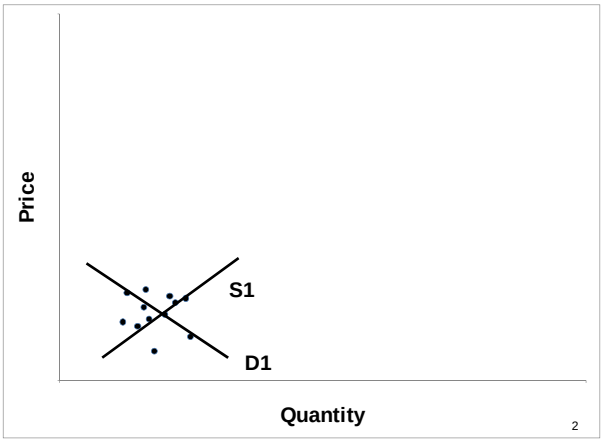
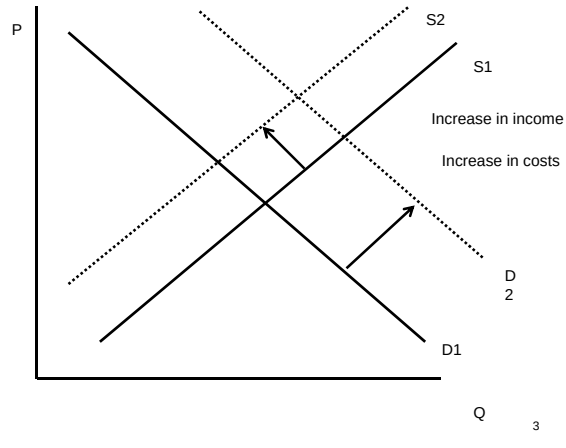
2SLS

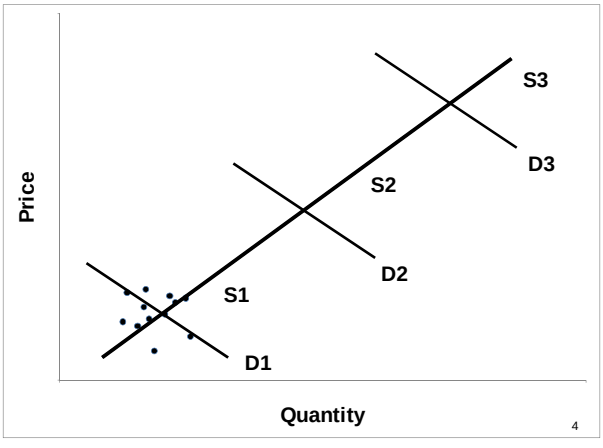
1


Price
S1
D1
Quantity

2


P
S2
S1
Increase in income
Increase in costs
D
2
D1
Q

3


Price
S3
D3
S2
D2
S1
D1
Quantity

4

Births in 2010

Means of Variables

Variable	Smoked during pregnancy	Did not smoke
Birthweight of child (grams)	3111	3281
Mom Unmarried	68.1%	37.8%
Mom a teen	12.5%	9.0%
Received < adequate pre. Care	16.3%	9.3%
Mom < HS degree*	25.0%	11.2%
Mom HS degree *	64.4%	31.2%

5

Births in 2010

% smoked during pregnancy

Variable	Answered Yes	Answered No
Mom Unmarried	11.6%	3.6%
Mom a teen	9.2%	6.6%
Received < adequate pre. Care	11.4%	6.3%
Mom < HS degree*	13.4%	5.6%
Mom HS degree *	12.8%	3.4%

6

Births in 2010

Average Birthweight

Variable	Answered Yes	Answered No
Smoked	3111	3281
Mom Unmarried	3196	3318
Mom a teen	3159	3127
Received < adequate pre. Care	3006	3298
Mom < HS degree*	3245	3286
Mom HS degree *	3261	3290

7

$Z_i=0$ $Z_i=1$

$\bar{x}_0 = 0.8$ $\bar{x}_1 = 0.57$

Table 2.—Smoking Characteristics of Control and Treatment Groups at Eighth Month of Pregnancy*

	Control Group	Treatment Group
% reporting smoking, cigarettes/day†		
N	395‡	388§
0	20.0	43.0
1-5	12.7	19.1
6-10	22.0	16.2
11-20	31.4	17.8
≥ 20	13.9	3.9
Mean ± SD‖	12.8 ± 11.5	6.4 ± 8.1
Thiocyanate, μmole/L		
N	389	380
Mean ± SD¶	2,452 ± 1,228	2,094 ± 1,209

8

$\bar{y}_0 = 3186$ $\bar{y}_1 = 3278$

Table 4.—Measurement of Status of Newborn*

	Control Group (N=438)		Treatment Group (N=429)		
	N	Mean ± SD	N	Mean ± SD	t
Primary factor					
Birth weight, g	438	3,186 ± 586	429	3,278 ± 627	2.28†
<2,500, %	438	6.9	429	6.8	...
<1,500, %	438	1.1	429	1.9	...
Other factors					
Birth length, cm	410	49.70 ± 4.16	406	50.30 ± 3.67	2.18†
Head circumference, cm	421	34.06 ± 2.44	413	34.14 ± 1.79	0.54
Apgar score (1 min)	430	8.01 ± 1.55	425	7.96 ± 1.65	0.49
<7, %	430	10.6	425	10.6	...
Apgar score (5 min)	431	9.00 ± 0.92	425	8.97 ± 1.09	0.45
<7, %	431	1.9	425	2.8	...
Gestational age, wk	438	39.09 ± 2.77	429	39.65 ± 2.77	0.22

*Includes only single, live births.
†P<.05 by Student's *t* test.

