Toulouse School of Economics

TSE 22nd February 2017

Applied Econometrics for Development: Instrumental Variables III

Ana GAZMURI

Paul SEABRIGHT

Paper for presentation (I)

- Kristin F. Butcher & Anne Case: « The Effect of Sibling Sex Composition on Women's Education and Earnings », *Quarterly Journal of Economics* 1992.

- Find that women's education in the US between 1920 and 1965 was systematically affected by sex composition of a woman's siblings – women raised with boys received more education than women raised with girls

- Sex composition of siblings is likely therefore to be a good instrument for education in earnings equations

- Instrumenting raises estimates of returns to women's education

Why should sibling sex composition matter?

- If returns to education differ for boys and girls, they will receive different education levels, but these will be independent of siblings
- Through budget constraints in the presence of limits on the family’s ability to borrow: if boys receive more education than girls, then girls raised with boys should have fewer resources for their own education
- Through within-family inequality aversion coupled with preferences for higher education for boys
- Through family socialization – girls may be more assertive if raised with brothers, or parents may group their aspirations

Data

- Panel Study of Income Dynamics (PSID)
- National Longitudinal Survey of Women (NLSW)
- Current Population Survey (CPS)

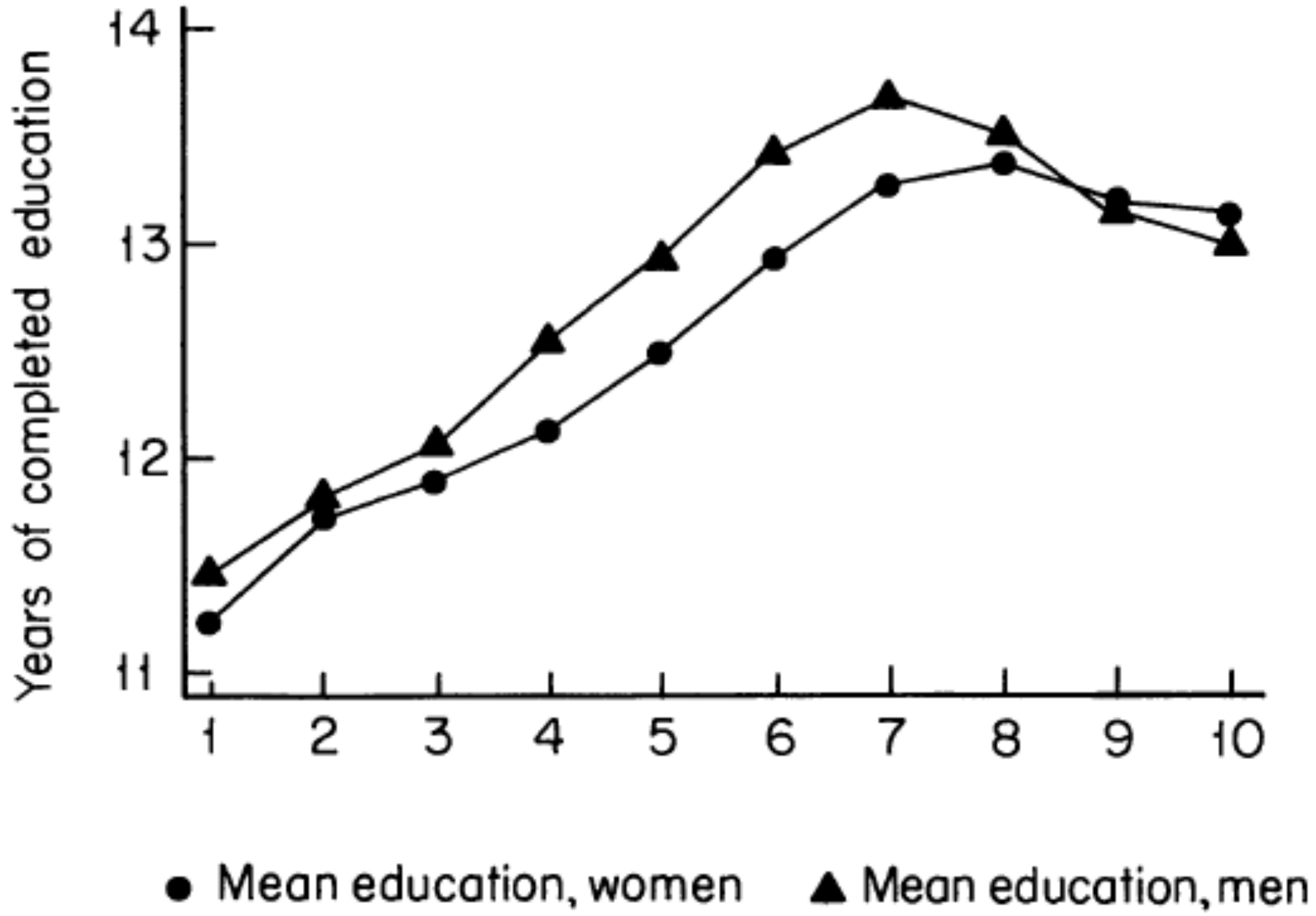

FIGURE I

Cohort 1 = birth year 1916–1920; cohort 2 = 1921–1925.
CPS November 1989, Mean Educational Attainment, Whites

TABLE III

YEARS OF COMPLETED SCHOOLING: PSID 1985 AND NLSW 1984[a]

(STANDARD ERRORS IN PARENTHESES)

	PSID: ages 24–44		PSID: ages 45–65	
	Men	Women	Men	Women
Number of siblings	−0.342	−0.280	−0.771	−0.230
	(0.084)	(0.082)	(0.150)	(0.099)
Number of siblings^2	0.016	0.017	0.045	0.004
	(0.008)	(0.008)	(0.013)	(0.008)
High school father[b]	0.619	0.410	0.325	0.729
	(0.156)	(0.152)	(0.346)	(0.257)
College father	1.580	1.325	1.102	1.371
	(0.170)	(0.187)	(0.376)	(0.316)

From family size to sibling sex composition

- The effects of sibling numbers suggest budget constraints may play an important role in educational decisions
