^{IV} = \frac{\sum (z_i - \bar{z}) Y_i}{\sum (z_i - \bar{z}) X_i}$$

z - eligibility
x - service
y - earnings

$b^{IV} \xrightarrow{p} \beta^c$ because $Cov(z, \varepsilon) = 0$

$$= \frac{\sum (z_i - \bar{z}) Y_i \Big/ \sum (z_i - \bar{z})^2}{\sum (z_i - \bar{z}) X_i \Big/ \sum (z_i - \bar{z})^2} = \frac{b_{YZ} \approx \$500}{b_{XZ} \approx .16} = \frac{\text{"Reduced Form"}}{\text{"First Stage"}}$$

$\approx \$3000$

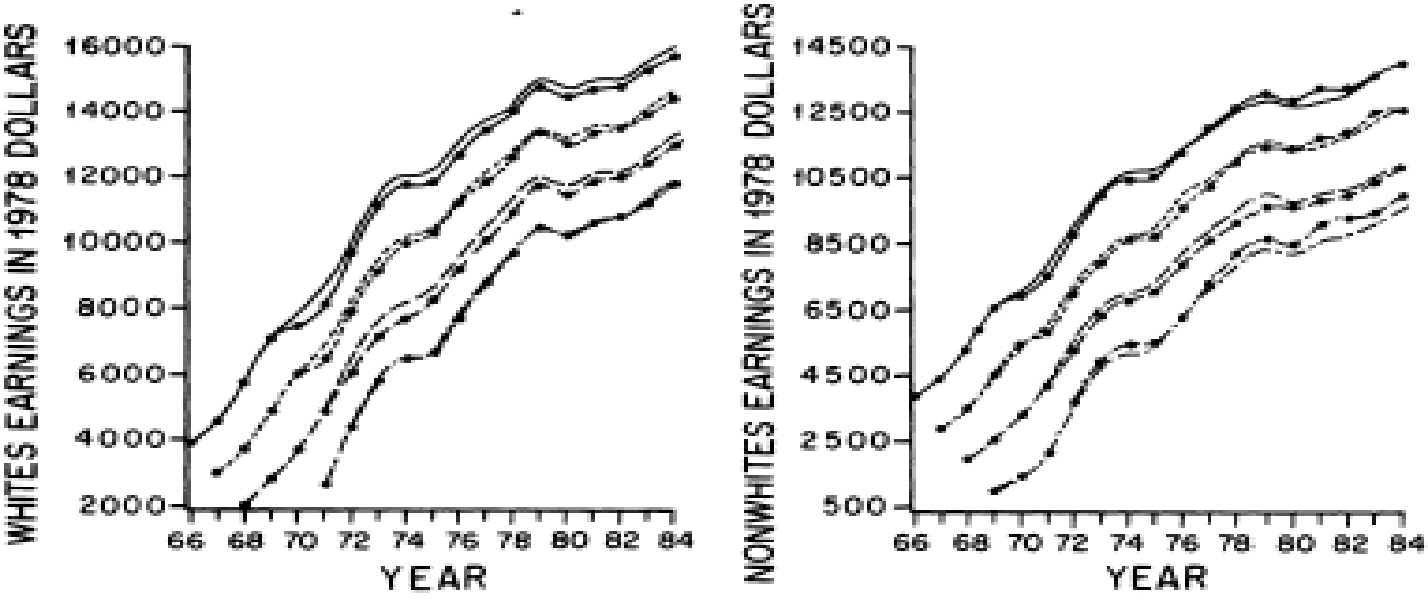
$b_{XZ} = \bar{X}_{Z=1} - \bar{X}_{Z=0}$