9

structural equation: $y = \beta_0 + x_i\beta_1 + \varepsilon_i$

first – stage: $x_i = \theta_0 + z_i\theta_1 + u_i$

reduced – form: $y_i = \pi_0 + z_i\pi_1 + v_i$

y = *birth weight in grams*

$x = 1$ *if a mom smoked*, 0 *otherwise*

$z = 1$ *if assigned to treatment*, 0 *otherwise*

10

$$\hat{\theta}_1 = \frac{\sum_i (z_i - \bar{z})(x_i - \bar{x})}{\sum_i (z_i - \bar{z})^2} = \bar{x}_1 - \bar{x}_0 = 0.57 - 0.80 = -0.23$$

$$\hat{\pi}_1 = \frac{\sum_i (z_i - \bar{z})(y_i - \bar{y})}{\sum_i (z_i - \bar{z})^2} = \bar{y}_1 - \bar{y}_0 = 3278 - 3186 = 92$$

$$\hat{\beta}_1 = \hat{\pi}_1 / \hat{\theta}_1 = 92 / -0.23 = 400$$

11

- bw_example.dta
- Three variables
  - treatment (1=yes, 0=no)
  - smoker (1=yes, 0=no)
  - birthweight (birth weight in grams)

12

```
. * get breakdown of sample
. tab treatment

    =1 if in |
   treatment |
   sample, 0 |
   otherwise |      Freq.     Percent        Cum.
-------------+-----------------------------------
           0 |        438       50.64       50.64
           1 |        427       49.36      100.00
-------------+-----------------------------------
       Total |        865      100.00
```

13

```
. ttest smoker, by(treatment)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     438    .7990868    .0191673    .4011415    .7614152    .8367583
       1 |     427    .5690867    .0239927    .4957849    .5219278    .6162455
---------+--------------------------------------------------------------------
combined |     865    .6855491    .0157957    .4645654    .6545467    .7165516
---------+--------------------------------------------------------------------
    diff |            .2300001    .0306274                .1698872     .290113
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   7.5096
Ho: diff = 0                                     degrees of freedom =      863

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

$$\bar{x}_1 - \bar{x}_0$$

14

```
. * get reduced form and 1st stage by ttest
. ttest birthweight, by(treatment)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     438        3186    28.00016         586    3130.968    3241.032
       1 |     427        3278    30.34266         627     3218.36     3337.64
---------+--------------------------------------------------------------------
combined |     865    3231.415    26.67189    607.9784    3190.842    3271.988
---------+--------------------------------------------------------------------
    diff |              -92    41.25236                -172.9667    -11.0333
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -2.2302
Ho: diff = 0                                     degrees of freedom =      863

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0130         Pr(|T| > |t|) = 0.0260          Pr(T > t) = 0.9870
```

$$\bar{y}_1 - \bar{y}_0$$

15

```
. * run the reduced form
. reg smoker treatment

      Source |       SS       df       MS              Number of obs =     865
-------------+------------------------------           F(  1,   863) =   56.39
       Model |  11.4377857     1  11.4377857           Prob > F      =  0.0000
    Residual |  175.031578   863  .202817588           R-squared     =  0.0613
-------------+------------------------------           Adj R-squared =  0.0603
       Total |  186.469364   864  .215821023           Root MSE      =  .45035

------------------------------------------------------------------------------
      smoker |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   treatment |  -.2300001   .0306274    -7.51   0.000     -.290113   -.1698872
       _cons |   .7990868   .0215107    37.13   0.000     .7568517    .8413218
------------------------------------------------------------------------------
```

16

```
. * run the reduced form
. reg birthweight treatment

      Source |       SS       df       MS              Number of obs =     865
-------------+------------------------------           F(  1,   863) =    4.97
       Model |     1830044      1     1830044           Prob > F      =  0.0260
    Residual |   317537006    863  367945.546           R-squared     =  0.0057
-------------+------------------------------           Adj R-squared =  0.0046
       Total |   319367050    864  369637.789           Root MSE      =  606.59

------------------------------------------------------------------------------
 birthweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   treatment |          92   41.25236     2.23   0.026      11.0333    172.9667
       _cons |        3106   28.98376   109.92   0.000     3129.113    3242.887
------------------------------------------------------------------------------
```

17

Instruments

Other exogenous control variables

```
ivregress 2sls y w (x=z)
```

Outcome of interest

Endogenous right hand side variables

18

```
.
. * run the 2sls
. ivregress 2sls birthweight (smoker=treatment)

Instrumental variables (2SLS) regression              Number of obs =     865
                                                      Wald chi2(1)  =    4.53
                                                      Prob > chi2   =  0.0332
                                                      R-squared     =       .
                                                      Root MSE      =  635.29