- In principle we would expect women with sisters to have more education, controlling for total family size
- That is not what the estimates show
- Similarly, if the explanation were intra-household inequality aversion, we would expect sibling sex composition to impact boys' education as well; that is not what we find
- Other estimates not reported suggest that childhood socialization is not the explanation (older brothers have no greater impact than younger brothers)
- A reference group model is more consistent with these findings

TABLE V

SIBLING SEX COMPOSITION AND EDUCATIONAL ATTAINMENT

(STANDARD ERRORS IN PARENTHESES)

Explanatory variables:	PSID[a]							
	Men				Women			
Indicator variable, any sisters	0.052 (0.146)	—	—	0.244 (0.224)	−0.302 (0.134)	—	—	−0.278 (0.195)
Indicator variable, any brothers	—	0.094 (0.155)	—	—	—	0.227 (0.134)	—	—
Percentage of siblings female (including respondent)	—	—	−0.123 (0.251)	−0.433 (0.387)	—	—	−0.399 (0.220)	−0.057 (0.319)
Number of siblings	−0.507 (0.079)	−0.513 (0.082)	−0.491 (0.078)	−0.515 (0.079)	−0.186 (0.067)	−0.272 (0.066)	−0.258 (0.063)	−0.194 (0.076)
Number of siblings^2	0.027 (0.007)	0.028 (0.007)	0.026 (0.007)	0.028 (0.007)	0.006 (0.006)	0.011 (0.006)	0.010 (0.006)	0.006 (0.006)
Number of observations	1816	1816	1816	1816	2010	2010	2010	2010
R^2	0.2989	0.2990	0.2990	0.2994	0.2880	0.2870	0.2872	0.2880

TABLE VI

EDUCATIONAL TRANSITIONS FOR WOMEN, PSID 1985[a]

(STANDARD ERRORS IN PARENTHESES)

	Older women (45–65 years old)			Younger women (24–44 years old)		
Explanatory variables:	HS degree[b]	Attend college	College degree	HS degree	Attend college	College degree
Indicator variable, any sisters	−0.093 (0.042)	−0.048 (0.056)	−0.132 (0.079)	−0.001 (0.022)	−0.028 (0.037)	−0.080 (0.050)
Number of siblings	−0.011 (0.022)	−0.028 (0.029)	−0.059 (0.051)	−0.007 (0.016)	−0.041 (0.023)	−0.040 (0.030)
Number of siblings^2	−0.001 (0.002)	0.003 (0.003)	0.002 (0.005)	−0.001 (0.002)	0.003 (0.002)	0.003 (0.003)
Number of observations	762	528	231	1267	1058	651
R^2	0.1444	0.1503	0.1541	0.1596	0.1505	0.0740

Can sibling sex composition be a good instrument for labor market outcomes?