Draft Lottery Estimates

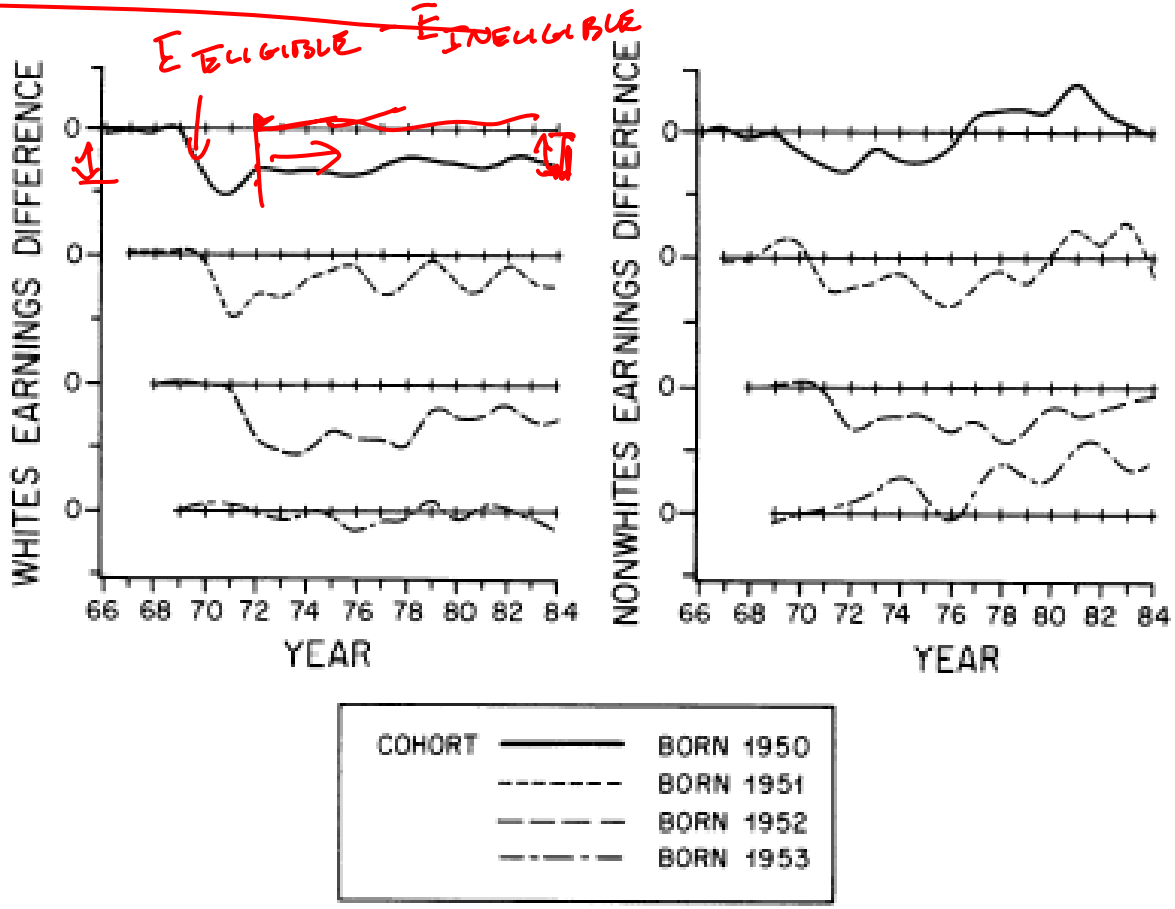
Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records

Joshua D. Angrist

The American Economic Review, Vol. 80, No. 3 (Jun., 1990), 313-336.



Differences in Earnings w/ trend removed [Reduced Form]



Notes: The figure plots the difference in FICA taxable earnings by draft-eligibility status for the four cohorts born 1950–53. Each tick on the vertical axis represents $500 real (1978) dollars.

FIGURE 2. THE DIFFERENCE IN EARNINGS BY DRAFT-ELIGIBILITY STATUS

Effect of Eligibility on Service [1st Stage]