------------------------------------------------------------------------------
 birthweight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smoker |  -399.9998   187.8466    -2.13   0.033    -768.1724    -31.8272
       _cons |   3505.635   130.5771    26.85   0.000     3249.708    3761.561
------------------------------------------------------------------------------
Instrumented:  smoker
Instruments:   treatment
```

19

Two-stage least squares examples

20

Angrist: Vietnam Draft Lottery

21

Vietnam era service

- Defined as 1964-1975
- Estimated 8.7 million served during era
- 3.4 million were in SE Asia
- 2.6 million served in Vietnam
- 1.6 million saw combat
- 203K wounded in action, 153K hospitalized
- 58,000 deaths
- http://www.history.navy.mil/library/online/american%20war%20casualty.htm#t7

22

1980 Men, 1940-1952 Cohorts

Variable	Non-veterans	Veterans
In labor force	93.2%	95.9%
Unemployed	5.0%	4.7%
Labor earnings	$15,155	$15,875
Nonwhite	16.8%	12.3%
< HS degree	21.5%	8.8%
HS degree	49.4%	67.8%
College degree	28.9%	23.3%
Married	72.1%	75.5%

23

OLS Estimates
Impact of Vietnam Vet Status

Independent Variable	Labor Earnings	Unemployed
Age	510 (6.5)	-0.0021 (0.0001)
Non-white	-3446 (67)	0.029 (0.0014)
< high school	-9449 (74)	0.078 (0.0015)
High school	-4800 (57)	0.0032 (0.0011)
Vietnam era vet	523 (53)	-0.000 (0.0010)
Mean of Outcome	$15,372	4.9%
R^2	0.12	0.02

24

Vietnam Era Draft

- 1st part of war, operated liked WWII and Korean War
- At age 18 men report to local draft boards
- Could receive deferment for variety of reasons (kids, attending school)
- If available for service, pre-induction physical and tests
- Military needs determined those drafted

25

- Everyone drafted went to the Army
- Local draft boards filled army.
- Priorities
  - Delinquents, volunteers, non-vol. 19-25
  - For non-vol., determined by age
- College enrollment powerful way to avoid service
  - Men w. college degree 1/3 less likely to serve

26

Draft Lottery

- Proposed by Nixon
- Passed in Nov 1969, 1st lottery Dec 1, 1969
- 1st lottery for men age 19-26 on 1/1/70
  - Men born 1944-1950.
- Randomly assigned number 1-365, Draft Lottery number (DLN)
- Military estimates needs, sets threshold T
- If DLN<=T, drafted

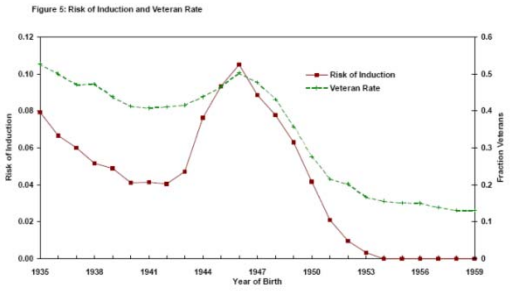
27

- If volunteer, could get better assignment
- Thresholds for service

Draft Year	Year of Birth	Threshold
• 1970	1946-50	195
• 1971	1951	125
• 1972	1952	95

- Draft suspended in 1973

28



29

Model

- Sample, men from 1950-1953 birth cohorts
- Y_i = earnings
- X_i = Vietnam military service (1=yes, 0=no)
- Z_i = draft eligible, that is DLN <=T
  - (1=yes, 0=no)

30

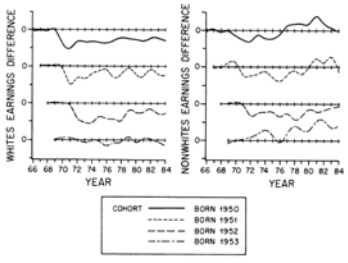
$\overline{x}_1$ $\overline{x}_0$

TABLE 2—VETERAN STATUS AND DRAFT ELIGIBILITY

			Whites			
Data Set	Cohort	Sample	P(Veteran)	$\hat{p}^e$	$\hat{p}^n$	$\hat{p}^e - \hat{p}^n$
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)

31

Graph of $\overline{y}_1 - \overline{y}_0$



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

32

$\overline{y}_1 - \overline{y}_0$ *in numbers*

TABLE 1—DRAFT-ELIGIBILITY TREATMENT EFFECTS FOR EARNINGS

	Whites							
	FICA Taxable Earnings				Total W-2 Compensation			
Year	1950	1951	1952	1953	1950	1951	1052	1953
76	−314.2	−86.8	−370.7	−145.5				
	(106.6)	(102.9)	(98.2)	(93.0)				
77	−262.6	−274.2	−396.9	−85.5				
	(117.9)	(112.2)	(111.1)	(107.1)				
78	−205.3	−203.8	−467.1	−65.3	1,059.3	233.2	175.3	−1,974.5
	(132.7)	(127.0)	(127.3)	(123.1)	(2,159.3)	(1,609.4)	(1,567.9)	(912.1)
79	−263.6	−60.5	−236.8	89.2	−1,588.7	523.6	−580.8	−557.9
	(160.5)	(152.3)	(153.9)	(148.7)	(1,575.6)	(1,590.5)	(736.7)	(750.1)
80	−339.1	−267.9	−312.1	−93.8	−1,028.1	85.6	−581.3	−428.7
	(183.2)	(175.3)	(178.2)	(170.7)	(756.8)	(599.8)	(309.1)	(341.5)
81	−435.8	−358.3	−342.8	34.3	−589.6	−71.6	−440.5	−109.5
	(210.5)	(203.6)	(206.8)	(199.0)	(299.4)	(423.4)	(265.0)	(245.2)
82	320.2	117.3	235.1	29.4	305.5	72.7	514.7	18.7
	(235.8)	(229.1)	(232.3)	(222.6)	(345.4)	(372.1)	(296.5)	(281.9)
83	−349.5	−314.0	−437.7	−96.3	−512.9	−896.5	−915.7	30.1
	(261.6)	(253.2)	(257.5)	(248.7)	(441.2)	(426.3)	(395.2)	(318.1)
84	−484.3	−398.4	−436.0	−228.6	−1,143.3	−809.1	−767.2	−164.2
	(286.8)	(279.2)	(281.9)	(272.2)	(492.2)	(380.9)	(376.0)	(366.0)

TABLE 3—WALD ESTIMATES

		Draft-Eligibility Effects in Current $				
Cohort	Year	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)

$\beta_1^{2sls} = (\overline{y}_1 - \overline{y}_0)/(\overline{x}_1 - \overline{x}_0)$ = -487.8/0.159 = $3067.9

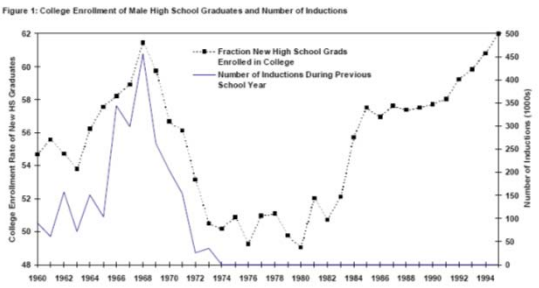
CPI_{78} = 65.2 CPI_{81}=90.9 65.2/90.9 = .7173

.717*3067.92 = $2199

34

- Although DLN is random, what are some ways that a low DLN could DIRECTLY change wages

35



36

Angrist and Evans: The impact of children on labor supply

37

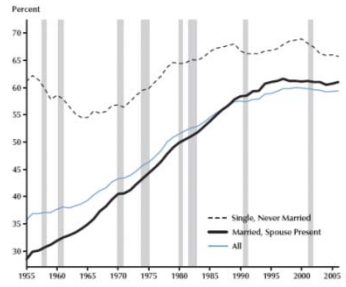
Introduction

- 2 key labor market trends in the past 40 years
  - Rising labor force participation of women
  - Falling fertility
- These two fact are intimately linked, but how?
  - Are women working more because they are having less children
  - Are women having less children because they are working more