- It clearly has a significant explanatory effect on women's education

- To be a valid instrument it would have to be uncorrelated with labor market outcomes except via education

- Table VII suggests that it is not correlated with labor market participation

- This does not show that it is uncorrelated with factors (eg personality characteristics) that affect earnings, but it at least rules out some rather obvious channels of influence

TABLE VII

SIBLING COMPOSITION AND OTHER SOCIOECONOMIC OUTCOMES

WHITE WOMEN PSID 1985[a]

	Socioeconomic outcomes Older women (45–65 years old)			
	Currently working[b]	Currently married	Never married	Any children[c]
Mean of dependent var (standard deviation)	0.505 (0.500)	0.744 (0.437)	0.029 (0.168)	0.907 (0.291)
Explanatory variables:[a]				
Indicator variable, any sisters (standard error)	−0.015 (0.052)	0.029 (0.050)	−0.011 (0.024)	0.004 (0.038)
Number of observations	666	756	756	756
	Socioeconomic outcomes Younger women (24–44 years old)			
Mean of dependent var (standard deviation)	0.685 (0.465)	0.726 (0.446)	0.117 (0.322)	0.774 (0.418)
Explanatory variables:				
Indicator variable, any sisters (standard error)	0.063 (0.039)	−0.008 (0.035)	0.009 (0.026)	0.030 (0.030)
Number of observations	1085	1254	1254	1254

TABLE VIII
ORDINARY AND TWO-STAGE-LEAST-SQUARES ESTIMATES OF RETURNS TO EDUCATION WHITE WOMEN, PSID 1985
DEPENDENT VARIABLE: LOG HOURLY EARNINGS[b]
(STANDARD ERRORS IN PARENTHESES)

Explanatory variables:[a]	OLS	Reduced form	TSLS[c] (anysis)	TSLS[d] (anysis, numsib, numsib^2)
Years of education	0.091	—	0.184	0.182
	(0.007)		(0.113)	(0.055)
Any sisters	−0.033	−0.066	—	—
	(0.037)	(0.040)		
Number of siblings	−0.025	−0.039	−0.009	—
	(0.020)	(0.021)	(0.032)	
Number of siblings^2	0.002	0.003	0.001	—
	(0.002)	(0.002)	(0.002)	
Number of observations	1061	1061	1061	1061
R^2	0.2366	0.1168		

Why should instrumenting increase the estimated return to education?

- If there is a bias due to the effect of unobserved ability, this would normally bias returns upward (higher ability students obtain more education)
- Alternatively attenuation bias due to measurement error in the education variable may be responsible
- How big does the measurement error have to be to halve the coefficient? If it's the only regressor, measurement error would have to account for half the variance, which is implausible
- If other covariates are correlated with the true component of schooling they may absorb part of the signal
- Still, the effect of instrumenting is large – the authors suggest that ability bias may be attenuating rather than increasing the coefficient estimate

Paper for presentation (II)

- Christina Paxson: “Using Weather Variability to Estimate the response of Savings to Transitory Income in Thailand”, *American Economic Review* 1992.
- Directly estimates transitory income in order to compare savings responses to transitory versus permanent income.
- Finds that households save significantly more out of transitory than out of permanent income.
- Also contributes to improved methods of measuring savings in household surveys in developing countries.

The permanent income hypothesis

- Underlying idea very simple: concavity of utility function in consumption implies that any income that arrives in one period only will yield higher utility if the consumption is "smoothed" across periods

- Although concavity is rarely discussed in detail, it is not an incontrovertible assumption (why not?)

- Life-cycle hypothesis additionally implies that, modulo uncertainty about length of life, individuals will aim to consume all income over life time

- Again, not incontrovertible – what about bequests?