TABLE 2—VETERAN STATUS AND DRAFT ELIGIBILITY

Data Set	Cohort	Sample	P(Veteran)	$\hat{p}^e$	$\hat{p}^n$	$\hat{p}^e - \hat{p}^n$
Whites						
SIPP (84)[a]	1950	351	0.2673	0.3527	0.1933	0.1594
			(0.0140)	(0.0325)	(0.0233)	(0.0400)
	1951	359	0.1973	0.2831	0.1468	0.1362
			(0.0127)	(0.0390)	(0.0180)	(0.0429)
	1952	336	0.1554	0.2310	0.1257	0.1053
			(0.0114)	(0.0473)	(0.0146)	(0.0495)
	1953	390	0.1298	0.1581	0.1153	0.0427
			(0.0106)	(0.0339)	(0.0152)	(0.0372)
DMDC/CWHS[b]	1950	16119	0.0633	0.0936	0.0279	0.0657
			(0.0019)	(0.0032)	(0.0019)	(0.0037)
	1951	16768	0.1176	0.2071	0.0708	0.1362
			(0.0025)	(0.0053)	(0.0024)	(0.0059)
	1952	17703	0.1515	0.2683	0.1102	0.1581
			(0.0027)	(0.0065)	(0.0027)	(0.0071)
	1953	17749	0.1343	0.1548	0.1268	0.0280
			(0.0026)	(0.0053)	(0.0029)	(0.0060)
Nonwhites						
SIPP (84)[a]	1950	70	0.1625	0.1957	0.1354	0.0603
			(0.0292)	(0.0699)	(0.0491)	(0.0854)
	1951	63	0.1703	0.2014	0.1514	0.0500
			(0.0292)	(0.0827)	(0.0448)	(0.0940)
	1952	52	0.1332	0.1449	0.1287	0.0161
			(0.0275)	(0.1040)	(0.0373)	(0.1105)
	1953	55	0.1749	0.1980	0.1612	0.0367
			(0.0305)	(0.0865)	(0.0470)	(0.0984)
DMDC/CWHS[b]	1950	5447	0.0417	0.0548	0.0271	0.0276
			(0.0027)	(0.0042)	(0.0032)	(0.0053)
	1951	5258	0.0794	0.1173	0.0599	0.0574
			(0.0037)	(0.0076)	(0.0040)	(0.0086)
	1952	5493	0.0953	0.1439	0.0794	0.0644
			(0.0040)	(0.0095)	(0.0042)	(0.0104)
	1953	5303	0.0925	0.0984	0.0904	0.0080
			(0.0040)	(0.0079)	(0.0046)	(0.0092)

Notes: Standard errors in parentheses. $\hat{p}^e$ is the probability of being a veteran conditional on being draft eligible; $\hat{p}^n$ is the probability of being a veteran conditional on being ineligible.

Effect of Service on Earnings [2nd Stage]

TABLE 3—WALD ESTIMATES

Cohort	Year	Draft-Eligibility Effects in Current $: FICA Earnings (1)	Adjusted FICA Earnings (2)	Total W-2 Earnings (3)	$\hat{p}^e - \hat{p}^n$ (4)	Service Effect in 1978 $ (5)
1950	1981	−435.8	−487.8	−589.6	0.159	−2,195.8
		(210.5)	(237.6)	(299.4)	(0.040)	(1,069.5)
	1982	−320.2	−396.1	−305.5		−1,678.3
		(235.8)	(281.7)	(345.4)		(1,193.6)
	1983	−349.5	−450.1	−512.9		−1,795.6
		(261.6)	(302.0)	(441.2)		(1,204.8)
	1984	−484.3	−638.7	−1,143.3		−2,517.7
		(286.8)	(336.5)	(492.2)		(1,326.5)
1951	1981	−358.3	−428.7	−71.6	0.136	−2,261.3
		(203.6)	(224.5)	(423.4)	(0.043)	(1,184.2)
	1982	−117.3	−278.5	−72.7		−1,386.6
		(229.1)	(264.1)	(372.1)		(1,312.1)
	1983	−314.0	−452.2	−896.5		-2,181.8
		(253.2)	(289.2)	(426.3)		(1,395.3)
	1984	−398.4	−573.3	−809.1		−2,647.9
		(279.2)	(331.1)	(380.9)		(1,529.2)
1952	1981	−342.8	−392.6	−440.5	0.105	−2,502.3
		(206.8)	(228.6)	(265.0)	(0.050)	(1,556.7)
	1982	−235.1	−255.2	−514.7		−1,626.5
		(232.3)	(264.5)	(296.5)		(1,685.8)
	1983	−437.7	−500.0	−915.7		−3,103.5
		(257.5)	(294.7)	(395.2)		(1,829.2)
	1984	−436.0	−560.0	−767.2		−3,323.8
		(281.9)	(330.1)	(376.0)		(1,959.3)

Notes: Standard errors in parentheses.
Columns (1) and (3) are taken from Table 1.
Column (2) reports draft-eligibility treatment effects on earnings adjusted for censoring at the FICA taxable maximum. The adjustment procedure is described in the Appendix. Column (4) reports SIPP estimates of the effect of draft eligibility on veteran status, taken from Table 2. Column (5) reports estimates of the effect of military service on civilian earnings is implied by columns (2) and (4).