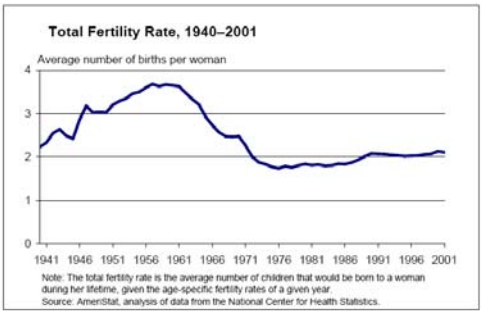
38

Figure 2

LFPR by Marital Status (Women)



39



40

TABLE 1—FERTILITY AND LABOR-SUPPLY MEASURES

Sample	1970 PUMS	1980 PUMS	1990 PUMS
Women aged 21–35			
Mean children ever born	1.78	1.27	1.18
Percent with 2 or more children	52.10	40.40	37.60
Percent worked last year	60.00	73.40	79.30
Observations	203,918	1,326,631	1,478,546

34% decline in children ever born -0.34=(1.18-1.78)/1.78

32% increase in the fraction of women that worked last year 0.32=(79.3-60)/60

- Note that between 1970 and 1990
  - Mean children ever born has fallen by 34%, from 1.78 to 1.18
  - % worked last year increased by 32%, from 60 to 79%
- Hundreds have studies have attempted to address these questions
- Lots of persistent relationships, but what have we measured?

- Women with children are not randomly assigned
- Who is most likely to have large families?
  - Lower educated
  - Those with lower wages
  - Certain minority groups
  - Certain religious groups
  - Those who want more children

- Problem is, many of these same groups are also those most likely to be out of the labor force
- Of the lower labor supply women among women with young children, how much is due to the kids, how much is attributable to some of these other factors?

To identify labor supply effects

- Need an instrument that
  - Alters fertility
  - Does not directly enter labor supply equation
- Ideas???

45

Table 1. Preference for sex of next child, according to current parity and sex composition of past births. (Number of women shown in parentheses.)

Parity	Sex preference for next child	Total (%)	Sex preference if past births were: All or most boys (%)	Equal number of boys and girls (%)	All or most girls (%)
2	Boy	48.8	15.0	50.9	84.2
	Girl	51.2	84.9	49.1	15.8
		(1505)	(392)	(777)	(336)
3	Boy	49.1	20.8		81.1
	Girl	50.9	81.2		18.8
		(1052)	(548)		(504)
4	Boy	50.7	18.4	50.9	84.2
	Girl	49.3	81.6	49.2	15.8
		(611)	(198)	(218)	(195)
5 +	Boy	50.3	26.6	55.2	77.8
	Girl	49.7	73.4	44.8	22.2
		(656)	(335)	(55)	(226)

46

TABLE 3—FRACTION OF FAMILIES THAT HAD ANOTHER CHILD BY PARITY AND SEX OF CHILDREN

Sex of first child in families with one or more children	All women: 1980 PUMS (649,887 observations) Fraction of sample	All women: 1980 PUMS (649,887 observations) Fraction that had another child	All women: 1990 PUMS (627,362 observations) Fraction of sample	All women: 1990 PUMS (627,362 observations) Fraction that had another child	Married women: 1980 PUMS (410,333 observations) Fraction of sample	Married women: 1980 PUMS (410,333 observations) Fraction that had another child	Married women: 1990 PUMS (477,798 observations) Fraction of sample	Married women: 1990 PUMS (477,798 observations) Fraction that had another child
(1) one girl	0.488	0.694 (0.001)	0.489	0.665 (0.001)	0.485	0.720 (0.001)	0.487	0.698 (0.001)
(2) one boy	0.512	0.694 (0.001)	0.511	0.667 (0.001)	0.515	0.720 (0.001)	0.513	0.699 (0.001)
difference (2) − (1)	—	0.000 (0.001)	—	0.002 (0.001)	—	0.000 (0.001)	—	0.001 (0.001)

47

Sex of first two children in families with two or more children	All women: 1980 PUMS (394,835 observations) Fraction of sample	All women: 1980 PUMS (394,835 observations) Fraction that had another child	All women: 1990 PUMS (380,007 observations) Fraction of sample	All women: 1990 PUMS (380,007 observations) Fraction that had another child	Married women: 1980 PUMS (254,654 observations) Fraction of sample	Married women: 1980 PUMS (254,654 observations) Fraction that had another child	Married women: 1990 PUMS (301,588 observations) Fraction of sample	Married women: 1990 PUMS (301,588 observations) Fraction that had another child
one boy, one girl	0.494	0.372 (0.001)	0.495	0.344 (0.001)	0.494	0.346 (0.001)	0.497	0.331 (0.001)
two girls	0.242	0.441 (0.002)	0.241	0.412 (0.002)	0.239	0.425 (0.002)	0.239	0.408 (0.002)
two boys	0.264	0.423 (0.002)	0.264	0.401 (0.002)	0.266	0.404 (0.002)	0.264	0.396 (0.002)
(1) one boy, one girl	0.494	0.372 (0.001)	0.495	0.344 (0.001)	0.494	0.346 (0.001)	0.497	0.331 (0.001)
(2) both same sex	0.506	0.432 (0.001)	0.505	0.407 (0.001)	0.506	0.414 (0.001)	0.503	0.401 (0.001)
difference (2) − (1)	—	0.060 (0.002)	—	0.063 (0.002)	—	0.068 (0.002)	—	0.070 (0.002)

48

Exactly identified model
With 1 instrument

TABLE 6—OLS ESTIMATES OF *MORE THAN 2 CHILDREN* EQUATIONS

Independent variable	All women (1)	All women (2)	All women (3)
1980 PUMS			
Boy 1st	—	−0.0080 (0.0015)	0.0001 (0.0021)
Boy 2nd	—	−0.0081 (0.0015)	—
Same sex	0.0600 (0.0016)	0.0617 (0.0015)	—
Two boys	—	—	0.0536 (0.0021)
Two girls	—	—	0.0698 (0.0021)
With other covariates	no	yes	yes
R^2	0.004	0.084	0.084

49

TABLE 7—OLS AND 2SLS ESTIMATES OF LABOR-SUPPLY MODELS USING 1980 CENSUS DATA

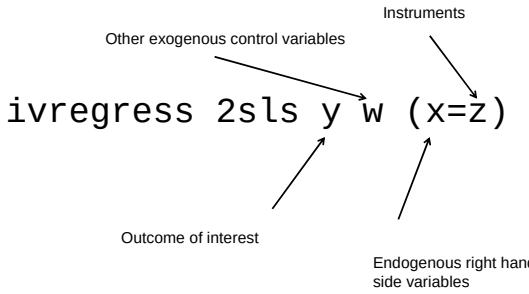
	All women (1)	All women (2)	All women (3)
Estimation method	OLS	2SLS	2SLS
Instrument for *More than 2 children*	—	*Same sex*	*Two boys, Two girls*
Dependent variable:			
Worked for pay	−0.176 (0.002)	−0.120 (0.025)	−0.113 (0.025) [0.013]
Weeks worked	−8.97 (0.07)	−5.66 (1.11)	−5.37 (1.10) [0.017]
Hours/week	−6.66 (0.06)	−4.59 (0.95)	−4.37 (0.94) [0.030]

50