The permanent income hypothesis (II)

- Suppose savings are a linear function of permanent and transitory income as follows (consistent with CARA or quadratic utility and normally distributed income):

$$(1) \quad S_{irt} = \alpha_0 + Y_{irt}^{P}\alpha_1 + Y_{irt}^{T}\alpha_2 + VAR_{ir}\alpha_3 + W_{irt}\alpha_4 + \varepsilon_{irt}$$

- The permanent income hypothesis implies that

$$\alpha_1 \approx 0, \alpha_2 \approx 1$$

- But variance has an ambiguous effect (CARA utility implies it is positive, quadratic utility that it is zero).

Defining permanent and transitory income

- Estimating permanent income is very difficult, so transitory income is estimated directly through influence of known transitory variable (rainfall deviations from regional mean)

- Let permanent and transitory income be respectively

$$(2) \quad Y_{irt}^{P} = \beta_{t}^{P} + \beta_{0r} + X_{irt}^{P}\beta_{1} + \varepsilon_{irt}^{P}$$

$$(3) \quad Y_{irt}^{T} = \beta_{t}^{T} + X_{rt}^{T}\beta_{2} + \varepsilon_{irt}^{T}$$

- Yielding total income given by

$$(4) \quad Y_{irt} = \beta_{t} + \beta_{0r} + X_{irt}^{P}\beta_{1} + X_{rt}^{T}\beta_{2} + \varepsilon_{irt}$$

Defining permanent and transitory income (II)

- We can then write the structural and reduced form equations for savings as follows:

$$(5) \quad S_{irt} = \alpha_{0t} + \left[\beta_{0r} + X_{irt}^{P}\beta_1\right]\alpha_1$$

$$+\left[X_{rt}^{T}\beta_2\right]\alpha_2 + VAR_{ir}\alpha_3 + W_{irt}\alpha_4 + u_{irt}$$

$$(6) \quad S_{irt} = \gamma_t + \gamma_{0r} + X_{irt}^{P}\gamma_1 + X_{rt}^{T}\gamma_2 + v_{irt}$$

- Notice that the region fixed effects absorb the variance term, and the demographic variables W are absorbed in the X^T

Estimation issues

- To distinguish transitory rainfall from region-specific effects, need at least two periods of data
- Measurement error will be a big issue, both for income and for savings since savings are measured as the difference between income and consumption – thus the errors in income and savings will be correlated
- Also savings will tend to be biased downwards if inflation is not taken into account, since consumption measured more recently than income

Estimation issues (II)

- Two distinct strategies for estimating equation (5)
- Two-step procedure consists of estimating (4) and using the predicted estimates of permanent and transitory income as regressors in place of the X variables in (5)
- Residual from estimate of (4) is also included as regressor (called “unexplained income” – NOT the same as transitory income, because includes error terms from both equations). Its coefficient should therefore lie between the values of the two others
- Alternative estimation strategy is simultaneous estimation using maximum likelihood

TABLE 2—SAVINGS RATES BY INCOME QUARTILE, USING UNADJUSTED AND ADJUSTED SAVINGS MEASURES, 1981

	SAVE1/income			SAVE2/income		
Quartile	Unadjusted	Adjusted	Difference	Unadjusted	Adjusted	Difference
Income less than 1,658 baht	−0.69	−0.61	0.07	−0.51	−0.44	0.07
Income between 1,658 and 2,850 baht	−0.24	−0.18	0.06	−0.09	−0.04	0.05
Income between 2,850 and 4,915 baht	−0.10	−0.05	0.05	0.04	0.09	0.05
Income greater than 4,915 baht	0.08	0.12	0.04	0.21	0.24	0.03

Notes: SAVE1 is income minus expenditure on all goods; SAVE2 is income minus expenditure on nondurable goods.

Note that savings still seem implausibly low after adjustment – so some income under-reporting seems likely

Data

- Sample of rice farmers from central, north and northeastern Thailand in 1975/6, 1981 and 1986
- Data on regional rainfall 1951-1985; rainfall does not seem to be serially correlated
- Rainfall deviations are calculated from means for each agricultural season:
  - Season 1: off-season
  - Season 2: planting
  - Season 3: growing
  - Season 4: harvest