```
. * in the data set;
. desc;

Contains data from pums80.dta
  obs:       254,654
 vars:            15                          17 Aug 2006 12:18
 size:     6,621,004 (73.3% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
kidcount        byte   %9.0g                  number of kids
morekids        byte   %9.0g                  =1 if mom had more than 2 kids
boy1st          byte   %9.0g                  =1 if 1st kid was a boy
boy2nd          byte   %9.0g                  =1 if 2nd kid was a boy
samesex         byte   %9.0g                  =1 if 1st two kids same sex
multi2nd        byte   %9.0g                  =1 if 2nd and 3rd kidss are
                                                twins
agem1           byte   %9.0g                  age of mom at census
agefstm         byte   %9.0g                  moms age when she 1st gave birth
black           byte   %9.0g                  =1 if mom is black
hispan          byte   %9.0g                  =1 if mom is hispanic
othrace         byte   %9.0g                  =1 if mom is othrace
workedm         byte   %9.0g                  did mom work for pay i 1979
weeksm1         byte   %9.0g                  moms weeks worked in 1979
hourswm         byte   %9.0g                  hours of work per week in 1979
incomem         float  %9.0g                  labor income per week, 1979,
                                                constant $
-------------------------------------------------------------------------------
```

51



52

```
. * get correlation coefficient between;
. * instrument and endogenous RHS variable;
. * correlation coefficient is 0.0695;
. corr morekids samesex;
(obs=254654)

             | morekids  samesex
-------------+------------------
    morekids |   1.0000
     samesex |   0.0695   1.0000

. * OLS of bivariate regression;
. * model assuming OLS model is correct;
. * specification;
. reg worked morekids;

      Source |       SS       df       MS              Number of obs =  2546
-------------+------------------------------           F(  1,254652) = 3237.
       Model |  796.712284     1  796.712284           Prob > F      =  0.00
    Residual |  62664.0083254652  .246077032           R-squared     =  0.01
-------------+------------------------------           Adj R-squared =  0.01
       Total |  63460.7206254653  .249204685           Root MSE      =  .496

------------------------------------------------------------------------------
      workedm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interva
-------------+----------------------------------------------------------------
    morekids |  -.1152029   .0020246   -56.90   0.000    -.1191712   -.11123
       _cons |   .5720607    .001249   458.02   0.000     .5696127    .57450
------------------------------------------------------------------------------
```

53

```
. * wald estimate;
. * using the notation from class, if we have y,x,z,w;
. * syntax for ivregress;
. * ivregress 2sls y w (x=z);
. * in this case, w=null,y=worked, x=morekids, z=samesex;
.  ivregress 2sls worked (morekids=samesex);

Instrumental variables (2SLS) regression              Number of obs =  254654
                                                      Wald chi2(1)  =   22.33
                                                      Prob > chi2   =  0.0000
                                                      R-squared     =  0.0121
                                                      Root MSE      =  .49618

------------------------------------------------------------------------------
      workedm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    morekids |  -.1376139   .0291242    -4.73   0.000    -.1946962   -.0805315
       _cons |   .5805895   .0111271    52.18   0.000     .5587807    .6023983
------------------------------------------------------------------------------
Instrumented:  morekids
Instruments:   samesex
```

$$Var(\hat{\beta}_1^{2SLS}) = Var(\hat{\beta}_1^{ols}) / \hat{\rho}(x,z)^2$$

$$= 0.0020246^2 / 0.0695^2 = 0.0291^2$$

$$Se(\hat{\beta}_1^{2SLS}) = 0.0291$$

54

Exactly Identified Model

```
. * demonstrate 1st stage and reduced form results for;
. * exactly identified model;
. * 1st stage;
. reg morekids samesex boy1st boy2nd agem1 agefstm black hispan othrace;

      Source |       SS       df       MS              Number of obs =  254654
-------------+------------------------------           F(  8,254645) = 2825.70
       Model |  4894.61525     8  611.826907           Prob > F      =  0.0000
    Residual |  55136.2215254645  .216521909           R-squared     =  0.0815
-------------+------------------------------           Adj R-squared =  0.0815
       Total |  60030.8368254653  .235735832           Root MSE      =  .46532

------------------------------------------------------------------------------
    morekids |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     samesex |   .0693854   .0018456    37.59   0.000      .065768    .0730028
      boy1st |  -.0111225   .0018456    -6.03   0.000    -.0147398   -.0075051
      boy2nd |  -.0095472   .0018456    -5.17   0.000    -.0131646   -.0059298
       agem1 |   .0304246    .000298   102.09   0.000     .0298405    .0310087
     agefstm |  -.0435676   .0003462  -125.85   0.000    -.0442461   -.0428891
       black |   .0679715   .0041853    16.24   0.000     .0597684    .0761747
      hispan |    .125998   .0038974    32.33   0.000     .1183591    .1336369
     othrace |   .0479479   .0044209    10.85   0.000      .039283    .0566127
       _cons |   .3234167   .0092616    34.92   0.000     .3052642    .3415692
------------------------------------------------------------------------------
```

55

```
. * there are 4 variables, y,x,w and z as we have defined them in class.
> * the syntax is ivregress 2sls y w (x=z);
. ivregress 2sls workedm boy1st boy2nd agem1 agefstm black hispan othrace
> (morekids=samesex);

Instrumental variables (2SLS) regression              Number of obs =  254654
                                                      Wald chi2(8)  = 6922.17
                                                      Prob > chi2   =  0.0000
                                                      R-squared     =  0.0482
                                                      Root MSE      =  .48703

------------------------------------------------------------------------------
      workedm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    morekids |  -.1203151   .0278407    -4.32   0.000    -.1748818   -.0657483
      boy1st |   .0009211   .0019489     0.47   0.636    -.0028986    .0047409
      boy2nd |  -.0048314   .0019425    -2.49   0.013    -.0086386   -.0010241
       agem1 |   .0219352   .0009013    24.34   0.000     .0201687    .0237018
     agefstm |  -.0264911   .0012647   -20.95   0.000    -.0289698   -.0240124
       black |   .1899764   .0047674    39.85   0.000     .1806325    .1993203
      hispan |  -.0139081   .0053812    -2.58   0.010    -.0244551   -.0033611
     othrace |   .0443545   .0048137     9.21   0.000     .0349198    .0537891
       _cons |   .4498966   .0138562    32.47   0.000     .4227389    .4770543
------------------------------------------------------------------------------
Instrumented:  morekids
Instruments:   boy1st boy2nd agem1 agefstm black hispan othrace samesex
```

56

```
. * there are 4 variables, y,x,w and z as we have defined them in class.
> * the syntax is ivregress 2sls y w (x=z);
. ivregress 2sls workedm boy1st boy2nd agem1 agefstm black hispan othrace
> (morekids=samesex);

Instrumental variables (2SLS) regression              Number of obs =  254654
                                                      Wald chi2(8)  = 6922.17
                                                      Prob > chi2   =  0.0000
                                                      R-squared     =  0.0482
                                                      Root MSE      =  .48703

------------------------------------------------------------------------------
     workedm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   morekids |  -.1203151   .0278407    -4.32   0.000    -.1748818   -.0657483
     boy1st |   .0009211   .0019489     0.47   0.636    -.0028986    .0047409
     boy2nd |  -.0048314   .0019425    -2.49   0.013    -.0086386   -.0010241
      agem1 |   .0219352   .0009013    24.34   0.000     .0201687    .0237018
   agefstm |  -.0264911   .0012647   -20.95   0.000    -.0289698   -.0240124
      black |   .1899764   .0047674    39.85   0.000     .1806325    .1993203
     hispan |  -.0139081   .0053812    -2.58   0.010    -.0244551   -.0033611
    othrace |   .0443545   .0048137     9.21   0.000     .0349198    .0537891
      _cons |   .4498966   .0138562    32.47   0.000     .4227389    .4770543
------------------------------------------------------------------------------
Instrumented:  morekids
Instruments:   boy1st boy2nd agem1 agefstm black hispan othrace samesex
```

57

```
. * reduced form;
. * look at the t-stat on the same sex variable and compare later on;
. * to the t-stat in the 2sls model;
. reg worked samesex boy1st boy2nd agem1 agefstm black hispan othrace;

      Source |       SS       df       MS              Number of obs =  254654
-------------+------------------------------           F(  8,254645) =  845.42
       Model |   1641.9059     8  205.238237            Prob > F      =  0.0000
    Residual |  61818.8147254645  .242764691           R-squared     =  0.0259
-------------+------------------------------           Adj R-squared =  0.0258
       Total |  63460.7206254653  .249204685           Root MSE      =  .49271