TABLE 3—REDUCED-FORM INCOME AND SAVINGS EQUATIONS

Variable	Income Estimate	Income t	SAVE1 Estimate	SAVE1 t	SAVE2 Estimate	SAVE2 t
Intercept	2,455.6	(16.30)	767.30	(2.88)	1,062.0	(4.03)
Year = 1981	301.68	(6.39)	44.774	(0.54)	37.450	(0.45)
Year = 1986	−402.26	(4.85)	−616.08	(4.20)	−725.18	(5.00)
Rainfall variables:						
$(R_1 - \bar{R}_1)$	1.9093	(2.52)	3.2338	(2.42)	2.9861	(2.26)
$(R_1 - \bar{R}_1)^2$	−0.0450	(3.99)	−0.0654	(3.28)	−0.0493	(2.50)
$(R_2 - \bar{R}_2)$	1.2502	(5.55)	1.2077	(3.03)	1.2888	(3.27)
$(R_2 - \bar{R}_2)^2$	0.0009	(0.66)	−0.0009	(0.40)	−0.0002	(0.09)
$(R_3 - \bar{R}_3)$	0.2282	(1.00)	−0.7973	(1.98)	−0.6963	(1.75)
$(R_3 - \bar{R}_3)^2$	0.0004	(0.62)	0.0008	(0.63)	0.0009	(0.72)
$(R_4 - \bar{R}_4)$	1.6097	(2.57)	0.5466	(0.49)	0.6314	(0.58)
$(R_4 - \bar{R}_4)^2$	−0.0095	(2.85)	−0.0090	(1.53)	−0.0087	(1.50)

TABLE 3—REDUCED-FORM INCOME AND SAVINGS EQUATIONS

Variable	Income		SAVE1		SAVE2	
	Estimate	t	Estimate	t	Estimate	t
Intercept	2,455.6	(16.30)	767.30	(2.88)	1,062.0	(4.03)
Year = 1981	301.68	(6.39)	44.774	(0.54)	37.450	(0.45)
Year = 1986	−402.26	(4.85)	−616.08	(4.20)	−725.18	(5.00)
Rainfall variables:						
$(R_1 - \bar{R}_1)$	1.9093	(2.52)	3.2338	(2.42)	2.9861	(2.26)
$(R_1 - \bar{R}_1)^2$	−0.0450	(3.99)	−0.0654	(3.28)	−0.0493	(2.50)
$(R_2 - \bar{R}_2)$	1.2502	(5.55)	1.2077	(3.03)	1.2888	(3.27)
$(R_2 - \bar{R}_2)^2$	0.0009	(0.66)	−0.0009	(0.40)	−0.0002	(0.09)
$(R_3 - \bar{R}_3)$	0.2282	(1.00)	−0.7973	(1.98)	−0.6963	(1.75)
$(R_3 - \bar{R}_3)^2$	0.0004	(0.62)	0.0008	(0.63)	0.0009	(0.72)
$(R_4 - \bar{R}_4)$	1.6097	(2.57)	0.5466	(0.49)	0.6314	(0.58)
$(R_4 - \bar{R}_4)^2$	−0.0095	(2.85)	−0.0090	(1.53)	−0.0087	(1.50)

Similar effects on income and saving

TABLE 4—TWO-STEP AND MAXIMUM-LIKELIHOOD ESTIMATES OF SAVINGS EQUATIONS

Variable	Two-step SAVE1	Two-step SAVE2	Two-step SAVE3	Maximum likelihood SAVE1	Maximum likelihood SAVE2	Maximum likelihood SAVE3
$\hat{Y}^{P}$ (α_1)	0.2773	0.4400	0.1824	0.2514	0.4210	0.1649
	(5.40)	(8.94)	(2.73)	(4.86)	(8.51)	(2.45)
$\hat{Y}^{T}$ (α_2)	0.7362	0.8039	0.7340	0.7546	0.8015	0.8294
	(4.28)	(4.87)	(3.21)	(4.32)	(4.84)	(3.50)
$\hat{\varepsilon}$	0.6015	0.6925	0.3801			
	(24.89)	(29.71)	(11.91)			

Implications

- Propensity to save out of transitory income are high (and hypothesis that they are equal to one cannot be rejected)
- Propensity to save out of permanent income are significantly lower, but still significantly positive
- Propensity to save out of unexplained income seem very sensitive to measurement error (since SAV3 – a measure of financial saving – is measured with larger error)
- Effect of rainfall variance on saving ambiguous