------------------------------------------------------------------------------
     workedm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    samesex |  -.0083481   .0019543    -4.27   0.000    -.0121785   -.0045178
     boy1st |   .0022593   .0019543     1.16   0.248    -.001571    .0060897
     boy2nd |  -.0036827   .0019543    -1.88   0.060    -.0075131    .0001477
      agem1 |   .0182747   .0003156    57.91   0.000     .0176562    .0188932
   agefstm |  -.0212493   .0003666   -57.97   0.000    -.0219677   -.0205308
      black |   .1817984   .0044317    41.02   0.000     .1731124    .1904845
     hispan |  -.0290676   .0041269    -7.04   0.000    -.0371561   -.020979
    othrace |   .0385856   .0046811     8.24   0.000     .0294107    .0477605
      _cons |   .4109847   .0098068    41.91   0.000     .3917636    .4302058
------------------------------------------------------------------------------
```

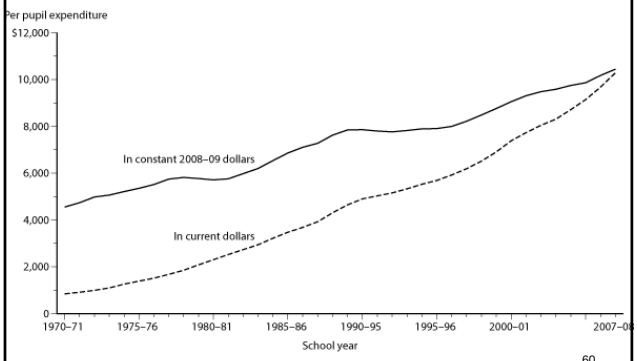
$$\hat{\beta}_1^{2SLS} = -0.0083481 / 0.0693854$$
$$= -0.1203$$

58

Angrist/Lavy

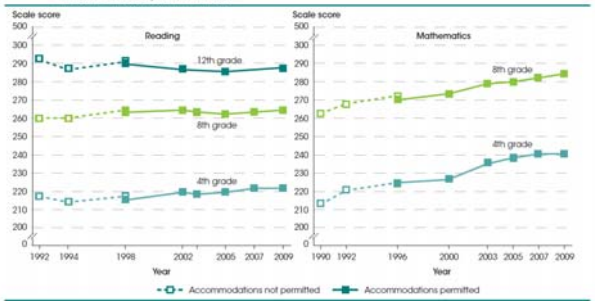
59

Figure 10. Current expenditure per pupil in fall enrollment in public elementary and secondary schools: 1970–71 through 2007–08



60

Figure 2. (Figures 10-1 and 12-1) Average reading and mathematics scale scores of 4th-, 8th-, and 12th-grade students: Selected years, 1990–2009



61

Figure A: High School Completion Rates, by Race and Cohort

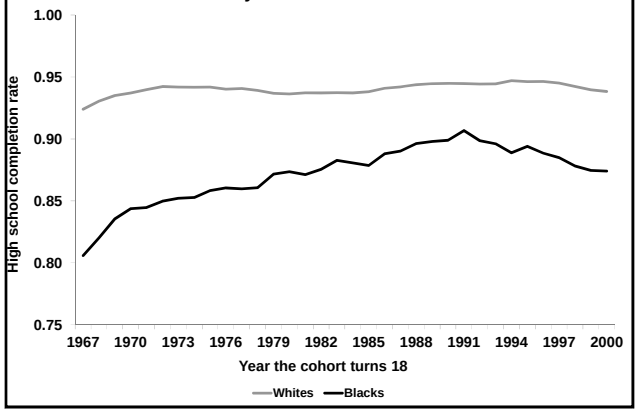
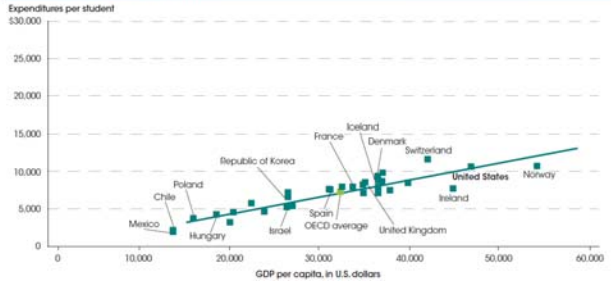


Figure 38-1. Annual expenditures per student for elementary and secondary education in selected OECD countries, by GDP per capita: 2007



63

Figure 15-1. Average scores of 15-year-old students on combined reading literacy scale, by country: 2009

OECD country and average score					
Korea, Republic of	539	Canada	524	Japan	520
Finland	536	New Zealand	521	Australia	515
Netherlands	508	Iceland	500	France	496
Belgium	506	**United States**	**500**	Denmark	495
Norway	503	Sweden	497	United Kingdom	494
Estonia	501	Germany	497	Hungary	494
Switzerland	501	Ireland	496	**OECD average**	**493**
Poland	500				
Portugal	489	Czech Republic	478	Austria	470
Italy	486	Slovak Republic	477	Turkey	464
Slovenia	483	Israel	474	Chile	449
Greece	483	Luxembourg	472	Mexico	425
Spain	481				
Non-OECD country or other education system and average score					
Shanghai-China	556	Hong Kong-China	533	Singapore	526
Liechtenstein	499	Chinese Taipei	495		
Macao-China	487	Romania	424	Argentina	398
Latvia	484	Thailand	421	Kazakhstan	390
Croatia	476	Trinidad and Tobago	416	Albania	385
Lithuania	468	Colombia	413	Qatar	372
Dubai-UAE	459	Brazil	412	Panama	371
Russian Federation	459	Montenegro, Republic of	408	Peru	370
Serbia, Republic of	442	Jordan	405	Azerbaijan	362
Bulgaria	429	Tunisia	404	Kyrgyz Republic	314
Uruguay	426	Indonesia	402		

Average is higher than the U.S. average | Average is not measurably different from the U.S. average | Average is lower than the U.S. average

64

Figure 16-1. Average scores of 15-year-old students on mathematics and science literacy scales, by country: 2009

Mathematics literacy scale					
OECD country and average score					
Korea, Republic of	546	New Zealand	519	Denmark	503
Finland	541	Belgium	515	Slovenia	501
Switzerland	534	Australia	514	Norway	498
Japan	529	Germany	513	France	497
Canada	527	Estonia	512	Slovak Republic	497
Netherlands	526	Iceland	507	**OECD average**	**496**
Austria	496	United Kingdom	492	Ireland	487
Poland	495	Hungary	490	Portugal	487
Sweden	494	Luxembourg	489	Spain	483
Czech Republic	493	**United States**	**487**	Italy	483
Greece	466	Turkey	445	Mexico	419
Israel	447	Chile	421		
Non-OECD country or other education system and average score					
Shanghai-China	600	Hong Kong-China	555	Liechtenstein	536
Singapore	562	Chinese Taipei	543	Macao-China	525
Latvia	482				
Lithuania	477	Uruguay	427	Colombia	381
Russian Federation	468	Thailand	419	Albania	377
Croatia	460	Trinidad and Tobago	414	Tunisia	371
Dubai-UAE	453	Kazakhstan	405	Indonesia	371
Serbia, Republic of	442	Montenegro, Republic of	403	Qatar	368
Azerbaijan	431	Argentina	388	Peru	365
Bulgaria	428	Jordan	387	Panama	360
Romania	427	Brazil	386	Kyrgyz Republic	331

65

TABLE 3
Percentage Distribution of Estimated Effect of Key Resources on Student Performance, Based on 377 Studies

Resources	Number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
Real classroom resources						
Teacher-pupil ratio	277	15%	13%	27%	25%	20%
Teacher education	171	9	5	33	27	26
Teacher experience	207	29	5	30	24	12
Financial aggregates						
Teacher salary	119	20%	7%	25%	20%	28%
Expenditure per pupil	163	27	7	34	19	13

Note: Source—Author's tabulations.

66

study. The great twelfth century Rabbinic scholar, Maimonides, interprets the Talmud's discussion of class size as follows: "Twenty-five children may be put in charge of one teacher. If the number in the class exceeds twenty-five but is not more than forty, he should have an assistant to help with the instruction. If there are more than forty, two teachers must be appointed" [Hyamson 1937, p. 58b].[1] Interestingly, while Maimonides' maximum of 40 students

67

- 1-40 students, one class
- 41-80 students, 2 classes
- 81 to 120 students, 3 classes

- Addition of one student can generate large changes in average class size

68

The class-size function derived from Maimonides' rule can be stated formally as follows. Let e_s denote beginning-of-the-year enrollment in school s in a given grade, and let f_{sc} denote the class size assigned to class c in school s, for that grade. Assuming that cohorts are divided into classes of equal size, we have

(1) $$f_{sc} = e_s/[\text{int}\,((e_s - 1)/40) + 1],$$

e_S= 80

f_{sc}= 80/[int((80-1)/40) +1]

= 80/[int(1.975) + 1]

= 80/[1+1] = 40

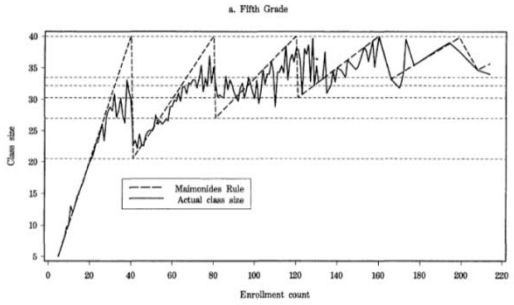
e_S= 81

f_{sc}= 81/[int((81-1)/40) +1]

= 81/[int(2) + 1]

= 81/[2+1] = 27

69

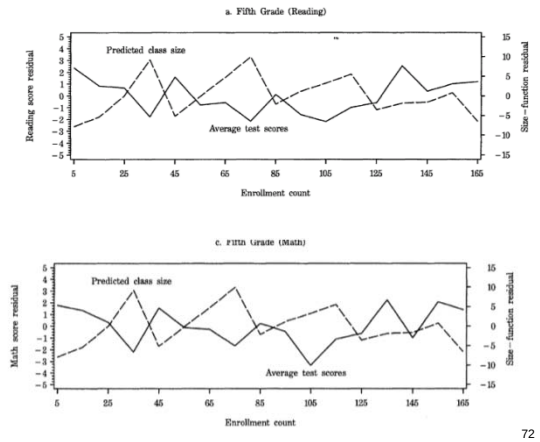


70

TABLE I
UNWEIGHTED DESCRIPTIVE STATISTICS

Variable	Mean	S.D.	Quantiles 0.10	0.25	0.50	0.75	0.90
A. Full sample 5th grade (2019 classes, 1002 schools, tested in 1991)							
Class size	29.9	6.5	21	26	31	35	38
Enrollment	77.7	38.8	31	50	72	100	128
Percent disadvantaged	14.1	13.5	2	4	10	20	35
Reading size	27.3	6.6	19	23	28	32	36
Math size	27.7	6.6	19	23	28	33	36
Average verbal	74.4	7.7	64.2	69.9	75.4	79.8	83.3
Average math	67.3	9.6	54.8	61.1	67.8	74.1	79.4

71



72

TABLE II
OLS Estimates for 1991

	5th Grade					
	Reading comprehension			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean score		74.3			67.3	
(s.d.)		(8.1)			(9.9)	
Regressors						
Class size	.221	−.031	−.025	.322	.076	.019
	(.031)	(.026)	(.031)	(.039)	(.036)	(.044)
Percent disadvantaged		−.350	−.351		−.340	−.332
		(.012)	(.013)		(.018)	(.018)
Enrollment			−.002			.017
			(.006)			(.009)
Root MSE	7.54	6.10	6.10	9.36	8.32	8.30
R^2	.036	.369	.369	.048	.249	.252
N		2,019			2,018	

73

TABLE III
Reduced-form Estimates for 1991

	5th Graders					
	Class size		Reading comprehension		Math	
	(1)	(2)	(3)	(4)	(5)	(6)
A. Full sample						
Means	29.9		74.4		67.3	
(s.d.)	(6.5)		(7.7)		(9.6)	
Regressors						
f_{sc}	.704	.542	−.111	−.149	−.009	−.124
	(.022)	(.027)	(.028)	(.035)	(.039)	(.049)
Percent disadvantaged	−.076	−.053	−.360	−.355	−.354	−.338
	(.010)	(.009)	(.012)	(.013)	(.017)	(.018)
Enrollment		.043		.010		.031
		(.005)		(.006)		(.009)
Root MSE	4.56	4.38	6.07	6.07	8.33	8.28
R^2	.516	.553	.375	.377	.247	.255
N	2,019		2,019		2,018	

IV estimates reading = -0.111/0.704 = -0.1576

IV estimates math = -0.009/0.704 = -0.01278

74

TABLE IV
2SLS Estimates for 1991 (Fifth Graders)

	Reading comprehension					
	Full sample				+/− 5 Discontinuity sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Mean score		74.4			74.5	
(s.d.)		(7.7)			(8.2)	
Regressors						
Class size	−.158	−.275	−.260	−.186	−.410	−.582
	(.040)	(.066)	(.081)	(.104)	(.113)	(.181)
Percent disadvantaged	−.372	−.369	−.369		−.477	−.461
	(.014)	(.014)	(.013)		(.037)	(.037)
Enrollment		.022	.012			.053
		(.009)	(.026)			(.028)
Enrollment squared/100			.005			
			(.011)			
Piecewise linear trend				.136		
				(.032)		
Root MSE	6.15	6.23	6.22	7.71	6.79	7.15
N		2019		1961	471	

75

TABLE IV
2SLS Estimates for 1991 (Fifth Graders)

	Math					
	Full sample				+/− 5 Discontinuity sample	
	(7)	(8)	(9)	(10)	(11)	(12)
Mean score		67.3			67.0	
(s.d.)		(9.6)			(10.2)	
Regressors						
Class size	−.013	−.230	−.261	−.202	−.185	−.443
	(.056)	(.092)	(.113)	(.131)	(.151)	(.236)
Percent disadvantaged	−.355	−.350	−.350		−.459	−.435
	(.019)	(.019)	(.019)		(.049)	(.049)
Enrollment		.041	.062			.079
		(.012)	(.037)			(.036)
Enrollment squared/100			−.010			
			(.016)			
Piecewise linear trend				.193		
				(.040)		
Root MSE	8.34	8.40	8.42	9.49	8.79	9.10
N		2018		1960	471	

76